

12-2017

## The functions of SETD5 and miR-221 in embryonic stem cell differentiation

Tsai-Yu Chen

Follow this and additional works at: [https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations](https://digitalcommons.library.tmc.edu/utgsbs_dissertations)



Part of the [Medicine and Health Sciences Commons](#), [Molecular Biology Commons](#), and the [Other Cell and Developmental Biology Commons](#)

---

### Recommended Citation

Chen, Tsai-Yu, "The functions of SETD5 and miR-221 in embryonic stem cell differentiation" (2017). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 817.

[https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations/817](https://digitalcommons.library.tmc.edu/utgsbs_dissertations/817)

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact [digitalcommons@library.tmc.edu](mailto:digitalcommons@library.tmc.edu).

**THE FUNCTIONS OF SETD5 AND *MIR-221***  
**IN EMBRYONIC STEM CELL DIFFERENTIATION**

**Tsai-Yu Chen, M.S.**

**APPROVED:**

---

Min Gyu Lee, Ph.D.  
Advisory Professor

---

Xiaobing Shi, Ph.D.

---

Jae-Il Park, Ph.D.

---

Min Sup Song, Ph.D.

---

Liuqing Yang, Ph.D.

**APPROVED:**

---

Dean, The University of Texas  
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

**THE FUNCTIONS OF SETD5 AND *MIR-221***  
**IN EMBRYONIC STEM CELL DIFFERENTIATION**

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Tsai-Yu Chen, M.S.

Houston, Texas

December, 2017

## ACKNOWLEDGEMENT

I would like to first thank my mentor, Dr. Min Gyu Lee, for giving me the opportunity to pursue my degree in his laboratory. He brought me into the epigenetic field and really opened my eyes as a scientist. I always remembered the most important thing he told me, “Control! Control! Control! Always include controls in the experiment.” By following his guidance, I learned numerous new things, not only in the scientific field but also precious life lessons. I would like to express my sincere gratitude to all the faculty members that have served on my committees, Dr. Hui-Kuan Lin, Dr. Jessica Tyler, Dr. Xiaobing Shi, Dr. Jae-Il Park, Dr. Min Sup Song and Dr. Liuqing Yang. Your suggestions and guidance really provided me great directions and inspirations to finish my projects and my presentation skills.

I would like to thank all the former and current members in Dr. Lee’s laboratory. Dr. Na Li, Dr. Hunain Alam, Dr. Jaehwan Kim, Dr. Bingnan Gu, Dr. Eunah Kim and Sarah Wu. Thanks for everything during the past 7 years in the lab. It has been a privilege to work with you everyday in the lab. I’ve learned a lot from every one of you. Moreover, I want to especially thank Dr. Sunghun Lee and Dr. Shilpa Dhar. I’ve learned everything about my projects from you and you two provided me not only your knowledge but also the great patience to help me during the past several years. I could not finish my PhD without your help.

I would like to thank my parents, Mei-Yin Lee and Chu-I Chen for your unconditional love and support to allow me to pursue my dream in the USA. Last but not least, I would like to thank my lovely wife Tsun-Hsuan Chen. I am the luckiest person in the world to meet you and marry you. You bring me happiness every day and accompany me through all the difficulties. Thank you and I love you.

**THE FUNCTION OF SETD5 AND *MIR-221***  
**IN EMBRYONIC STEM CELL DIFFERENTIATION**

**Tsai-Yu Chen, M.S.**

**Advisory Professor: Min Gyu Lee, Ph.D.**

Embryonic stem cells (ESCs) are a widely used model system to study cellular differentiation because of their pluripotent characteristics, and ESC differentiation is an epigenetic process. In an effort to identify a new epigenetic factor that is required for ESC differentiation, the function of SETD5 in ESCs was studied for this thesis. Results show that SETD5 is essential for retinoic acid (RA)-induced differentiation of mouse ESCs and for RA-induced expression of critical developmental genes (e.g., *Hoxa1* and *Hoxa2*) and neuron-related genes (e.g., *Nestin* and *Pax6*). SETD5 was upregulated during ESC differentiation. Additional results demonstrated that SETD5 bound to RAR- $\alpha$  upon RA treatment and was recruited to a retinoic acid response element (RARE) for *Hoxa1* and *Hoxa2* activation. Methyltransferase assay using recombinant SETD5 and SETD5 complex showed that SETD5 was catalytically inactive, although it has a putative catalytic domain called SET. The transcription coactivator HCF1 was identified as a major SETD5-interacting protein. Depletion of HCF1 inhibited RA-induced ESC differentiation and RA-induced expression of SETD5-regulated genes. Chromatin immunoprecipitation assay provided evidence that HCF1 was localized to the SETD5-bound RARE region after RA treatment. These findings reveal a previously unknown ESC differentiation mechanism in which SETD5 facilitates mouse ESC differentiation by activating differentiation-specific genes via cooperation with HCF1.

In a related but independent study, we show that miR-221-3p and miR-221-5p, which are encoded by the *miR-221* gene, are new anti-stemness miRNAs whose expression levels in mouse ESCs are directly repressed by the epigenetic modifier and pluripotent factor PRMT7. Notably, both miR-221-3p and miR-221-5p can target the 3' untranslated regions of the major pluripotent factors Oct4, Nanog and Sox2 to antagonize mouse ESC stemness while miR-221-5p additionally silences expression of its transcriptional repressor PRMT7. Transfection of miR-221-3p and miR-221-5p mimics induced spontaneous differentiation of mouse ESCs. CRISPR-mediated *miR-221* deletion and anti-sense *miR-221* inhibitors inhibited spontaneous differentiation of PRMT7-depleted mouse ESCs. These results reveal that PRMT7-mediated repression of miR-221-3p and miR-221-5p is critical for maintaining mouse ESC stemness. Taken together, these two studies establish SETD5 and *miR-221* as novel anti-stemness regulators that play a pro-differentiation role in ESC differentiation.

# TABLE OF CONTENTS

Title Page .....	II
Acknowledgements .....	III
Abstract .....	IV
Table of Contents .....	VI
List of Figures .....	X
List of Tables .....	XII
CHAPTER 1 Introduction.....	1
1.1 General Background of embryonic stem cells .....	2
1.2 General Background of Epigenetics.....	3
1.3 Post-translational Modifications of Histones .....	4
1.4 Histone Methylation.....	5
1.5 SET Domain-containing Histone Methyltransferases and diseases .....	8
1.6 Histone Lysine Methylation as a Signature for Gene Regulation .....	10
CHAPTER 2 Function of SETD5 in mouse embryonic stem cell differentiation.....	13
2.1 Introduction .....	14
2.1.1 SET Domain-containing protein 5 (SETD5).....	14
2.1.2 Rationale and hypothesis .....	17
2.2 Materials and Methods .....	20
2.2.1 Antibodies, plasmids and other reagents .....	20
2.2.2 Mouse ESC culture and AP staining .....	21
2.2.3 Retinoic acid (RA)-induced differentiation of mouse ESCs .....	21
2.2.4 RNA interference .....	23
2.2.5 Western blot analysis.....	23

2.2.6 Quantitative PCR for mRNA expression.....	23
2.2.7 Rescue experiments by ectopic expression of SETD5.....	24
2.2.8 CRISPR/Cas9 gene editing.....	24
2.2.9 Chromatin Immunoprecipitation (ChIP) assay .....	25
2.2.10 Immunoprecipitation assay.....	25
2.2.11 Recombinant protein purification.....	26
2.2.12 Affinity purification of SETD5 complex.....	27
2.2.13 <i>In vitro</i> histone methyltransferase (HMT) assay .....	27
2.2.14 Radioactive S-adenosyl-L-methionine (SAM) binding assay .....	28
2.2.15 Statistical analysis.....	28
2.3 Results .....	29
2.3.1 SETD5 is required for proper mouse ESC differentiation. ....	29
2.3.2 SETD5 activates the expression of differentiation-specific genes and the neuron- related genes. ....	32
2.3.3 Generation of <i>Setd5</i> -2XHA knock-in mouse ESCs by CRISPR-Cas9 system.....	36
2.3.4 <i>Setd5</i> depletion also inhibits RA-induced differentiation of <i>Setd5</i> -2XHA knock-in mouse ESCs.....	39
2.3.5 <i>Setd5</i> -2XHA was recruited to RA Responsive Element (RARE) after RA treatment	42
2.3.6 SETD5 contains a SET domain but is catalytically inactive.....	45
2.3.7 SETD5 interacts with HCF1.....	49
2.3.8 Depletion of HCF1 inhibits RA-induced differentiation of mouse ESCs .....	52
2.3.9 HCF1 is recruited to <i>Hoxa1</i> RARE region upon RA treatment .....	55
2.4 Discussion .....	57
2.4.1 SETD5 and HCF1 in cellular differentiation .....	57



2.4.2 SETD5 is catalytically inactive but functional .....	58
2.4.3 SETD5 and neurodevelopmental disorders .....	59
2.4.4 The regulation of Hox genes by SETD5 .....	62
2.5 perspective and future directions.....	64
CHAPTER 3 PRMT7-mediated repression of <i>miR-221</i> Is Required for Maintaining Oct4,	
Nanog and Sox2 Levels in Mouse Embryonic Stem Cells .....	65
3.1 Introduction .....	66
3.1.1 General background of microRNA.....	66
3.1.2 MicroRNAs regulate ESC stemness and differentiation .....	67
3.1.3 General background of protein arginine methylation .....	68
3.1.4 Protein arginine methyltransferase .....	69
3.1.5 Rationale .....	69
3.2 Materials and Methods .....	71
3.2.1 Antibodies, plasmids, and other reagents .....	71
3.2.2 Quantitative PCR for miRNA and mRNA expression.....	71
3.2.3 Chromatin Immunoprecipitation (ChIP) assay .....	73
3.2.4 Transfections of miRNA mimics and LNA oligonucleotides.....	73
3.2.5 Luciferase reporter assays.....	73
3.2.6 CRISPR/Cas9 gene editing.....	74
3.2.7 Statistical analysis.....	74
3.3 Results .....	75
3.3.1 The expression of miR-221-3p and miR-221-5p is directly repressed by PRMT7 .....	75
3.3.2 <i>miR-221</i> has an anti-stemness function.....	78

3.3.3 <i>miR-221</i> can target the 3'UTRs of several pluripotent factors, including <i>Oct4</i> , <i>Nanog</i> , <i>Sox2</i> , and <i>PRMT7</i> .....	81
3.3.4 Both miR-221-3p and miR-221-5p can target 3'UTRs of <i>Oct4</i> , <i>Nanog</i> , <i>Sox2</i> , and <i>Klf4</i> while miR-221-5p can also silence <i>Prmt7</i> . ....	84
3.3.5 The repression of miR-221-3p and miR-221-5p expression is required for maintaining mouse ESC stemness.....	87
3.3.6 CRISPR-mediated deletion of the miR-221 gene impedes spontaneous differentiation of PRMT7-depleted mouse ESCs.....	90
3.3.7 <i>miR-221</i> loss blocks spontaneous differentiation of PRMT7-depleted mouse ESCs..	93
3.4 Discussion .....	96
3.4.1 The anti-stemness function of <i>miR-221</i> .....	96
3.4.2 PRMT7 negatively regulates the expression of <i>miR-221</i> .....	97
3.4.3 Summary .....	98
3.5 perspective and future directions.....	100
References .....	101
VITA .....	126

# LIST OF FIGURES

Figure 1 Lysine and arginine methylation .....	7
Figure 2. Six major histone lysine methylation sites .....	12
Figure 3. SETD5 is a putative histone methyltransferase which has been linked to 3p25 microdeletion syndrome.....	16
Figure 4. The expression level of 36 epigenetic regulators in different stage of RA-induced mouse ESC differentiation .....	19
Figure 5. Retinoic acid (RA)-induced Differentiation of Mouse ESCs .....	22
Figure 6. <i>Setd5</i> is required for RA-induced differentiation of mouse ESCs. ....	31
Figure 7. <i>Setd5</i> is necessary for RA-induced expression of differentiation-specific genes ( <i>Hoxa1</i> and <i>Hoxa2</i> ) and neuron-relevant genes ( <i>Nestin</i> and <i>Pax6</i> ).....	35
Figure 8. Generation of <i>Setd5</i> -2XHA knock-in mouse ESC by CRISPR-Cas9 system.....	38
Figure 9. SETD5 depletion also inhibits RA-induced differentiation of <i>Setd5</i> -2XHA knock-in mouse ESCs .....	41
Figure 10. <i>Setd5</i> -2XHA is recruited to RA responsive element (RARE) near <i>Hoxa1</i> after RA treatment .....	44
Figure 11. SETD5 does not have a detectable catalytic activity. ....	48
Figure 12. SETD5 interacts with HCF1 .....	51
Figure 13. Depletion of HCF1 inhibits mouse ESC differentiation.....	54
Figure 14. HCF1 is recruited by SETD5 to HOXA1 RARE region upon RA treatment .....	56
Figure 15. De-novo loss-of-function mutations in SETD5.....	61
Figure 16. A hypothetical model for SETD5-mediated gene regulation .....	63
Figure 17. PRMT7 represses the expression of the miR-221 gene by establishing repressive	

arginine methylation. ....	77
Figure 18. <i>miR-221</i> mimics induces mouse ESC differentiation. ....	80
Figure 19. <i>miR-221</i> can target 3'UTRs of the pluripotent factors <i>Oct4</i> , <i>Nanog</i> , <i>Sox2</i> , <i>Klf4</i> and <i>Prmt7</i> . ....	83
Figure 20. Both miR-221-3p and miR-221-5p can target 3'UTRs of <i>Oct4</i> , <i>Nanog</i> , <i>Sox2</i> , and <i>Klf4</i> while miR-221-5p can also silence <i>Prmt7</i> . ....	86
Figure 21. PRMT7-mediated repression of miR-221-3p and miR-221-5p levels is indispensable for maintaining mouse ESC stemness. ....	89
Figure 22. Generation of <i>miR-221</i> -null V6.5 mouse ESCs by a CRISPR/Cas9 strategy. ....	92
Figure 23. <i>miR-221</i> loss blocks spontaneous differentiation of PRMT7-depleted mouse ESCs. ....	95

# LIST OF TABLES

Table 1 SETD5 Oligonucleotide sequences.....	20
Table 2 <i>miR-221</i> Oligonucleotide sequences.....	72

# **CHAPTER 1**

## **INTRODUCTION**

## 1.1 GENERAL BACKGROUND OF EMBRYONIC STEM CELLS

In 1954, Stevens *et al.* reported that the spontaneous testicular teratomas from an inbred strain of mice contained nervous tissues, cartilage, bone, muscle, fat and undifferentiated embryonic cells (1). Dr. Stevens further described that the undifferentiated embryonic cells are pluripotent and “appear to give rise to both rapidly differentiating cells and others which, like themselves, remain undifferentiated.” This was the first report of the pluripotent cells, which were later known as embryonic carcinoma cells due to the cells’ origin from carcinoma cells.

The term of embryonic stem cells (ESCs) was coined by Dr. Gail Martin. In 1981, she and another group led by Dr. Martin Evans were first to derive mouse ESCs from mouse embryos. (2,3) These cells are pluripotent and can differentiate into a wide variety of cell types. Therefore, she gave the term of ESCs to denote their origin directly from embryos and to distinguish them from embryonic carcinoma cells. Nowadays, mouse or human ESCs can be derived from inner cell masses of blastocysts (4) and are defined by two main characteristics: 1) capability of long-term self-renewal and 2) ability to form all three germ layers and to differentiate into all kinds of different cell types (pluripotency). (5) The stemness of ESCs is maintained by pluripotent transcription factors, including Oct4, Nanog, Sox2, and their regulatory network (6-11). These factors co-occupy and activate numerous genes that are important for maintaining ESC pluripotency (e.g., *STAT3*, *Klf4* and *c-Myc*) while inhibiting lineage-specific genes to prevent ESC differentiation (e.g., *Hox* clusters, *Pax6* and *Meis1*) (12,13). Oct4, Nanog and Sox2 can also bind to their own promoters and form positive feedback loops to auto-regulate their own expression (14).

Under the right conditions, we can induce ESC differentiation into three different embryo germ layers: mesoderm, endoderm, and ectoderm (15). ESC differentiation requires rapid downregulation of pluripotent networks as well as upregulation of lineage-specific markers. The

downregulation of pluripotent factors depends on several different epigenetic mechanisms such as histone modification, DNA methylation and miRNA regulation (14). The *in vitro* differentiation of ESCs is an excellent model to mimic the early stage of embryo development at cellular level. Several technologies to manipulate genes (e.g., RNA interference and CRIPR/Cas9 gene editing) have become useful tools to study how developmental genes be regulated in ESC differentiation.

## **1.2 GENERAL BACKGROUND OF EPIGENETICS**

All cell types that originate from the fertilized egg during development in principle share the same DNA sequence. However, various cell types have diverse phenotypes with different gene expression profiles. In addition, they may have different responses to same stimuli, and some genes are only transcribed and expressed under certain stimuli or environmental conditions. The mechanisms underlying this phenomenon are a fundamental question in the field of biology. In 1942, Conrad Waddington first coined the word “epigenetics” as he proposed a concept to study the mechanisms of developmental processes between genotype and phenotype (16). He later defined epigenetics as “the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being” (17). Now, with the advances in knowledge of this field, epigenetics is defined as “a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence” (18). The detailed mechanism of epigenetics can be further characterized into several different categories, including the post-translational modifications of histones, DNA methylation, ATP-dependent chromatin remodeling, histone variants incorporation, and non-coding RNA pathways (19).



### 1.3 POST-TRANSLATIONAL MODIFICATIONS OF HISTONES

Chromatin is a complex structure that is composed of DNA, proteins and RNA. Surprisingly, two-meter long DNA is packaged via chromatin in the nucleus (approximately 6  $\mu\text{m}$  in diameter) (20,21). The compact structure of chromatin is dynamically regulated and is important for gene regulation (22). Chromatin can be further characterized into euchromatin or heterochromatin. Chromatin is loosely packed in euchromatin regions. Therefore, transcription machinery can easily access DNA in euchromatin regions so that genes in the regions are actively transcribed (23). In contrast, heterochromatin is a tightly packed chromatin, and its compact structure prevents the access of transcription machinery while resulting in gene silencing (24).

The nucleosome is the basic unit of chromatin and consists of a histone octamer wrapped around 147 base pairs of DNA. Histone proteins contain high amounts of basic amino acids (e.g., lysine and arginine) and therefore are positively charged. The positively charged histones play a pivotal role in controlling the structure of the chromatin and accessibility of DNA. Two H2A-H2B dimers and one H3-H4 tetramer form the core histone octamer whereas histone H1 serves as a linker histone (25). The N-terminal tails of the histones are more flexible and are subjected to post-translational modifications, including methylation, phosphorylation, acetylation, ADP ribosylation, ubiquitylation, and sumoylation (26). These modifications can change the structure of the nucleosome through altering the charge on histones or serving as binding sites to recruit downstream proteins to modulate several biological processes, such as gene expression, DNA repair, and DNA replication (27).

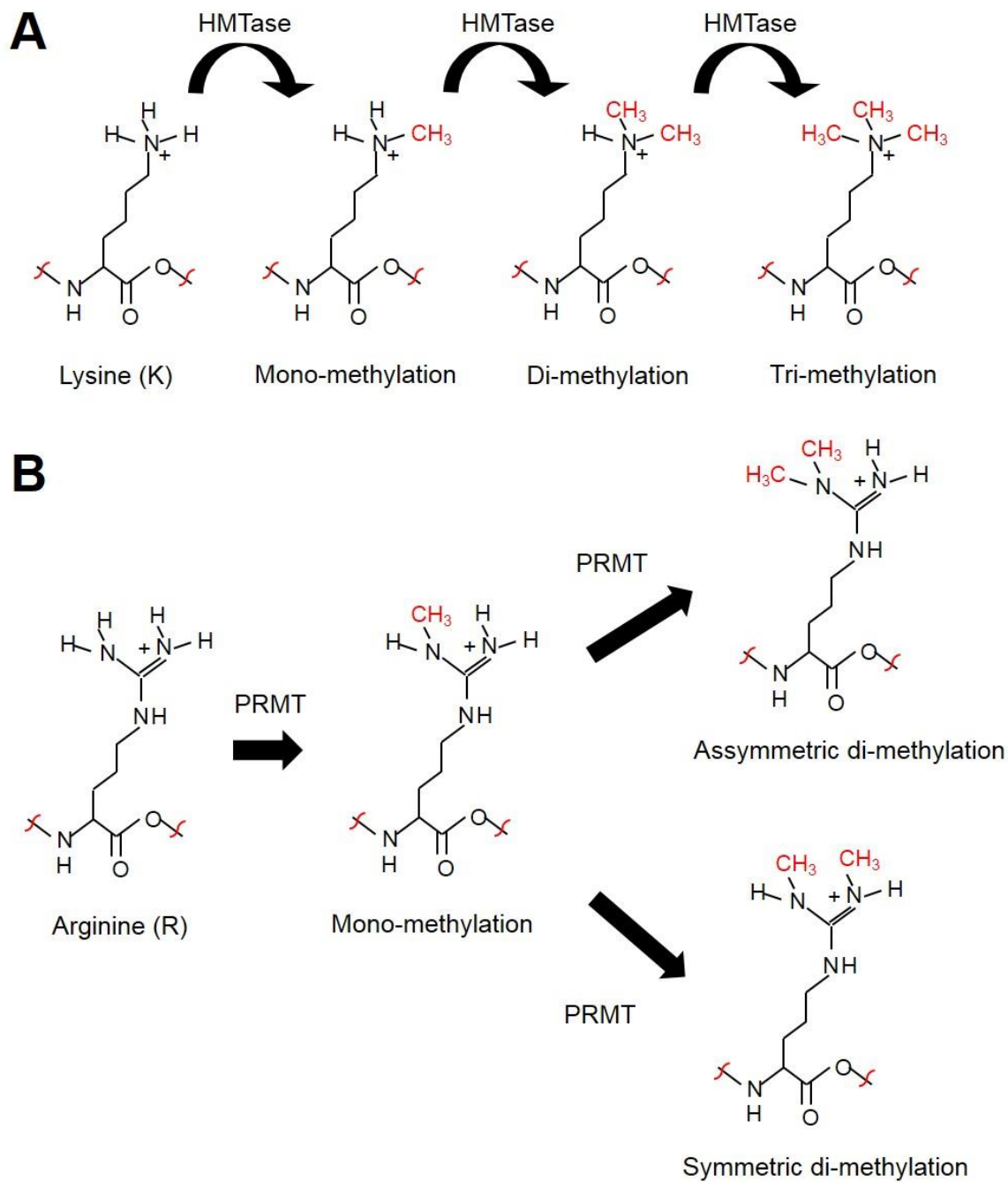
The histone code hypothesis proposes that the different combinations of histone modifications may have distinct readouts for gene regulation (28). The histone modifications are dynamically regulated by epigenetic modifiers. The enzymes that add the modifications to the histone are called “histone writers”; whereas, the enzymes that remove the modifications are

“histone erasers”. These epigenetic modifiers play important roles in regulating gene expression. Specific modifications can be recognized by proteins known as “histone readers”, which can further recruit transcription machinery or change chromatin conformation to activate or repress the expression of target genes. Emerging evidence shows that dysregulation of these epigenetic regulators may lead to several severe developmental diseases and cancer development (29).

## 1.4 HISTONE METHYLATION

Histone methylation is an important epigenetic modification that regulates gene expression and consequently affects numerous biological processes, such as embryonic development, cell differentiation, and cancer development (30). Histones can mainly be methylated on lysine or arginine residues. There are three different states of methylation on lysine  $\epsilon$ -amine group - *i.e.*, mono-methylation (me1), di-methylation (me2) and tri-methylation (me3) (31) (**Figure 1A**). For arginine methylation, there are also three different types of methylation: mono-methylation (me1), asymmetric di-methylation (me2a) and symmetric di-methylation (me2s) (32) (**Figure 1B**). The positive charge on histone lysine and arginine residues are important for DNA packaging and chromatin structure. Histone acetylation and phosphorylation change the net charge on histones, thereby altering the chromatin accessibility to several cellular machineries such as transcription and DNA repair (33,34). In contrast, histone methylation has a minimal effect on the overall charge of lysine or arginine residues of histones, but creates a binding motif to recruit downstream effector proteins to achieve specific biological outcomes (35). These downstream effector proteins contain methylation binding domains, including PHD finger, chromodomain, WD40 domain, tudor domain and PWWP domains (36). For example, TATA box-binding protein associated factor 3 (TAF3) contains a PHD finger domain which can strongly bind to tri-methylated histone H3 lysine 4 (H3K4me3) at the promoters and recruit the TFIID complex to the promoters for transcription initiation (37,38).

**Figure 1**



### **Figure 1 Lysine and arginine methylation**

**(A)** Histone methyltransferases methylate the  $\epsilon$ -amine group of lysine into mono-methylation, di-methylation and tri-methylation. **(B)** Arginine residues can be methylated by PRMTs into mono-methylation, asymmetric di-methylation and symmetric di-methylation.

## 1.5 SET DOMAIN-CONTAINING HISTONE METHYLTRANSFERASES AND DISEASES

Histone methylation is established by histone methyltransferases (HMTase) which can be classified into three different families, including SET domain-containing methyltransferase, seven- $\beta$ -strand methyltransferase, and protein arginine methyltransferase (PRMT). The SET domain-containing proteins can transfer methyl groups from S-Adenosyl methionine (SAM) to lysine residues of target proteins through their catalytic SET domain (39). The seven- $\beta$ -strand methyltransferase family can methylate a variety of substrates, including lipids, nucleic acids and proteins. Until now, Dot/DOTIL was the only histone-specific methyltransferase that belonged to this group (40). The third group of HMTases are protein arginine methyltransferases which can also use SAM as a cofactor and transfer methyl groups to the arginine residues on histones (32).

The SET domain-containing proteins are major histone lysine methyltransferases and are highly conserved from yeast to humans. So far, 51 SET domain-containing proteins have been identified according to the SMART database. SET is an acronym comprising three different *Drosophila melanogaster* proteins: Su(var)3-9, Enhancer-of-zest and Trithorax, which all contain approximately 130-amino acid long conserved sequences and possess HMTase activity (41,42). According to sequence alignment, two conserved sequence motifs, ELxF/YDY and NHS/CxxPN (x being any amino acid) have been identified in the SET domain. It has been shown that the NHS/CxxPN motifs play important roles in the binding of cofactor SAM (43-45). In general, the structure of the SET domain can be divided into three parts: core SET, pre-SET (N-SET, n-terminal), and post-SET (C-SET, c-terminal). The pre-SET and post-SET provide scaffold structures to complement the core SET for specific substrate recognition and cofactor binding. Based on crystal structure analysis, the substrate peptide and the cofactor SAM bind to different sides of the SET domain and meet at the active site located at the core SET (46).

Numerous reports have shown that mutations on HMTases are associated with severe developmental disorder and tumorigenesis. For example, Mixed lineage leukemia 1 (MLL1) is a histone H3 lysine 4 (H3K4) methyltransferase which locates on chromosome 11q23. Chromosome translocations on the MLL1 gene have been linked to several types of leukemia, including MLL, acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) (47,48). It has been shown that the translocation leads to the N-terminal of MLL1 to fuse with other proteins (e.g. ENL, ELL and AF9) and creates a new chimeric protein (e.g. MLL-ENL, MLL-ELL and MLL-AF9) (49). Because the catalytic SET domain of MLL1 is at the C-terminal end, the MLL1 translocation results in the loss of the SET domain as well as methyltransferase activity. Moreover, several MLL1 translocation partners are members of the super elongation complex (SEC), including ENL, ELL and AF9. Therefore, the new chimeric protein may lead to dysregulation of MLL target genes and transcription elongation, which play an important role in the pathogenesis of leukemia (50,51). Another example is MLL4 (KMT2D/MLL2/ALR), an H3K4 methyltransferase, which acts as a key regulator by establishing the enhancer mark H3K4me1 (52). Several different mutations on MLL4 have been found in patients with Kabuki syndrome, which is a congenital disorder characterized by mental retardation, distinctive facial characteristics, congenital heart disease, and skeletal abnormalities (53,54). Moreover, several studies also showed that the loss-of-function mutations on MLL4 often occur in childhood medulloblastoma (55,56). These reports further support the importance of SET domain-containing HMTases in regulating development and tumorigenesis. Therefore, it is critical to further understand the detailed molecular mechanisms and biological functions of these SET domain-containing HMTases.

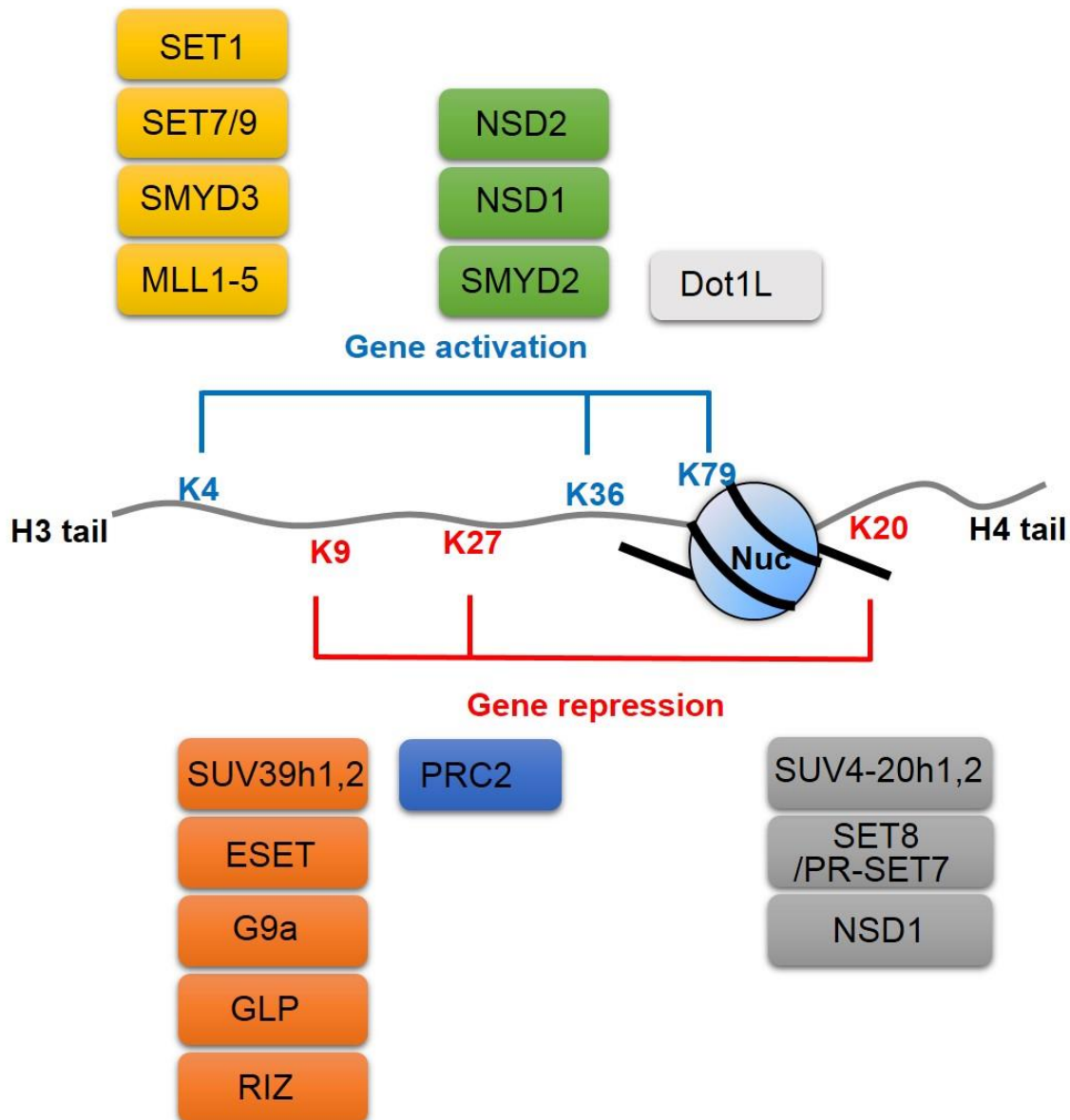
## 1.6 HISTONE LYSINE METHYLATION AS A SIGNATURE FOR GENE REGULATION

During past several years, histone modifications and their functions have been extensively studied. There are six major lysine methylation sites on histones, including H3K4, H3K9, H3K27, H3K36, H3K79, and H4K20. Functional studies have revealed the important roles of these histone lysine methylation marks in gene regulation. In general, methylation on H3K4, H3K36, and H3K79 are associated with gene activation while H3K9, H3K27, and H4K20 are associated with gene repression (57-59) (**Figure 2A**). Trimethyl H3K4 decorates transcriptional start sites (TSS), and its levels are positively correlated with transcription frequency (60). Trimethyl H3K36 is enriched near the 3' end of active gene and has been implicated in transcription elongation (61). H3K79 methylation is strongly correlated with gene activity and is localized at active genes (62). In contrast, H3K9 methylation has been shown to regulate heterochromatin formation (e.g. X inactivation) and gene silencing (63). H3K27 methylation serves as a binding site for the polycomb repressive complex (PRC), which represses gene expression. H4K20 methylation prevents the formation of the active mark H4K16 acetylation (H4K16Ac), which leads to RNA PolII promoter-proximal pausing and gene repression (64).

With the development of chromatin immunoprecipitation (ChIP) followed by genome-wide sequencing (ChIP-seq), the genome-wide distribution of histone modifications has been determined (65). Computational analysis has shown that specific histone modifications may correlate with different functional elements in the genome (also known as chromatin states, e.g. promoters, enhancers, gene bodies and repressed genes) (66). For example, the transcriptional start sites (TSS) of active promoters are generally marked by H3K4 trimethylation (H3K4me3) (67), while H3K4 monomethylation (H3K4me1) and acetylation on H3K27 generally marks active enhancers (68,69). These histone modifications often serve as binding sites to recruit downstream effector proteins. For example, the chromatin remodelers CHD1 and BPTF are

recruited to the TSS through binding to H3K4me3 to relax chromatin structure and allow the access of transcriptional machinery (70,71). In addition, the TFIID complex component TAF3 can strongly interact with H3K4me3 and recruit the pre-initiation complex (PIC) to the TSS for transcription activation (38).

**Figure 2**





**Figure 2. Six major histone lysine methylation sites**

H3K4, H3K9, H3K27, H3K36, H3K79, and H4K20 are major histone lysine methylation sites that play important roles in gene regulation. Methylation on H3K4, H3K36, and H3K79 is associated with gene activation. Methylation on H3K9, H3K27, and H4K20 is associated with gene repression.

**CHAPTER 2**

**FUNCTION OF SETD5 IN MOUSE EMBRYONIC STEM  
CELL DIFFERENTIATION**

## 2.1 INTRODUCTION

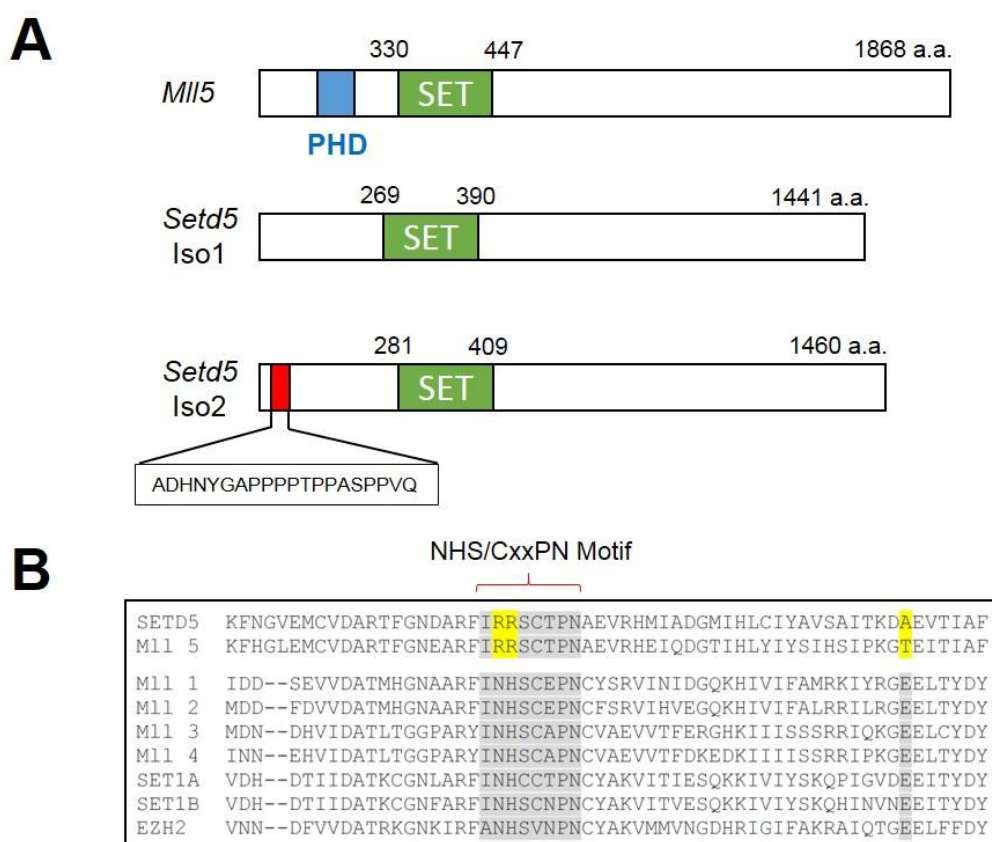
### 2.1.1 SET Domain-containing protein 5 (SETD5)

SET domain-containing 5 (SETD5)/KIAA1757 was originally identified from a human cDNA sequencing project and was shown to be ubiquitously expressed across human tissues with predominantly high levels in brain tissue (72). At least two SETD5 isoforms have been identified in human and mouse. SETD5 isoform 1 (canonical sequence) consists of 23 exons that encode a protein of 1442 amino acids with an estimated molecular weight of 163 kDa. SETD5 isoform 2 also consists of 23 exons but encodes a longer protein product with 1661 amino acids. This longer isoform has an alternative 3' splicing site on exon 4 which generates a 19-amino acid-insertion at codon 59 (A → ADHNYGAPPPPTPPASPPVQ) (Figure 3A). SETD5 contains one SET domain near its N-terminal and is considered a putative HMTase (Figure 3A). However, the enzymatic activity and functions of SETD5 have not been well characterized.

Based on phylogenetic analysis and sequence alignment, SETD5 and another SET domain-containing protein, MLL5 (KMT2E), are paralogs with similar domain architecture and have highly conserved SET domains (73) (Figure 3A and 3B). MLL5 was originally identified as a SET domain-containing tumor suppressor in myeloid leukemia and therefore was categorized into the MLL family, which includes MLL1, MLL2, MLL3, and MLL4 (74). However, a recent report by Mas-y-Mas *et al.* suggested that MLL5 is not a member of the MLL family but belongs in its own family with SETD5 since the SET domain of MLL5 is distinct from other MLL family proteins (75). Knockout mice of *MLL5* showed partial lethality and multiple hematopoietic defects, suggesting the functional importance of MLL5 in development (76). Several groups have shown that MLL5 does not have intrinsic HMT activity but can indirectly regulate H3K4 methylation status through inhibiting the H3K4 demethylase LSD1 (76,77). Moreover, Mas-y-Mas *et al.* showed that the SET domain of MLL5 does not have a conserved NHS/CxxPN motif (NH is

replaced by RR) and may not bind SAM. In this study, crystal structure analysis demonstrated that the SET domain of MLL5 contains a large loop structure to prevent the binding of histone peptides (75) (**Figure 3B**). Based on the fact that the SET domains of MLL5 and SETD5 are highly conserved, it implies that SETD5 may also lack intrinsic HMT activity.

**Figure 3**



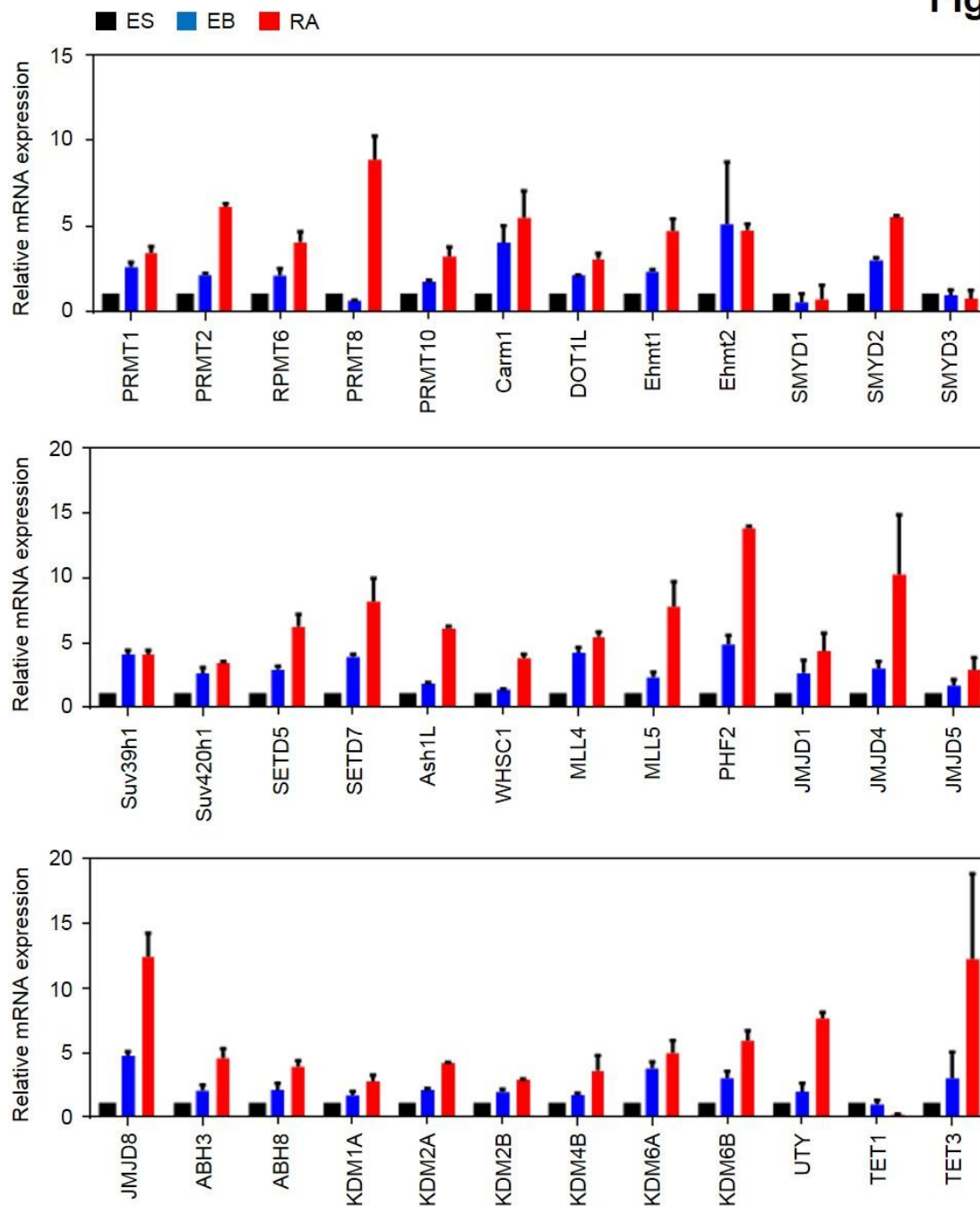
**Figure 3. SETD5 is a putative histone methyltransferase which has been linked to 3p25 microdeletion syndrome**

**(A)** Diagrammatic representation of mouse MLL5, SETD5 isoform 1, and SETD5 isoform 2. PHD, plant homeodomain. SET, Su(var)3-9, Enhancer-of-zest and Trithorax domain. SETD5 isoform 2 contains a 19 amino acid insertion at the N-terminal of SETD5. **(B)** Sequence alignment of the SET domain of SETD5 and MLL5 with other SET domain-containing histone methyltransferases, including MLL1, MLL2, MLL3, MLL4, SET1A, SET1B, and EZH2. The conserved ELxF/YDY and NHS/CxxPN motifs are highlighted in grey. The SET domain of SETD5 and MLL5 do not have conserved NHS/CxxPN motifs, as the asparagine (N) and histidine (H) are replaced by two arginine residues (RR) (highlighted in yellow).

### 2.1.2 Rationale and hypothesis

As mentioned previously, epigenetic regulation is essential in ESC differentiation and embryonic development. In our effort to identify epigenetic regulators that are critical for ESC differentiation, we examined the expression profile of 36 histone methylation modifiers to determine which histone methylation modifiers are up-regulated during ESC differentiation. We found that expression levels of SETD5 increased across stages of RA-induced mouse ESC differentiation (**Figure 4**). SETD5 is a putative histone methyltransferase, although its function and enzymatic activity remain elusive. Interestingly, Osipovich *et al.* showed that *Setd5*-null mice were embryonic lethal (78). Moreover, several clinical reports indicated that *de novo* loss-of-function mutations in SETD5 can cause intellectual disability and autism, further pointing to an important role for SETD5 in ESC differentiation, embryogenesis, and development (79-82). Therefore, we hypothesized that SETD5 is required for mouse ESC differentiation. In this study, we utilized biochemical and molecular biological approaches to determine SETD5's catalytic activity and its function in mouse ESC differentiation.

**Figure 4**



**Figure 4. The expression level of 36 epigenetic regulators in different stages of RA-induced mouse ESC differentiation**

Comparison of relative mRNA level of 36 epigenetic regulators in different stages of RA-induced mouse ESC differentiation. EB, 5 days after embryoid body generation; RA, 5 days after retinoic acid treatment.



## 2.2 MATERIALS AND METHODS

### 2.2.1 Antibodies, plasmids and other reagents

Anti-HA antibodies were purchased from Abcam (ChIP grade, ab91110) and Biolegend (#901501). Anti-HCF1 (A301-399A), was from Bethyl Laboratories. Anti-Oct4 (#2840 antibody) was from Cell Signaling. Anti- $\beta$ -actin antibody (A5441) and anti-FLAG (F1804) was from Sigma-Aldrich. Anti-H3 antibody (ab1791) was from Abcam. Anti-SETD5 antibody was custom generated by Thermo Fisher. FLAG-tagged expression vector pFlagCMV2 and GST-tagged bacterial expression vector pGEX6P1 was obtained as previously described (83). Baculoviral expression vector pFastBac-HTa was obtained as previously described (84). Lentiviral expression vector pCDH-EF1-IRES-Neo was a gift from Dr. Mien-Chie Hung. pGEX6P1-SET7 was kindly provided by Dr. Xiaobing Shi. Oligonucleotides used for cloning, site-directed mutagenesis, RT-PCR, ChIP-PCR and CRISPR-Cas9 are listed in [Table 1](#).

**Table 1 SETD5 Oligonucleotide sequences.**

**Table 1**

		5'-3' Forward Sequence	5'-3' Reverse Sequence
<b>Cloning &amp; mutation</b>	pCDH-mSETD5-Flag	TACGGAGATCTGCCACCATGGACTACAAAGACGATGACGACAA GAGCATTGCAATCCCTCTGGGA	TACGGGCGCCGCTCACTAGGAAAGTCCTGTCTGAGT
	pFlagCMV2-mSETD5	TACGGGCGCCGCGAGCATTGCAATCCCTCTGGGA	TACGGGTCGACTCACTAGGAAAGTCCTGTCTGAGT
	pFastBac-mSETD5	TACGGGTCGACGAAGCATTGCAATCCCTCTGGGA	TACGGGCGCCGCTCACTAGGAAAGTCCTGTCTGAGT
	pGEX6P1-mSETD5-N	TACGGGTCGACTCATGAGCATTGCAATCCCTCTG	TACGGGCGCCGCTCAAACCTCCTCATGCTCATTGGA
<b>RT-PCR</b>	delta-SET domain	AGTAATCGGTTCCCAAATGCAGTTAGAATATAGTAAGTGAATTA CAAAG	CTTTGTAATTACAGTTACTATATTCTAACTGCATTGGGAACCGA TTACT
	mGAPDH	CATGGCCTTCGGTGTCTCTA	GCCTGCTTCACCACCTTCTT
	mOct4	AGAGGATCACCTTGGGGTACA	CGAAGCGACAGATGGTGGTC
	mNanog	TCTTCCTGGTCCCCACAGTTT	GCAAGAATAGTTCTCGGGATGAA
	mHOXA1	CCCAGACGGCTACTTACCAGA	CATAAGGCGCACTGAAGTTCT
	mHOXA2	TACGAATTTGAGCGAGAGATTGG	GTCGAGGTCTTGATTGATGAAC
	mNestin	GCCTATAGTTCAACGCCCCC	AGACAGGCAGGGCTAGCAAG
	mPax6	AAGGAGGGGGAGAGAACC	TCTGAGCTTCATCCGAGTCTT
	mSETD5	TGGGACCACTCAGAGGCAT	CCACAGCGTTCTCTACGG
	HA-SETD5	ACACCAGTACCGACTCCA	AGCGTAATCTGGAACATCGTA
<b>ChIP</b>	Hoxa1-TSS	GTCTATGGAGGAAGTGAGAA	GGGGTATTCCAGAAAGGAGTTCA
	Hoxa2-TSS	CCCAGCAGCGATCTTCTATG	CGTAATTCATGGCCTTCTCT
	Hoxa1-RARE	AAAGGCTGCTAACAACTG	ACTGCTGGGACTCATTCTAAAG
<b>CRISPR-Cas9</b>	Cas9n-SETD5-2XHA-F1	CACCGTTGACTCCTGACCCTTGCA	AAACCTGCAAGGGTCAGGAGTCAAC
	Cas9n-SETD5-2XHA-F2	CACCGCAGGACTTTCTAGGGCTTC	AAACGAAGCCCTAGGAAAGTCCTGC
<b>cloning</b>	Cas9-SETD5-2XHA-seq	GGACTATCAGTCTGCCCAATG	TGTCCTCACTGCACAAGTGAC

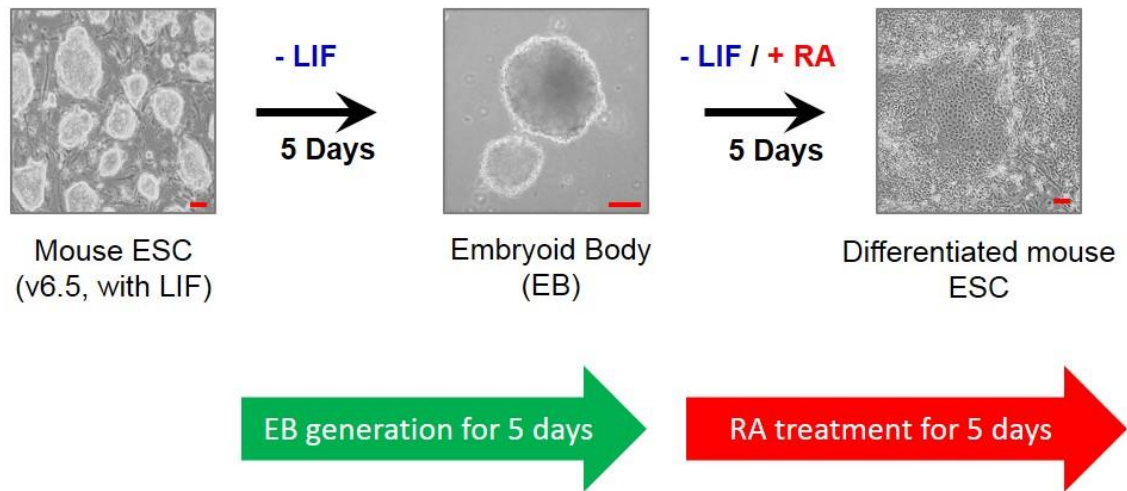
### **2.2.2 Mouse ESC culture and AP staining**

V6.5 mouse ESC were cultured on gelatin-coated plates in complete knockout Dulbecco modified eagle's medium (Life technology) supplemented with 20% ES grade FBS (GenDEPOT), 2 mM L-glutamine, 50 µg/ml penicillin and 50 µg/ml streptomycin (Life technologies), 0.1 mM β-mercaptomethanol, 0.1 mM nonessential amino acid (NEAA), and 1,000 U/ml leukemia inhibitory factor (LIF, homemade). mESC were trypsinized and split every 3 days while medium was changed daily. Alkaline phosphatase is a universal pluripotent marker for all types of pluripotent stem cells. Red-Color AP Staining kit was purchased from System Biosciences and mouse ESC were stained according to manufacturer's instruction.

### **2.2.3 Retinoic acid (RA)-induced differentiation of mouse ESCs**

In order to induce differentiation, we first generate embryoid body (EB) from V6.5 mouse ESC. The generation of EB allow ESC to grow under three dimensional condition which mimic the real embryogenesis. Briefly, V6.5 mouse ESC were trypsinized and suspended in 10 cm low attachment petri dish (Fisher) with ESC media without adding LIF. Under this condition, the suspended mouse ESC will not attach to culture dish and form round-shaped cell aggregates called as EB. After 5 days of incubation, the suspended EB were transferred to gelatin-coated tissue culture dish allowing the attachment of EB and 0.5 µM RA was then added to induce further differentiation. After 5 days of RA treatment, the differentiated cells spread out from the attached cell aggregates and were harvested for further analysis ([Figure 5](#)).

**Figure 5**



**Figure 5. Retinoic acid (RA)-induced Differentiation of Mouse ESCs**

Schematic diagram representation for the procedure of retinoic acid-induced mouse ESCs differentiation.

#### **2.2.4 RNA interference**

Mouse shSETD5 (shSETD5-1, TRCN0000098701; shSETD5-5, TRCN0000098700) and mouse shHCF1 (shHCF1-1 TRCN0000235784; shHCF1-5, TRCN0000086292) in the puromycin-resistant PLKO.1 vector were purchased from Sigma-Aldrich. The shRNA plasmids (30 µg) were transfected into mouse ESCs (5 X 10<sup>5</sup> cells in 0.8 mL) by using Gene Pulser Xcell Electroporation System (500 µF, 250 volts; BioRad) according to the manufacturer's protocol and were plated on a 6 cm dish. The cells were treated with puromycin (1.0 µg/ml) 48 h later to obtain stable clones.

#### **2.2.5 Western blot analysis**

Mouse ESCs were lysed by mammalian lysis buffer (20 mM Tris-HCl, 137 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 1mM EDTA, 10% glycerol, 1% TritonX-100, 0.2 mM PMSF) to obtain total cell lysates. Protein concentration was determined by Bradford Protein Assay (BioRad). 20 µg of total protein were subjected to standard western blot analysis. Antibodies against HA, SETD5, OCT4, NANOG, NESTIN and β-ACTIN are used for immunoblotting.

#### **2.2.6 Quantitative PCR for mRNA expression**

Total RNAs were isolated by using Quick-RNA MiniPrep kit (Zymo Research). To detect relative mRNA level, cDNA was synthesized by using iScript cDNA Synthesis Kit (BioRad) according to manufacturer's instruction. Quantitative PCR was performed by using CFX384 real-time PCR detection system (BioRad) and GAPDH level were used as internal control.

### 2.2.7 Rescue experiments by ectopic expression of SETD5

To ectopic express SETD5 in V6.5 mouse ESC, FLAG-tagged full length mouse SETD5 cDNA was cloned into pCDH-EF1-IRES-Neo vector to obtain FLAG-SETD5 expression construct (pCDH-EF1-FLAG-SETD5-IRES-Neo). Next, pCDH-EF1-IRES-Neo empty vector and pCDH-EF1-FLAG-SETD5-IRES-Neo were transfected into shSETD5-5-treated cells by using Lipofectamine 3000 (Life technology) according to manufacturer's protocol. 48 h after transfection, cells were treated with puromycin (InvivoGen, 1.0 µg/ml) and genitacin (Life technology, 500 µg/ml) for 14 days to obtain stable clones. To perform rescue experiment, shLuc-treated cells, shSETD5-5-treated cells, and two genitacin-resistant cells (shSETD5-5+Vector and shSETD5-5+FLAG-SETD5) were subjected to RA-induced ESC differentiation assay for further analysis.

### 2.2.8 CRISPR/Cas9 gene editing

pSpCas9n(BB)-2A-GFP (PX461) plasmid with nCas9 (D10A nickase mutant) were obtained from Addgene. To generate *Setd5*-2XHA knock-in V6.5 mouse ESC, two guide RNA (sgRNA) sequences around the stop codon (inside exon 22) of SETD5 (F1: TTGACTCCTGACCCTTGCAG; F2: CAGGACTTTCCTAGGGCTTC) were cloned separately into pSpCas9n(BB)-2A-GFP. Single-stranded donor oligonucleotides (ssODN) repair template which contains 2XHA sequences was synthesized by IDT (ssODN sequence: GGGGGTGTACACCAGTACCGACTCCAGCCACTGCAAGGGTCAGGAGTCAAGACTCA GACAGGACTTTCGGCGGAGGTTACCCATACGATGTTCCAGATTACGCTGGCTATCCA TATGACGTTCCAGATTACGCTGTCTGACTAGGGCTTCTGGATATGGGCAAACCTGAACTT AATGAGCCCATAGCTGCTTCCTTCC). sgRNA-containing plasmids and ssODN repair template were then co-transfected into V6.5 mouse ESCs by using Lipofectamine 3000 according

to manufacturer's protocol. After 48 h of incubation, mouse ESCs were trypsinized and sorted by GFP signals. GFP-positive mouse ESCs were plated into 96 well plate (one cell per well) to obtain single colony for further selection.

### **2.2.9 Chromatin Immunoprecipitation (ChIP) assay**

ChIP assay was performed as previously described with small modification (85). Mouse ESCs were first fixed by 1% formaldehyde. Cell pellets were then lysed with ChIP lysis buffer and sonicated by using Bioruptor (Diagenode) for 5 min (30 sec on and 30 sec off for 5 cycles) to shear the DNA. The DNA was further digested by adding 10U of Benzonase (Sigma) and incubated for 15 min on ice to obtain a smear of DNA from 200 bp to 500 bp. Antibodies were added and incubate for overnight at 4 °C. Pre-blocked protein A beads were added and incubated for another 1-2 hr to capture protein-DNA complex. The beads were then washed once with following buffer: low-salt buffer, high-salt buffer, LiCl buffer and TE buffer. ChIP DNA were then eluted by ChIP elution buffer (1% SDS and 0.1M NaHCO<sub>3</sub>). The eluate were reverse-crosslinked and ChIP DNA were purify by phenol/chloroform extraction method.

### **2.2.10 Immunoprecipitation assay**

Immunoprecipitation assay was performed as previously described (83). To test the interaction between SETD5 and RAR- $\alpha$ , FLAG-RAR- $\alpha$  was transiently expressed in HEK 293T cells by Lipofectamine 2000 (Life technologies) according to manufacturer's protocol. 48 h after transfection, DMSO or 10  $\mu$ M retinoic acid was added to cells and incubated for another 1 h. Total cell lysates were then harvested by adding mammalian lysis buffer and Anti-FLAG M2 affinity gel (Sigma) was added and incubated in 4 °C for 5 h to perform FLAG-immunoprecipitation. The

affinity gel was washed with BC500 (20 mM Tris-HCl, 500 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 10% glycerol, 0.2mM PMSF at pH 8.0).and eluted by by using 0.4 µg/mL of FLAG peptides in HMT buffer (50 mM Tris-HCl, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 4 mM DTT and 10% glycerol at pH 8.5). The eluates were subjected to SDS-PAGE followed by western blot analysis and stored for further analysis.

To verify the endogenous interaction between SETD5 and HCF1 in mouse ESC, total cell lysates from wild-type and *Setd5*-2XHA knock-in mouse ESCs were harvested to perform HA-IP. Briefly, Anti-HA agarose (Pierce) was added to total cell lysates and incubated overnight. The agarose gel was washed with mammalian lysis buffer and eluted by using 1 µg/mL of HA peptides (Sigma). The eluates were subjected to SDS-PAGE followed by western blot analysis.

### **2.2.11 Recombinant protein purification**

To obtain full length SETD5 recombinant protein, FLAG-tagged mouse SETD5 cDNA was cloned into pFastBac-HTa vector. The construct was transformed into DH10Bac competent cells (Invitrogen) to generate recombinant Bacmid. Recombinant Bacmid DNA was then transfected into Sf9 insect cells by using Cellfectin (Invitrogen) to generate recombinant baculoviruses. The full-length FLAG-SETD5 recombinant protein was expressed by infecting Sf21 insect cells with recombinant baculoviruses. 72 h after infection, the Sf21 cells was harvested and FLAG-immunoprecipitation was performed to obtain full length FLAG-SETD5 recombinant protein.

To obtain GST-tagged SETD5-SET domain recombinant protein, SETD5 N-terminal fragment (1-500 a.a.; SETD5 SET domain: 291 ~415 a.a.) was cloned into pGEX-6P-1 GST expression vector (pGEX-6P-1-SETD5-SET). GST-SETD7 expression plasmid (pGEX-6P-1-SETD7) was kindly provided by Dr. Xiaobing Shi. The plasmids were transformed into BL21

competent cells and recombinant protein expression was induced by adding Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) at 16 °C. 20 h after induction, the bacteria was harvested and lysed by mammalian lysis buffer. Cell lysates were then incubated with Glutathione Sepharose 4B (GE Healthcare) for overnight to perform GST pull-down assay. GST fusion protein were purified and eluted by elution buffer (10 mM glutathione, 100 mM Tris HCl [pH 8.0]).

#### **2.2.12 Affinity purification of SETD5 complex**

Mouse SETD5 cDNA was cloned into pFlagCMV2 (FLAG-SETD5) and SET domain deletion mutant in the same expression vector (FLAG-SETD5- $\Delta$ SET) was generated by using QuickChange Site-Directed Mutagenesis Kit mutagenesis assay (Agilent). The plasmids were transiently expressed in HEK 293T cells and nuclear extracts were isolated 48 h after transfection. SETD5 complex were purified by using FLAG-immunoprecipitation. FLAG-IP eluates were then subjected to silver staining, *in vitro* histone methyltransferase (HMT) assay and liquid chromatography-tandem mass spectrometry.

#### **2.2.13 *In vitro* histone methyltransferase (HMT) assay**

Recombinant protein and FLAG-IP eluates were incubated with 1  $\mu$ g of recombinant H2A, H2B, H3, H4 (New England BioLabs) or recombinant nucleosomes. [ $^3$ H]-labeled S-adenosyl-L-methionine ([ $^3$ H]-SAM) was added to the reaction mix to serve as methyl donor. All reaction were performed in HMT buffer (50 mM Tris-HCL, 100 mM KCL, 5 mM MgCl<sub>2</sub>, 4 mM DTT and 10% glycerol at pH 8.5) and incubated for 16 h in 30 °C. The reaction was stopped by adding SDS sample buffer and was subjected to SDS-PAGE. The HMT signal was then detected by autoradiography.



#### **2.2.14 Radioactive S-adenosyl-L-methionine (SAM) binding assay**

1 µg of GST fusion protein (GST, GST-SETD7 and GST-SETD5-SET) were incubated with 2.5 µCi of [<sup>3</sup>H]-labeled S-adenosyl-L-methionine ([<sup>3</sup>H]-SAM) in SAM binding buffer (100 mM Tris HCl [pH 8.0]) for 1 h at 30 °C. Binding mixtures were then pull-down by Glutathione Sepharose 4B (GE Healthcare) and unbound [<sup>3</sup>H]-SAM was removed by washing 3 times with binding buffer. GST fusion protein were eluted by elution buffer (10 mM glutathione, 100 mM Tris HCl [pH 8.0]). The bound [<sup>3</sup>H]-SAM was detected by Liquid Scintillation Spectrometry (LS 6000 Scintillation Counter, Beckman).

#### **2.2.15 Statistical analysis**

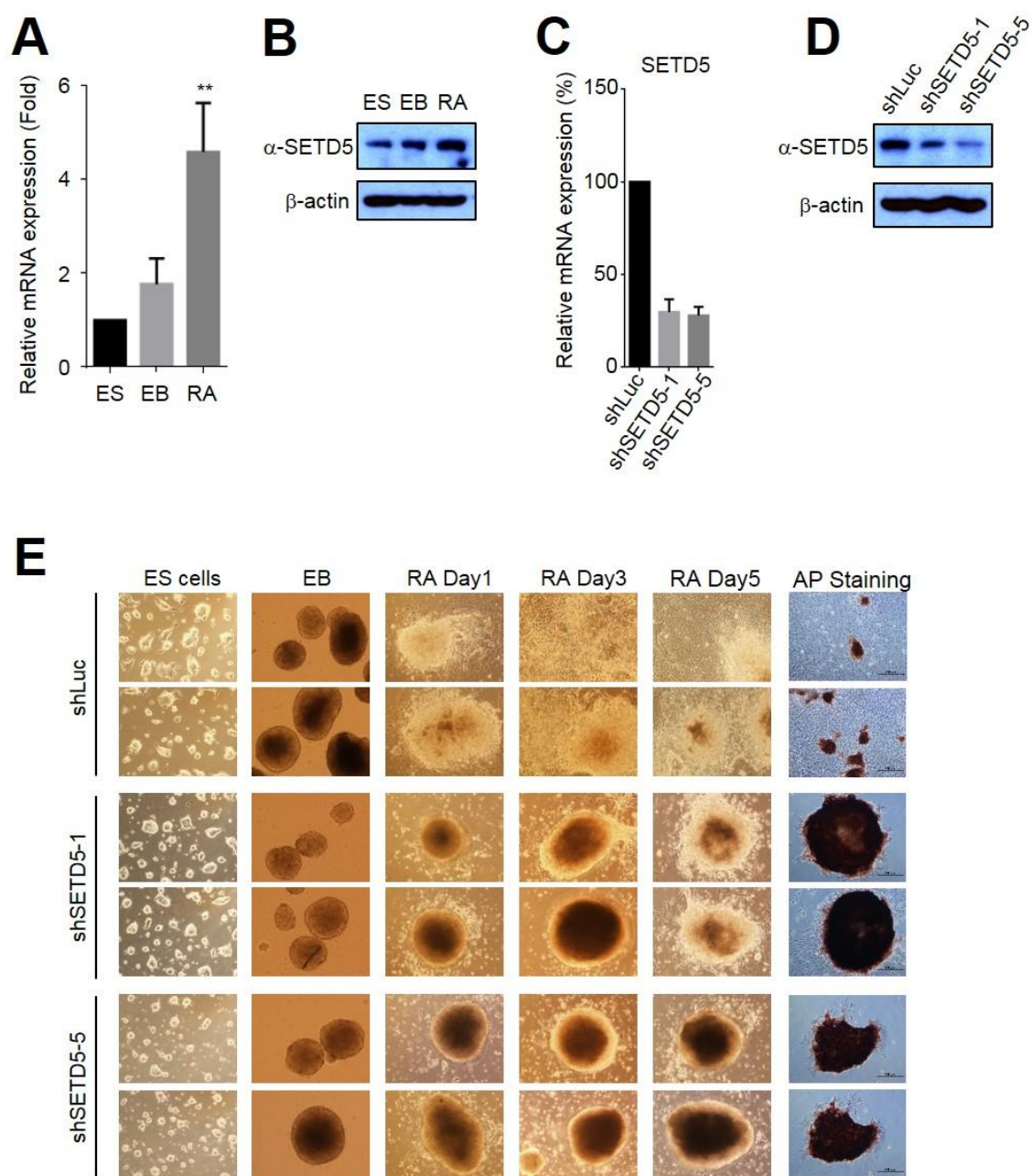
The statistical significance between the two groups was analyzed by student's t-test using Prism software (GraphPad Software Inc.). Data are presented as the mean ± standard deviation (SD) of at least three independent experiments. P < 0.05 (\*), P < 0.01 (\*\*), and P < 0.001(\*\*\*) indicate statistically significant changes.

## 2.3 RESULTS

### 2.3.1 SETD5 is required for proper mouse ESC differentiation.

In order to determine the role of SETD5 in ESC differentiation, we employed RA-induced differentiation of mouse ESCs, as this is a well-established *in vitro* differentiation model system (85,86). We first examined the expression level of SETD5 across different stages of RA-induced mouse ESC differentiation. Quantitative RT-PCR results showed that *Setd5* mRNA levels gradually increased during RA-induced mouse ESC differentiation (Figure 6A). Consistent with this, Western blot analysis showed that SETD5 protein levels also increased (Figure 6B). To assess the role of SETD5 in RA-induced mouse ESC differentiation, we depleted SETD5 in V6.5 mouse ESCs by using two independent shRNAs against SETD5 (shSETD5-1 and shSETD5-5). SETD5 knockdown was confirmed by quantitative RT-PCR and Western blot analysis (Figure 6C and 6D). SETD5 knockdown appeared not to affect the pluripotency of mouse ESCs because SETD5-depleted cells were morphologically normal compared to the control, as they formed round-shaped colonies after several passages (Figure 6E). Next, we compared RA-induced mouse ESC differentiation between shLuc-treated and SETD5-depleted cells. During the generation of EB, SETD5 depletion had no obvious effect on the size and shape of EBs (Figure 6E). After RA treatment, round-shaped colonies of the shLuc-treated cells spread out and negatively stained for alkaline phosphatase (AP), a pluripotent marker. In contrast, as most SETD5-depleted ESCs still formed round-shaped colonies and stained positively for AP, SETD5 knockdown impeded RA-induced differentiation of mouse ESCs (Figure 6E). These results indicate that SETD5 is essential for mouse ESC differentiation.

**Figure 6**



**Figure 6. *Setd5* is required for RA-induced differentiation of mouse ESCs.**

**(A)** Comparison of relative mRNA level of SETD5 in different stage of RA-induced mouse ESC differentiation. EB, 5 days after embryoid body generation; RA, 5 days after retinoic acid treatment. **(B)** Western blot analysis of SETD5 and  $\beta$ -actin (loading control) protein levels in different stages of RA-induced mouse ESC differentiation. **(C)** Analysis of relative mRNA expression levels between shLuc-treated and two SETD5-depleted (shSETD5-1 and shSETD5-5) mouse ESCs. **(D)** Western blot analysis of SETD5 and  $\beta$ -actin (loading control) protein levels between shLuc-treated and two SETD5-depleted (shSETD5-1 and shSETD5-5) mouse ESCs. **(E)** Microscopic and AP staining images of shLuc-treated and two SETD5-depleted (shSETD5-1 and shSETD5-5) mouse ESCs in different stages of RA-induced mouse ESC differentiation. AP staining was performed 5 days after retinoic acid treatment. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*)

### 2.3.2 SETD5 activates the expression of differentiation-specific genes and the neuron-related genes.

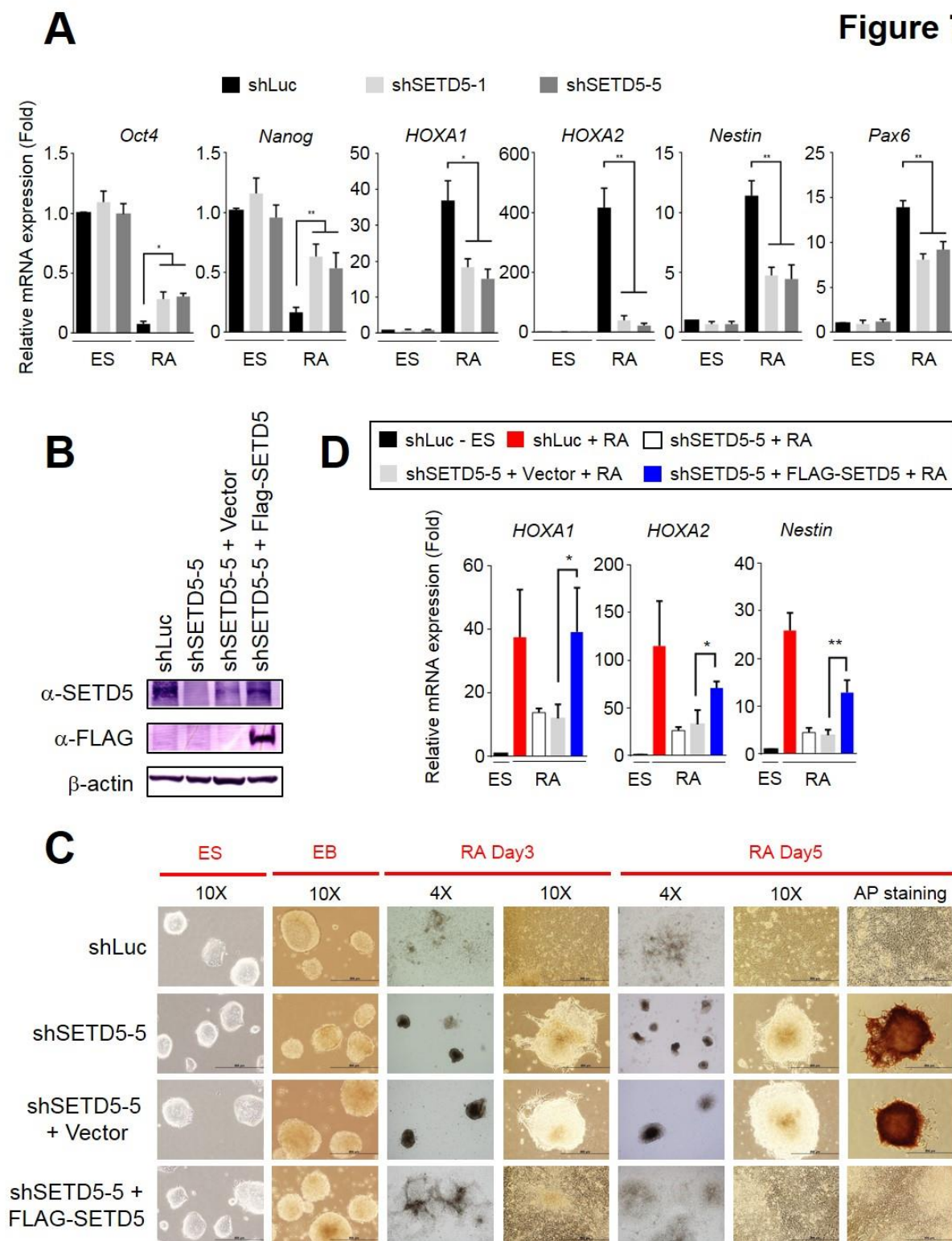
The differentiation process of mouse ESCs requires rapid down-regulation of pluripotent factors (e.g. *Oct4* and *Nanog*) and induction of lineage-specific genes to further differentiate into the precursors of different germ layers (14). The treatment of RA induces the neuronal lineage commitment of mouse ESC, which then differentiates into the ectodermal precursor cells (87). To understand how SETD5 regulates mouse ESC differentiation, we examined the effect of SETD5 depletion on the expression level of several pluripotent factors and neuronal lineage-specific markers during RA-induced differentiation of mouse ESCs. Quantitative RT-PCR results showed the mRNA levels of pluripotent factors *Oct4* and *Nanog* were less reduced in SETD5-depleted cells than in shLuc-treated cells after RA treatment (Figure 7A). These results are consistent with the above results that SETD5-depleted cells were positively stained with AP after RA treatment. Moreover, RA-induced activation of the differentiation-specific genes (*Hoxa1* and *Hoxa2*) and the neuron-related genes (*Nestin* and *Pax6*) was also inhibited by depletion of SETD5 (Figure 7A). Together, these data suggest that SETD5 depletion inhibited RA-induced differentiation of mouse ESCs.

To confirm the specificity of shSETD5 and the role of SETD5 in mouse ESC differentiation, we performed a rescue experiment by stably expressing mouse SETD5 in stably SETD5-depleted cells. We first cloned FLAG-tagged full length mouse SETD5 into a pCDH-EF1-IRES-Neo vector to obtain a FLAG-SETD5 expression construct (pCDH-EF1-FLAG-SETD5-IRES-Neo). Next, we transfected empty vector or pCDH-EF1-FLAG-SETD5-IRES-Neo into stably SETD5-depleted cells and selected with genitacin to obtain genitacin-resistant stable clones. For stably SETD5-depleted cells, we chose shSETD5-5-containing stable cells to perform a rescue experiment, because shSETD5-5 is predicted to target the 3'UTR of mouse *Setd5* and

thus may not have an effect on SETD5 expression. The stable expression of SETD5 was confirmed by anti-FLAG western blot analysis (**Figure 7B**). We compared RA-induced differentiation of the following 4 groups of ESC clones: (1) shLuc, (2) shSETD5-5, (3) shSETD5-5+Vector, and (4) shSETD5-5+FLAG-SETD5.

After RA treatment, the shSETD5-5+Vector group had an impaired differentiation morphology and was positively stained with AP. In contrast, the shSETD5-5+FLAG-SETD5 group showed almost normal differentiation morphologies similar to those of the shLuc group and were negatively stained with AP (**Figure 7C**). Consistent with phenotypic observation and AP staining, quantitative RT-PCR results showed that RA-induced expression levels of differentiation-specific genes (*Hoxa1* and *Hoxa2*) and the neuron-related genes (*Nestin* and *Pax6*) in the shSETD5-5+FLAG-SETD5 group were highly increased as compared to those in the shSETD5-5+Vector group while being similar to those in the shLuc group (**Figure 7D**). These results showed that the impaired RA-induced differentiation of SETD5-depleted cells can be rescued largely by SETD5 re-expression, confirming an essential role for SETD5 in proper mouse ESC differentiation.

**Figure 7**



**Figure 7. *Setd5* is necessary for RA-induced expression of differentiation-specific genes (*Hoxa1* and *Hoxa2*) and neuron-relevant genes (*Nestin* and *Pax6*).**

**(A)** Analysis of relative mRNA level of *Oct4*, *Nanog*, *Hoxa1*, *Hoxa2*, *Nestin* and *Pax6* between shLuc-treated and two SETD5-depleted (shSETD5-1 and shSETD5-5) mouse ESCs in different stage of RA-induced mouse ESC differentiation. **(B)** Ectopic expression of SETD5 for rescue experiment. SETD5-depleted cells (shSETD5-5) were transfected with pCDH-EF1-IRES-Neo vector (shSETD5-5+Vector) or pCDH-EF1-FLAG-SETD5-IRES-Neo (shSETD5-5+FLAG-SETD5) and selected with genitacin to obtain genitacin-resistant stable clones. The protein level of SETD5, FLAG and  $\beta$ -actin (loading control) were examined by western blot analysis. **(C)** Microscopic and AP staining images between shLuc-treated, SETD5-depleted (shSETD5-5), shSETD5-5+Vector and shSETD5-5+FLAG-SETD5 cells in different stage of RA-induced mouse ESC differentiation. AP staining was performed 5 days after retinoic acid treatment. **(D)** Analysis of relative mRNA level of *Hoxa1*, *Hoxa2*, *Nestin* and *Pax6* between shLuc-treated, SETD5-depleted (shSETD5-5), shSETD5-5+Vector and shSETD5-5+FLAG-SETD5 cells in different stage of RA-induced mouse ESC differentiation. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*).



### 2.3.3 Generation of *Setd5*-2XHA knock-in mouse ESCs by CRISPR-Cas9 system

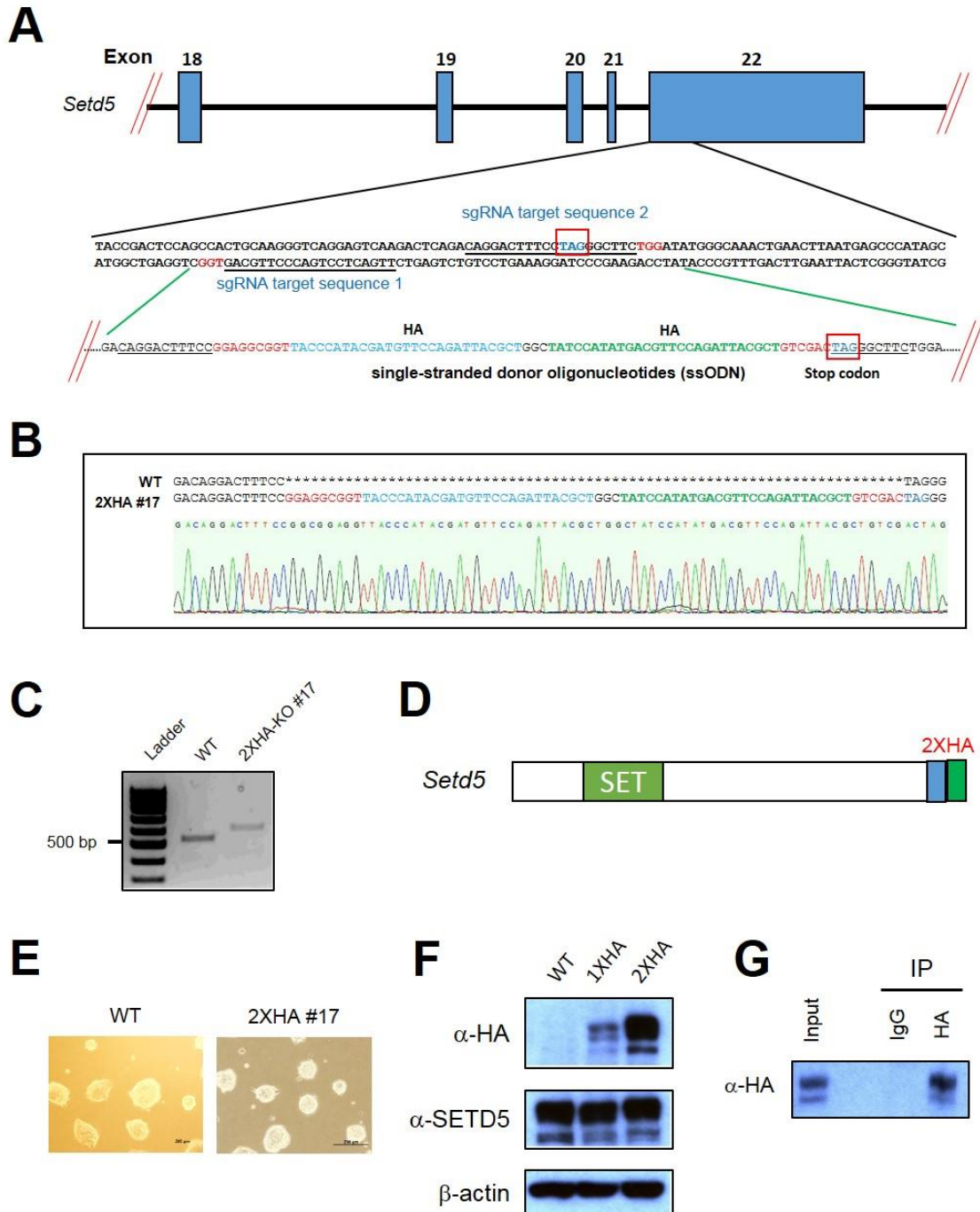
To assess whether SETD5 directly regulates *Hoxa1-2*, *Nestin*, and *Pax6* genes, we performed quantitative chromatin immunoprecipitation (qChIP) for SETD5 using commercially available SETD5 antibodies. However, we failed to obtain successful ChIP results in these genes, perhaps because of the poor quality of commercially available SETD5 antibodies. In the end, we decided to insert a tandem HA tag in the c-terminal end of *Setd5* gene (*Setd5*-2xHA) in V6.5 mouse ESCs, because an HA tag would allow us to use anti-HA antibody to perform qChIP assay for SETD5.

For the knock-in strategy, we took advantage of the recently developed CRISPR-Cas9 gene editing technology. Cas9 is an RNA-guided DNA endonuclease that can create specific DNA double strand breaks at target sites in living cells. We cloned two 20 bp-long single-guide RNAs (sgRNA) sequences into a Cas9 D10A nickase-GFP expression plasmid to target exon 22 of the mouse *Setd5* gene and to generate a DNA double strand break near the stop codon (Figure 8A). We also designed a repair template containing homologous arms flanking the insertion sequence so that the desired insertion at the target locus occurs in cells via homology-directed repair (HDR) (88). For the insertion sequence in the repair template, we placed the tandem HA tag at the C-terminal end of SETD5 (before the stop codon) to avoid interfering with endogenous SETD5 enzymatic activity because the SET domain of SETD5 is close to N-terminal (Figure 8D).

Next, we transfected sgRNA-Cas9-GFP plasmids along with single-stranded donor oligonucleotides (ssODN) as a repair template which contains 2XHA sequences into V6.5 mouse ESCs and performed GFP-based cell sorting to obtain Cas9-expressing cells (Figure 8A). Of the 84 clones screened, we were able to obtain a homozygous insertion clone (clone #17) which was confirmed by genomic PCR followed by DNA sequencing (Figure 8B and 8C). The SETD5-2XHA clone #17 was morphologically normal, because it formed round-shaped colonies and was

positively stained by AP (**Figure 8E**). The expression of SETD5-2XHA was further confirmed by immunoblotting and immunoprecipitation (**Figure 8F and 8G**). It should be noted that the insertion of 2XHA tag did not affect endogenous SETD5 protein level (**Figure 8F**).

**Figure 8**



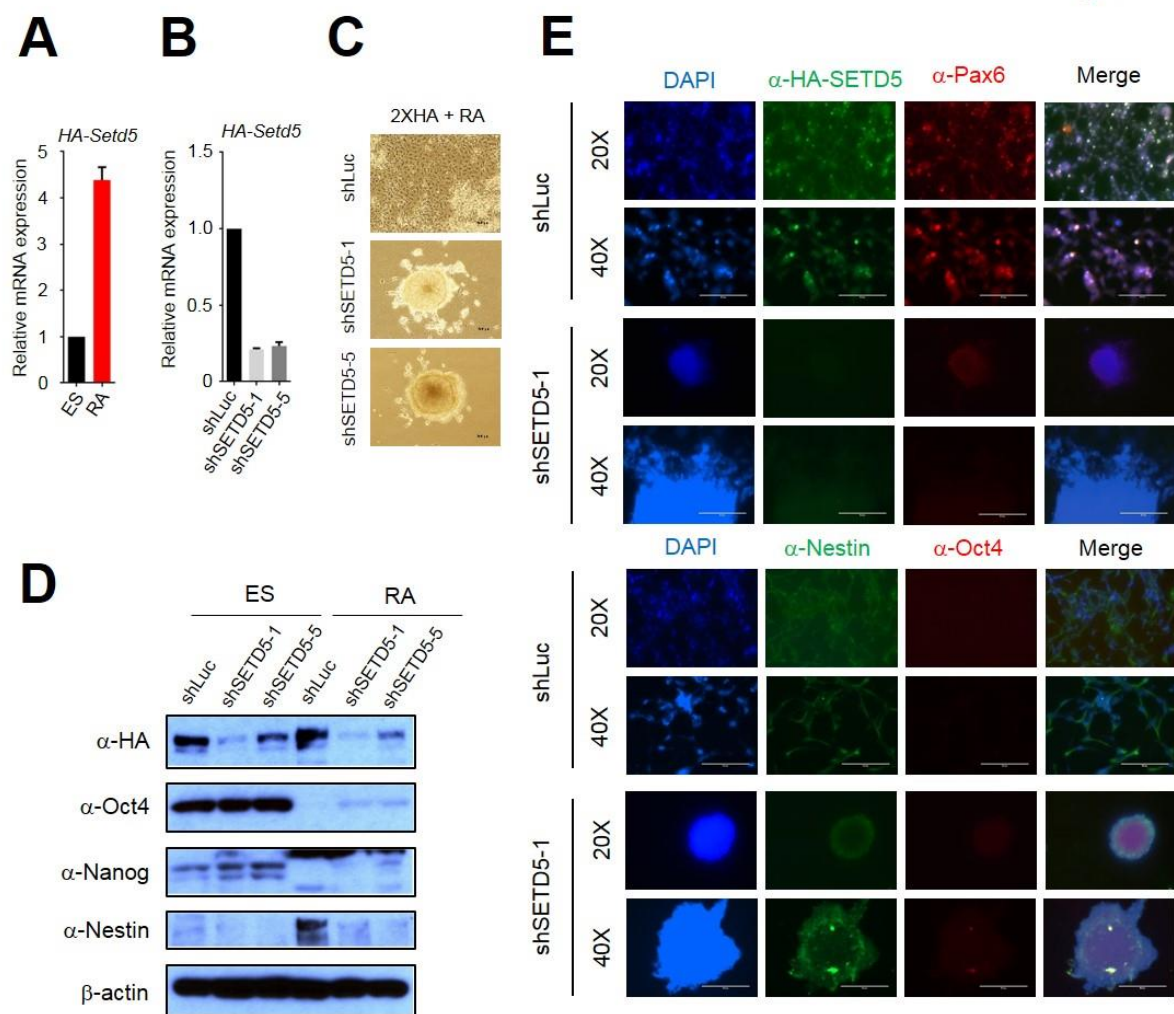
**Figure 8. Generation of Setd5-2XHA knock-in mouse ESCs by CRISPR-Cas9 system**

**(A)** Schematic representation the design of CRISPR/Cas9 targeting sites for generating *Setd5*-2XHA knock-in V6.5 mouse ESCs. Two guide RNAs (sgRNAs) sequences in the last exon (exon 22) of *Setd5* are labelled with underlines. The PAM sequences are labelled in red and the stop codon is labelled in blue. **(B)** Sequencing results and chromatograms for *Setd5*-2X-HA clone #17. **(C)** PCR using primers flanking the last exon of *Setd5* in wild-type and *Setd5*-2XHA knock-in V6.5 mouse ESCs. The wild-type PCR band is predicted to be 500 bp. *Setd5*-2X-HA clone #17 showed homozygous knock-in. **(D)** Schematic representation the protein product of *Setd5*-2X-HA knock-in. **(E)** Microscopic images of wild-type and *Setd5*-2X-HA clone #17 mouse ESCs. **(F)** Western blot analysis of HA, SETD5 and  $\beta$ -actin protein levels between wild-type and *Setd5*-2X-HA knock-in mouse ESCs. **(G)** Total cell lysate from *Setd5*-2X-HA knock-in mouse ESCs was immunoprecipitated by using IgG or HA antibody (Abcam), followed by western blot analysis with HA antibody (Biolegend).

#### **2.3.4 *Setd5* depletion also inhibits RA-induced differentiation of *Setd5*-2XHA knock-in mouse ESCs**

To assess whether cellular characteristics of *Setd5*-2XHA mouse ESCs are similar to those of wildtype mouse ESCs, we first examined *Setd5* levels during RA-induced differentiation of *Setd5*-2XHA mouse ESCs. Quantitative RT-PCR by using HA-specific primers showed the *Setd5* mRNA levels were increased after RA treatment (**Figure 9A**). We next examined the effect of SETD5 knockdown on RA-induced differentiation of SETD5-2XHA mouse ESCs. The knockdown of SETD5 was confirmed by quantitative RT-PCR using HA-specific primers (**Figure 9B**). RA-induced differentiation of *Setd5*-2XHA mouse ESCs were inhibited upon SETD5 depletion (**Figure 9C**). We further examined the protein level of pluripotent factors (OCT4 and NANOG) and neuronal markers (NESTIN and PAX6) by immunoblotting and immunofluorescence (**Figure 9E**). As in WT mouse ESCs, OCT4 and NANOG proteins levels in *Setd5*-2XHA mouse ESCs after RA treatment were less reduced in SETD5-depleted cells than in shLuc-treated cells (**Figures 9D and 9E**). The RA-mediated induction of NESTIN and PAX6 levels in *Setd5*-2XHA mouse ESCs, as in WT mouse ESCs, was reduced by SETD5 knockdown (**Figures 9D and 9E**). These results indicated that *Setd5*-2XHA mouse ESCs behave as WT mouse ESCs.

**Figure 9**



**Figure 9. SETD5 depletion also inhibits RA-induced differentiation of *Setd5*-2XHA knock-in mouse ESCs**

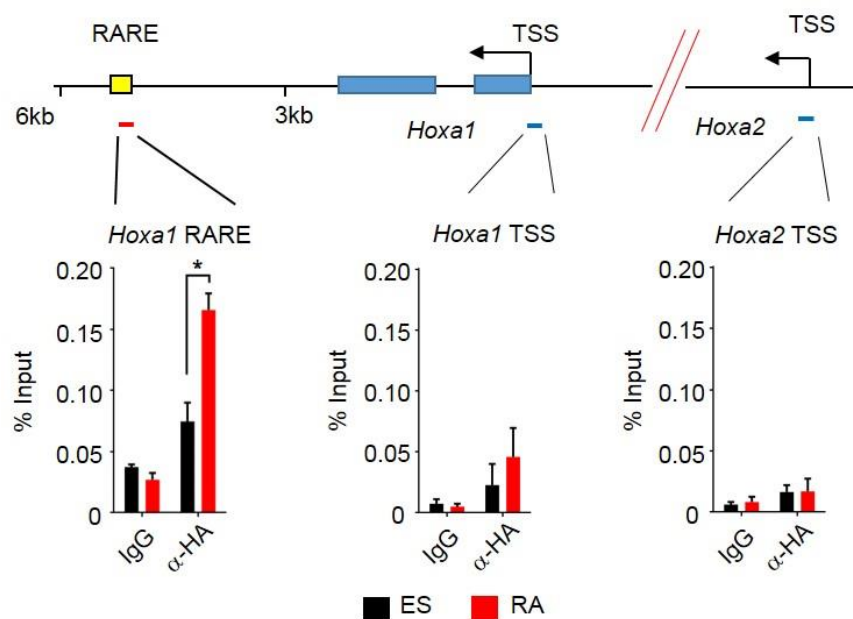
**(A)** Analysis of relative mRNA level of *Setd5* in Microscopic and AP staining images of shLuc-treated and two SETD5-depleted (shSETD5-1 and shSETD5-5) mouse ESCs in different stage of RA-induced mouse ESC differentiation. mouse ESCs by using HA-specific primers before (ES) and after RA-induced mouse ESC differentiation (RA). **(B)** Analysis of relative mRNA expression level of *Setd5* between shLuc-treated and two SETD5-depleted (shSETD5-1 and shSETD5-5) *Setd5*-2XHA knock-in mouse ESCs by using HA-specific primers. **(C)** Microscopic images of shLuc-treated and two SETD5-depleted (shSETD5-1 and shSETD5-5) *Setd5*-2XHA knock-in mouse ESCs after RA-induced mouse ESC differentiation. **(D)** Western blot analysis of HA-SETD5, Oct4, Nanog, NESTIN and  $\beta$ -actin protein levels between shLuc-treated and two SETD5-depleted (shSETD5-1 and shSETD5-5) *Setd5*-2XHA knock-in mouse ESCs before (ES) and after RA-induced mouse ESC differentiation (RA). **(E)** Immunofluorescence staining with DAPI, anti-HA, anti-PAX6, anti-NESTIN, and anti-OCT4 between shLuc-treated and SETD5-depleted (shSETD5-1) *Setd5*-2XHA knock-in mouse ESCs after RA-induced mouse ESC differentiation. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*)).

### 2.3.5 *Setd5*-2XHA was recruited to RA Responsive Element (RARE) after RA treatment

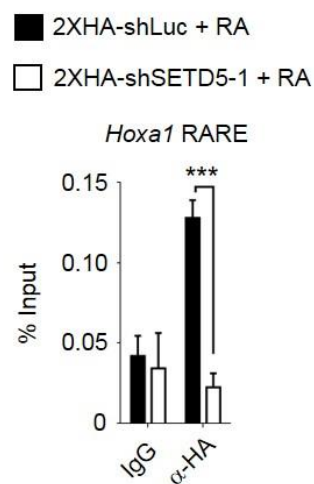
To determine whether SETD5 directly regulates RA-mediated induction of *Hoxa1-2* and *Nestin*, we performed quantitative ChIP using SETD5-2XHA mouse ESCs. However, our HA-ChIP results showed that SETD5 was not recruited to the promoter regions of *Hoxa1-2* or *Nestin* (Figure 10A). *Hox* genes are responsive to RA and several *Hox* genes contain RA response elements (RARE) (89-92). Moreover, Fujiki *et al.* showed that MLL5 can bind to RAR- $\alpha$  upon RA treatment (93). Therefore, we examined the possibility that SETD5 binds to RAR- $\alpha$  and is recruited to the *Hoxa1* RARE region after RA treatment. Our HA-ChIP results showed that SETD5 was recruited to *Hoxa1* RARE region upon RA treatment (Figure 10A). Moreover, SETD5 recruitment at the *Hoxa1* RARE region upon RA treatment was abolished by SETD5 depletion (Figure 10B). To assess whether SETD5 physically interacts with RAR- $\alpha$  upon RA treatment, we transfected an expression plasmid encoding FLAG-tagged RAR- $\alpha$  into HEK 293T cells and performed FLAG-immunoprecipitation (FLAG-IP) before and after RA treatment. Western blot analysis of the FLAG-IP eluates showed that RAR- $\alpha$  interacted with endogenous SETD5 upon RA treatment (Figure 10C).

**Figure 10**

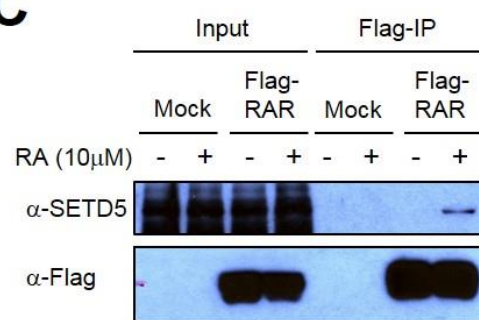
**A**



**B**



**C**





**Figure 10. *Setd5*-2XHA is recruited to RA responsive element (RARE) near *Hoxa1* after RA treatment**

**(A)** Analysis of SETD5 occupancy at RA responsive element near *Hoxa1*, *Hoxa1* promoter and *Hoxa2* promoter region by using anti-HA antibody to perform HA-ChIP quantitative PCR in *Setd5*-2XHA knock-in mouse ESCs before (ES) and after RA-induced mouse ESCs differentiation (RA). **(B)** Comparison of SETD5 occupancy at RA responsive element near *Hoxa1* between shLuc-treated and SETD5-depleted (shSETD5-1) *Setd5*-2XHA knock-in mouse ESCs by using ChIP qPCR after RA-induced mouse ESC differentiation. **(C)** FLAG-tagged RAR- $\alpha$  was expressed in HEK 293T cells. 48 h after transfection, DMSO or 10  $\mu$ M retinoic acid was added to cells and incubated for another 1 h. Total cell lysates were then harvested to perform FLAG-immunoprecipitation, followed by western blot analysis with anti-SETD5 and anti-FLAG. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*)).

### 2.3.6 SETD5 contains a SET domain but is catalytically inactive.

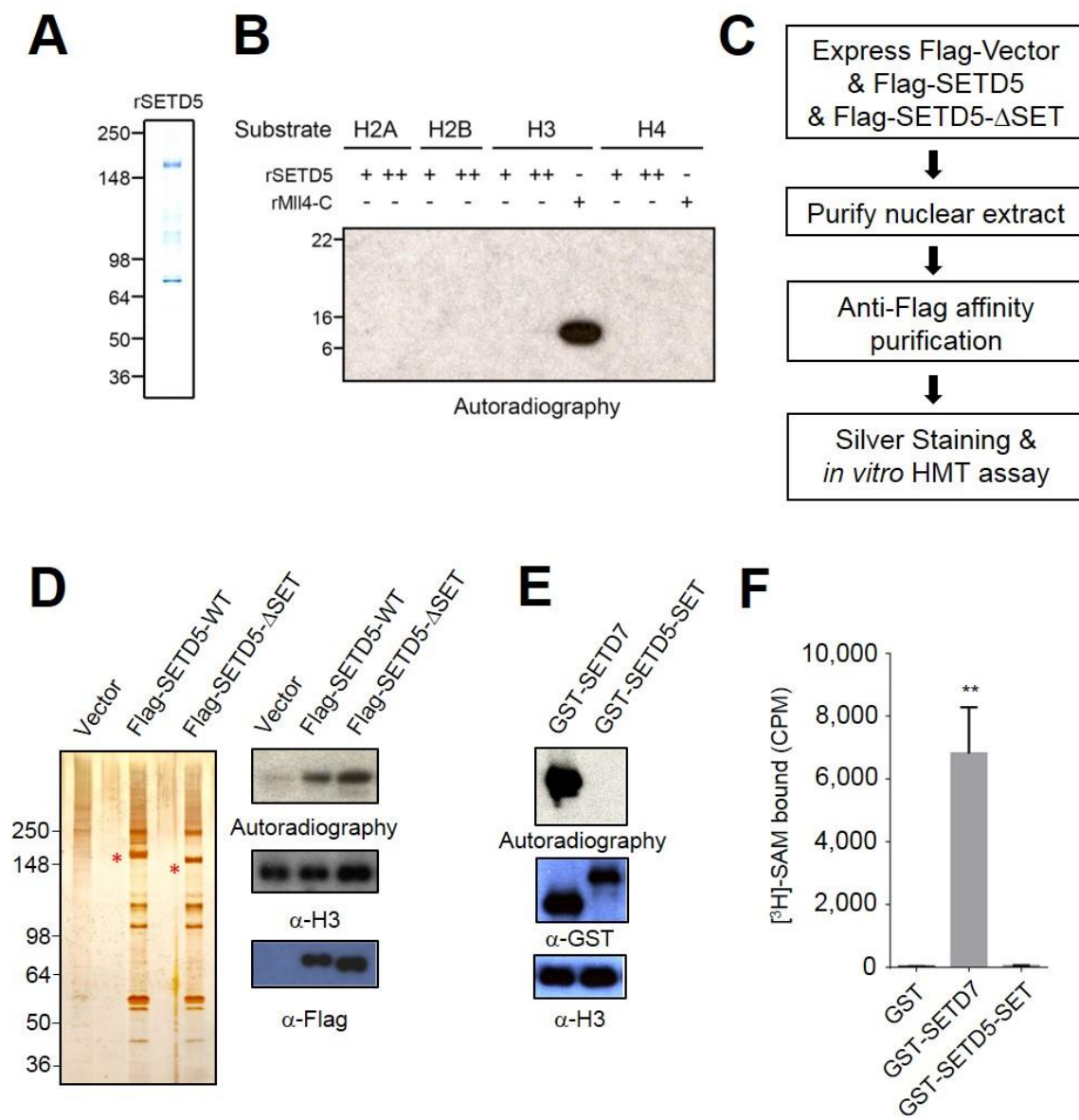
Although SETD5 contains one SET domain near its N-terminus, the catalytic activity of SETD5 has not been clearly defined. In an effort to further understand the role of SETD5 in RA-induced differentiation of mouse ESCs, we sought to determine the catalytic activity and the substrate of SETD5. To obtain recombinant SETD5 protein, we cloned FLAG-tagged full length SETD5 into a baculoviral expression plasmid, expressed FLAG-tagged SETD5 in Sf9 insect cells, and purified recombinant SETD5 protein using FLAG-immunoprecipitation (Figure 11A). We then performed an *in vitro* HMT assay using recombinant SETD5 and four different recombinant histone substrates (H2A, H2B, H3 and H4). In this experiment, recombinant MLL4-C protein that methylates H3 Lysine 4 was included as a positive control (83). The HMT assay results showed that the full length recombinant SETD5 protein did not have any detectable methyltransferase activity toward four histone substrates while recombinant MLL4-C robustly methylated recombinant H3 (Figure 11B).

Histone-modifying enzymes isolated from mammalian cells are often more active than their recombinant forms isolated from baculoviral expression system. One reason is that histone-modifying enzymes often associate with other proteins to form functional complexes in which some protein components play essential roles in regulating the catalytic activity or substrate recognition of the histone-modifying enzymes (27). For example, Ash2L, RBBP5 and WDR5 are the core components of the MLL protein complex that are indispensable for the full catalytic activity of MLL (94). The histone H3K27 methyltransferase EZH2 associates with Suz12 and EED to form the PRC2 complex, and the presence of Suz12 and EED are required for EZH2-mediated catalytic activity (95,96). To determine the possibility that SETD5 isolated from mammalian cells is catalytically active, we cloned mouse SETD5 into FLAG-tagged expression plasmid (FLAG-SETD5), expressed FLAG-SETD5 in HEK293T cells, and isolated FLAG-

SETD5 eluate using FLAG-IP (**Figure 11C**). We also generated a SET domain deletion mutant (FLAG-SETD5-ΔSET) as a negative control for the HMT assay and isolated FLAG-SETD5-ΔSET eluate (**Figure 11C**). Our *in vitro* HMT assay using recombinant nucleosomes as substrates showed that Flag-SETD5 and FLAG-SETD5-ΔSET eluates had similar HMT activities. These results indicate that SETD5 does not have a detectable HMT activities and that these HMT activities were not from the SET domain of SETD5 (**Figure 11D**).

It has been shown that the conserved NHS/CxxPN motif in SET domain is essential for the binding of SAM, which is critical for HMT activity. However, the sequence alignment showed that the NHS/CxxPN motif in SET domain of SETD5 is not conserved because the asparagine (N) and histidine (H) are replaced by two arginine residues (RR) (**Figure 3B**). To determine whether the SET domain of SETD5 interacts with SAM, we cloned the SET domain of SETD5 into a GST expression plasmid and purified GST-fusion protein (GST-SETD5-SET) from bacteria. We then performed a SAM binding assay by incubating [<sup>3</sup>H]-labeled-SAM with GST-fusion proteins. GST-SETD7 was included as a positive control that can also methylate recombinant H3 (97) (**Figure 11E**). The GST pull down assay of the incubated mixtures showed that GST-SET7 but not GST-SETD5-SET interacted with [<sup>3</sup>H]-labeled-SAM (**Figure 11F**). These results suggest that the non-conserved status of the NHS/CxxPN motif in the SET domain of SETD5 may negatively affect the recruitment of the cofactor SAM and thereby result in the inactivity of SETD5.

**Figure 11**



**Figure 11. SETD5 does not have detectable catalytic activity.**

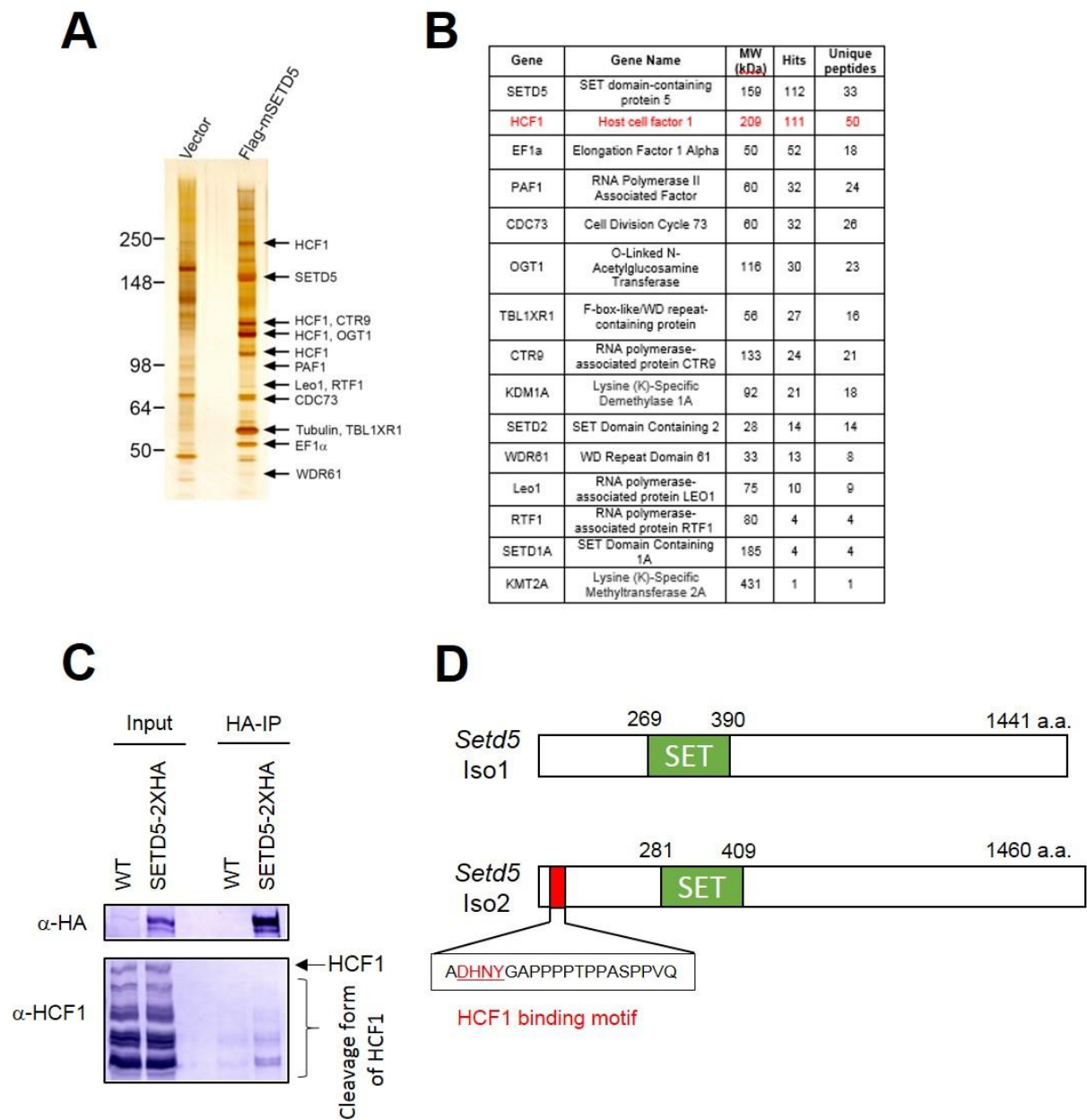
**(A)** Colloidal blue staining of full length recombinant SETD5 protein (rSETD5) isolated from SF21 insect cells. **(B)** *In vitro* HMT assay for full length recombinant SETD5 protein. Recombinant c-terminal end of MLL4 protein (rMLL4-C) was used as a positive control. Recombinant H2A, H2B, H3 and H4 were used as substrates. [<sup>3</sup>H]-SAM was used as methyl donor and the HMT activity was detected by autoradiography. **(C)** Flow chart of experimental design to isolate SETD5 complex. **(D)** Silver staining and *in vitro* HMT assay for SETD5 complex (FLAG-SETD5) and SET domain deletion mutant (FLAG-SETD5-ΔSET). Recombinant nucleosomes were used as substrates. [<sup>3</sup>H]-SAM was used as methyl donor and the HMT activity was detected by autoradiography. Antibodies against FLAG and H3 were used for western analysis as indicated. **(E)** *In vitro* HMT assay for SET domain of SETD5 (GST-SETD5-SET). GST-SETD7 was used as a positive control. Recombinant H3 was used as substrates. [<sup>3</sup>H]-SAM was used as methyl donor and the HMT activity was detected by autoradiography. Antibodies against GST and H3 were used for western analysis as indicated. **(F)** Radioactive SAM binding assay for SET domain of SETD5 (GST-SETD5-SET). GST-SETD7 was used as a positive control. The bound [<sup>3</sup>H]-SAM was detected by Liquid Scintillation Spectrometry.

### 2.3.7 SETD5 interacts with HCF1

In an effort to understand how SETD5 regulates RA-induced gene expression without its histone-modifying activity, we rationalized that SETD5-interacting proteins play a role in SETD5-mediated regulation of RA-induced gene expression, because FLAG-SETD5 IP eluates contain other protein bands (**Figure 11F**). Therefore, we examined the identity of SETD5-interacting proteins using mass spectrometry analysis (**Figure 12A and B**). Results showed that FLAG-SETD5 IP eluates contain host cell factor 1 (HCF1) in a one-to-one ratio (112 hits for SETD5 and 111 hits for HCF1). Consistent with a previous study by Osipovich *et al.*, mass spectrometric results also identified all six components of the PAF1 complex (PAF1, CDC73, CTR9, WDR51, LEO1 and RTF1) and O-linked N-acetylglucosamine transferase (OGT) (78). Interestingly, several histone methyltransferases SETD2, SETD1A and KMT2A were also present in the FLAG-SETD5 IP eluate (**Figure 12B**), suggesting that the HMT activity of the SETD5 complex might result from these enzymes.

We chose to focus on studying HCF1 because our mass spectrometric results identified HCF1 as a protein with the highest and most unique peptide numbers. HCF1 is a transcriptional coactivator that plays an important role in cell proliferation and cell cycle regulation (98-100). It has been shown that HCF1 interacts with several histone modifying enzymes, including the MLL1 complex, Sin3-HDAC complex, MLL5 and OGT (101-103). To validate the interaction between SETD5 and HCF1, we performed HA-immunoprecipitation (HA-IP) using whole cell lysates isolated from wildtype and SETD5-2XHA mouse ESCs. Western blot analysis of IP eluates showed that HCF1 was immunoprecipitated by SETD5-2XHA (**Figure 12C**). Moreover, analysis of the SETD5 protein sequence showed that SETD5 isoform 2 contains the HCF1 binding motif (HBM), (D/E)HxY (**Figure 12D**). These data confirmed that SETD5 interacts with HCF1 in mouse ESCs.

Figure 12



## Figure 12. SETD5 interacts with HCF1

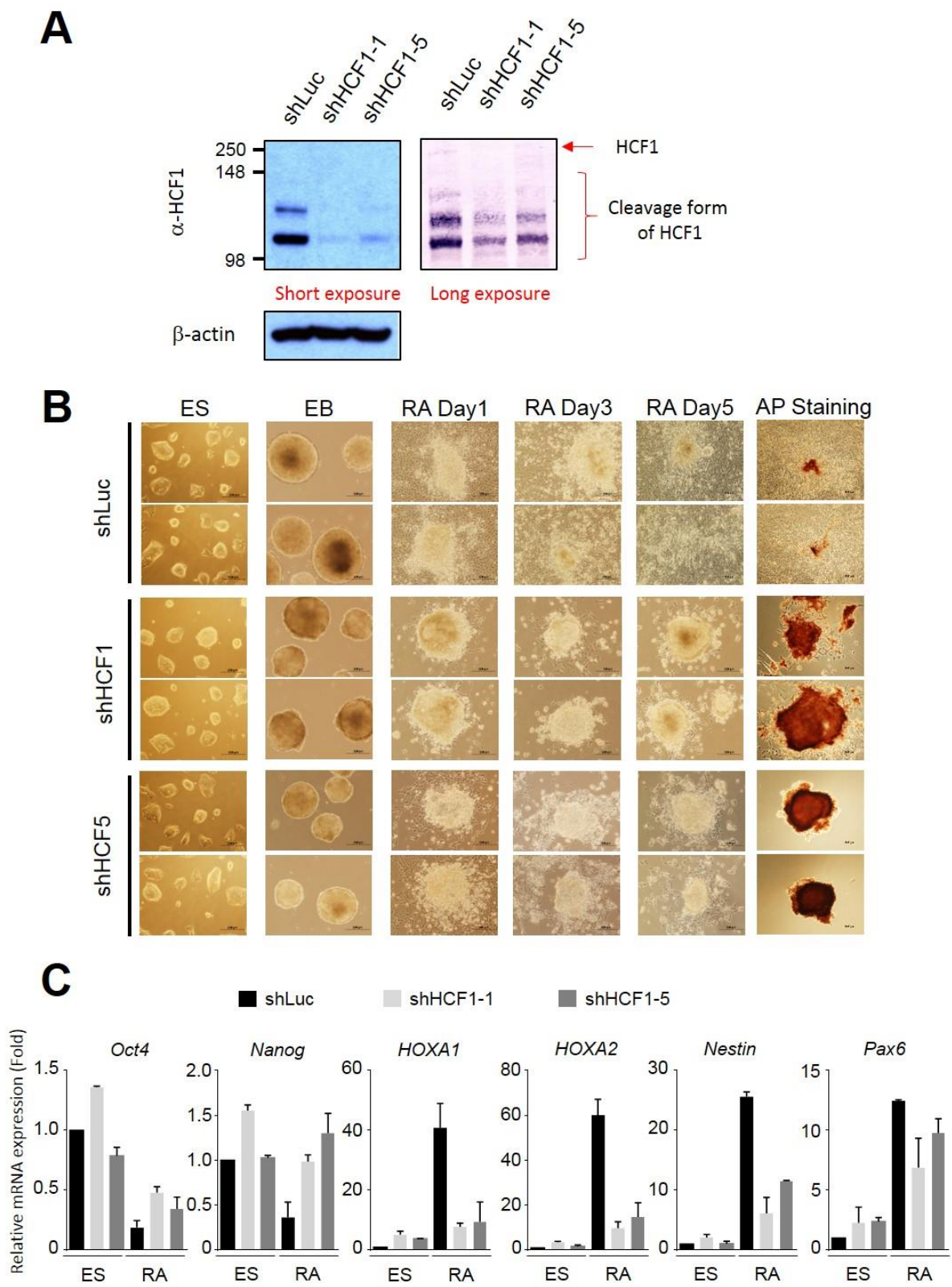
**(A)** Analysis of SETD5 complex by silver staining. Nuclear extracts were isolated from HEK293T cells expressing either empty vector or FLAG-tagged SETD5 followed by FLAG-IP. FLAG-IP eluate was further analyzed by mass spectrometry. The top ranked SETD5-interacting proteins were labelled as indicated. **(B)** SETD5-interacting proteins and numbers of peptides identified by mass spectrometry are listed. **(C)** HA-IP to confirm the endogenous interaction between SETD5 and HCF1. Total cell lysates from wildtype and *Setd5*-2XHA knock-in mouse ESCs were used to perform HA-IP. Antibodies against HA and HCF1 were used for western analysis as indicated. **(D)** Diagrammatic representation of mouse SETD5 isoform 1 and SETD5 isoform 2. SETD5 isoform 2 contains a 19 amino acid insertion at the N-terminal of SETD5. The HCF1 binding motif inside the 19 amino acid insertion is labelled in red.



### 2.3.8 Depletion of HCF1 inhibits RA-induced differentiation of mouse ESCs

To assess whether the SETD5-interacting protein HCF1 acts as a transcriptional coactivator of SETD5, we first examined the effect of HCF1 knockdown on RA-induced differentiation of mouse ESCs. HCF1 knockdown in V6.5 mouse ESC using two shRNAs against HCF1 (shHCF1-1 and shHCF1-5) were confirmed by Western blot analysis (Figure 13A). Interestingly, HCF1 knockdown, similar to SETD5 knockdown, impeded morphological changes during RA-induced differentiation of mouse ESCs, as also confirmed by AP staining (Figure 13B). In line with this, HCF1 depletion also inhibited RA-induced expression of SETD5 target genes, including *Hoxa1*, *Hoxa2*, *Nestin*, and *Pax6* (Figure 13C). These results indicate that HCF1 acts as a transcriptional co-activator of SETD5 in mouse ESCs.

Figure 13



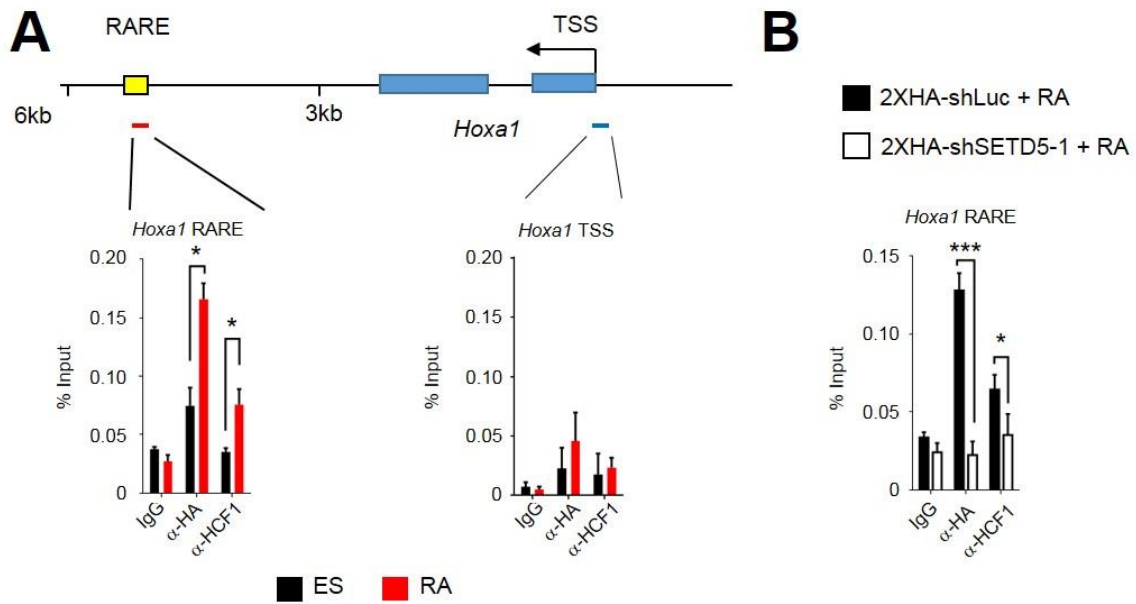
### **Figure 13. Depletion of HCF1 inhibits mouse ESC differentiation**

**(A)** Western blot analysis of HCF1 and  $\beta$ -actin (loading control) protein levels between shLuc-treated and two HCF1-depleted (shHCF1-1 and shHCF1-5) mouse ESCs. **(B)** Microscopic and AP staining images of shLuc-treated and two HCF1-depleted (shHCF1-1 and shHCF1-5) mouse ESCs in different stage of RA-induced mouse ESC differentiation. AP staining was performed 5 days after retinoic acid treatment. **(C)** Analysis of relative mRNA level of *Oct4*, *Nanog*, *Hoxa1*, *Hoxa2*, *Nestin* and *Pax6* between shLuc-treated and two HCF1-depleted (shHCF1-1 and shHCF1-5) mouse ESCs in different stage of RA-induced mouse ESC differentiation. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*)).

### **2.3.9 HCF1 is recruited to *Hoxa1* RARE region upon RA treatment**

Since SETD5 was recruited to the *Hoxa1* RARE region upon RA treatment, we assessed using ChIP assay whether HCF1 was also localized at the *Hoxa1* upon RA treatment. Our ChIP results showed that SETD5 and HCF1 were both recruited to the *Hoxa1* RARE region upon RA treatment (**Figure 14A**). Moreover, the recruitment of HCF1 at the *Hoxa1* RARE region was abolished by SETD5 depletion (**Figure 14B**). These results indicate that HCF1 is recruited to the *Hoxa1* RARE region in a SETD5-dependent manner during RA-induced differentiation of mouse ESCs.

**Figure 14**



**Figure 14. HCF1 is recruited by SETD5 to HOXA1 RARE region upon RA treatment**

**(A)** Analysis of SETD5 and HCF1 occupancy at RA responsive element near *Hoxa1* and *Hoxa1* promoter by using antibodies against HA and HCF1 to perform ChIP quantitative PCR in *Setd5*-2XHA knock-in mouse ESCs before (ES) and after RA-induced mouse ESCs differentiation (RA).

**(B)** Comparison of SETD5 and HCF1 occupancy at RA responsive element near *Hoxa1* between shLuc-treated and SETD5-depleted (shSETD5-1) *Setd5*-2XHA knock-in mouse ESCs by using ChIP qPCR after RA-induced mouse ESC differentiation. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*).

## 2.4 DISCUSSION

### 2.4.1 SETD5 and HCF1 in cellular differentiation

In current study, we showed that the depletion of SETD5 inhibited RA-induced differentiation of mouse ESCs. This is in line with the following previous studies that suggest a role for SETD5 in development: 1) SETD5 is highly expressed in brain tissue, suggesting a functional importance of SETD5 in central nervous system (72), 2) the loss of one copy of SETD5 is frequently included in deleted regions in 3p25 microdeletion syndrome, intellectual disability and autism (79-82,104-108), and 3) Osipovich *et al.* indicated that *Setd5*-null mice are embryonic lethal and that their embryos exhibit severe defects in neural tube formation (78). Our results showed that SETD5 is required for RA-driven mouse ESC differentiation and the RA-induced expression of developmentally important genes, such as *Hoxa1* and *Hoxa2*. Therefore, our findings may shed light on the molecular mechanism by which the clinical mutations of SETD5 cause developmental disorders.

We identified HCF1 as a major SETD5-interacting protein that can be recruited to the SETD5-occupied RARE region near *Hoxa1* upon RA treatment. Interestingly, our results showed that SETD5 and HCF1 depletion similarly caused defective morphological changes during RA-induced differentiation of mouse ESCs (**Figure 6E and 13B**). Moreover, the expression of several RA-induced genes, including *Hoxa1*, *Hoxa2*, *Nestin*, and *Pax6*, was commonly inhibited by SETD5 and HCF1 depletion (**Figure 7A and 13C**). Such similar effects between SETD5 and HCF1 knockdown on mouse ESC differentiation indicate that they have similar functions during mouse ESC differentiation. This is further supported by several clinical reports implicating that loss-of-function mutations in both SETD5 and HCF1 are linked to neurodevelopmental disorders (109-111). Our results, together with other reports, indicate that HCF1 acts as a transcriptional co-activator of SETD5 during mouse ESC differentiation.

#### 2.4.2 SETD5 is catalytically inactive but functional

Based on the phylogenetic analysis, Mas-y-Mas *et al.* proposed that MLL5, SETD5, the yeast protein SET3 and the *Drosophila* protein UpSET should belong to the same subfamily of SET domain-containing proteins (75). In fact, their SET domains are highly conserved although they are distinct from SET domains in other MLL family members, and the asparagine (N) and histidine (H) residues in NHS/CxxPN motif in the SET domains are all replaced by two arginines (RR). In the current study, our *in vitro* HMT assay using purified recombinant SETD5 and SETD5 complex showed no intrinsic catalytic activity for SETD5. Moreover, our sequence analysis and SAM binding assay indicate that the lack of a conserved NHS/CxxPN motif in the SET domain of SETD5 may cause its inability to bind cofactor SAM. These results are in line with several previous studies showing that MLL5, SET3, and UpSET are all catalytically inactive (75,112,113).

Although SETD5 and other family members have no intrinsic catalytic activity, several functional studies still indicate their roles in regulating development. For example, the knockout mice of MLL5 showed partial lethality and exhibited multiple hematopoietic defects (76). SETD5 knockout mice are embryonic lethal (78). Deletion of *set3* in yeast results in reduced ascus formation (112). Flies with UpSET homozygous mutation are viable but females with mutations are sterile because of severe defects in oogenesis (113). It is likely that developmental gene regulation by MLL5, SET3, and UpSET, despite their lack of catalytic activity, can be mediated by protein-protein interactions. Indeed, MLL5 associates with HCF1 to regulate E2F1-responsive gene expression (103). It was also reported that MLL5 regulates H3K4 methylation status through inhibiting the H3K4 demethylase LSD1 to alter target gene expression (76,77). Both *set3* and UpSET have been shown to form protein complexes with histone deacetylases to regulate gene expression (112,113). Our results showed that SETD5 interacts with HCF1 and that both proteins co-activate SETD5 target genes. In addition, our MS analysis showed that SETD5 co-

immunoprecipitates with two H3K4 methyltransferases, SETD1A and KMT2A (MLL1) (Figure 12B). The interaction between SETD5 and these H3K4 methyltransferases may be either direct or mediated by HCF1 because HCF1 has been shown to interact with several histone modifying enzymes, including SETD1A and MLL1 complexes (101,102). Nevertheless, it is possible that SETD5 indirectly affects histone methylation status, because the purified SETD5 complex (FLAG-SETD5 IP eluate) showed *in vitro* HMT activity toward recombinant nucleosomes (Figure 11F) and the FLAG-SETD5-ΔSET (SET deletion) IP eluates also showed similar HMT activity to that of the wildtype SETD5 complex (Figure 11F). Furthermore, we showed that SETD5 interacts with RAR-α after RA treatment and recruits HCF1 to the *Hoxa1* RARE (Figure 14). Taken together, our findings suggest that SETD5 functions as an adaptor protein to regulate gene expression via the interaction with HCF1, RAR-α and other H3K4 methyltransferase, such as SETD1A and KMT2A.

### 2.4.3 SETD5 and neurodevelopmental disorders

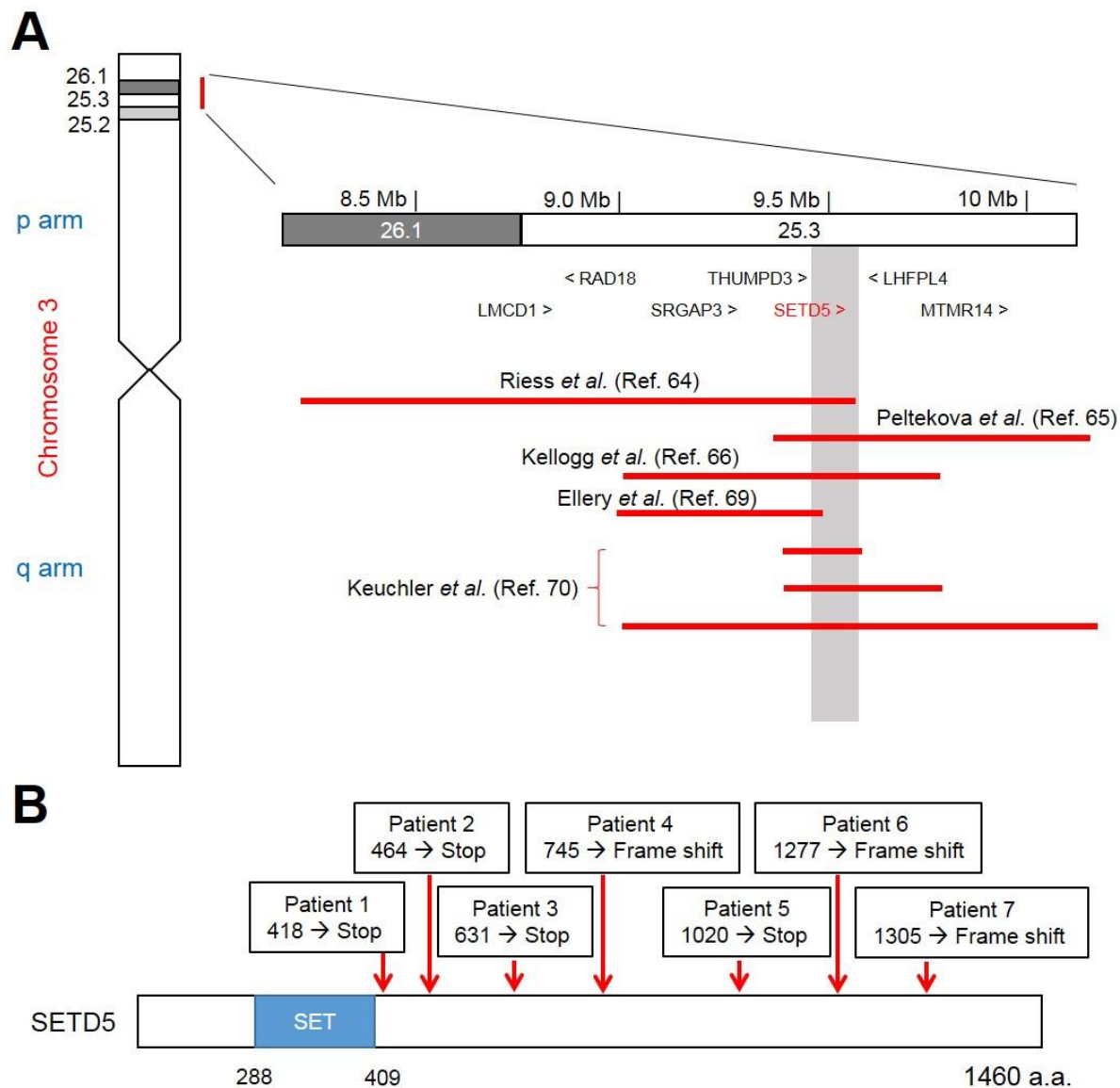
Several clinical reports showed the connection between SETD5 and 3p25 microdeletion syndrome (Figure 15A). 3p25 microdeletion syndrome is a rare chromosomal disorder that shows multiple symptoms including growth retardation, microcephaly, facial dysmorphia, seizures, congenital heart disease, and intellectual disability. Individuals with 3p25 microdeletion have one intact allele of chromosome 3 while the short arm of the other allele has an interstitial deletion (79,80). Recently, a clinical case study first suggested that the *SETD5* gene is located in the region of 3p25 microdeletion (81). Grozeva *et al.* further reported that *de novo* loss-of-function mutations in SETD5 may be linked to intellectual disability (82,104). In line with these studies, several other reports further indicate that SETD5 is a candidate gene to cause 3p25 microdeletion syndrome and intellectual disability (105-107). In addition, Pinto *et al.* has shown that both copy-number



variation (CNV) and single-nucleotide variation of *SETD5* occur in autism spectrum disorder (ASD) patients, suggesting SETD5 as an ASD candidate gene (108). Taken together, these studies strongly suggest that SETD5 plays an important role in regulating embryogenesis and neuronal development.

Most, if not all, SETD5 mutations lead to premature stop codons and generate truncated SETD5 proteins. Interestingly, such mutations are all located after the SET domain (**Figure 15B**). Therefore, truncated SETD5 proteins likely contain an intact SET domain, suggesting that the function of the SET domain of SETD5 is dispensable. It is possible that the truncated SETD5 may lose the ability to associate with some SETD5-interacting proteins, leading to dysregulation of SETD5 target genes and subsequently abnormal differentiation and development.

**Figure 15**



**Figure 15. De-novo loss-of-function mutations in SETD5**

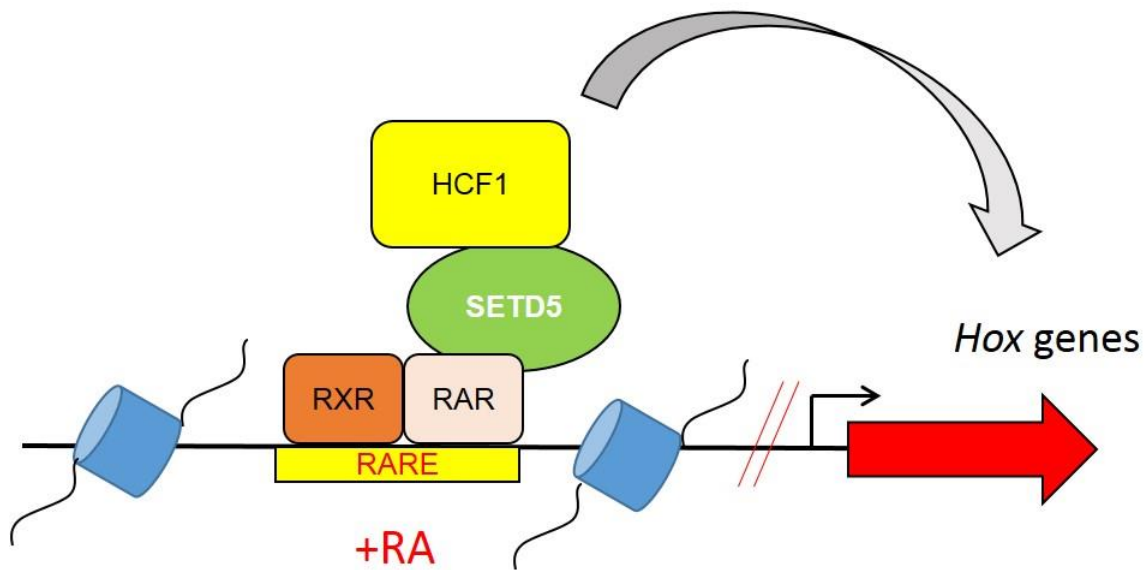
(A) SETD5 is located in the critical region of chromosome 3p25 microdeletion syndrome. Each red bar shows the interstitial deletion on the chromosome 3 from the individual with 3p25 microdeletion syndrome. The overlap region is highlighted by light grey. (B) Diagrammatic representation individuals with *de-novo* loss-of-function mutations in SETD5 reported by Grozeva *et al.* (82). The numbers indicate the amino acid position where the mutations occur.

#### 2.4.4 The regulation of Hox genes by SETD5

The Homeobox (*Hox*) genes exist in four genomic clusters (*HoxA* to *HoxD*) and encode transcriptional factors that play important roles in determining the anterior-posterior axis during embryogenesis and development (114). The expression pattern of *Hox* genes show spatial and temporal co-linearity (in 3' to 5' order) during embryogenesis. Specifically, the 3' end of *Hox* gene clusters are expressed in the anterior segments while the 5' end of *Hox* gene clusters are expressed in the posterior segments (115). For example, the 3' end of *Hox* genes, such as *Hoxa1*, *Hoxa2*, *Hoxb1* and *Hoxd1*, that are essential for neurogenesis of brain are located in the anterior segments (116). *Hox* gene expression is also regulated in ESC differentiation. *Hox* genes are transcriptionally repressed and undetectable in ESCs. Interestingly, most of the *Hox* genes clusters in ESCs are decorated simultaneously by so-called bivalent domains (repressed but activatable states) containing both active H3K4me3 marks and repressive H3K27me3 marks (117). During RA-induced ESC differentiation, the expression of *Hox* genes is highly induced and many bivalent *Hox* promoters lose H3K27me3 marks while being converted to active promoters containing H3K4me3 marks (85).

In the current study, we demonstrated that SETD5 is recruited to *Hoxa1* RARE region through the interaction with RAR- $\alpha$ . We also showed that the induction of *Hoxa1* and *Hoxa2* expression in mouse ESCs by RA was significantly inhibited by SETD5 knockdown. It has been shown that *Hoxa1* is required for RA-induced differentiation of ESCs into neurons (90). Moreover, the induction of ectoderm markers by RA are reduced in *Hoxa1* knockout ESCs (118). Therefore, it is possible that the dysregulated induction of *Hoxa1* and *Hoxa2* in SETD5 knockdown during RA-driven mouse ESC differentiation contributes to the impaired differentiation phenotype of SETD5-depleted cells (**Figure 16**).

**Figure 16**



**Figure 16. A hypothetical model for SETD5-mediated gene regulation**

A proposed model for SETD5-mediated gene regulation. During RA-induced mouse ESCs differentiation, SETD5 is recruited to the *Hox* genes RARE region through interaction with RAR- $\alpha$ . SETD5 further recruits transcriptional coactivator HCF1 to the *Hox* genes RARE region to activate the expression of *Hox* genes.

## 2.5 PERSPECTIVE AND FUTURE DIRECTIONS

CRISPR-Cas9 gene editing technology allowed us to generate *Setd5*-2XHA knock-in mouse ESCs. We further utilized this technology to generate *Setd5* knockout mouse ESCs to strengthen the results for our study. However, when we performed RA-induced mouse ESC differentiation, *Setd5* knockout mouse ESCs showed no significant defective differentiation phenotype (data not shown). These results are not consistent with our RNA interference experiment which showed that depletion of SETD5 inhibited the RA-induced mouse ESC differentiation. Interestingly, we observed that our *Setd5*-2XHA knock-in mouse ESCs showed multiple bands when we performed western by using HA antibody. This result indicates that there are multiple isoforms for SETD5 protein in V6.5 mouse ESCs. Our *Setd5* knockout mouse ESCs were generated by targeting 1<sup>st</sup> or 4<sup>th</sup> exons of *Setd5*. Therefore, we suspected that other functional SETD5 isoforms may still exist in *Setd5* knockout mouse ESCs which may explain why our *Setd5* knockout mouse ESCs showed no defective differentiation phenotype. We can further generate *Setd5* knockout mouse ESCs by targeting two separated exons and perform differentiation assays to examine the phenotype and to further determine the functions of other SETD5 isoforms and their expression in mouse ESCs.

**CHAPTER 3**

**PRMT7-MEDIATED REPRESSION OF *MIR-221* IS  
REQUIRED FOR MAINTAINING OCT4, NANOG AND  
SOX2 LEVELS IN MOUSE EMBRYONIC STEM CELLS**

### 3.1 INTRODUCTION

#### 3.1.1 General background of microRNA

After the completion of human genome project, researchers were surprised that there are only approximately 22,333 protein-coding genes in entire human genome (119,120). The rest of DNA sequences were non-protein coding and were once considered junk sequences. In 1993, Dr. Rosalind Lee and colleagues discovered a *C. elegans* gene, *lin-4*, does not encode protein but can generate a pair of functional small RNAs (121). They further showed that these small RNAs have antisense complimentary to 3'UTR of *lin-14* gene and can negatively regulate *lin-14* expression. We now know that these small RNAs belong to a class of non-coding RNA called microRNA (miRNA). They are single-stranded RNA which are normally 20-24 nucleotide in length and can negatively regulate gene expression through post-transcriptional mechanisms (122).

MicroRNAs will go through a serial maturation process to become single-stranded functional RNA (122,123). After transcription by RNA polymerase, miRNA will first form a stem loop structure called primary microRNA transcripts (pri-miRNA) (124). Pri-miRNA will further be cleaved by the Drosha/DGCR8 microprocessor complex in the nucleus to form precursor miRNA (pre-miRNA) of approximately 70 nucleotides in length (125). The cleavage by Drosha/DGCR8 microprocessor complex creates a 2- nucleotide overhang in 3' end of the pre-miRNA hairpin structure, which can be recognized by Exportin-5 and result in the nuclear exportation of pre-miRNA to the cytosol (126,127). In the cytosol, Dicer can cleave the pre-miRNA hairpin structure into a mature miRNA duplex (128). Dicer, together with TRBP and Argonaute-2, uses mature miRNA to induce gene silencing by cleaving mRNAs (129).

MicroRNAs can negatively regulate gene expression by interacting with 3' untranslated regions (UTR) of target mRNA (130). This interaction is dependent on the complementary pairing of miRNA with its target mRNAs. In particular, several studies have shown that the seed sequence

of miRNA (nucleotides 2-7 from 5' end of miRNAs) is important for pairing with its target mRNAs (131,132). Mature miRNA can lead to post-transcriptional gene repression through Argonaute-2-mediated mRNA degradation, translational repression, and mRNA deadenylation (129,133-135).

### **3.1.2 MicroRNAs regulate ESC stemness and differentiation**

To understand the biological function of miRNA pathways, knockout mice for the genes involved in the biogenesis of miRNA have been generated. The knockout mice for Dicer, DGCR8, and Argonaute-2 are all embryonic lethal (136-138). The Dicer- and DGCR8-null mouse ESCs are both viable but have severe defects in proliferation and differentiation (137,139). These studies suggest that proteins associated with miRNA processing play important roles in early embryogenesis, development, and ESC maintenance.

miRNA-mediated regulation of stemness and differentiation often results from changes in miRNA levels between ESC state and differentiated state (140,141). Several miRNAs are identified as ESC-specific miRNAs that are highly expressed in ESC and significantly decreased during differentiation. For example, the mouse miR-290-295 family (human homolog miR-371-373) is the most abundant miRNAs in mouse ESCs (142). These miRNAs have been shown to target the cell cycle inhibitor p21 (Cdkn1a) to promote G1/S cell cycle transition in ESCs (143). This may partially explain why ESCs can rapidly proliferate. In contrast to ESC-specific miRNAs, there are groups of differentiation-associated miRNAs that contribute to ESC differentiation and early development processes by targeting the pluripotent factor network. For example, miR-27a, miR-24, and let-7 have been identified as differentiation-associated miRNAs that are induced during ESC differentiation and directly target pluripotent factors (e.g., Oct4 and Foxo1) or signal transducers (gp130 and smad) to inhibit ESC pluripotency (144).



The expression patterns of miRNAs can be highly cell type-specific and are important in regulating cellular differentiation and development (145-147). It has been shown that the expression of miRNAs is regulated by several different mechanisms, including transcriptional factors, epigenetic regulation, and post-transcriptional regulation (148,149). Dysregulation of miRNAs can lead to cancer or other human diseases (150,151). For example, the repression of miRNA expression by DNA hypermethylation at certain *miRNA* promoters (e.g. miR-124, miR-34, miR-9 and miR-200 family) has been shown in various type of cancers (152,153).

### 3.1.3 General background of protein arginine methylation

Protein arginine methylation is a post-translational modification mediated by a family of enzymes called protein arginine methyltransferases (PRMTs) and has been linked to various cellular processes, including signal transduction, DNA repair, RNA processing, transcriptional regulation, and subcellular transport (154). Three different types of arginine methylation occurs on arginine residues in eukaryotes:  $N^G$ -monomethylarginine (MMA),  $\omega$ - $N^G$ ,  $N^G$ -dimethylarginine (asymmetric DMA), and  $\omega$ - $N^G$ ,  $N^G$ -dimethylarginine (symmetric DMA) (32,154). **(Figure 1B)** Arginine is a positively-charged amino acid. The addition of methyl groups on the arginine side chain does not change the charge but increases the hydrophobicity of the arginine residue and creates a binding motif for protein-protein interaction (154-156). For example, the tudor domain can specifically bind to methylated arginine through their aromatic cage. Several tudor domain-containing proteins, including SND1 and TDRD1, have been shown to interact with methyl arginine and to serve as reader proteins to regulate downstream cellular process (157).

### 3.1.4 Protein arginine methyltransferase

At least nine protein arginine methyltransferases (PRMT1-9) have been identified in mammals (32). These enzymes can all use S-adenosyl methionine (AdoMet) as a cofactor and transfer methyl groups from AdoMet to the arginine residue of target protein (158). PRMTs can be further classified into three different subtypes depending on their arginine methylation products. Type I PRMTs, including PRMT1, 2, 3, 6, 8 and PRMT4/CARM1, catalyze arginine mono-methylation and asymmetric di-methylation. Type II PRMTs, such as PRMT5 and PRMT9, catalyze arginine mono-methylation and symmetric di-methylation. Type III PRMT only catalyzes arginine mono-methylation and PRMT7 is the sole member of this type (32,154,159). In addition, we and others additionally showed that PRMT7 represses gene expression by indirectly establishing symmetric dimethylation at H4R3 (H4R3me2s) (83,85,160,161).

### 3.1.5 Rationale

We previously reported that protein arginine methyltransferase 7 (PRMT7), a transcriptional co-repressor, is essential for maintaining mouse ESC stemness. In the same study, we showed that PRMT7 maintains mouse ESC stemness, at least in part, by repressing the expression of the *miR-24-2* gene (i.e., miR-24-3p and miR-24-2-5p) (85). We also characterized miR-24-3p and miR-24-2-5p as anti-stemness miRNAs that directly inhibit the expression of the major pluripotent factors *Oct4*, *Nanog*, *Sox2*, *Klf4*, and *c-Myc*. We further showed that miR-24-3p and miR-24-2-5p levels are increased during mouse ESC differentiation and are highly upregulated by PRMT7 knockdown (85). In our effort to discover new anti-stemness miRNAs, we re-analyzed our previous miRNA expression profile data to determine which miRNAs in mouse ESCs are highly upregulated by PRMT7 knockdown. In the present study, we found that miR-221-3p and miR-221-5p are novel anti-stemness miRNAs that target the 3' untranslated

regions (3'UTRs) of the major pluripotent factors *Oct4*, *Nanog*, and *Sox2*. Our results also revealed that negative regulation of miR-221-3p and miR-221-5p by PRMT7 is necessary to maintain ESC pluripotency.

## **3.2 MATERIALS AND METHODS**

### **3.2.1 Antibodies, plasmids, and other reagents**

Anti-PRMT7 antibodies were purchased from Santa Cruz Biotechnology (SC9882). Anti-Nanog (#61419), anti-Sox2 (#39823), and anti-H4R3me2s (#61187) antibodies were from Active Motif. Anti-Oct4 (#2840), anti-c-Myc (#5605), and anti-Klf4 (#4038) antibodies were from Cell Signaling. Anti- $\beta$ -actin antibody (A5441) was from Sigma-Aldrich. Anti-H4R3me1 antibody (PA5-27065) was from Thermo Scientific. Anti-H3 antibody (ab1791) was from Abcam. Mouse shPRMT7s (shPRMT7-7, TRCN0000097477; shPRMT7-8, TRCN0000097478) in the puromycin-resistant PLKO.1 vector were previously reported from this laboratory (85). Oligonucleotides used for site-directed mutagenesis, RT-PCR, ChIP-PCR and CRISPR-Cas9 guide RNAs are listed in [Table 2](#).

### **3.2.2 Quantitative PCR for miRNA and mRNA expression**

Total RNAs were isolated by using TRIzol RNA Isolation Reagents (Life technology). To detect relative mRNA levels, cDNA was synthesized by using iScript cDNA Synthesis Kit (BioRad) according to manufacturer's instruction. Quantitative PCR was performed by using CFX384 real-time PCR detection system (BioRad). GAPDH level were used as internal control. For miRNA detection, qScript microRNA cDNA synthesis Kit (Quanta Bio) was used to synthesize microRNA cDNA. Briefly, total RNAs (1  $\mu$ g) were used in a poly(A) polymerase reaction to add Poly(A) tail to miRNA. Polyadenylated miRNAs were further reverse-transcribed with qScript reverse transcriptase to synthesize miRNA cDNA. Quantitative PCR was performed and normalized to sno66 to determine relative miRNA levels.

Table 2 *miR-221* Oligonucleotide sequences.

Table 2

		5'-3' Forward Sequence	5'-3' Reverse Sequence
	L-m. <i>Oct4-3p</i> (Mut-01)	AAGACTGGGACACACAGTAGAATTAAGAATTTTGTTCCTTCAGTTC	GGAACTGAAGGAAAACAAATTTCTTAATTTCTACTGTGTGCCAGTCTT
	L-m. <i>Oct4-3p</i> (Mut-02)	ATTTTATTTATATGGGTACACAAATTAAGTCACCTTCTAGACACACCAG	CTGGTGTGTCTAGAAGGTGACTTAATTTGTGTACCCATATAAATAAAAT
	L-m. <i>Oct4-5p</i> (Mut-01)	GCCAGTAATGGATTCTCAAACCTTTTAAGATCTTCAAAACAGGCGCCAT	ATGGCGCCTGTTTGAAGATCTTAAAGGTTTGAGAATCCATTACTGGC
	L-m. <i>Oct4-5p</i> (Mut-02)	GGGTCTGTACACATGTCTACAATAATTAATAGGTCAAAGGCTTTAGAT	ATCTAAAGCCTTTGACCTATTAATTTATGTAGACATGTGTACAGACCC
	L-m. <i>Nanog-3p</i> (Mut-01)	TTTTAAAGATTTTATTATTCATTATATGTAAGTACACAAAAATGTCTTC	CTGACGCCCTCTTCTGGAGTGTCTGAAGACATTTTTTGTGTACTTACATA
	L-m. <i>Nanog-3p</i> (Mut-02)	AGACACTCCAGAAGAGGGCGTCAG	TAATGAATAAATAAATCTTTAAAA
	L-m. <i>Nanog-3p</i> (Mut-02)	GACAGAAGGACAGGAGTTTGAGGAAAAATCAGATATGCAATAAGTTCAA	GCCTTGAACCTATTGCATATCTGATTTTTCCTCAAACCTCGTCCCTCT
	L-m. <i>Nanog-5p</i> (Mut-01)	TACGCAACATCTGGGCTTAAAGTCAGGGCAAAAAAATTCCTTCCTCT	AAATATGAAAAATTTTGAAGAAGGAATTTTTTTTGGCCCTGACTTT
	L-m. <i>Nanog-5p</i> (Mut-02)	TCCAAATATTTTCATATT	AAGCCAGATGTTGCGTA
Mutation	L-m. <i>Sox2-3p</i> (Mut-01)	TGATGTTACCAGCAATGTCTGTATTTATTAACATGCAAGAAAAAATG	CATTTTTTCTTCTGCAGTTAATAAATACAGACATTGCTGGTAACATCA
	L-m. <i>Sox2-3p</i> (Mut-02)	AGTCAATTCTGAAAAATTTTCGAAATTAATAGCAGATGAGATACTGCGCC	GGCGCAGTATCTCATCTGCTATTAATTTTCGAAAAATTTTTCAGAAATGACT
	L-m. <i>Sox2-5p</i> (Mut-01)	TGAGATACTGCGCCAAATCAGAAATAATTTGGCTGGGTGTTTTCATT	AAAAATGAAAAACCAACCCAGCAATTTTCTGATTGGCGCAGTATCTC
	L-m. <i>Sox2-5p</i> (Mut-02)	GCAGCTGGTACACGCAAAAGAGAAATTTCAACAGACATATTACAGAA	TTCTGTAATATGTCTGTTGAAATTTTCTCTTTTTCGCTGTACCAGCTGC
	L-m. <i>Klf4-3p</i>	AGGCAAAACAAACCTAAGAACAGTTAAATTTTCCACAGTGCCTCCTGAG	CTCAGGAGGGCACTGTGAAAAATTTAACTGTTCTTAGGGTTGTTTGCCCT
	L-m. <i>Klf4-5p</i> (Mut-01)	CCCACGTAGTGGATGTGACCCACACAATTAAGAGAGAGTTCAATATT	AAATACTGAACCTCTCTTTAAATTTGTGTGGGTACATCCACTACGTGGG
	L-m. <i>Klf4-5p</i> (Mut-02)	TCCGATCTACATTGATGACCTAAAAATTTAAATAAGCCTGGTTTATTTTC	GAATAAAACAGGCTTATTTTAAATTTTTAGGTATCAATGTAGATCGGA
	L-m. <i>c-Myc-3p</i>	TTCCCATGTAAATAGGGCCTTGATTAATAATAACTTTAATAAAACGTT	AACGTTTTATTAAAGTTATTATTAACTCAAGGCCCTATTACATGGGAA
	L-m. <i>c-Myc-5p</i> (Mut-01)	CGTTTGTTCCTTGCCCTCTTGCTCAATAAAGATAGTCCCTTCACATCAGTA	GATACTGATGTGAAGGACTATCTTTATTGAGCAAGAGGCAAGGAACAAA
	L-m. <i>c-Myc-5p</i> (Mut-02)	GGTACGAACACACAAAAATAACAATAAGGCCCTGTCTTGAAGAACAGCC	AGGCTTGTTCCTCAAGACAGGGGCTTATTGTTATTTTGTGTGTTCTGAC
	L-m. <i>Prt7-3p</i> (Mut-01)	GCACCTGGGAAGAACTAGGAGAAATAACGCTCCTGTCTCTCTGTCTGT	GACAGACAGAGAGACAGGAGGCTTATTCTCCTAGTTTCTTCCAGTGC
	L-m. <i>Prt7-3p</i> (Mut-02)	GGCATGGCCTCAGTGCTGAGATGAATAAGGACCTGAAGGCAGGCAGGCT	CAGCCTGCCTGCCTTCAGGTCCTTATTCATCTCAGCACTGAGGCCATGC
	L-m. <i>Prt7-5p</i> (Mut-01)	GCCTATCTGCTTTCAAGGAAGTTAATAAACTCCAGGGAAAAATGGCCTCC	TGGGAGGCCATTTTCCCTGGGAGTTTATTAACTTCCTTGAAGCAGATAG
	L-m. <i>Prt7-5p</i> (Mut-02)	GATTGCTAAGCCCTGAGACCCAAATAAATACTGTGGTTTCAAGGCCGT	GATGACGGCCTTGTGAACCACAGTTATTTTATTGGGTCTCAGGGCTTAGC
	mGAPDH	CATGGCCTTCCGTGTTCTTA	GCCTGCTTACCACCTTCTT
	mPRMT7	CAGGTAGCGTCTGCCTTTGT	CAACTCAGGCCGTTTATTGATGA
	mNanog	TCTTCCTGGTCCCCACAGTTT	GCAAGAATAGTTCTCGGGATGAA
	mSox2	GCTCGCAGACCTACATGAAC	GCCTCGGACTTGACCACAG
	mOct4	AGAGGATCACCTTGGGGTACA	CGAAGCGACAGATGGTGGTC
RT-PCR	mCMyC	CACCATGCCCCCTCAACGTGAACCTTCACC	TTATGCACACAGATTTCGAAGCTGTTCG
	mKlf4	GTGCCCCGACTAACCGTTG	GTCGTTGAACCTCCTCGGTCT
	miR-221-3p	AGCTACATTGTCTGCTGGGTTTC	*PerfeCTa Universal PCR Primer
	miR-221-5p	ACCTGGCATAACAATGTAGATTTCTGT	*PerfeCTa Universal PCR Primer
	SNORD66	GTCAGTGCCACGTGTCTGGGCCACTGAGACCACATGATGGATTGAGGACCTGAGGAA	*PerfeCTa Universal PCR Primer
	mir221-S1	TCCCTTCCAGAACTCTCTTCA	GGTAATCAGCAGCTACATCTGG
ChIP	mir221-S2	ACTTCCATACATCTCTGTCTTTA	CACAATTCCTGTGACTTACCAATAG
	mir221-S3	GATTTCTGTGTTGTTAGGC	AAGAAATGATTCCAGTAGCC
CRISPR-Cas9	Cas9-mir221-F1	CACCGCTGCTGGGTTTCAGGCTACC	AAACGGTACCGCTGAAACCCAGCAGC
	Cas9-mir221-F2	CACCGAGAAATCTACATTGTATGCC	AAACGGCATACATGTAGATTTCTC
	Cas9-mir221-seq	TCACCTATTGGTAAGTCACA	AGCAGTATGCTAAGATGATTT

### **3.2.3 Chromatin Immunoprecipitation (ChIP) assay**

ChIP assay was performed as previously described with small modifications (83). Mouse ESCs were first fixed by 1% formaldehyde. Cell pellets were then lysed with ChIP lysis buffer and sonicated for 15 min (30 sec on and 30 sec off for 15 cycles) to shear DNA by using Bioruptor (Diagenode). Antibodies were added and incubated for overnight at 4 °C. Pre-blocked protein A beads were added and incubated for another 1-2 hr to capture protein-DNA complex. The beads were then washed once with the following buffers: low-salt buffer, high-salt buffer, LiCl buffer and TE buffer. ChIP DNA were then eluted by ChIP elution buffer (1% SDS and 0.1M NaHCO<sub>3</sub>). The eluate were reverse-crosslinked, and ChIP DNA were purify by phenol/chloroform extraction method.

### **3.2.4 Transfections of miRNA mimics and LNA oligonucleotides**

Mouse ESCs ( $5 \times 10^5$  cells) were trypsinized and transferred to 0.4 cm Gene Pulser Electroporation Cuvettes (Biorad). Mouse miR-221-3p mimic, miR-221-5p mimic (Ambion) or mouse LNA-miR-221-3p and LNA-miR-221-5p (Exiqon) (150 pmole) were added to the cuvettes. Electroporation (250 V and 500  $\mu$ F) were performed by using Gene Pulser Xcell Electroporation Systems (Biorad). After electroporation, cells were rested at room temperature for 10 min and seeded into 6 cm dishes.

### **3.2.5 Luciferase reporter assays**

The 3'UTRs of mouse *Oct4*, *Nanog*, *Sox2*, *Klf4*, *c-Myc* and *Prmt7* genes in pMIR-REPORT (Ambion) vector have been previously described (85). The predicted miR-221-3p and miR-221-5p target sites in the 3'UTRs were mutated by using QuickChange Site-Directed

Mutagenesis Kit (Stratagene). The pre-mature *miR-221* sequence was synthesized and cloned into retroviral-based miRNA expression vector (pMDH1-PKG-*miR-221*-GFP). To perform luciferase assay, pMIR-REPORT vector containing wild type (WT) or mutant 3'UTRs of *Oct4*, *Nanog*, *Sox2*, *Klf4*, *c-Myc* and *Prmt7* and pMDH1-PKG-*miR-221*-GFP were transfected into HEK293T cells with Renilla vector (Promega). After 48h incubation, transfected cells were harvested. Luciferase activities were measured using Dual-Luciferase Reporter assay system (Promega) according to the manufacturer's instructions. The firefly luciferase activities were normalized to Renilla luciferase activities.

### 3.2.6 CRISPR/Cas9 gene editing

pSpCas9n(BB)-2A-GFP (PX461) plasmid with nCas9 (D10A nickase mutant) was obtained from Addgene. To generate *miR-221*-null V6.5 mouse ESCs, two guide RNA (sgRNA) sequences that target *miR-221* (F1: CTGCTGGGTTTCAGGCTACC; F2: AGAAATCTACATTGTATGCC) were cloned separately into pSpCas9n(BB)-2A-GFP. sgRNA-containing plasmids were then transfected into V6.5 mouse ESCs by using Lipofectamine 3000 (Life technology) according to manufacturer's protocol. After 48 h incubation, mouse ESCs were trypsinized and sorted by GFP signals. GFP-positive mouse ESCs were plated into 96 well plate (one cell per well) to obtain single colony.

### 3.2.7 Statistical analysis

The statistical significance between the two groups was analyzed by student's t-test using Prism software (GraphPad Software Inc.). Data are presented as the mean  $\pm$  standard deviation (SD) of at least three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*) indicate statistically significant changes.

### 3.3 RESULTS

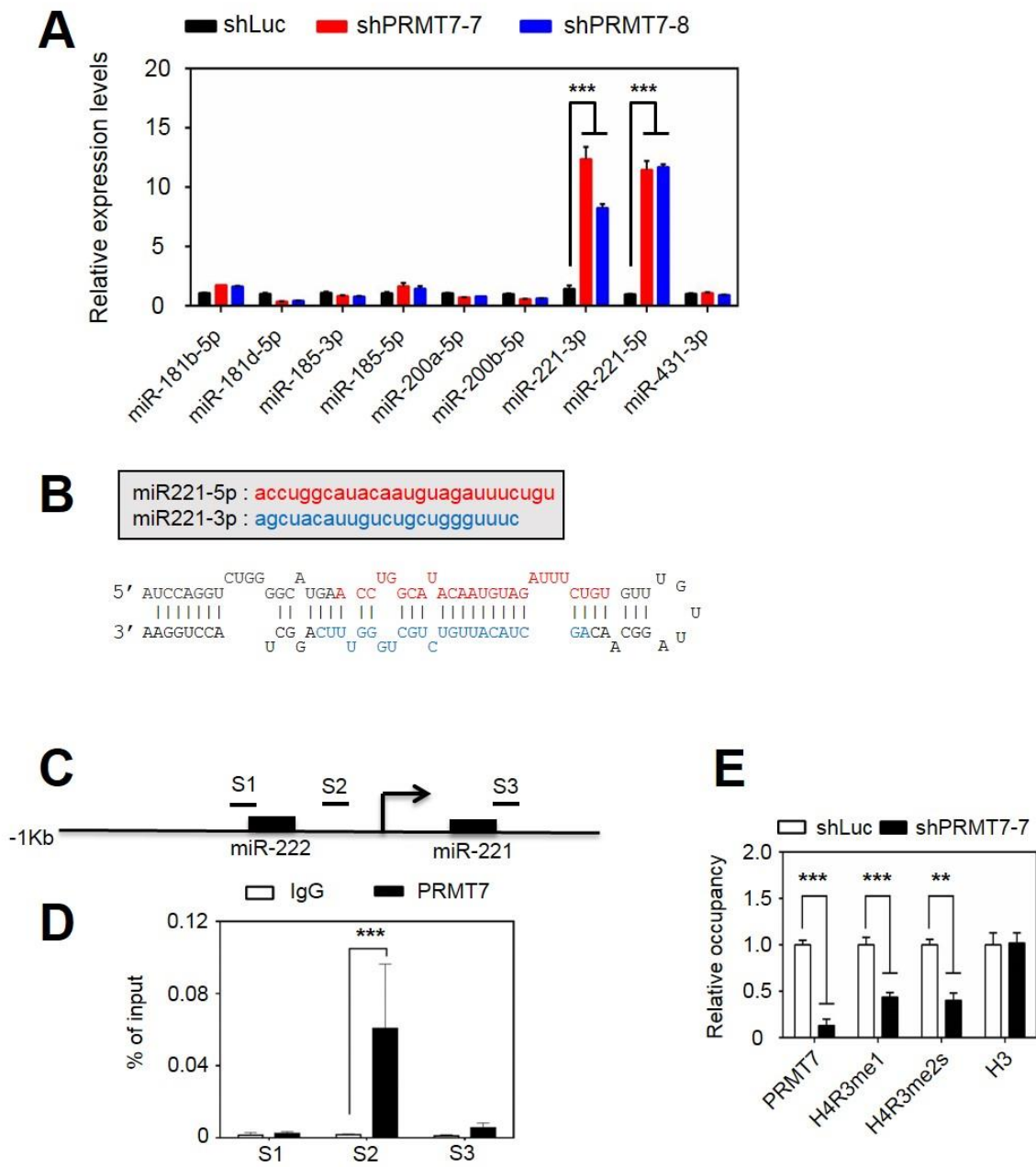
#### 3.3.1 The expression of miR-221-3p and miR-221-5p is directly repressed by PRMT7

In our effort to identify anti-stemness miRNAs, we searched for miRNAs in mouse ESCs that are highly upregulated by knockdown of the pluripotent factor PRMT7 using our previously reported microRNA microarray data (85). This search resulted in several candidate miRNAs, and miRNA-specific quantitative PCR results showed that miR-221-3p and miR-221-5p in V6.5 ESCs were highly up-regulated by two independent PRMT7 shRNAs (shPRMT7-7 and shPRMT7-8) (**Fig. 17A**). The two mature microRNAs, miR-221-3p and miR-221-5p, are from the stem loop structure of the precursor miRNA *miR-221* (**Fig. 17B**). The *miR-221* gene is located on the X chromosome and is overexpressed in many types of cancers (e.g. glioblastomas, breast cancer, hepatocellular carcinoma, thyroid papillary carcinoma, and lung cancer) (162). However, little is known about anti-stemness function of *miR-221*.

It has been generally accepted that PRMT7 monomethylates arginine residues, such as H4R3, for gene repression (32). We and others additionally showed that PRMT7 represses gene expression by indirectly establishing symmetric dimethylation at H4R3 (H4R3me2s) (83,85,160,161). To examine whether *miR-221* expression is directly repressed by PRMT7, we performed quantitative ChIP experiments. ChIP results showed that PRMT7 occupied a promoter region of the *miR-221* gene in V6.5 mouse ESCs (**Fig. 17C and 17D**). We also examined the effect of PRMT7 knockdown on monomethylated H4R3 (H4R3me1) and H4R3me2s levels at the *miR-221* promoter. Our results showed that H4R3me1 and H4R3me2s levels at the *miR-221* promoter were decreased by PRMT7 depletion (**Fig. 17E**). Together, these results indicate that PRMT7 represses the expression of *miR-221* by upregulating the repressive histone marks, such as H4R3me1 and H4R3me2s, at the *miR-221* promoter in V6.5 mouse ESCs.



### Figure 17



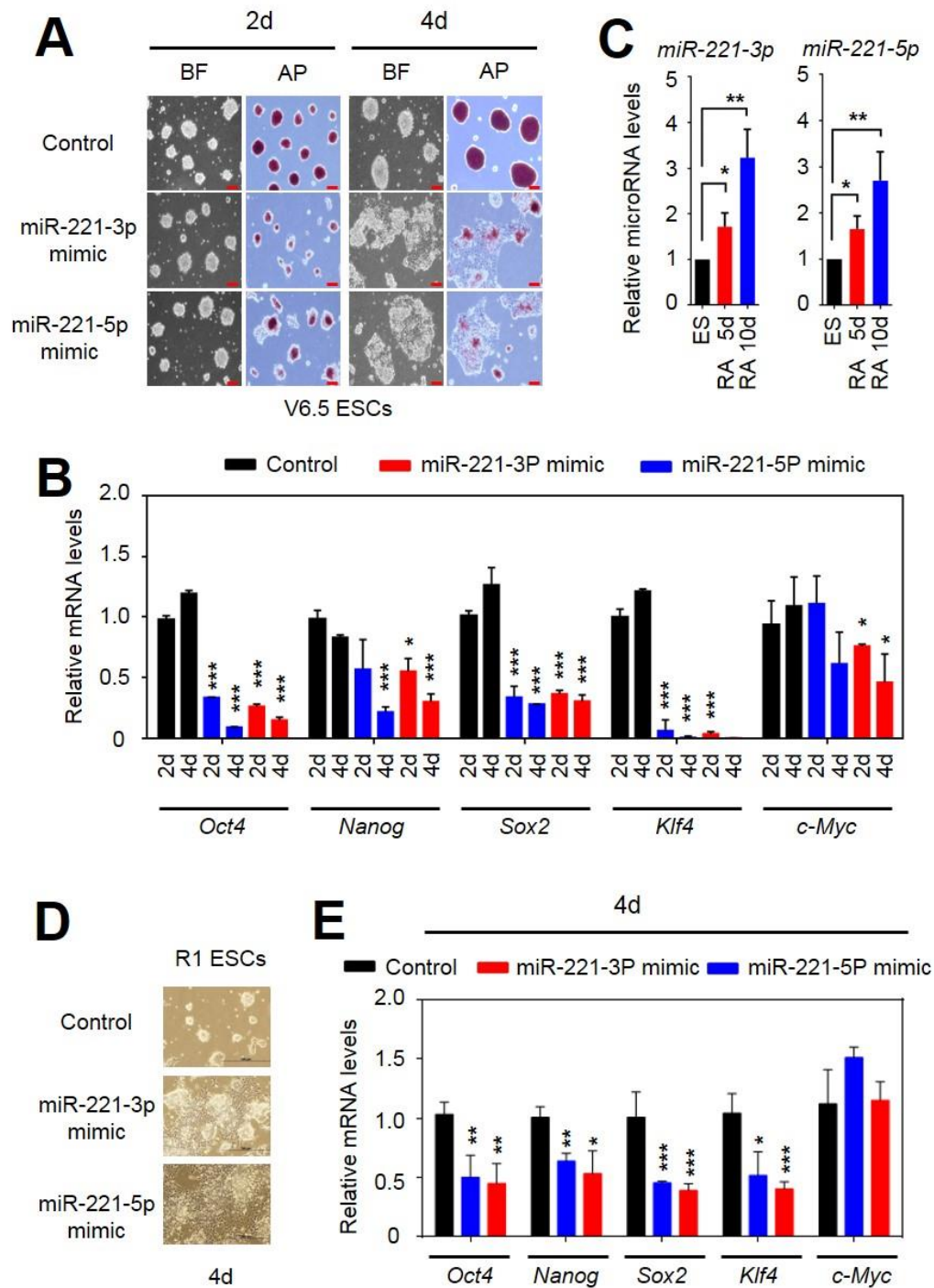
**Figure 17. PRMT7 represses the expression of the *miR-221* gene by establishing repressive arginine methylation.**

(A) Analysis of relative miRNA expression levels between shLuc-treated and two PRMT7-depleted (shPRMT7-7 and shPRMT7-8) mouse ESCs using miRNA-specific quantitative PCR. (B) The sequences of mature miR-221-5p (Red) and miR-221-3p (Blue) and the predicted stem loop structure form of pre-mature *miR-221*. (C) Schematic representation of mouse *miR-221* gene localization. S1, S2 and S3 indicate the PCR-amplified regions for ChIP. (D) Analysis of relative PRMT7 occupancy at *miR-221* promoter using quantitative ChIP. (E) Comparison of relative occupancy of PRMT7, H4R3me1, H4R3me2s and total H3 between shLuc-treated and PRMT7-depleted mouse ESCs at the *miR-221* promoter region in V6.5 mouse ESCs. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*). Experiments in Figure 17A were conducted by Sunghun Lee and used with his permission. Experiments in Figure 17 D and E were conducted by Shilpa Dhar and used with her permission.

### 3.3.2 *miR-221* has an anti-stemness function

To determine whether miR-221-3p and miR-221-5p have an anti-stemness function, we examined the effects of miR-221-3p and miR-221-5p mimics on mouse ESC stemness. AP staining analysis showed that the transfection of miR-221-3p and miR-221-5p mimics induced spontaneous differentiation of V6.5 mouse ESCs (**Fig. 18A**). Consistent with this, quantitative RT-PCR results showed the mRNA levels of the pluripotent factors *Oct4*, *Nanog*, *Sox2*, *Klf4*, and *c-Myc* were downregulated by miR-221-3p and miR-221-5p mimics (**Fig. 18B**). We also compared the expression levels of miR-221-3p and miR-221-5p between V6.5 mouse ESCs and RA-induced differentiated cells. As shown in **Fig. 18C**, the expression levels of miR-221-3p and miR-221-5p were increased upon RA treatment. To further validate the effect of miR-221-3p and miR-221-5p mimics on mouse ESC stemness, we used another mouse ESC line R1. Similar to results obtained using V6.5 mouse ESCs, the transfection of miR-221-3p and miR-221-5p into R1 mouse ESCs induced spontaneous differentiation of R1 ESCs (**Fig. 18D**) while downregulating the major pluripotent factors *Oct4*, *Nanog*, *Sox2*, and *Klf4* (**Fig. 18E**). Our results indicate that *miR-221* negatively regulates ESC stemness.

**Figure 18**



**Figure 18. *miR-221* mimics induces mouse ESC differentiation.**

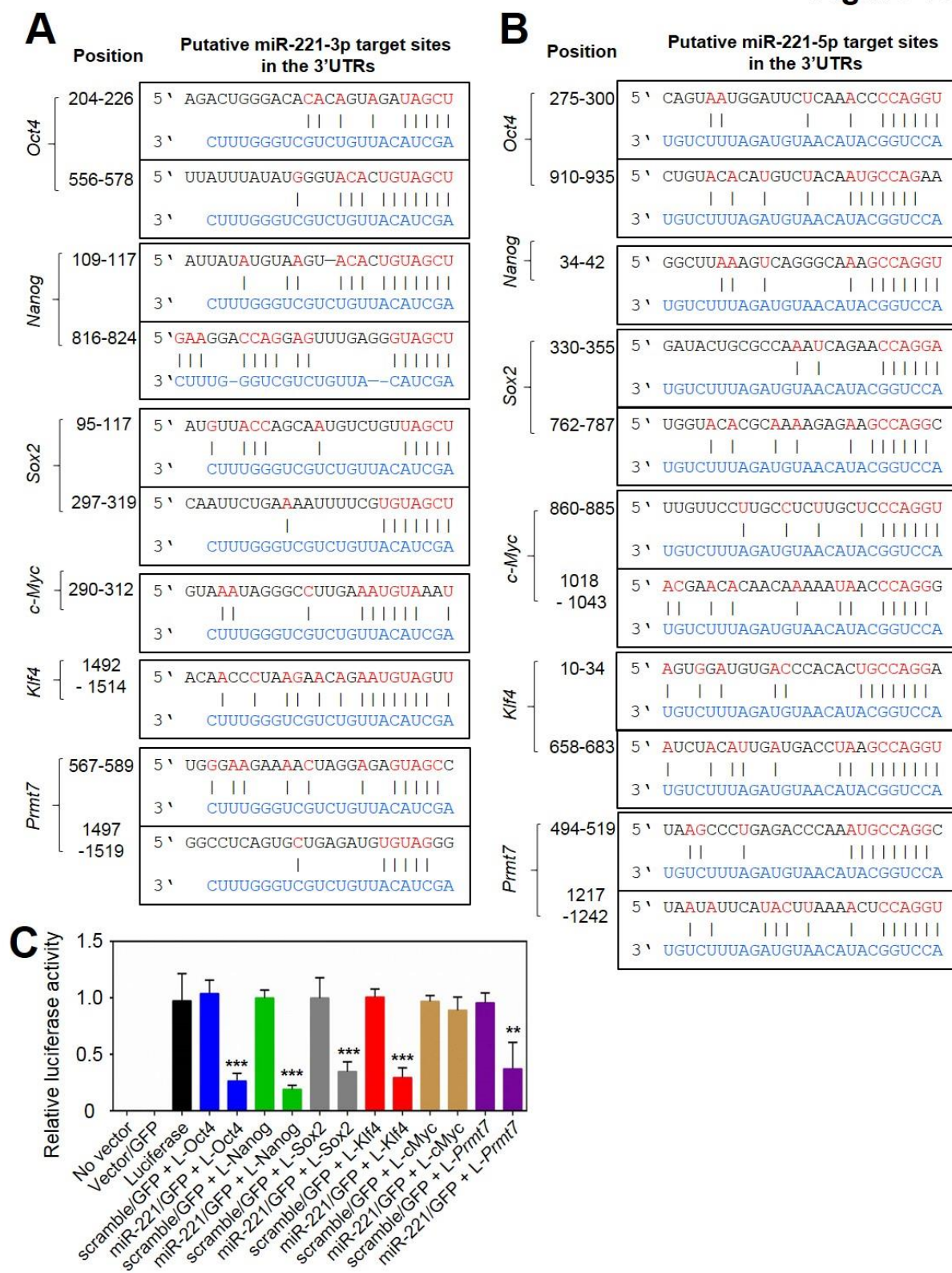
**(A & D)** Microscopic images of V6.5 (**A**) and R1 (**D**) mouse ESCs after the treatment of miR-221-3p and miR-221-5p mimics for 2 days (2d) or 4 days (4d). BF, bright field; AP; alkaline phosphatase staining. **(B & E)** Analysis of relative *Oct4*, *Nanog*, *Sox2*, *Klf4*, and c-Myc mRNA levels in V6.5 (**B**) and R1 (**E**) mouse ESCs after the treatment of miR-221-3p and miR-221-5p mimics. R1 cell line was treated with miRNA mimics for 4 days. **(C)** Relative miR-221-3p and miR-221-5p level during RA-induced mouse ESC differentiation. RA 5d, retinoic acid treatment for 5 days; RA 10d, retinoic acid treatment for 10 days. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*). Experiments in Figure 18 A, B and C were conducted by Sunghun Lee and used with his permission.

### 3.3.3 *miR-221* can target the 3'UTRs of several pluripotent factors, including *Oct4*, *Nanog*, *Sox2*, and *PRMT7*

Because miR-221-3p and miR-221-5p act as anti-pluripotent miRNAs, we sought to find their potential targets. Because *Oct4*, *Nanog*, *Sox2*, *Klf4*, and *c-Myc* were downregulated by miR-221-3p and miR-221-5p mimics and are major pluripotent factors that are critical for both stemness maintenance and stemness induction from differentiated cells, we focused on determining whether *miR-221* targets any of these factors. In addition, we included PRMT7 to test the possibility that *miR-221* silences its transcriptional repressor PRMT7. The miRNA-mediated mRNA targeting requires base pairing between a miRNA and its target mRNAs. Such base pairing is largely based on the complementarity between the nucleotide positions 2–8 (known as seed sequences) in miRNAs and their corresponding mRNA sequences. It has been known that the miRNA target sites in mRNAs are present in the 5' UTRs, open reading frames, and the 3' UTRs (130). Using these criteria, we examined putative target sites for miR-221-3p and miR-221-5p in *Oct4*, *Nanog*, *Sox2*, *Klf4*, *c-Myc*, and *Prmt7* genes. Our analysis using several software and manual examination suggested that there were putative target sites for miR-221-3p and miR-221-5p in 3'UTRs of these factors (Fig. 19A and 19B).

To determine whether *miR-221* can directly target *Oct4*, *Nanog*, *Sox2*, *Klf4* and *c-Myc* 3'UTRs, we transfected both one of luciferase expression plasmids with the 3'UTRs of these pluripotent factors along with a *miR-221* expression plasmid encoding miR-221-3p and miR-221-5p into HEK293T cells. Our results showed that *miR-221* expression substantially reduced the luciferase activities of *Oct4*, *Nanog*, *Sox2*, *Klf4*, and *Prmt7* 3'UTR but not *c-Myc* 3'UTR (Fig. 19C), suggesting that *miR-221* can directly target *Oct4*, *Nanog*, *Sox2*, *Klf4*, and *Prmt7* 3'UTR.

**Figure 19**



**Figure 19. *miR-221* can target 3'UTRs of the pluripotent factors *Oct4*, *Nanog*, *Sox2*, *Klf4* and *Prmt7*.**

**(A and B)** Putative target sites of miR-221-3p **(A)** and miR-221-5p **(B)** in 3'UTRs of mouse *Oct4*, *Nanog*, *Sox2*, *c-Myc*, *Klf4* and *Prmt7*. miR-221-3p and miR-221-5p sequences are shown in blue.

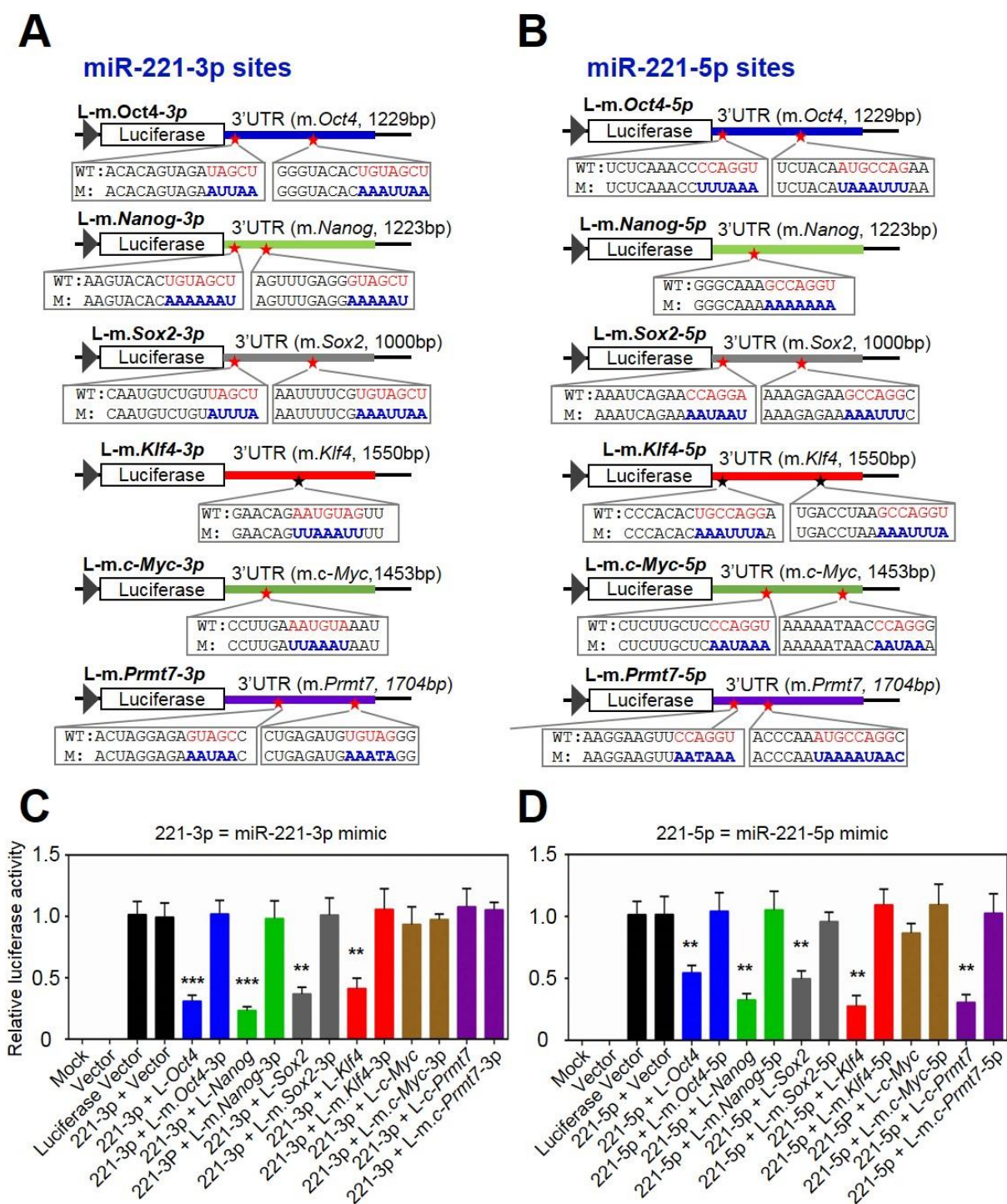
**(C)** The effect of *miR-221* expression on activities of reporter constructs containing *Oct4*-3'UTR (L-*Oct4*), *Nanog*-3'UTR (L-*Nanog*), *Sox2*-3'UTR (L-*Sox2*), *Klf4*-3'UTR (L-*Klf4*), *c-Myc*-3'UTR (L-*c-Myc*) and *Prmt7*-3'UTR (L-*Prmt7*). The *miR-221* expression plasmid (pMDH1-PKG-miR221-GFP) encoding both miR-221-3p and miR-221-5p were transfected with each luciferase reporter construct into HEK293T cells. Firefly luciferase activities were normalized to the internal transfection control (Renilla luciferase). Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*). Experiments in Figure 19 A, B and C were conducted by Sunghun Lee and used with his permission.



### **3.3.4 Both miR-221-3p and miR-221-5p can target 3'UTRs of *Oct4*, *Nanog*, *Sox2*, and *Klf4* while miR-221-5p can also silence *Prmt7*.**

To determine specific target sites of miR-221-3p and miR-221-5p in *Oct4*, *Nanog*, *Sox2*, *Klf4*, and *Prmt7* 3'UTR, we individually mutated putative target sites of miR-221-3p and miR-221-5p in the 3'UTRs in the reporter plasmids (Fig. 20A and 20B). We then co-transfected WT or mutant 3'UTR reporter plasmids with a miR-221-3p mimic or miR-221-5p mimic into HEK 293T cells. Both miR-221-3p and miR-221-5p mimics inhibited the luciferase activities of *Oct4*, *Nanog*, *Sox2*, and *Klf4* 3'UTRs whereas the mutant reporter plasmids resisted the inhibition mediated by miR-221-3p and miR-221-5p mimics (Fig. 20C and 20D). Interestingly, miR-221-5p mimics but not miR-221-3p mimics reduced the reporter activity of *Prmt7* 3'UTR. Together, these results indicate that target sites of miR-221-3p and miR-221-5p are located in *Oct4*, *Nanog*, *Sox2*, and *Klf4* 3'UTRs whereas miR-221-5p but not miR-221-3p targets *Prmt7* 3'UTR.

Figure 20



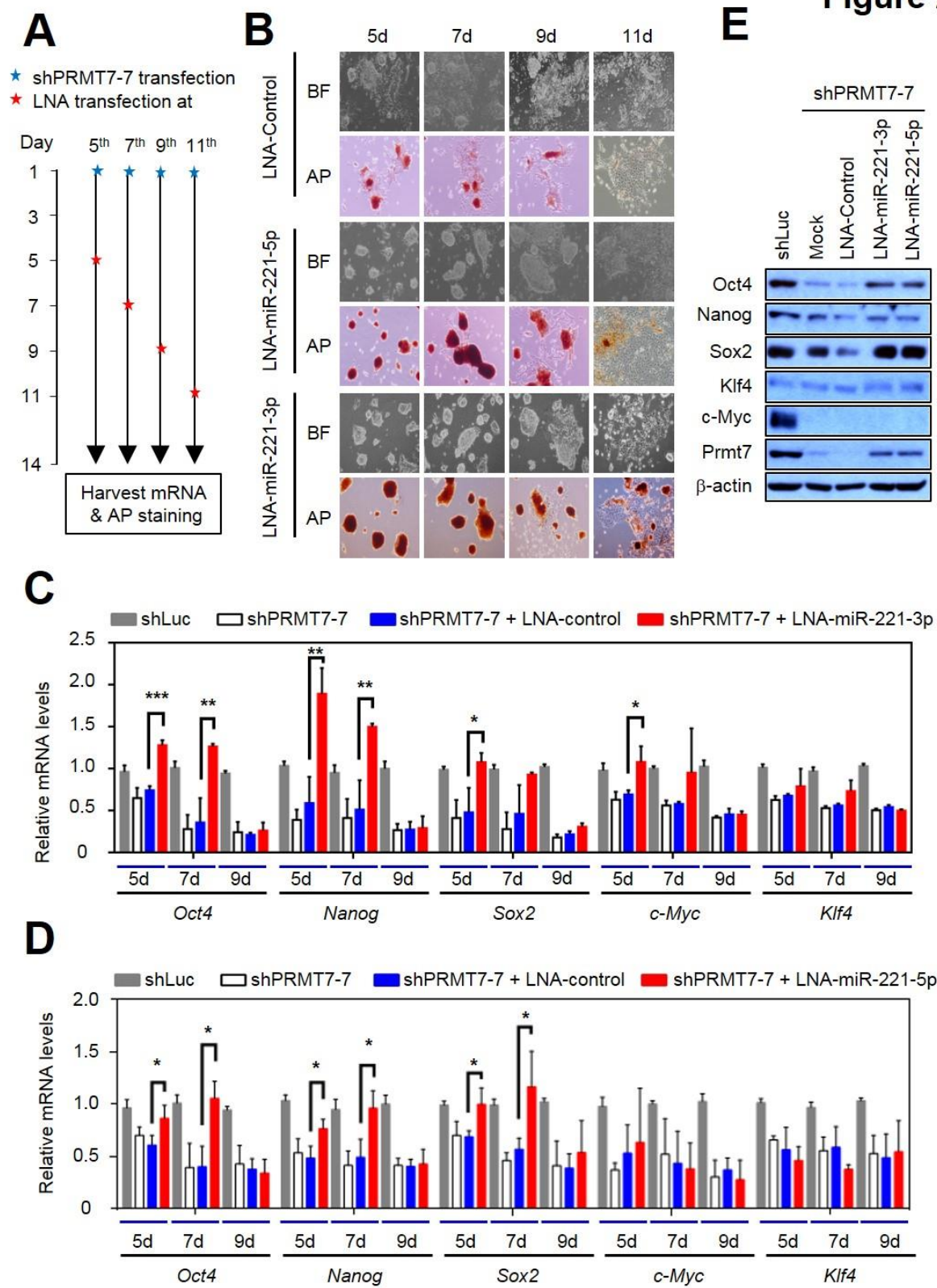
**Figure 20. Both miR-221-3p and miR-221-5p can target 3'UTRs of Oct4, Nanog, Sox2, and Klf4 while miR-221-5p can also silence *Prmt7*.**

**(A and B)** Schematic representation of luciferase reporter constructs containing *Oct4*-3'UTR, *Nanog*-3'UTR, *Sox2*-3'UTR, *Klf4*-3'UTR, *c-Myc*-3'UTR and *Prmt7*-3'UTR. Mutations in putative target sites for miR-221-3p **(A)** and miR-221-5p **(B)** are shown. The mutated sequences are showed in blue. WT, wild type; M, mutant. **(C and D)** The effect of miR-221-3p and miR-221-5p mimics on reporter activities of *Oct4*-3'UTR, *Nanog*-3'UTR, *Sox2*-3'UTR, *Klf4*-3'UTR, *c-Myc*-3'UTR and their mutants. Each WT or mutated reporter constructs were transfected with miR-221-3p mimic **(C)** or miR-221-5p mimic **(D)** into HEK293T cells. Firefly luciferase activities were normalized to the internal transfection control (Renilla luciferase). Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*). Experiments in Figure 20 were conducted by Sunghun Lee and used with his permission.

### 3.3.5 The repression of miR-221-3p and miR-221-5p expression is required for maintaining mouse ESC stemness

To determine whether the repression of miR-221-3p and miR-221-5p is necessary for maintaining mouse ESC stemness, we examined whether LNA inhibitors of miR-221-3p and miR-221-5p (LNA-miR-221-3p and LNA-miR-221-5p) delay or inhibit spontaneous differentiation of PRMT7-depleted mouse ESCs in which miR-221-3p and miR-221-5p levels are increased by PRMT7 depletion (LNA-miRNAs strongly bind to and inhibit their target miRNAs). We treated cells with LNA-control, LNA-miR-221-3p, or LNA-miR-221-5p at 5<sup>th</sup>, 7<sup>th</sup>, 9<sup>th</sup> and 11<sup>th</sup> day after transfecting shPRMT7-7 into V6.5 mouse ESCs (**Fig. 21A**) and examined mouse ESC morphology, alkaline phosphatase (AP) staining, protein levels, and mRNA expression. Our results showed that the treatment of PRMT7-depleted mouse ESCs with LNA-miR-221-3p or LNA-miR-221-5p (as compared to LNA-control) at the 5<sup>th</sup> and 7<sup>th</sup> day inhibited spontaneous ESC differentiation and restored AP staining (**Fig. 21B**). Quantitative RT-PCR results showed that expression levels of *Oct4*, *Nanog* and *Sox2* but not *Klf4* in shPRMT7-treated mouse ESCs were substantially recovered by both LNA-miR-221-3p and LNA-miR-221-5p at 5<sup>th</sup> and 7<sup>th</sup> day (**Fig. 21C and 21D, red bar**). Consistent with this, protein levels of these pluripotent factors and PRMT7 were increased by the treatment of LNA-miR-221-3p and LNA-miR-221-5p (**Fig. 21E**). In contrast, the transfection of LNA-miR-221-3p or LNA-miR-221-5p at the 9<sup>th</sup> or 11<sup>th</sup> day had insignificant effects on the differentiation of PRMT7-depleted cells and barely reversed the expression of *Oct4*, *Nanog*, and *Sox2*, suggesting that treatment of *miR-221* inhibitors at these two time points may be too late to inhibit spontaneous differentiation of mouse ESCs by PRMT7 knockdown. These results indicate that transcriptional repression of miR-221-3p and miR-221-5p by PRMT7 is necessary for the maintenance of *Oct4*, *Nanog*, *Sox2*, and *Prmt7* levels and the stemness of mouse ESCs.

**Figure 21**



**Figure 21. PRMT7-mediated repression of miR-221-3p and miR-221-5p levels is indispensable for maintaining mouse ESC stemness.**

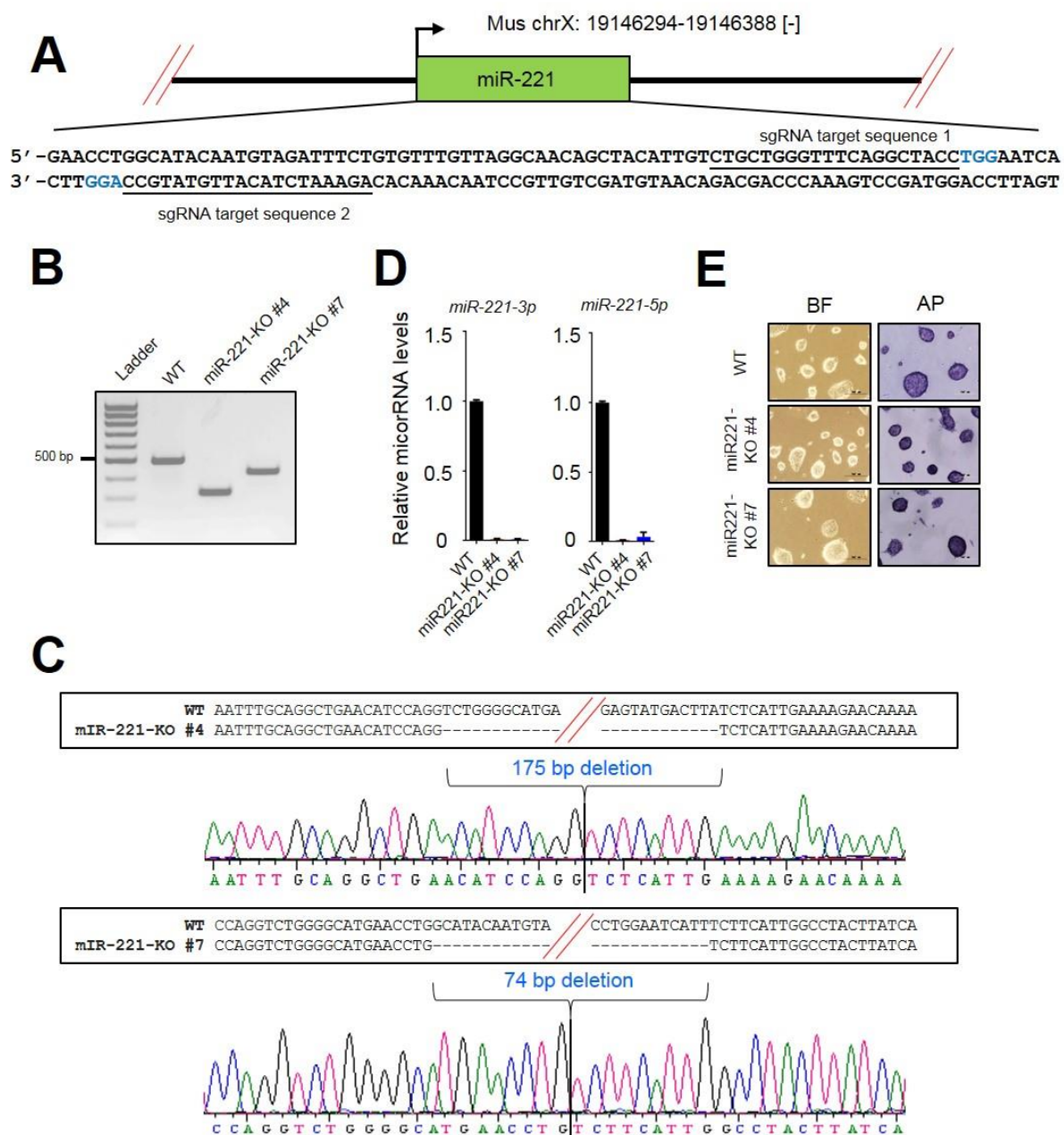
**(A)** Schematic representation of the procedure for the treatment of LNA-miR-221-3p and LNA-miR-221-5p after transfection of shPRMT7-7. **(B)** Microscopic and AP staining images of PRMT7-depleted V6.5 mouse ESCs at the 5th, 7th, 9th, or 11th day after treatment with LNA-control, LNA-miR-221-5p and LNA-miR-221-3p. **(C and D)** Analysis of relative mRNA levels of *Oct4*, *Nanog*, *Sox2*, *c-Myc* and *Klf4* after treatment of shLuc-treated or PRMT7-depleted V6.5 mouse ESCs with LNA-miR-221-3p **(C)** or LNA-miR-221-5p **(D)** at different time point (5<sup>th</sup>, 7<sup>th</sup>, and 9<sup>th</sup> day). **(E)** Western blot analysis of OCT4, NANOG, SOX2, C-MYC, KLF4, PRMT7 and  $\beta$ -ACTIN (loading control) levels after the treatment of shLuc-treated and PRMT7-depleted V6.5 mouse ESCs with LNA-control, LNA-miR-221-3p and LNA-miR-221-5p. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*).

### 3.3.6 CRISPR-mediated deletion of the miR-221 gene impedes spontaneous differentiation of PRMT7-depleted mouse ESCs

To vigorously confirm that the repression of miR-221-3p and miR-221-5p expression is required for mouse ESC stemness, we sought to examine whether *miR-221* loss impedes spontaneous differentiation of PRMT7-depleted mouse ESCs. We first generated *miR-221*-null mouse ESCs by using a CRISPR-Cas9 strategy that is a powerful tool for genome editing in living cells (88). In particular, we employed a double nicking strategy involving two guide RNAs and Cas9-D10A nickase (a mutant form of the RNA-guided, double-strand cleaving DNA endonuclease Cas9), because this strategy introduces DNA double strand breaks with significantly increased specificity around the target region (163). Two 20 bp-long single-guide RNAs (sgRNAs) was individually cloned into a Cas9-D10A nickase expression plasmid that was used to target the mouse *miR-221* gene (Fig. 22A). We transfected these plasmids into V6.5 mouse ESCs and screened mouse ESC colonies to obtain *miR-221*-null mouse ESC clones. Since V6.5 mouse ESCs was derived from a male mouse embryo and *miR-221* gene is located on the X chromosome, one allele of the X chromosome needs to be deleted for the generation of *miR-221*-null mouse ESCs. Genomic PCR results showed that the *miR-221* gene was deleted in two clones (*miR-221*-KO #4 and *miR-221*-KO #7) (Fig. 22B). DNA sequencing of genomic PCR products also confirmed the knockout of *miR-221* (Fig. 22C). Furthermore, miRNA-specific quantitative PCR showed that the expression of miR-221-3p and miR-221-5p was undetectable (Fig. 22D). These two *miR-221* knockout clones (*miR-221*-KO #4 and *miR-221*-KO #7) were morphologically normal and were positively stained by alkaline phosphatase (Fig. 22E), in line with a previous report showing a minor effect of *miR-221* inhibition on ESC proliferation (164).



**Figure 22**





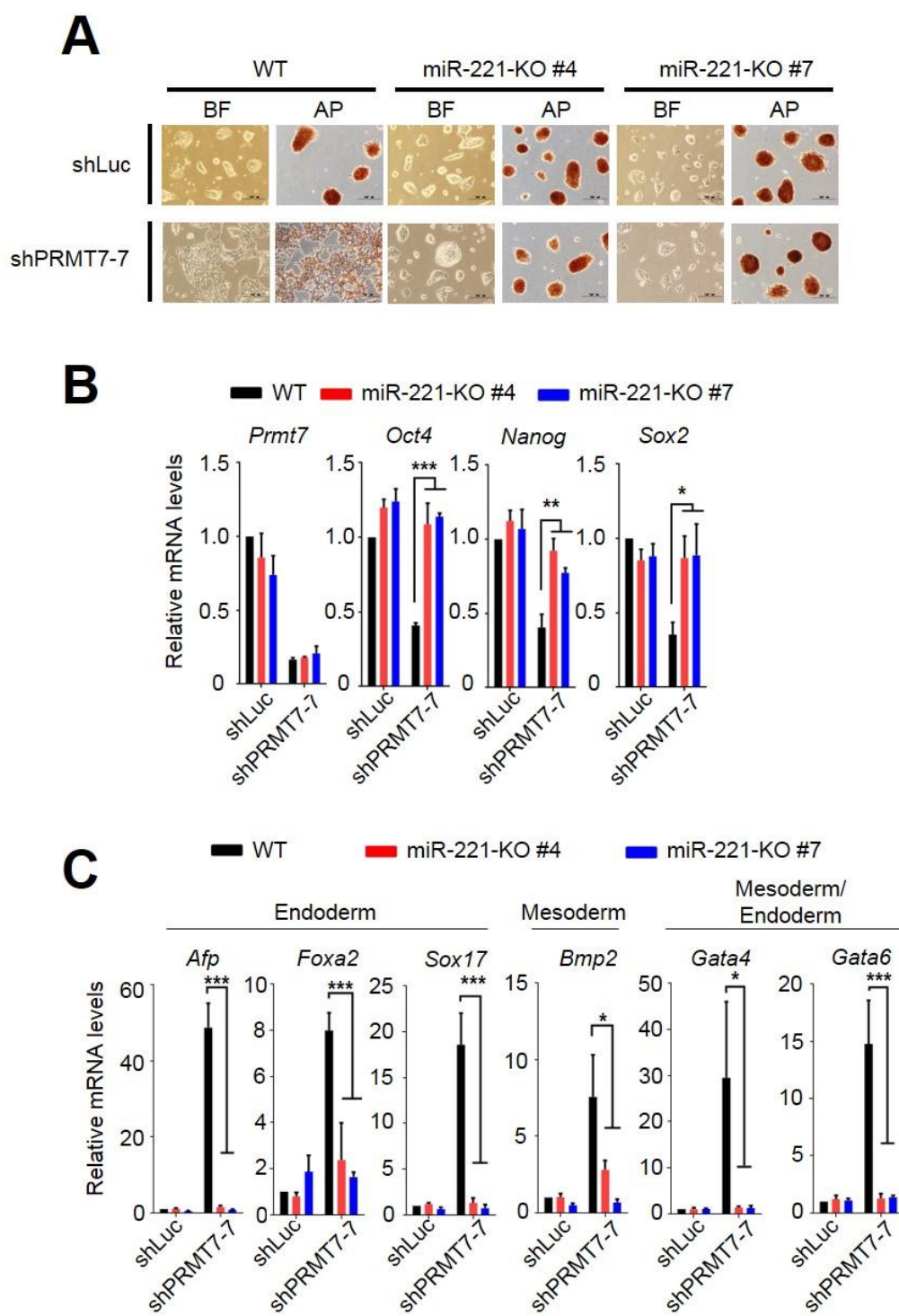
**Figure 22. Generation of *miR-221*-null V6.5 mouse ESCs by a CRISPR/Cas9 strategy.**

**(A)** CRISPR/Cas9 targeting sites for generating *miR-221*-null V6.5 mouse ESCs. The two guide RNA (sgRNA) sequences are underlined. The PAM sequences are labelled in blue. **(B)** PCR analysis of WT and *miR-221*-null V6.5 mouse ESCs by using primers flanking the *miR-221* deletion region. The PCR band size for WT *miR-221* is predicted to be 500 bp. **(C)** DNA sequencing chromatograms for *miR-221*-KO #4 and *miR-221*-KO-#7. **(D)** Analysis of relative miRNA levels between WT and *miR-221*-null V6.5 mouse ESCs (Clone #4 and #7) by using miRNA-specific quantitative PCR. **(E)** Microscopic and AP staining images of WT, *miR-221*-KO #4 and *miR-221*-KO #7 mouse ESCs. BF, bright field; AP; alkaline phosphatase staining. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*)).

### 3.3.7 *miR-221* loss blocks spontaneous differentiation of PRMT7-depleted mouse ESCs.

To determine whether *miR-221* loss blocks spontaneous differentiation of PRMT7-depleted mouse ESCs, we knocked down PRMT7 in *miR-221*-null ESCs (*miR-221*-KO #4 and *miR-221*-KO #7). *miR-221* loss inhibited spontaneous differentiation of PRMT7-depleted mouse ESCs, as evident by mouse ESC morphology as well as positive AP staining (Fig. 23A). Consistent with this, *miR-221* loss restored *Oct4*, *Nanog*, and *Sox2* mRNA levels in PRMT7-depleted mouse ESCs (Fig. 23B). Because we previously showed that PRMT7 knockdown induced endoderm (e.g., *Afp*, *Foxa2*, and *Sox17*), mesoderm (e.g., *Bmp2*) and mesoderm/endoderm (e.g., *Gata4* and *Gata6*) markers (85), we examined whether *miR-221* loss prevents their induction by PRMT7 knockdown. In fact, *miR-221* loss inhibited shPRMT7-mediated induction of *Afp*, *Foxa2*, *Sox17*, *Bmp2*, *Gata4* and *Gata6* (Fig. 23C). Together, our results indicate that the repression of miR-221-3p and miR-221-5p is critical for maintaining mouse ESC stemness.

**Figure 23**



**Figure 23. *miR-221* loss blocks spontaneous differentiation of PRMT7-depleted mouse ESCs.**

**(A)** Microscopic and AP-staining images of shLuc-treated and PRMT7-depleted WT and *miR-221*-null V6.5 mouse ESCs. BF, bright field; AP; alkaline phosphatase staining. **(B)** Comparison of relative mRNA level of *Prmt7*, *Oct4*, *Nanog*, *Sox2*, *Klf4*, and *c-Myc* between WT and *miR-221*-null V6.5 mouse ESCs after treatment of shLuc or shPRMT7-7. **(C)** Comparison of relative mRNA level of *Afp*, *Foxa2*, *Sox17*, *Bmp2*, *Gata4*, and *Gata6* between WT and *miR-221*-null V6.5 mouse ESCs after treatment of shLuc or shPRMT7-7. Data are presented as the mean  $\pm$  SD of three independent experiments.  $P < 0.05$  (\*),  $P < 0.01$  (\*\*), and  $P < 0.001$  (\*\*\*)).

### 3.4 DISCUSSION

#### 3.4.1 The anti-stemness function of *miR-221*

miR-221 is overexpressed in various types of cancers. Several targets of miR-221 have been identified (162). For example, miR-221 directly targets the cell cycle inhibitor p27 (Kip1) and positively affects proliferation potential in prostate cancer (165). miR-221 also targets the tumor suppressor and cell cycle inhibitor p57 (CDKN1C) in hepatocellular carcinoma (166). Garofalo *et al.* showed that miR-221 targets the tumor suppressors PTEN and TIMP3 to enhance tumorigenicity of non-small cell lung cancer and hepatocellular carcinoma (167). Overexpression of miR-221 in several cancers is linked to the resistance to various cancer therapies in addition to growth advantage of cancer cells (162). For instance, expression of miR-221 is increased in several chemo-resistant cancer cells (168,169). For these reasons, miR-221 is considered an oncomiR that plays an important role in cancer development, and the inhibition of miR-221 in combination with other cancer treatments may be relevant to a new therapeutic strategy for cancer treatment (170,171). Distinct from these studies, results reported here uncover that *miR-221* has an anti-stemness function to enhance the differentiation of mouse ESCs.

The anti-stemness and pro-differentiation function of miR-221-3p and miR-221-5p are supported by several lines of evidence. In fact, we showed that transfection of miR-221-3p or miR-221-5p mimics in two different ESC cell lines induced ESC differentiation and that the loss of the *miR-221* gene almost completely inhibited spontaneous differentiation of PRMT7-depleted mouse ESCs. We also found that the expression levels of miR-221-3p and miR-221-5p were low in mouse ESCs as compared to differentiated somatic cells (3T3 fibroblasts) (data not shown) and were increased during RA-induced differentiation. In line with our findings, *miR-221* is upregulated in fully differentiated neurons (172) and plays a role in neuron differentiation (173). The differentiation of ESCs requires both downregulation of pluripotent factors and upregulation

of lineage-specific markers (14,174). In support of this, our results showed that *miR-221* loss not only recovers the expression of the pluripotent markers (e.g., *Oct4*, *Nanog*, and *Sox2*) but also downregulates the expression of mesoderm and endoderm markers (Fig. 23).

There are several ESC-specific miRNAs that target anti-stemness genes (140,141). For example, the mouse *miR-290-295* family members (human homolog miR-371-373) are among the most abundant miRNAs in mouse ESCs (142) and target the cell cycle inhibitor p21 (*Cdkn1a*) to promote G1/S cell cycle transition and rapid proliferation of ESCs (143). In contrast to pro-stemness miRNAs, certain miRNAs inhibit ESC stemness and induce cell differentiation by targeting pluripotent networks. For example, let-7, miR-21, miR-134, and miR-145 are anti-stemness miRNAs that target pluripotent factors, such as *Oct4*, *Nanog*, *Sox2*, c-Myc, *Sal4*, *Lin28*, and *Klf4* (175-178). In the present study, our reporter assay in combination with mutagenesis indicate that *miR-221* miRNAs can target the 3' UTRs of the major pluripotent factors *Oct4*, *Nanog*, *Sox2*, *Klf4*, and *Prmt7*. Moreover, LNA-miRNA-mediated inhibition or CRISPR-mediated deletion of miR-221-3p and miR-221-5p restored the expression of *Oct4*, *Nanog*, *Sox2*, and *Prmt7* in PRMT7-depleted mouse ESCs. These findings indicate that *miR-221* acts as an anti-stemness miRNA by targeting mRNAs of multiple pluripotent genes, including at least *Oct4*, *Nanog*, *Sox2*, and *Prmt7*.

### 3.4.2 PRMT7 negatively regulates the expression of *miR-221*

It has been shown that the *miR-221* gene is transcriptionally regulated by several transcriptional factors. For example, ER $\alpha$  binds to the *miR-221* promoter to repress *miR-221* expression, consistent with the overexpression of *miR-221* in ER-negative breast cancer (179). The expression of *miR-221* in breast cancer is positively regulated by EGFR-RAS-RAF-MEK-ERK2-FOSL1 axis, and FOSL1 activates *miR-221* expression through its interaction with the

miR-221 promoter (180). Fornari *et al.* showed that p53 can regulate miR-221 levels in hepatocellular carcinoma (181). In the current study, our results indicate that PRMT7 binds to the *miR-221* promoter and increases the repressive epigenetic marks H4R3me1 and H4R3me2s to downregulate *miR-221* expression. Thus, our findings provide a new *miR-221*-regulatory mechanism in which the expression of *miR-221* is epigenetically repressed by the histone arginine methyltransferase PRMT7 in mouse ESCs.

miR-221-5p targets its own repressor PRMT7 besides the well-known pluripotent factors *Oct4*, *Nanog*, and *Sox2* (Figs. 20D), and PRMT7 represses the *miR-221* gene by establishing repressive arginine methylation marks (Fig. 17). Therefore, there is a mutual antagonistic relationship between miR-221-5p and PRMT7. Interestingly, we previously reported an additional antagonistic relationship that is formed by PRMT7 and miR-24-3p/miR-24-2-5p (85). Although PRMT7 represses both *miR-221* and *miR-24-2* genes in mouse ESCs, results reported here showed that CRISPR-mediated deletion of the *miR-221* gene alone was enough to block spontaneous differentiation of PRMT7-depleted mouse ESCs. Thus, it is plausible that PRMT7-mediated repression of the *miR-221* gene may play a more predominant role in maintaining mouse ESC stemness than PRMT7-mediated repression of the *miR-24-2* gene, although both *miR-221* and *miR-24-2* might be required for spontaneous differentiation of PRMT7-depleted mouse ESCs. Perhaps, this could be addressed by future experiments involving both CRISPR-mediated *miR-24-2* deletion and PRMT7 knockdown.

### 3.4.3 Summary

In summary, our results show that miR-221-3p and miR-221-5p can target the major pluripotent factors *Oct4*, *Nanog*, and *Sox2* in mouse ESCs, indicating an anti-stemness and pro-differentiation role for miR-221-3p and miR-221-5p. Because our results also uncovered that

PRMT7 epigenetically represses the expression of miR-221-3p and miR-221-5p in mouse ESCs and that miR-221-5p can target the expression of *Prmt7*, it is possible that miR-221-5p and *Prmt7* form a negative feedback loop. Finally, we provide evidence that PRMT7-mediated repression of miR-221-3p and miR-221-5p is necessary for maintaining mouse ESC stemness.



### 3.5 PERSPECTIVE AND FUTURE DIRECTIONS

In our current study, we showed that PRMT7 represses the expression of *miR-221* gene in mouse ESCs to maintain the stemness of mouse ESCs. Our results further showed that both miR-221-3p and miR-221-5p are anti-stemness miRNAs that can target several major pluripotent factors in mouse ESCs to induce spontaneous differentiation. Furthermore, we also found that the expression levels of *miR-221* is low in mouse ESCs as compared to differentiated somatic cells and were increased during RA-induced differentiation. It should be interesting to further determine whether we can target *miR-221* to generate induced pluripotent stem cells (iPSCs). This could be addressed by future experiments to inhibit the expression of *miR-221* and determine the efficiency of iPSCs generation.

## REFERENCES

1. Stevens, L. C., and Little, C. C. (1954) Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proceedings of the National Academy of Sciences of the United States of America* **40**, 1080-1087
2. Martin, G. R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 7634-7638
3. Evans, M. J., and Kaufman, M. H. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156
4. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. (1998) Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* **282**, 1145-1147
5. Donovan, P. J., and Gearhart, J. (2001) The end of the beginning for pluripotent stem cells. *Nature* **414**, 92-97
6. Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998) Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell* **95**, 379-391
7. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003) The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* **113**, 631-642
8. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003) Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic

- Stem Cells. *Cell* **113**, 643-655
9. Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & Development* **17**, 126-140
  10. Loh, Kyle M., and Lim, B. A Precarious Balance: Pluripotency Factors as Lineage Specifiers. *Cell Stem Cell* **8**, 363-369
  11. Thomson, M., Liu, Siyuan J., Zou, L.-N., Smith, Z., Meissner, A., and Ramanathan, S. Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell* **145**, 875-889
  12. Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K.-Y., Sung, K. W., Lee, C. W. H., Zhao, X.-D., Chiu, K.-P., Lipovich, L., Kuznetsov, V. A., Robson, P., Stanton, L. W., Wei, C.-L., Ruan, Y., Lim, B., and Ng, H.-H. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-440
  13. Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. (2005) Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell* **122**, 947-956
  14. Young, Richard A. (2011) Control of the Embryonic Stem Cell State. *Cell* **144**, 940-954
  15. Keller, G. M. (1995) In vitro differentiation of embryonic stem cells. *Current Opinion in Cell Biology* **7**, 862-869
  16. Waddington, C. H. (2012) The Epigenotype. *International Journal of Epidemiology* **41**, 10-

17. Waddington, C. H. (1968) Towards a Theoretical Biology. *Nature* **218**, 525-527
18. Berger, S. L., Kouzarides, T., Shiekhata, R., and Shilatifard, A. (2009) An operational definition of epigenetics. *Genes & Development* **23**, 781-783
19. Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007) Epigenetics: A Landscape Takes Shape. *Cell* **128**, 635-638
20. Kornberg, R. D. (1974) Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **184**, 868-871
21. Olins, D. E., and Olins, A. L. (2003) Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* **4**, 809-814
22. Li, G., and Reinberg, D. (2011) Chromatin higher-order structures and gene regulation. *Current opinion in genetics & development* **21**, 175-186
23. Human Genome Sequencing, C. (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945
24. Grewal, S. I. S., and Elgin, S. C. R. (2002) Heterochromatin: new possibilities for the inheritance of structure. *Current Opinion in Genetics & Development* **12**, 178-187
25. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260
26. Bannister, A. J., and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395
27. Bannister, A. J., and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Research* **21**, 381-395
28. Strahl, B. D., and Allis, C. D. (2000) The language of covalent histone modifications. *Nature*

**403**, 41-45

29. Chi, P., Allis, C. D., and Wang, G. G. (2010) Covalent histone modifications — miswritten, misinterpreted and mis-erased in human cancers. *Nat Rev Cancer* **10**, 457-469
30. Greer, E. L., and Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* **13**, 343-357
31. Black, Joshua C., Van Rechem, C., and Whetstone, Johnathan R. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Molecular Cell* **48**, 491-507
32. Bedford, M. T., and Clarke, S. G. (2009) Protein Arginine Methylation in Mammals: Who, What, and Why. *Molecular Cell* **33**, 1-13
33. Turner, B. M. Decoding the nucleosome. *Cell* **75**, 5-8
34. Hans, F., and Dimitrov, S. (2001) Histone H3 phosphorylation and cell division. *Oncogene* **20**, 3021-3027
35. Martin, C., and Zhang, Y. (2005) The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* **6**, 838-849
36. Yun, M., Wu, J., Workman, J. L., and Li, B. (2011) Readers of histone modifications. *Cell Research* **21**, 564-578
37. Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W. M. P., van Schaik, F. M. A., Varier, R. A., Baltissen, M. P. A., Stunnenberg, H. G., Mann, M., and Timmers, H. T. M. (2007) Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **131**, 58-69
38. van Ingen, H., van Schaik, F. M. A., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A. C., Kruijzer, J. A. W., Liskamp, R. M. J., Timmers, H. T. M., and Boelens, R. (2008) Structural

- Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3. *Structure* **16**, 1245-1256
39. Dillon, S. C., Zhang, X., Trievel, R. C., and Cheng, X. (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biology* **6**, 227-227
  40. Falnes, Pål Ø., Jakobsson, Magnus E., Davydova, E., Ho, A., and Malecki, J. (2016) Protein lysine methylation by seven- $\beta$ -strand methyltransferases. *Biochemical Journal* **473**, 1995-2009
  41. Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. (1998) SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cellular and Molecular Life Sciences CMLS* **54**, 80-93
  42. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599
  43. Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M., and Gamblin, S. J. (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**, 652-656
  44. Qian, C., and Zhou, M.-M. (2006) SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cellular and Molecular Life Sciences CMLS* **63**, 2755-2763
  45. Jacobs, S. A., Harp, J. M., Devarakonda, S., Kim, Y., Rastinejad, F., and Khorasanizadeh, S. (2002) The active site of the SET domain is constructed on a knot. *Nat Struct Mol Biol* **9**, 833-838
  46. Schapira, M. (2011) Structural Chemistry of Human SET Domain Protein Methyltransferases. *Current Chemical Genomics* **5**, 85-94

47. Thirman, M. J., Gill, H. J., Burnett, R. C., Mbangkollo, D., McCabe, N. R., Kobayashi, H., Ziemer-van der Poel, S., Kaneko, Y., Morgan, R., Sandberg, A. A., Chaganti, R. S. K., Larson, R. A., Le Beau, M. M., Diaz, M. O., and Rowley, J. D. (1993) Rearrangement of the MLL Gene in Acute Lymphoblastic and Acute Myeloid Leukemias with 11q23 Chromosomal Translocations. *New England Journal of Medicine* **329**, 909-914
48. Krivtsov, A. V., and Armstrong, S. A. (2007) MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* **7**, 823-833
49. Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De Braekeleer, E., De Braekeleer, M., Delabesse, E., de Oliveira, M. P., Cave, H., Clappier, E., van Dongen, J. J. M., Balgobind, B. V., van den Heuvel-Eibrink, M. M., Beverloo, H. B., Panzer-Grumayer, R., Teigler-Schlegel, A., Harbott, J., Kjeldsen, E., Schnittger, S., Koehl, U., Gruhn, B., Heidenreich, O., Chan, L. C., Yip, S. F., Krzywinski, M., Eckert, C., Moricke, A., Schrappe, M., Alonso, C. N., Schafer, B. W., Krauter, J., Lee, D. A., zur Stadt, U., Te Kronnie, G., Sutton, R., Izraeli, S., Trakhtenbrot, L., Lo Nigro, L., Tsaur, G., Fechina, L., Szczepanski, T., Strehl, S., Ilencikova, D., Molkentin, M., Burmeister, T., Dingermann, T., Klingebiel, T., and Marschalek, R. (2009) New insights to the MLL recombinome of acute leukemias. *Leukemia* **23**, 1490-1499
50. Mohan, M., Lin, C., Guest, E., and Shilatifard, A. (2010) Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nat Rev Cancer* **10**, 721-728
51. Smith, E., Lin, C., and Shilatifard, A. (2011) The super elongation complex (SEC) and MLL in development and disease. *Genes & Development* **25**, 661-672
52. Hu, D., Gao, X., Morgan, M. A., Herz, H.-M., Smith, E. R., and Shilatifard, A. (2013) The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4

53. Adam, M. P., and Hudgins, L. (2005) Kabuki syndrome: a review. *Clinical Genetics* **67**, 209-219
54. Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., Lee, C., Turner, E. H., Smith, J. D., Rieder, M. J., Yoshiura, K.-i., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D. A., Bamshad, M. J., and Shendure, J. (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-793
55. Parsons, D. W., Li, M., Zhang, X., Jones, S., Leary, R. J., Lin, J. C.-H., Boca, S. M., Carter, H., Samayoa, J., Bettegowda, C., Gallia, G. L., Jallo, G. I., Binder, Z. A., Nikolsky, Y., Hartigan, J., Smith, D. R., Gerhard, D. S., Fults, D. W., VandenBerg, S., Berger, M. S., Marie, S. K. N., Shinjo, S. M. O., Clara, C., Phillips, P. C., Minturn, J. E., Biegel, J. A., Judkins, A. R., Resnick, A. C., Storm, P. B., Curran, T., He, Y., Rasheed, B. A., Friedman, H. S., Keir, S. T., McLendon, R., Northcott, P. A., Taylor, M. D., Burger, P. C., Riggins, G. J., Karchin, R., Parmigiani, G., Bigner, D. D., Yan, H., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., and Velculescu, V. E. (2011) The Genetic Landscape of the Childhood Cancer Medulloblastoma. *Science* **331**, 435
56. Pugh, T. J., Weeraratne, S. D., Archer, T. C., Pomeranz Krummel, D. A., Auclair, D., Bochicchio, J., Carneiro, M. O., Carter, S. L., Cibulskis, K., Erlich, R. L., Greulich, H., Lawrence, M. S., Lennon, N. J., McKenna, A., Meldrim, J., Ramos, A. H., Ross, M. G., Russ, C., Shefler, E., Sivachenko, A., Sogoloff, B., Stojanov, P., Tamayo, P., Mesirov, J. P., Amani, V., Teider, N., Sengupta, S., Francois, J. P., Northcott, P. A., Taylor, M. D., Yu, F., Crabtree, G. R., Kautzman, A. G., Gabriel, S. B., Getz, G., Jager, N., Jones, D. T. W., Lichter, P., Pfister,



- S. M., Roberts, T. M., Meyerson, M., Pomeroy, S. L., and Cho, Y.-J. (2012) Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106-110
57. Sims, R. J., Nishioka, K., and Reinberg, D. (2003) Histone lysine methylation: a signature for chromatin function. *Trends in Genetics* **19**, 629-639
  58. Shilatifard, A. (2006) Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. *Annual Review of Biochemistry* **75**, 243-269
  59. Kouzarides, T. (2007) Chromatin Modifications and Their Function. *Cell* **128**, 693-705
  60. Ruthenburg, A. J., Allis, C. D., and Wysocka, J. (2007) Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Molecular Cell* **25**, 15-30
  61. Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2005) Spatial Distribution of Di- and Tri-methyl Lysine 36 of Histone H3 at Active Genes. *Journal of Biological Chemistry* **280**, 17732-17736
  62. Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823-837
  63. Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A. H. F. M. (2003) Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology* **13**, 1192-1200
  64. Kapoor-Vazirani, P., Kagey, J. D., and Vertino, P. M. (2011) SUV420H2-Mediated H4K20 Trimethylation Enforces RNA Polymerase II Promoter-Proximal Pausing by Blocking hMOF-Dependent H4K16 Acetylation. *Molecular and Cellular Biology* **31**, 1594-1609

65. Park, P. J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**, 669-680
66. Ernst, J., and Kellis, M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotech* **28**, 817-825
67. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., and Young, R. A. (2007) A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* **130**, 77-88
68. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318
69. Creighton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., and Jaenisch, R. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**, 21931-21936
70. Flanagan, J. F., Mi, L.-Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**, 1181-1185
71. Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D., and Patel, D. J. (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91-95
72. Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K., and Ohara, O. (2000) Prediction of the Coding Sequences of Unidentified Human Genes. XIX. The complete Sequences of

- 100 New cDNA Clones from Brain Which Code for Large Proteins in vitro. *DNA Research* **7**, 347-355
73. Sun, X.-J., Xu, P.-F., Zhou, T., Hu, M., Fu, C.-T., Zhang, Y., Jin, Y., Chen, Y., Chen, S.-J., Huang, Q.-H., Liu, T. X., and Chen, Z. (2008) Genome-Wide Survey and Developmental Expression Mapping of Zebrafish SET Domain-Containing Genes. *PLOS ONE* **3**, e1499
  74. Emerling, B. M., Bonifas, J., Kratz, C. P., Donovan, S., Taylor, B. R., Green, E. D., Le Beau, M. M., and Shannon, K. M. (2002) MLL5, a homolog of *Drosophila trithorax* located within a segment of chromosome band 7q22 implicated in myeloid leukemia. *Oncogene* **21**, 4849-4854
  75. Mas-y-Mas, S., Barbon, M., Teyssier, C., Déméné, H., Carvalho, J. E., Bird, L. E., Lebedev, A., Fattori, J., Schubert, M., Dumas, C., Bourguet, W., and le Maire, A. (2016) The Human Mixed Lineage Leukemia 5 (MLL5), a Sequentially and Structurally Divergent SET Domain-Containing Protein with No Intrinsic Catalytic Activity. *PLOS ONE* **11**, e0165139
  76. Madan, V., Madan, B., Brykczynska, U., Zilbermann, F., Hogeveen, K., Döhner, K., Döhner, H., Weber, O., Blum, C., Rodewald, H.-R., Sassone-Corsi, P., Peters, A. H. F. M., and Fehling, H. J. (2009) Impaired function of primitive hematopoietic cells in mice lacking the Mixed-Lineage-Leukemia homolog Mll5. *Blood* **113**, 1444
  77. Sebastian, S., Sreenivas, P., Sambasivan, R., Cheedipudi, S., Kandalla, P., Pavlath, G. K., and Dhawan, J. (2009) MLL5, a trithorax homolog, indirectly regulates H3K4 methylation, represses cyclin A2 expression, and promotes myogenic differentiation. *Proceedings of the National Academy of Sciences* **106**, 4719-4724
  78. Osipovich, A. B., Gangula, R., Vianna, P. G., and Magnuson, M. A. (2016) *Setd5* is essential for mammalian development and the co-

- transcriptional regulation of histone acetylation. *Development* **143**, 4595
79. Riess, A., Grasshoff, U., Schäferhoff, K., Bonin, M., Riess, O., Horber, V., and Tzschach, A. (2012) Interstitial 3p25.3–p26.1 deletion in a patient with intellectual disability. *American Journal of Medical Genetics Part A* **158A**, 2587-2590
  80. Peltekova, I. T., Macdonald, A., and Armour, C. M. (2012) Microdeletion on 3p25 in a patient with features of 3p deletion syndrome. *American Journal of Medical Genetics Part A* **158A**, 2583-2586
  81. Kellogg, G., Sum, J., and Wallerstein, R. (2013) Deletion of 3p25.3 in a patient with intellectual disability and dysmorphic features with further definition of a critical region. *American Journal of Medical Genetics Part A* **161**, 1405-1408
  82. Grozeva, D., Carss, K., Spasic-Boskovic, O., Parker, Michael J., Archer, H., Firth, Helen V., Park, S.-M., Canham, N., Holder, Susan E., Wilson, M., Hackett, A., Field, M., Floyd, James A. B., Hurles, M., and Raymond, F. L. (2014) De Novo Loss-of-Function Mutations in SETD5, Encoding a Methyltransferase in a 3p25 Microdeletion Syndrome Critical Region, Cause Intellectual Disability. *The American Journal of Human Genetics* **94**, 618-624
  83. Dhar, S. S., Lee, S.-H., Kan, P.-Y., Voigt, P., Ma, L., Shi, X., Reinberg, D., and Lee, M. G. (2012) Trans-tail regulation of MLL4-catalyzed H3K4 methylation by H4R3 symmetric dimethylation is mediated by a tandem PHD of MLL4. *Genes & Development* **26**, 2749-2762
  84. Lee, M. G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Di Croce, L., and Shiekhhattar, R. (2007) Demethylation of H3K27 Regulates Polycomb Recruitment and H2A Ubiquitination. *Science* **318**, 447-450
  85. Lee, S.-H., Chen, T.-Y., Dhar, S. S., Gu, B., Chen, K., Kim, Y. Z., Li, W., and Lee, M. G. (2016)

- A feedback loop comprising PRMT7 and miR-24-2 interplays with Oct4, Nanog, Klf4 and c-Myc to regulate stemness. *Nucleic Acids Research* **44**, 10603-10618
86. Nishikawa, S.-I., Jakt, L. M., and Era, T. (2007) Embryonic stem-cell culture as a tool for developmental cell biology. *Nat Rev Mol Cell Biol* **8**, 502-507
  87. Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.-A. (2004) Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat Neurosci* **7**, 1003-1009
  88. Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281-2308
  89. Pöpperl, H., and Featherstone, M. S. (1993) Identification of a retinoic acid response element upstream of the murine Hox-4.2 gene. *Molecular and Cellular Biology* **13**, 257-265
  90. Martinez-Ceballos, E., and Gudas, L. J. (2008) Hoxa1 is required for the retinoic acid-induced differentiation of embryonic stem cells into neurons. *Journal of Neuroscience Research* **86**, 2809-2819
  91. Mahony, S., Mazzoni, E. O., McCuine, S., Young, R. A., Wichterle, H., and Gifford, D. K. (2011) Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biology* **12**, R2-R2
  92. Cunningham, T. J., and Duester, G. (2015) Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat Rev Mol Cell Biol* **16**, 110-123
  93. Fujiki, R., Chikanishi, T., Hashiba, W., Ito, H., Takada, I., Roeder, R. G., Kitagawa, H., and Kato, S. (2009) GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. *Nature* **459**, 455-459

94. Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006) Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* **13**, 713-719
95. Cao, R., and Zhang, Y. (2004) SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex. *Molecular Cell* **15**, 57-67
96. Montgomery, N. D., Yee, D., Montgomery, S. A., and Magnuson, T. (2007) Molecular and Functional Mapping of EED Motifs Required for PRC2-Dependent Histone Methylation. *Journal of Molecular Biology* **374**, 1145-1157
97. Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., and Zhang, Y. (2001) Purification and Functional Characterization of a Histone H3-Lysine 4-Specific Methyltransferase. *Molecular Cell* **8**, 1207-1217
98. Tyagi, S., Chabes, A. L., Wysocka, J., and Herr, W. (2007) E2F Activation of S Phase Promoters via Association with HCF-1 and the MLL Family of Histone H3K4 Methyltransferases. *Molecular Cell* **27**, 107-119
99. Machida, Y. J., Machida, Y., Vashisht, A. A., Wohlschlegel, J. A., and Dutta, A. (2009) The Deubiquitinating Enzyme BAP1 Regulates Cell Growth via Interaction with HCF-1. *Journal of Biological Chemistry* **284**, 34179-34188
100. Luciano, R. L., and Wilson, A. C. (2003) HCF-1 Functions as a Coactivator for the Zinc Finger Protein Krox20. *The Journal of biological chemistry* **278**, 51116-51124
101. Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N., and Herr, W. (2003) Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes & Development* **17**, 896-911

102. Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W., and Cleary, M. L. (2004) Leukemia Proto-Oncoprotein MLL Forms a SET1-Like Histone Methyltransferase Complex with Menin To Regulate Hox Gene Expression. *Molecular and Cellular Biology* **24**, 5639-5649
103. Zhou, P., Wang, Z., Yuan, X., Zhou, C., Liu, L., Wan, X., Zhang, F., Ding, X., Wang, C., Xiong, S., Wang, Z., Yuan, J., Li, Q., and Zhang, Y. (2013) Mixed Lineage Leukemia 5 (MLL5) Protein Regulates Cell Cycle Progression and E2F1-responsive Gene Expression via Association with Host Cell Factor-1 (HCF-1). *Journal of Biological Chemistry* **288**, 17532-17543
104. Grozeva, D., Carss, K., Spasic-Boskovic, O., Tejada, M.-I., Gecz, J., Shaw, M., Corbett, M., Haan, E., Thompson, E., Friend, K., Hussain, Z., Hackett, A., Field, M., Renieri, A., Stevenson, R., Schwartz, C., Floyd, J. A. B., Bentham, J., Cosgrove, C., Keavney, B., Bhattacharya, S., Italian, X. I. M. R. P., Consortium, U. K., Consortium, G., Hurles, M., and Raymond, F. L. (2015) Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Human Mutation* **36**, 1197-1204
105. Ellery, P. M., Ellis, R. J., and Holder, S. E. (2014) Interstitial 3p25 deletion in a patient with features of 3p deletion syndrome: further evidence for the role of SRGAP3 in mental retardation. *Clinical Dysmorphology* **23**, 29-31
106. Kuechler, A., Zink, A. M., Wieland, T., Ludecke, H.-J., Cremer, K., Salviati, L., Magini, P., Najafi, K., Zweier, C., Czeschik, J. C., Aretz, S., Endeley, S., Tamburrino, F., Pinato, C., Clementi, M., Gundlach, J., Maylahn, C., Mazzanti, L., Wohlleber, E., Schwarzmayr, T., Kariminejad, R., Schlessinger, A., Wieczorek, D., Strom, T. M., Novarino, G., and Engels, H. (2015) Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. *Eur J Hum Genet* **23**, 753-760

107. Szczałuba, K., Brzezinska, M., Kot, J., Rydzanicz, M., Walczak, A., Stawiński, P., Werner, B., and Płoski, R. (2016) SETD5 loss-of-function mutation as a likely cause of a familial syndromic intellectual disability with variable phenotypic expression. *American Journal of Medical Genetics Part A* **170**, 2322-2327
108. Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., Vorstman, Jacob A. S., Thompson, A., Regan, R., Pilorge, M., Pellecchia, G., Pagnamenta, Alistair T., Oliveira, B., Marshall, Christian R., Magalhaes, Tiago R., Lowe, Jennifer K., Howe, Jennifer L., Griswold, Anthony J., Gilbert, J., Duketis, E., Dombroski, Beth A., De Jonge, Maretha V., Cuccaro, M., Crawford, Emily L., Correia, Catarina T., Conroy, J., Conceição, Inês C., Chiocchetti, Andreas G., Casey, Jillian P., Cai, G., Cabrol, C., Bolshakova, N., Bacchelli, E., Anney, R., Gallinger, S., Cotterchio, M., Casey, G., Zwaigenbaum, L., Wittemeyer, K., Wing, K., Wallace, S., van Engeland, H., Tryfon, A., Thomson, S., Soorya, L., Rogé, B., Roberts, W., Poustka, F., Mougá, S., Minshew, N., McInnes, L. A., McGrew, Susan G., Lord, C., Leboyer, M., Le Couteur, Ann S., Kolevzon, A., Jiménez González, P., Jacob, S., Holt, R., Guter, S., Green, J., Green, A., Gillberg, C., Fernandez, B. A., Duque, F., Delorme, R., Dawson, G., Chaste, P., Café, C., Brennan, S., Bourgeron, T., Bolton, Patrick F., Bölte, S., Bernier, R., Baird, G., Bailey, Anthony J., Anagnostou, E., Almeida, J., Wijsman, Ellen M., Vieland, Veronica J., Vicente, Astrid M., Schellenberg, Gerard D., Pericak-Vance, M., Paterson, Andrew D., Parr, Jeremy R., Oliveira, G., Nurnberger, John I., Monaco, Anthony P., Maestrini, E., Klauck, Sabine M., Hakonarson, H., Haines, Jonathan L., Geschwind, Daniel H., Freitag, Christine M., Folstein, Susan E., Ennis, S., Coon, H., Battaglia, A., Szatmari, P., Sutcliffe, James S., Hallmayer, J., Gill, M., Cook, Edwin H., Buxbaum, Joseph D., Devlin, B., Gallagher, L., Betancur, C., and Scherer,



- Stephen W. (2014) Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. *The American Journal of Human Genetics* **94**, 677-694
109. Huang, L., Jolly, Lachlan A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., Shoubbridge, C., Wieczorek, D., Tzschach, A., Cohen, M., Hackett, A., Field, M., Froyen, G., Hu, H., Haas, Stefan A., Ropers, H.-H., Kalscheuer, Vera M., Corbett, Mark A., and Gecz, J. (2012) A Noncoding, Regulatory Mutation Implicates HCFC1 in Nonsyndromic Intellectual Disability. *The American Journal of Human Genetics* **91**, 694-702
  110. Yu, H.-C., Sloan, Jennifer L., Scharer, G., Brebner, A., Quintana, Anita M., Achilly, Nathan P., Manoli, I., Coughlin, Curtis R., Geiger, Elizabeth A., Schneck, U., Watkins, D., Suormala, T., Van Hove, Johan L. K., Fowler, B., Baumgartner, Matthias R., Rosenblatt, David S., Venditti, Charles P., and Shaikh, Tamim H. (2013) An X-Linked Cobalamin Disorder Caused by Mutations in Transcriptional Coregulator HCFC1. *The American Journal of Human Genetics* **93**, 506-514
  111. Jolly, L. A., Nguyen, L. S., Domingo, D., Sun, Y., Barry, S., Hancarova, M., Plevova, P., Vlckova, M., Havlovicova, M., Kalscheuer, V. M., Graziano, C., Pippucci, T., Bonora, E., Sedlacek, Z., and Gecz, J. (2015) HCFC1 loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain. *Human Molecular Genetics* **24**, 3335-3347
  112. Pijnappel, W. W. M. P., Schaft, D., Roguev, A., Shevchenko, A., Tekotte, H., Wilm, M., Rigaut, G., Séraphin, B., Aasland, R., and Stewart, A. F. (2001) The *S. cerevisiae* SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. *Genes & Development* **15**, 2991-3004
  113. Rincon-Arano, H., Halow, J., Delrow, J. J., Parkhurst, S. M., and Groudine, M. (2012) UpSET recruits HDAC complexes and restricts chromatin accessibility and histone acetylation at

- promoter regions. *Cell* **151**, 1214-1228
114. Seifert, A., Werheid, D. F., Knapp, S. M., and Tobiasch, E. (2015) Role of Hox genes in stem cell differentiation. *World Journal of Stem Cells* **7**, 583-595
  115. Pearson, J. C., Lemons, D., and McGinnis, W. (2005) Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* **6**, 893-904
  116. Gavalas, A., Ruhrberg, C., Livet, J., Henderson, C. E., and Krumlauf, R. (2003) Neuronal defects in the hindbrain of *Hoxa1*, *Hoxb1* and *Hoxb2* mutants reflect regulatory interactions among these Hox genes. *Development* **130**, 5663
  117. Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., and Lander, E. S. (2006) A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315-326
  118. Martinez-Ceballos, E., Chambon, P., and Gudas, L. J. (2005) Differences in Gene Expression between Wild Type and *Hoxa1* Knockout Embryonic Stem Cells after Retinoic Acid Treatment or Leukemia Inhibitory Factor (LIF) Removal. *Journal of Biological Chemistry* **280**, 16484-16498
  119. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921
  120. Pertea, M., and Salzberg, S. L. (2010) Between a chicken and a grape: estimating the number of human genes. *Genome Biology* **11**, 206
  121. Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843-854
  122. Winter, J., Jung, S., Keller, S., Gregory, R. I., and Diederichs, S. (2009) Many roads to

- maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* **11**, 228-234
123. Bartel, D. P. (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281-297
  124. Lee, Y., Jeon, K., Lee, J. T., Kim, S., and Kim, V. N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO Journal* **21**, 4663
  125. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V. N. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419
  126. Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development* **17**, 3011-3016
  127. Zeng, Y., and Cullen, B. R. (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Research* **32**, 4776-4785
  128. Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Bálint, É., Tuschl, T., and Zamore, P. D. (2001) A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the *let-7* Small Temporal RNA. *Science* **293**, 834
  129. Gregory, R. I., Chendrimada, T. P., Cooch, N., and Shiekhattar, R. (2005) Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing. *Cell* **123**, 631-640
  130. Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**, 102-114
  131. Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* **120**, 15-20

132. Lewis, B. P., Shih, I. h., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. Prediction of Mammalian MicroRNA Targets. *Cell* **115**, 787-798
133. Hutvágner, G., and Zamore, P. D. (2002) A microRNA in a Multiple-Turnover RNAi Enzyme Complex. *Science* **297**, 2056
134. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004) Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs. *Molecular Cell* **15**, 185-197
135. Wu, L., Fan, J., and Belasco, J. G. (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4034-4039
136. Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V., and Hannon, G. J. (2003) Dicer is essential for mouse development. *Nat Genet* **35**, 215-217
137. Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* **39**, 380-385
138. Morita, S., Horii, T., Kimura, M., Goto, Y., Ochiya, T., and Hatada, I. (2007) One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. *Genomics* **89**, 687-696
139. Kanellopoulou, C., Muljo, S. A., Kung, A. L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D. M., and Rajewsky, K. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes & Development* **19**, 489-501
140. Morin, R. D., O'Connor, M. D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-L.,

- Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Eaves, C. J., and Marra, M. A. (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Research* **18**, 610-621
141. Stadler, B., Ivanovska, I., Mehta, K., Song, S., Nelson, A., Tan, Y., Mathieu, J., Darby, C., Blau, C. A., Ware, C., Peters, G., Miller, D. G., Shen, L., Cleary, M. A., and Ruohola-Baker, H. (2010) Characterization of microRNAs Involved in Embryonic Stem Cell States. *Stem Cells and Development* **19**, 935-950
  142. Marson, A., Levine, S. S., Cole, M. F., Frampton, G. M., Brambrink, T., Johnstone, S., Guenther, M. G., Johnston, W. K., Wernig, M., Newman, J., Calabrese, J. M., Dennis, L. M., Volkert, T. L., Gupta, S., Love, J., Hannett, N., Sharp, P. A., Bartel, D. P., Jaenisch, R., and Young, R. A. (2008) Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells. *Cell* **134**, 521-533
  143. Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J. E., Baehner, L., and Blelloch, R. (2008) Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet* **40**, 1478-1483
  144. Ma, Y., Yao, N., Liu, G., Dong, L., Liu, Y., Zhang, M., Wang, F., Wang, B., Wei, X., Dong, H., Wang, L., Ji, S., Zhang, J., Wang, Y., Huang, Y., and Yu, J. (2015) Functional screen reveals essential roles of miR-27a/24 in differentiation of embryonic stem cells. *The EMBO Journal* **34**, 361-378
  145. Wienholds, E., Kloosterman, W. P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., Horvitz, H. R., Kauppinen, S., and Plasterk, R. H. A. (2005) MicroRNA Expression in Zebrafish Embryonic Development. *Science* **309**, 310
  146. Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., Rheinheimer, S.,

- Meder, B., Stähler, C., Meese, E., and Keller, A. (2016) Distribution of miRNA expression across human tissues. *Nucleic Acids Research* **44**, 3865-3877
147. Alvarez-Garcia, I., and Miska, E. A. (2005) MicroRNA functions in animal development and human disease. *Development* **132**, 4653
  148. Obernosterer, G., Leuschner, P. J. F., Alenius, M., and Martinez, J. (2006) Post-transcriptional regulation of microRNA expression. *RNA* **12**, 1161-1167
  149. Gulyaeva, L. F., and Kushlinskiy, N. E. (2016) Regulatory mechanisms of microRNA expression. *Journal of Translational Medicine* **14**, 143
  150. Kloosterman, W. P., and Plasterk, R. H. A. (2006) The Diverse Functions of MicroRNAs in Animal Development and Disease. *Developmental Cell* **11**, 441-450
  151. Sayed, D., and Abdellatif, M. (2011) MicroRNAs in Development and Disease. *Physiological Reviews* **91**, 827
  152. Croce, C. M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nature reviews. Genetics* **10**, 704-714
  153. Suzuki, H., Maruyama, R., Yamamoto, E., and Kai, M. (2012) DNA methylation and microRNA dysregulation in cancer. *Molecular Oncology* **6**, 567-578
  154. Pahlich, S., Zakaryan, R. P., and Gehring, H. (2006) Protein arginine methylation: Cellular functions and methods of analysis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1764**, 1890-1903
  155. Hughes, R. M., and Waters, M. L. (2006) Arginine Methylation in a  $\beta$ -Hairpin Peptide: Implications for Arg- $\pi$  Interactions,  $\Delta C_p^\circ$ , and the Cold Denatured State. *Journal of the American Chemical Society* **128**, 12735-12742
  156. Boisvert, F.-M., Chénard, C. A., and Richard, S. (2005) Protein Interfaces in Signaling

Regulated by Arginine Methylation. *Science* **2005**, re2

157. Gayatri, S., and Bedford, M. T. (2014) Readers of histone methylarginine marks. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1839**, 702-710
158. Lee, H. W., Kim, S., and Paik, W. K. (1977) S-Adenosylmethionine:protein-arginine methyltransferase. Purification and mechanism of the enzyme. *Biochemistry* **16**, 78-85
159. Blanc, R. S., and Richard, S. (2017) Arginine Methylation: The Coming of Age. *Molecular Cell* **65**, 8-24
160. Lee, J.-H., Cook, J. R., Yang, Z.-H., Mirochnitchenko, O., Gunderson, S. I., Felix, A. M., Herth, N., Hoffmann, R., and Pestka, S. (2005) PRMT7, a New Protein Arginine Methyltransferase That Synthesizes Symmetric Dimethylarginine. *Journal of Biological Chemistry* **280**, 3656-3664
161. Zurita-Lopez, C. I., Sandberg, T., Kelly, R., and Clarke, S. G. (2012) Human Protein Arginine Methyltransferase 7 (PRMT7) Is a Type III Enzyme Forming  $\omega$ -N(G)-Monomethylated Arginine Residues. *The Journal of Biological Chemistry* **287**, 7859-7870
162. Howe, E. N., Cochrane, D. R., and Richer, J. K. (2012) The miR-200 and miR-221/222 microRNA Families: Opposing Effects on Epithelial Identity. *Journal of Mammary Gland Biology and Neoplasia* **17**, 65-77
163. Ran, F. A., Hsu, Patrick D., Lin, C.-Y., Gootenberg, Jonathan S., Konermann, S., Trevino, A. E., Scott, David A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380-1389
164. Li, J., Bei, Y., Liu, Q., Lv, D., Xu, T., He, Y., Chen, P., and Xiao, J. (2015) MicroRNA-221 is Required for Proliferation of Mouse Embryonic Stem Cells via P57 Targeting. *Stem Cell Reviews and Reports* **11**, 39-49

165. Galardi, S., Mercatelli, N., Giorda, E., Massalini, S., Frajese, G. V., Ciafrè, S. A., and Farace, M. G. (2007) miR-221 and miR-222 Expression Affects the Proliferation Potential of Human Prostate Carcinoma Cell Lines by Targeting p27Kip1. *Journal of Biological Chemistry* **282**, 23716-23724
166. Fornari, F., Gramantieri, L., Ferracin, M., Veronese, A., Sabbioni, S., Calin, G. A., Grazi, G. L., Giovannini, C., Croce, C. M., Bolondi, L., and Negrini, M. (2008) MiR-221 controls CDKN1C/p57 and CDKN1B/p27 expression in human hepatocellular carcinoma. *Oncogene* **27**, 5651-5661
167. Garofalo, M., Leva, G. D., Romano, G., Nuovo, G., Suh, S.-S., Ngankee, A., Taccioli, C., Pichiorri, F., Alder, H., Secchiero, P., Gasparini, P., Gonelli, A., Costinean, S., Acunzo, M., Condorelli, G., and Croce, C. M. (2009) MiR-221&222 regulate TRAIL-resistance and enhance tumorigenicity through PTEN and TIMP3 down-regulation. *Cancer cell* **16**, 498-509
168. Pogribny, I. P., Filkowski, J. N., Tryndyak, V. P., Golubov, A., Shpileva, S. I., and Kovalchuk, O. (2010) Alterations of microRNAs and their targets are associated with acquired resistance of MCF-7 breast cancer cells to cisplatin. *International Journal of Cancer* **127**, 1785-1794
169. Zhou, M., Liu, Z., Zhao, Y., Ding, Y., Liu, H., Xi, Y., Xiong, W., Li, G., Lu, J., Fodstad, O., Riker, A. I., and Tan, M. (2010) MicroRNA-125b Confers the Resistance of Breast Cancer Cells to Paclitaxel through Suppression of Pro-apoptotic Bcl-2 Antagonist Killer 1 (Bak1) Expression. *The Journal of Biological Chemistry* **285**, 21496-21507
170. Park, J.-K., Lee, E. J., Esau, C., and Schmittgen, T. D. (2009) Antisense Inhibition of microRNA-21 or -221 Arrests Cell Cycle, Induces Apoptosis, and Sensitizes the Effects of



Gemcitabine in Pancreatic Adenocarcinoma. *Pancreas* **38**, e190-e199

171. Gullà, A., Di Martino, M. T., Cantafio, M. E. G., Morelli, E., Amodio, N., Botta, C., Pitari, M. R., Lio, S. G., Britti, D., Stamato, M. A., Hideshima, T., Munshi, N. C., Anderson, K. C., Tagliaferri, P., and Tassone, P. (2016) A 13 mer LNA-i-miR-221 inhibitor restores drug-sensitivity in melphalan-refractory multiple myeloma cells. *Clinical cancer research : an official journal of the American Association for Cancer Research* **22**, 1222-1233
172. Pandey, A., Singh, P., Jauhari, A., Singh, T., Khan, F., Pant, A. B., Parmar, D., and Yadav, S. (2015) Critical role of the miR-200 family in regulating differentiation and proliferation of neurons. *Journal of Neurochemistry* **133**, 640-652
173. Hamada, N., Fujita, Y., Kojima, T., Kitamoto, A., Akao, Y., Nozawa, Y., and Ito, M. (2012) MicroRNA expression profiling of NGF-treated PC12 cells revealed a critical role for miR-221 in neuronal differentiation. *Neurochemistry International* **60**, 743-750
174. Ivey, K. N., and Srivastava, D. (2010) MicroRNAs as Regulators of Differentiation and Cell Fate Decisions. *Cell Stem Cell* **7**, 36-41
175. Melton, C., Judson, R. L., and Blelloch, R. (2010) Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**, 621-626
176. Singh, S. K., Kagalwala, M. N., Parker-Thornburg, J., Adams, H., and Majumder, S. (2008) REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature* **453**, 223-227
177. Tay, Y. M. S., Tam, W.-L., Ang, Y.-S., Gaughwin, P. M., Yang, H., Wang, W., Liu, R., George, J., Ng, H.-H., Perera, R. J., Lufkin, T., Rigoutsos, I., Thomson, A. M., and Lim, B. (2008) MicroRNA-134 Modulates the Differentiation of Mouse Embryonic Stem Cells, Where It Causes Post-Transcriptional Attenuation of Nanog and LRH1. *STEM CELLS* **26**, 17-29

178. Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J. A., and Kosik, K. S. (2009) MicroRNA-145 Regulates OCT4, SOX2, and KLF4 and Represses Pluripotency in Human Embryonic Stem Cells. *Cell* **137**, 647-658
179. Di Leva, G., Gasparini, P., Piovan, C., Nganheu, A., Garofalo, M., Taccioli, C., Iorio, M. V., Li, M., Volinia, S., Alder, H., Nakamura, T., Nuovo, G., Liu, Y., Nephew, K. P., and Croce, C. M. (2010) MicroRNA Cluster 221-222 and Estrogen Receptor  $\alpha$  Interactions in Breast Cancer. *JNCI: Journal of the National Cancer Institute* **102**, 706-721
180. Stinson, S., Lackner, M. R., Adai, A. T., Yu, N., Kim, H.-J., O'Brien, C., Spoerke, J., Jhunjhunwala, S., Boyd, Z., Januario, T., Newman, R. J., Yue, P., Bourgon, R., Modrusan, Z., Stern, H. M., Warming, S., de Sauvage, F. J., Amler, L., Yeh, R.-F., and Dornan, D. (2011) TRPS1 Targeting by miR-221/222 Promotes the Epithelial-to-Mesenchymal Transition in Breast Cancer. *Science Signaling* **4**, ra41
181. Fornari, F., Milazzo, M., Galassi, M., Callegari, E., Veronese, A., Miyaaki, H., Sabbioni, S., Mantovani, V., Marasco, E., Chieco, P., Negrini, M., Bolondi, L., and Gramantieri, L. (2014) p53/mdm2 Feedback Loop Sustains miR-221 Expression and Dictates the Response to Anticancer Treatments in Hepatocellular Carcinoma. *Molecular Cancer Research* **12**, 203

## VITA

Tsai-Yu Chen was born in Taipei, Taiwan on September 20, 1982, the son of Mei-Yin Lee and Chu-I Chen. He entered National Taiwan University in Taipei, Taiwan in 2001. He received his degree of Bachelor of Science in Pharmacy from National Taiwan University in 2005 and later obtained the degree of Master of Science from National Taiwan University in 2007 with a major in Pharmacology. From 2007 to 2010, he worked as a registered pharmacist in Koo Foundation Sun Yat-Sen Cancer Center, Taipei, Taiwan. In August of 2010 he entered the Ph.D. program in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences.

Permanent address:

7F.-2, No.2-6, Aly. 9, Ln. 244, Sec. 3, Roosevelt Rd., Taipei, 100, Taiwan