

**THE EPIGENETIC MECHANISMS OF POLYCOMB AND TRITHORAX PROTEINS
IN STEM CELL MAINTENANCE AND LEUKEMOGENESIS**

By

Mohammad Aljazi

A DISSERTATION

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

Biochemistry and Molecular Biology - Doctor of Philosophy

2022

ABSTRACT

THE EPIGENETIC MECHANISMS OF POLYCOMB AND TRITHORAX PROTEINS IN STEM CELL MAINTENANCE AND LEUKEMOGENESIS

By

Mohammad Aljazi

Functionally, Polycomb repressive complex 2 (PRC2) mediates transcriptional repression of differentiation genes critical for mouse embryonic stem cell (mESC) maintenance. Culturing mESCs in 2i serum-free medium inhibits FGF/ERK signaling and activates the Wnt/ β -catenin pathway, which induces a naive cell state characterized by a reduced expression of lineage-specific genes. Interestingly, in naive mESCs, both PRC2 chromatin occupancy and the repressive histone3 lysine 27 trimethylation (H3K27me3) modification they mediate are largely depleted. To explore the molecular mechanism attributing to the transcriptional changes in naive cells, we performed RNA-sequencing for mESCs cultured in serum versus 2i medium. Gene expression analysis revealed reduced *Jarid2* levels in naive mESCs. Reactivation of FGF/ERK signaling caused elevated *Jarid2* expression, whereas ERK1/2 deletion decreased its expression levels. Ectopic expression of ERK depleted cells restored *Jarid2* expression, showing that *Jarid2* expression dependent on ERK signaling. Using ChIP-seq analysis, we observed reduced occupancy for *Jarid2* and PRC2 and decreased H3K27me3 levels in both naive and Erk1/2 depleted mESCs. Expression of *Jarid2* in Erk1/2 depleted cells reestablished PRC2 occupancy and H3K27me3 modifications. Taken together, these results reveal the molecular mechanism associated with FGF/ERK signaling and PRC2 recruitment in mESCs.

The TrxG (Trithorax) group member ASH1L serves as a regulator of cell development. However, its functional role in MLL-rearranged leukemia initiation and maintenance is not well understood. Using an Ash1L conditional knockout mouse model, we demonstrated that ASH1L in

hematopoietic progenitor cells (HPCs) impaired initiation of MLL-AF9-induced leukemic transformation *in vitro*. Ablation of ASH1L in the MLL-AF9-transformed leukemia cells impeded maintenance *in vitro* and leukemia progression *in vivo*. Furthermore, Ash1L depleted cells expressing wild-type ASH1L rescued MLL-AF9-induced leukemia transformation, while the transformation of cells expressing enzymatically inactive ASH1L inhibited their maintenance. Implementation of RNA-sequencing analysis revealed that ASH1L controls the expression of MLL-AF9 target genes by occupying their promoters and depositing H3K36me2 marks at these sites. Altogether, these results demonstrate that the enzymatic activity of ASH1L is crucial for MLL-AF9-induced leukemic transformation and maintenance. In addition, our study identifies a potential therapeutic target in MLL-AF9-induced leukemias.

Histone post-translational modifications are vital for epigenetic mediated gene regulation. While past studies have characterized the functional role of many histone H3 lysine residues modifications, the post-translational modification of histone H3 lysine 37 and the factors contributing to these modifications remain undefined in mammals. Using *in vitro* methyltransferase assays, we found that SMYD family member 5 (SMYD5) catalyzes mono-methylation of H3 lysine 36 and 37 (H3K36/K37me1). Mutation of the conserved histidine within the catalytic SET domain abolished SMYD5 methyltransferase activity *in vitro*. Additionally, loss of Smyd5 in mESCs reduces the global histone H3K37me1 level in cells. Thus, our data functionally identifies SMYD5 as an H3 specific methyltransferase that mediates H3K36me/H3K37me1 *in vitro*. It also reveals that SMYD5 serves as one of the histone methyltransferases catalyzing histone H3K37me1 *in vivo*.

ACKNOWLEDGMENTS

First, I would like to thank my advisor Dr. Jin He for giving me the opportunity to join your laboratory. I am deeply grateful for your mentorship, support, and patience with me. Your insight and critical feedback helped me grow as a graduate student and scientist. My achievements would not have been possible without your invaluable expertise and guidance. I cannot thank you enough for the time, guidance, and encouragement you provided me throughout my graduate career.

I would also like to thank my committee members, Dr. Ronald Henry, Dr. Jason Knott, Dr. Min-Hao Kuo, and Dr. George Mias. I cannot express how grateful I am for your guidance and insight. Your advice helped direct me through the difficulties I faced throughout my experiments.

I would like to thank the members of my lab, Yuen Gao and Yan Wu. The experimental work conducted would not have been achieved without your technical support. Thank you for your time and dedication in providing me with the assistance I needed. It was a pleasure having the opportunity to work alongside you.

Last but not least, I would like to thank all my family and friends that have supported me throughout my graduate school years. Thank you for always being there for me. Your encouragement and support, especially through difficult times, help motivate me.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
KEY TO ABBREVIATIONS	xi
CHAPTER 1: INTRODUCTION.....	1
INTRODUCTION.....	2
POLYCOMB REPRESSIVE COMPLEX STRUCTURE	3
CATALYTIC FUNCTION OF POLYCOMB REPRESSIVE COMPLEXES.....	5
POLYCOMB REPRESSIVE COMPLEX RECRUITMENT	5
THE ROLE POLYCOMB REPRESSIVE COMPLEXES IN TRANSCRIPTION	
CONTROL.....	8
FUNCTION OF POLYCOMB REPRESSIVE COMPLEXES IN STEM CELLS.....	8
TRITHORAX GROUP COMPLEXES	10
RECRUITMENT OF TRITHORAX GROUP COMPLEXES	13
FUNCTIONAL ROLE TRITHORAX GROUP COMPLEXES IN TRANSCRIPTION	
CONTROL.....	14
THE ROLE OF TRITHORAX GROUP COMPLEXES IN STEM CELL	
REGULATION AND CANCER DEVELOPMENT	15
MLL1 COMPLEX POST-RANSCRIPTIONAL PROCESSING	16
ROLE MLL REARRANGMENTS IN LEUKEMOGENESIS	18
REFERENCES.....	22
CHAPTER 2: CELL SIGNALING COORDINATES GLOBAL PRC2 RECRUITMENT	
AND DEVELOPMENTAL GENE EXPRESSION IN MURINE EMBRYONIC STEM	
CELLS	30
ABSTRACT	31
INTRODUCTION.....	32
MATERIALS METHODS	34
Mouse embryonic stem cell culture.....	34
Western blot analysis.....	34
RT-qPCR assays	35
Cell cycle analysis	37
Lentiviral vector generation and infection.....	37
Crispr-mediate Erk1/Erk2 gene knock-out in mESCs	37
ChIP-Seq sample preparation	38
ChIP DNA preparation for HISEQ4000 sequencing	39
ChIP-Seq data analysis	39
RNA-seq sample preparation for HiSeq4000 sequencing.....	40
RNA-Seq data analysis.....	40
RESULTS.....	41

<i>Jarid2</i> expression is significantly reduced in naive mESCs	41
FGF/ERK signaling positively regulates <i>Jarid2</i> expression in mESCs.....	45
Knockout of <i>Erk1/Erk2</i> reduces <i>Jarid2</i> expression in mESCs.....	50
The global PRC2 occupancy at CGIs is largely reduced in naive mESCs	54
Ectopic expression of <i>Jarid2</i> restores the global PRC2 occupancy in naive mESCs.....	58
Ectopic expression of <i>Jarid2</i> restores the global PRC2 occupancy in the <i>Erk1/Erk2</i> - dKO mESCs	61
De-repression of bivalent genes appears to be determined by the presence of signaling- associated transcription factors but not the status of PRC2 occupancy in naive mESCs	64
DISCUSSION	70
LIMITATIONS OF THE STUDY	77
REFERENCES	78
 CHAPTER 3: HISTONE H3K36ME2-SPECIFIC METHYLTRANSFERASE ASH1L PROMOTES MLL-AF9-INDUCED LEUKEMOGENESIS.....	
ABSTRACT	83
INTRODUCTION.....	84
MATERIALS METHODS	85
Mice.....	87
Hematopoietic progenitor isolation and culture	87
Retroviral and lentiviral vector production and transduction.....	88
Serial methylcellulose replating assay and leukemia transplantation.....	88
FACS analysis	89
Western Blot analysis	89
Quantitative RT-PCR and ChIP-qPCR assays.....	90
RNA-seq sample preparation for HiSeq4000 sequencing.....	92
RNA-seq data analysis.....	92
Statistical analysis.....	93
RESULTS.....	93
ASH1L promotes the initiation of MLL-AF9-induced leukemic transformation <i>in vitro</i>	93
ASH1L facilitates the maintenance of MLL-AF9-induced leukemic cells <i>in vitro</i>	96
ASH1L promotes the MLL-AF9-induced leukemia development <i>in vivo</i>	98
The enzymatic activity of ASH1L is required for its function in promoting MLL-AF9- induced leukemic transformation	100
ASH1L facilitates the MLL-AF9-induced leukemogenic gene expression.....	102
ASH1L binds and mediates the histone H3K36me2 modification at <i>Hoxa9</i> and <i>Hoxa10</i> gene promoters.....	106
DISCUSSION	108
REFERENCES	111
 CHAPTER 4: SMYD5 IS A HISTONE H3-SPECIFIC METHYLTRANSFERASE MEDIATING MONO-METHYLATION OF HISTONE H3 LYSINE 36 AND 37	
ABSTRACT	115
INTRODUCTION.....	116
	117

MATERIALS METHODS	118
Mouse embryonic stem cell culture.....	118
Crispr/Cas9-mediated <i>Smyd5</i> gene knockout in mESCs	118
Recombinant SMYD5 protein purification	118
Recombinant histone H3 and H4 peptide purification.....	119
<i>In vitro</i> histone methyltransferase assays	120
Histone extraction.....	120
Western blot analysis.....	121
Mass spectrometry analysis	121
RESULTS.....	123
SMYD5 mediates methyl group transferring towards histone H3 <i>in vitro</i>	123
SMYD5 catalyzes mono-methylation of histone H3 lysine 36 and lysine 37 <i>in vitro</i> ...	125
A species-conserved histidine in the SET domain is required for the SMYD5 methyltransferase activity.....	128
Deletion of <i>Smyd5</i> in mESCs partially reduces the global histone H3K37me1 level...	130
DISCUSSION	132
REFERENCES	134
 CHAPTER 5: CONCLUSION.....	 137
OVERVIEW	138
CHAPTER 2: CELL SIGNALING COORDINATES GLOBAL PRC2 RECRUITMENT AND DEVELOPMENTAL GENE EXPRESSION IN MURINE EMBRYONIC STEM CELLS	138
CHAPTER 3: HISTONE H3K36ME2-SPECIFIC METHYLTRANSFERASE ASH1L PROMOTES THE MLL-AF9-INDUCED LEUKEMOGENESIS.....	139
CHAPTER 4: SMYD5 IS A HISTONE H3-SPECIFIC METHYLTRANSFERASE MEDIATING MONO-METHYLATION OF HISTONE H3 LYSINE 36 AND 37....	140
FUTURE WORK	140
REFERENCES	143

LIST OF TABLES

Table 1: Mammalian Trithorax Group Complexes and their associated factors	12
Table 2: Sequences of all primers used in PRC2 study	36
Table 3: Result of KEGG pathway analysis on the genes upregulated in response to the reactivation of FGF/ERK signaling.	48
Table 4: Result of gene ontology analysis on the downregulated bivalent genes in naive mESCs	67
Table 5: Result of gene ontology analysis on the upregulated bivalent genes in naive mESCs...	69
Table 6: Primes used in ASH1L study.....	91
Table 7: Result of gene ontology enrichment analysis of genes upregulated in the MLL-AF9-transformed cells.	104
Table 8:Result of gene ontology enrichment analysis of genes down-regulated in the MLL-AF9-transformed cells.	105

LIST OF FIGURES

Figure 1: Assembly of Polycomb repressive complexes.	4
Figure 2: Recruitment mechanisms of PcG complexes.	7
Figure 3 :Schematic representation of MLL1 and its interaction partners.	17
Figure 4: MLL Fusion Proteins.....	19
Figure 5: <i>Jarid2</i> expression is significantly reduced in naive mESCs.	43
Figure 6 : <i>Jarid2</i> expression is significantly reduced in naive mESCs.	44
Figure 7: FGF/ERK signaling positively regulates <i>Jarid2</i> expression in mESCs.....	47
Figure 8: FGF/ERK signaling positively regulates the <i>Jarid2</i> expression.	49
Figure 9: Knockout of <i>Erk1/Erk2</i> reduces <i>Jarid2</i> expression in mESCs.....	52
Figure 10: Knockout of <i>Erk1/Erk2</i> reduces <i>Jarid2</i> expression in mESCs.....	53
Figure 11: The global PRC2 occupancy at CGIs is largely reduced in naive mESCs.....	56
Figure 12: The global PRC2 occupancy at CGIs is largely reduced in naive mESCs.....	57
Figure 13: Ectopic expression of <i>Jarid2</i> restores the global PRC2 occupancy in naive mESCs...59	
Figure 14: Ectopic expression of <i>Jarid2</i> restores the global PRC2 occupancy in naive mESCs...60	
Figure 15: Ectopic expression of <i>Erk2</i> or <i>Jarid2</i> restores the global PRC2 occupancy at bivalent promoters in <i>Erk1/Erk2</i> -dKO mESCs..	62
Figure 16: Ectopic expression of <i>Erk2</i> or <i>Jarid2</i> restores the global PRC2 occupancy in <i>Erk1/Erk2</i> -dKO mESCs.....	63
Figure 17: De-repression of bivalent genes is determined by the presence of signaling-associated transcription factors but not the status of PRC2 occupancy in naive mESCs.	66
Figure 18: De-repression of bivalent genes is determined by the presence of signaling-associated transcription factors but not the status of PRC2 occupancy in naive mESCs.	68
Figure 19: Proposed model: cell signaling coordinates PRC2 recruitment and developmental gene expression in mESCs.....	76

Figure 20:ASH1L is required for the initiation of MLL-AF9-induced leukemic transformation.	95
Figure 21: ASH1L is required for the maintenance of MLL-AF9-induced leukemic cells <i>in vitro</i>	97
Figure 22: ASH1L promotes the MLL-AF9-induced leukemia development <i>in vivo</i>	99
Figure 23: The enzymatic activity of ASH1L is required for its function in promoting MLL-AF9-induced leukemic transformation.	101
Figure 24: ASH1L facilitates the MLL-AF9-induced leukemogenic gene expression.	103
Figure 25: ASH1L binds and mediates histone H3K36me2 modification at <i>Hoxa9</i> and <i>Hoxa10</i> gene promoters.....	107
Figure 26: SMYD5 mediates methyl group transferring towards histone H3 <i>in vitro</i>	124
Figure 27: SMYD5 catalyzes mono-methylation of histone H3 lysine 36 and lysine 37 <i>in vitro</i>	127
Figure 28:.. A species-conserved histidine in the SET domain is required for the SMYD5 methyltransferase activity.	129
Figure 29: Deletion of <i>Smyd5</i> in mESCs partially reduces the global histone H3K37me1 level.	131

KEY TO ABBREVIATIONS

PcG, Polycomb group

Hox, Homeotic

TrxG, Trithorax group

PRC1, Polycomb Repressive Complex 1

PRC2, Polycomb Repressive Complex 2

PCGRF, Polycomb group ring finger

cPRC1, canonical Polycomb Repressive Complex 1

nPRC1, non-canonical Polycomb Repressive Complex 1

EZH1, enhancer of zeste homolog 1

EZH1, enhancer of zeste homolog 2

EED, embryonic ectoderm development

SUZ12, suppressor of zeste

CBX, chromobox

HPH, human polyhomeotic homologue

H2AK119u1, histone H2A mono-ubiquitylation of lysine 119

H3K27me2, H3 lysine 27 di-methylation

H3K27me3, H3 lysine 27 tri-methylation

H3K36me1, H3 lysine 36 mono-methylation

H3K36me2, H3 lysine 36 di-methylation

H3K4me3, H3 lysine 4 tri-methylation

H4K20me3, H4 lysine 20 tri-methylation

TFs, transcriptional factors

nRNAs, non-coding RNAs

CGIs, CpG islands

PREs, Polycomb response elements

CXXC-ZF, CXXC zinc finger

KDM2B, histone H3 lysine 36 demethylase 2B

SWI/ SNF, switch/sucrose non-fermentable complexes

BRM, Brahma

ISWI, imitation switch

CHD, Chromodomain helicase DNA-binding

COMPASS, complex proteins associated with Set1

ASH1L, Absent small and homeotic discs 1 Like

TREs, Trithorax group response elements

PAF1, polymerase-associated factor 1

LEDGF, lens epithelium-derived growth factor

HOTTIP, HOXA transcript at the distal tip

KDM2A, lysine demethylase 2A

SEC, super elongation complex

TASP1, Taspase1

HATs, various histone acetyltransferases

hMBM, high affinity MENIN binding motif

mESCs, murine embryonic stem cells

ERK, extracellular signal-regulated kinase

FGFs, fibroblast growth factors

RNA-seq, RNA-sequencing

MLLr, MLL rearrangement

KMTase, histone lysine methyltransferase

SET, Su(var)3-9, Enhancer-of-zeste and Trithorax

HPCs, hematopoietic progenitor cells

4-OHT, 4-hydroxytamoxifen

qRT-PCR, quantitative RT-PCR

TBI, total-body-irradiated

ChIP, chromatin immunoprecipitation

TSS, transcriptional starting sites

LTR, long terminal repeat

LAP, intracisternal A-particle

PTMs, post-translational modifications

SMYD5, SMYD family member 5

KMT, lysine methyltransferase

CHAPTER 1:
INTRODUCTION

INTRODUCTION

Nearly half century ago, the first Polycomb group (PcG) gene was discovered in *Drosophila*¹. Phenotypic analysis revealed that PcG proteins are crucial for embryogenesis, coordinating the expression of Homeotic (*Hox*) genes^{1,2}. Several years later, the first Trithorax group (TrxG) protein was identified, revealing another factor involved in *Hox* gene regulation. Studies showed that TrxG mutations caused transformations in embryonic segmentation by antagonizing PcG proteins^{3,4}. These findings supported the hypothesis that PcG and TrxG complexes are linked to cell memory and are important for *Hox* gene control. Since then, numerous investigations studying PcG and TrxG complexes have revealed their biological significance beyond the epigenetic control of *Hox* gene expression. PcG and TrxG complexes are associated with a broad range of cellular and developmental processes. The intricate mechanisms between PcG mediated gene repression and TrxG mediated gene activation play an essential role in regulating these processes. An abundance of PcG and TrxG interacting factors also enhance or modulate their function, attributing to greater complexity in epigenetic control. Although previous studies have provided a wealth of information on PcG and TrxG complex functions and their biological significance, further investigations are vital for elucidating the processes involved in recruitment, gene regulation, cell biology, and cancer development.

POLYCOMB REPESSIVE COMPLEX STRUCTURE

Structurally Polycomb Repressive Complexes (PRCs) assemble into two distinct groups of chromatin-modifying complexes, Polycomb repressive complex 1 (PRC1) and 2 (PRC2)⁵. Among mammals, the core subunits of PRC1 complexes are conserved, consisting of either RING1A or RING1B and a member of the Polycomb group ring finger proteins (PCGF1 – PCGF6)^{6,7} (Figure 1A). PRC1 complexes can further be subdivided into canonical complexes (cPRC1) that associate with chromobox (CBX) and human polyhomeotic homologue (HPH) proteins, whereas non-canonical complexes (ncPRC1) interact with RYBP or YAF2⁸. The three main core subunits that compose PRC2 are either enhancer of zeste homolog 1 (EZH1) or EZH2, embryonic ectoderm development (EED), suppressor of zeste (SUZ12)⁹. Similarly, PRC2 core subunits are associated with various accessory proteins, such as JARID2, PCL1-3, and AEBP2 (Figure 1B), that manipulate the function of PRC2¹⁰.

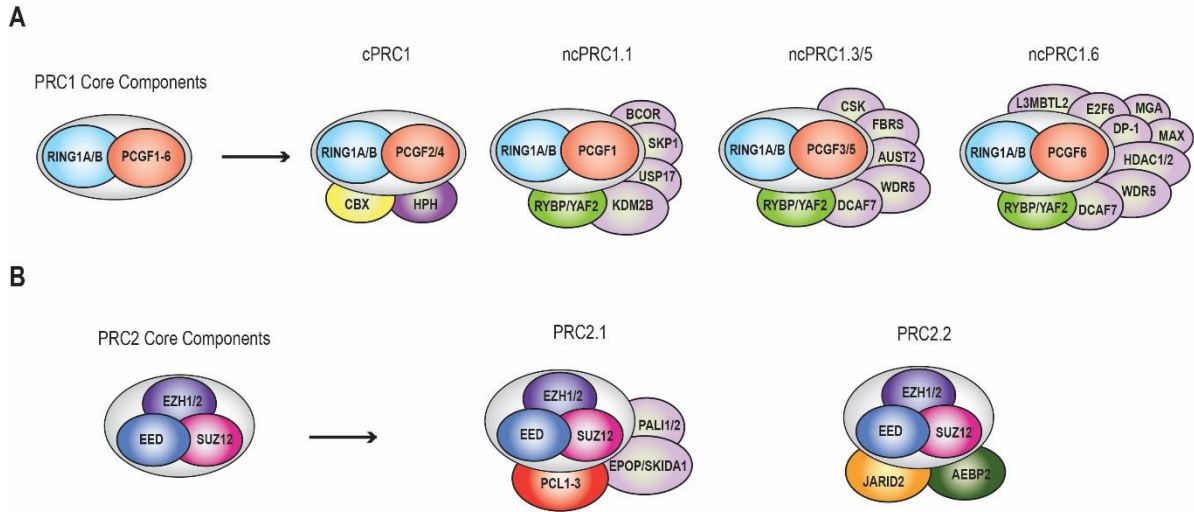


Figure 1: Assembly of Polycomb repressive complexes. (A) PRC1 consists of the core factors RING1A/1B and PCGFs. PRC1 complexes are sub-divided into the cPRC1 sub-group, consisting of PCGF2 or PCGF4 along with a CBX and HPH factor, whereas the ncPRC1 sub-groups associate with RYBP/YAF2. (B) The PRC2 core components are EZH1 or EZH2, embryonic EED, and SUZ12. These core components interact with several accessory factors, such as JARID2, AEBP2, and PCL proteins.

CATALYTIC FUNCTION OF POLYCOMB REPESSIVE COMPLEXES

Both PRC1 and PRC2 are involved in catalyzing covalent modification of histone tails. In mammalian cells, the RING1 proteins of PRC1 complexes facilitate histone H2A mono-ubiquitylation of lysine 119 (H2AK119u1) through their E3 ligase activity, and the core EZH subunit of PRC2 catalyzes histone H3 lysine 27 di- (H3K27me2), and tri-methylation (H3K27me3)^{6,11}. The functional role of PRC histone modifications is associated with transcriptional silencing, and it is well established that H3K27me3 histone modifications are markers associated with genomic repression. Yet, it remains unclear how these PRC-mediated histone modifications directly contribute to transcriptional repression.

POLYCOMB REPESSIVE COMPLEX RECRUITMENT

Though previous studies have shown that PRC complexes localize within repressed genomic regions, the mechanisms associated with their recruitment are not fully understood. In recent years, reports have provided greater insight into the processes involved in chromatin targeting and their role in gene function. Analysis of PRC core components failed to identify any distinct DNA binding motifs, suggesting interaction factors mediate PRC chromatin recruitment. Studies identified several transcriptional factors (TFs), such as REST, RUNX1, and YY1, which directly interact with PRC complexes, contributing to site-specific chromatin targeting^{12–14} (Figure 2A). Other reports have shown that non-coding RNAs (ncRNAs) also mediate PRC recruitment. For instance, XIST ncRNA involved X chromosome inactivation recruits PRC2 to the inactive X chromosome¹⁵ (Figure 2B). Although interactions between PcG complexes with various factors mediate PRC recruitment to specific genomic sites, they do not fully address what contributes to the colocalization of PRCs with CpG islands (CGIs)¹⁶.

In *Drosophila*, Polycomb response elements (PREs) serve as the DNA elements that recruit PcG complex to their target sites¹⁷. While DNA sites complementary to PREs have not been discovered in mammalian cells, genome-wide analysis of PcG proteins has revealed a strong correlation between CGIs and PRC chromatin occupancy¹⁶. PRC interacting factors harboring CXXC zinc finger (CXXC-ZF) domains that specifically recognize and bind to unmethylated CGIs play a role in their occupancy at the sites. For instance, histone H3 lysine 36 demethylase 2B (KDM2B) associates with PRC1, recruiting the complexes to these sites, and loss of KDM2B effectively reduces RING1B occupancy at CGIs¹⁸ (Figure 2C). However, while KDM2B localizes at nearly all CGIs, less than half the PRCs co-occupy these regions in ESCs¹⁹. This data suggests that alternative mechanisms contribute to chromatin recruitment of PcG proteins. Additionally, PRC1 recruitment is also dependent on initial PRC2 targeting to chromatin and depositing H3K27me3 for the binding of CBX interaction factors in PRC1 (Figure 2D)²⁰.

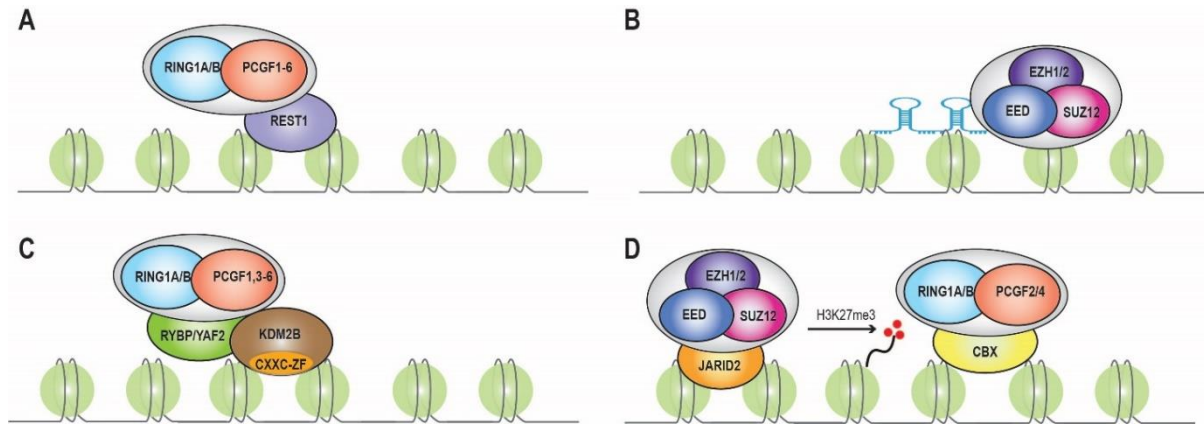


Figure 2: Recruitment mechanisms of PcG complexes. (A) Interaction of various transcription factors such as REST1 with PRC complexes directs their recruitment to target genes. (B) During X chromosome inactivation, ncRNA XIST promotes recruitment of PRC2 to the inactive X chromosome. (C) PRC complexes are recruited by interaction factors such as KDM2B, which have CXXC-ZF domains specifically recognize and bind to unmethylated CGIs. (D) PRC2 mediated H3K27me3 modifications are recognized by CBX factors and promote the recruitment of PRC1.

THE ROLE POLYCOMB REPESSIVE COMPLEXES IN TRANSCRIPTION CONTROL

Transcription factors play a crucial role in PRC-mediated gene regulation and the dynamic nature of their chromatin occupancy. Although genome-wide analysis shows that KDM2B is generally bound to most CGIs, PRC1 occupancy is predominantly observed at transcriptionally inactive promoters, demonstrating that transcriptional activity can disrupt the binding of PRC²¹. Moreover, transcriptional inhibition in mouse ESCs recruits PRC2 to the CGIs associated with transcriptionally repressed promoters²². In stem cells, genomic regions bound by PcG factors tend to harbor bivalent domains modified with H3K4me3 activating markers and H3K27me3 inactivating markers catalyzed by MLL2 and PRC2, respectively. Once the cell receives transcriptional activating cues, H3K27me3 marks are depleted from CGIs, and bivalent domains resolve to monovalent regions marked by H3K4me3. This demonstrates that PRC-mediated gene silencing results from the absence of transcription factors that promote gene activation rather than direct repression by PRCs and maintain gene suppression by increasing the transcriptional threshold.

FUNCTION OF POLYCOMB REPESSIVE COMPLEXES IN STEM CELLS

Initial PcG mutation studies in *Drosophila* lead to developmental defects, revealing that these factors were essential for homeotic gene regulation. Further investigations involving mice showed that loss of PcG function in mice is embryonic lethal, demonstrating their crucial role in the development and cell fate. Genomic mapping in stem cells found that PcG occupancy of the core subunits of PRC1 and PRC2 were depleted in terminally differentiated cells relative to their immature cell state, supporting their essential contribution to stem cell self-renewal²³. ESC lines from mice genetically mutated for PRC1 or PRC2 core components or their association factors

promoted transcriptional activation of differentiation markers²⁴. For instance, the loss of Ring1B and Ring1B proteins compromised ESC self-renewal, and cells depleted of PRC2 accessory factors, Pcl2 and Pcl3, altered the expression of pluripotency markers^{25,26}. Interestingly, loss of PRC2 core components did have a similar effect on pluripotency genes observed in Pcl2 and Pcl3 depletion, suggesting that accessory factors play a significant role in PRC2 mediated gene regulation.

In addition to their association as factors important for maintaining self-renewal in ESCs, PcG complexes are also essential in coordinating proper cellular differentiation. MLL2 and PRC2 maintain a poised repressed chromatin state by targeting promoters CGIs, establishing bivalent domains by depositing H3K4me3 and H3K27me3 marks at these sites. Activation of differentiation genes coincides with PRC eviction from gene promoters and resolution of bivalent genomic sites into monovalent H3K4me3 marked domains¹⁹. Loss of Ezh2, the catalytic subunit of PRC2 mediating H3K27me3, is required during mesodermal cell lineage differentiation²⁷. These findings support the dual role of bivalency in maintaining gene silencing in ESCs, while also serving as a process for rapid induction of cell differentiation upon transcriptional activating signals²⁸. Therefore, precise regulation of bivalent domains drives proper cell development and provides an intricate mechanism for gene expression during differentiation.

The process of PRC2 in self-renewal and differentiation largely relies on its dynamic interactions with various accessory proteins. Generally, complexes that interact with PHF1 and AEBP2 are found in differentiated cells, whereas complexes associated with MTF2, PHF19, and JARID2, are found in undifferentiated cells^{29,30}. Thus, accessory proteins contribute to the specificity by which PRC2 recruitment is targeted genomic regions and promote fine-tuned transcriptional control. Future studies still need to address how the expression of PRC complex

core proteins and their associated factors are regulated in various cell types and the molecular mechanisms involved in assembling and regulating their function.

TRITHORAX GROUP COMPLEXES

Like the PcG family of proteins, the TrxG complex also assembles into a diverse array of proteins subdivided into three classes. These include the SET domain-containing factors, the ATP chromatin remodelers, and the DNA binding TrxG factors. Many TrxG complexes regulate gene expression as part of multiprotein complexes that possess histone-modifying or nucleosome remodeling activities.

ATP-dependent chromatin remodeling complexes are conserved factors that are subdivided into three different families based on the structure of their ATPase subunit. First are the switch/sucrose non-fermentable complexes (SWI/ SNF), originally were first discovered in yeast. In *Drosophila*, TrxG member Brahma (BRM), a bromodomain-containing protein homologous SWI, was identified as a suppressor of PcG homeotic transformations³¹. The discovery of BRM and BRG1, the mammalian homologs of SWI/SNF complexes, revealed several subunits linked to the proliferation and cell cycle regulation³². Another member of ATP chromatin remodelers is the imitation switch (ISWI) family that functions in DNA repair, DNA replication, and transcriptional regulation. In humans, the ISWI complexes, known as NURF (BPTF) complexes, consist of one of the ATPase subunits, SMARCA1 or SMARCA5, which associate with several noncatalytic subunits, forming the core of these ATP chromatin remodeling complex³³. The Chromodomain helicase DNA-binding (CHD) family of chromatin modifiers comprises the third group of TrxG ATP-dependent remodeling factors. Members of the CHD

family are characterized by the presence of chromodomains and bind methylated histone lysine marks³⁴.

The Histone modifying TrxG factors are implicated in transcriptional regulation. Set1 was the first enzyme found in Yeast identified within a protein complex called complex proteins associated with Set1 (COMPASS), which catalyzes H3K4me. Later investigations found that mammalian organisms have homologous factors of Set1 that assemble into COMPASS-like complexes. COMPASS-like complexes have highly conserved core subunits that include the factors ASH2L, DPY30 RBBP5, and WDR5 that are important for their function³⁵. Additional accessory factors associate with TrxG complexes, further contributing to their diversity. For instance, MENIN and HCF1 interact with MLL1 and MLL2, whereas NCOA6, PA1, PTIP, and UTX associate with MLL3 and MLL4³⁶. The TrxG protein Absent small and homeotic discs 1 Like (ASH1L) is an epigenetic factor that also possesses a SET domain, although it does not assemble into COMPASS-like complexes. Nevertheless, the SET domain of ASH1L and MLL factors catalyze histone modifications crucial for regulating transcription³⁷.

Table 1: Mammalian Trithorax Group Complexes and their associated factors

ATP-Dependent Remodeling Complexes					
SWI/SNF (BAF/PBAF)		ISWI		CHD	
Complex	Associated Factors	Complex	Associated Factors	Complex	Associated Factors
BRM/BRG1	BAF45A-D	BPTF	RBAP46 RBAP48 SNF2L	CHD1-8	ASH2L
	BAF47				RBBP5
	BAF53A/B				WDR5
	BAF57				RBAP46/48
	BAF60A-C				MBD2/3
	BAF155				MTA1-3
	BAF170				p66
	BAF180/BAF200 BAF250A/B				
Histone-Modifying Complexes					
COMPASS-Like		COMPASS-Like		ASH1	
Complex	Associated Factors	Complex	Associated Factors	Complex	Associated Factors
MLL1/MLL2	Menin	MLL3/4	UTX	ASH1L	---
	MOF		NCOA6		
	ASH2L		PA1		
	DYP30		PTIP		
	HCF1		ASH2L		
	RBBP5		DYP30		
	WDR5		RBBP5 WDR5		

RECRUITMENT OF TRITHORX GROUP COMPLEXES

Like PcG complexes, in *Drosophila*, TrxG complexes are bound to specific DNA sites known as TrxG response elements (TREs) and are frequently found within PRE genomic regions. In mammals, investigations have not revealed the existence of TREs homologous to *Drosophila*, and the mechanisms involved in TrxG protein recruitment are far less understood. MLL1 and MLL2 both have CXXC-ZF domains, allowing them to recognize and bind to unmethylated CpG sequences. COMPASS-like complexes are also recruited indirectly to unmethylated CGIs through their association with additional factors that contain CXXC-ZF domains, such as CFP1³⁸. Thus, CGIs can serve as recruiting genomic sites for TrxG complexes and contribute to their potential colocalization with PcG complexes. The RNA polymerase II interacting factor polymerase-associated factor 1 (PAF1) associates with MLL complexes through their CXXC domain flanking regions, functioning as another recruiting mechanism between epigenetic and transactional mediating factors, which further shows that TrxG complexes attribute to transcriptional activation³⁹. Preexisting histone modifications are another mechanism for TrxG protein chromatin targeting. For instance, MLL1 can recognize and bind H3K4me3 modified histones through its PHD finger domain and binds to H3K36me2 histone marks through its association with lens epithelium-derived growth factor (LEDGF). Lastly, lncRNAs such as HOXA transcript at the distal tip (HOTTIP) play a role in TrxG complex recruitment through interaction with the COMPASS-like core component WDR5, targeting complexes to the HOXA gene locus⁴⁰. Altogether, these findings reveal that TrxG complexes, like PcG complexes, are dependent on various interacting factors to mediate their precise recruitment.

FUNCTIONAL ROLE TRITHORAX GROUP COMPLEXES IN TRANSCRIPTION CONTROL

Biochemically TrxG members of COMPASS-like family methylate H3K4 of histone tails, catalyzed through their highly conserved SET domain. In mESCs, MLL2 deposits H3K4me3 at bivalent promoters, establishing the primed genomic landscape for transcriptional regulation⁴¹. MLL1 also facilitates H3K4me3, though it modifies a specific subset of gene loci, such as *Hox* genes. Furthermore, mouse studies for MLL1 and MLL2 mutations revealed that these complexes have unique functions in ESC transcriptional gene regulation^{42,43}. MLL can also interact with males-absent-on-the-first (MOF), a key member of the MYST family that can acetylate H4K16, a chromatin marker associated with gene activation⁴⁴. Although MLL3 and MLL4 complexes are unessential for H3K4me at gene loci, these complexes deposit H3K4me on gene enhancers⁴⁵. In addition, demethylase UTX, a key component of MLL3/4 complexes, facilitates the removal of H3K27me histone marks originally deposited by PcG complexes⁴⁵. Another member of the TrxG proteins, ASH1L, catalyzes the modification of H3K36me2 within the *Hox* gene locus. The chromatin reader LEDGF, an MLL1 complex interacting factor, can recognize H3K36me2 histone marks, promoting MLL1 recruitment to their target *Hox* genes⁴⁶. On the other hand, lysine demethylase 2A (KDM2A) mediated removal of H3K36me2 histone modifications reduced LEDGF and MLL chromatin occupancy and inhibited expression of MLL target genes⁴⁷. As observed in the PcG complex, these results demonstrate that TrxG mediated transcriptional regulation is dependent on their interaction with a variety of factors.

THE ROLE OF TRITHORAX GROUP COMPLEXES IN STEM CELL REGULATION AND CANCER DEVELOPMENT

Control of self-renewal, proliferation, and differentiation in stem cells is dependent on the function of TrxG, but their roles are far less understood. In ESCs, developmental genes are acted upon by both PcG and TrxG factors, establishing bivalent promoters that function to maintain a poised repressed state. Transcriptional activating cues promote the reduction of PRC2 mediated H3K27me3 marks at bivalent domains while maintaining MLL2 H3K4me3 established markers, inducing gene activation. Loss of MLL2 in ESC causes depletion of H3K4me3 modifications and increased H3K27me3 at transcriptional start sites. In ESCs, the MLL complex subunit WDR5 interacts with the pluripotency transcription factor OCT4, targeting it to chromatin regions associated with self-renewal regulatory genes⁴⁸. Additional MLL complex members, such as DPY40 and RBBP5, were found as necessary components in ESC to neural progenitor cell differentiation⁴⁹. Targeted MLL1 knockout in the mouse hematopoietic system develops deficiencies in hematopoiesis, whereas loss of MLL2 causes infertility^{50–52}. These outcomes demonstrate the significance of TrxG function in ESC and adult stem cell regulation.

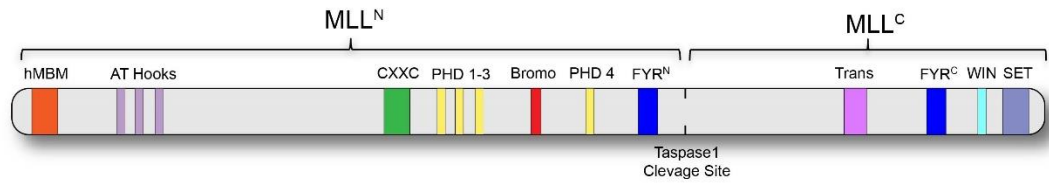
Mutations observed in TrxG complexes are associated with cancer development in various tissue types. For instance, several subunits constituting the SWI/SNF chromatin remodeling complexes are mutated in renal, lung, ovarian, and colorectal tumors. MLL complex mutations and chromosomal translocations also have been associated with tumorigenesis. *MLL* genes harbor genomic sites that are readily targeted for chromosomal translocations that create fusion genes coding novel fusion proteins. ELL elongation factor was the first fusion partner of MLL1 discovered, forming a chimeric MLL-ELL fusion protein found in human leukemia⁵³. Later studies identified several other MLL fusion partners involved in leukemogenesis, including AFF, AF9, and ENL⁵⁴. Intriguingly, analysis of MLL fusion-proteins shows that the catalytic SET domain of

MLL is lost during translocation, indicating leukemia development is possibly not dependent on H3K4me. MLL fusion-proteins can also mediate gene regulation via their interaction and recruitment with the super elongation complex (SEC), promoting aberrant gene activation⁵⁵. Mutations in the MLL3 and MLL4 were identified in lymphomas, medulloblastomas⁵⁶.

MLL1 COMPLEX POST-TRANSCRIPTIONAL PROCESSING

MLL1 is post-translationally cleaved by the enzyme Taspase1 (TASP1), generating MLL^N and MLL^C, the N-terminal and C-terminal fragments of MLL, respectively^{57,58}. The MLL^N and MLL^C subunits associate with one another through intramolecular protein-protein interactions⁵⁹. The MLL^C subunit transactivation domain interacts with several histone acetyltransferases (HATs), including p300/CBP and MOF⁶⁰⁻⁶². The SET domain found downstream of the transactivation domain mediates the H3K36m3 methyltransferase activity of MLL1 (Figure 3A)⁶³. MLL^C interacts with WRD5 via its WIN motif⁶⁴. Association with WRD5 recruits RBBP5 and ASH2L, forming a complex that facilitates H3K4 methylation (Figure 3B)⁶⁵⁻⁶⁷. Found on the MLL^N subunit are three consecutive AT hooks that preferentially bind to AT-rich DNA sequences, which may serve to regulate its nuclear localization. Following these hooks is a CXXC-ZF domain that binds to unmethylated CpG islands. MENIN directly interacts with MLL through a high affinity MENIN binding motif (hMBM) of MLLN⁶⁸. The MLL/MENIN complex promotes interaction with lens epithelium derived growth factor (LEDGF), a protein that binds to H3K36 methylation through its PWWP (pro-trp-trp-pro) motif⁶⁹.

A



B

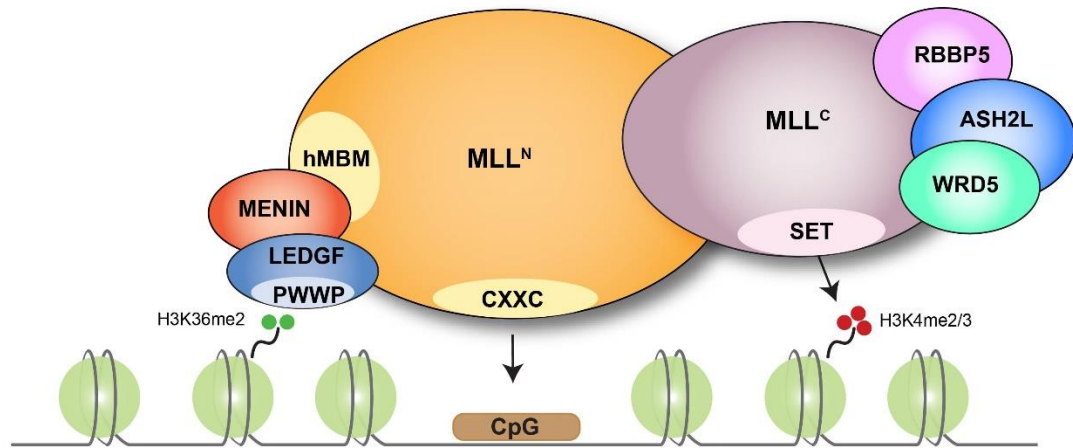


Figure 3: Schematic representation of MLL1 and its interaction partners. (A) Gene annotation of MLL1 (B) MLL1 is post-transcriptionally cleaved by taspase1 generating two MLL subunits, MLL^N and MLL^C, that interact with each other and associate with various factors. The MLL^N subunit indirectly associates with LEDGF through its interaction with MENIN promoting binding to H3K36me2 modification. The MLL^N CXXC domain binds to CpG sites. The MLL^C subunit SET domain catalyzes H3K4me2/3 modifications.

ROLE MLL REARRANGMENTS IN LEUKEMOGENESIS

During hematopoiesis, *Hox* gene expression gradually declines as hematopoietic stem cells differentiate. MLL serves as essential regulators of *Hox* genes and hematopoietic development⁷⁰. Notably, the activity of MLL is critical for the maintenance of the hematopoietic stem/progenitor cell pool, and the loss of its function leads to bone marrow failure^{71,72}. However, MLL fusion genes generated by 11q23 translocations lose the C-terminal end of the wild-type *MLL* gene, resulting in a protein lacking all the domains downstream of the CXXC motif. In exchange, MLL acquires fusion partners with strong transcriptional activation capabilities that promote aberrant *Hox* gene expression, which is essential for driving leukemic transformation (Figure 4A). A majority of observed clinical cases result from MLL-AEP (AF4/ENL/P-TEFb), which are fusion proteins containing a member of the AF4 family/ENL family/P-TEFb complex (Figure 4B)^{73,74}. MLL-AEP fusion proteins form an SEC involved in rapid transcriptional induction that drives leukemogenesis⁵⁵.

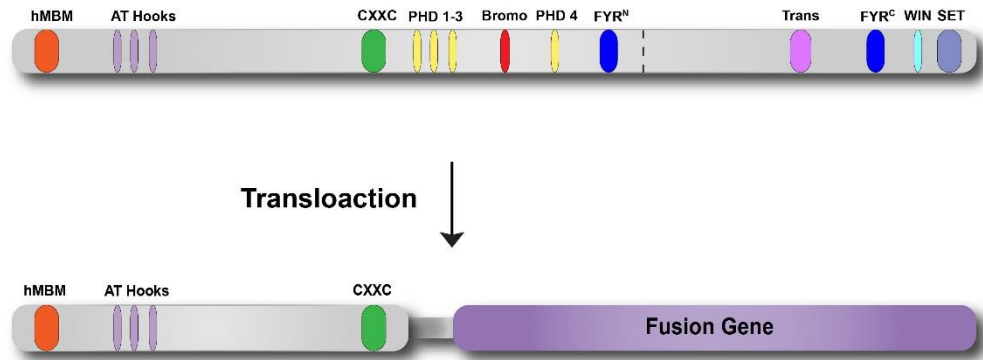
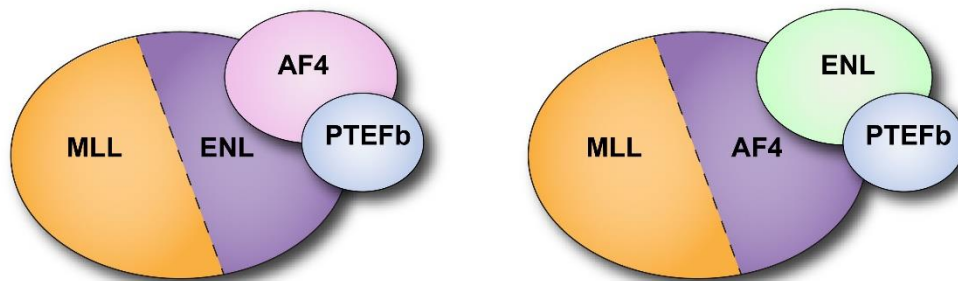
A**B**

Figure 4: MLL Fusion Proteins. (A) Chromosomal translocations generate MLL fusion genes that lose all functional domains downstream of the CXXC domain. (B) MLL-AEP proteins are fused with a member of the ENL or AF4 family and interact with P-TEFb, forming an SEC that mediates transcriptional activation of target genes.

More than 70 MLL fusion proteins are identified among patients diagnosed with MLL associated leukemia⁷⁵. Yet, despite the vast number of fusion proteins generated by 11q23 translocations, the CXXC domain and hMBM motif of wild-type MLL1 are preserved in all MLL fusion proteins currently identified. Interestingly, reports have shown that MLL-fusion protein-mediated leukemogenesis is dependent on the function of the retained CXXC and hMBM domains^{76–78}. As previously stated, the CXXC domain aids in MLL recruitment by binding to unmethylated CpG islands in DNA. Studies reported that mutation of the CXXC domain in MLL-rearrangement (MLLr) leukemia resulted in the loss of cell immortalization and inhibited cellular transformation in mice^{79,80}. The retained hMBM domain in MLL fusion proteins can interact with MENIN, enabling its association with LEDGF, a factor that reads H3K36 methylation marks through its PWWP domain⁸¹. Thus, MLL fusion proteins can also indirectly recognize H3K36me2 through their interaction with LEDGF. While addressing the function of MENIN in MLL fusion proteins, a study found that substitution of the MENIN binding domain with the PWWP domain of LEDGF was sufficient for the maintenance of *Hox* gene expression and MLLr leukemogenesis⁸². These results show that the N-terminal region of MLL retained in their fusion proteins provides essential chromatin targeting capabilities.

Recently ASH1L, another member of the TrxG family of proteins, was implicated in hematopoietic development⁸³. Like other TrxG proteins, ASH1L occupies regions of active genes while also counteracting the repressive function of the PCG family of proteins^{84,85}. The *ASH1L* gene encodes a 333KDa protein that contains four AT hooks followed by a bromodomain, a SET domain, a PHD domain, and a BAH domain. Although the functions of the majority of ASH1L protein domains are not clearly defined, biochemical studies have revealed that ASH1L deposits H3K36me2 marks on chromatin, mediated by its catalytic SET domain^{83,86}. Studies have shown

that ASH1L occupies the same genomic regions as MLL, suggesting ASH1L functions synergistically with MLL in regulating gene expression in cells⁸⁷. Another study demonstrated that loss of either ASH1L or MLL results in decreased expression of similar genes⁸⁸. These findings suggest that ASH1L may play a critical role in MLLr leukemia development and maintenance. However, the significance of H3K36me2 in the context of MLL fusion protein recruitment and MLLr leukemogenesis is yet to be defined.

REFERENCES

REFERENCES

1. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* 276(): 565--570.
2. de Ayala Alonso AG, Gutiérrez L, Fritsch C, et al. A Genetic Screen Identifies Novel Polycomb Group Genes in *Drosophila*. *Genetics*. 2007;176(4):2099–2108.
3. Struhl G, Akam M. Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of *Drosophila*. *The EMBO Journal*. 1985;4(12):3259–3264.
4. Grimaud C, Nègre N, Cavalli G. From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. *Chromosome Res*. 2006;14(4):363–375.
5. He J. Function of Polycomb repressive complexes in stem cells. *Front. Biol*. 2016;11(2):65–74.
6. McGinty RK, Henrici RC, Tan S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature*. 2014;514(7524):591–596.
7. Aranda S, Mas G, Di Croce L. Regulation of gene transcription by Polycomb proteins. *Sci. Adv*. 2015;1(11):e1500737.
8. Rose NR, King HW, Blackledge NP, et al. RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. *eLife*. 2016;5:e18591.
9. Margueron R, Li G, Sarma K, et al. Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Molecular Cell*. 2008;32(4):503–518.
10. Yu J-R, Lee C-H, Oksuz O, Stafford JM, Reinberg D. PRC2 is high maintenance. *Genes Dev*. 2019;33(15–16):903–935.
11. Cao R, Wang L, Wang H, et al. Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science*. 2002;298(5595):1039–1043.
12. Dietrich N, Lerdrup M, Landt E, et al. REST–Mediated Recruitment of Polycomb Repressor Complexes in Mammalian Cells. *PLoS Genet*. 2012;8(3):e1002494.
13. Yu M, Mazor T, Huang H, et al. Direct Recruitment of Polycomb Repressive Complex 1 to Chromatin by Core Binding Transcription Factors. *Molecular Cell*. 2012;45(3):330–343.
14. Woo CJ, Kharchenko PV, Daheron L, Park PJ, Kingston RE. A Region of the Human HOXD Cluster that Confers Polycomb-Group Responsiveness. *Cell*. 2010;140(1):99–110.

15. Zhao J, Sun BK, Erwin JA, Song J-J, Lee JT. Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. *Science*. 2008;322(5902):750–756.
16. Ku M, Koche RP, Rheinbay E, et al. Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *PLoS Genet*. 2008;4(10):e1000242.
17. Müller J, Kassis JA. Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Current Opinion in Genetics & Development*. 2006;16(5):476–484.
18. Farcas AM, Blackledge NP, Sudbery I, et al. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife*. 2012;1:e00205.
19. Bernstein BE, Mikkelsen TS, Xie X, et al. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*. 2006;125(2):315–326.
20. Min J, Zhang Y, Xu R-M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev*. 2003;17(15):1823–1828.
21. He J, Shen L, Wan M, et al. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat Cell Biol*. 2013;15(4):373–384.
22. Riising EM, Comet I, Leblanc B, et al. Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Molecular Cell*. 2014;55(3):347–360.
23. Morey L, Pascual G, Cozzuto L, et al. Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. *Cell Stem Cell*. 2012;10(1):47–62.
24. Leeb M, Wutz A. Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *Journal of Cell Biology*. 2007;178(2):219–229.
25. Walker E, Chang WY, Hunkapiller J, et al. Polycomb-like 2 Associates with PRC2 and Regulates Transcriptional Networks during Mouse Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell*. 2010;6(2):153–166.
26. Hunkapiller J, Shen Y, Diaz A, et al. Polycomb-Like 3 Promotes Polycomb Repressive Complex 2 Binding to CpG Islands and Embryonic Stem Cell Self-Renewal. *PLoS Genet*. 2012;8(3):e1002576.
27. Shen X, Liu Y, Hsu Y-J, et al. EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. *Molecular Cell*. 2008;32(4):491–502.
28. Azuara V, Perry P, Sauer S, et al. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol*. 2006;8(5):532–538.

29. Oliviero G, Brien GL, Waston A, et al. Dynamic Protein Interactions of the Polycomb Repressive Complex 2 during Differentiation of Pluripotent Cells. *Molecular & Cellular Proteomics*. 2016;15(11):3450–3460.
30. Kloet SL, Makowski MM, Baymaz HI, et al. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol*. 2016;23(7):682–690.
31. Kennison JA, Tamkun JW. Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* 1988;85(21):8136–8140.
32. Hodges C, Kirkland JG, Crabtree GR. The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. *Cold Spring Harb Perspect Med*. 2016;6(8):a026930.
33. Erdel F, Rippe K. Chromatin remodelling in mammalian cells by ISWI-type complexes - where, when and why?: ISWI chromatin remodellers in mammalian cells. *FEBS Journal*. 2011;278(19):3608–3618.
34. Murawska M, Brehm A. CHD chromatin remodelers and the transcription cycle. *Transcription*. 2011;2(6):244–253.
35. Patel A, Dharmarajan V, Vought VE, Cosgrove MS. On the Mechanism of Multiple Lysine Methylation by the Human Mixed Lineage Leukemia Protein-1 (MLL1) Core Complex. *Journal of Biological Chemistry*. 2009;284(36):24242–24256.
36. Hu D, Garruss AS, Gao X, et al. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat Struct Mol Biol*. 2013;20(9):1093–1097.
37. Yuan W, Xu M, Huang C, et al. H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation. *Journal of Biological Chemistry*. 2011;286(10):7983–7989.
38. Thomson JP, Skene PJ, Selfridge J, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*. 2010;464(7291):1082–1086.
39. Muntean AG, Tan J, Sitwala K, et al. The PAF Complex Synergizes with MLL Fusion Proteins at HOX Loci to Promote Leukemogenesis. *Cancer Cell*. 2010;17(6):609–621.
40. Wang KC, Yang YW, Liu B, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011;472(7341):120–124.
41. Denissov S, Hofemeister H, Marks H, et al. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development*. 2014;141(3):526–537.

42. Lubitz S, Glaser S, Schaft J, Stewart AF, Anastassiadis K. Increased Apoptosis and Skewed Differentiation in Mouse Embryonic Stem Cells Lacking the Histone Methyltransferase Mll2 \square D. *Molecular Biology of the Cell*. 2007;18:11.
43. Yu BD, Hess JL, Horning SE, Brown GAJ, Korsmeyer SJ. Altered Hox expression and segmental identity in Mll-mutant mice. 1995;4.
44. Dou Y, Milne TA, Ruthenburg AJ, et al. Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol*. 2006;13(8):713–719.
45. Yan J, Chen S-AA, Local A, et al. Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res*. 2018;28(2):204–220.
46. Swigut T, Wysocka J. H3K27 Demethylases, at Long Last. *Cell*. 2007;131(1):29–32.
47. Yokoyama A, Cleary ML. Menin Critically Links MLL Proteins with LEDGF on Cancer-Associated Target Genes. *Cancer Cell*. 2008;14(1):36–46.
48. Zhu L, Li Q, Wong SHK, et al. ASH1L Links Histone H3 Lysine 36 Dimethylation to MLL Leukemia. *Cancer Discov*. 2016;6(7):770–783.
49. Ang Y-S, Tsai S-Y, Lee D-F, et al. Wdr5 Mediates Self-Renewal and Reprogramming via the Embryonic Stem Cell Core Transcriptional Network. *Cell*. 2011;145(2):183–197.
50. Jiang H, Shukla A, Wang X, et al. Role for Dpy-30 in ES Cell-Fate Specification by Regulation of H3K4 Methylation within Bivalent Domains. *Cell*. 2011;144(4):513–525.
51. Gan T, Jude CD, Zaffuto K, Ernst P. Developmentally induced Mll1 loss reveals defects in postnatal haematopoiesis. *Leukemia*. 2010;24(10):1732–1741.
52. McMahon KA, Hiew SY-L, Hadjur S, et al. Mll Has a Critical Role in Fetal and Adult Hematopoietic Stem Cell Self-Renewal. *Cell Stem Cell*. 2007;1(3):338–345.
53. Andreu-Vieyra CV, Chen R, Agno JE, et al. MLL2 Is Required in Oocytes for Bulk Histone 3 Lysine 4 Trimethylation and Transcriptional Silencing. *PLoS Biol*. 2010;8(8):e1000453.
54. Poel SZ-VD, PATEL^t Y, HARDEN^t A, et al. Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proc. Natl. Acad Sci. USA*. 1991;5.
55. Ayton PM, Cleary ML. Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene*. 2001;20(40):5695–5707.
56. Smith E, Lin C, Shilatifard A. The super elongation complex (SEC) and MLL in development and disease. *Genes & Development*. 2011;25(7):661–672.

57. Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science*. 2016;352(6290):aad9780.
58. Hsieh JJ-D, Ernst P, Erdjument-Bromage H, Tempst P, Korsmeyer SJ. Proteolytic Cleavage of MLL Generates a Complex of N- and C-Terminal Fragments That Confers Protein Stability and Subnuclear Localization. *MCB*. 2003;23(1):186–194.
59. Capotosti F, Hsieh JJ-D, Herr W. Species Selectivity of Mixed-Lineage Leukemia/Trithorax and HCF Proteolytic Maturation Pathways. *Mol Cell Biol*. 2007;27(20):7063–7072.
60. Pless B, Oehm C, Knauer S, et al. The heterodimerization domains of MLL—FYRN and FYRC—are potential target structures in t(4;11) leukemia. *Leukemia*. 2011;25(4):663–670.
61. Ernst P, Wang J, Huang M, Goodman RH, Korsmeyer SJ. MLL and CREB Bind Cooperatively to the Nuclear Coactivator CREB-Binding Protein. *Mol. Cell. Biol*. 2001;21(7):2249–2258.
62. Dou Y, Milne TA, Tackett AJ, et al. Physical Association and Coordinate Function of the H3 K4 Methyltransferase MLL1 and the H4 K16 Acetyltransferase MOF. *Cell*. 2005;121(6):873–885.
63. Paggetti J, Largeot A, Aucagne R, et al. Crosstalk between leukemia-associated proteins MOZ and MLL regulates HOX gene expression in human cord blood CD34+ cells. *Oncogene*. 2010;29(36):5019–5031.
64. Milne TA, Briggs SD, Brock HW, et al. MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters. *Molecular Cell*. 2002;10(5):1107–1117.
65. Dharmarajan V, Lee J-H, Patel A, Skalnik DG, Cosgrove MS. Structural Basis for WDR5 Interaction (Win) Motif Recognition in Human SET1 Family Histone Methyltransferases. *Journal of Biological Chemistry*. 2012;287(33):27275–27289.
66. Steward MM, Lee J-S, O'Donovan A, et al. Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol*. 2006;13(9):852–854.
67. Senisterra G, Wu H, Allali-Hassani A, et al. Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5. *Biochemical Journal*. 2013;449(1):151–159.
68. Cao F, Chen Y, Cierpicki T, et al. An Ash2L/RbBP5 Heterodimer Stimulates the MLL1 Methyltransferase Activity through Coordinated Substrate Interactions with the MLL1 SET Domain. *PLoS ONE*. 2010;5(11):e14102.
69. Shi A, Murai MJ, He S, et al. Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia. *Blood*. 2012;120(23):4461–4469.

70. Botbol Y, Raghavendra NK, Rahman S, Engelman A, Lavigne M. Chromatinized templates reveal the requirement for the LEDGF/p75 PWWP domain during HIV-1 integration in vitro. *Nucleic Acids Research*. 2007;36(4):1237–1246.
71. Schraets D, Lehmann T, Dingermann T, Marschalek R. MLL-mediated transcriptional gene regulation investigated by gene expression profiling. *Oncogene*. 2003;22(23):3655–3668.
72. Jude CD, Climer L, Xu D, et al. Unique and Independent Roles for MLL in Adult Hematopoietic Stem Cells and Progenitors. *Cell Stem Cell*. 2007;1(3):324–337.
73. Ernst P, Fisher JK, Avery W, et al. Definitive Hematopoiesis Requires the Mixed-Lineage Leukemia Gene. *Developmental Cell*. 2004;6(3):437–443.
74. Yokoyama A, Lin M, Naresh A, Kitabayashi I, Cleary ML. A Higher-Order Complex Containing AF4 and ENL Family Proteins with P-TEFb Facilitates Oncogenic and Physiologic MLL-Dependent Transcription. *Cancer Cell*. 2010;17(2):198–212.
75. Yokoyama A. Transcriptional activation by MLL fusion proteins in leukemogenesis. *Experimental Hematology*. 2017;46:21–30.
76. Meyer C, Hofmann J, Burmeister T, et al. The MLL recombinoome of acute leukemias in 2013. *Leukemia*. 2013;27(11):2165–2176.
77. Yokoyama A, Somervaille TCP, Smith KS, et al. The Menin Tumor Suppressor Protein Is an Essential Oncogenic Cofactor for MLL-Associated Leukemogenesis. *Cell*. 2005;123(2):207–218.
78. Slany RK, Lavau C, Cleary ML. The Oncogenic Capacity of HRX-ENL Requires the Transcriptional Transactivation Activity of ENL and the DNA Binding Motifs of HRX. *Mol Cell Biol*. 1998;18(1):122–129.
79. Ayton PM, Chen EH, Cleary ML. Binding to Nonmethylated CpG DNA Is Essential for Target Recognition, Transactivation, and Myeloid Transformation by an MLL Oncoprotein. *MCB*. 2004;24(23):10470–10478.
80. Risner LE, Kuntimaddi A, Lokken AA, et al. Functional Specificity of CpG DNA-binding CXXC Domains in Mixed Lineage Leukemia. *Journal of Biological Chemistry*. 2013;288(41):29901–29910.
81. Cierpicki T, Risner LE, Grembecka J, et al. Structure of the MLL CXXC domain–DNA complex and its functional role in MLL-AF9 leukemia. *Nat Struct Mol Biol*. 2010;17(1):62–68.
82. Shun M-C, Botbol Y, Li X, et al. Identification and Characterization of PWWP Domain Residues Critical for LEDGF/p75 Chromatin Binding and Human Immunodeficiency Virus Type 1 Infectivity. *J Virol*. 2008;82(23):11555–11567.

83. Yokoyama A, Cleary ML. Menin Critically Links MLL Proteins with LEDGF on Cancer-Associated Target Genes. *Cancer Cell*. 2008;14(1):36–46.
84. Jones M, Chase J, Brinkmeier M, et al. Ash1l controls quiescence and self-renewal potential in hematopoietic stem cells. *J. Clin. Invest.* 2015;125(5):2007–2020.
85. Miyazaki H, Higashimoto K, Yada Y, et al. Ash1l Methylates Lys36 of Histone H3 Independently of Transcriptional Elongation to Counteract Polycomb Silencing. *PLoS Genet.* 2013;9(11):e1003897.
86. Gregory GD, Vakoc CR, Rozovskaia T, et al. Mammalian ASH1L Is a Histone Methyltransferase That Occupies the Transcribed Region of Active Genes. *MCB*. 2007;27(24):8466–8479.
87. Tanaka Y, Katagiri Z, Kawahashi K, Kioussis D, Kitajima S. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene*. 2007;397(1–2):161–168.
88. Brinkmeier ML, Geister KA, Jones M, et al. The Histone Methyltransferase Gene Absent, Small, or Homeotic Discs-1 Like Is Required for Normal Hox Gene Expression and Fertility in Mice¹. *Biology of Reproduction*. 2015;93(5):.
89. Tanaka Y, Kawahashi K, Katagiri Z-I, et al. Dual Function of Histone H3 Lysine 36 Methyltransferase ASH1 in Regulation of Hox Gene Expression. *PLoS ONE*. 2011;6(11):e28171.

CHAPTER 2:

CELL SIGNALING COORDINATES GLOBAL PRC2 RECRUITMENT AND DEVELOPMENTAL GENE EXPRESSION IN MURINE EMBRYONIC STEM CELLS

Published: iScience. 2020 23(11):101646.

Title: Cell Signaling Coordinates Global PRC2 Recruitment and Developmental Gene Expression in Murine Embryonic Stem Cells

Authors: Mohammad B Aljazi, Yuen Gao, Yan Wu, George I Mias, Jin He

ABSTRACT

The recruitment of Polycomb repressive complex 2 (PRC2) to gene promoters is critical for its function in repressing gene expression in murine embryonic stem cells (mESCs). However, previous studies have demonstrated that although the expression of early lineage-specific genes is largely repressed, the genome-wide PRC2 occupancy is unexpectedly reduced in naive mESCs. In this study, we provide evidence that fibroblast growth factor/extracellular signal-regulated kinase signaling determines the global PRC2 occupancy through regulating the expression of PRC2-recruiting factor JARID2 in naive mESCs. At the transcriptional level, the de-repression of bivalent genes is predominantly determined by the presence of cell signaling-associated transcription factors but not the status of PRC2 occupancy at gene promoters. Hence, this study not only reveals a key molecular mechanism by which cell signaling regulates the PRC2 occupancy in mESCs but also elucidates the functional roles of transcription factors and Polycomb-mediated epigenetic mechanisms in transcriptional regulation.

INTRODUCTION

Polycomb repressive complexes (PRCs) are critical for maintaining mESCs in an undifferentiated state by repressing lineage-specific gene expression¹⁻³. Previous studies have shown that both PRC1 and PRC2 are recruited to gene promoters and mediate monoubiquitylation of histone H2A lysine 119 (H2AK119ub1) and trimethylation of histone H3 lysine 27 (H3K27me3) respectively⁴⁻⁶, which are essential for their function in repressing gene expression^{4,7,8}.

Although the genome-wide binding sites of PRCs in mESCs have been well documented, the molecular mechanisms for recruiting PRCs to specific genomic loci in mammalian cells have not been fully elucidated⁹. A variety of factors, such as JARID2 and MTF2, have been reported to mediate the PRC2 recruitment¹⁰⁻¹⁵. On the other hand, histone lysine demethylase 2b (KDM2B) recruits a non-canonical PRC1.1 (ncPRC1.1) variant to unmethylated CpG islands (CGIs) in mammalian cells through its CXXC-zinc finger (CXXC-ZF) domain¹⁶⁻¹⁸. Recently, KDM2B was also reported to recruit PRC2 through the interaction of JARID2/PRC2 with KDM2B/ncPRC1.1-mediated histone H2AK119ub1 modification¹⁹.

The repression of differentiation gene expression is essential for mESCs to maintain their pluripotency²⁰. However, differentiation signals such as the extracellular signal-regulated kinase (ERK) signaling triggered by fibroblast growth factors (FGFs) in serum induce stochastic expression of early lineage-specific genes. After being cultured in 2i serum-free medium in which the FGF/ERK signaling is inhibited and the Wnt/ β -catenin signaling is activated, mESCs turn into a naive state with reduced basal expression of early lineage-specific genes^{21,22}. Along with the gene expression changes, naive mESCs display two major genome-wide epigenetic changes²³: (i) a global reduction of DNA methylation mainly due to decreased UHRF1 expression and loss of

DNA methylation maintenance^{24,25}, and (ii) both genome-wide PRC2 occupancy and its mediated histone H3K27me3 modification are drastically reduced²⁶⁻²⁸. Although the underlying molecular mechanisms for the global reduction of PRC2 occupancy are unclear, several recent studies suggest it might be due to the relocation of PRC2 to new DNA demethylated non-CGI regions²⁹⁻³¹, or the reduced chromatin accessibility to PRC2³⁵.

To examine the molecular mechanisms leading to the reduced PRC2 occupancy and H3K27me3 modification in naive mESCs, we performed RNA-sequencing (RNA-seq) analyses to compare the gene expression in wild-type mESCs cultured in serum versus 2i medium. The result showed that the expression of PRC2 recruiting factor JARID2 was significantly reduced in naive mESCs. Reactivation of FGF/ERK signaling increased the expression of *Jarid2*, suggesting the FGF/ERK signaling positively regulated its expression. Similarly, genetic deletion of ERK signaling molecules ERK1 and ERK2 reduced the *Jarid2* expression, which could be rescued by ectopic expression of wild-type ERK2. ChIP-seq analyses showed that the global occupancy of JARID2/EZH2 and histone H3K27me3 modification were largely reduced at CGIs in both naive and *Erk1/Erk2* double knockout (*Erk1/Erk2*-dKO) mESCs, which correlated with the global reduction of JARID2 occupancy. Importantly, ectopic expression of *Jarid2* fully restored the global EZH2 occupancy and the H3K27me3 modification at CGIs in both naive and *Erk1/Erk2*-dKO mESCs. At the transcriptional level, although the PRC2 occupancy and H3K27me3 modification were reduced at bivalent promoters, the FGF/ERK-regulated lineage-specific genes remained silenced while the Wnt signaling-regulated genes were de-repressed in naive mESCs, suggesting the presence of transcription factors, but not the status of PRC2 occupancy, played a predominant role in determining transcriptional activation. Thus, this study not only revealed a main molecular mechanism by which the FGF/ERK signaling regulated the global PRC2

occupancy in mESCs, but also elucidated a fundamental question regarding the functional roles of transcription factors and Polycomb-mediated epigenetic mechanisms in transcriptional regulation.

MATERIALS METHODS

Mouse embryonic stem cell culture

The wild-type E14 mESC line and wild-type E14 mESCs with 3xFlag knock-in to the endogenous Kdm2b gene (He, Shen et al. 2013) were maintained on the 0.1% gelatin coated plates in "2i" medium that contained 50% Neurobasal medium (Life Technologies) and 50% DMEM/F12 medium (Life Technologies) supplemented with 1x N2-supplement (Life Technologies), 1x B27 supplement (Life Technologies), 7.5% bovine serum albumin (Life Technologies), 1x GlutaMAX (Life Technologies), 1x beta-mercaptoethanol (Life Technologies), 1mM PD03259010, 3mM CHIR99021, 1000 units/ml leukemia inhibitory factor (ESGRO, EMD Millipore), and 100U/ml penicillin/streptomycin (Life Technologies). The serum-containing mESC culture condition included DMEM medium (Life Technologies) supplemented with 100U/ml penicillin/streptomycin (Life Technologies), 15% fetal bovine serum (Sigma), 1x nonessential amino acid, 1x sodium pyruvate (Life Technologies), 1x GlutaMAX (Life Technologies), 1x beta-mercaptoethanol (Life Technologies) and 1000 units/ml leukemia inhibitory factor (ESGRO, EMD Millipore).

Western blot analysis

Total proteins were extracted by RIPA buffer and separated by electrophoresis by 8-10% PAGE gel. The protein was transferred to the nitrocellulose membrane and blotted with primary antibodies. The antibodies used for Western Blot and IP-Western Blot analyses included: rabbit

anti-FLAG (1:2000, Cell Signaling Technology, Cat#14793); rabbit anti-EZH2 (1:1000 Cell Signaling Technology, Cat# 5246); rabbit anti-RING1B (1:1000, Cell Signaling Technology, Cat# 5694); rabbit anti-SUZ12 (1:1000 Cell Signaling Technology, Cat# 3737); rabbit anti-JARID2 (1:1000, Cell Signaling Technology ,Cat# 13594); rabbit anti-p42/44 MAPKs (1:1000 Cell Signaling Technology, Cat# 4695); rabbit anti-phospho-p42/44 MAPKs ((1:1000 Cell Signaling Technology, Cat# 9101), rabbit anti-MTF2 (1:1000, Proteintech, Cat# 16208-1-AP), rabbit anti-EZH1 (1:1000, Cell Signaling Technology, Cat# 42088); rabbit anti-EED (1:1000, Cell Signaling Technology, Cat# 85322); rabbit anti-AEBP2 (1:1000, Cell Signaling Technology, Cat# 14129); rabbit anti-RING1A(1:1000, Cell Signaling Technology, Cat# 13069); rabbit anti-RYBP (1:1000, Abcam, Cat# ab5976); rabbit anti-PHF19 (1:1000, Cell Signaling Technology, Cat# 77271); rabbit anti-PHF1 (1:1000, Abcam, Cat# ab184951); rabbit anti-EPOP (1:1000, Invitrogen, Cat# 703052); mouse anti-PCGF1 (1:1000, Santa Cruz Biotechnology, Cat# 515371); rabbit anti-SKP1 (1:1000, Cell Signaling Technology, Cat# 2156); rabbit anti-BCOR (1:1000, Abgent, Cat# AP7359c); rabbit anti-TUBULIN (1:2000, Proteintech, Cat# 11224-1-AP), and IRDye 680 donkey anti-rabbit or anti-mouse second antibody (1: 10000, Li-Cor). The images were developed by Odyssey Li-Cor Imager (Li-Cor).

RT-qPCR assays

RNA was extracted and purified from cells using QI shredder (Qiagen) and RNeasy (Qiagen) spin columns. Total RNA (1 µg) was subjected to reverse transcription using Iscript reverse transcription supermix (Bio-Rad). cDNA levels were assayed by real-time PCR using iTaq universal SYBR green supermix (Bio-Rad) and detected by CFX386 Touch Real-Time PCR detection system (Bio-Rad). Primer sequences for qPCR are listed in Table 2.

Table 2: Sequences of all primers used in PRC2 study

Name	Sequence (5'-3')	Assays
GAPDH-F	GCAGTGGCAAAGTGGAGATT	RT-qPCR
GAPDH-R	GAATTTGCCGTGAGTGGAGT	RT-qPCR
Ezh2-F	CCTGTTCCCACTGAGGATGT	RT-qPCR
Ezh2-R	GAGCCGTCCTTTTTCAGTTG	RT-qPCR
Eed-F	CTGTGGGAAGCAACAGAGTAA	RT-qPCR
Eed-F	TAGGTCCATGCACAAGTGTAAG	RT-qPCR
Jarid2-F	CTTTCTCTGCCTTCGAGGTT	RT-qPCR
Jarid2-R	CCGTCTCATCATCTTCCTCTTC	RT-qPCR
Suz12-F	GAGCAACATGGGAGACAAT	RT-qPCR
Suz12-R	TCCTGTCCATCGAAGAGTAA	RT-qPCR
Mtf2-F	AGAAAGGCATCCAAACCTACA	RT-qPCR
Mtf2-R	AGGAGGACGACCTACAGATT	RT-qPCR
Kdm2b-F	AGCAGACAGAAGCCACCAAT	RT-qPCR
Kdm2b-R	AGGTGCCTCCAAAGTCAATG	RT-qPCR
Ring1b-F	TTGCGCGGATTGTATTATCA	RT-qPCR
Ring1b-R	GTGCATCAAAGTTCGGGTCT	RT-qPCR
Ezh1-F	GTCTTCCACGGCACCTATTT	RT-qPCR
Ezh1-R	TGTTGGCAGCTTTAGGATAAGT	RT-qPCR
Rbbp4-F	CTGCTCCACGAGTCTCTATTTG	RT-qPCR
Rbbp4-R	TGGCTTGGCTTGGAAGTATT	RT-qPCR
Rbbp7-F	GGTGGTTGCTCGAGTTCATATT	RT-qPCR
Rbbp7-R	ACCAAAGCCACCGAATTCTC	RT-qPCR
EPOP-F	CTTCCTACATCGAGCACCTTC	RT-qPCR
EPOP-R	AGGTCTCCGTTCTCTTCCA	RT-qPCR
Phf1-F	ATGAATTTGAGTGCTGTGTGTG	RT-qPCR
Phf1-R	TGAGGTGGTAGAGGACAAGAT	RT-qPCR
Phf19-F	CCTGTGTGGCAAGGAAATTAAG	RT-qPCR
Phf19-R	CCTTTGTCACTTGGCATCAAC	RT-qPCR
Aebp2-F	TGACACACAGTGGAGACAAG	RT-qPCR
Aebp2-R	GACTGAAGTGTGTGGGTACAT	RT-qPCR
Pcgf1-F	GCCGTTGCTCAATCTCAAAC	RT-qPCR
Pcgf1-R	TTCCCGAATCCGTTTCTCTTC	RT-qPCR
Ring1a-F	TGAGCTCCAGTATCGAGGAA	RT-qPCR
Ring1a-R	CTCATTTGTGGCGGTCTGAT	RT-qPCR
BcoR-F	AGACAACGGTTCAAGACAGAAA	RT-qPCR
BcoR-R	CCGCATCCACCACTTTAGAA	RT-qPCR
Rybp-F	CAGGAAACCTCGCATCAATTC	RT-qPCR
Rybp-R	GACCTTCTCCTTCTTCTCCTTC	RT-qPCR
Skp1-F	CAAGACCATGCTGGAAGATTTG	RT-qPCR
Skp1-R	GCACCACTGAATGACCTTCT	RT-qPCR

Cell cycle analysis

The cells were fixed with ice-cold 70% ethanol, stained with 7-AAD at 25 µg/ml (Sigma), and analyzed by BS LSR II instrument at the MSU flow cytometry core facility. The data was (van Mierlo, Dirks et al. 2019) analyzed by FlowJo 7.6.1 software (FlowJo, LLC).

Lentiviral vector generation and infection

The lentiviral system was obtained from the National Institutes of Health AIDS Research and Reference Reagent Program. To generate mouse *Erk2* and *Jarid2* expression vectors, the complementary DNAs were PCR amplified, fused with P2A and puromycin resistant cassette and cloned into the SpeI/EcoRI sites under the EF1α promoter. To generate lentiviral viruses, the transducing vectors pTY, pHP and pHEF1α-VSVG were co-transfected into HEK293T cells. The supernatant was collected at 24, 36 and 48 hours after transfection, filtered through a 0.45 µm membrane and concentrated using a spin column (EMD Millipore). The mESCs were infected with lentiviral vectors at MOI 1.0. 48 hours after infection, the infected cells are selected by adding puromycin into the medium (1 µg/ml).

Crispr-mediate Erk1/ERk2 gene knock-out in mESCs

The mouse *Erk1* gene gRNA (GGTAGAGGAAGTAGCAGATG) and mouse *Erk2* gene gRNA (GGTTCTTTGACAGTAGGTC and CTTAGGGTTCTTTGACAGT) were cloned into pX330 vector obtained from Addgene. The target vector and pEF1a-pac vector were co-transfected (5:1 ratio) to E14 mESCs using Xfect according to the manufacture's instruction (TaKaRa, Inc). 48 hours after transfection, Puromycin (1 µg/ml) was added to the medium to select transfected cells. The individual clones were manually picked and expanded. The correct knockout clones

were selected based on the Sanger sequencing on the targeting sites of genomic DNA, cDNA, and Western blot analysis.

ChIP-Seq sample preparation

For EZH2, JARID2, RING1B, KDM2B, H3K27me3 and H3K4me3 ChIP, Kdm2b 3xFlag knock-in wild-type E14 cells (He, Shen et al. 2013) were fixed with 2mM Ethylene glycol bis[succinimidylsuccinate] (Thermo Scientific) for 1 hour, followed by 10 min in 1% formaldehyde and 5 min 0.125 M glycine to sequence the reaction. Cells were lysed in 1% SDS, 10 mM EDTA, 50 mM Tris-HCl (pH 8.0) and the DNA was fragmented to approximately 200-400 bp by sonification (Branson Sonifier 450). Immunoprecipitation was performed with 3 μ g rabbit polyclonal anti-FLAG (Sigma, Cat# F7425), rabbit anti-EZH2 antibody (1:100, Cell Signaling Technology, Cat# 5246), rabbit anti-JARID2 antibody (1:100, Cell Signaling Technology, Cat# 13594), rabbit anti-RING1B antibody (1:100, Cell Signaling Technology, Cat# 5694), rabbit anti- β -CATENIN (1:100, Cell Signaling Technology, Cat# 8480), 2ug rabbit anti-H3K27me3 antibody (Diagenode, Cat# C15410195), 2ug rabbit anti-H3K4me4 antibody (Diagenode, Cat# C15410003) overnight at 4°C. Antibody bound DNA-proteins were isolated by protein G plus/protein A agarose beads (EMD Millipore), washed, eluted and reverse cross-linked DNA was extracted by phenol/chloroform and precipitated.

ChIP DNA preparation for HiSeq4000 sequencing

ChIP DNA library was constructed for HiSeq4000 (Illumina) sequencing using NEBNext UltraII DNA library Prep Kit for Illumina (New England BioLabs, Inc) according to the manufacturer's instructions. Adapter-ligated DNA was amplified by PCR for 12-14 cycles and followed by size selection using agarose gel electrophoresis. The DNA was purified using QIAquick gel extraction kit (Qiagen) and quantified both with an Agilent Bioanalyzer and Invitrogen Qubit. The DNA was diluted to a working concentration of 20nM prior to sequencing. Sequencing on an Illumina HiSeq4000 instrument was carried out by the Genomics Core Facility at Michigan State University.

ChIP-Seq data analysis

For the ChIP-Seq data analysis, all sequencing reads were mapped to NCBI build 37 (mm9) of the mouse genome using Bowtie2 (Langmead and Salzberg 2012). Mapped reads were analyzed using the MACS program and bound-regions (peaks) were determined using sequencing reads from input DNA as negative controls (Zhang, Liu et al. 2008). When multiple reads mapped to the same genomic position, a maximum of two reads were retained. The statistical cutoff used for peak calling was $P\text{-value} < 10^{-8}$ and >5 -fold enrichment over the control. The mapped sequencing reads were normalized as Counts Per Million Reads (CPM). The normalized reads were binned into 50-bp windows along the genome using the bamCoverage of deepTools program and visualized in the IGV genome browser (Robinson, Thorvaldsdottir et al. 2011, Ramirez, Dundar et al. 2014). The datasets of CpG islands and Refseq genes of mm9 mouse reference genome were retrieved from the UCSC table browser. The heatmap and plot of ChIP-seq reads in the 10kb CGI-flanking regions or Refseq genes were generated using plotHeatmap and plotProfile in the deepTools

program. The subset of bivalent promoters was determined by the 2kb promoter regions upstream of transcriptional start sites that contain both H3K27me3 and H3K4me3 using the bedtools program (Quinlan and Hall 2010).

RNA-seq sample preparation for HiSeq4000 sequencing

RNA was extracted and purified from cells using QI shredder (Qiagen) and RNeasy (Qiagen) spin columns. Total RNA (1 µg) was used to generate RNA-seq library using NEBNext Ultra Directional RNA library Prep Kit for Illumina (New England BioLabs, Inc) according to the manufacturer's instructions. Adapter-ligated cDNA was amplified by PCR and followed by size selection using agarose gel electrophoresis. The DNA was purified using Qiaquick gel extraction kit (Qiagen) and quantified both with an Agilent Bioanalyzer and Invitrogen Qubit. The libraries were diluted to a working concentration of 10nM prior to sequencing. Sequencing on an Illumina HiSeq4000 instrument was carried out by the Genomics Core Facility at Michigan State University.

RNA-Seq data analysis

RNA-Seq data analysis was performed essentially as described previously. All sequencing reads were mapped mm9 of the mouse genome using Tophat2(Kim, Pertea et al. 2013). The mapped reads were normalized to reads as Reads Per Kilobase of transcript per Million mapped reads (RPKM). The differential gene expression was calculated by Cuffdiff program and the statistic cutoff for identification of differential gene expression is $q < 0.05$ and 1.5-fold RPKM change between samples. The normalized mapped reads (RPKM) of each RNA-seq experiments were binned into 50bp windows along the genome using the bamCoverage of deepTools program

and visualized in the IGV genome browser. The heatmap and plot of gene expression were generated using plotHeatmap and plotProile in the deepTools program. The differential expressed gene lists were input into the DAVID Functional Annotation Bioinformatics Microarray Analysis for KEGG pathway and gene ontology enrichment analyses (<https://david.ncifcrf.gov/>).

RESULTS

***Jarid2* expression is significantly reduced in naive mESCs**

To examine whether the reduced global PRC2 occupancy in naive mESCs is caused by reduced expression of PRC2 components, we performed RNA-seq analyses to examine the differential gene expression between wild-type E14 mESCs cultured in serum-containing medium (ESC-S) and 2i medium (ESC-2i). The results identified 1905 upregulated genes and 2371 downregulated genes in ESC-2i respectively (cutoff: 1.5 fold expression change and $q < 0.05$) (Figure 5A, 5B). Consistent with previous reports, the expression of *Nanog*, *Klf4*, and *Prdm14* was found to be upregulated in ESC-2i, validating the proper 2i culture condition and the ground state of mESCs in this study (Figure 5C)³²⁻³⁴. Among all PRC2 core and PRC2.1/PRC2.2-specific genes, *Jarid2* expression was found to be significantly reduced in mESC-2i compared to that in mESC-S (cutoff: 1.5 fold expression change and $q < 0.05$) (Figure 6A), which was further confirmed by quantitative reverse transcription PCR (qRT-PCR) and western blot (WB) analyses at both mRNA and protein levels (Figure 6B, 6C, 5D). *Eed*, to a less extent, also showed significant reduced expression in mESC-2i (Figure 6A). To corroborate our findings, we re-analyzed the expression of *Jarid2* and *Eed* in naive mESCs from a published RNA-seq dataset²⁶. Consistent with our findings, *Jarid2* expression was largely reduced in both E14 and TNGA mESC lines cultured in 2i medium while the change of *Eed* expression varied in these two mESC lines (Figure 5E, 5F).

Therefore, the observed reduced *Eed* expression in the E14 mESC line was likely to be a cell line-specific effect. In contrast to reduced *Jarid2* expression, *Suz12* expression was found to be increased in mESC-2i (cutoff: 1.5 fold expression change and $q < 0.05$). Other PRC2 genes including catalytic subunits *Ezh1* and *Ezh2* did not show statistically significant difference in expression between mESC-S and mESC-2i (Figure 6A-C).

In addition to JARID2 and MTF2-mediated PRC2 recruitment, ncPRC1.1 was reported to recruit PRC2¹⁹. However, RNA-seq analysis did not reveal any significant expression difference of ncPRC1.1 genes in naive mESCs, which was further confirmed by qRT-PCR and WB analysis (Figure 6B, 6D, 6E). Based on these results, we concluded that the expression of PRC2 recruiting factor JARID2 was significantly reduced in naive mESCs.

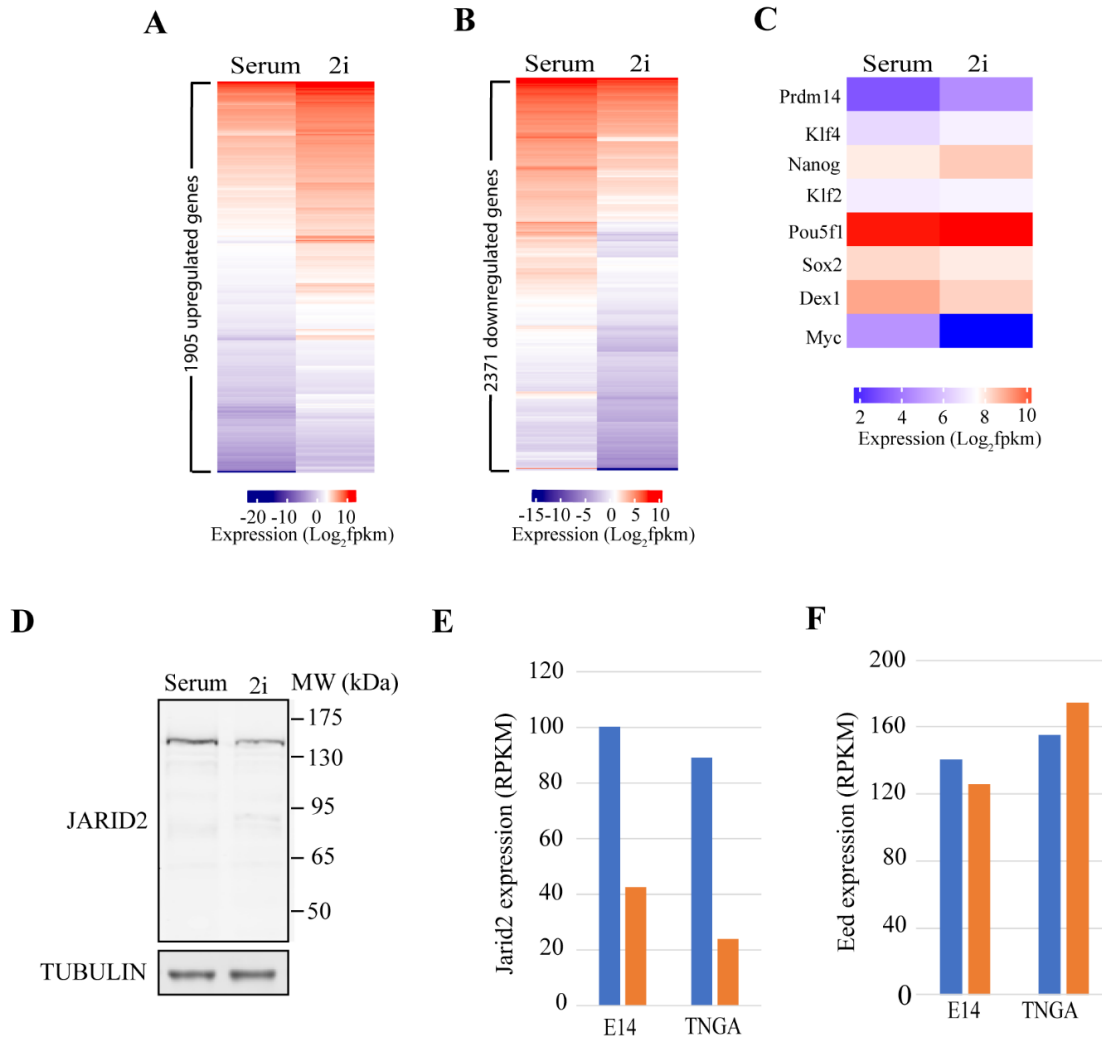


Figure 5: *Jarid2* expression is significantly reduced in naive mESCs. (A) Heatmap showing 1905 genes upregulated in mESC-2i. (B) Heatmap showing 2371 genes downregulated in mESC-2i. (C) Heatmap showing the expression of pluripotent genes in mESC-S and mESC-2i. (D) Western blot analysis showing the protein level of JARID2 in mESC-S (serum) and mESC-2i (2i). (E) Re-analysis on a published dataset showing the expression of *Jarid2* in E14 and TNGA mESC lines²⁶. (F) Re-analysis on a published dataset showing the expression of *Eed* in E14 and TNGA mESC lines²⁶.

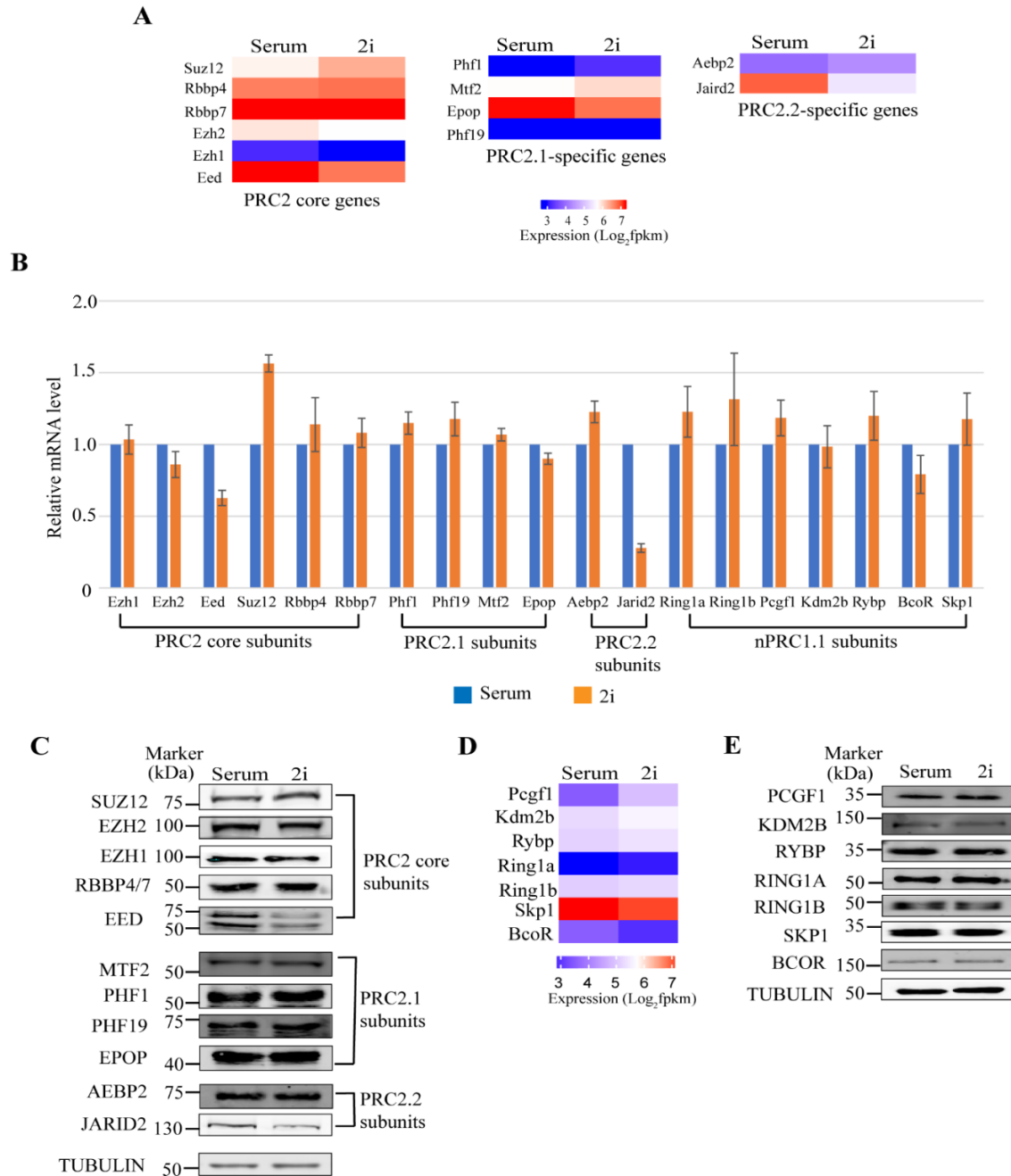


Figure 6: *Jarid2* expression is significantly reduced in naive mESCs. (A) Heatmap showing the expression of PRC2 core and PRC2.1/PRC2.2-specific genes analyzed by RNA-seq. (B) qRT-PCR analysis showing the expression levels of PRC2 core, PRC2.1/PRC2.2-specific genes, and ncPRC1.1 genes in mESC-S and mESC-2i. The results were normalized against levels of *Gapdh* and the expression level in mESC-S was arbitrarily set to 1. The error bars represent the standard deviation (n=3). (C) Western blot analysis showing the protein levels of PRC2 core and PRC2.1/PRC2.2-specific components. (D) Heatmap showing the expression of PRC2 core and PRC2.1/PRC2.2-specific genes analyzed by RNA-seq. (E) Western blot analysis showing the protein levels of ncPRC1.1 components.

FGF/ERK signaling positively regulates *Jarid2* expression in mESCs

The FGF/ERK signaling is blocked by removal of serum and supplement of MEK inhibitor PD0325901 in 2i medium. To examine whether the reduced expression of *Jarid2* was caused by the deficiency of FGF/ERK signaling in 2i medium, we reactivated the FGF/ERK signaling by adding serum and removing the MEK inhibitor PD0325901 from 2i medium. RNA-seq analyses were performed to examine the gene expression at different time points after the culture medium switched (Figure 7A). The reactivation of FGF/ERK signaling was monitored by WB analyses on the phosphorylated MAPK p42/44. The results showed that the phosphorylated MAPK p42/44 were not detected in ESC-2i. However, the phosphorylated proteins quickly increased and reached to a constant level at 6 hours after the medium switched, confirming the FGF/ERK signaling was reactivated (Figure 7B). Kinetic RNA-seq analyses identified 309 genes upregulated in response to the activation of FGF/ERK signaling (Figure 8A). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis revealed that the upregulated genes were enriched in MAPK signaling pathway, signaling involved in regulating stem cell pluripotency, signaling pathway in cancer, and HIF signaling pathway (cutoff: Benjamini adjusted $p < 0.05$), further confirming that the FGF/ERK signaling pathway was properly reactivated in mESCs after the culture condition changed (Figure 8B, Table 3). Importantly, RNA-seq results showed that the *Jarid2* expression was increased in response to the reactivation of FGF/ERK signaling (Figure 7C). The increased *Jarid2* expression was further confirmed by qRT-PCR and WB analyses at both mRNA and protein levels (Figure 7D, 8C, 8D). Other PRC2.1 and PRC2.2 subunits did not show significant difference in expression at both mRNA and protein levels in response to the reactivation of FGF/ERK signaling (Figure 7C, 7D).

To examine whether the increased *Jarid2* expression was caused by cell cycle changes, we performed the cell cycle analysis during the reactivation of FGF/ERK signaling. The results showed that there were no significant changes in cell cycle within 12 hours after the FGF/ERK signaling reactivation (Figure 8E). In contrast, *Jarid2* expression increased quickly and reached to a stable level in 6 hours after the ERK signaling reactivation (Figure 7C, 7D, 8C, 8D). The results suggested the changed *Jarid2* expression in mESC-2i was less likely caused by cell cycle difference. Based on these results, we concluded that the FGF/ERK signaling positively regulated the expression of *Jarid2* and the inhibition of this signaling pathway in 2i medium reduced the *Jarid2* expression at both mRNA and protein levels in naive mESCs.

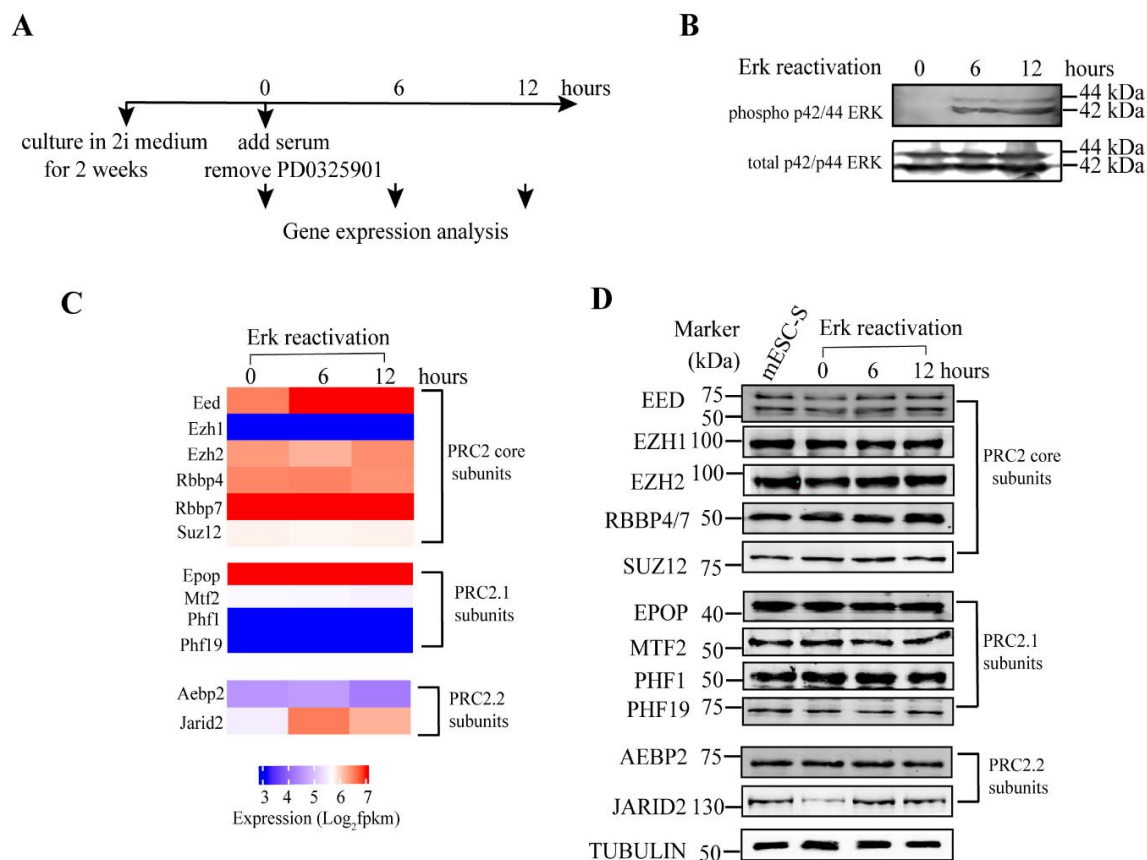


Figure 7: FGF/ERK signaling positively regulates *Jarid2* expression in mESCs. (A) Schematic chart showing the experimental design for the analysis of gene expression in response to the reactivation of FGF/ERK signaling. (B) Western blot analysis showing phosphorylated p42/44 MAPKs were increased in response to the reactivation of FGF/ERK signaling. (C) Heatmap showing the expression of PRC2 core and PRC2.1/PRC2.2-specific genes after the reactivation of FGF/ERK signaling. (D) Western blot analysis showing the protein levels of PRC2 core and PRC2.1/PRC2.2-specific components in mESCs after the reactivation of FGF/ERK signaling.

Table 3: Result of KEGG pathway analysis on the genes upregulated in response to the reactivation of FGF/ERK signaling.

Category	Term	Count	P-Value	FDR
KEGG_PATHWAY	mmu04550:Signaling pathways regulating pluripotency of stem cells	12	1.16E-05	0.0143205
KEGG_PATHWAY	mmu05200:Pathways in cancer	20	1.43E-05	0.0177318
KEGG_PATHWAY	mmu04010:MAPK signaling pathway	15	4.36E-05	0.0539604
KEGG_PATHWAY	mmu05205:Proteoglycans in cancer	13	9.09E-05	0.1124996
KEGG_PATHWAY	mmu04066:HIF-1 signaling pathway	9	2.13E-04	0.2632131
KEGG_PATHWAY	mmu05166:HTLV-I infection	13	0.0014808	1.818569
KEGG_PATHWAY	mmu05217:Basal cell carcinoma	6	0.0016304	2.0005531
KEGG_PATHWAY	mmu05222:Small cell lung cancer	7	0.002177	2.6630183
KEGG_PATHWAY	mmu04390:Hippo signaling pathway	9	0.002812	3.4274873
KEGG_PATHWAY	mmu04380:Osteoclast differentiation	8	0.0039576	4.7925018
KEGG_PATHWAY	mmu05211:Renal cell carcinoma	6	0.0045007	5.4334555
KEGG_PATHWAY	mmu05220:Chronic myeloid leukemia	6	0.0057434	6.8851429
KEGG_PATHWAY	mmu05161:Hepatitis B	8	0.0087528	10.316134
KEGG_PATHWAY	mmu04510:Focal adhesion	9	0.0176625	19.804536
KEGG_PATHWAY	mmu05210:Colorectal cancer	5	0.0189346	21.08132
KEGG_PATHWAY	mmu05214:Glioma	5	0.0199334	22.070566
KEGG_PATHWAY	mmu04068:FoxO signaling pathway	7	0.0202172	22.349647
KEGG_PATHWAY	mmu04115:p53 signaling pathway	5	0.022027	24.107457
KEGG_PATHWAY	mmu04310:Wnt signaling pathway	7	0.0252419	27.140135
KEGG_PATHWAY	mmu05219:Bladder cancer	4	0.0273704	29.086217
KEGG_PATHWAY	mmu04917:Prolactin signaling pathway	5	0.0290931	30.62614
KEGG_PATHWAY	mmu04350:TGF-beta signaling pathway	5	0.0468547	44.80656
KEGG_PATHWAY	mmu04012:ErbB signaling pathway	5	0.0502904	47.220613
KEGG_PATHWAY	mmu04151:PI3K-Akt signaling pathway	11	0.0540293	49.737143
KEGG_PATHWAY	mmu04015:Rap1 signaling pathway	8	0.0553402	50.592996
KEGG_PATHWAY	mmu05169:Epstein-Barr virus infection	6	0.0665377	57.37628
KEGG_PATHWAY	mmu05160:Hepatitis C	6	0.0665377	57.37628
KEGG_PATHWAY	mmu04370:VEGF signaling pathway	4	0.0708032	59.726732
KEGG_PATHWAY	mmu04916:Melanogenesis	5	0.0737176	61.26352
KEGG_PATHWAY	mmu05230:Central carbon metabolism in cancer	4	0.0824117	65.533506
KEGG_PATHWAY	mmu05142:Chagas disease (American trypanosomiasis)	5	0.0825707	65.607389
KEGG_PATHWAY	mmu05212:Pancreatic cancer	4	0.0854349	66.913804
KEGG_PATHWAY	mmu05145:Toxoplasmosis	5	0.087186	67.689889
KEGG_PATHWAY	mmu05412:Arrhythmogenic right ventricular cardiomyopathy (ARVC)	4	0.0916211	69.581254
KEGG_PATHWAY	mmu04668:TNF signaling pathway	5	0.0967842	71.654623
KEGG_PATHWAY	mmu04931:Insulin resistance	5	0.0992588	72.601565

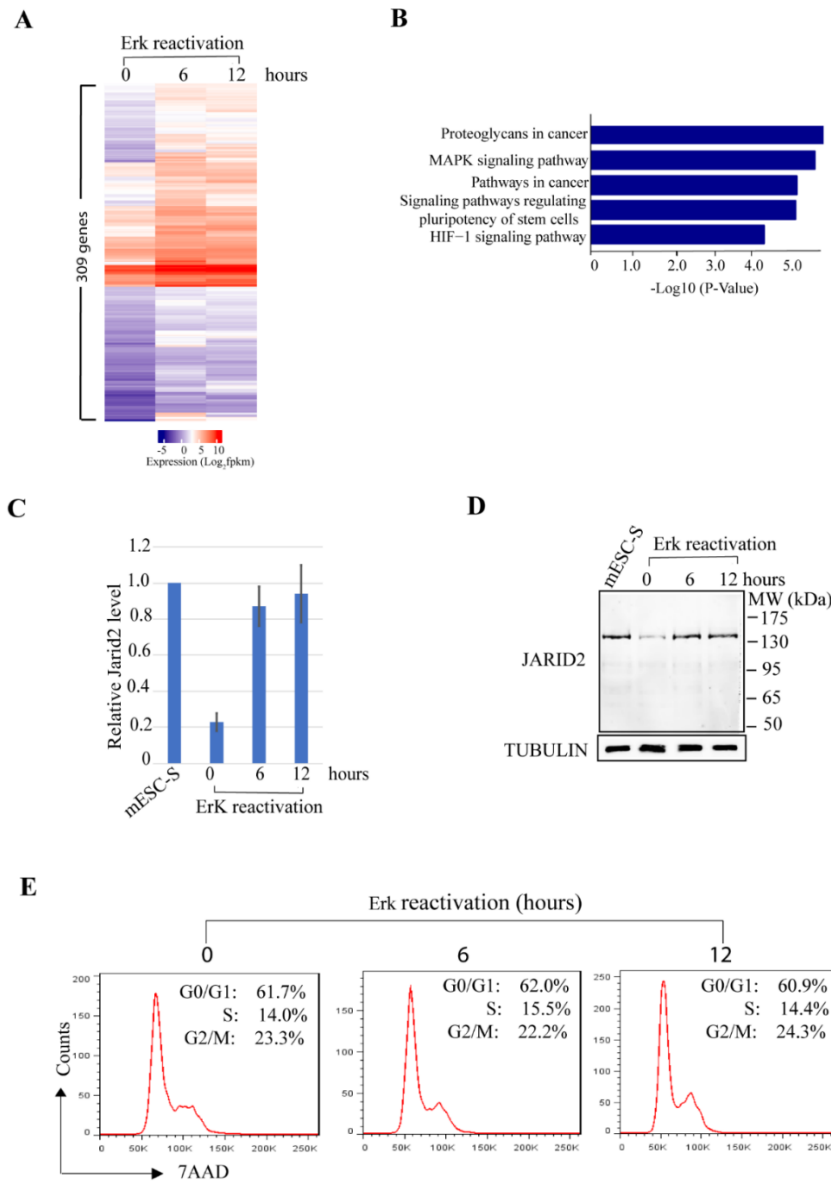


Figure 8: FGF/ERK signaling positively regulates the *Jarid2* expression. (A) Heatmap showing 309 genes were upregulated in response to the reactivation of FGF/ERK signaling. (B) KEGG signal pathway analysis showing the five top signaling pathways that were activated in response to the reactivation of FGF/ERK signaling (cutoff: Benjamini adjusted $p < 0.05$). (C) qRT-PCR analysis showing the *Jarid2* expression was upregulated at the mRNA level in response to the reactivation of FGF/ERK signaling. The results were normalized against levels of *Gapdh* and the expression level in mESC-S is arbitrarily set to 1. The error bars represent the standard deviation ($n=3$). (D) Western blot analysis showing JARID2 expression was upregulated at the protein level in response to the reactivation of FGF/ERK signaling in mESC-S and different time points after reactivation of FGF/ERK signaling. (E) Flow cytometry analysis on the cell cycle at different time points after the reactivation of FGF/ERK signaling.

Knockout of *Erk1/Erk2* reduces *Jarid2* expression in mESCs

A previous study reported the genetic depletion of MAPK signaling molecules ERK1 and ERK2 in mESCs led to reduced PRC2 occupancy and H3K27me3 modifications at bivalent promoters³⁵. Since the study did not examine the *Jarid2* expression in the *Erk1/Erk2*-depleted mESCs, we asked whether the observed reduction of PRC2 occupancy at bivalent promoters in the *Erk1/Erk2*-depleted mESCs was due to reduced JARID2 expression as observed in naive mESCs.

To examine this possibility, we established multiple *Erk1/Erk2*-dKO mESC lines by the Crispr-Cas9-mediated gene knockout approach. The depletion of ERK1/ERK2 proteins in two independent *Erk1/Erk2*-dKO lines were confirmed by WB analyses (Figure 9A, lane 1-3). Compared to wild-type mESCs, the *Erk1/Erk2*-dKO mESCs cultured in serum-containing medium supplemented with a GSK3 inhibitor CHIR99021 were able to maintain normal colony morphology (Figure 10A). RNA-seq analyses showed that compared to mESC-S, the *Erk1/Erk2*-dKO mESCs had similar expression of *Prdm14*, *Nanog*, *Klf4*, *c-myc*, *Pou5f1*, and *Sox2* as naive mESCs (Figure 5C, 10B). Importantly, RNA-seq analyses showed that *Jarid2* expression was significantly reduced while other key PRC1 and PRC2 genes had similar or increased expression in the *Erk1/Erk2*-dKO mESCs (Figure 9B, 9C). The decreased *Jarid2* expression in the *Erk1/Erk2*-dKO cells is confirmed by qRT-PCR and WB analyses at both mRNA and protein levels (Figure 9A lane 1-3, 9D). To corroborate the findings, we rescued the ERK signaling by lentiviral vectors expressing wild-type ERK2 fused with a self-cleaved P2A peptide-linked puromycin resistant protein. After transduction, the transduced cells were selected by puromycin and the expression of exogenous ERK2 was confirmed by WB analysis (Figure 9A lane 4). RNA-seq analysis showed

the *Jarid2* expression was restored in the wild-type *Erk2*-rescued cells (Figure 9B, 9C), which was further confirmed by qRT-PCR and WB analyses (Figure 9A lane 4, 9D).

Based on these results, we concluded that *Jarid2* expression in mESCs was positively regulated by the ERK signaling. Same as the chemical inhibition of FGF/ERK signaling in 2i medium, the genetic disruption of ERK signaling led to reduced *Jarid2* expression in mESCs cultured in serum-containing medium.

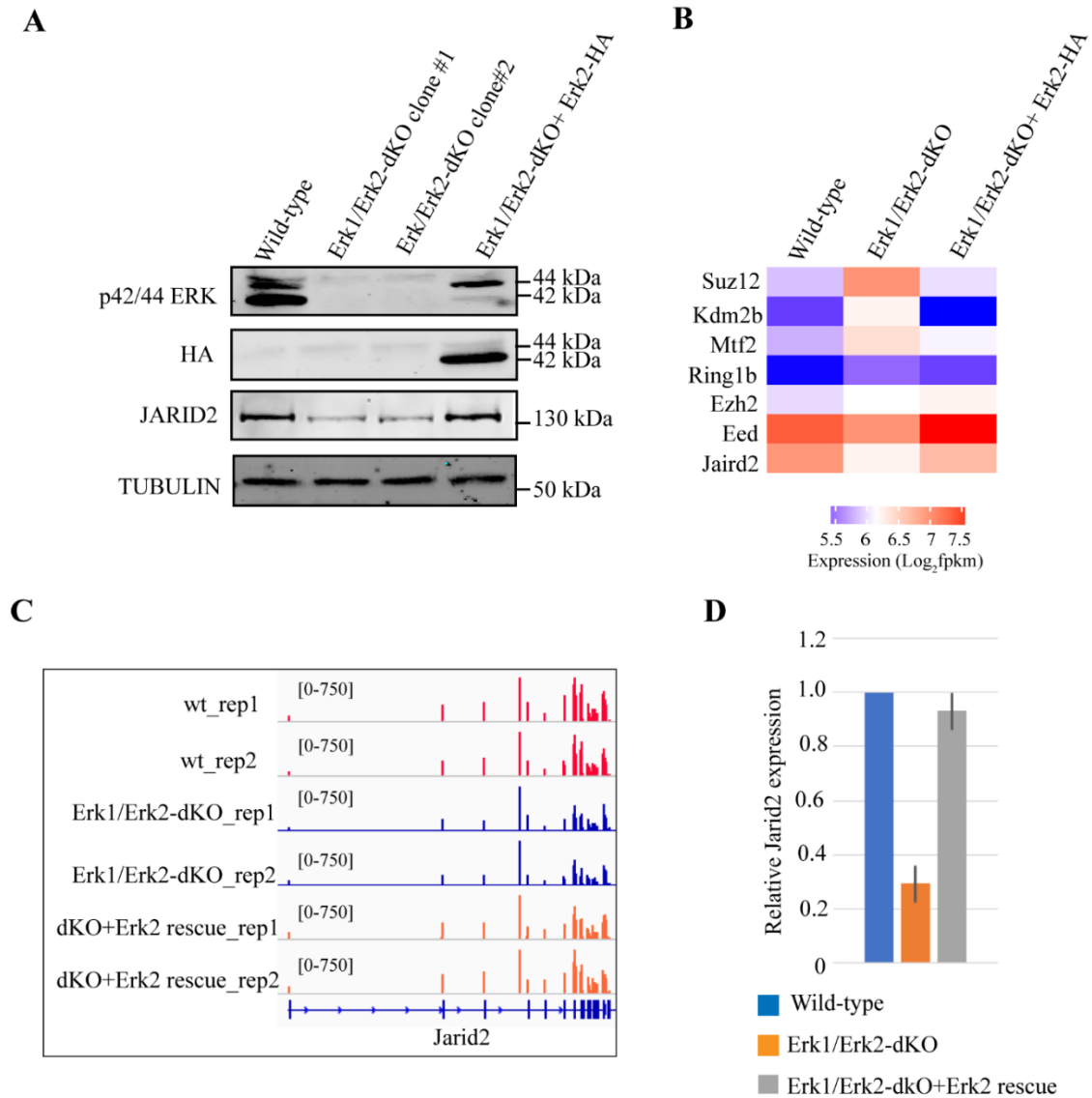


Figure 9: Knockout of Erk1/Erk2 reduces Jarid2 expression in mESCs. (A) Western blot analysis showing p42/44 MAPKs, ectopic expressed wild-type ERK2-HA, and JARID2 proteins in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO rescued with wild-type ERK2-HA mESCs. (B) Heatmap showing the expression of Polycomb core genes in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO rescued with wild-type ERK2-HA mESCs. (C) The IGV genome browser view of *Jarid2* expression in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO rescued with wild-type Erk2-HA mESCs. (D) qRT-PCR analysis showing the *Jarid2* expression in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO rescued with wild-type ERK2-HA mESCs. The results were normalized against levels of *Gapdh* and the expression level in wild-type mESCs was arbitrarily set to 1. The error bars represent standard deviation (n=3).

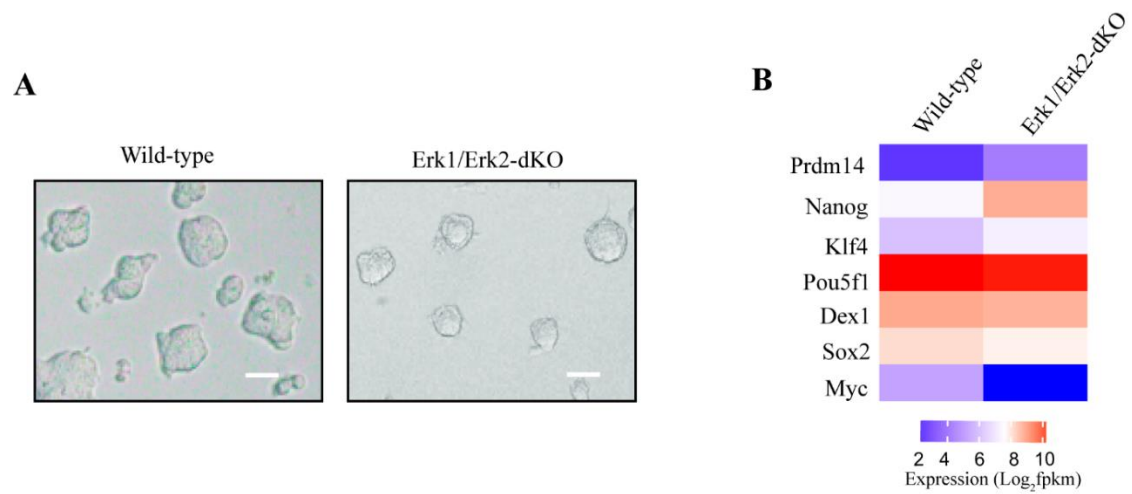


Figure 10: Knockout of *Erk1/Erk2* reduces *Jarid2* expression in mESCs. (A) Photos showing the colony morphology of wild-type and *Erk1/Erk2*-dKO mESCs. Scale bar = 100 μ m. (B) Heatmap showing the expression of pluripotent genes in wild-type and *Erk1/Erk2*-dKO mESCs.

The global PRC2 occupancy at CGIs is largely reduced in naive mESCs

To examine the global epigenetic changes in naive mESCs, we performed ChIP-seq analyses to compare the genome-wide occupancy of PRC2 components JARID2/EZH2, ncPRC1.1 components KDM2B/RING1B, and PRC2-mediated histone H3K27me3 modification as well as Trithorax MLL1/MLL2 complexes-mediated H3K4me3 modification in mESC-S and mESC-2i. Since Polycomb complexes and Trithorax MLL1/MLL2 complexes were globally recruited to CGIs in mammalian cells, we measured the overall Polycomb occupancy and histone H3K27me3/H3K4me3 modifications by calculating the total normalized mapped reads within CGIs and 10kb CGI-flanking regions. Consistent with previous reports^{26, 28}, both genome-wide PRC2 occupancy and histone H3K27me3 modification were largely reduced at CGIs in naive mESCs, which correlated well with the global decreased JARID2 occupancy (Figure 11A-C, 11E). However, KDM2B, one ncPRC1.1 component that binds to unmethylated CGIs through its CXXC-ZF domain, had similar genome-wide occupancy at CGIs in mESC-S and mESC-2i (Figure 11D, 11E). In contrast to the reduced H3K27me3 modification, the global H3K4me3 modification at CGIs was slightly increased in mESC-2i (Figure 12A). Also, in contrast to the reduced EZH2 occupancy at both CGIs and 10-kb CGI-flanking regions, the occupancy of PRC1 core component RING1B was slightly reduced at the CGI flanking regions but not the CGI center (Figure 12B).

To examine whether the PRC2 occupancy and H3K27me3 modification were redistributed to non-CGI regions in naive mESCs^{26, 31}, we calculated the overall occupancy of JARID2/EZH2 and H3K27me3 modification over extended 50kb CGI-flanking regions. The results did not reveal any significant increased PRC2 occupancy and H3K27me3 modification within 50kb CGI-flanking regions (Figure 12C-E). Since JARID2 is a known PRC2 recruiting factor^{10, 12, 36}, these results suggested that reduced *Jarid2* expression was likely to be a key molecular mechanism

leading to the reduced global PRC2 occupancy and histone H3K27me3 modification at CGIs in naive mESCs.

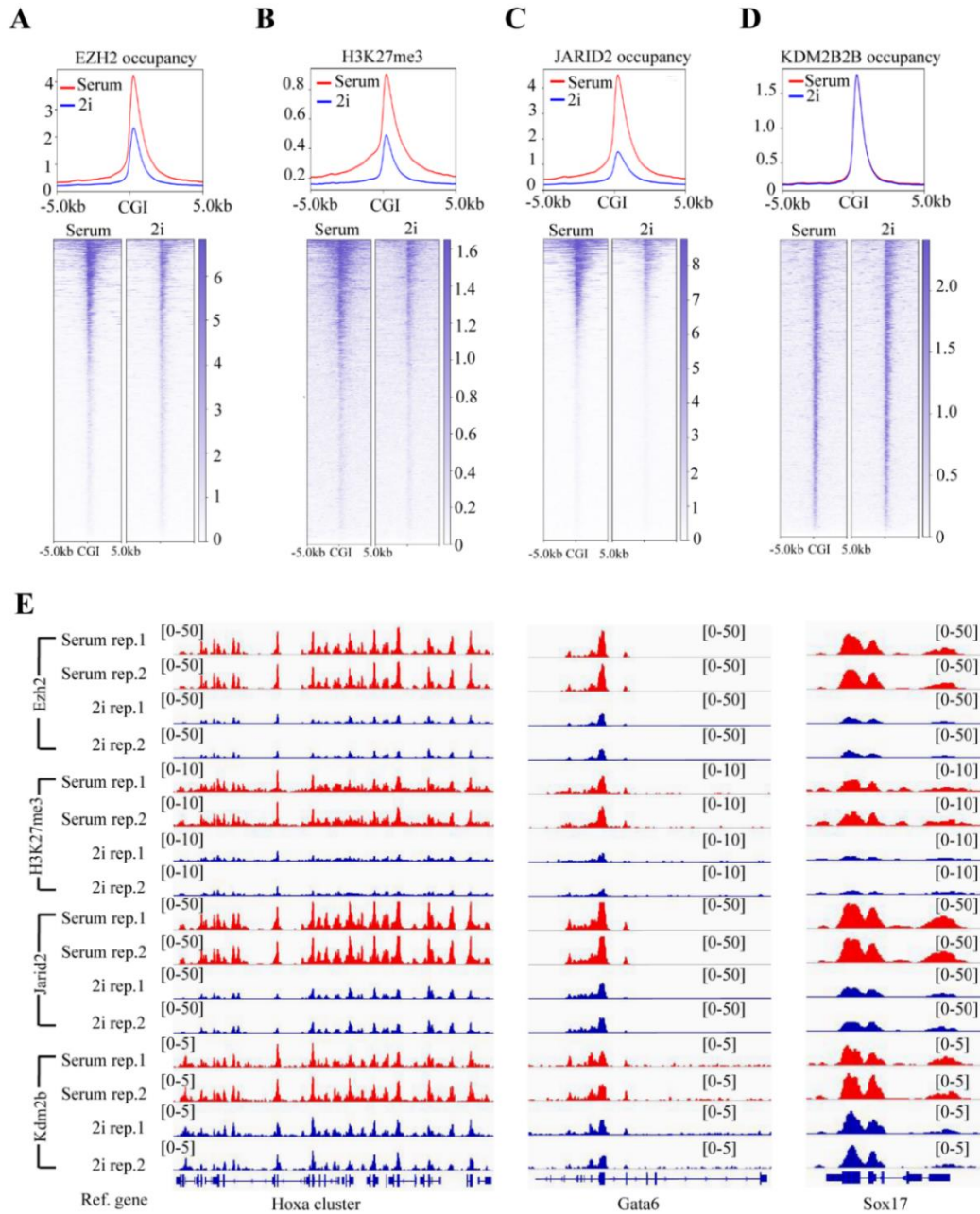


Figure 11: The global PRC2 occupancy at CGIs is largely reduced in naive mESCs. (A) The plot (upper) and heatmap (bottom) showing the EZH2 occupancy at CGIs and 10kb CGI-flanking regions in mESC-S and mESC-2i. (B) The plot (upper) and heatmap (bottom) showing the histone H3K27me3 modification at CGIs and 10kb CGI-flanking regions in mESC-S and mESC-2i. (C) The plot (upper) and heatmap (bottom) showing the JARID2 occupancy at CGIs and 10kb CGI-flanking regions in mESC-S and mESC-2i. (D) The plot (upper) and heatmap (bottom) showing the KDM2B occupancy at CGIs and 10kb CGI-flanking regions in mESC-S and mESC-2i. (E) The IGV genome browser view of EZH2, JARID2, KDM2B occupancy and histone H3K27me3 modification at the representative *Hoxa* gene locus (left panel), *Gata6* (middle panel), and *Sox17* (right panel).

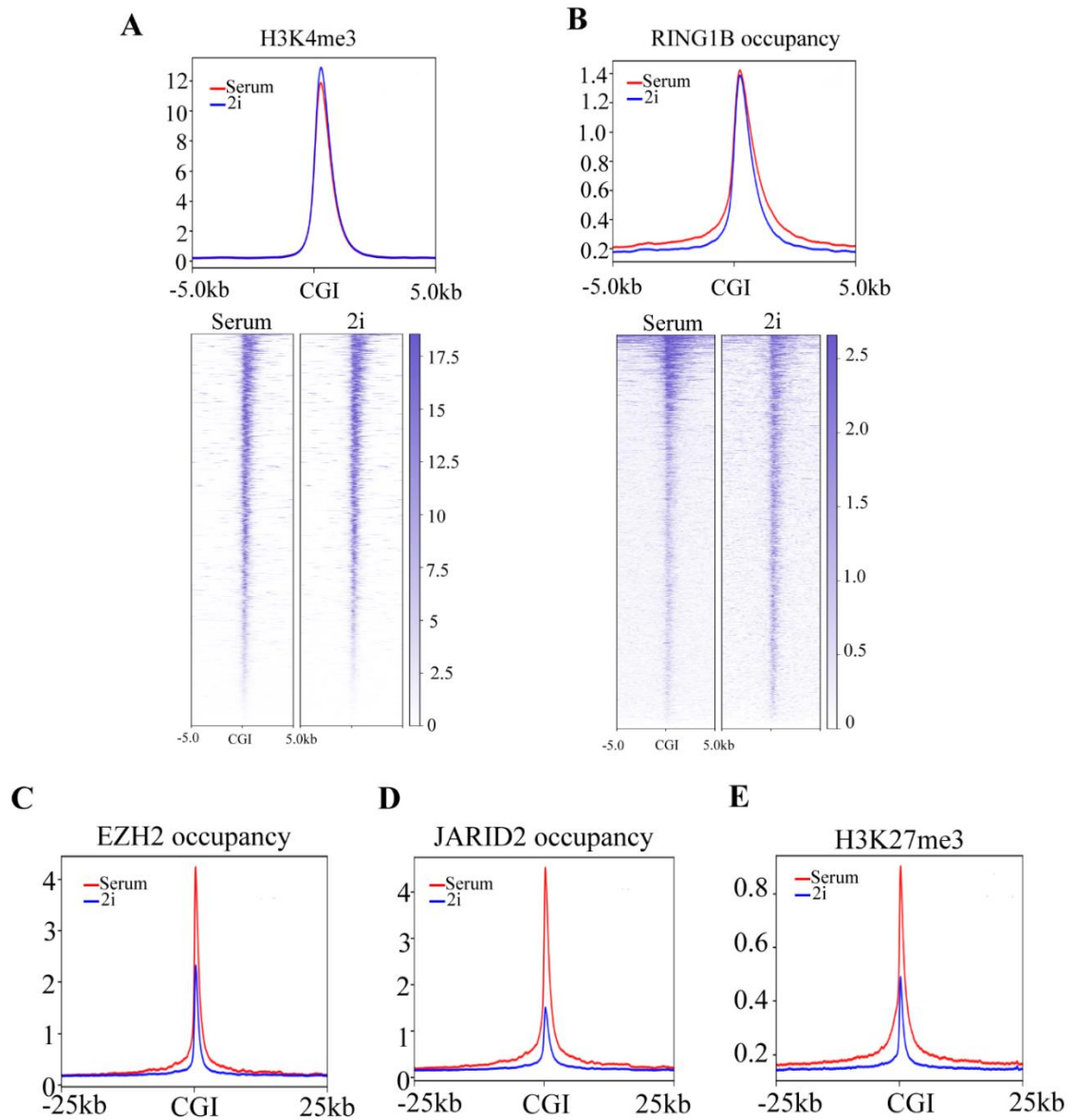


Figure 12: The global PRC2 occupancy at CGIs is largely reduced in naive mESCs. (A) The plot (upper) and heatmap (bottom) showing the histone H3K4me3 modification at CGIs and 10kb-CGI flanking regions in mESC-S and mESC-2i. (B) The plot (upper) and heatmap (bottom) showing the RING1B occupancy at CGIs and 10kb-CGI flanking regions in mESC-S and mESC-2i. (C) The plot showing the EZH2 occupancy at CGIs and 50kb-CGI flanking regions in mESC-S and mESC-2i. (D) The plot showing the JARID2 occupancy at CGIs and 50kb-CGI flanking regions in mESC-S and mESC-2i. (E) The plot showing the histone H3K27me3 at CGIs and 50kb-CGI flanking regions in mESC-S and mESC-2i.

Ectopic expression of *Jarid2* restores the global PRC2 occupancy in naive mESCs

To further examine whether the reduced *Jarid2* expression was the main reason causing the global reduction of PRC2 occupancy at CGIs in naive mESCs, we rescued the *Jarid2* expression in naive mESCs by lentiviral viruses expressing JARID2 fused with a self-cleaved P2A peptide-linked puromycin resistant protein. After transduction, the transduced cells were selected by puromycin in the medium. RNA-seq analyses showed that the *Jarid2* expression increased to a comparable level as that in mESC-S (Figure 13A). The rescued *Jarid2* expression is confirmed by qRT-PCR and WB analyses at both mRNA and protein levels (Figure 13B, 13C). The mESCs with ectopic *Jarid2* expression displayed normal mESC morphology under both serum-containing and 2i medium, as well as had similar expression of pluripotent genes and early lineage-specific genes as wild-type mESCs (Figure 13D-F). The ChIP-seq analyses demonstrated that the global EZH2 occupancy and histone H3K27me3 modification were fully restored, which correlated with the restored global JARID2 occupancy in naive mESCs (Figure 14A-D). Next, we extracted a total of 2830 unique bivalent promoters containing both H3K27me3 and H3K4me3 modifications and analyzed the JARID2/EZH2 occupancy as well as H3K27me3 modifications on these promoters (Figure 13G). Same as CGIs, the bivalent promoters had reduced JARID2/EZH2 occupancy and H3K27me3 modification in naive mESCs, which could be fully rescued by ectopic expression of *Jarid2* (Figure 13H-J). Based on these results, we concluded that the reduced *Jarid2* expression was the main molecular mechanism leading to the reduction of global PRC2 occupancy and histone H3K27me3 modification at CGIs and bivalent promoters in naive mESCs.

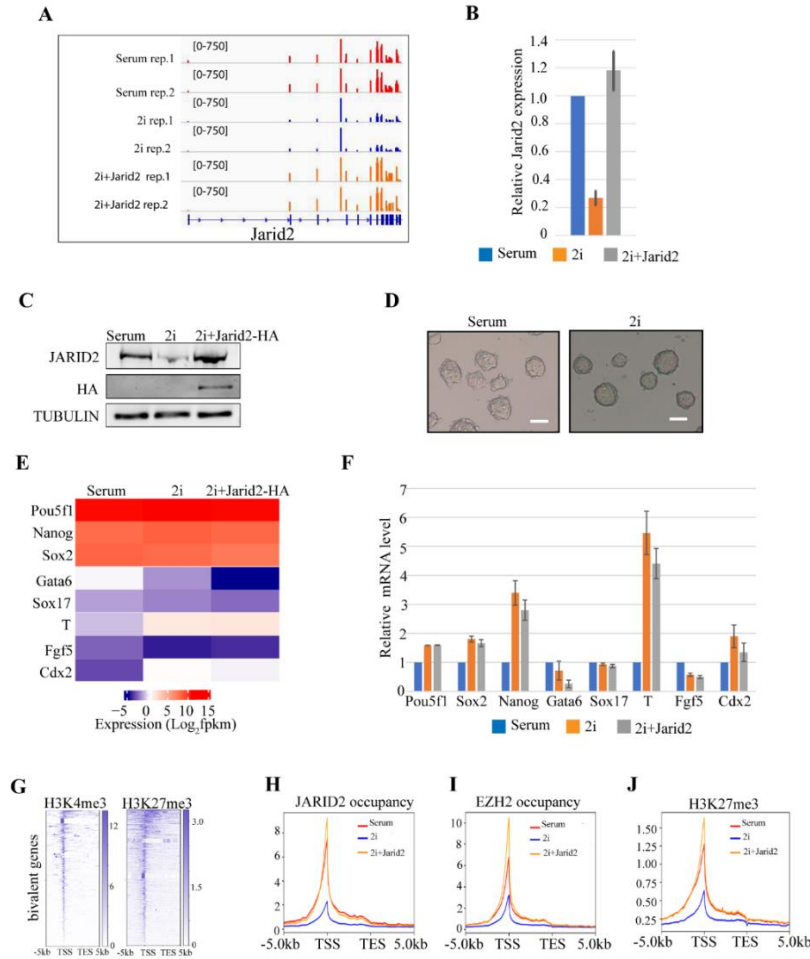


Figure 13: Ectopic expression of *Jarid2* restores the global PRC2 chromatin occupancy in naive mESCs. (A) The IGV genome browser view of *Jarid2* expression in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). (B) qRT-PCR analysis showing the *Jarid2* mRNA levels in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). The results were normalized against levels of *Gapdh* and the expression level in mESC-S was arbitrarily set to 1. The error bars represent standard deviation (n=3). (C) Western blot analysis showing the JARID2 protein level in mESC-S (serum), mESC-2i (2i), and the mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). (D) Photos showing the colony morphology of mESC-2i with ectopically expressed *Jarid2* in serum-containing medium and 2i medium. Scale bar = 100 μ m. (E) The Heatmap showing the expression of pluripotent genes and early lineage-specific genes in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). (F) qRT-PCR analysis showing the expression of pluripotent genes and early lineage-specific genes in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). (G) The heatmap showing 2830 bivalent gene promoters contain both H3K27me3 and H3K4me3 modifications. (H) The plot showing the JARID2 occupancy at 2830 bivalent gene promoters in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). (I) The plot showing the EZH2 occupancy at 2830 bivalent gene promoters in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). (J) The plot showing the H3K27me3 modification at 2830 bivalent gene promoters in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). TSS: transcription starting sites; TES: transcription ending sites.

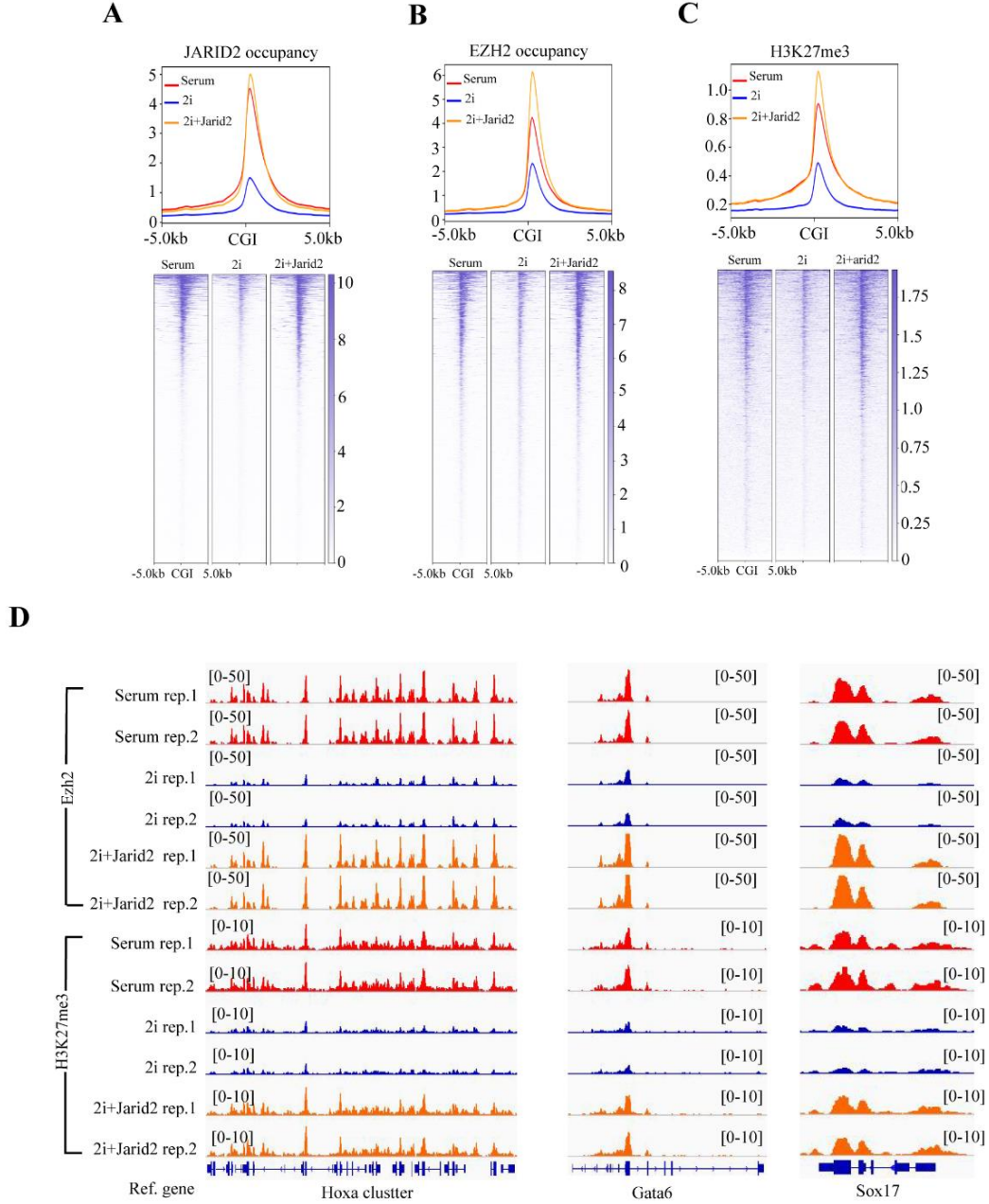


Figure 14: Ectopic expression of *Jarid2* restores the global PRC2 occupancy in naive mESCs. (A) The plot (upper) and heatmap (bottom) showing the JARID2 occupancy at CGIs and 10kb CGI-flanking regions in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). (B) The plot (upper) and heatmap (bottom) showing the EZH2 occupancy at CGIs and 10kb CGI-flanking regions in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). (C) The plot (upper) and heatmap (bottom) showing the histone H3k27me3 at CGIs and 10kb CGI-flanking regions in mESC-S (serum), mESC-2i (2i), and mESC-2i with ectopically expressed *Jarid2* (2i+Jarid2). (D) The IGV genome browser view of EZH2 and histone H3K27me3 modification at the representative *Hoxa* gene locus (left panel), *Gata6* (middle panel), and *Sox17* (right panel).

Ectopic expression of *Jarid2* restores the global PRC2 occupancy in the *Erk1/Erk2*-dKO mESCs

A previous study reported that genetic depletion of *Erk1/Erk2* resulted in reduced PRC2 occupancy and histone H3K27me3 modification at bivalent promoters in mESCs³⁵. Using the established *Erk1/Erk2*-dKO mESCs (Figure 9), we performed ChIP-seq analyses to examine the genome-wide occupancy of JARID2/EZH2 and H3K27me3 modifications in these cells. The results showed that the occupancy of JARID2/EZH2 and H3K27me3 modification were largely reduced not only at the bivalent promoters (Figure 15A-C), but more broadly at CGIs (Figure 16A-C). Importantly, the ectopic expression of wild-type *Erk2* in the *Erk1/Erk2*-dKO mESCs not only increased the *Jarid2* expression (Figure 9), but also restored the JARID2/EZH2 occupancy and histone H3K27me3 modification at both bivalent promoters and CGIs (Figure 15A-C, 16A-C).

To further examine whether the reduced JARID2-mediated PRC2 recruitment led to the reduction of JARID2/EZH2 occupancy and H3K27me3 modification at CGIs in the *Erk1/Erk2*-dKO cells, we ectopically expressed *Jarid2* in the *Erk1/Erk2*-dKO cells (Figure 15D). ChIP-seq analyses showed that the occupancy of JARID2/EZH2 and histone H3K27me3 modification at CGIs as well as bivalent promoters were fully restored in the ectopically *Jarid2*-expressed *Erk1/Erk2*-dKO mESCs (Figure 16D-F, 15E-G). Thus, same as in the naive mESCs, the reduced genome-wide PRC2 occupancy at CGIs and bivalent promoters appeared to be mainly caused by the reduced *Jarid2* expression and decreased JARID2-mediated PRC2 recruitment in the *Erk1/Erk2*-dKO mESCs.

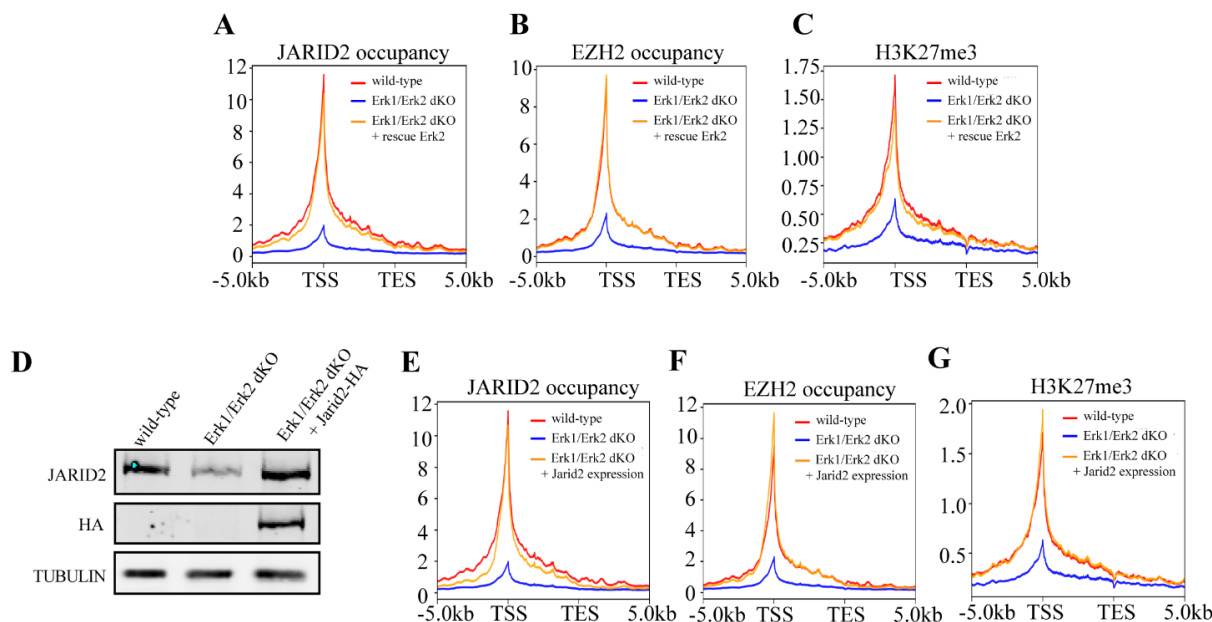


Figure 15: Ectopic expression of *Erk2* or *Jarid2* restores the global PRC2 occupancy at bivalent promoters in *Erk1/Erk2*-dKO mESCs. (A) The plot showing the JARID2 occupancy at 2830 bivalent gene promoters in wild-type mESCs (wild-type), *Erk1/Erk2*-dKO mESCs (*Erk1/Erk2*-dKO), and *Erk1/Erk2*-dKO cells rescued with wild-type *Erk2* (*Erk1/Erk2*-dKO + rescue *Erk2*). (B) The plot showing the EZH2 occupancy at 2830 bivalent gene promoters in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO + rescue *Erk2* mESCs. (C) The plot showing the histone H3k27me3 modification at 2830 bivalent gene promoters in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO + rescue *Erk2* mESCs. (D) Western blot analysis showing the JARID2 protein level in wild-type mESCs (wild-type), *Erk1/Erk2*-dKO mESCs (*Erk1/Erk2*-dKO), and *Erk1/Erk2*-dKO cells with ectopically expressed *Jarid2* (*Erk1/Erk2*-dKO + *Jarid2* expression). (E) The plot showing the JARID2 occupancy at 2830 bivalent gene promoters in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO + *Jarid2* expression mESCs. (F) The plot showing the EZH2 occupancy at 2830 bivalent gene promoters in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO + *Jarid2* expression mESCs. (G) The plot showing the histone H3k27me3 modification at 2830 bivalent gene promoters in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO + *Jarid2* expression mESCs. TSS: transcription starting sites; TES: transcription ending sites.

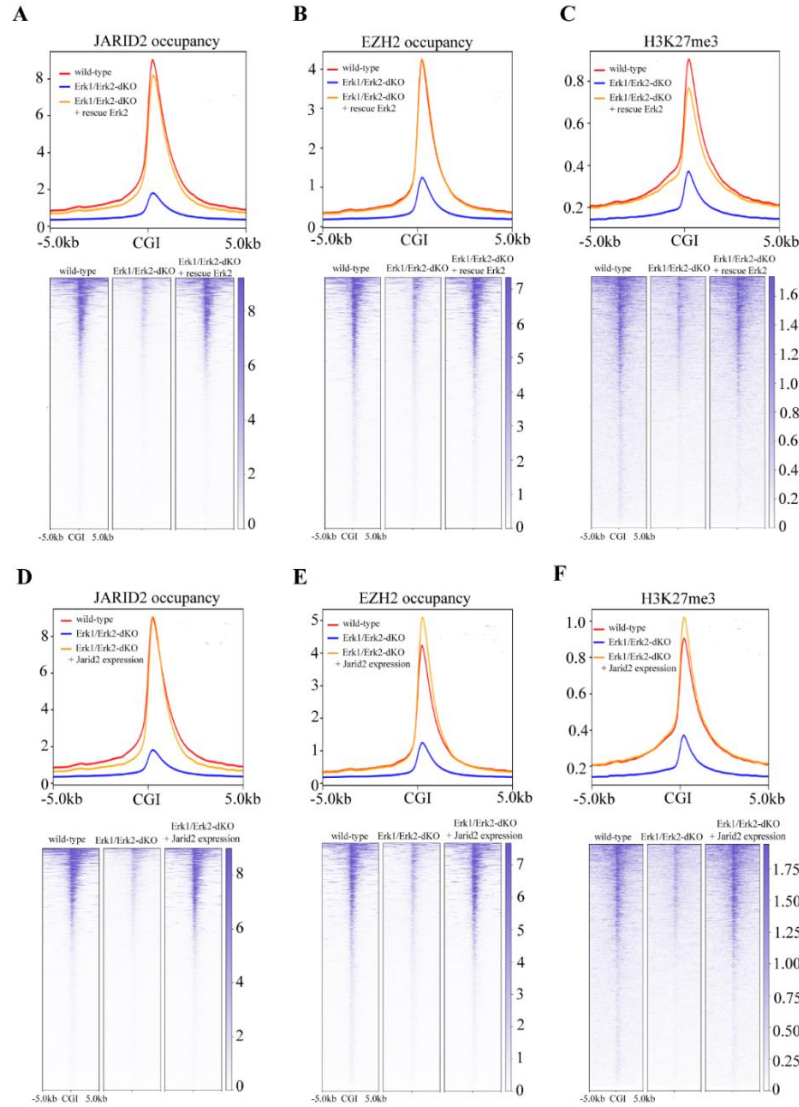


Figure 16: Ectopic expression of *Erk2* or *Jarid2* restores the global PRC2 occupancy in *Erk1/Erk2*-dKO mESCs. (A) The plot (upper) and heatmap (bottom) showing the JARID2 occupancy at CGIs and 10kb CGI-flanking regions in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO + rescue *Erk2* mESCs. (B) The plot (upper) and heatmap (bottom) showing the EZH2 occupancy at CGIs and 10kb CGI-flanking regions in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO + rescue *Erk2* mESCs. (C) The plot (upper) and heatmap (bottom) showing the histone H3k27me3 modification at CGIs and 10kb CGI-flanking regions in wild-type, *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO + rescue *Erk2* mESCs. (D) The plot (upper) and heatmap (bottom) showing the JARID2 occupancy at CGIs and 10kb CGI-flanking regions in wild-type mESCs (wild-type), *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO with ectopically expressed *Jarid2* (*Erk1/Erk2*-dKO + *Jarid2* expression) mESCs. (E) The plot (upper) and heatmap (bottom) showing the EZH2 occupancy at CGIs and 10kb CGI-flanking regions in wild-type mESCs (wild-type), *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO with ectopically expressed *Jarid2* (*Erk1/Erk2*-dKO + *Jarid2* expression) mESCs. (F) The plot (upper) and heatmap (bottom) showing the histone H3k27me3 modification at CGIs and 10kb CGI-flanking regions in wild-type mESCs (wild-type), *Erk1/Erk2*-dKO, and *Erk1/Erk2*-dKO with ectopically expressed *Jarid2* (*Erk1/Erk2*-dKO + *Jarid2* expression) mESCs.

De-repression of bivalent genes appears to be determined by the presence of signaling-associated transcription factors but not the status of PRC2 occupancy in naive mESCs

To examine whether the reduced PRC2 occupancy led to transcriptional changes of its regulated genes in naive mESCs, we compared the expression of 2830 bivalent genes in mESC-S versus mESC-2i by RNA-seq analyses (Figure 13G). The results identified 589 downregulated genes and 253 de-repressed genes in mESC-2i (cutoff: 1.5 fold expression change and $q < 0.05$) (Fig. 7A, S7A, S7B). Since all bivalent gene promoters had the same reduced PRC2 occupancy and histone H3K27me3 modification in mESC-2i (Figure 13G-J), the distinct transcriptional status of these two gene groups suggested that the transcriptional activation was determined by factors other than the PRC2 occupancy at promoters. The gene ontology (GO) analysis showed that the downregulated genes had enriched GO terms involved in cell differentiation, organ morphogenesis, and multicellular development (cutoff: Benjamini adjusted $p < 0.05$), which included multiple primitive endoderm-specific genes such as *Gata6*, *Dab1*, *Dab2*, *Sox7*, and *Sox17* that were known to be activated by the FGF/ERK signaling (Figure 17B, 18C, Table 4). On the other hand, the enriched GO terms of de-repressed genes were involved in multicellular development, transcriptional regulation, and canonical WNT signaling (cutoff: Benjamini adjusted $p < 0.05$) (Figure 18D, Table 5). Notably, multiple known Wnt/ β -catenin direct target genes such as *Axin2*, *T*, *Tcf7*, *Snai1*, *Cdx1*, and *Cdx2* were de-repressed in naive mESCs³⁷⁻⁴², suggesting the de-repression of these genes might be induced by transcription factors associated with the activated Wnt/ β -catenin signaling in 2i medium.

To further examine whether these genes were activated by the Wnt/ β -catenin signaling in naive mESCs, we inactivated the Wnt/ β -catenin signaling by removing GSK3 inhibitor CHIR99021 and leaving a single MEK inhibitor PD0325901 in the medium (mESC-PD). RNA-seq analysis showed that 136 out of 253 genes that were de-repressed in mESC-2i, including all

known Wnt/ β -catenin direct target genes such as *Axin2*, *T*, *Tcf7*, *Snai1*, *Cdx1*, and *Cdx2*, lost the transcriptional de-repression and were re-silenced in mESC-PD (Figure 17C, 17D). Further ChIP-seq analysis showed that compared to mESC-S, both mESC-2i and mESC-PD had reduced EZH2/JARID2 occupancy and histone H3K27me3 modification at these 136 gene promoters, while β -CATENIN occupancy at gene promoters increased in mESC-2i but not in mESC-S or mESC-PD, supporting the transcriptional activation of these genes in mESC-2i relied on the transcription factor binding to the promoters (Figure 17E, 17F).

Hence, all these results suggested that the reduced PRC2 occupancy at gene promoters alone was insufficient to activate transcription of bivalent genes in naive mESCs. Compared to the reduced PRC2 occupancy at gene promoters, the presence of transcription factors appeared to be necessary and play a predominant role in inducing transcriptional activation.

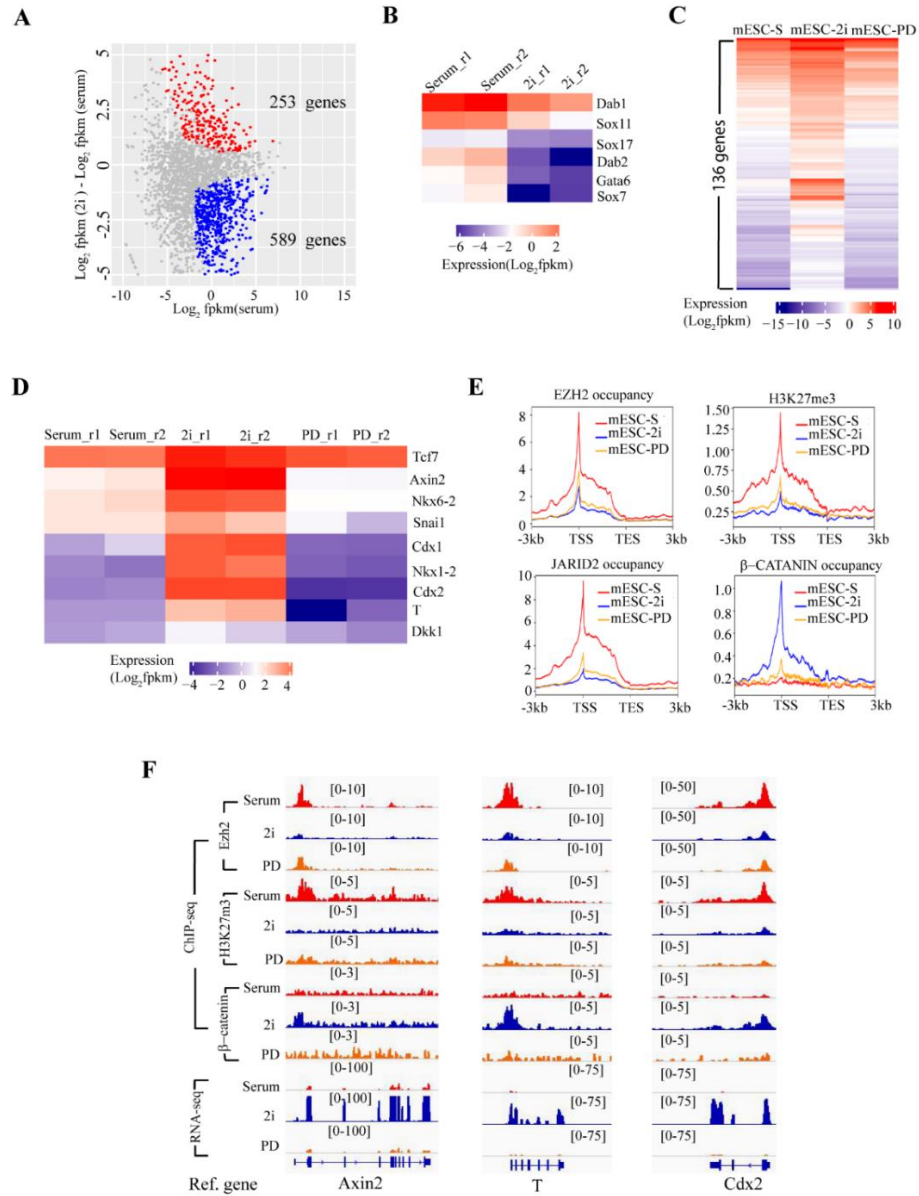


Figure 17: De-repression of bivalent genes is determined by the presence of signaling-associated transcription factors but not the status of PRC2 occupancy in naive mESCs. (A) Plot showing 253 up- and 589 down-regulated bivalent genes in mESC-2i compared to mESC-S. (B) Heatmap showing the expression of representative primitive endoderm-specific genes in mESC-S and mESC-2i. (C) Heatmap showing 136 bivalent genes de-repressed in mESC-2i and re-silenced after inactivation of Wnt/β-catenin signaling (mESC-PD). (D) Heatmap showing the representative Wnt/β-catenin direct target genes de-repressed in mESC-2i and re-silenced after inactivation of Wnt/β-catenin signaling (mESC-PD). (E) The plot showing the EZH2 occupancy, JARID2 occupancy, H3K27me3 modification, and β-CATENIN occupancy at 136 Wnt/β-catenin target genes in mESC-S, mESC-2i, and mESC-PD. TSS: transcription starting sites; TES: transcription ending sites. (F) The IGV genome browser view of EZH2 occupancy, H3K27me3 modification, β-CATENIN occupancy, and gene expression of representative Wnt/β-catenin direct target genes *Axin2*, *T*, and *Cdx2* in mESC-S, mESC-2i and mESC-PD.

Table 4: Result of gene ontology analysis on the downregulated bivalent genes in naive mESCs

Category	Term	Count	P-Value	FDR
GOTERM_BP_DIRECT	GO:0007275~multicellular organism development	61	7.57E-13	1.32E-09
GOTERM_BP_DIRECT	GO:0030154~cell differentiation	51	2.56E-12	4.47E-09
GOTERM_BP_DIRECT	GO:0009887~organ morphogenesis	14	5.25E-07	9.17E-04
GOTERM_BP_DIRECT	GO:0007399~nervous system development	25	1.83E-06	0.0031884
GOTERM_BP_DIRECT	GO:0071371~cellular response to gonadotropin stimulus	5	6.59E-06	0.0115095
GOTERM_BP_DIRECT	GO:0043046~DNA methylation involved in gamete generation	6	2.83E-05	0.0494691
GOTERM_BP_DIRECT	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	34	3.61E-05	0.06303
GOTERM_BP_DIRECT	GO:0042493~response to drug	21	3.96E-05	0.0690304
GOTERM_BP_DIRECT	GO:0045893~positive regulation of transcription, DNA-templated	29	4.21E-05	0.0735148
GOTERM_BP_DIRECT	GO:0090090~negative regulation of canonical Wnt signaling pathway	11	6.10E-05	0.106346
GOTERM_BP_DIRECT	GO:0048557~embryonic digestive tract morphogenesis	6	6.39E-05	0.1114602
GOTERM_BP_DIRECT	GO:0001701~in utero embryonic development	19	7.41E-05	0.1292034
GOTERM_BP_DIRECT	GO:0007626~locomotory behavior	11	9.91E-05	0.1727516
GOTERM_BP_DIRECT	GO:0045766~positive regulation of angiogenesis	11	2.54E-04	0.4422398
GOTERM_BP_DIRECT	GO:0048661~positive regulation of smooth muscle cell proliferation	9	2.89E-04	0.5030369
GOTERM_BP_DIRECT	GO:0045104~intermediate filament cytoskeleton organization	5	2.95E-04	0.5131394
GOTERM_BP_DIRECT	GO:0038063~collagen-activated tyrosine kinase receptor signaling pathway	4	4.88E-04	0.8480843
GOTERM_BP_DIRECT	GO:0001938~positive regulation of endothelial cell proliferation	8	5.28E-04	0.9177038
GOTERM_BP_DIRECT	GO:0008285~negative regulation of cell proliferation	20	5.80E-04	1.0082991
GOTERM_BP_DIRECT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	38	6.61E-04	1.1480858
GOTERM_BP_DIRECT	GO:0007605~sensory perception of sound	11	7.23E-04	1.2541734
GOTERM_BP_DIRECT	GO:0007420~brain development	14	7.28E-04	1.2639078
GOTERM_BP_DIRECT	GO:0050680~negative regulation of epithelial cell proliferation	8	8.17E-04	1.4169911

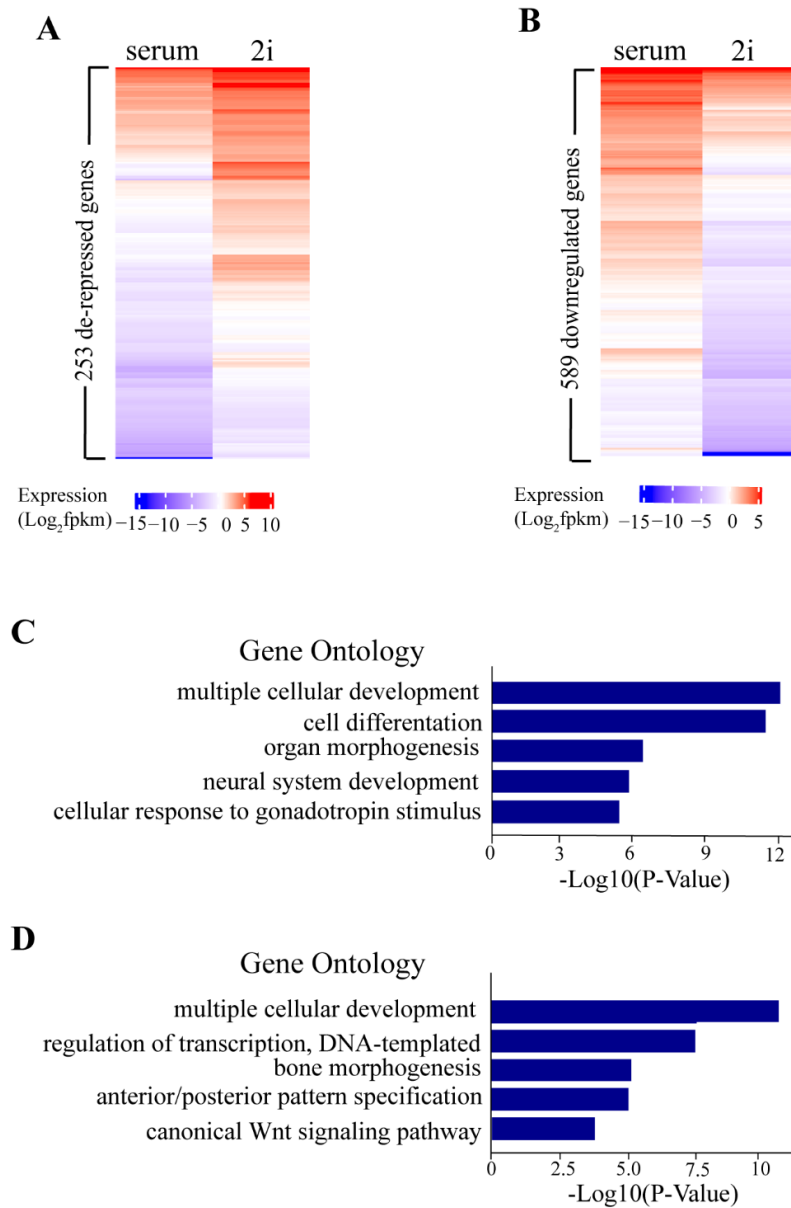


Figure 18: De-repression of bivalent genes is determined by the presence of signaling-associated transcription factors but not the status of PRC2 occupancy in naive mESCs. (A) Heatmap showing 589 bivalent genes downregulated in mESC-2i. (B) Heatmap showing 253 bivalent genes de-repressed in mESC-2i. (C). Gene ontology analysis showing the enriched GO terms of 589 downregulated bivalent genes in mESC-2i (cutoff: Benjamini adjusted $p < 0.05$). (D) Gene ontology analysis showing the enriched GO terms of 253 de-repressed bivalent genes in mESC-2i (cutoff: Benjamini adjusted $p < 0.05$).

Table 5: Result of gene ontology analysis on the upregulated bivalent genes in naive mESCs

Category	Term	Count	P-Value	FDR
GOTERM_BP_DIRECT	GO:0007275~multicellular organism development	36	1.28E-10	2.08E-07
GOTERM_BP_DIRECT	GO:0006355~regulation of transcription, DNA-templated	50	9.49E-08	1.54E-04
GOTERM_BP_DIRECT	GO:0006351~transcription, DNA-templated	43	4.03E-07	6.54E-04
GOTERM_BP_DIRECT	GO:0060349~bone morphogenesis	6	1.50E-05	0.0242873
GOTERM_BP_DIRECT	GO:0009952~anterior/posterior pattern specification	9	1.63E-05	0.0263717
GOTERM_BP_DIRECT	GO:0030509~BMP signaling pathway	8	2.58E-05	0.0417793
GOTERM_BP_DIRECT	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	21	4.05E-05	0.0657075
GOTERM_BP_DIRECT	GO:0060070~canonical Wnt signaling pathway	7	2.63E-04	0.4250707
GOTERM_BP_DIRECT	GO:0016055~Wnt signaling pathway	10	2.90E-04	0.4689491
GOTERM_BP_DIRECT	GO:0035556~intracellular signal transduction	13	6.88E-04	1.1103687
GOTERM_BP_DIRECT	GO:0009653~anatomical structure morphogenesis	4	9.76E-04	1.5720754

DISCUSSION

In mammalian cells PRC1 and PRC2 are recruited to unmethylated CGIs and CGI-associated bivalent gene promoters. The occupancy of Polycomb complexes and their mediated covalent histone modifications at gene promoters are critical for maintaining mESCs in an undifferentiated state by preventing stochastic transcription of lineage differentiation genes.

Previous studies have revealed that both JARID2 and MTF2 are required for the recruitment of PRC2¹⁰⁻¹⁵, while KDM2B recruits a ncPRC1.1 variant to unmethylated CGIs through its CXXC-ZF domain¹⁶⁻¹⁸. Although the exact molecular mechanisms underlying the PRC2 recruitment to its targets are not fully elucidated, it is believed that the preferential binding to unmethylated CG-rich DNA sequences by the Polycomb recruiting factors JARID2, MTF2, and KDM2B is important to determine the genome-wide Polycomb occupancy at CGIs in mammalian cells^{9,43}.

Compared to the mESCs cultured in serum-containing medium, the naive mESCs in 2i medium have reduced basal expression of early differentiation genes^{21,22}. Unexpectedly, the PRC2-mediated histone H3K27me3, which functions as a major epigenetic mechanism mediating gene silencing in mESCs, has been found to have a global reduction in naive mESCs^{26,28}. Previous studies suggested the reduced PRC2 occupancy could be caused by relocation of PRC2 from CGIs to non-CGI DNA demethylated regions in naive mESCs, or reduced chromatin accessibility to PRC2 at bivalent promoters in the ERK1/ERK2 signaling-deficient mESCs^{29-31,35}.

For the first time, this study provides compelling evidence to demonstrate that the genome-wide decreased PRC2 occupancy at CGIs and bivalent promoters in naive mESCs is mainly caused by the reduced JARID2-mediated PRC2 recruitment. First, the RNA-seq, qRT-PCR, and WB analyses show that the *Jarid2* expression is significantly reduced in naive mESCs (Figure 5, 6),

and the reduced JARID2 expression is not due to different isoforms expressed in naive mESCs since WB analyses only detect a single long isoform protein (Figure 5D, 8D)⁴⁴. JARID2 is a known factor to recruit PRC2 to CGIs in mammalian cells^{10,12,36}. Therefore, the decreased JARID2 expression provides an important molecular basis for the global reduced PRC2 occupancy in naive mESCs. Second, the FGF/ERK signaling positively regulates the *Jarid2* expression since the reactivation of FGF/ERK signaling in naive mESCs upregulates the *Jarid2* expression at both mRNA and protein levels (Figure 17). Third, similar to the chemical inhibition of ERK signaling in the 2i medium, the genetic disruption of ERK signaling by deletion of *Erk1* and *Erk2* largely reduces the *Jarid2* expression in the mESCs cultured in the serum-containing medium, which can be rescued by the ectopic expression of wild-type *Erk2*, further confirming that ERK signaling positively regulates the expression of *Jarid2* (Figure 9, 10). Fourth, the ChIP-seq analyses show the global reduced EZH2 occupancy and histone H3K27me3 modification at CGIs and bivalent promoters correlate with the global reduced JARID2 occupancy, which can be fully rescued by the ectopic expression of *Jarid2* in both naive and *Erk1/Erk2*-dKO mESCs (Figure 11-16). Finally, although the PRC2 core component *Suz12* showed increased expression in mESC-2i, the findings that ectopic *Jarid2* expression fully restored the PRC2 occupancy in mESCs-2i suggest the reduced PRC2 occupancy in naive mESCs is mainly due to the reduced JARID2-mediated PRC2 recruitment, but unlikely caused by changed expression of *Suz12* or other PRC2 components. Collectively, all these results strongly indicate the FGF/ERK signaling determines the global PRC2 occupancy at CGIs and bivalent promoters mainly through regulating *Jarid2* expression in mESCs.

Several recent studies report that although the H3K27me3 modification is reduced at CGIs, the overall abundance of PRC2 and H3K27me3 in naive mESCs is increased^{30,31}. Furthermore, the H3K27me3 is redistributed to non-CGI euchromatin and heterochromatin regions, which

correlates with the local CpG density³¹. Since naive mESCs have global DNA demethylation, these results suggest the reduced PRC2 occupancy at CGIs in naive mESCs could be caused by relocation of PRC2 from CGIs to new DNA demethylated regions³¹. Although our results do not reveal any relocation of EZH2/JARID2 and H3K27me3 modification to 50kb CGI-flanking regions (Figure 12C-E), our current study does not preclude these conclusions due to the practical difficulty in mapping ChIP-seq reads to heterochromatin. To monitor whether there exists Polycomb redistribution in response to the global DNA demethylation in naive mESCs, we include KDM2B in our ChIP-seq assay since its CXXC-ZF domain has a high binding affinity to unmethylated CG-rich DNA such as CGIs, and importantly KDM2B is known to bind DNA demethylated pericentric heterochromatin in the DNA methyltransferases triple knockout (DNMT-TKO) cells^{17,19,45}. However, our study finds both mESC-S and mESC-2i have the same KDM2B occupancy at CGIs (Figure 11D). Since KDM2b has the same expression level in mESC-S and mESC-2i (Figure 13, 16), the results suggest that KDM2B is unlikely to have a significant redistribution or to mediate PRC2 relocation to non-CGI demethylated regions in naive mESCs. In addition, the histone H3K4me3 modification at CGIs, what is mainly mediated by MLL2 that binds to CGIs through a similar CXXC-ZF domain as KDM2B^{46,47}, is increased at CGIs in naive mESCs (Figure 12), further arguing against that the CXXC-ZF domain-containing proteins are relocated from CGIs to new DNA demethylated regions. Finally but importantly, the ectopic expression of *Jarid2* in naive mESCs fully restores the global PRC2 occupancy and the histone H3K27me3 modification at CGIs (Figure 13, 14), further supporting that the reduced PRC2 occupancy at CGIs in naive mESCs is mainly caused by the reduced JARID2-mediated PRC2 recruitment.

A previous study has reported that ERK1/ERK2 determine the PRC2 occupancy at bivalent promoters through regulating local chromatin accessibility to PRC2³⁵. Consistent with this report, we find that the global PRC2 occupancy and H3K27me3 modification are reduced in the *Erk1/Erk2*-dKO mESCs, which is rescued by ectopic expression of wild-type *Erk2* (Figure 16). However, we find that the same as naive mESCs, *Erk1/Erk2*-dKO mESCs have significant reduced *Jarid2* expression, which can be rescued by ectopic expression of wild-type *Erk2* (Figure 9). Importantly, the ectopic expression of *Jarid2* fully restores the EZH2/JARID2 occupancy and H3K27me3 modification at both CGIs and bivalent promoters in the *Erk1/Erk2*-dKO mESCs (Figure 15, 16). Since majority of bivalent promoters is associated with CGIs, our study suggests that the reduced JARID2-mediated PRC2 occupancy at CGIs is the main molecular mechanism leading to the reduction of PRC2 occupancy at bivalent promoters in the *Erk1/Erk2*-depleted mESCs. On the other hand, Tee et al. reported that the depletion of *Jarid2* impaired the phosphorylation of ERK1/ERK2, suggesting there might exist mutual regulation of JARID2 expression and ERK signaling activation³⁵. It is worth to examine how the interaction of JARID2 and ERK signaling affects the global PRC2 occupancy and gene expression in mESCs in future studies.

Depletion of Polycomb induces de-repression of some bivalent genes in mESCs cultured in serum-containing medium, in which mESCs express low-level transcription factors associated with FGF/ERK signaling^{17,48}, suggesting Polycomb binding to promoters is essential for setting high transcriptional thresholds to prevent stochastic transcription⁴⁹. However, previous studies showed that majority of bivalent genes remains silenced although the PRC2 occupancy at their promoters are reduced in naive mESCs^{21,26,27,50}. To solve this unexpected discrepancy, our study separates two groups of bivalent genes with distinct transcriptional states, either downregulated or

de-repressed, in naive mESCs (Figure 17A, 18A, 18B). Since all bivalent promoters have the same reduced PRC2 occupancy in mESC-2i (Figure 13G-J), the distinct transcriptional status of these two gene groups suggests that the transcriptional activation of bivalent genes is determined by factors other than the PRC2 occupancy at their promoters in naive mESCs. Further analyses reveal that the downregulated gene group contains multiple primitive endoderm-specific genes known to be activated by the FGF signaling^{48,51,52}, while the de-repressed gene group includes all known direct targets of Wnt/ β -catenin signaling, suggesting the de-repression of these genes is induced by transcription factors associated with the activated Wnt/ β -catenin signaling in 2i medium (Figure 17B-D, 18C-D)³⁷⁻⁴². Notably, more than half de-repressed genes in naive mESCs, including all known Wnt/ β -catenin direct target genes, are re-silenced after inactivation of Wnt/ β -catenin signaling by removal of GSK3 inhibitor from 2i medium (Figure 17C, 17D). In addition, the activation of Wnt signaling target genes in naive mESCs correlates with the increased β -CATENIN occupancy at their promoters (Figure 7E, 7F), further supporting that the presence of transcription factors, but not the reduced PRC2 occupancy at promoters, plays a predominant role in activating of bivalent genes in naive mESCs. Of note, this result is consistent with previous reports that the promoter-bound PRC2 does not actively repress transcription but rather sets high transcriptional thresholds for gene activation in cells^{49,50}.

In summary, our study provides compelling evidence to reveal a key molecular mechanism by which the FGF/ERK signaling regulates the global PRC2 occupancy at CGIs in naive mESCs, and to elucidate a fundamental question regarding the function of transcription factors and PRC2-mediated epigenetic mechanisms in transcriptional regulation. Based on the evidence, we propose a model to explain how cell signaling coordinates both global PRC2 recruitment and developmental gene expression in mESCs. In this model, the FGF/ERK signaling positively

regulates the *Jarid2* expression and promotes the JARID2-mediated PRC2 recruitment to CGIs in mESCs cultured in serum-containing medium. The PRC2 binding to promoters increases the overall transcriptional thresholds, which is essential for preventing stochastic transcription of lineage differentiation genes induced by the FGF/ERK signaling in serum-containing medium. In naive mESCs, the chemical inhibition of FGF/ERK signaling reduces the *Jarid2* expression and leads to a global reduction of JARID2-mediated PRC2 recruitment to CGIs and bivalent promoters, which subsequently reduces the overall transcriptional thresholds of bivalent genes. However, in the absence of FGF/ERK signaling-associated transcription factors, the reduced PRC2 occupancy at promoters alone is insufficient to activate FGF/ERK signaling target genes. In contrast, the Wnt/ β -catenin signaling target genes are de-repressed by the activated Wnt/ β -catenin signaling and its-associated transcription factors in naive mESCs (Figure 19).

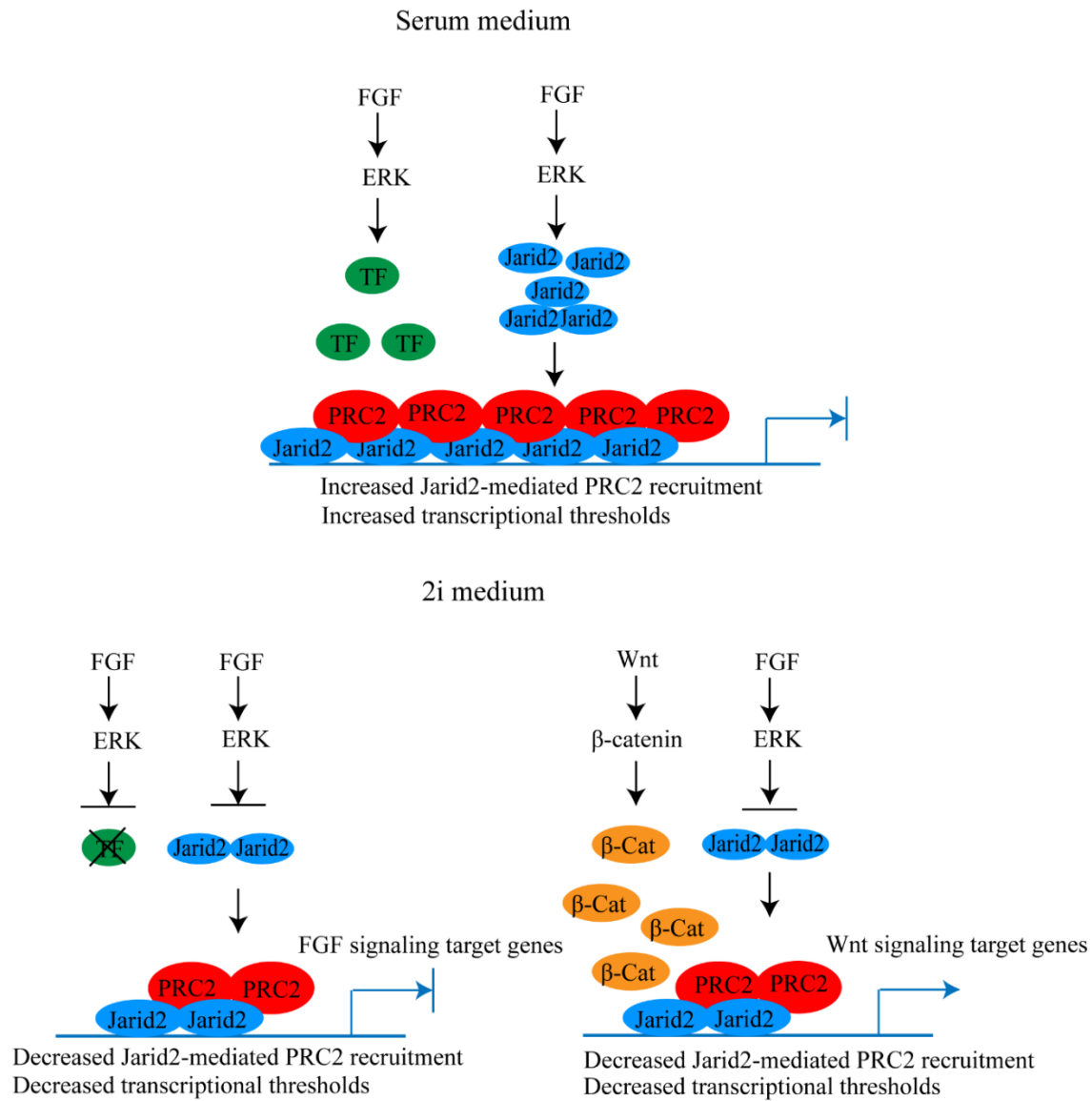


Figure 19: Proposed model: cell signaling coordinates PRC2 recruitment and developmental gene expression in mESCs. In serum-containing medium, FGF/ERK signaling increases both *Jarid2* expression and JARID2-mediated PRC2 recruitment to bivalent promoters, which increases the thresholds of transcriptional activation (upper panel). In 2i medium, the deficient FGF/ERK signaling decreases the *Jarid2* expression and JARID2-mediated PRC2 recruitment to bivalent promoters, which reduces the thresholds of transcriptional activation. In the absence of FGF/ERK signaling-associated transcriptional factors, the FGF/ERK signaling target genes remain silenced. In contrast, the activated Wnt/β-catenin signaling-associated transcription factors (β-catenin) activate the Wnt signaling target genes (lower panel).

LIMITATIONS OF THE STUDY

Previous studies show in naive mESCs the H3K27me3 is redistributed to non-CGI euchromatin and heterochromatin regions, which correlates with the local CpG density, suggesting the reduced PRC2 occupancy at CGIs in naive mESCs could be caused by the relocation of PRC2 from CGIs to new DNA demethylated regions including pericentric heterochromatin³¹. Due to the practical difficulty in mapping ChIP-seq reads to heterochromatin, in this study we were unable to quantify the change of PRC2 occupancy at heterochromatin and further determined the extent of PRC2 re-distribution to heterochromatin in naive mESCs. Mass spectrometry-assisted quantitative measurement of heterochromatin-bound PRC2 and H3K27me3 modification in naive mESCs should be conducted to address this question in future studies.

REFERENCES

REFERENCES

1. Boyer L.A., Plath K., Zeitlinger J., et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*. 2006;441:349–353.
2. Lee T.I., Jenner R.G., Boyer L.A., et al. Control of developmental regulator's by polycomb in human embryonic stem cells. *Cell*. 2006;125:301–313.
3. Aloia L., Di Stefano B., Di Croce L. Polycomb complexes in stem cells and embryonic development. *Development*. 2013;140:2525–2534.
4. Cao R., Wang L., Wang H., et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*. 2002;298:1039–1043
5. Wang H., Wang L., Erdjument-Bromage H., et al. Role of histone H2A ubiquitination in Polycomb silencing. *Nature*. 2004;431:873–878.
6. Di Croce L., Helin K. Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* 2013;20:1147–1155.
7. Blackledge N.P., Fursova N.A., Kelley J.R., et al. PRC1 catalytic activity is central to polycomb system function. *Mol. Cell*. 2019;77:857–874.e9
8. Tamburri S., Lavarone E., Fernandez-Perez D., et al. Histone H2AK119 mono-ubiquitination is essential for polycomb-mediated transcriptional repression. *Mol. Cell*. 2019;77:840–856.e5.
9. Laugesen A., Hojfeldt J.W., Helin K. Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. *Mol. Cell*. 2019;74:8–18.
10. Shen X., Kim W., Fujiwara Y., et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell*. 2009;139:1303–1314.
11. Landeira D., Sauer S., Poot R., et al. Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* 2010;12:618–624.
12. Pasini D., Cloos P.A., Walfridsson J., et al. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature*. 2010;464:306–310.
13. Casanova M., Preissner T., Cerase A., et al. Polycomblake 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. *Development*. 2011;138:1471–1482.

14. Li H., Liefke R., Jiang J., et al. Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature*. 2017;549:287–291.
15. Oksuz O., Narendra V., Lee C.H., et al. Capturing the onset of PRC2-mediated repressive domain formation. *Mol. Cell*. 2018;70:1149–1162.e5.
16. Farcas A.M., Blackledge N.P., Sudbery I., et al. KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *Elife*. 2012;1:e00205.
17. He J., Shen L., Wan M., et al. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat. Cell Biol*. 2013;15:373–384.
18. Wu X., Johansen J.V., Helin K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol. Cell*. 2013;49:1134–1146.
19. Cooper S., Dienstbier M., Hassan R., et al. Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep*. 2014;7:1456–1470.
20. Smith A.G. Embryo-derived stem cells: of mice and men. *Annu. Rev. Cell Dev. Biol*. 2001;17:435–462.
21. Silva J., Barrandon O., Nichols J., et al. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol*. 2008;6:e253
22. Ying Q.L., Wray J., Nichols J., et al. The ground state of embryonic stem cell self-renewal. *Nature*. 2008;453:519–523.
23. Takahashi S., Kobayashi S., Hiratani I. Epigenetic differences between naive and primed pluripotent stem cells. *Cell Mol. Life Sci*. 2018;75:1191–1203.
24. Leitch H.G., McEwen K.R., Turp A., et al. Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol*. 2013;20:311–316.
25. von Meyenn F., Iurlaro M., Habibi E., et al. Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell*. 2016;62:983.
26. Marks H., Kalkan T., Menafrá R., et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*. 2012;149:590–604.
27. Galonska C., Ziller M.J., Karnik R., et al. Ground state conditions induce rapid reorganization of core pluripotency factor binding before global epigenetic reprogramming. *Cell Stem Cell*. 2015;17:462–470.

28. Guo G., Pinello L., Han X., et al. Serum-based culture conditions provoke gene expression variability in mouse embryonic stem cells as revealed by single-cell analysis. *Cell Rep.* 2016;14:956–965.
29. Walter M., Teissandier A., Perez-Palacios R., et al. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife.* 2016;5
30. Kumar B., Elsasser S.J. Quantitative multiplexed ChIP reveals global alterations that shape promoter bivalency in ground state embryonic stem cells. *Cell Rep.* 2019;28:3274–3284.e5.
31. van Mierlo G., Dirks R.A.M., De Clerck L., et al. Integrative proteomic profiling reveals PRC2-dependent epigenetic crosstalk maintains ground-state pluripotency. *Cell Stem Cell.* 2019;24:123–137.e8.
32. Guo G., Yang J., Nichols J., Hall J.S., et al. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development.* 2009;136:1063–1069.
33. Munoz Descalzo S., Rue P., Garcia-Ojalvo J., et al. Correlations between the levels of Oct4 and Nanog as a signature for naive pluripotency in mouse embryonic stem cells. *Stem Cells.* 2012;30:2683–2691.
34. Yamaji M., Ueda J., Hayashi K., et al. PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell.* 2013;12:368–382.
35. Tee W.W., Shen S.S., Oksuz O., et al. Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs. *Cell.* 2014;156:678–690.
36. Peng J.C., Valouev A., Swigut T., et al. Sidow A., Wysocka J. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell.* 2009;139:1290–1302.
37. Roose J., Huls G., van Beest M., et al. Synergy between tumor suppressor APC and the beta-catenin-Tcf4 target Tcf1. *Science.* 1999;285:1923–1926.
38. Lickert H., Domon C., Huls G., et al. Wnt/(beta)-catenin signaling regulates the expression of the homeobox gene Cdx1 in embryonic intestine. *Development.* 2000;127:3805–3813.
39. Yan D., Wiesmann M., Rohan M., et al. Elevated expression of axin2 and hnk4 mRNA provides evidence that Wnt/beta -catenin signaling is activated in human colon tumors. *Proc. Natl. Acad. Sci. U S A.* 2001;98:14973–14978.

40. Lustig B., Jerchow B., Sachs M., et al. Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol. Cell Biol.* 2002;22:1184–1193.
41. ten Berge D., Koole W., Fuerer C., et al. Wnt signaling mediates self-organization and axis formation in embryoid bodies. *Cell Stem Cell.* 2008;3:508–518.
42. Horvay K., Casagrande F., Gany A., et al. Wnt signaling regulates Snail expression and cellular localization in the mouse intestinal epithelial stem cell niche. *Stem Cells Dev.* 2011;20:737–745
43. Deaton A.M., Bird A. CpG islands and the regulation of transcription. *Genes Dev.* 2011;25:1010–1022.
44. Al-Raawi D., Jones R., Wijesinghe S., et al. A novel form of JARID2 is required for differentiation in lineage-committed cells. *EMBO J.* 2019;38:e98449.
45. Blackledge N.P., Zhou J.C., Tolstorukov M.Y., et al. CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell.* 2010;38:179–190.
46. Long H.K., Blackledge N.P., Klose R.J. ZF-CXXC domain-containing proteins, CpG islands and the chromatin connection. *Biochem. Soc. Trans.* 2013;41:727–740.
47. Denissov S., Hofemeister H., Marks H., et al. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development.* 2014;141:526–537.
48. Illingworth R.S., Holzespies J.J., Roske F.V., et al. Polycomb enables primitive endoderm lineage priming in embryonic stem cells. *Elife.* 2016;5:e14926.
49. Laugesen A., Hojfeldt J.W., Helin K. Role of the polycomb repressive complex 2 (PRC2) in transcriptional regulation and cancer. *Cold Spring Harb. Perspect. Med.* 2016;6:a026575.
50. Riising E.M., Comet I., Leblanc B., et al. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol. Cell.* 2014;55:347–360.
51. Chazaud C., Yamanaka Y., Pawson T., et al. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev. Cell.* 2006;10:615–624.
52. Hamilton W.B., Mosesson Y., Monteiro R.S., et al. Dynamic lineage priming is driven via direct enhancer regulation by ERK. *Nature.* 2019;575:355–360.

CHAPTER 3:

HISTONE H3K36ME2-SPECIFIC METHYLTRANSFERASE ASH1L PROMOTES MLL-AF9-INDUCED LEUKEMOGENESIS

Published: Frontiers in Oncology. 2021; 11: 754093

Title: Histone H3K36me2-specific methyltransferase ASH1L promotes MLL-AF9-induced leukemogenesis

Authors: Mohammad B Aljazi, Yuen Gao, Yan Wu, George I Mias, Jin He

ABSTRACT

ASH1L and MLL1 are two histone methyltransferases that facilitate transcriptional activation during normal development. However, the roles of ASH1L and its enzymatic activity in the development of MLL-rearranged leukemias are not fully elucidated in *Ash1L* gene knockout animal models. In this study, we used an *Ash1L* conditional knockout mouse model to show that loss of ASH1L in hematopoietic progenitor cells impaired the initiation of MLL-AF9-induced leukemic transformation *in vitro*. Furthermore, genetic deletion of ASH1L in the MLL-AF9-transformed cells impaired the maintenance of leukemic cells *in vitro* and largely blocked the leukemia progression *in vivo*. Importantly, the loss of ASH1L function in the *Ash1L*-deleted cells could be rescued by wild-type but not the catalytic-dead mutant ASH1L, suggesting the enzymatic activity of ASH1L was required for its function in promoting MLL-AF9-induced leukemic transformation. At the molecular level, ASH1L enhanced the MLL-AF9 target gene expression by directly binding to the gene promoters and modifying the local histone H3K36me2 levels. Thus, our study revealed the critical functions of ASH1L in promoting the MLL-AF9-induced leukemogenesis, which provides a molecular basis for targeting ASH1L and its enzymatic activity to treat MLL-AF9-induced leukemias.

INTRODUCTION

The MLL rearrangement (MLLr) caused by 11q23 chromosomal translocations creates a variety MLL fusion proteins that drive the acute lymphoblastic and myeloid leukemia development, which accounts for approximate 5-10% acute leukemias in human patients¹⁻⁵. Despite recent progression in the development of chemotherapies against leukemias, the overall prognosis for the MLLr leukemias remains poor^{6,7}.

MLL1 protein is a histone lysine methyltransferase (KMTase) that contains a SET (Su(var)3-9, Enhancer-of-zeste and Trithorax) domain to catalyze trimethylation of histone H3 lysine 4 (H3K4me3)⁸. Functionally, MLL1 belongs to the Trithorax-group (TrxG) proteins that antagonize the Polycomb-group (PcG)-mediated gene silencing and facilitate transcriptional activation⁹. In 11q23 chromosomal translocations, the *N*-terminal portion of MLL1 is fused with a variety of fusion partners to generate different oncogenic MLL fusion proteins that function as disease drivers leading to leukemia development¹⁰⁻¹². Previous studies have revealed that the *N*-terminal portion of MLL fusion proteins interacts with MENIN and LEDGF (Lens Epithelium-Derived Growth Factor), which is critical for the recruitment of MLL fusion proteins to chromatin, whereas the *C*-terminal fusion partners interact with various trans-activators to induce transcriptional activation¹³⁻¹⁷. However, since the MLL fusion proteins lack the intrinsic histone H3K4 methyltransferase activity due to loss of the SET domain located in the *C*-terminal portion of MLL1¹⁰, it is unclear whether other histone modifications are required for the MLL fusion proteins-induced gene expression and leukemogenesis.

Recently, another member of TrxG proteins, ASH1L (Absent, Small, or Homeotic-Like I), was found to play important roles in normal hematopoiesis and leukemogenesis^{8,18,19}. Biochemically, ASH1L is a histone KMTase that mediates dimethylation of histone H3 lysine 36

(H3K36me2)²⁰. Similar to MLL1, ASH1L facilitates gene expression through antagonizing PcG-mediated gene silencing⁸. Previous studies have shown that ASH1L and MLL1 co-occupies the same transcriptional regulatory regions, and loss of either ASH1L or MLL1 reduces the expression of common genes²¹⁻²³, suggesting ASH1L and MLL1 function synergistically to activate gene expression during normal development. However, the significance of ASH1L and its-mediated histone H3K36me2 in the MLLr-associated leukemogenesis has not been addressed in the *Ash1L* gene knockout animal models.

In this study, we used an *Ash1L* conditional knockout mouse model to show that loss of ASH1L in hematopoietic progenitor cells (HPCs) impaired the initiation of MLL-AF9-induced leukemic transformation *in vitro*. Furthermore, genetic deletion of ASH1L in the MLL-AF9-transformed cells impaired the maintenance of leukemic cells *in vitro* and largely blocked the leukemia progression *in vivo*. Importantly, the loss of ASH1L function in the *Ash1L*-deleted cells could be rescued by ectopic expression of wild-type but not the catalytic-dead mutant ASH1L, suggesting the enzymatic activity of ASH1L was required for its function in promoting MLL-AF9-induced leukemic transformation. At the molecular level, ASH1L activated the MLL-AF9 target gene expression by directly binding to the gene promoters and modifying the local histone H3K36me2 levels. Thus, our study revealed the critical functions of ASH1L in MLL-AF9-induced leukemogenesis and raised the possibility that ASH1L might serve as a potential therapeutic target for the treatment of MLL-AF9-induced leukemias.

MATERIALS METHODS

Mice

The *Ash1L* conditional knockout mice were generated as previously reported²⁴. To generate inducible *Ash1L* deletion, mice were crossed with Rosa26-CreER^{T2} mice that were obtained from The Jackson Laboratory. All mice for this study were backcrossed to C57BL/6 mice for at least five generations to reach pure genetic background prior to conducting experiments. All mouse experiments were performed with the approval of the Michigan State University Institutional Animal Care & Use Committee.

Hematopoietic progenitor isolation and culture

Hematopoietic progenitor cells were isolated from femurs of 4- to 6-week C57BL/6 mice. The red blood cells in the bone marrows were lysed by ammonium chloride solution (Stem Cell Technologies 07800) and filtered with a 70- μ m nylon filter. The c-KIT⁺ HPCs were isolated using c-KIT antibody-conjugated IMag (BD Biosciences) beads. HPCs cells were maintained in RPMI1640 medium supplemented with 10% FBS, 1% MEM non-essential amino acids, 1% Glutamax, 10 ng/mL, 2-mercaptoethanol, and 50 ng/mL mSCF (PeproTech), 10 ng/mL mIL-6 (PeproTech), and 10 ng/mL mIL-3 (PeproTech). To induce CRE-mediated recombination *in vitro*, 4-hydroxy-tamoxifen (Sigma-Aldrich) was resuspended in DMSO and supplemented into the culture medium with concentration of 250 nM.

Retroviral and lentiviral vector production and transduction

The pMIG-FLAG-MLL-AF9 retroviral vectors as obtained from Addgene (Plasmid #71443). Retroviral vectors were generated by co-transfection of retroviral vectors with pGag-pol, pVSVG 293T cells using CalPhos mammalian transfection kit (TaKaRa). After 48hrs post transfection, viral supernatant was harvested, filtered through a 0.45 μ m membrane, and concentrated by ultracentrifugation. The lentiviral system was obtained from the National Institutes of Health AIDS Research and Reference Reagent Program. To generate GFP expression vectors, the GFP cDNA was PCR amplified, fused with P2A and puromycin resistant cassette and cloned into the SpeI/EcoRI sites under the EF1 α promoter. To generate lentiviral viruses, the transducing vectors pTY, pHP and pHEF1 α -VSVG were co-transfected into HEK293T cells. The supernatant was collected at 24, 36 and 48 hours after transfection, filtered through a 0.45 μ m membrane and concentrated by ultracentrifugation. Retroviral and lentiviral transduction of HPCs was performed by spin inoculation for 1 hour at 800g, in RPMI1640 medium supplemented with 10% FBS, 1x MEM non-essential amino acids (Life Technologies), 1x Glutamax (Life Technologies), 1x sodium pyruvate (Life Technologies), and 10 ng/mL mIL-3 (PeproTech).

Serial methylcellulose replating assay and leukemia transplantation

The colony formation assays were conducted by plating 500 cells into methylcellulose media consisting of Iscove MDM (Life Technologies) supplemented with FBS, BSA, insulin-transferrin (Life Technologies), 2-mercaptoethanol, 50 ng/mL mSCF (PeproTech), 10 ng/mL mIL-6 (PeproTech), 10 ng/mL mIL-3 (PeproTech), and 10 ng/mL GM-CSF (PeproTech). After 7-10 days, the colony numbers were counted under a microscope. The colonies were picked up, and cells were pooled and replated onto secondary methylcellulose plates. Three rounds of replating

were performed for each experiment. For leukemia transplantation, recipient C57BL/6 mice were subjected to total body irradiation at a dose of 11 Gy with the use of a X-RAD 320 biological irradiator. Donor cells (5×10^5) and radiation protector cells (5×10^5) isolated from BM were mixed in $1 \times$ PBS and transplanted into the recipient mice through retro-orbital injection. The mice were fed with water supplemented with trimethoprim/sulfamethoxazole for 4 weeks after transplantation.

FACS analysis

For FACS analysis, cells were stained with antibodies in staining buffer ($1 \times$ PBS, 2% FBS) and incubated at 4°C for 30 minutes. The samples were washed once with staining buffer before subjected to FACS analysis with the use of a BD LSRII. The antibodies used in this study include anti-Mac-1(eBioscience), anti-Gr-1(eBioscience), anti-c-KIT (eBioscience).

Western Blot analysis

Total proteins were extracted by RIPA buffer and separated by electrophoresis by 8-10% PAGE gel. The protein was transferred to the nitrocellulose membrane and blotted with primary antibodies. The antibodies used for Western Blot and IP-Western Blot analyses included: rabbit anti-ASH1L (1:1000, in house)²⁴ and IRDye 680 donkey anti-rabbit second antibody (1: 10000, Li-Cor). The images were developed by Odyssey Li-Cor Imager (Li-Cor).

Quantitative RT-PCR and ChIP-qPCR assays

RNA was extracted and purified from cells with the use of Qiashredder (QIAGEN) and RNeasy (QIAGEN) spin columns. Total RNA (1 µg) was subjected to reverse transcription using Iscript reverse transcription supermix (Bio-Rad). cDNA levels were assayed by real-time PCR using iTaq universal SYBR green supermix (Bio-Rad) and detected by CFX386 Touch Real-Time PCR detection system (Bio-Rad). Primer sequences for qPCR are listed in Table 6. The expression of individual genes is normalized to expression level of *Gapdh*. ChIP assays that used rabbit anti-ASH1L antibody (in house), rabbit anti-H3K36me2 antibody (Abcam), rabbit anti-Flag antibody (Cell Signaling) were carried out according to the previously reported protocol with the following modifications²⁵: ~2 µg antibodies were used in the immunoprecipitation, and chromatin-bound beads were washed 3 times each with TSEI, TSEII, and TSEIII followed by 2 washes in 10mM Tris, pH 7.5, 1mM EDTA. Histone modification ChIPs were carried out as previously reported²⁶. DNA that underwent ChIP was analyzed by quantitative PCR (qPCR), and data are presented as the percentage of input as determined with CFX manager 3.1 software. The amplicons were designed to locate at 1.0-kb upstream of transcriptional starting sites (TSS) and transcription ending sites (TES) of *Hoxa9/Hoxa10* genes. The mouse intracisternal A-particle LTR repeat elements were included as a negative control for the ASH1L binding. The ChIP primers for the mouse IAP LTR were purchased from Cell Signaling (85916, Cell Signaling). Other qPCR and ChIP primers are listed in Tables 6, respectively.

Table 6: Primes used in ASH1L study

Name	Sequence (5'-3')	Purpose
mHoxa9_F1	CGC CGG CAA CTT ATT AGG TG	ChIP
mHoxa9_R1	CCG ACC CGC CGA AAT TAT GA	ChIP
mHoxa9_F2	CTA CGC TCC AGG GAC CCT	ChIP
mHoxa9_R2	CTG GAA GCT GCA AGG ACT GA	ChIP
mHoxa9_F3	GTA TAT GCG CTC CTG GCT GG	ChIP
mHoxa9_R3	GCG GTT CAG GTT TAA TGC CA	ChIP
mHoxa9_F4	AGG ACC GAG CAA AAG ACG AG	ChIP
mHoxa9_R4	GGG AGA GGA GAC AGA GGG AG	ChIP
mHoxa10_F1	CTG CTA CAG GGC CCG TTT AA	ChIP
mHoxa10_R1	TTA TGA TGT GCA CCC CAG CC	ChIP
mHoxa10_F2	GCG TCT TCT GGC CCA TCA AT	ChIP
mHoxa10_R2	CGA CCA CTC CCA GTT TGG TT	ChIP
mHoxa10_F3	TCT GCT CCC TTC GCC AAA TT	ChIP
mHoxa10_R3	CCT CTG CCT GAG CTG ATG AG	ChIP
mHoxa10_F4	GGA GTG CTG GGC TGT GTT TA	ChIP
mHoxa10_R4	CTC TGC TCT TGG CCA AGG AA	ChIP
mAsh1_1F	CCCACACAAATGTAAGTTTGGA	Genotype
mAsh1_1R	ACATGGAGTTATTAGATCCTG	Genotype
mAsh1_2F	AGC CTG ACT GGC CTA GAA TG	Genotype
mAsh1_2R	TCC AAA CTG TAG AGC AGA AAA CA	Genotype

RNA-seq sample preparation for HiSeq4000 sequencing

RNA was extracted and purified from cells using QI shredder (Qiagen) and RNeasy (Qiagen) spin columns. Total RNA (1 µg) was used to generate RNA-seq library using NEBNext Ultra Directional RNA library Prep Kit for Illumina (New England BioLabs, Inc) according to the manufacturer's instructions. Adapter-ligated cDNA was amplified by PCR and followed by size selection using agarose gel electrophoresis. The DNA was purified using Qiaquick gel extraction kit (Qiagen) and quantified both with an Agilent Bioanalyzer and Invitrogen Qubit. The libraries were diluted to a working concentration of 10nM prior to sequencing. Sequencing on an Illumina HiSeq4000 instrument was carried out by the Genomics Core Facility at Michigan State University.

RNA-seq data analysis

RNA-Seq data analysis was performed essentially as described previously. All sequencing reads were mapped mm9 of the mouse genome using Tophat2²⁷. The mapped reads were normalized to reads as Reads Per Kilobase of transcript per Million mapped reads (RPKM). The differential gene expression was calculated by Cuffdiff program and the statistic cutoff for identification of differential gene expression is $p < 0.01$ and 1.5-fold RPKM change between samples²⁸. The heatmap and plot of gene expression were generated using plotHeatmap and plotProfile in the deepTools program²⁹. The differential expressed gene lists were input into the David Functional Annotation Bioinformatics Microarray Analysis for the GO enrichment analyses (<https://david.ncifcrf.gov/>).

Statistical analysis

All statistical analyses were performed using GraphPad Prism 9 (GraphPad Software). Parametric data were analyzed by a two-tailed *t* test or two-way ANOVA test for comparisons of multiple samples. The post-transplantation survivals were analyzed by the Gehan-Breslow-Wilcoxon test. *P* values < 0.05 were considered statistically significant. Data are presented as mean \pm SEM.

RESULTS

ASH1L promotes the initiation of MLL-AF9-induced leukemic transformation *in vitro*

To examine the function of ASH1L in MLL-AF9-induced leukemogenesis, we generated an *Ash1L* conditional knockout (*Ash1L*-cKO) mouse line in which two *LoxP* elements inserted into the exon 4 flanking regions²⁴. A CRE recombinase-mediated deletion of exon 4 resulted in altered splicing of mRNA that created a premature stop codon before the sequences encoding the first functional AWS (Associated With SET) domain (Figure 20A, B). The *Ash1L*-cKO mice were further crossed with the Rosa26-CreER^{T2} mice to generate a tamoxifen-inducible *Ash1L* knockout line (*Ash1L*^{2f/2f};Rosa26-CreER^{T2}), which allowed us to study the function of ASH1L in leukemogenesis *in vitro* and *in vivo*.

Using this *Ash1L*-cKO mouse model, we investigate the role of ASH1L in the initiation of MLL-AF9-induced leukemic transformation. To this end, we isolated the bone marrow cells from wild-type (*Ash1L*^{+/+};Cre-ER^{T2}) and *Ash1L*-cKO (*Ash1L*^{2f/2f};Cre-ER^{T2}) mice, respectively. The c-KIT⁺ HPCs were further enriched by the c-KIT antibody-conjugated magnetic beads (Figure 20C). The HPCs were cultured in the HPC medium supplemented with murine IL-3, IL-6, and SCF for three days, and followed by transduction of retroviral vectors expressing a *MLL1-AF9* fusion gene or control empty viruses (EV). After transduction, the cells were cultured in the suspension

medium with 4-hydroxytamoxifen (4-OHT) for five days to induce *Ash1L* gene deletion in the *Ash1L*-cKO HPCs (Figure 20D). The quantitative RT-PCR (qRT-PCR) analysis showed that the *Ash1L* expression reduced to less than 5% at the mRNA level in the *Ash1L*-deleted cells (Figure 20E). To investigate the effect of *Ash1L* loss on the initiation of MLL-AF9-induced leukemic transformation *in vitro*, we performed serial colony replating assays by plating the cells on the semi-solid methylcellulose medium to examine the leukemic transformation. The results showed that although the cells transduced with *MLL-AF9* or empty vectors had comparable colony numbers in the first round of plating, the cells transduced with control empty vectors did not form colonies in the following rounds of replating. In contrast, both wild-type and *Ash1L*-cKO HPCs transduced with *MLL-AF9* retroviruses formed colonies in all three rounds of plating, indicating successful leukemic transformation by the *MLL-AF9* transgene *in vitro*. Notably, compared to the MLL-AF9-transduced wild-type cells, the *Ash1L*-deleted cells had reduced colony numbers in the second and third rounds of plating, suggesting that loss of *Ash1L* in HPCs compromised the MLL-AF9-induced leukemic transformation (Figure 20F, G), suggesting ASH1L promotes the MLL-AF9-induced leukemic transformation *in vitro*.

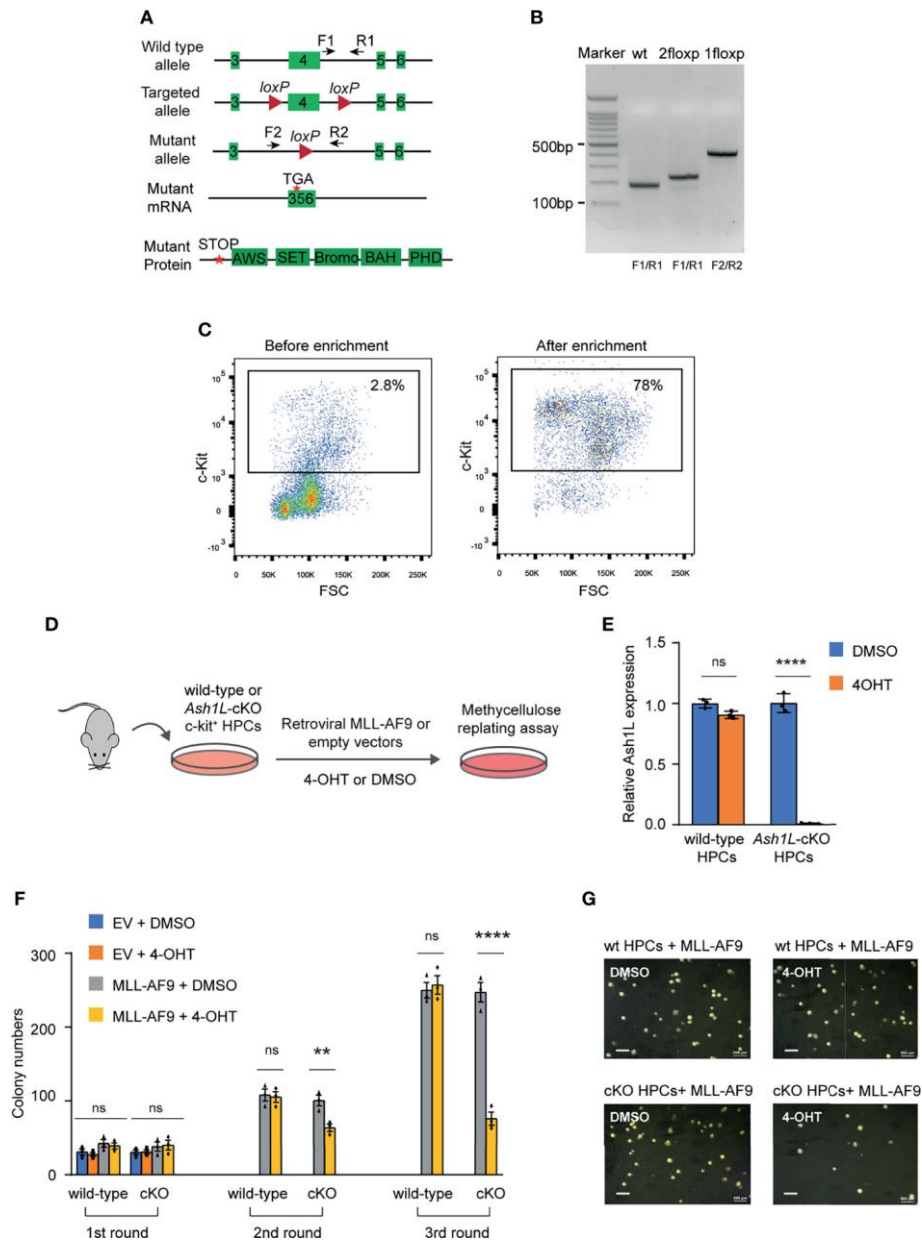


Figure 20: ASH1L is required for the initiation of MLL-AF9-induced leukemic transformation. (A) Diagram showing the strategy for the generation of *Ash1L* conditional knockout mice. CRE-mediated deletion of exon 4 results in an altered spliced mRNA with a premature stop codon, which generates a truncated protein without all functional AWS, SET, Bromo, BAH and PHD domains. The arrows labeled as F and R represent the genotyping primers. (B) Genotyping results showing the PCR results of wild-type, 2 floxP, and 1 floxP alleles. (C) FACS analysis showing the c-KIT⁺ HPC populations before and after enrichment with c-KIT antibody-conjugated beads. (D) Schematic experimental procedure. (E) qRT-PCR analysis showing the *Ash1L* expression levels in wild-type and *Ash1L*-cKO cells after treated with 4-OHT or DMSO. The results were normalized against levels of *Gapdh* and the expression level in DMSO-treated cells was arbitrarily set to 1. The error bars represent mean \pm SEM, $n = 3$ per group. **** $P < 0.0001$, ns, not significant. (F) Methylcellulose replating assays showing the colony numbers for each round of plating. The error bars represent mean \pm SEM, $n = 3$ per group. ** $P < 0.01$; **** $P < 0.0001$, ns, not significant. (G) Photos showing the representative colony formation on methylcellulose plates for each group. Bar = 0.5 mm.

ASH1L facilitates the maintenance of MLL-AF9-induced leukemic cells *in vitro*

Next, we examined the functional role of *Ash1L* in maintaining the MLL-AF9-transformed cells. To this end, we transduced both wild-type and *Ash1L*-cKO HPCs with *MLL-AF9* retroviruses and plated the transduced cells onto the methylcellulose medium. After three rounds of replating, the transformed colonies were manually picked and cultured in the suspension medium supplemented with 4-OHT for 5 days to induce deletion of *Ash1L* in the *Ash1L*-cKO cells. The cells were further maintained in suspension culture without 4-OHT for 5 days before plated onto the methylcellulose to examine the colony formation (Figure 21A). The results showed that compared to the wild-type MLL-AF9-transformed cells, the *Ash1L*-deleted cells had marked reduced colony formation (Figure 21B, C), suggesting that ASH1L facilitated the maintenance of MLL-AF9 transformed cells *in vitro*.

To examine cellular responses to the *Ash1L* depletion, we performed the FACS analysis to examine cell death in response to the loss of *Ash1L* in the MLL-AF9-transformed cells. The results showed that compared to the wild-type cells, the *Ash1L*-deleted cells had increased populations of both early apoptotic cells (Annexin V+/DAPI-) and late dead cells (Annexin V+/DAPI+) (Figure 21D), suggesting that the loss of *Ash1L* induced cell death of MLL-AF9-transformed cells. Moreover, FACS analyses showed that compared to the wild-type transformed cells, the *Ash1L*-deleted cells had increased expression of myeloid differentiation surface markers CD11b and GR-1 (Figure 21E). Morphologically, the wild-type transformed cells displayed leukoblast-like morphology with enlarged dark stained nuclei, while the *Ash1L*-deleted cells had light-stained and segmented nuclei, a feature indicating the differentiation towards matured myeloid cells (Figure 21F). Taken together, these results suggested that ASH1L facilitated the maintenance of MLL-AF9-transformed cells through suppressing cell death and differentiation.

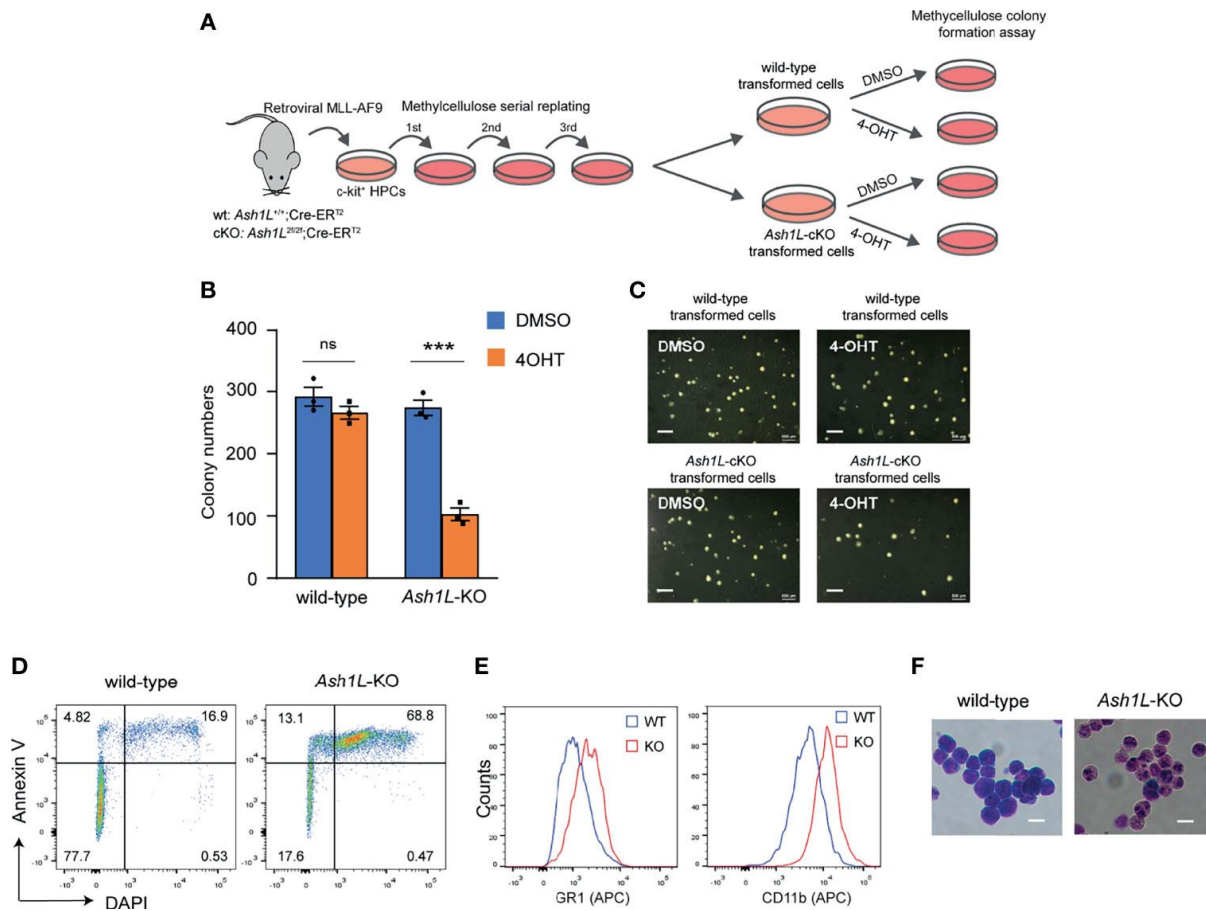


Figure 21: ASH1L is required for the maintenance of MLL-AF9-induced leukemic cells *in vitro*. (A) Schematic experimental procedure. (B) Methylcellulose colony formation assays showing the colony numbers. The error bars represent mean ± SEM, n = 3 per group. ****P* < 0.001; ns, not significant. (C) Photos showing the representative colony formation on methylcellulose plates for each group. Bar = 0.5 mm. (D) Representative FACS results showing the Annexin V⁺ and DAPI⁺ populations of wild-type and *Ash1L*-KO MLL-AF9-transformed cells. (E) Representative FACS results showing the GR-1 and CD11b expression of wild-type and *Ash1L*-KO MLL-AF9-transformed cells. (F) Photos showing the Wright-Giemsa staining of wild-type and *Ash1L*-KO MLL-AF9-transformed cells. Bar = 10 µm.

ASH1L promotes the MLL-AF9-induced leukemia development *in vivo*

To determine the role of ASH1L in the MLL-AF9-induced leukemogenesis *in vivo*, we performed leukemia transplantation assays and monitor the leukemia development in recipient mice. To this end, the wild-type and *Ash1L*-deleted MLL-AF9-transformed cells were labeled with GFP by transduction with lentiviral-GFP vectors, mixed with normal protective bone marrow cells, and transplanted into the total-body-irradiated (TBI) syngeneic recipient mice (Figure 22A). Four weeks after transplantation, FACS analysis showed that the mice transplanted with wild-type leukemic cells had higher GFP+ leukemic cell populations in the peripheral blood compared to the mice received with *Ash1L*-KO leukemic cells (Figure 22B), which was consistent with the higher leukemic cell numbers in the peripheral blood smears and splenomegaly found in the mice transplanted with wild-type leukemic cells (Figure 22C, D). All mice transplanted with wild-type leukemic cells died within 3 months after transplantation, and the median survival time was around 8.5 weeks. In contrast, the mice transplanted with *Ash1L*-deleted cells had significant longer survival time (Chi square = 10.73, df = 1, $p = 0.0011$) compared to the mice transplanted with wild-type leukemic cells (Figure 22E). These results suggested that ASH1L in the MLL-AF9-transformed leukemic cells promoted the development and progression of leukemia *in vivo*.

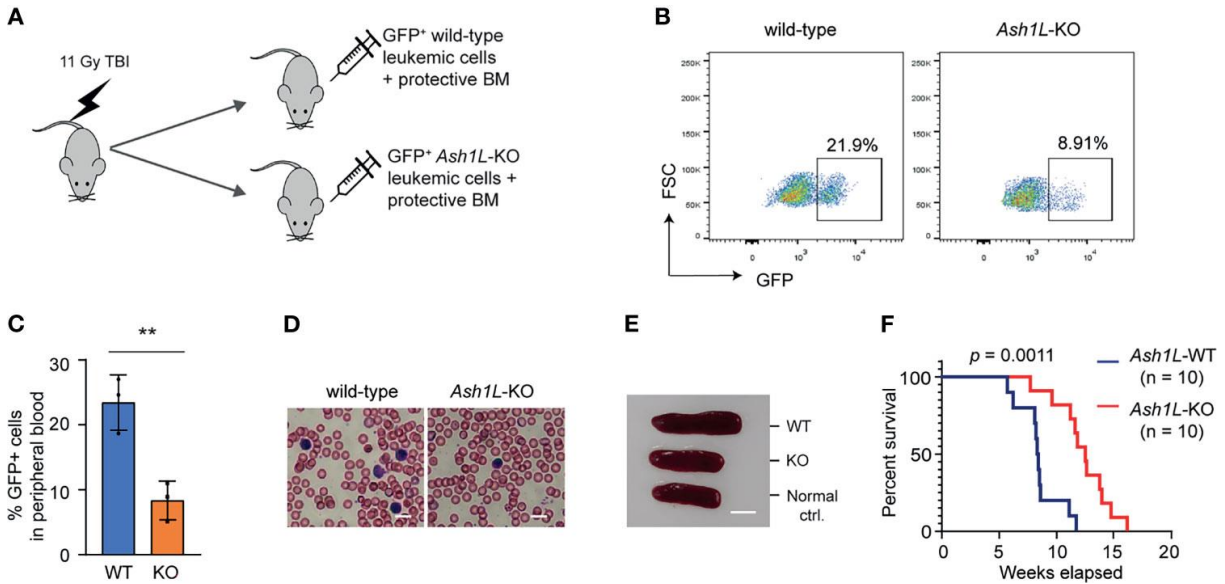


Figure 22: ASH1L promotes the MLL-AF9-induced leukemia development *in vivo*. (A) Schematic experimental procedure. (B) Representative FACS analysis showing the GFP⁺ leukemic cell populations in the peripheral blood of mice transplanted with wild-type or *Ash1L*-KO MLL-AF9-transformed cells. (C) Quantitative results showing the percentage of GFP⁺ leukemic cell populations in the peripheral blood of mice transplanted with wild-type or *Ash1L*-KO MLL-AF9-transformed cells. The error bars represent mean \pm SEM, $n = 3$ per group. $**P < 0.01$. (D) Photos showing the leukemic cells in the peripheral blood smear of mice transplanted with wild-type or *Ash1L*-KO MLL-AF9-transformed cells. Bar = 10 μ m. (E) Photos showing the representative spleen size from the normal control mice (Normal ctrl.), mice transplanted with wild-type (WT) or *Ash1L*-KO (KO) MLL-AF9-transformed cells. The samples were collected at post-transplantation 4 weeks. Bar = 5 mm. (F) Kaplan-Meier survival curve of mice transplanted with wild-type or *Ash1L*-KO MLL-AF9-transformed cells. P value calculated using a Gehan-Breslow-Wilcoxon test. $n = 10$ mice per group.

The enzymatic activity of ASH1L is required for its function in promoting MLL-AF9-induced leukemic transformation

Next, we set out to determine whether the histone methyltransferase activity of ASH1L was required for its function in promoting MLL-AF9-induced leukemic transformation. To this end, the *Ash1L*-cKO HPCs were infected with retroviruses expressing *MLL-AF9* transgene, followed by transduced with lentiviral vectors expressing either wild-type ASH1L or catalytic-dead mutant ASH1L(H2214A)²¹. The transformed cells were treated with 4-OHT to induce deletion of endogenous *Ash1L* gene (Figure 23A). Western blot analysis showed that both wild-type and mutant exogenous ASH1L had a similar expression level (Figure 23B). The cells were further plated onto the methylcellulose medium to examine the colony formation (Figure 23A). The results showed that compared to the wild-type ASH1L-expressed cells, the cells with ectopic expression of catalytic-dead mutant ASH1L had reduced colony formation (Figure 23C, D). Similar to the *Ash1L*-deleted cells, the *Ash1L*-deleted cells rescued with mutant ASH1L had increased cell death and upregulated expression of myeloid differentiation markers of CD11b and GR-1 (Figure 23E, F). These results suggested that ASH1L histone methyltransferase activity was required for its function in promoting MLL-AF9-induced leukemogenesis by inhibiting cell death and blocking myeloid differentiation.

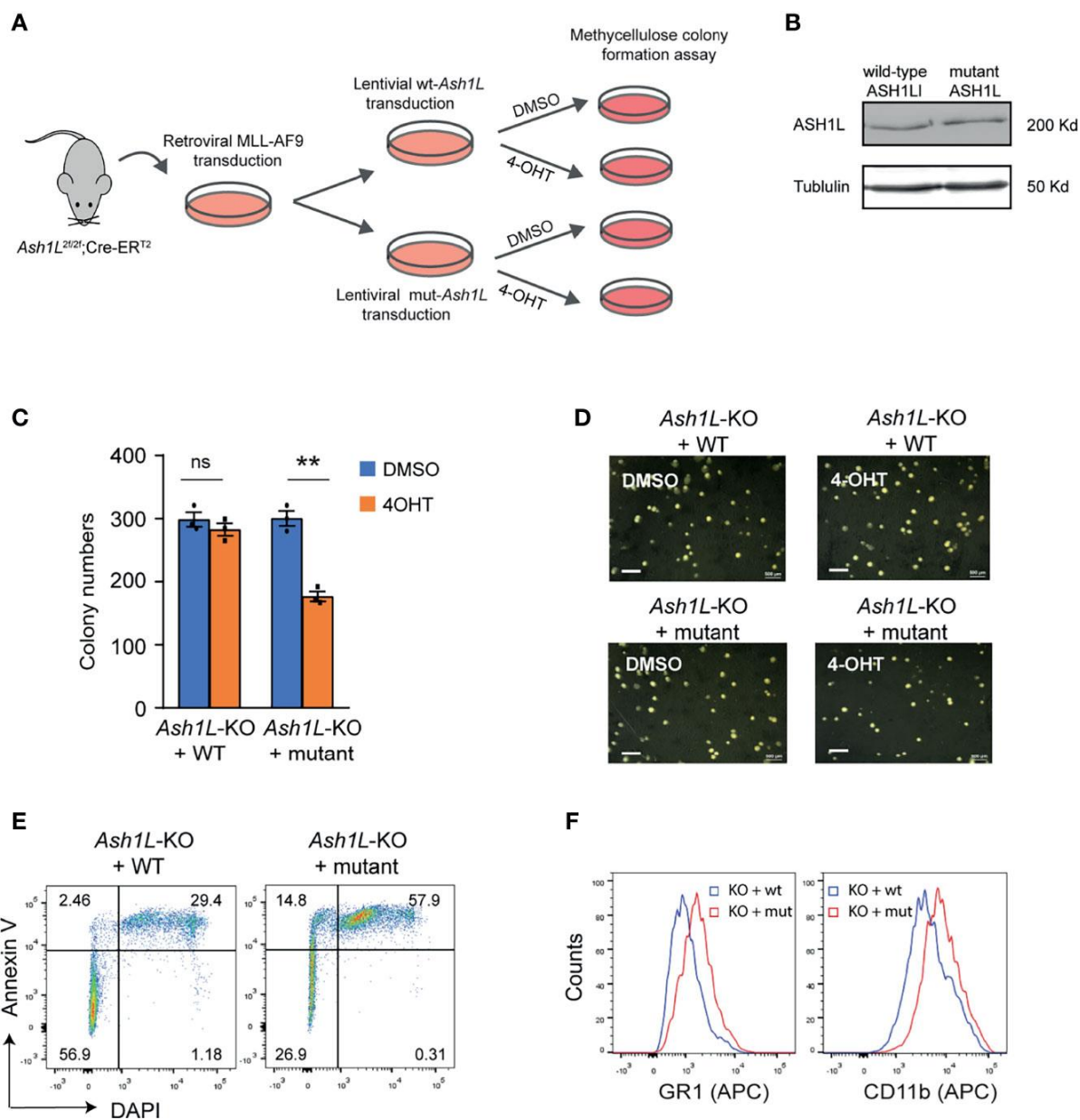


Figure 23: The enzymatic activity of ASH1L is required for its function in promoting MLL-AF9-induced leukemic transformation. (A) Schematic experimental procedure. (B) WB analysis showing the ectopic expression of wild-type and mutant ASH1L. (C) Methylcellulose colony formation assays showing the colony numbers. The error bars represent mean \pm SEM, $n = 3$ per group. $**P < 0.01$; ns, not significant. (D) Photos showing the representative colony formation on methylcellulose plates for each group. Bar = 0.5 mm. (E) Representative FACS results showing the Annexin V⁺ and DAPI⁺ populations of *Ash1L*-KO cells rescued with wild-type and mutant ASH1L. (F) Representative FACS results showing the GR1 and CD11b expression of *Ash1L*-KO cells rescued with wild-type and mutant ASH1L.

ASH1L facilitates the MLL-AF9-induced leukemogenic gene expression

To examine the molecular mechanisms underlying the function of ASH1L in promoting MLL-AF9-induced leukemogenesis, we performed RNA-seq analyses to examine the transcriptome changes in normal HPCs, wild-type and *Ash1L*-deleted MLL-AF9-transformed cells. The results showed that compared to normal HPCs, the MLL-AF9-transformed cells had 1,021 upregulated and 1,228 downregulated genes (cutoff: fold changes > 1.5, FDR < 0.05), respectively (Figure 24A). The gene ontology (GO) enrichment analysis showed that both upregulated and downregulated genes were involved in immune processes and inflammatory responses (cutoff: FDR < 0.05) (Tables 7 and 8), suggesting that MLL-AF9 fusion proteins disrupted the normal differentiation and mis-regulated the normal function of myeloid cells. Notably, multiple genes, such as *Hoxa5*, *Hoxa7*, *Hoxa9*, *Hoxa10* and *Meis1* that were known to mediate the MLL-AF9-induced leukemogenesis, were highly expressed in the MLL-AF9-transformed cells (Figure 24B). Further RNA-seq analysis showed that compared to MLL-AF9-transformed wild-type cells, the *Ash1L*-deleted cells had 372 upregulated gene and 472 downregulated genes (cutoff: fold changes > 1.5, FDR < 0.05), respectively (Figure 24C). Cross-examining these two data sets revealed that 105 genes, including *Hoxa5*, *Hoxa7*, *Hoxa9*, *Hoxa10*, and *Meis1* that were highly expressed in the wild-type MLL-AF9-transformed cells, were downregulated in the *Ash1L*-deleted cells (Figure 24D, E). Altogether, these results suggested that ASH1L promoted the MLL-AF9-induced leukemogenesis by facilitating the MLL-AF9-induced leukemic gene expression.

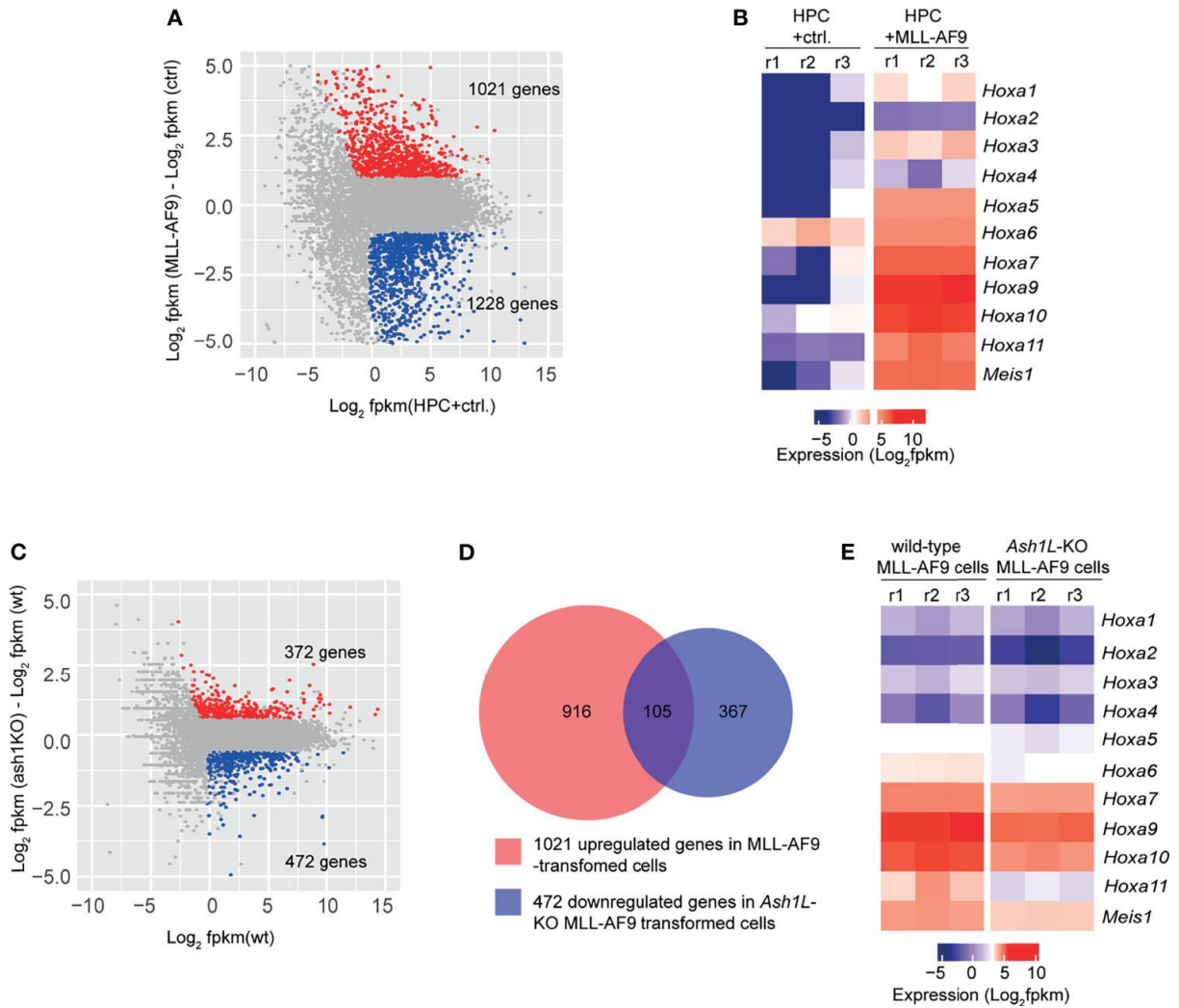


Figure 24: ASH1L facilitates the MLL-AF9-induced leukemogenic gene expression. (A) Plot showing 1021 up- and 1228 down-regulated genes in the MLL-AF9-transformed cells compared to the normal HPCs. (B) Heatmap showing the upregulation of *Hoxa* gene cluster and *Meis1* in the MLL-AF9-transformed cells compared to normal HPCs. (C) Plot showing 372 up- and 472 down-regulated genes in the *Ash1L*-KO MLL-AF9-transformed cells compared to the wild-type MLL-AF9-transformed cells. (D) Venn diagram showing the 105 genes upregulated in the MLL-AF9-transformed cells and downregulated in the *Ash1L*-KO cells. (E) Heatmap showing the *Hoxa* gene cluster and *Meis1* downregulated in the *Ash1L*-KO cells compared to the wild-type MLL-AF9-transformed cells.

Table 7: Result of gene ontology enrichment analysis of genes upregulated in the MLL-AF9-transformed cells.

Term	Count	P-Value	FDR
GO:0002376~immune system process	64	5.89E-19	1.94E-15
GO:0045087~innate immune response	57	9.25E-14	1.52E-10
GO:0006954~inflammatory response	46	2.60E-10	2.85E-07
GO:0006897~endocytosis	28	6.75E-08	5.55E-05
GO:0001974~blood vessel remodeling	12	2.76E-06	0.001812751
GO:0002224~toll-like receptor signaling pathway	8	5.57E-06	0.003049695
GO:0001774~microglial cell activation	8	8.74E-06	0.004107539
GO:0070269~pyroptosis	6	4.26E-05	0.017495751
GO:0002755~MyD88-dependent toll-like receptor signaling pathway	7	5.05E-05	0.018451926
GO:0050707~regulation of cytokine secretion	6	7.51E-05	0.024686387
GO:0043277~apoptotic cell clearance	7	1.08E-04	0.0318869
GO:0001525~angiogenesis	26	1.16E-04	0.0318869
GO:0016042~lipid catabolic process	16	1.40E-04	0.035497258
GO:0048704~embryonic skeletal system morphogenesis	11	1.81E-04	0.042510487

Table 8: Result of gene ontology enrichment analysis of genes down-regulated in the MLL-AF9-transformed cells.

Term	Count	P-Value	FDR
GO:0006955~immune response	52	5.56E-14	2.19E-10
GO:0006954~inflammatory response	50	4.13E-09	8.14E-06
GO:0000188~inactivation of MAPK activity	10	2.47E-08	3.25E-05
GO:0045766~positive regulation of angiogenesis	25	8.71E-08	6.78E-05
GO:0032496~response to lipopolysaccharide	33	9.88E-08	6.78E-05
GO:0007155~cell adhesion	59	1.03E-07	6.78E-05
GO:0002376~immune system process	50	1.32E-07	7.45E-05
GO:0006935~chemotaxis	23	9.09E-07	4.48E-04
GO:0030593~neutrophil chemotaxis	17	1.40E-06	6.12E-04
GO:0008285~negative regulation of cell proliferation	47	2.04E-06	8.04E-04
GO:0045785~positive regulation of cell adhesion	15	2.43E-06	8.72E-04
GO:0010628~positive regulation of gene expression	47	5.76E-06	0.001894
GO:0042102~positive regulation of T cell proliferation	15	1.30E-05	0.003949
GO:0050900~leukocyte migration	11	1.42E-05	0.004009
GO:0018108~peptidyl-tyrosine phosphorylation	15	1.57E-05	0.004133
GO:0006915~apoptotic process	59	1.82E-05	0.004475
GO:0001525~angiogenesis	32	1.99E-05	0.004609
GO:0070374~positive regulation of ERK1 and ERK2 cascade	27	3.01E-05	0.006592
GO:0042127~regulation of cell proliferation	30	4.82E-05	0.010003
GO:0050731~positive regulation of peptidyl-tyrosine phosphorylation	18	6.08E-05	0.011981
GO:0042493~response to drug	39	6.78E-05	0.012734
GO:0000122~negative regulation of transcription from RNA polymerase II	68	1.04E-04	0.017976
GO:0097192~extrinsic apoptotic signaling pathway in absence of ligand	11	1.05E-04	0.017976
GO:0043065~positive regulation of apoptotic process	38	1.14E-04	0.018508
GO:0007204~positive regulation of cytosolic calcium ion concentration	22	1.17E-04	0.018508
GO:0002687~positive regulation of leukocyte migration	8	1.64E-04	0.023961
GO:0043306~positive regulation of mast cell degranulation	7	1.75E-04	0.024678
GO:0007169~transmembrane receptor protein tyrosine kinase signaling	17	1.83E-04	0.024914
GO:0009612~response to mechanical stimulus	13	2.19E-04	0.028781
GO:0014068~positive regulation of phosphatidylinositol 3-kinase signaling	13	2.56E-04	0.032509
GO:0006334~nucleosome assembly	17	2.91E-04	0.03588
GO:0048711~positive regulation of astrocyte differentiation	6	3.55E-04	0.042358
GO:0001569~patterning of blood vessels	10	3.66E-04	0.04242
GO:0001938~positive regulation of endothelial cell proliferation	13	3.98E-04	0.044845
GO:0050728~negative regulation of inflammatory response	15	4.28E-04	0.046884
GO:0038083~peptidyl-tyrosine autophosphorylation	10	4.45E-04	0.04742

ASH1L binds and mediates the histone H3K36me2 modification at *Hoxa9* and *Hoxa10* gene promoters

To determine whether ASH1L directly regulated the expression of MLL-AF9 target genes, we performed chromatin immunoprecipitation (ChIP) coupled with quantitative PCR (ChIP-qPCR) assays to examine the ASH1L occupancy, MLL-AF9 occupancy, and histone H3K36me2 modification at the gene promoters, transcriptional starting sites (TSS), transcriptional ending sites (TES) of *Hoxa9* and *Hoxa10*, two MLL-AF9 target genes that were shown to be activated in the wild-type transformed cells and have reduced expression in the *Ash1L*-deleted cells (Figure 24B, E). The results showed that both ASH1L occupancy and histone H3K36me2 were enriched at the *Hoxa9* and *Hoxa10* promoters compared to that on the TES and the long terminal repeat (LTR) of intracisternal A-particle (IAP) (Figure 25A-E). Furthermore, compared to wild-type MLL-AF9-transformed cells, both ASH1L occupancy and histone H3K36me2 modification were reduced at the gene promoters in the *Ash1L*-deleted cells (Figure 25A-E), suggesting that ASH1L bound to the *Hoxa9* and *Hoxa10* gene promoters directly and mediated local histone H3K36me2 modification. However, the MLL-AF9 occupancy at both gene promoters did not show significant difference between wild-type and *Ash1L*-deleted MLL-AF9-transformed cells (Figure 25F-G), suggesting the ASH1L-mediated histone H3K36me2 did not affect the binding of MLL-AF9 fusion protein to the gene promoters.

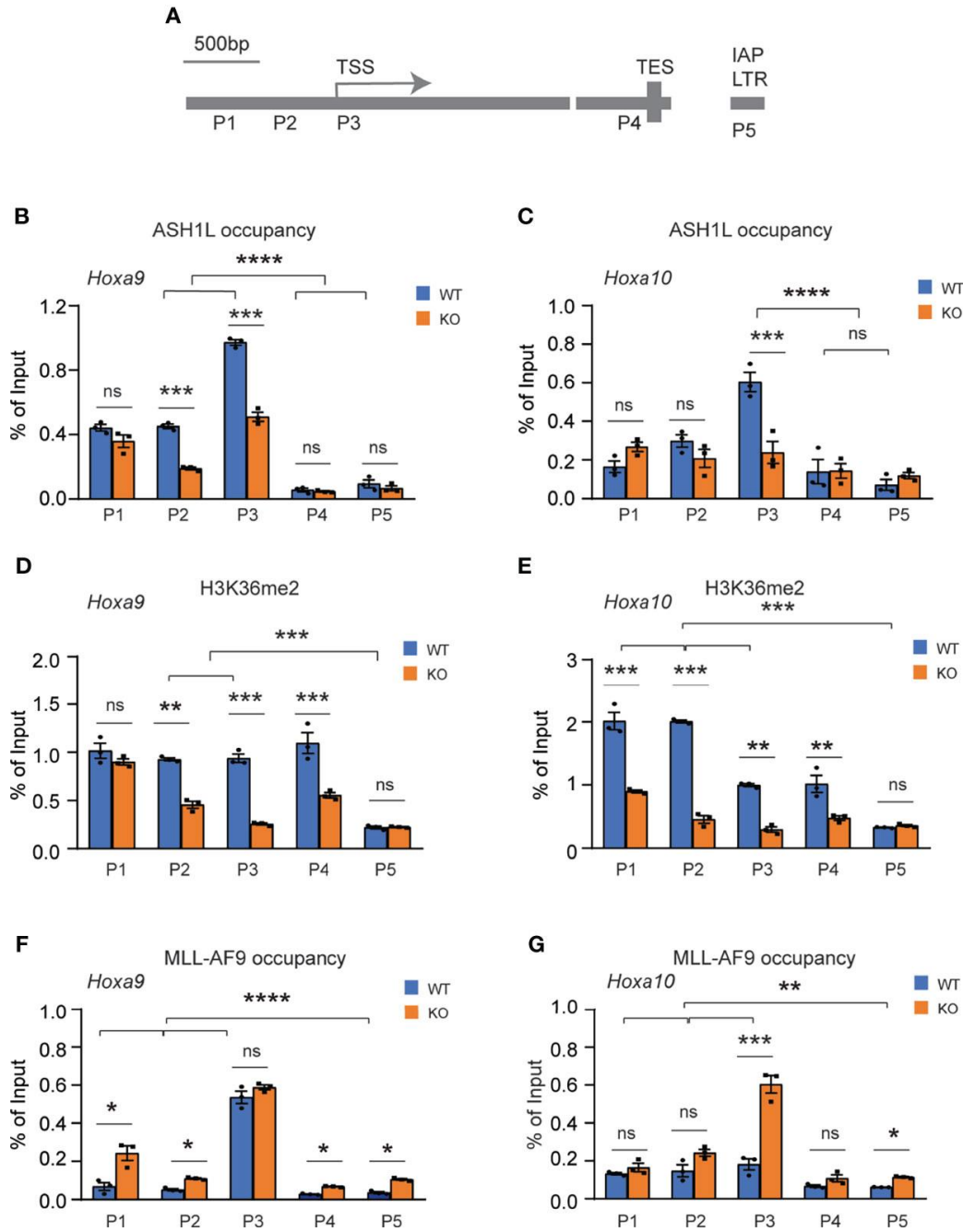


Figure 25: ASH1L binds and mediates histone H3K36me2 modification at *Hoxa9* and *Hoxa10* gene promoters. (A) Plot showing the locations of ChIP-qPCR amplicons at the *Hoxa9* and *Hoxa10* gene loci and LTR of intracisternal A-particle (IAP). (B-C) ChIP-qPCR analysis showing the ASH1L occupancy at *Hoxa9* and *Hoxa10* gene loci in the wild-type and *Ash1L*-KO MLL-AF9-transformed cells. (D-E) ChIP-qPCR analysis showing the histone H3K36me2 at *Hoxa9* and *Hoxa10* gene loci in the wild-type and *Ash1L*-KO MLL-AF9-transformed cells. (F-G) ChIP-qPCR analysis showing the MLL-AF9 occupancy at *Hoxa9* and *Hoxa10* gene loci in the wild-type and *Ash1L*-KO MLL-AF9-transformed cells. Note: for panels B-E, the error bars represent mean \pm SEM, $n = 3$ biological replicates. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; **** $P < 0.0001$; ns, not significant

DISCUSSION

Chromosomal 11q23 translocations generate various MLL fusion proteins that contain the *N*-terminal portion of MLL1 and different fusion partners including AF9^{30,31}. Previous studies have demonstrated that the *N*-terminal MLL1 is critical for the recruitment of MLL fusion proteins to chromatin through its CXXC-zinc finger (CXXC-zf) domain and its interacting proteins MENIN and LEDGF, while the *C*-terminal fusion partners interact with multiple trans-activators to induce transcriptional activation¹⁶. Since the MLL fusion proteins lose the MLL1 *C*-terminal SET domain and its-associated histone H3K4 methyltransferase activity, it is unclear whether other histone KMTase-mediated histone modifications are required for the MLL fusion proteins to activate leukemogenic gene expression and induce leukemia development.

ASH1L is another member of TrxG proteins that facilitate transcriptional activation⁸. Biochemically, ASH1L is a histone KMTase mediating histone H3K36me2 modification²⁰. Recent studies reported that ASH1L and MLL1 co-occupied the same gene promoters to activate gene expression, suggesting ASH1L and MLL function synergistically in activating gene expression in normal development and leukemogenesis^{19,21-23}. However, the functional roles of ASH1L and its-mediated histone H3K36me2 in the MLLr-associated leukemogenesis have not been addressed using *Ash1L* gene knockout animal models.

In this study, we used an *Ash1L* conditional knockout mouse model to show that ASH1L and its histone methyltransferase activity are required for promoting the MLL-AF9-induced leukemogenesis. First, genetic deletion of ASH1L in normal HPCs largely impairs the MLL-AF9-induced colony formation in serial methylcellulose replating assays (Figure 20), suggesting ASH1L promotes the initiation of MLL-AF9-induced leukemic transformation. Second, loss of ASH1L in the MLL-AF9-transformed cells largely impaired the colony formation *in vitro* and

delayed the leukemia development in the recipient mice transplanted with leukemic cells (Figures 21 and 22), suggesting ASH1L facilitates the maintenance of MLL-AF9-transformed cells *in vitro* and leukemia progression *in vivo*. Importantly, the impaired ASH1L's function in the *Ash1L*-KO cells could be rescued by the wild-type but not the catalytic-dead mutant ASH1L (Figure 4), suggesting that the histone methyltransferase activity is required for its function in promoting MLL-AF9-induced leukemogenic transformation, which is consistent with a recent study showing that the SET domain is required for the MLL-AF9-induced leukemic transformation³².

At the cellular level, we observed that the loss of ASH1L in MLL-AF9-transformed cells induced cell death and myeloid differentiation, which could be rescued by the wild-type but not the catalytic-dead mutant ASH1L (Figures 21, 23), suggesting that ASH1L promotes MLL-AF9-induced leukemic transformation through inhibiting cell apoptosis and blocking cell differentiation. The results are consistent with the molecular findings that ASH1L is required for the full activation of MLL-AF9 target genes including *Hoxa* gene cluster and *Meis1* (Figure 24), which are known to play important roles in leukemogenesis through inhibiting cell death and blocking normal cell differentiation³³⁻³⁵. Finally, the ChIP assays showed that both ASH1L occupancy and histone H3K36me2 modification were enriched at the promoters of MLL-AF9 target genes *Hoxa9* and *Hoxa10* in the wild-type transformed cells (Figure 25), indicating the ASH1L regulates the MLL-AF9 target genes through directly chromatin binding and its-mediated histone H3K36me2 modification.

Previous studies have shown that the PWWP domain of LEDGF is required for the recruitment of MLL fusion proteins through its binding to histone H3K36me2^{13,15,16}. However, our ChIP analysis did not reveal reduction of MLL-AF9 occupancy at the *Hoxa9* and *Hoxa10* promoters in the *Ash1L*-KO cells (Figure 25F, G), suggesting the MLL-AF9 fusion protein could

bind to its target regions through other recruiting mechanisms, such as the CXXC-zf domain-mediated binding to unmethylated CpG-rich promoters³⁶, and the reduced H3K36me2 at gene promoters in the *Ash1L*-KO cells impaired the *Hoxa* gene expression through mechanisms other than the recruitment of MLL-AF9 fusion protein.

Our current study has some limitations: (i) since this study includes a single type of MLLr, MLL-AF9 fusion protein, to induce leukemia development in mice, it is unclear whether ASH1L has the similar function in promoting other MLLr-induced leukemogenesis; (ii) although *Ash1L* deletion induces cell death, some MLL-AF9-transformed cells survive *in vitro* and *in vivo*, suggesting the MLL-AF9-transformed cells have heterogeneous responses to the *Ash1L* depletion. However, the underlying mechanisms are not addressed by our current study. These fundamental questions merit further investigation for a better understanding of the function of ASH1L in broad MLLr-associated leukemogenesis.

In summary, our study reveals that the histone H3K36me2-specific methyltransferase ASH1L and its enzymatic activity play an important role in promoting the MLL-AF9-induced leukemogenesis, which provides an important molecular basis for targeting ASH1L and its enzymatic activity to treat MLL-AF9-induced leukemias.

REFERENCES

REFERENCES

1. Berger, R., et al., Acute monocytic leukemia chromosome studies. *Leuk Res.* 1982;6(1):17-26.
2. Nakamura, T., et al., Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. *Proc Natl Acad Sci U S A*, 1993;90(10): p. 4631-5.
3. De Braekeleer, M., et al., The MLL gene and translocations involving chromosomal band 11q23 in acute leukemia. *Anticancer Res.* 2005;25(3B):1931-44.
4. Vermaelen, K., et al., Anomalies of the long arm of chromosome 11 in human myelo- and lymphoproliferative disorders. I. Acute nonlymphocytic leukemia. *Cancer Genet Cytogenet.* 1983;10(1):05-16.
5. Ayton, P.M. and M.L. Cleary, Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene.* 2001;20(40):5695-707.
6. Chessells, J.M., et al., Clinical features, cytogenetics and outcome in acute lymphoblastic and myeloid leukaemia of infancy: report from the MRC Childhood Leukaemia working party. *Leukemia* 2002;16(5):776-84.
7. Hilden, J.M., et al., Analysis of prognostic factors of acute lymphoblastic leukemia in infants: report on CCG 1953 from the Children's Oncology Group. *Blood* 2006;108(2):441-51.
8. Kingston, R.E. and J.W. Tamkun, Transcriptional regulation by trithorax-group proteins. *Cold Spring Harb Perspect Biol.* 2014;6(10):a019349.
9. Schuettengruber, B., et al., Genome regulation by polycomb and trithorax proteins. *Cell.* 2007;128(4):735-45.
10. Slany, R.K., The molecular biology of mixed lineage leukemia. *Haematologica.* 2009;94(7):984-93.
11. Winters, A.C. and K.M. Bernt, MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front Pediatr.* 2017;5:4.
12. Chan, A.K.N. and C.W. Chen, Rewiring the Epigenetic Networks in MLL-Rearranged Leukemias: Epigenetic Dysregulation and Pharmacological Interventions. *Front Cell Dev Biol.* 2019;7:81.

13. Yokoyama, A., et al., The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell*. 2005;123(2):207-18.
14. Shun, M.C., et al., Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. *J Virol*. 2008;82(23):11555-67.
15. Yokoyama, A. and M.L. Cleary, Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell*. 2008;14(1):36-46.
16. Yokoyama, A., Transcriptional activation by MLL fusion proteins in leukemogenesis. *Exp Hematol*. 2017;46:21-30.
17. Okuda, H., et al., MLL fusion proteins link transcriptional coactivators to previously active CpG-rich promoters. *Nucleic Acids Res*, 2014. 42(7): p. 4241-56.
18. Jones, M., et al., Ash1l controls quiescence and self-renewal potential in hematopoietic stem cells. *J Clin Invest*. 2015;125(5):2007-20.
19. Zhu, L., et al., ASH1L Links Histone H3 Lysine 36 Dimethylation to MLL Leukemia. *Cancer Discov*. 2016;6(7):770-83.
20. Yuan, W., et al., H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem*. 2011;286(10):7983-7989.
21. Miyazaki, H., et al., Ash1l methylates Lys36 of histone H3 independently of transcriptional elongation to counteract polycomb silencing. *PLoS Genet*. 2013. 9(11):e1003897.
22. Gregory, G.D., et al., Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. *Mol Cell Biol*. 2007;27(24):8466-79.
23. Tanaka, Y., et al., Dual function of histone H3 lysine 36 methyltransferase ASH1 in regulation of Hox gene expression. *PLoS One*. 2011;6(11):e28171.
24. Gao, Y., et al., Loss of histone methyltransferase ASH1L in the developing mouse brain causes autistic-like behaviors. *Commun Biol*. 2021;4(1):756.
25. Aljazi, M.B., et al., Cell Signaling Coordinates Global PRC2 Recruitment and Developmental Gene Expression in Murine Embryonic Stem Cells. *iScience*. 2020;23(11):101646.
26. He, J., et al., Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat Cell Biol*. 2013;15(4):373-84.

27. Kim, D., et al., TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14(4):R36.
28. Trapnell, C., et al., Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol.* 2013;31(1):46-53.
29. Ramirez, F., et al., deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 2014;42(Web Server issue):W187-91.
30. Krivtsov, A.V. and S.A. Armstrong, MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer.* 2007;7(11):823-33.
31. Dobson, C.L., et al., The mll-AF9 gene fusion in mice controls myeloproliferation and specifies acute myeloid leukaemogenesis. *EMBO J.* 1999;18(13):3564-74.
32. Rogawski, D.S., et al., Discovery of first-in-class inhibitors of ASH1L histone methyltransferase with anti-leukemic activity. *Nat Commun.* 2021;12(1):2792.
33. Sitwala, K.V., M.N. Dandekar, and J.L. Hess, HOX proteins and leukemia. *Int J Clin Exp Pathol.* 2008;1(6):461-74.
34. Domsch, K., F. Papagiannouli, and I. Lohmann, The HOX-Apoptosis Regulatory Interplay in Development and Disease. *Curr Top Dev Biol.* 2015;114:121-58.
35. Magli, M.C., C. Largman, and H.J. Lawrence, Effects of HOX homeobox genes in blood cell differentiation. *J Cell Physiol.* 1997;173(2):168-77.
36. Ayton, P.M., E.H. Chen, and M.L. Cleary, Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. *Mol Cell Biol.* 2004;24(23):10470-8.

CHAPTER 4:

SMYD5 IS A HISTONE H3-SPECIFIC METHYLTRANSFERASE MEDIATING MONO-METHYLATION OF HISTONE H3 LYSINE 36 AND 37

Published: Biochem Biophys Res Commun. 2022 Apr 9;599:142-147

Title: SMYD5 is a histone H3-specific methyltransferase mediating mono-methylation of histone H3 lysine 36 and 37

Authors: Mohammad B Aljazi, Yuen Gao, Yan Wu, Jin He

ABSTRACT

Although post-translational modifications (PTMs) of some histone H3 lysine residues are well studied, the PTMs of histone H3 lysine 37 in mammalian cells remain largely unknown. In this study, we provide evidence to show that SMYD family member 5 (SMYD5) is a histone H3-specific methyltransferase that catalyzes mono-methylation of H3 lysine 36 and 37 (H3K36/K37me1) *in vitro*. The site-mutagenesis analysis shows that a species-conserved histidine in its catalytic SET domain is required for its histone methyltransferase activity. Genetic deletion of *Smyd5* in murine embryonic stem cells (mESCs) partially reduces the global histone H3K37me1 level in cells, suggesting SMYD5 is one of histone methyltransferases catalyzing histone H3K37me1 *in vivo*. Hence, our study reveals that SMYD5 is a histone H3-specific methyltransferase that mediates histone H3K36/K37me1, which provides a biochemical basis for further studying its functions in mammalian cells.

INTRODUCTION

Histone H3 methylation at transcriptional regulatory regions plays critical roles in regulating gene expression^{1,2}. Specifically, Polycomb repressive complex 2 (PRC2)-mediated tri-methylation of histone H3 lysine 27 (H3K27me3) at gene promoters silences gene expression^{3,4}, while Trithorax-group (TxG) proteins MLL1/MLL2-mediated tri-methylation of histone H3 lysine 4 (H3K4me3) and ASH1L-mediated di-methylation of histone H3 lysine 36 (H3K36me2) facilitate transcription⁵⁻⁷. Although the PTMs of these histone lysine residues and their functions in transcriptional regulation are well studied, PTMs of other histone H3 lysine residues, such as histone H3 lysine 37 (H3K37), are not fully characterized.

SMYD5 (SET and MYND domain-containing protein 5) protein contains a methyltransferase catalytic SET domain and a MYND zinc finger domain^{8,9}. A previous study has reported that SMYD5 is a histone lysine methyltransferase (KMT) that specifically mediates tri-methylation of histone H4 lysine 20 (H4K20me3)¹⁰. Functionally, SMYD5 is involved in transcriptional repression of cytokine genes in macrophages, mESC pluripotency and differentiation, as well as primitive and definitive hematopoiesis in zebrafish¹⁰⁻¹².

To study the function of SMYD5 in mammalian cells, we intended to confirm its enzymatic activity towards histone H4 as reported in the previous study¹⁰. Unexpectedly, our results showed that SMYD5, instead of using histone H4 as its substrate, mediated mono-methylation of histone H3 lysine 36 and lysine 37 (H3K36/K37me1) *in vitro*. The mutagenesis analysis identified a species-conserved histidine in the SET domain was required for its enzymatic activity. Genetic deletion of *Smyd5* in mESCs partially reduced the global histone H3K37me1 level, suggesting SMYD5 is likely to be one of histone KMTs that mediate histone H3K37me1 in mammalian cells.

MATERIALS METHODS

Mouse embryonic stem cell culture

The wild-type E14 mESC line were maintained on the 0.1% gelatin coated plates in the serum-containing mESC culture condition included DMEM medium (Life Technologies) supplemented with 100U/ml penicillin/streptomycin (Life Technologies), 15% fetal bovine serum (Sigma), 1x nonessential amino acid, 1x sodium pyruvate (Life Technologies), 1x GlutaMAX (Life Technologies), 1x β -mercaptoethanol (Life Technologies) and 1000 units/ml leukemia inhibitory factor (ESGRO, EMD Millipore).

Crispr/Cas9-mediated *Smyd5* gene knockout in mESCs

The experiments were carried out according to our previous report¹³. Briefly, the mouse *smyd5* gene gRNA (5'-GCTCTGGGTGTGGTAGAATC-3') was cloned into pX330 vector obtained from Addgene. The target vector and pEF1a-pac vector were co-transfected (5:1 ratio) to E14 mESCs using Xfect according to the manufacture's instruction (TaKaRa, Inc). 48 h after transfection, Puromycin (1 μ g/ml) was added to the medium to select transfected cells. The individual clones were manually picked and expanded. The correct knockout clones were selected based on the Sanger sequencing on the targeting sites of genomic DNA, cDNA, and Western blot analysis.

Recombinant SMYD5 protein purification

Wild-type and mutant SMYD5 cDNAs were cloned into pFASTBAC as described previously¹⁴. Each baculovirus expressing recombinant proteins was generated and amplified following the manufacturer's protocol (Invitrogen). To purify the recombinant proteins, infected

insect SF9 cells were collected and resuspended in F lysis buffer (20 mM Tris–HCl, at pH 7.9, 500 mM NaCl, 4 mM MgCl₂, 0.4 mM EDTA, 2 mM dithiothreitol, 20% glycerol and 0.1% NP40) with proteinase inhibitors. Cells were homogenized with pestle A. The supernatant was recovered by centrifuging and was adjusted to 300 mM NaCl by adding dilution buffer (20 mM Tris–HCl, at pH 7.9), and subsequently incubated with M2 agarose (Sigma) for 4 h at 4 °C. After washing with F washing buffer (20 mM Tris–HCl at pH7.9, 150 mM NaCl, 2 mM MgCl₂, 0.2 mM EDTA, 1 mM dithiothreitol, 15% glycerol and 0.01% NP40), the bound proteins were eluted with Flag peptide (0.2 mg ml⁻¹).

Recombinant histone H3 and H4 peptide purification

Recombinant histone H3 and H4 peptides were purified according to a previous report¹⁵. Briefly, BL21 cells were transformed with the human histones H3 and H4 expression plasmids and grown at 37 °C to a density of OD600 0.6 in LB in shaking cultures. Histone expression was induced by addition of 1 mM IPTG for 2 h. The bacteria pellet was resuspended in SAU buffer (40 mM NaOAc pH 5.2, 6 M urea, 1 mM EDTA pH 8, 5 mM β-Mercaptoethanol, 10 mM lysine) supplemented with 200 mM NaCl (SAU 200) and protease inhibitors (1 mM PMSF, 1 mg/l Aprotinin, 1 mg/l Leupeptin, 0.7 mg/l Pepstatin). Cells were lysed by three passes through a French Press and sonication on ice. The extract was cleared by centrifugation for 20–30 min at 41,000 g and filtration through a glass-fiber prefilter (HPF Millex, Millipore). The pre-cleared cell extract was passed over HiTrap Q HP column (GE Healthcare) followed by a HiTrap SP HP column (5 ml; GE Healthcare) that was pre-equilibrated in SAU 200 buffer. The SP column was washed with 200 mM NaCl for 5 column volumes. Histones were eluted with a NaCl gradient. Pooled histone-containing fractions were dialyzed against cold water over night in dialysis tubing

with a molecular weight cut-off of 6000–8000 Da. After dialysis, the sample was passed over a 1 ml Q HP column. The flow-through containing purified histone peptides was collected and saved in –80 °C for later use.

***In vitro* histone methyltransferase assays**

To test SMYD5 methyltransferase activity *in vitro*, assays were conducted by preparing 20 µl reactions on ice in methyltransferase buffer (50 mM Tris-HCl pH 8.0, 2 mM MgCl₂, 1 mM DTT, 10% glycerol) or Bicine buffer (50 mM Bicine pH 9.0, 2.5 mM EDTA, 0.5 mM DTT, 5% glycerol), with 1–5 µg full-length or short-form SMYD5 enzyme, 3 µg recombinant human histone H3 and H4 peptide, and 1 µCi of [³H]SAM (PerkinElmer Life Sciences). Reactions were incubated for 16 h at 30 °C. Methyltransferase assay samples were resolved by SDS-PAGE gels and transferred on PVDF then stained with Coomassie Blue. The membrane was dried and exposed to a Kodak Tridium Storage Phosphor Screen for 24–48 h and then scanned with a PharosFX Plus Molecular Imager (Bio-Rad).

Histone extraction

1 × 10⁶ mESCs were lysed in 1 ml buffer A [0.25 M sucrose, 10 mM sodium HEPES (pH 7.5), 3 mM CaCl₂, 10 mM NaCl, 1 mM PMSF, 1 mM DTT, 0.25% Nonidet 40] on ice for 30min. The lysate was then centrifuged at 3000 rpm for 10 min to pellet the nuclei. The nuclei were then washed by buffer A one more time before being suspended in 300 µl buffer B [0.25 M sucrose, 10 mM sodium HEPES (pH 7.5), 3 mM CaCl₂, 10 mM NaCl, 1 mM PMSF, 1 mM DTT]. 2 N HCl was then added to a final concentration of 0.2 N. The suspension was extracted at 4 °C overnight and was then centrifuged for 10 min at 13,000 rpm. The supernatant was mixed with equal volume of 50% trichloroacetic acid (Sigma) and kept on ice for 1 h. The precipitated proteins were

collected and washed with cold acetone (Sigma). Proteins were then dissolved in SDS loading buffer and analyzed by Western blotting.

Western blot analysis

Total proteins were extracted by RIPA buffer and separated by electrophoresis by 10–18% PAGE gel. The protein was transferred to the nitrocellulose membrane and blotted with primary antibodies. The antibodies used for Western blot included: rabbit anti-SMYD5 (1:1000, in house), rabbit anti-H3 (1:1000, ab1792, Abcam), rabbit anti-H3K36me1 (1:1000, C15410089, Diagenode), rabbit anti-H3K37me1 (1:1000, C15410295, Diagenode), and IRDye 680 donkey anti-rabbit second antibody (1: 10000, Li-Cor). The images were developed by Odyssey Li-Cor Imager (Li-Cor).

Mass spectrometry analysis

Gel bands were digested in-gel according to Shevchenko et al. with modifications¹⁶. Briefly, gel bands were washed with 100 mM ammonium bicarbonate and dehydrated using 100% acetonitrile. Sequencing grade modified trypsin was prepared to 0.01ug/ μ L in 50 mM ammonium bicarbonate and ~100uL of this was added to each gel band so that the gel was completely submerged. Bands were then incubated at 37C overnight. Peptides were extracted from the gel by water bath sonication in a solution of 60%ACN/1%TFA and vacuum dried to ~2uL. Peptides were then re-suspended in 2% acetonitrile/0.1%TFA to 20uL. From this, 5uL were automatically injected by a Thermo (www.thermo.com) EASYnLC 1200 onto a Thermo Acclaim PepMap RSLC C18 peptide trap (5um, 0.1 mm \times 20 mm) and washed with buffer A for ~5min. Bound peptides were then eluted onto a Thermo Acclaim PepMap RSLC 0.075 mm \times 250 mm C18 resolving

column and eluted over 35min with a gradient of 8%B to 40%B in 24min, ramping to 90%B at 25min and held at 90%B for the duration of the run (Buffer A = 99.9% Water/0.1% Formic Acid, Buffer B = 80% Acetonitrile/0.1% Formic Acid/19.9% Water) at a constant flow rate of 300 nL/min. Column temperature was maintained at 50C using an integrated column heater (PRSO-V2, www.sonation.com). Eluted peptides were sprayed into a ThermoFisher Q-Exactive HF-X mass spectrometer (www.thermo.com) using a FlexSpray spray ion source. Survey scans were taken in the Orbitrap (60000 resolution, determined at m/z 200) and the top 15 ions in each survey scan are then subjected to automatic higher energy collision induced dissociation (HCD) with fragment spectra acquired at 15,000 resolution. The resulting MS/MS spectra are converted to peak lists using Mascot Distiller, v2.7 (www.matrixscience.com) and searched against a database containing all *H.sapiens* protein entries available from UniProt (downloaded from www.uniprot.org on 2020-01-15) appended with common laboratory contaminants (downloaded from www.thegpm.org). Searches were performed using the Mascot searching algorithm, v 2.7, on an in-house server. The Mascot output was then analyzed using Scaffold, v4.11.0 (www.proteomesoftware.com) to probabilistically validate protein identifications. Assignments validated using the Scaffold 1% FDR confidence filter are considered true.

RESULTS

SMYD5 mediates methyl group transferring towards histone H3 *in vitro*

The mammalian SMYD5 proteins contain a MYND zinc finger domain, a lysine methyltransferase catalytic SET domain, and an acidic poly-glutamic acid (poly-E) region at its C-terminus (Figure 26A). A previous study reported SMYD5 possessed a histone methyltransferase activity towards histone H4 lysine 20¹⁰. To confirm the substrate specificity of its histone methyltransferase activity, we purified recombinant human SMYD5 proteins from baculovirus-infected SF9 insect cells and performed *in vitro* methyltransferase assays using full-length recombinant histone H3 or H4 peptides as substrates. Unexpectedly, the results showed that SMYD5 mediated the methyl group transferring towards histone H3 but not H4 peptides (Figure 26B).

To confirm our results, we changed the reaction condition to a bicine buffer that was reported to enhance the enzymatic activity of lysine methyltransferases¹⁷. However, the results showed that under the modified reaction condition, SMYD5 mediated the methyl group transferring only towards histone H3 but not histone H4 peptides (Figure 26C). Furthermore, to exclude the possibility that the C-terminal poly-E acidic region of SMYD5 interfered its enzymatic activity towards histone H4, we generated a short form of recombinant SMYD5 by removing its C-terminal poly-E region (SMYD5^{Δ384-418}) and performed the methyltransferase assays. Same as the full-length SMYD5, the short-form SMYD5 mediated the methyl group transferring towards histone H3 only (Figure 26D). Collectively, these results suggested that SMYD5 is a histone H3-specific methyltransferase and mediates lysine methylation towards histone H3 *in vitro*.

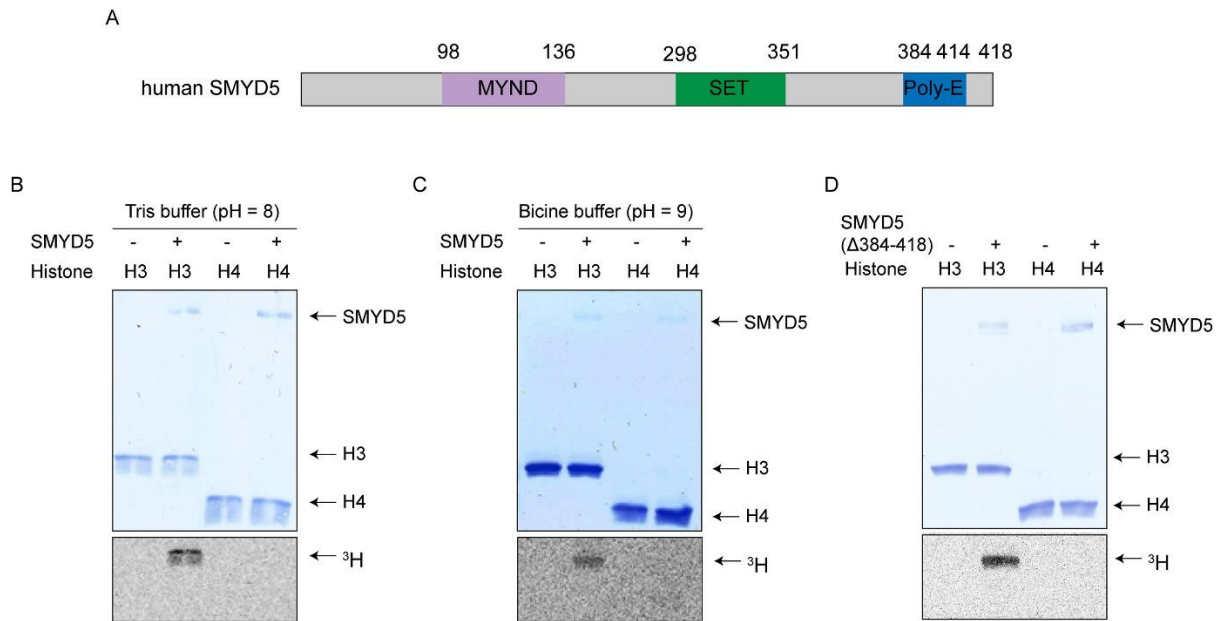


Figure 26: SMYD5 mediates methyl group transferring towards histone H3 *in vitro*. (A) Diagram showing the MYND domain, the SET domain, and the poly-E acidic region of human SMYD5 protein. (B) The results of *in vitro* histone methyltransferase assay (Tris buffer) showing SMYD5 mediates methyl group transferring towards histone H3 but not H4 peptides. (C) The results of *in vitro* histone methyltransferase assay (Bicine buffer) showing SMYD5 mediates methyl group transferring towards histone H3 but not H4 peptides. (D) The results of *in vitro* histone methyltransferase assay showing short-form SMYD5^{Δ384-418} mediates methyl group transferring towards histone H3 but not H4 peptides.

SMYD5 catalyzes mono-methylation of histone H3 lysine 36 and lysine 37 *in vitro*

To identify the lysine residues of histones that are methylated by SMYD5, we performed the mass spectrometry assays to examine the lysine methylation of histone H3 and H4 peptides after incubating histone peptides in the SMYD5-mediated histone methyltransferase reactions *in vitro*. The results showed that compared to negative controls that were incubated in the SMYD5-free reaction buffer, the histone H3 peptides obtained mono-methylation at both lysine 36 and lysine 37 (H3K36/K37me1) after subjected to the SMYD5-mediated methyltransferase reaction (Figure 27A). In contrast, the mass spectrometry assays did not detect methylation at lysine 20 or other lysine residues of histone H4 peptides, consistent with the results of methyltransferase assays *in vitro* (Figure 26).

To confirm that histone H3 K36 and K37 were the lysine residues methylated by SMYD5, we generated the full-length histone H3 peptides with single K36, single K37, and double K36/K37 mutated to alanine (H3K36A, H3K37A, H3K36A/K37A) and performed *in vitro* methyltransferase assays using the mutant histone H3 peptides as substrates. The results showed that single lysine mutation (H3K36A or H3K37A) partially reduced the signal intensity. In contrast, the H3K36/K37 double mutant histone H3 peptides (H3K36A/K37A) largely reduced the signal to the baseline (Figure 27B and C).

To further confirm that SMYD5 mediated the histone H3K36/K37me1 modification, we performed the western blot (WB) analyses using both histone H3K36me1 and histone H3K37me1-specific antibodies to detect the modification after incubation of histone H3 peptides with SMYD5-containing methyltransferase reaction *in vitro*. The results showed both antibodies detected the H3K36me1 and H3K37me1 modifications, and the signal intensity was positively correlated with the SMYD5 concentration in the reactions (Figure 27D). Collectively, these results suggested that

SMYD5 is a histone H3-specific methyltransferase catalyzing mono-methylation of histone H3 lysine 36 and lysine 37 *in vitro*.

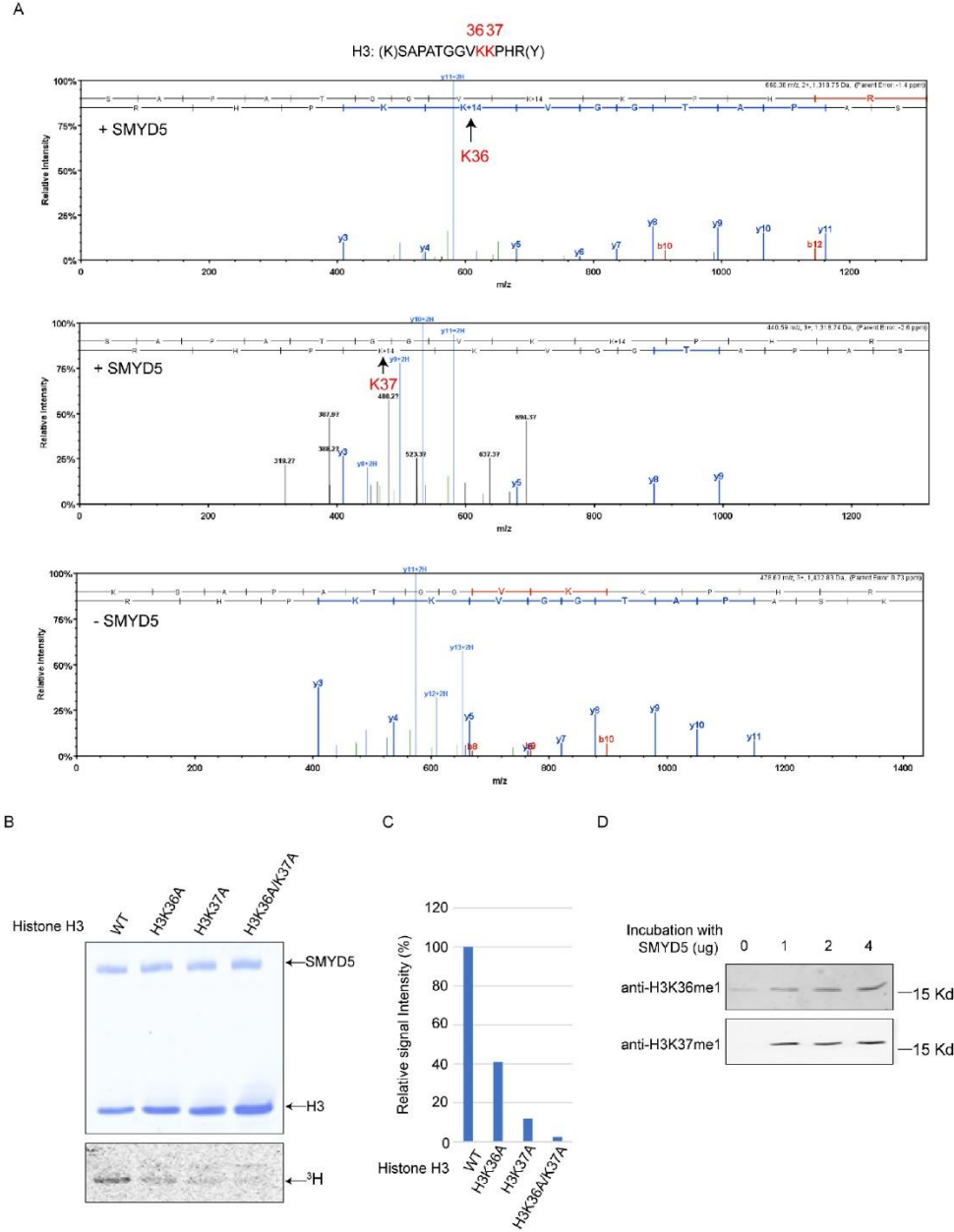


Figure 27: SMYD5 catalyzes mono-methylation of histone H3 lysine 36 and lysine 37 *in vitro*. (A) The spectrum of mass spectrometry analysis showing histone H3 K36 (top panel) and K37 (middle panel) obtain mono-methylation after subjected to the SMYD5-mediated methyltransferase reaction *in vitro*. The histone H3 peptides incubated in the SMYD5-free reaction buffer serve as negative controls (bottom panel). (B–C) *In vitro* histone methyltransferase assays (B) and quantitative analysis (C) showing mutations of histone K36 and K37 largely reduce SMYD5-mediated methyl group transferring to histone H3 peptides. Relative intensity of ³H signal for different histone H3 peptides is normalized to both input (Coomassie staining) and signal intensity of wild-type H3 peptides: (Intensity of ³H/Intensity of H3 input)/(Intensity of ³H of wild-type H3/Intensity of wild-type H3 input) x 100%. (D) WB analysis showing the histone H3K36me1 and H3K37me1-specific antibodies detect the H3K36me1 and H3K37me1 modifications after histone H3 peptides are subjected to the SMYD5-mediated methyltransferase reaction.

A species-conserved histidine in the SET domain is required for the SMYD5 methyltransferase activity

To identify the critical amino acids required the methyltransferase activity of SMYD5, we compared the amino acid sequences of catalytic SET domains of SMYD5 proteins from different species. The analysis showed a histidine (H316 of human SMYD5) was conserved in different species (Figure 28A), suggesting its importance for the SMYD5 methyltransferase activity. To test this hypothesis, we generated the recombinant human SMYD5 with its histidine mutated to alanine (SMYD5^{H316A}) and performed *in vitro* methyltransferase assays to examine its enzymatic activity. The results showed that the mutant SMYD5 lost its methyltransferase activity towards histone H3 peptides, suggesting that the conserved histidine in the SET domain is required for the histone methyltransferase activity of SMYD5.

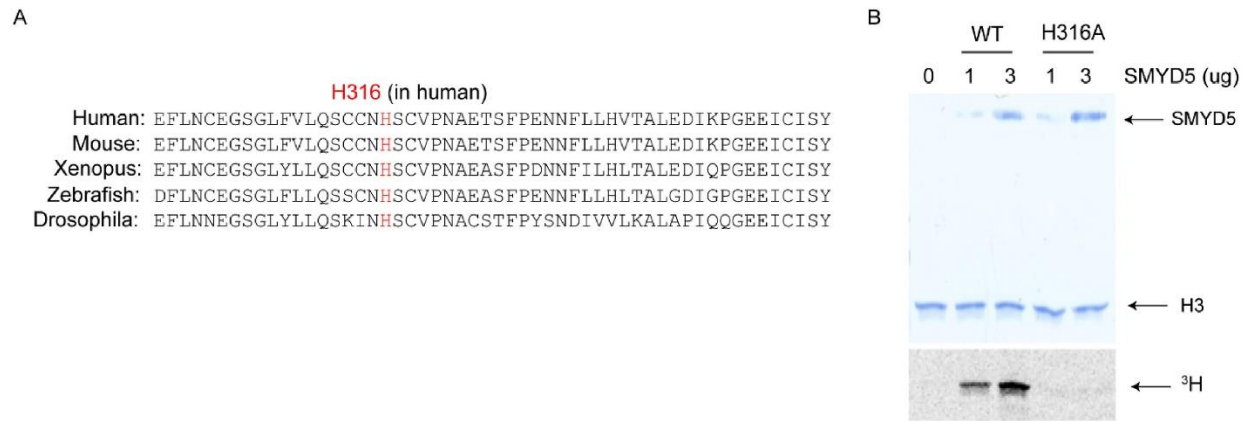


Figure 28: A species-conserved histidine in the SET domain is required for the SMYD5 methyltransferase activity. (A) Sequence alignment showing the conserved histidine in the SET domain of SMYD5 from different species. (B) The results of histone methyltransferase assay showing that mutation of histidine (H316A) of human SMYD5 abolishes its histone KMT activity.

Deletion of *Smyd5* in mESCs partially reduces the global histone H3K37me1 level

To examine the SMYD5-mediated histone H3 modifications in mammalian cells, we knockout the *Smyd5* gene in mESCs (*Smyd5*-KO) by the Crispr/Cas9-mediated gene knockout approach¹³. The deletion of *Smyd5* gene in two cell clones was confirmed by WB analysis at the protein level (Figure 29A). Further WB analysis showed that compared to wild-type cells, the *Smyd5*-KO mESCs had partially reduced global histone H3K37me1 levels, while the global histone H3K36me1 did not show markedly changes (Figure 29B), suggesting that SMYD5 was likely to be one of KMTs contributing to the histone H3K37me1 modification in mammalian cells.

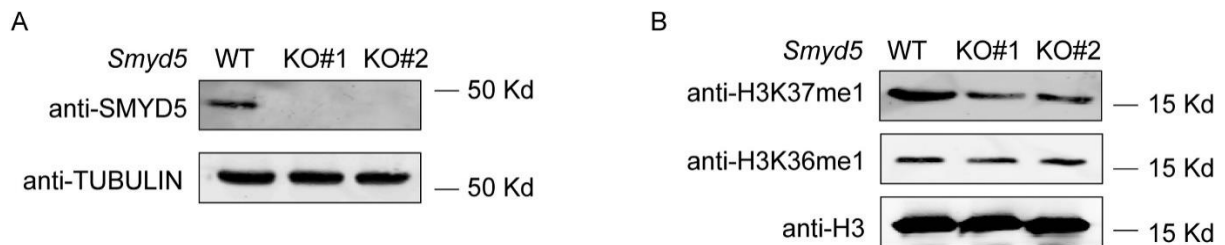


Figure 29: Deletion of *Smyd* in mESCs partially reduces the global histone H3K37me1 level. (A) WB analysis showing the SMYD5 expression in wild-type (WT) and *Smyd5*-knockout mESC clone 1 and 2 (KO #1 and KO #2). (B) WB analysis showing the global histone H3, H3K36me1, and H3K37me1 levels in wild-type (WT) and *Smyd5*-knockout mESC clone 1 and 2 (KO #1 and KO #2).

DISCUSSION

Different from a previous study showing that SMYD5 is a histone H4K20me3-specific methyltransferase¹⁰, our current study demonstrates that SMYD5 is a histone H3-specific lysine methyltransferase that catalyzes mono-methylation of histone H3 K36 and K37. Our conclusion is supported by *in vitro* methyltransferase assays showing that SMYD5 mediates the methyl group transferring towards histone H3 peptides only under two reaction conditions as well as using both full-length and short-form recombinant SMYD5 proteins for the reactions (Figure 26A–C). Moreover, the mass spectrometry assays show that histone H3 peptides, but not histone H4 peptides, obtain mono-methylation at its K36 and K37 residues after subjected to the SMYD5-mediated methyltransferase reactions (Figure 27A), which is further confirmed by the *in vitro* methyltransferase assays showing that mutation of both histone H3 K36 and K37 diminishes the methyl group transferring towards histone H3 peptides (Figure 27B and C). In addition, WB analyses show that H3K36me1 and H3K37me1-specific antibodies detect both modifications of histone H3 peptides after subjected to the SMYD5-mediated methyl group transferring reactions (Figure 27D). Compared to the H3K37me1 signal, there existed a background signal of histone H3K36me1 in the WB analysis (Figure 27D, lane 1), which might be caused by non-specific recognition of histone H3 by the antibody. The collective results suggest that SMYD5 is a histone H3-specific methyltransferase catalyzing mono-methylation of histone H3 K36 and K37. The discrepancy between our results and the earlier report regarding the histone substrates of SMYD5 could be due to the antibody-based methods used for detecting histone modifications in the previous study¹⁰, which might generate a non-specific H4K20me3 signal and mislead the conclusion.

Although our *in vitro* biochemical assays suggest that SMYD5 is a histone H3K36/K37-specific methyltransferase (Figure 26, Figure 27, Figure 28), we notice that the deletion of *Smyd5* in mESCs does not completely abolish histone H3K36/K37me1 in cells (Figure 29). The discrepancy between *in vitro* and *in vivo* results could be caused by two possible reasons: (i) histone H3K36me1 and H3K37me1 modifications in mammalian cells are catalyzed by both SMYD5 and other histone KMTs, which compensate the loss of SMYD5 function in depositing H3K36/K37me1 in the *Smyd5*-KO cells; or (ii) SMYD5 mediates locus-specific histone H3K36/K37me1, thus loss of SMYD5 has limited effects on the global H3K36/K37me1 level.

Recently, SET7 and SET1/SET2 are found to catalyze histone H3K37 methylation in *S. pombe* and *S. cerevisiae*, respectively^{18,19}, which are functionally involved in gametogenesis and DNA replication. Our current study demonstrates that histone H3K37me1 also exists in mammalian cells and SMYD5 is likely to be one of histone KMTs depositing this histone modification in cells. Although the function of SMYD5 remains largely unelucidated, it is worth to note that SMYD5 is found to localize in both nuclei and mitochondria in mammalian cells²⁰, suggesting its intra- and extra-nuclear functions. The identification of SMYD5-mediated histone H3K36/K37me1 in this study provides a biochemical basis for further studying the functions of SMYD5 and its-mediated histone modifications in mammalian cells.

REFERENCES

REFERENCES

1. Kouzarides, T., Histone methylation in transcriptional control. *Curr Opin Genet Dev.* 2002;12(2):198-209.
2. Zhang, Y. and D. Reinberg, Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* 2001;15(18):2343-60.
3. Di Croce, L. and K. Helin, Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol.* 2013;20(10):1147-55.
4. Aranda, S., G. Mas, and L. Di Croce, Regulation of gene transcription by Polycomb proteins. *Sci Adv.* 2015;1(11):e1500737.
5. Kingston, R.E. and J.W. Tamkun, Transcriptional regulation by trithorax-group proteins. *Cold Spring Harb Perspect Biol.* 2014;6(10):a019349.
6. Shilatifard, A., The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem.* 2012;81:65-95.
7. Gregory, G.D., et al., Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. *Mol Cell Biol.* 2007; 27(24):8466-79.
8. Rueda-Robles, A., et al., Functions of SMYD proteins in biological processes: What do we know? An updated review. *Arch Biochem Biophys.* 2021;712:109040.
9. Spellmon, N., et al., Structure and function of SET and MYND domain-containing proteins. *Int J Mol Sci.* 2015;16(1):1406-28.
10. Stender, J.D., et al., Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. *Mol Cell.* 2012;48(1):28-38.
11. Kidder, B.L., et al., SMYD5 regulates H4K20me3-marked heterochromatin to safeguard ES cell self-renewal and prevent spurious differentiation. *Epigenetics Chromatin.* 2017;10:8.
12. Fujii, T., et al., Smyd5 plays pivotal roles in both primitive and definitive hematopoiesis during zebrafish embryogenesis. *Sci Rep.* 2016;6:29157.
13. Aljazi, M.B., et al., Cell Signaling Coordinates Global PRC2 Recruitment and Developmental Gene Expression in Murine Embryonic Stem Cells. *iScience.* 2020;23(11):101646.

14. He, J., et al., Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat Cell Biol.* 2013;15(4):373-84.
15. Klinker, H., et al., Rapid purification of recombinant histones. *PLoS One.* 2014;9(8):e104029.
16. Shevchenko, A., et al., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc.* 2006;1(6):2856-60.
17. Eram, M.S., et al., Trimethylation of histone H3 lysine 36 by human methyltransferase PRDM9 protein. *J Biol Chem.* 2014;289(17):12177-12188.
18. Shen, Y., et al., Set7 Is a H3K37 Methyltransferase in *Schizosaccharomyces pombe* and Is Required for Proper Gametogenesis. *Structure.* 2019;27(4):631-638 e8.
19. Santos-Rosa, H., et al., Methylation of histone H3 at lysine 37 by Set1 and Set2 prevents spurious DNA replication. *Mol Cell.* 2021;81(13):2793-2807 e8.
20. Hou, Y., et al., Methyltransferase SMYD5 Exaggerates IBD by Downregulating Mitochondrial Functions via Post-translational Control of PGC-1 α Stability. 2020:202011.16.385765.

CHAPTER 5:
CONCLUSION

OVERVIEW

This thesis explored the molecular mechanisms involved in PcG and TrxG protein recruitment and assessed their functional role in transcriptional regulation within stem cell maintenance and cancer development. Our results demonstrated that FGF/ERK signaling regulated the global PRC2 occupancy in mESCs through Jarid2 mediated recruitment, and transcriptional activation is predominantly affected by transcription factors rather than PRC2 occupancy (Chapter 2). In chapter 3, we found that MLL-AF9 leukemia is dependent on the enzymatic activity of the TrxG factor ASH1L by promoting leukemogenic initiation and maintenance. In chapter 4, we found that SMYD5 mediates H3K36me/H3K37me1 *in vitro*, and functions as one of the histone methyltransferases catalyzing histone H3K37me1 *in vivo*. A summary of the findings from each chapter is described below.

CHAPTER 2: CELL SIGNALING COORDINATES GLOBAL PRC2 RECRUITMENT AND DEVELOPMENTAL GENE EXPRESSION IN MURINE EMBRYONIC STEM CELLS

Previous work has shown that while the expression of early lineage genes is repressed PRC2 occupancy is unexpectedly reduced in naive mESCs. In this study, we investigated the role of FGF/ERK signaling in mediating PRC2 recruitment. Analysis revealed that Jarid2 expression levels were reduced in naive mESCs, whereas FGF/ERK signaling reestablished Jarid2 expression. Moreover, loss of Erk1/Erk2 reduced Jarid2 expression in mESCs, confirming its dependence on ERK signaling function. ChIP-seq for PRC2 showed reduced occupancy at CGIs in naive mESCs, while ectopic expression of Jarid2 could sufficiently restore PRC2 chromatin binding. In addition, our results suggest that transcription of bivalent genes in naive mESCs is dependent on the function of transcription factors rather than PRC2 chromatin occupancy. Thus, the work outlined in this

study revealed the molecular mechanism by which FGF/ERK signaling regulates PRC2 occupancy in mESCs, while also addressing the functional role of transcription factors and Polycomb-mediated epigenetic mechanisms in transcriptional regulation.

CHAPTER 3: HISTONE H3K36ME2-SPECIFIC METHYLTRANSFERASE ASH1L PROMOTES THE MLL-AF9-INDUCED LEUKEMOGENESIS

Although ASH1L is a factor associated with MLL-rearrangement leukemogenesis, the significance of its enzymatic activity in the development of MLL-rearranged leukemias remains unclear in the *Ash1L* gene knockout animal models. Using an *Ash1L* conditional knockout mouse model, we showed that loss of *Ash1L* impairs the initiation of MLL-AF9-induced leukemic transformation *in vitro*. In addition, *in vitro* studies also revealed that Ash1L is essential for MLL-AF9 leukemia maintenance. Our leukemia transplantation experiments showed that *Ash1L* deletion in MLL-AF9 transformed cells increased survival rates and hindered leukemia progression *in vivo*. Notably, the loss of ASH1L function in the *Ash1L* deficient cells was rescued by the expression of wild-type but not the enzymatically inactive ASH1L, demonstrating the importance of ASH1L catalytic activity in the maintenance of MLL-AF9 leukemia. RNA-seq analysis showed that ASH1L regulates several *HOXA* genes necessary for MLL-rearrangement leukemia initiation and maintenance. Furthermore, ChIP analysis results showed that ASH1L occupies MLL-AF9 target genes and deposits H3K36me2 marks at the transcriptional start sites of these genes. Collectively, these findings reveal the significance of ASH1L and its catalytic activity in MLL-AF9 and identify a potential therapeutic target for MLL-AF9 leukemia.

CHAPTER 4: SMYD5 IS A HISTONE H3-SPECIFIC METHYLTRANSFERASE MEDIATING MONO-METHYLATION OF HISTONE H3 LYSINE 36 AND 37

Previous studies identified SMYD5 as histone H4 methyltransferase involved in transcriptional repression genes. Interestingly, our assessment of SMYD5 catalytic function revealed that it mediated H3K36/K37me1 *in vitro* rather than targeting histone H4. Mutation of the conserved histidine within the catalytic SET domain abolished SMYD5 methyltransferase activity *in vitro*. Additionally, loss of Smyd5 in mESCs reduces the global histone H3K37me1 level in cells. Implementation of mass spectrometry analysis confirmed that H3 peptides were mono-methylation K36 and K37 residues in SMYD5 methyltransferase reactions assays. Thus, our data functionally identifies SMYD5 as an H3 specific methyltransferase that mediates H3K36me/H3K37me1 *in vitro*. It also reveals that SMYD5 serves as one of the histone methyltransferases catalyzing histone H3K37me1 *in vivo*.

FUTURE WORK

The results gathered from this dissertation provide insight into the molecular mechanisms involved in PcG and TrxG recruitment and address their significance in mESCs and MLLr leukemia. These studies also identified an unexpected catalytic function for the enzyme assessment of SMYD5, revealing that it mediated H3K36/K37me1. Although these studies characterize the underlying molecular mechanisms, additional work is needed, to fully elucidate the processes involved.

While this study addresses the molecular mechanism of PRC2 recruitment cell signaling, questions remain unaddressed. Previous findings shows that H3K27me3 becomes redistributed to non-CGI regions, which correlates with the local CpG density, indicating reduced PRC2 occupancy at CGIs in naive mESCs could be caused by the relocation of PRC2 from CGIs to new

DNA demethylated sites¹. Mass spectrometry-assisted quantitative measurement of heterochromatin-bound PRC2 and H3K27me3 modification in naive mESCs could potentially address this question. Our investigations also demonstrated that reduced PRC2 occupancy at promoters alone is insufficient for activating FGF/ERK signaling target genes, suggesting transcription factors serve a dominant role in transcription. Further studies should assess the dynamics of PcG complexes and transcription factor recruitment and their effects on transcriptional regulation in other cell types during cell differentiation and disease development.

Although our ASH1L studies demonstrated its catalytic significance within our Ash1L gene knockout animal model, questions remain unaddressed. Given that we studied the role of ASH1L specifically in MLL-AF9 leukemia, it remains unclear if ASH1L has a similar effect on other MLL fusion-protein-mediated leukemias. In our experiments, we found that Ash1L deletion promotes cell death. However, we found that its loss does not completely abolish their growth *in vitro* and *in vivo*. This indicates that the MLL-AF9-transformed cells have heterogeneous responses to Ash1L ablation. Future investigation into these underlying questions will provide a better understanding of the role of ASH1L in MLLr-associated leukemogenesis.

Finally, our investigation identified an unexpected catalytic function for the enzyme assessment of SMYD5, revealing that it mediated H3K36/K37me1. Further studies should focus on the biological significance of these modifications. Additionally, we found that the loss of Smyd5 in mESCs did completely deplete histone H3K36/K37me1 modifications. Uncovering which other factors mediate this function will aid in characterizing the role of these modifications in epigenetic gene regulation. Finally, reports have shown that SMYD5 localizes in both nuclei and mitochondria in mammalian cells². Future investigations should examine the underlying

molecular mechanisms for SMYD5 recruitment and localization to these distinct cellular compartments.

REFERENCES

REFERENCES

1. van Mierlo G., Dirks R.A.M., De Clerck L., et al. Integrative proteomic profiling reveals PRC2-dependent epigenetic crosstalk maintains ground-state pluripotency. *Cell Stem Cell*. 2019;24:123–137.e8.
2. Hou, Y., et al., Methyltransferase SMYD5 Exaggerates IBD by Downregulating Mitochondrial Functions via Post-translational Control of PGC-1 α Stability. 2020:202011.16.385765.