Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

11-1-2017 10:30 AM

Identification of novel binding partners and substrates of histone H3K4 specific demethylase KDM5B/JARID1B

Qi Fang, *The University of Western Ontario*

Supervisor: Dr. Shawn Li, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biochemistry

© Qi Fang 2017

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Biochemistry Commons

Recommended Citation

Fang, Qi, "Identification of novel binding partners and substrates of histone H3K4 specific demethylase KDM5B/JARID1B" (2017). *Electronic Thesis and Dissertation Repository*. 5061.
https://ir.lib.uwo.ca/etd/5061

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

It has been 50 years since the first histone methylation was identified. From then, numerous studies have verified that histone methylation neutralizes the positive charges of lysine, alters chromatin architectures, and recruits other DNA binding complexes. In this study, we examined the molecular basis of a histone demethylase – KDM5B in both its substrate and binding machineries. We hypothesized that KDM5B has alternative binding partners and substrates besides its known target H3K4me3. By utilizing biochemical approaches such as *in vitro* pull-down assays and fluorescence polarization assays, we were able to identify potential KDM5B binding partners – H2BK43me0/2. Also, we reinvestigated the recognition mechanisms and found that the unmodified H2BK43me0 was a potential binding site of PHD1 and PHD2 while H2BK43me2 was a potential partner of PHD1. In addition, we observed H2BK43me2 demethylation via KDM5B in MCF7 cells. Together, our preliminary results suggested that H2BK43me2 was an alternative binding partner and substrate of KDM5B.

Keywords

Histone H2BK43, histone H3K4, histone lysine methylation, stem cell epigenetics, pluripotency, stem cell differentiation, JARID1b, KDM5B, plant homeodomain

Co-Authorship Statement

Biochemistry experiments were performed by former member – Dr. Huadong Liu, Dr. Christopher Wynder, Dr. Marek Galka and Dr. Leanne Stalker; data were extracted from unpublished manuscript (Chapter 2 introduction section). Dr. Rui Wang helped the MS experiment for identifying histone modifications (**Figure 3.2**). Dr. Xiang Ruan cloned the H2BWT and mutant constructs (**Appendix D**). Dr. Kyle Biggar provided data for identifications of novel histone modifications (**Table 6**).

Acknowledgments

I would first like to thank Dr. Shawn Li for his guidance and support throughout my studies.

I would also like to thank Dr. Rui Wang, Dr. Mei Huang and Dr. Kyle Biggar for their helpful advice and support in this study.

I would like to extend my gratitude to Dr. Xiao Bing Shi for providing me KDM5B-related plasmids; Dr. Andras Nagy and Dr. Malgosia Kownacka for providing me the R1 stem cell line and related trainings; and Dr. Cindy Shao for providing me confocal microscopy trainings.

I would also like to give special thanks to all the members in the Li lab during my time of study. I sincerely appreciate all the help and support from our members, with special thanks to Courtney Voss for routine lab maintenances and Dr. Tomonori Kaneko for mentorship.

Table of Contents

Abstract i

Co-Authorship Statement ii

Acknowledgments ii

Table of Contents iii

List of Tables vi

List of Figures vii

List of Abbreviations ix

Chapter 1: Literature Review 1

1.1 DNA, histone and chromatin 1

1.2 Histone post-translational modifications and epigenetics 3

1.3 Histone code 5

1.4 Histone modification enzymes and epigenetic editors 6

1.5 KDM5B/JARID1B related diseases 9

1.6 Plant homeodomains of KDM5B 9

1.7 H3K4me2/3 demethylation by KDM5B 10

1.8 Stem cell biology and pluripotency 12

1.9 The role of KDM5B in neurogenesis 16

1.10 Previous study and preliminary data 17

1.11 Hypothesis and aims 24

Chapter 2: Method and results 25

2.1 Methods 25

2.1.1 Bacterial culture and protein purification 25

2.1.2 Peptide synthesis 25

2.1.3 Matrix Assisted Laser Desorption/Ionization (MALDI) 26

2.1.4 Mammalian cell culture and subcellular fractionation..... 26

2.1.5 GST and streptavidin pull-down analysis..... 27

2.1.6 SDS-PAGE and Coomassie staining..... 27

2.1.7 Western blot..... 28

2.1.8 Flourescence polarization assay..... 28

2.1.9 Mass spectrometry..... 29

2.1.10 Cell culture and neuronal differetiation..... 29

2.1.11 Histone extraction..... 30

2.1.12 Antibody validation..... 30

2.1.13 Plasmid purification and cloning..... 31

2.1.14 Transfection and viral infection..... 31

2.1.15 RT-qPCR.....32

2.1.16 Immunofluorescence and confocal microscopy.....32

2.2 Results..... 34

2.2.1 Protein purifications of GST-tagged KDM5B PHD domains.....34

2.2.2 Subcellular fractionation.....35

2.2.3 Histone H2BK43 peptdes intereacted with KDM5B PHD domains.....36

2.2.4 Identification of H2BK43 methylations under cell physiological conditions 39

2.2.5 Histone mutant H2BK43 to R showed deleterious effects on mESCs survival 41

2.2.6 KDM5B was downregulated at differentiation day 5, but not KDM5A and KDM5C.....42

2.2.7 KDM5B knockdown in breast cancer MCF7 cells suppressed cell proliferation.....45

2.2.8 H2Bk43me2 was a potenetial substrate of KDM5B in MCF7 cells but not mESCs.....46

Chapter 3: General discussion and future studies 47

3.1 General discussion and conclusions 47

3.2 Future studies 51

References 53

Tables 59

Appendices 63

Curriculum Vitae 70

List of Tables

Table 1: List of peptides used in this study 59

Table 2: List of shRNAs and siRNAs used in this study 60

Table 3: List of plasmids used in this study 60

Table 4: List of primers used in this study 61

Table 5: List of antibodies used in this study 61

Table 6: Computer-predicted new histone methylation sites 62

List of Figures

Figure 1.1: Eukaryotic genome compaction and organization..... 2

Figure 1.2: 2D depictions of four core histones and modifications 3

Figure 1.3: Major epigenetic mechanism that alters gene expression 4

Figure 1.4: Histone modifications that are acting in concert 5

Figure 1.5: Common histone modification readers for methyllysine, methylarginine, acetyllysine, phosphorylated serine and threonine 7

Figure 1.6: Major player in epigenetics – “Readers”, “Writers” and “Erasers” 8

Figure 1.7: Illustration of architectures of KDM5/JARID1 family in human, fruit fly and yeas**t** 11

Figure 1.8: Hypothetical KDM5B substrate (H3K4me3) binding model..... 11

Figure 1.9: Architectures of early embryonic development..... 14

Figure 1.10: Schemes of LIF induced JAK -STAT signaling and 2i/LIF induced signaling 15

Figure 1.11: Structure of nucleosome in cartoon representation..... 19

Figure 1.12: H2BK43me2 is an alternative substrate of recombinant KDM5B *in vitro*.. 19

Figure 1.13: H2BK43me2 is a substrate of KDM5B under cell physiological conditions21

Figure 1.14: Pull-down analysis of binding preference of different KDM5B PHD domains 23

Figure 2.1: Coomassie staining of KDM5B - PHD1, 2 and 3 domains from *E. coli* Rosetta DE2..... 34

Figure 2.2: Subcellular fractionations. 35

Figure 2.3: Histone H2BK43 peptides interacted with KDM5B PHD domains *in vitro*.. 38

Figure 2.4: KDM5B was found in histone peptide pull-downs from cell lysate..... 35

Figure 2.5: Identification of *in vitro* H2BK43 methylations by MS/MS..... 40

Figure 2.6: Histone H2BK43 to R was a potential deleterious trait in cell survival.... 41

Figure 2.7: KDM5B knockdown in mESCs displayed normal morphology but brought compensational upregulation of KDM5C..... 43

Figure 2.8: Downregulation of KDM5B during neuronal differentiation 44

Figure 2.9: KDM5B knockdown impaired cell proliferation in MCF7 cells..... 45

Figure 2.10: KDM5B expression affected H3K4me3 and H2BK43me2 levels MCF7 cells but only H3K4me3 in mESCs 46

List of Abbreviations

bHLH	Basic helix-loop-helix
ECL	Enhanced chemiluminescence
DMEM/F12	Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12
MEF	Mouse embryonic fibroblasts
RA	Retinoic acid
BSA	Bovine serum albumin
NEAA	Non-essential amino acids
DMEM	Dulbecco’s modified eagle medium
PBS	Phosphate buffered saline
FBS	Fetal bovine serum
FCS	Fetal calf serum
PHD	Plant homeodomain
PVDF	Polyvinylidene fluoride
mESC	Mouse embryonic stem cell
ICM	Inner cell mass
TE	Trophectoderm
PE	Primitive endoderm
EPI	Epiblast
TE (buffer)	Tris-EDTA

iPS	Induced pluripotent stem cell
JmjN	Jumonji N
JmjC	Jumonji C
KDM5B	Lysine demethylase 5B
JARID1B	Jumonji AT Rich Interactive Domain 1B
KMT	Lysine methyltransferase
LIF	Leukemia inhibitory factor
LSD1	Lysine specific demethylase 1
MALDI-MS	Matrix-assisted laser desorption/ionization mass spectrometry
MRM-MS	Multiple reaction monitoring mass spectrometry
RT-qPCR	Quantitative real-time polymerase chain reaction
siRNA	Short interfering RNA
shRNA	Short hairpin RNA
PMSF	Phenylmethylsulfonyl fluoride
DMSO	Dimethyl sulfoxide
PLL	Poly-L-lysine
PTM	Post-translational modification
GST	Glutathione S-transferase
IF	Immunofluorescence
PFA	Paraformaldehyde

TBST	Tris buffered saline with Tween 20
TBS	Tris buffered saline
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
APS	Ammonium Persulfate
TEMED	Tetramethylethylenediamine
Me1	Monomethyl
Me2	Dimethyl
Me3	Trimethyl
6-FAM	6-Carboxyfluorescein
FP	Fluorescence polarization
K_D	Equilibrium dissociation constant
BCA	Bicinchoninic acid assay
CBM	Chromatin binding modules
TOF	Time-of-flight
DTT	dithiothreiol
ACN	Acetonitrile
ABC	Ammonium bicarbonate
H2BK43me0	Histone H2B lysine 43 unmodified

Chapter 1

Literature review

1.1 DNA, histone and chromatin

In the eukaryotic genome, DNA is the basic molecule used to encrypt genetic information. DNA is a polymer of nucleotides, which consists of pentose sugars, nitrogenous bases and phosphate groups. From 1885 to 1901, Albrecht Kossel and his colleagues discovered nucleic acids contain five nitrogen-containing compounds: adenine, cytosine, guanine, thymine and uracil. The polymerization of nucleotides is held by the phosphodiester bonds between the phosphate groups at the 5'C position and the OH group at the 3'C position of pentose sugar. Unlike the prokaryotes, eukaryotic genome is organized in a nucleus; therefore, it is necessary to provide a substance that can arrange and anchor the DNA double helix. In the discovery of nucleic acids, Albrecht Kossel also showed that a protein portion was present in the isolated substance "nuclein", which he later termed as "histon" (Campos and Reinberg, 2009). Under the eukaryotic physiological conditions, approximately 147 base pairs of DNA wrap around a positively-charged octamer protein, which comprises of two copies of histones: H2A, H2B, H3 and H4. This structure is termed "nucleosome". Nucleosomes are organized into higher-order structures by linking to histone H1 with extra 10-70 base pairs of DNA. In 1976, evidence of higher-order structure than nucleosome was observed by John Finch and Aaron Klug and was later termed "solenoid". To date, no other higher-order structure beyond solenoid has been reported with clear evidence (Kornberg, 1974; van Holde *et al*., 1974; Woodcock, 2005).

The average size of a nucleus is approximately 10 micrometers in diameter, whereas the total length of DNA in human cells is around 2 meters. The mechanism of fitting the entire human genome into a nucleus is still poorly understood. In a nucleus, chromatin is

the highest order that results in DNA compression to date. Typically, a base pair in a DNA double helix is approximately 2 nanometers in width. Nucleosome core, which contains both DNA and histone cores, has a diameter of around 11 nanometers. Six nucleosomes join by histone H1, which results in a structure called solenoid that is approximately 30 nanometers in diameter. It is believed that from the solenoid, a 300-nanometer loop can be formed and further compressed into a 250-nanometer fiber. Hypothetically, a folding and tightening process after the 250-nanometer stage is required to construct a 700-nanometer chromatid, then a 1400-nanometer chromosome (**Figure 1.1**).

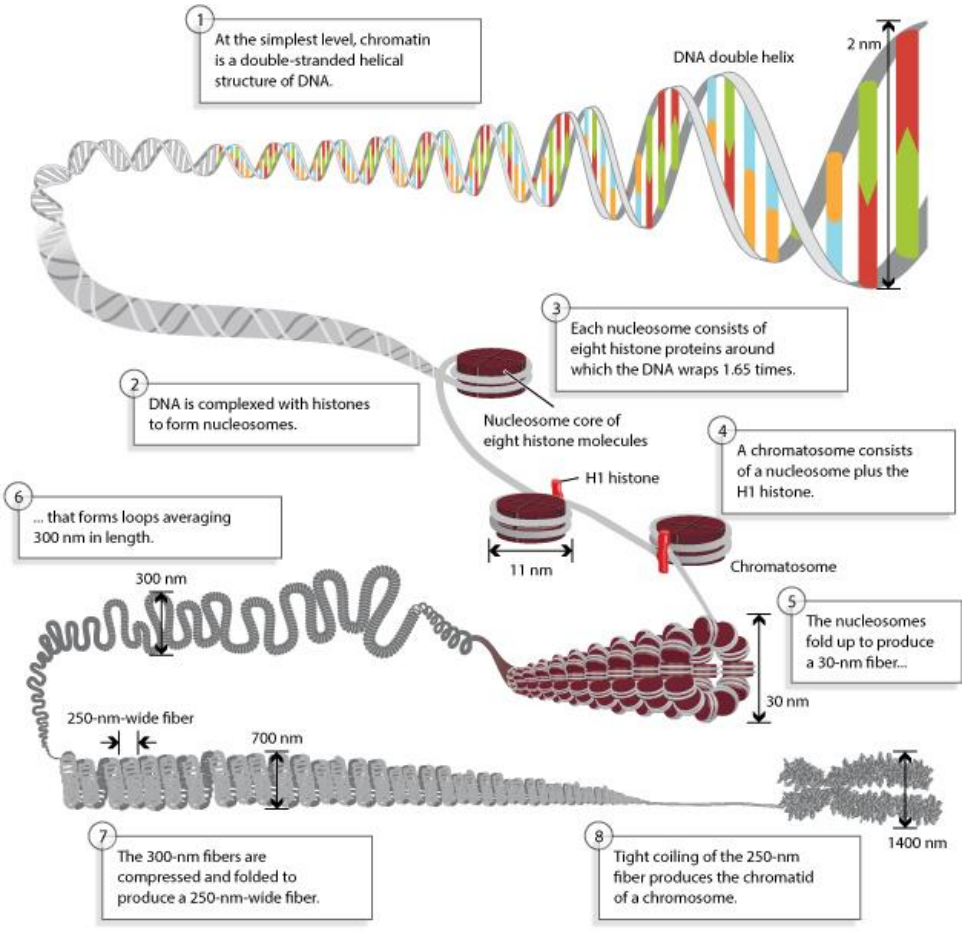


Figure 1.1: Eukaryotic genome compaction and organization. Hypothetical illustration of the compression of DNA double helix to chromosome (Annunziato, 2008). A total of $6.6*10^9$ base pairs of DNA is folded and compressed into a 10-micrometer nucleus.

1.2 Histone post-translational modifications and epigenetics

The reason that prokaryotes, such as bacteria, do not have nucleosomes is because they lack the structural protein - histones. There are four core histones: H2A, H2B, H3 and H4, and one linker histone H1. Except for histone H1, all four core histones have about 25% of its mass weighted in the N- and C- terminal domains.

Although these terminal domains do not provide structural functions, they can be modified to regulate gene expression (**Figure 1.2)**. Histone post-translational editors are able to directly modify these histone tails while chromatin-binding proteins recruit other transcriptional complexes for gene expression (Wolffe and Hayes, 1999).

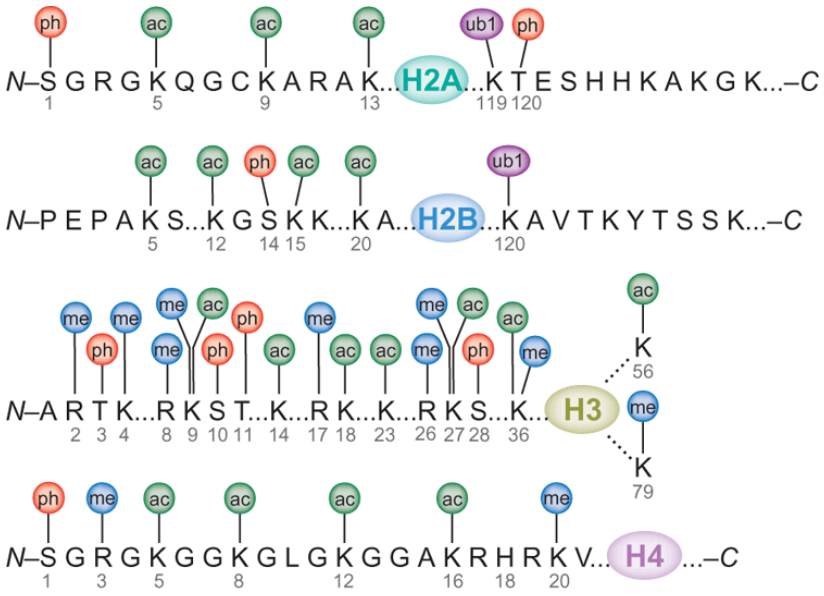


Figure 1.2: 2D depictions of four core histones and modifications. Histone modifications mostly take place in N- and C- terminal domains with the exception of histone H3 K56 and K79 positions. The modification code ac, me, ph and ub1 stand for acetylation, methylation, phosphorylation and ubiquitination respectively (Bhumik *et al.*, 2007).

Histone modifications have been known primarily since the 1960s, when Dr. Vincent Allfrey at Rockefeller University postulated that histone acetylation may be involved in

RNA synthesis (Allfrey *et al.*, 1964). Although the histone post-translational modifications (PTMs) is just one aspect of epigenetics, in the past 50 years, this field has grown massively along with the human ROADMAP epigenomics project. In fact, without the current progress of epigenetics, understanding and utilizing the ENCODE data would be more complicated. The concept of epigenetics was first mentioned in the 1940s by Conrad Waddington. The original idea of epigenetics was described as a biological phenomenon of heritable regulatory mechanism in gene expression that is independent of changes of the DNA sequence. Chromosome remodeling (ie: Xist), protein PTMs (ie: histone modifications), DNA methylations and ncRNAs related regulations (ie: miRNA silencing) are some of the examples of epigenetics (Woffle and Matzke, 1999).

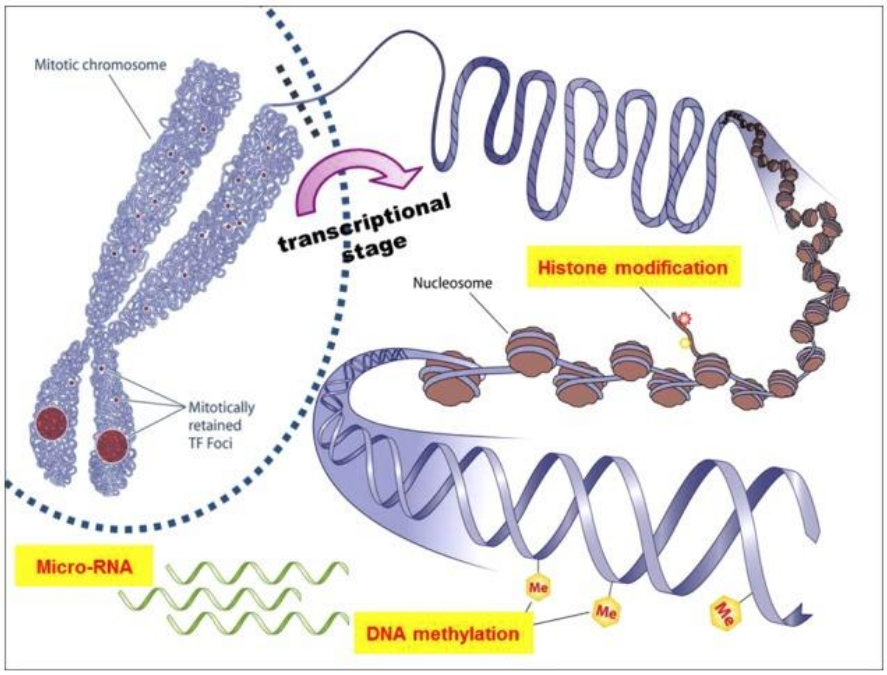


Figure 1.3: Major epigenetic mechanisms that alter gene expression. Histone modifications and DNA methylations directly regulate gene transcription whereas non-coding RNA such as micro-RNA inhibits mRNA translation (Kim, 2014).

1.3 Histone code

Like many other PTMs, histone modifications work coherently. The histone code hypothesis was first made by Strahl and Allis (**Figure 1.4**). The rationale behind the histone code hypothesis is based on some contradictory functions of certain histone modifications. For example, H3S10ph was first found to be a repressive mark during mitosis (Hendzel *et al.*, 1997). However, it was also reported with activation functions upon mitogen stimulation (Thomson *et al.*, 1999). This one modification exhibits paradoxical functions, implying that there is a potential mechanism that either overpowers this mark or alters its function. In fact, the latter scenario was evidenced by ChIP analysis with specific antibodies for both K9ac/S10ph and H3S10ph/K14ac. Both combined PTMs display transcriptional activation upon mitogen stimulation while acetylation becomes the major driver. Missing H3S10 phosphorylation delayed the acetylation process and therefore repressed gene expression (Cheung *et al.*, 2000 and Clayton *et al.*, 2000).

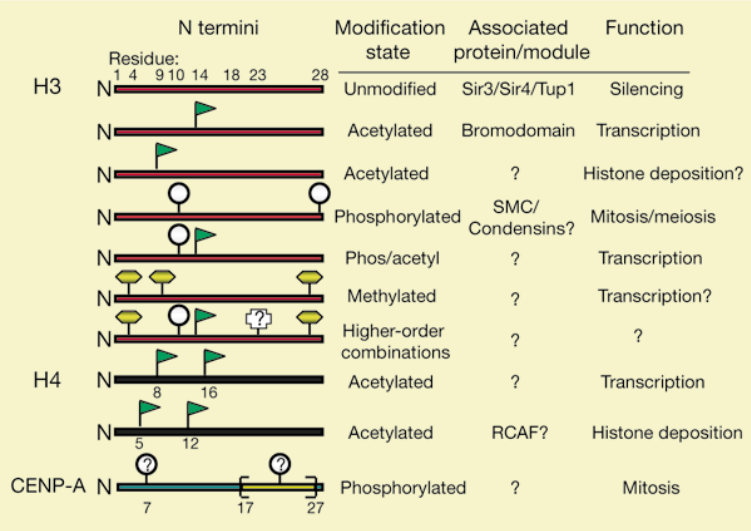


Figure 1.4: Histone modifications that are acting in concert. Normally, histone methylation occurs on arginine and lysine residues; acetylation occurs on lysine; phosphorylation takes place on serine. Combination of histone modifications require a complex reader (either multiple domains or proteins) to recognize them (Strahl and Allis, 2000).

1.4 Histone modification enzymes and epigenetic editors

The first histone modification enzyme was discovered on acetylation. In 1996, both histone acetyltransferase (HAT) and histone deacetylase (HDAC) were identified. The first HAT was identified as a homologue of the yeast protein Gcn5, which is a putative transcriptional activator. As a nuclear A-type HAT, it has a conserved acetyllysine-binding bromodomain that is not present in the cytoplasmic B-type HATs. Hence, it was also the first study that linked transcriptional activation to A-type HAT mediated acetylation (Brownell *et al.*, 1996). In contrast, the discovery of the first HDAC was identified as a homologue to the yeast transcriptional repressor protein Rpd3p. As a result, histone deacetylase was first linked to eukaryotic transcriptional repression (Taunton *et al.,* 1996). Subsequently, great interest arose in the field of histone editors, resulting in a major breakthrough in the chromatin field in late 1990s. The discovery of histone kinases took place in 1999 by multiple groups; histone arginine and lysine methyltransferases were found in 1999 and 2000 respectively; and histone ubiquitin ligase was identified in 2000 (Jin *et al.*, 1999; Thomson *et al.*, 1999; Chen *et al.*, 1999; Rea *et al*., 2000 and Robzyk *et al.*, 2000).

In addition, many histone modification recognition domains – such as acetyllysine-binding bromodomains, methyllysine binding chromodomains, plant homeodomains and Tudor domains – were identified in the late 1990s **(Figure 1.5**). Multiple histone modification readers were reported to be working in the complex for transcriptional regulation, suggesting a higher-order of histone code cross-talking was present. For example, the methyllysine-binding plant homeodomain containing finger protein 6 (PHF6) was co-purified with the nucleosome remodeling and deacetylation (NuRD) complex, which is known for transcriptional repression (Liu *et al.*, 2015).

To date, most of the histone modifiers and readers are called the epigenetic "readers", "writers" and "erasers" (**Figure 1.6**); although many of them have now progressively been identified with non-histone-binding and -modifying properties in cellular signal transduction. Notably, the study of epigenetics, especially PTMs, is no longer limited to histone proteins (Biggar and Li, 2015).

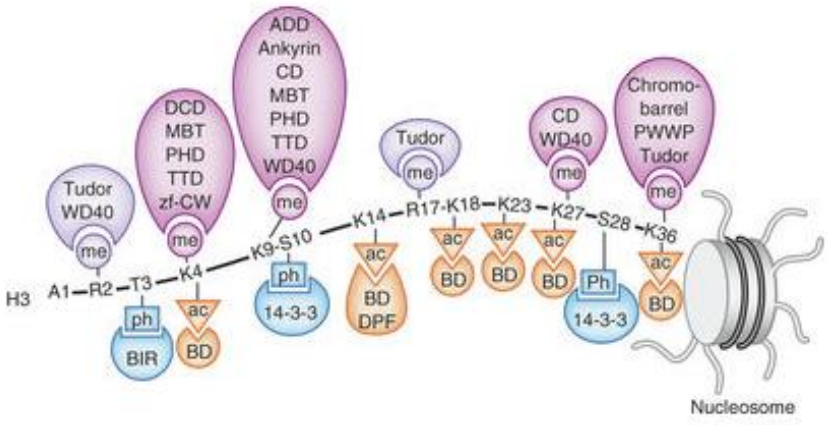


Figure 1.5: Common histone modification readers for methyllysine, methylarginine, acetyllysine, phosphorylated serine and threonine. Most methyl-binding domains contain an aromatic cage that interact specifically with methyllysine and methylarginine by hydrophobic and cation-pi interactions. Some PHD fingers can bind to both methyllysine and unmodified lysine. Acetyllysine is normally recognized by a highly conserved asparagine with hydrogen bonds. Histone phosphorylation is recognized by both hydrophobic interaction and hydrogen bonding with arginine (Musselman *et al.*, 2012).

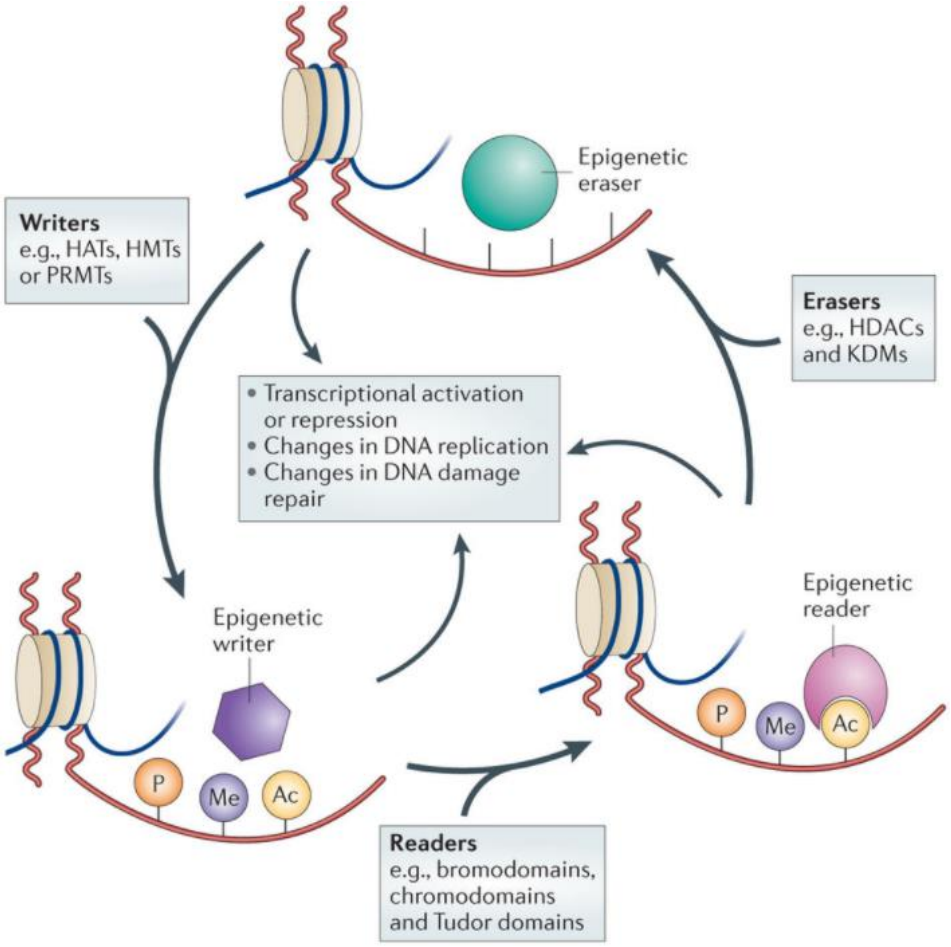


Figure 1.6: Major players in epigenetics - "Readers", "Writers" and "Erasers". Although many histone modification editors are found having non-histone catalytic activities, this figure is referring to enzymes that modify only histone proteins. Besides the enzyme showed here, other known histone "writers" and "erasers" include histone kinases/phosphotases, ubiquitin ligases, sumoylation ligases, ADP-ribosyltransferases and deiminases. Many of the histone "writers" and "erasers" in fact have "reader" property for detecting specific modifications (Falkenberg and Johnstone, 2014).

1.5 KDM5B/JARID1B and related diseases

The histone lysine demethylase KDM5B/JARID1B is a member of the KDM5/JARID1 family. All four family members of JARID1 contain JmjN and JmjC domains for demethylation, and ARID and Zinc finger binding domain for DNA binding. However, the type-3 PHD domain is missing in both KDM5C/JARID1C and KDM5D/JARID1D in humans. To date, the only known histone substrate of KDM5B is histone H3K4 mono-, di- and trimethylations (Klein *et al.*, 2014). KDM5B is known to be an oncoprotein as its expression is usually limited in most normal tissues besides testes and ovaries. Upregulation of KDM5B is commonly found in various human cancers such as breast, prostate, lung and cervical cancers (Hayami *et al.*, 2010). Recent studies also reported that KDM5B is linked to cancer stemness and hypoxia response (Lin *et al.*, 2015 and Salminen *et al.*, 2016).

The most well-studied KDM5B-related disease is breast cancer. Data was observed in previous breast cancer studies, suggesting that the functions of KDM5B in different types of breast cancer could be opposite. For example, in ER^+ breast cancer cell lines such as MCF7, KDM5B overexpression promoted cell proliferations whereas in ER^- MDA-MB 231 cells it suppressed cell proliferations (Hayami *et al.*, 2010 and Yamane *et al.*, 2007). This bivalent behavior suggests that KDM5B may be associated with transcriptional regulatory complexes in a cell-type-dependent manner (Klein *et al.*, 2014).

1.6 Plant homeodomains of KDM5B

As mentioned above, the KDM5/JARID1 family has relatively conserved constituents. The major difference among all family members is the number of PHD domains. KDM5B has a complete set of PHD domains in the KDM5 family (**Figure 1.7**) and its molecular weight is about 180 kDa. The type-1 PHD domain is located between the ARID DNA binding domain and JmjC domain. The type-2 PHD domain and type-3 PHD domains are placed in tandem after the Zinc finger binding domain. Recent study identified that the type-1 PHD domain of KDM5B has preferential binding to unmodified

histone H3K4, whereas the type-3 PHD domain preferentially binds to histone H3K4me2/3(di- and trimethylation). Surprisingly, the type-2 PHD domain has no binding specificity to most of the common histone marks, such as histone H3K9, K27, K36, K79 and the histone H4K20. Yet, binding partner of type-2 PHD domain in KDM5B has not been defined (Klein *et al.*, 2014).

1.7 H3K4me2/3 demethylation by KDM5B

Since the type-3 PHD domain of KDM5B binds to its substrate, it suggests that there is an uncommon demethylation mechanism for the H3K4me2/3. One hypothetical model suggested that both type-1 and type-3 PHD domains need to be anchored into their binding partner for the dimerized JmjN and JmjC domains to catalyze demethylation. However, in this model, the ARID, Zinc finger binding domain and type-2 PHD do not have any binding partners (**Figure 1.8**).

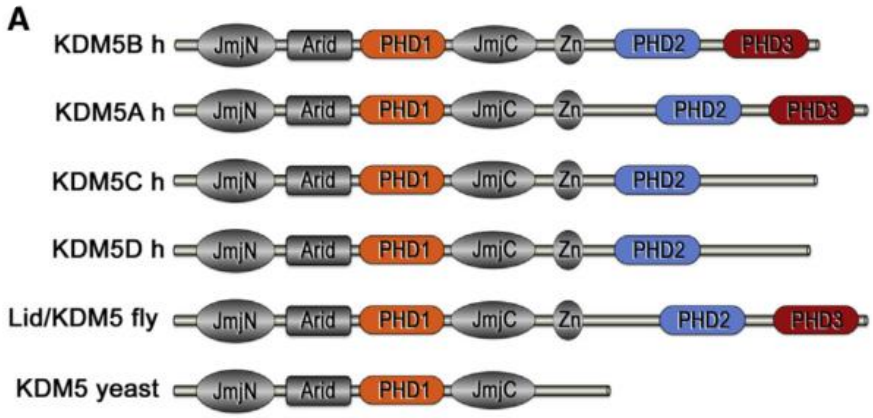


Figure 1.7: Illustration of architectures of KDM5/JARID1 family in human, fruit fly and yeast. The first four components from the N-terminal are most conserved among all KDM5/JARID1 members. Also, high sequence homology of PHD1 indicates that the binding mode of H3K4me0 by PHD1 is the most conserved.

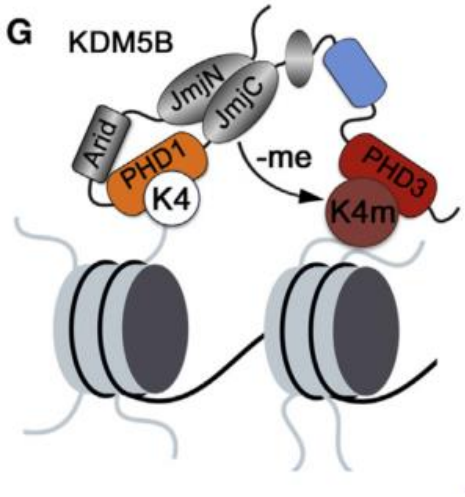


Figure 1.8: Hypothetical KDM5B substrate (H3K4me3) binding model. The coloured K4m circle represents histone H3K4 trimethylation while the K4 circle is unmodified H3K4. The hypothetical mechanism takes place not on the PHD3-bound H3K4me3, but on the adjacent nucleosomes that contain H3K4me3/2. During the demethylation, both PHD1 and PHD3 remain bound, while the catalytic JmjC domain slides alongside, searching for nearby substrates (Klein *et al.*, 2014).

1.8 Stem cell biology and pluripoptency

One significant breakthrough in stem cell biology came in 1981 when the first mouse stem cell line was isolated and maintained in a pluripotent state independently by Gail R. Martin and Matthew H. Kaufman (Martin, 1981; Evan and Kaufman, 1981). Since then, the field of stem cell biology flourished rapidly and led to comprehensive characterizations of stem cell potencies. Upon fertilization, oocyte quickly becomes a single-cell zygote and then a two-cell zygote with the unique activation of its zygotic genome, which is dependent on the typical cytoplasmic environment of oocyte. Accumulated evidence suggested that totipotency starts depleting from the two-cell (2C) stage to early morula (about 8 to 16 cells). Also, cells isolated after the 2C stage have not been successfully developed into a fully functional organism (Lu and Yi, 2015). Embryo beyond early morula is considered plenipotent until early blastocyst. Mouse embryonic stem cells (mESCs) are isolated from the inner cell mass (ICM) of blastocyst embryo (Condic, 2014). Cells at this stage are now widely considered pluripotent, that is, being able to differentiate into any cell types that are found in a mature organism (**Figure 1.9**).

The major challenge in culturing stem cell is maintaining pluripotency because in routine culture media, stem cells undergo spontaneous differentiation once they are isolated from ICM. The first successful attempt used the combination of fetal calf serum and mouse embryonic fibroblast (MEF) feeder layers (Martin, 1981; Evan and Kaufman, 1981). However, at that time, the substances that inhibit stem cell differentiation were not defined, and therefore this practice was highly variable. In late 1980s, Leukemia inhibitory factor (LIF) was first confirmed to have differentiation inhibitory activity as LIF was isolated from Buffalo-rat liver (BRL) cells that were used to substitute MEF (Smith and Hooper, 1987; Moreau *et al.*, 1988). Since then, LIF/serum became a routine medium for culturing stem cell although some mouse embryonic stem cell lines are strictly feeder dependent. Recently, a defined medium was found to have steady and a broad range of differentiation inhibitory effects. Instead of using serum, 2 small molecule inhibitors acting on Mitogen-activated protein kinases (MAPKs) and glycogen synthase kinase 3b (GSK3B) were used. This allowed all types of mESCs to retain pluripotency

without the need of feeder cells (**Figure 1.10**). This serum-free stem cell media is now referred to 2i/LIF (Ying *et al.*, 2008).

In 2006, Shinya Yamanaka and his colleagues demonstrated the regaining of pluripotency in reprogrammed somatic cells. The induced pluripotent stem cells (iPSCs), although inefficient, were the first evidence that somatic cells could be reprogrammed by the overexpression of four essential genes: Oct4. Sox2, Klf4 and c-Myc (Takahashi and Yamanaka, 2006). Later studies showed that this reprogramming did not require all four genes and they could be replaced (Okita *et al.*, 2008). Also, co-expressing other transcription factors such as Zic3 and Esrrb with these four genes showed increased reprogramming efficiency (Sone *et al*, 2017). Together, these discoveries suggested some degree of redundancy in the maintenance of stem cell pluripotency.

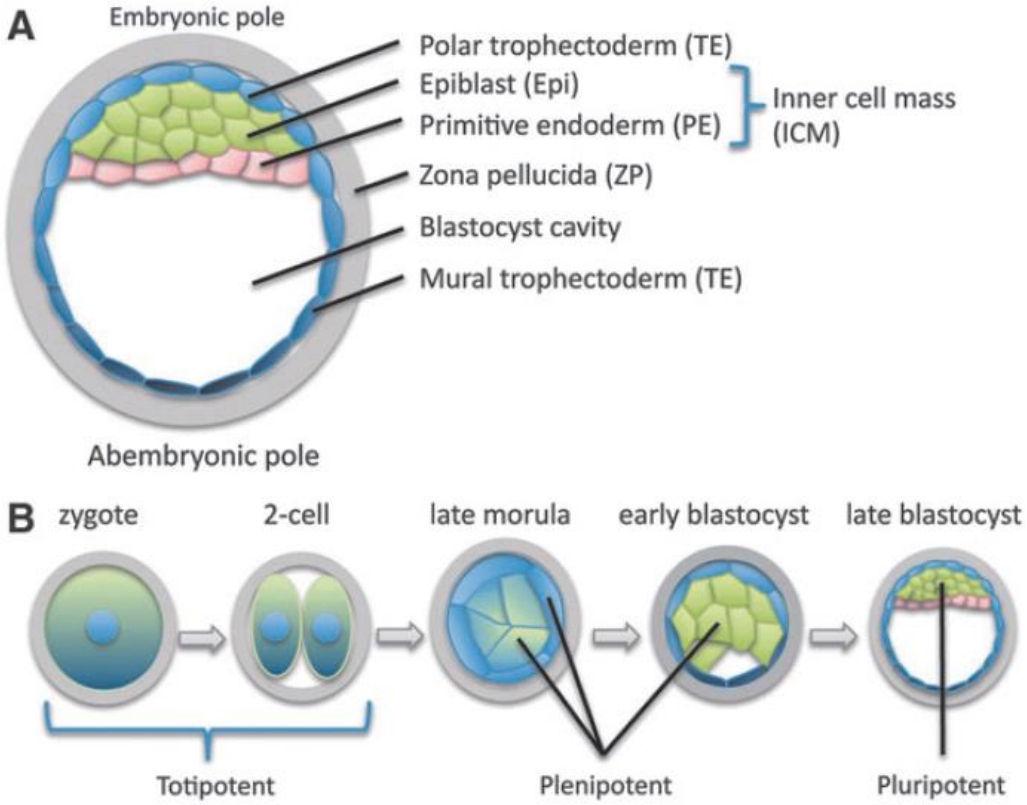


Figure 1.9: Architectures of early embryonic development. A) Annotated human blastocyst at 5-6 days. Embryonic stem cells are taken from ICM, which are considered pluripotent but not plenipotent since they are not able to differentiate into either the polar trophectoderm (PE) or mural trophectoderm (TE). Cells descended from ICM also exist in postnatal life. Both TE and PE are considered as extraembryonic stem cells and they differentiate into placental tissues (gestational functions). **B)** Chronological development stage from single cell zygote to late blastocyst. Despite only 2-C stage showing totipotent here, totipotency can remain until early morula stage (8-16 cells). In the developmental timeline, plenipotent stem cells can turn into ICM, PE and TE, that is, both embryonic and extraembryonic stem cells (Condic,2014).

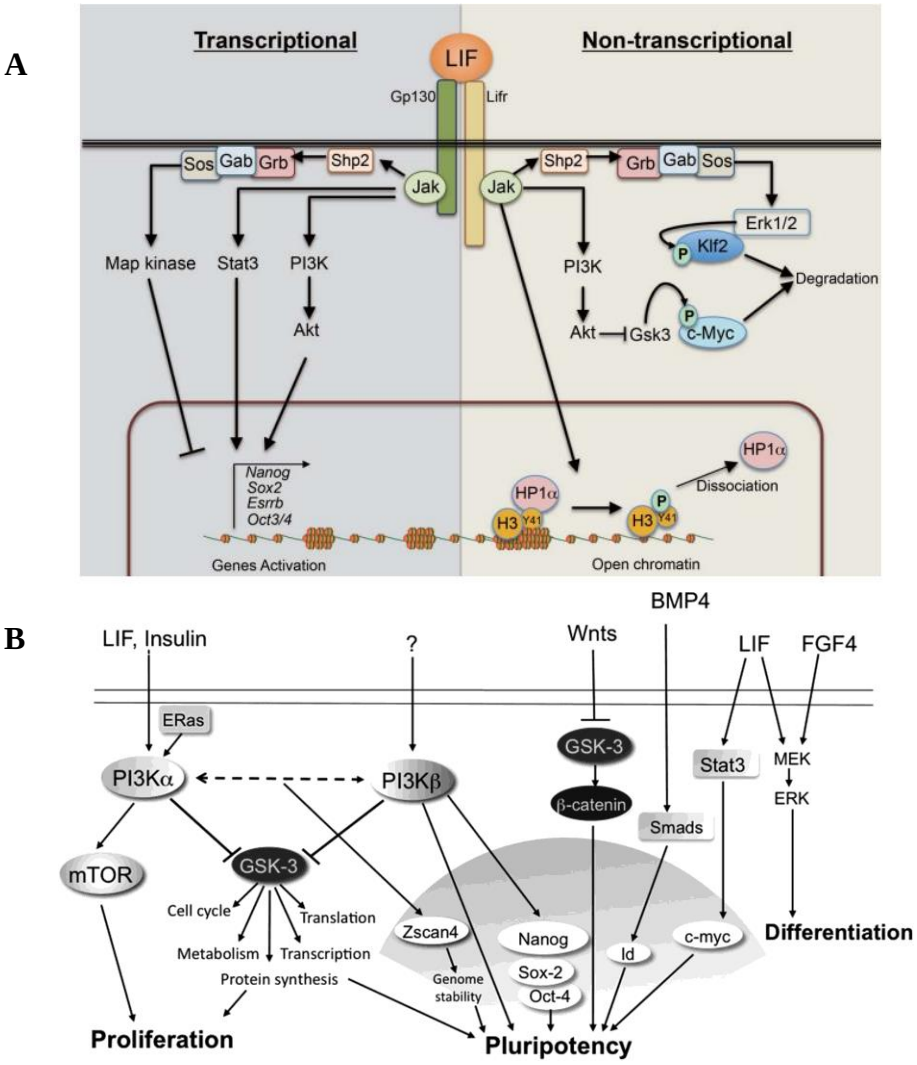


Figure 1.10: Schemes of LIF induced JAK -STAT signaling and 2i/LIF induced signaling. A) JAK-STAT activation by LIF. Upon LIF stimulation, PI3K-AKT, STAT3 and MAP Kinase are activated. The activation of STAT3 by JAK is known to sufficiently maintain pluripotency. Activation of PI3K-AKT can increase proliferation. The activation of MAP kinase can trigger differentiation but this effect is overwritten by STAT3 activation. **B)** Differentiation inhibitory activity with MEK and GSK3b inhibitors. Prolong activation of MAP kinase can induce differentiation. By including MEK inhibitor, it reduces the risk of LIF-induced differentiation. GSK3 inhibitor is used to mimic the activation of Wnts, which can further stabilize beta-catenin. The downstream responder is TCF/LEF, which is important for c-MYC and cycD expression. (Melanie *et al.*, 2011 and Ohtsuka *et al.*, 2015)

1.9 The role of KDM5B in neurogenesis

KDM5B/JARID1B has its known substrates H3K4me3/2, which are marks for gene activation. Since KDM5B removes H3K4 methylation, losing KDM5B/JARID1B can lead to potential latency of gene deactivation and misoriented gene expression. Complete knockdown of KDM5B is embryonic lethal between E4.5and E7.5, which is a relatively early embryonic stage right after implantation, suggesting severe developmental defects can take place (Catchpole *et al.,* 2011). Recent transcriptional and ChIP analysis revealed that, depletion of KDM5B introduced significant upregulation of self-renewal genes, thereafter impairing the differentiation landscape. Interestingly, genome-wide study showed that KDM5B and its substrate H3K4me3/2 co-occurred at same promoters, indicating that KDM5B had sole chromatin binding ability besides its catalytic function. Also, it is reported that KDM5B occupies the promoters at Nanog and Oct4. However, KDM5B itself did not play a critical role in pluripotency maintenance or self-renewal (Kidder *et al.*, 2014; Schmitz *et al.,* 2011 and Xie *et al.*, 2011).

Previous study found that KDM5B knockdown greatly impairs neuronal differentiation in mESCs by blocking the expression of genes in the neuronal axis, therefore resulting in the abnormal formation of embryoid bodies (Dey *et al*., 2008). Transcriptional studies identified that self-renewal associated genes of KDM5B knockdowns such as Oct4 and Nanog failed to downregulate during differentiation, which is in line with the previous findings. Cells with KDM5B knockdown do not pass differentiation day 8 after retinoic acid induction, and no apparent neuronal networks were observed. In contrast to the early knockdown, induced knockout of KDM5B after E12.5 show little to no effect to cell survival and later neuronal development. Intriguingly, the induced knockout of KDM5B also displayed no effect on H3K4me3 level, suggesting that demethylation of H3K4me3/2 by KDM5B acts in a developmental stage dependent manner (Schmitz *et al.*, 2011).

1.10 Previous study and preliminary data

Discovery of H2BK43me2

A previous study from our group in 2010 showed that methylated H2BK43 interacts with chromatin binding-modules (CBMs). Although the reason this location was chosen for study was not well explained, CBMs binding of this site were found to preferentially occur on dimethyllysine but not on mono- or trimethyllysine (Liu *et al.*, 2010). Also, in our recent unpublished data, H2BK43me1 was predicted to be a potential histone methylation site with high confidence (P-value 0.824, **Table 6**) by deep learning algorithms (Biggar *et al.*, unpublished data).

Although our previous works had provided biochemistry foundations, under cell physiological conditions, validation of KDM5B mediated H2BK43me2 demethylation was still required. Unlike histone H3 and H4, histone H2B has relatively shorter N-terminal tail. Many of the well-studied sites on histone H3 are located at its N- or C terminal, such as K4, K9, K27, K36 and K79, where they are freely accessible by other transcription factors or modification enzymes (Bhumik *et al.*, 2007). However, H2BK43me2 is located in its globular domain bound by the DNA double helix, which means it has less exposure to nucleoplasm. In fact, X-ray crystallography of histone H2B showed that lysine at position 43 is placed very close to the phosphodiester backbone (**Figure 1.11**), indicating there may be a binding between K43 and the DNA double helix which also limits its exposure to proteins in the nucleoplasm. Nevertheless, in our previous study (**Figure 1.12**), nucleosomes were dissolved in 1M guanidine hydrochloride (pH 7.6). Although this is not an extremely harsh condition, DNA double helix may be interrupted and H2BK43me2 exposed.

Our aim here was to confirm that histone H2BK43me2 was a true substrate of KDM5B under cell physiological conditions. To solve this problem, we adopted both human breast cancer cell lines and mESCs. Many studies showed that KDM5B was playing important roles in neuronal differentiation (Dey *et al.*, 2008 and Schmitz *et al.*, 2011). If

H2BK43me2 is one of the targets of KDM5B, its methylation level should be affected by the expression of KDM5B. Also, KDM5B overexpression is reported in MCF7 breast cancer cells (Bamodu *et al.*, 2016). KDM5B knockdown in MCF7 cells may cause accumulation of H2BK43me2. Therefore, these two systems are suitable for examining whether H2BK43me2 is a substrate of KDM5B or not.

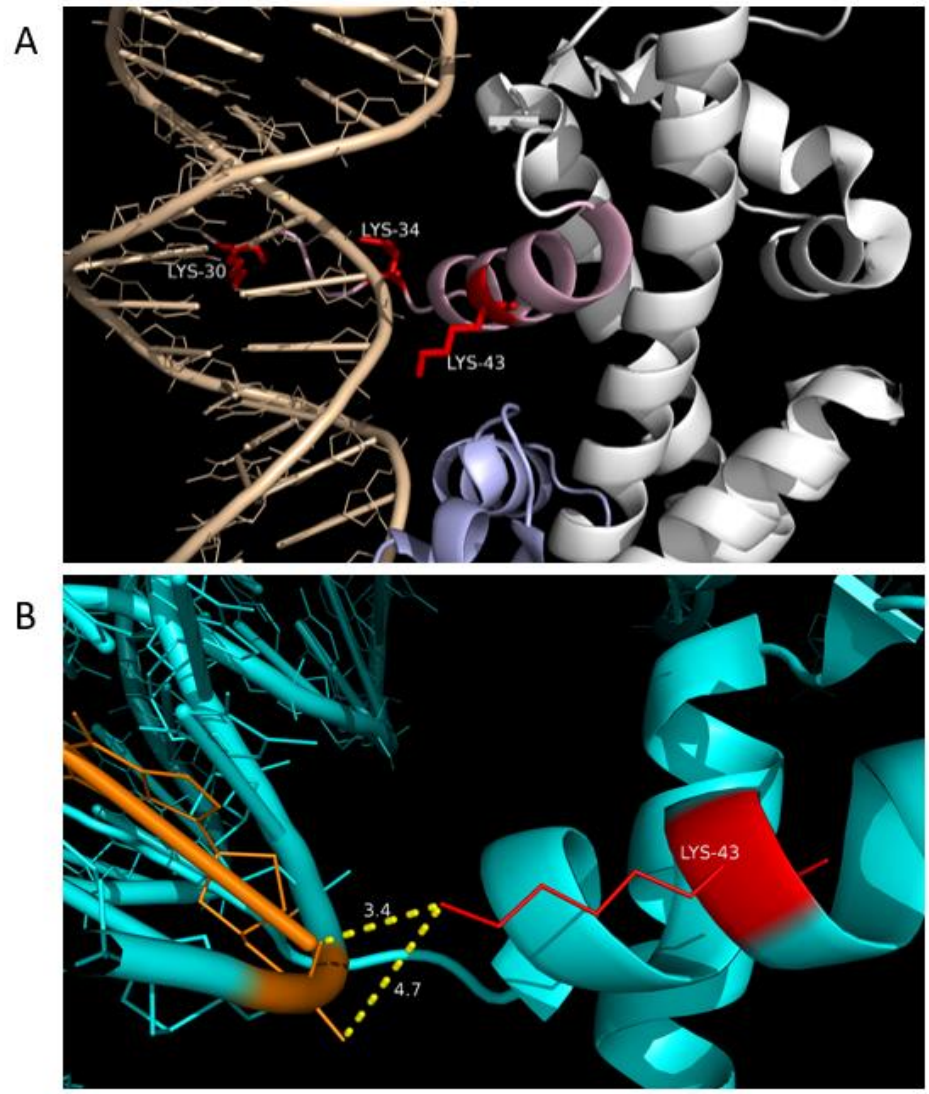


Figure 1.11: Structure of nucleosome in cartoon representation (PDB code 5AV8). A) H2A (grey and purple) - H2B (grey and pink) dimer associated with DNA double helix (beige) with lysine at position 43 of H2B showed in red stick. DNA double helix wind around the H2A – H2B dimer, thus making the H2BK43 less accessible to other proteins. **B)** Close-up image showing distance (indicated by dashed line) between H2B lysine 43 (red) and DNA double helix (cyan). The shortest distance between the positively charged ε-ammonium group and the negatively charged phosphodiester backbone (orange) is measured as 3.4 angstrom, suggesting there is potential for an interaction.

H2BK43me2 as a preferred substrate of KDM5B *in vitro*

Since we predicted that KDM5B may have an alternative substrate other than H3K4me3, we further screened peptides listed in our 2010 publication. Surprisingly, previous results from our group could not confirm histone H3K4me3 was the preferred substrate of KDM5B; instead, our data suggested that H2BK43me2 was the preferred substrate of KDM5B compared to H3K4me3 *in vitro* (**Figure 1.12 and Figure 1.13**).

The histone binding pattern of KDM5B had been examined recently **(Figure 1.14**). So far, most of known sites were identified on histone H3 and H4; yet not many studies had been done on histone H2B. Knowing that KDM5B could catalyze H2BK43me2 demethylation did not unveil enough details of how this recognition was achieved. Therefore, to complete the biochemistry study, we aim to elucidate the binding mechanism between KDM5B and H2BK43me2.

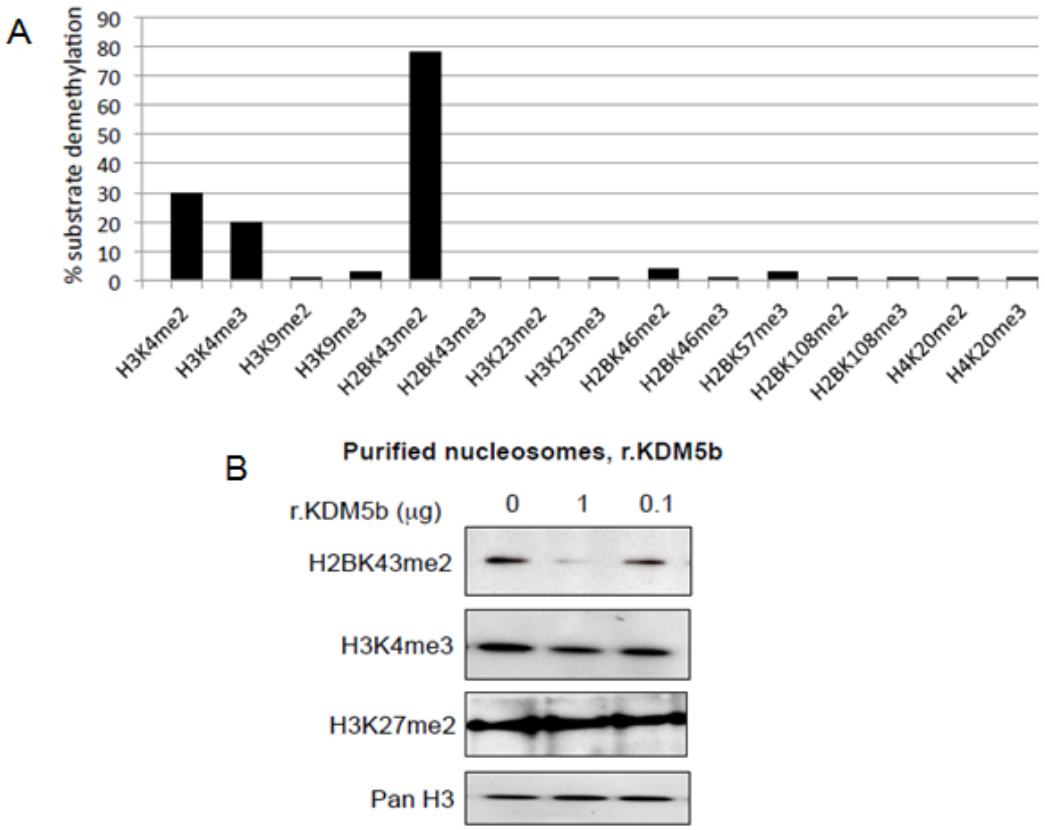


Figure 1.12: H2BK43me2 is an alternative substrate of recombinant KDM5B *in vitro*. A) Quantitative MS analysis reviewed substrate demethylation ratio. r. KDM5B was used to catalyze demethylations of synthetic peptides. After 60 minutes of reaction, remaining peptides were quantized by MS analysis. Bars represent the percentage of substrates catalyzed by r.KDM5B. H3K4me2 and H3K4me3 served as positive control in this analysis. **B)** r.KDM5B acted on purified nucleosomes. Purified nucleosomes were incubated with different amount (0, 0.1 and 1 μg) of r.KDM5B for 60 minutes. After reaction, proteins were separated by SDS-PAGE and stained with corresponding antibody for western blot analysis. Pan-H3 was used as reference of loading amount (Stalker *et al.*, unpublished data).

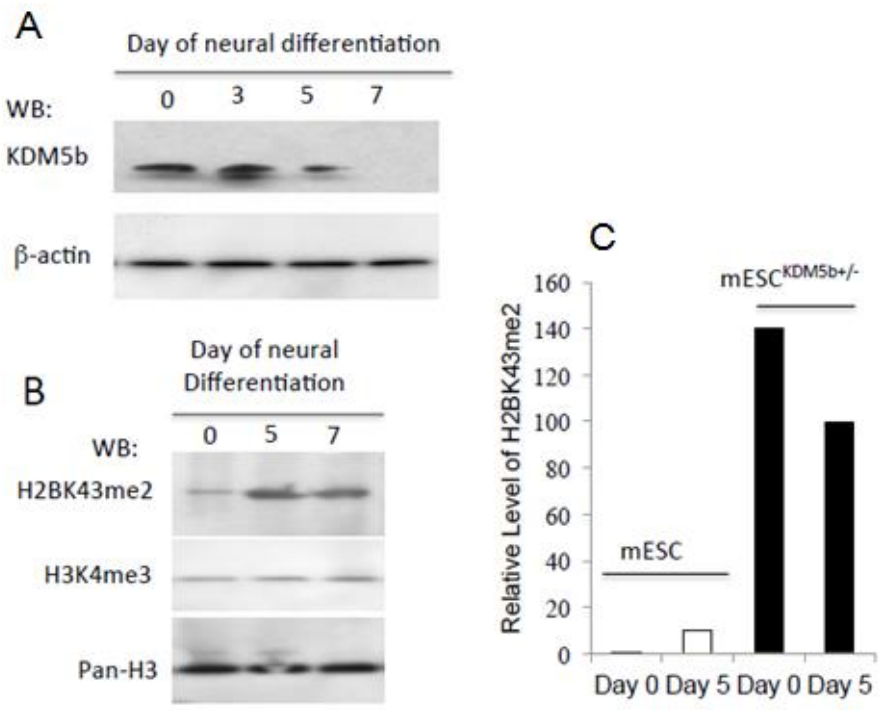


Figure 1.13: H2BK43me2 is a substrate of KDM5B under cell physiological conditions. A) KDM5B downregulation during neuronal differentiation. mESCs were induced to neurogenesis in specific media for 7 days. Total cell lysates were collected at day 0, 3, 5 and 7 for immunoblotting. Beta-actin was used as reference of loading amount. **B)** Accumulation of H2BK43me2 during neuronal differentiation. Same neuronal induction process was conducted but histone extracts were only collected at day 0, 5 and 7 for immunoblotting. Pan-H3 was used as reference of loading amount. **C.)** Quantitative MS analysis of H2BK43me2 level during differentiation in KDM5B knockdown and WT cells. Histone extracts collected at day 0 and 5 were sent for MS-MRM analysis to determine relative amount of H2BK43me2 in both WT mESCs and KDM5B knockdown mESCs. Bars represent the relative amount of H2BK43me2 to the corresponding H2BK43me2 standard peptide (Stalker *et al.*, unpublished data).

.

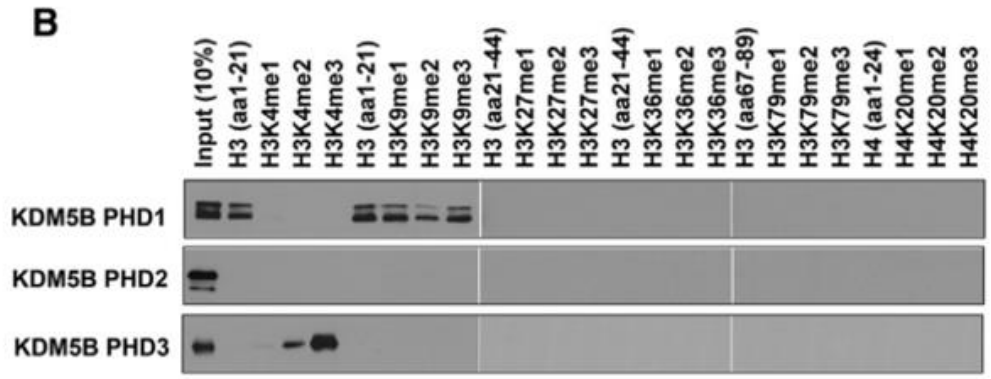


Figure 1.14: Pull-down analysis of binding preference of different KDM5B PHD domains. PHD domains were pulled down by streptavidin beads with biotinylated histone peptides. KDM5B PHD1 domain bound to unmodified H3 (aa 1-21) regardless of the condition of H3K9 methylation. KDM5B PHD2 domain had no known histone binding partner. KDM5B PHD3 domain specifically interacted with H3K4me3/2 (Klein *et al.*, 2014).

1.11 Hypothesis and aims

Given the fact that KDM5B mediated H3K4me3/2 demethylation is conditional and both KDM5B and H3K4me3/2 have same promoter occupancies, we hypothesize that there is an alternative chromatin binding partner or substrate of KDM5B present. To test our hypothesis, we used synthetic peptides to search for potential binding partners of KDM5B. Also, we investigated all three PHD domains of KDM5B separately to dissect the biochemistry of the binding partner recognition mechanism. To this end, we applied both peptide pull-down assay and fluorescence polarization assay. We also extended our research into cell physiological conditions. We generated both stable and transient KDM5B knockdown cell lines as well as differentiated stem cells to study the dynamics of histone methylation.

Our aims:

1. To identify the alternative binding partners of KDM5B

2. To examine whether KDM5B mediates demethylation of its alternative binding partners.

Chapter 2

Methods and results

2.1 Methods

2.1.1 Bacterial culture and protein purification

All three PHD domains (PHD-1, aa303-363; PHD-2, aa 1168-1239 and PHD-3, aa 1481-1538 of human KDM5B - UniProtKB -Q9UGL1) that were cloned into vector pGEX-6P-1 (gifts from Dr. Shi Xiaobing). In brief, plasmids amplified in *E. coli* DH5-α were purified using EZ-10 Spin Column Plasmid DNA Miniprep Kit (Bio Basic) according to manufacturer's instruction. Proteins were expressed in *E. coli* Rosetta-2 DE3 (Novagen) with Ampicillin (50 μg/ml) and Chloramphenicol (20 μg/ml) selection in Luria Broth (Bio Basic). Before induction, bacterial concentration was measured by OD600 in the range of 0.6~0.8 A; 0.5 mM IPTG and 150 μM $ZnCl_2$ were used for induction. After incubation at 18ºC for 16 hours, bacteria were first pelleted by 8000 rpm centrifugation at 4 ºC for 20 minutes then washed twice with ice-chilled PBS. Cell pellets were lysed by 1% Triton-X100 (Sigma), 10 unit/ml benzoase (Sigma), 1mg/ml lysozyme (Bio Basic), 1 mM PMSF in PBS at 4 ºC with mild agitation for 1 hour. Clear bacterial lysates were collected by 12000 rpm centrifugation at 4 ºC for 20 minutes. GST-tagged proteins were then purified on Glutathione Resin (GenScript) following manufacture's instruction. Eluted proteins were dialyzed in 20 mM Tris pH 7.4, 140 mM NaCl and 1 mM dithiothreiol (DTT) at 4 ºC overnight. Purified GST-tagged proteins were either concentrated by 10 kDa filters (Amicon) or used directly after concentration measurement. The quality of proteins was assessed by Coomassie staining SDS-PAGE.

2.1.2 Peptide synthesis

Peptides were synthesized on Tentagel resins (Intavis) by Intavis-AG MultiPep peptide synthesizer with Fmoc (N-(9-fluorenyl) methoxycarbonyl) chemistry (Carpino and Han 1970). In brief, peptides were synthesized in 2 μmol scale from their C-terminal. Either

biotin or 6-Carboxyfluorescein were attached to the 6-aminohexanoic acid at the N-terminal, enabling these peptides to pull-down and fluorescence polarization experiments. Post-synthetic peptides were cleaved by mixture of trifluoroacetic acid (TFA) and triisopropysilane (TIPS) then precipitated in -20 ^{o}C ether following manufacture's instruction. Peptide pellets were washed again with -20 ^{o}C ether twice and dried in fume hood before resuspended in distilled water. List of peptides used in this study were showed in **Table 1**.

2.1.3 Matrix Assisted Laser Desorption/Ionization (MALDI)

Quality of synthetic peptides were assessed by MALDI-TOF using Micromass M@LDI mass spectrometry (Waters). Briefly, 1 μl of undiluted peptide solution was mixed with 1 μl of provided matrix (10 μg/μl α-cyano-4-hydroxycinnamic acid in 3/5 CH_3CN/1% formic acid v/v). 2 μl of mixtures was doted and air-dried on MALDI plate before analyzed by mass spectrometer (see **appendix A** for example).

2.1.4 Mammalian cell culture and subcellular fractionation

MCF7 cells were culture in DMEM (Sigma) supplemented with 10% FBS (Wisent), 2 mM L-glutamine (Gibco), 100 U/ml penicillin-streptomycin (Gibco), 2 mM sodium pyruvate (Gibco) at 37 ^{o}C humidified incubator with 5% CO_2. Cells were passaged at subconfluent condition (80%) to avoid senescence. Cells were wash with room temperature PBS before using 0.1% trypsin (Gibco) for 5 minutes at 37 ^{o}C. Subcellular fractionation protocol was mainly adopted from Schechter *et al.*. In brief, trypsinized cells as described above were first wash with ice-chilled PBS twice at 1000 rpm for 5 minutes. Cell pellets were lysed in ice-chilled hypotonic buffer ($1 * 10^7$ cells/5ml) containing 10 mM Tris-Cl pH 7.4, 3 mM KCl, 1.5 mM $MgCl_2$, 0.1 % NP40 and 1 mM PMSF at 4 ^{o}C with mild agitation for 10 minutes. Supernatant (S1) was collected as cytoplasmic fraction by spinning at 2100 rpm, 4^{o}C for 5 minutes. Pellets after the first lysis containing mainly nucleus were further lysed by 3 mM EDTA, 0.2 mM EGTA, 1 mM PMSF and 0.2 % Tween-20 (Sigma). Supernatant (S2) was saved as nucleoplasmic fraction by spinning at 10000 rpm, 4 ^{o}C for 10 minutes (Schechter *et al.* 2007). Combined S1 and S2 was used for histone peptide pull-down analysis.

2.1.5 GST and streptavidin pull-down analysis

Histone peptide pull-down protocol was adopted from a publish protocol with minor modifications (**Figure 2.4**). In brief, biotinylated histone peptides were first incubated with ice-chilled PBS washed streptavidin beads (GE) at 4^{o}C with mild agitation for 2 hours. 10μg of peptides was immobilized by 40μl slurry. Subsequently, pre-bind beads were washed with ice-chilled PBS twice then PBS-T once. Beads were washed at 4^{o}C, 500 rpm. Beads were then resuspended in enough PBS with 1mM PMSF to achieve 50% slurry prior to use. For the subcellular lysate pull-down, lysates were pre-cleared by adding streptavidin beads (10μl slurry to 1ml lysate) and incubate at 4^{o}C with mild agitation for 1 hour. Pre-cleared lysates were subsequently mixed with different histone peptide conjugated streptavidin beads and incubate at 4^{o}C with mild agitation overnight. For the purified GST-tagged protein pull-down, pre-cleared GST-tagged proteins were first incubated with different histone peptides at 4 ^{o}C with mild agitation overnight (10 μg peptides to 1 mg protein). Peptide-protein complexes were then pull-down by adding streptavidin beads (40 μl slurry to 10 μg input peptides) at 4 ^{o}C with mild agitation for 2 hours. Beads were later collected at 4 ^{o}C, 500 rpm and washed twice with ice-chilled buffer containing 20 mM Hepes pH 7.9, 200 mM KCl, 0.1% Triton-X100 and 1 mM PMSF. Proteins on beads were eluted by boiling in 2X Laemmli buffer for 5 minutes (Wysocka, 2006).

2.1.6 SDS-PAGE and Coomassie staining

SDS-PAGE and Coomassie stain recipes were obtained from protocols published by Cold Spring Harbor Laboratory. Typically, 10% gels were used for protein separation. Running condition for SDS-PAGE was 120 V for 15 minutes following by 180 V for 45 minutes (Bio-Rad Mini-Protean) in 1X Towbin buffer. Gels were then incubated with Coomassie stain at room temperature for 30 minutes and destained in water overnight.

2.1.7 Western blot

After running SDS-PAGE, gels were transferred onto methanol activated PVDF membrane in 1X Towbin buffer plus 10% methanol via Trans-Blot® Semi-Dry (Bio-Rad) using constant milliampere (1 mA per cm^2) for 90 minutes. Transferred membrane was first block with TBST plus 2% BSA for 1 hour at room temperature with mild agitation. Blocked membrane was then probed with TBST, 1% BSA and 0.1% sodium azide diluted primary antibody according to specific working dilution. Typically, primary incubation was carried out at 4 ^oC with mild agitation overnight. Membrane was then washed with TBST twice for 10 minutes at room temperature with mild agitation before adding secondary antibody (1:2500 dilution in TBST plus 2% milk) and incubated for 2 hours at room temperature with mild agitation. After secondary incubation, membrane was washed again with TBST twice for 10 minutes at room temperature with mild agitation. Finally, membrane was developed using home-made substrate at room temperature for 2 minutes (solution A: 2.5 mM Luminol, 400 μM p-Coumaric acid and 100 mM Tris-Cl pH8.8; solution B: 0.02% Hydrogen Peroxide in 100 mM Tris-Cl pH8.8, all chemical purchased from Sigma). Images were taken with ChemiDoc™ XRS+ system (Bio-Rad).

2.1.8 Fluorescence polarization assay

Dialyzed GST-tagged proteins were adjusted concentration to a master stock of 200 μM. Different concentration of GST-tagged proteins were made from master stock. 6-FAM-tagged histone peptides concentration were measured according to fluorescence intensity. In principle, fluorescence signal intensity in each well was adjust to between 20000 to 60000 units when measuring emission at 520 nm with 35 μl total volume in 384-well plate on Perkin-Elmer 2103 multilabel plate reader. Adjusted histone peptides were incubated with GST-tagged proteins in dark for 10 minutes before taking fluorescence polarization measurement. Fluorescence polarization measurement was taken with 480 nm excitation and 520 nm emission. Biding curves and corresponding dissociation constants (K_d) were generated by Prism 3.0 software (GraphPad software, Inc.).

2.1.9 Mass spectrometry

Histone extraction was performed as describe below. Briefly, 20 μg histone extract was added in ABC pH 7.9 buffer. MS-grade Trypsin (Promega) was used for digestion according to manufacturer's instructions. Digested mixtures were cleaned by homemade C-18 column using Empore C18 Disk and subsequently eluted by 70% ACN in MS-Grade acidified water with 0.15 formic acid (Fisher). Elutes were then dried using speedvac to remove ACN residue. Prior to injection, samples were redissolved in MS-grade water with 0.1% FA. Mass spectrometry was carried on the Q Exactive Hybrid Quadrupole-Orbitrap (Thermo) using an EASY-Spray Column PepMap RSLC (C18, 75 μM x 500 mm) in 100% ACN and 0.1% FA. Data collected was processed by PEAKS (Bioinformatics Solution Inc.)

2.1.10 Cell culture and neuronal differentiation

Cell culture protocol for MCF7 cells was same as the method mentioned in the former chapter. Mouse embryonic stem cells E14 were cultured using feeder-independent protocol (Bibel *et al.*, 2007). In brief, mESCs were cultured on 0.1% gelatin-coated tissue culture treated dishes in DMEM (Sigma) supplemented with 15% ES grade FBS (Wisent), 2 mM sodium pyruvate (Gibco), 2 mM L-glutamine, 100 U/ml penicillin-streptomycin (Gibco), 0.1 mM 2-mercaptoethanol (Sigma), 1X NEAA supplement (Gibco) and 1000 U/ml LIF (Stem cell biotechnology), at 37 ^{o}C humidified incubator with 5% CO_2. Cells passaged at sub-confluent stage of 70% to avoid stem cell colonies touching each other. To passage, cells were first washed with PBS with 0.1 mM EDTA twice. Typically, 1 ml of 0.1% Trypsin was added to a 10-cm dish with 5 minutes incubation at 37 ^{o}C. Complete dissociation of clumps should be achieved to avoid spontaneous differentiation after passage. Neuronal differentiation protocol was adopted from our lab protocol. mESCs were grown in mESC media until cells became highly proliferative (doubling time less than 20 hours). Highly proliferative mESCs were lifted and 4 x 10^6 cells were seeded in bacterial Petri dish with 10 ml embryoid body media containing DMEM, 1% FBS (Gibco), 1X B-27 supplement (Gibco), 2 mM L-glutamine, 1X NEAA supplement, 100 U/ml penicillin-streptomycin and 2 mM sodium pyruvate. Cells were left untouched for

48 hours to allow formation of embryoid bodies. Embryoid bodies were then plated onto 1% gelatin coated tissue culture treated dishes (1:4 ratio) with neuronal differentiation media consisting of 5% FBS, 1X B27 supplement, 2 mM L-glutamine, 100 U/ml penicillin-streptomycin and 2 mM sodium pyruvate. Cells were maintained in this condition for further differentiation days.

2.1.11 Histone extraction

We followed the histone extraction strategy from Shechter *et al.* with minor modifications. Cells were first collected by washed by ice-chilled PBS twice to remove FBS completely. Typically, 1 x 10^7 cells were lysed with 5 ml ice-chilled lysis buffer containing 0.5% NP40 (Sigma), 10 mM Tris-Cl pH 7.6, 140 mM NaCl, 1.5 mM $MgCl_2$, 1 mM PMSF, on ice for 10 minutes with occasion inversions. Nuclei were pelleted by centrifugation at 2100 rpm, 4^{o}C for 5 minutes. 1 ml of 0.4N H_2SO_4 was added to the pellet and then incubated for 1 hour on ice. Supernatant was collected by centrifugation at 10000 rpm, 4^{o}C for 5 minutes. In a new Eppendorf tube, mix 250 μl of 100% TCA (w/v) with each 1 ml of supernatant for about 30 minutes to precipitate the histone proteins. Precipitates were pelleted by centrifugation at 14000 rpm, 4^{o}C for 10 minutes. Pelleted histones were first washed by acidified acetone (0.01% HCl, stored at -20 ^{o}C) then twice with 100% acetone (stored at -20 ^{o}C). Lastly, the pellet was air dried and resuspended in water. Quality of histone was assessed by SDS-PAGE and Coomassie staining (showed in **appendix C**).

2.1.12 Antibody validation

Methyllysine antibodies were assessed by dot-blot assays using diluted histone peptides (**Table 1**). Typically, histone peptides were dotted on nitrocellulose membrane and dried at room temperature. Membrane was then blocked with 3% milk at room temperature for 30 minutes and probe with corresponding antibody for two hours at room temperature subsequently. Procedures beyond this step were performed as regular western blot. Antibodies used in this study were shown in **Table 5**.

2.1.13 Plasmid purifications and cloning

All bacterial cultures, except for lentiviral plasmids, for plasmid (**Table 3**) production were grown as describe above. The lentiviral pLKO.1 vector was amplified at 32 ^{o}C overnight before purification. Plasmids were purified as showed above. Plasmids which were used for electroporation - histone H2BK43WT-3x Flag, H2BK43-A-3x Flag and H2BK43-R-3x Flag - were cloned in pEGFP (with GPF portion removed). In brief, oligonucleotides of H2BWT and mutants were purchased from Bio Basic and digested with NheI and BamHI (Thermo) followed by T4 ligase (NEB) mediated ligation into pEGFP. Constructs were then sequenced using on campus service; redesigned sequences of H2BWT and mutants were shown in **appendix D**. KDM5B constructs were created by Gateway cloning using Gateway LR Clonase II (Invitrogen). KDM5B-pENTR3C was subcloned into both pMAX-DEST and pLenti-CMV-Puro-DEST followed manufacturer's instructions. Plasmids for electroporation purposes were purified using QIAprep (QIAGEN) and manufacturer's protocol was followed. Most of plasmids in our study were sequenced with exception for the KDM5B overexpression construct.

2.1.14 Transfection and viral infection

All shRNA and siRNA used in this study were listed in **Table 2**. For siRNA transfection, 5 μg of siRNA (Sigma) was mixed with 50 μl X-tremeGENE siRNA transfection reagent (Roche) in 1 ml opti-MEM (Gibco) followed by 5 minutes incubation at room temperature. The mixture was added onto cells culturing in 6 cm dish to achieve total volume 2.5 ml. Transfection results were assessed after 72 hours incubation followed routine cell culture practices mentioned above. shRNA constructs were delivered by viral infection. In brief, Lentiviral particles were produced in HEK293T cells (provided by Dr. Zhao Lin). shRNA containing pLKO.1, packaging and envelop plasmids (1:2:2 molar ratio, **Table 2**) were delivered by X-tremeGENE HP transfection reagent (Roche) according to manufacturer's protocol. Media was replaced 24 hours post-transfection; viral particles were collected 48 hours post-transfection and filtered through 0.44 μm filter. Cells were infected with viral lysates with presence of polybrene (8 μg/ml, Sigma)

and selected in puromycin media (3 μg/ml, Invitrogen) 48 hours post-transfection for 5 days. Due to the hard-to-transfect nature of mESCs, a combined method of electroporation and cationic transfection was used. Cells were harvested and pooled in ice-chilled Gene Pulse 0.4 cm cuvettes (Bio-Rad). Typically, 40 μg linearized DNA was used in 1 x 10^7 cells in total volume of 1ml ice-chill PBS. Electroporation condition used was 220 V at 440 μF for two pulses. Cells were resuspended fully before seeded onto MEF coated dish. A reversed transfection (see transfection procedures above) was also performed before cells completely attached to MEF. Typically, 10 μg of linearized plasmids was used for 10 cm dish. Cells were selected 48 hours post-transfection for 7 days using neomycin (500 μg/ml, Bio Basic). For single cell colony isolation, cells were first limited diluted to 0.5 cells/well in 96-well plate. After 12 days incubation, single cell colonies were picked and transferred into 24-well plate supported with MEF for further expansion. Single cell colonies were examined by western blot to confirm the expression.

2.1.15 RT-qPCR

mRNAs were prepared using Absolutely Total and mRNA Purify Kits (Agilent) following manufacturer's protocol. mRNA quality was evaluated by both OD260/280 and 28S/18S ratio. To assess the integrity, crude mRNAs were loaded into 1% bleach agarose gel to look for degradation (example showed in **appendix E**). 2 μg of mRNAs was reverse transcribed into cDNA using iScript cDNA synthesis kit (Bio-Rad). cDNAs were further diluted 1:40 in water for qPCR analysis. Typically, 2 μl of diluted cDNAs were mixed with 6 μl of iTaq universal SYBR supermix (Bio-Rad), 2 μl ddH_2O, and 2 μl primer mix (3 μM). We used the Hard-Shell 96-well PCR Plates (Bio-Rad) for loading and the Bio-Rad CFX connect for qPCR. Data process was completed via CFX Manager (Bio-Rad). All primers used in this study were listed in **Table 4**.

2.1.16 Immunofluorescence and confocal microscopy

For immunofluorescence, cells were grown on glass-bottom dishes (Corning) to about 30% confluent. Cells were first washed with PBS twice then fixed in 4% paraformaldehyde at room temperature for 10 minutes (4 ml per 60 mm dish). Fixed cells were washed with

PBS twice to remove paraformaldehyde and then permeabilized with 4 ml 0.2% Triton X-100 in TBS at room temperature with mild shaking for 5 minutes. Cells were blocked in 2% BSA in TBST for 1 hour. 200 μl of diluted primary antibody (followed the antibody product sheets respectively) was added to the cells and then incubated at 37 ^{o}C in a humidified incubator for 1 hour. After incubation, cells were washed twice with 2% BSA in TBST then incubated with 200 μl of diluted secondary antibody for 1 hour at 37 ^{o}C. Cells were washed twice with 2% BSA in TBST before adding Vectashield antifade mounting medium with DAPI (Vector). Mounted dishes were covered with coverslips. Images were taken with LSM 880 with the Airyscan system (Zeiss) and processed via Zen Lite (Zeiss). All antibodies used in this study were listed in **Table 4**.

2.2 Results

2.2.1 Protein purifications of GST-tagged KDM5B PHD domains

In order to examine the binding mechanism of KDM5B PHD domains, cDNAs of different PHD domains were cloned into GST fusion vectors for purification. After induction, all three PHD domains showed more than one product in total cell lysates. GST protein itself has a molecular weight about 26 kDa and all three PHD domains are about 5.5 kDa. The purified product of PHD1 had two main bands at about 30 kDa and 35 kDa; PHD2 had two main bands at about 26 kDa and 30 kDa. The PHD3 domain had only one major band, which was about 36 kDa. Smaller bands under 25 kDa were also observed in all three purified proteins (**Figure 2.1**).

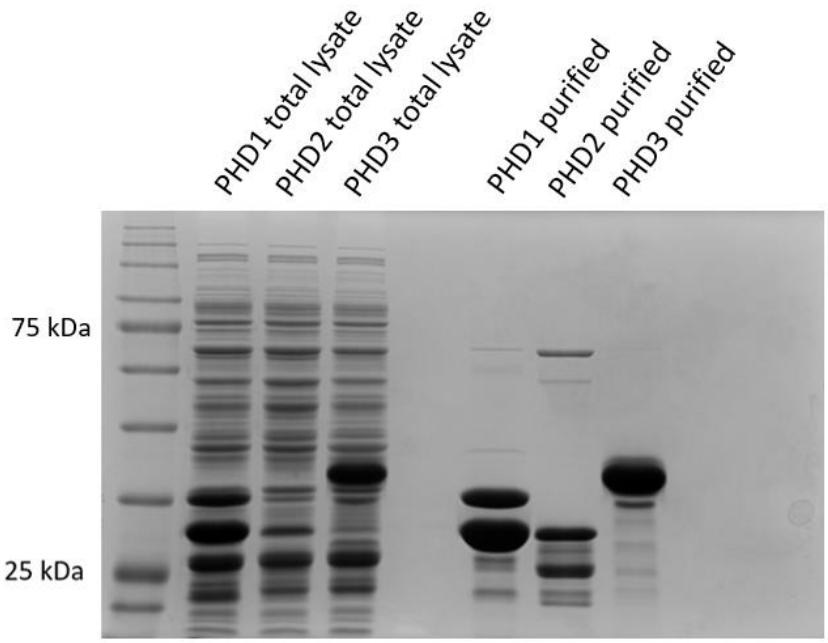


Figure 2.1: Coomassie staining of KDM5B - PHD1, 2 and 3 domains from *E. coli* Rosetta DE2. 1ml induced *E. coli* cells were directly lysed in Laemmli sample buffer and 10ul per lane was loaded. 10ul of total 1ml eluted GST-tagged protein was loaded on

each lane. Protein marker from bottom to top was 20 kDa, 25 kDa, 35 kDa, 48 kDa, 63 kDa and 75 kDa.

2.2.2 Subcellular fractionation

To ensure that histone proteins in cell lysate were not interfering histone peptide pull-down, subcellular fractionation was performed. Coomassie staining showed that in both S1 and S2 fractions, proteins (bands) below 17 kDa were considerably less than total cell lysate. Chromosome fraction showed enriched histone proteins at around 17 kDa (histone H2A, H2B, H3 and H4) and 35 kDa (histone H1). Immunoblotting using anti-histone H2B indicated that histone H2B were only present in total cell lysate and P1 fraction (**Figure 2.2**).

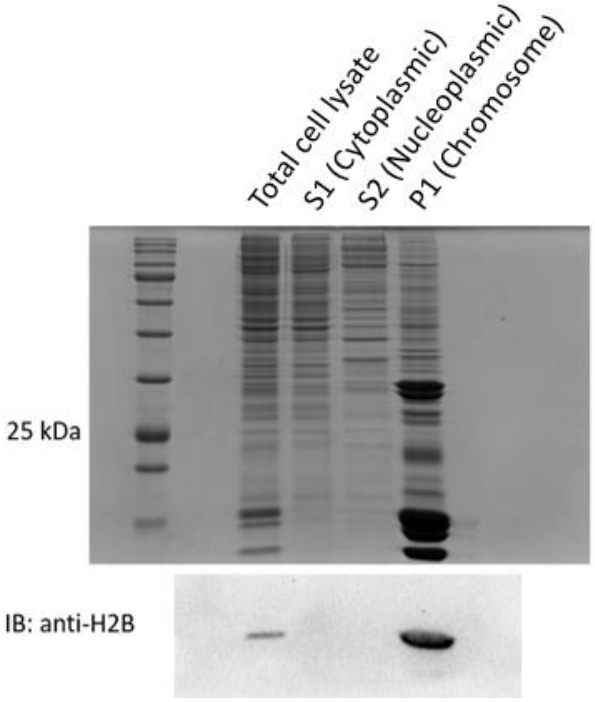


Figure 2.2: Subcellular fractionations. MCF7 cells were collected and fractionated into three different fractions. About 10 μg of proteins was loaded in each lane. Upper panel was Coomassie staining on 12% SDS-PAGE. Same amount of protein was used for the western blotting showed in lower panel.

2.2.3 Histone H2BK43 interacted with KDM5B PHD domains

Unlike PHD1 and PHD3, the PHD2 domain of KDM5B has not yet been reported with any binding partner. In the PHD1 and PHD3 binding study, only sites on histone H3 and H4 were examined, but not on H2B. Histone peptide pull-down assays of PHD3 confirmed previous findings (**Figure 2.3 B**). In addition, histone H2BK43 peptides failed to pull down GST-tagged PHD3. H2BK43 was the only histone H2B peptide tested on two other PHD domains of KDM5B - PHD1 and PHD2. The PHD1 domain showed better binding on unmodified H2BK43 and H2BK43 dimethyl than monomethyl and trimethyl. The PHD2 domain showed strong binding on H2BK43 unmodified peptide and weak binding on three other methylation stages (**Figure 2.3** and **Figure 2.4**). To evaluate the binding nature between KDM5B PHD domains and H2BK43 peptides, FP binding assays were performed using PHD1-H2BK43me0, PHD1-H2BK43me2 and PHD2-H2BK43me0. However, PHD1-H2BK43me0 and PHD1-H2BK43me2 FP did not result in specific binding affinity reads (data not showed). For PHD2-H2BK43me0, the specific kinetic dissociation constant was 102 μM.

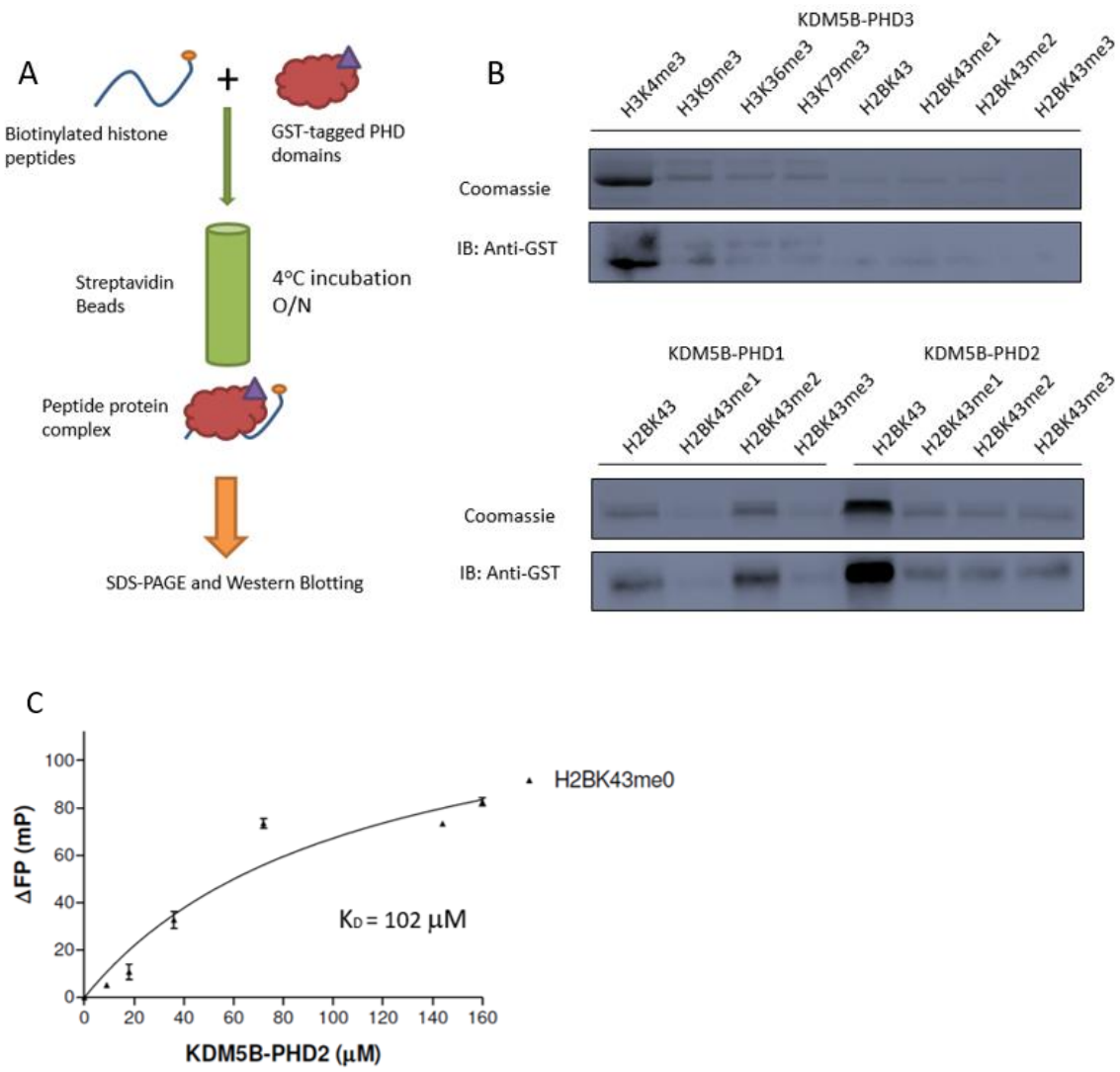


Figure 2.3: Histone H2BK43 peptides interacted with KDM5B PHD domains *in vitro*. **A)** Schematic workflow of biotinylated histone peptide pull-down assay. Amount of histone peptides added was meant to saturate the capacity of streptavidin beads to minimize variation of peptide amount used. **B)** Histone peptide binding preference of PHD1, 2 and 3. PHD3. 10μl of eluted proteins was loaded on each lane and both Coomassie gels and western blots were run in parallel on 12% SDS-PAGE. For western blot, transferred PVDF membranes were blocked with 3% non-fat milk in TBST before GST antibody incubation. **C)** Kinetic dissociation constant of PHD2 and H2BK43me0 binding. Calculation was conducted based on two replicates. Concentrations used is this assay from low to high were: 0.54 μM, 1.1 μM, 2.18 μM, 4.375 μM, 8.75 μM, 17.5 μM, 35 μM, 70 μM, 140 μM and 160 μM.

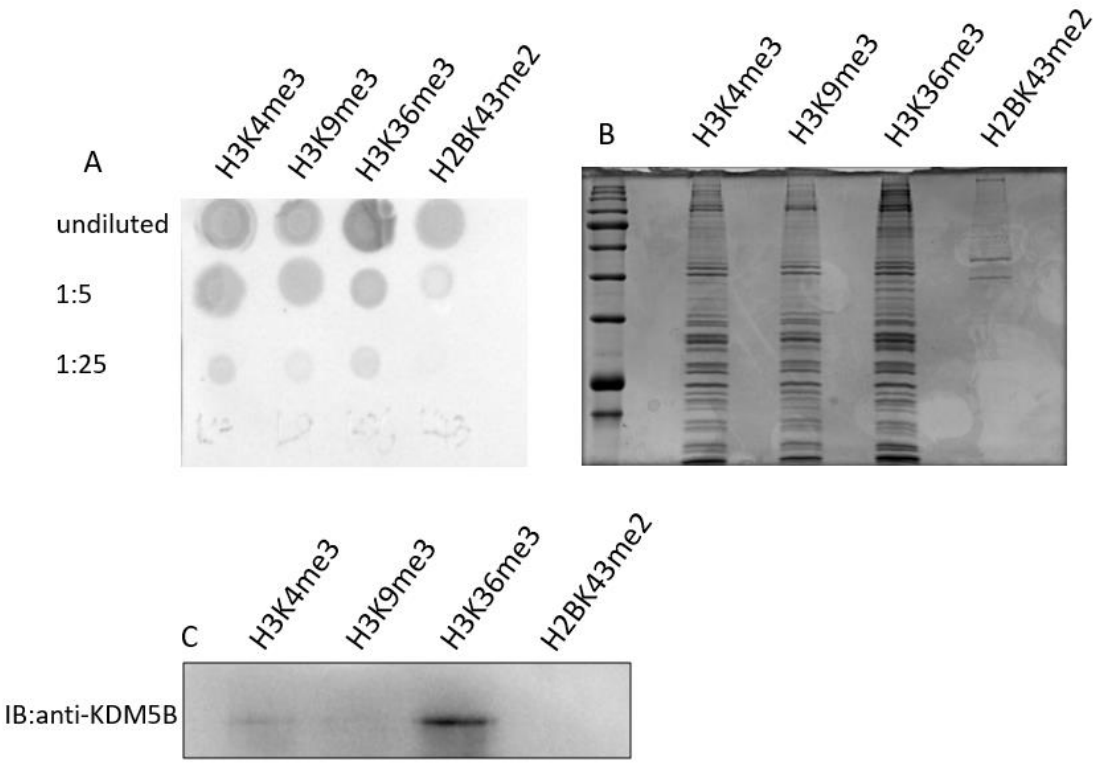


Figure 2.4: KDM5B was found in histone peptide pull-downs from cell lysate. A) Estimate of peptide concentration using Ponceau S staining. Peptides dissolved in H_2O were diluted and 2µl was dotted on nitrocellulose membrane. Membrane was stained with Ponceau S after air dried and subsequently rinsed with running water. Signal intensity correlated to peptide amount. **B)** Histone H3 N-terminal peptides interacts with more protein than H2BK43me2. Coomassie staining was performed on 12% SDS-PAGE gel. All four histone peptides could pull down proteins from fractionated cell lysates, regardless of efficiency. Due to the relatively low peptide concentration of H2BK43me2, double amount of peptide was used and comparable result was observed (**appendix B**). **C)** KDM5B interacted with histone H3K36me3 peptide. Western blot using anti-KDM5B was performed in the same loading amount as B. Membrane showed here was cropped at 180 kDa. Original uncropped membrane is shown in **appendix B**.

2.2.4 Identification of H2BK43 methylations under cell physiological conditions

To confirm the existence of H2BK43 methylation, we utilized mass spectrometry. Identification of histone post-translational modifications mainly relied on specific antibodies. Therefore, an alternative method such as mass spectrometry should be used to confirm western blot findings. In short, precursor ions and daughter ions were detected by MS/MS system and searched on pre-exist database. To H2BK43me1, y1 ion contained our target methyllysine, which has a molecular mass 161.13 Daltons. The detections of y2 to y7 ions also confirmed that the K43 lysine was monomethylated in its precursor ion ESYSVYVYK (**Figure 2.5 A**). To H2BK43me2, y1, y2 and y3 ions did not contain our target lysine. However, detections of dehydrated y5 and y7 ions confirmed that the K43 lysine was dimethylated in its precursor ion ESYSIYVYKVLK (**Figure 2.5 B**). Note that the sequence of H2BK43me2 precursor ions belonged to the Type2-E variant of H2B, which is different from the Type1 C/E/G variants of H2BK43me1 precursor. Lastly, H2BK43me3 was also confirmed by its deaminated b13 ions from its precursor ion KESYSIYIYKVLKQVHPDTGISSK (**Figure 2.5 C**). This specific sequence was from the Type1-A variant of H2B. Detailed data of all detected ions is shown in **appendix G**.

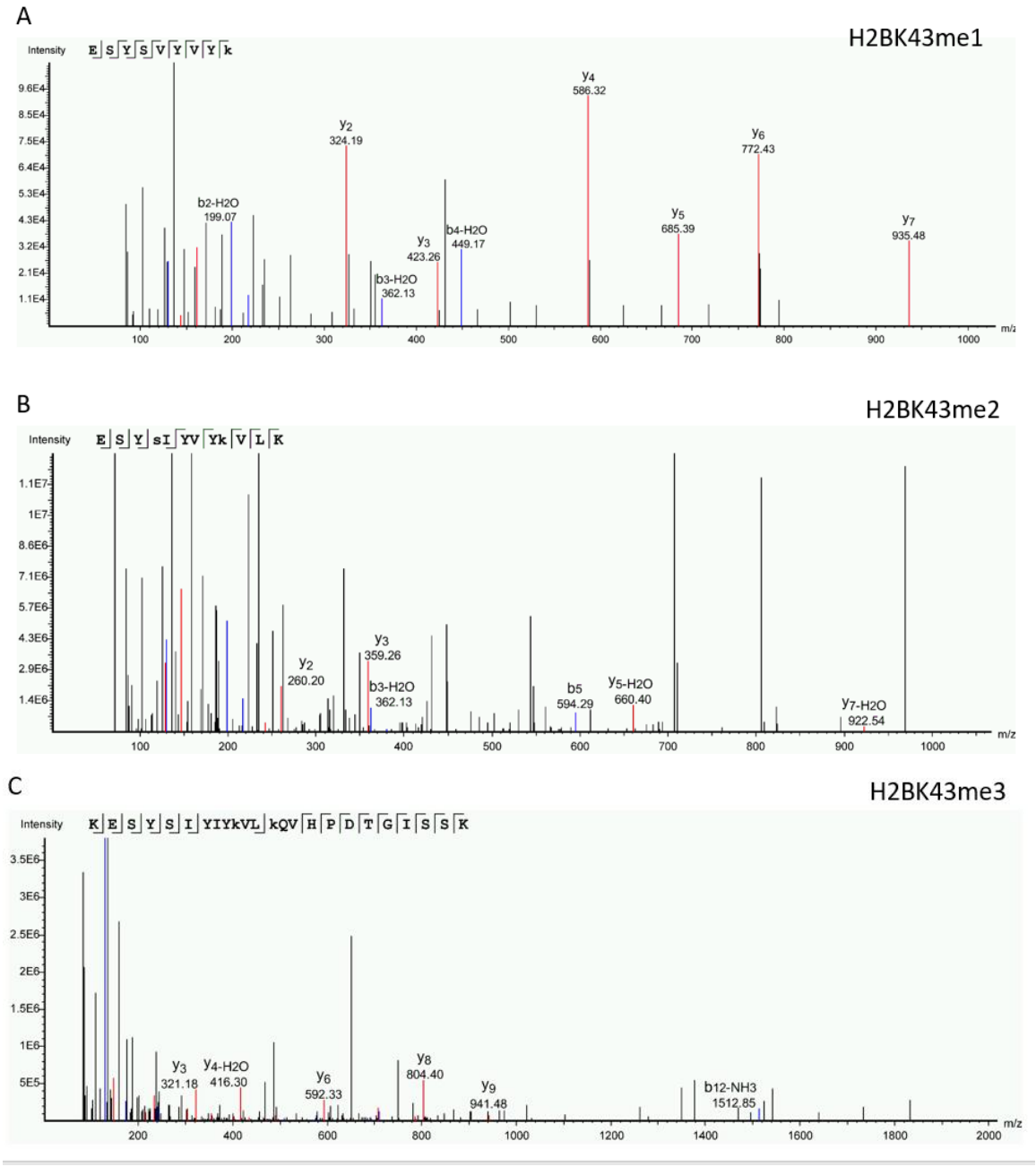


Figure 2.5: Identification of *in vitro* H2BK43 methylations by MS/MS. A-C) Spectra showing H2BK43 mono, di and trimethylation. MS data collection and analysis were performed by Dr. Rui Wong. Typically, b ions were fragments obtained from N-terminal of its corresponding precursors (for example, b5 ions in B section had sequence ESYSI); y ions were fragments obtained from C-terminal of its corresponding precursors (y3 ions in B section had sequence VLK). Peaks of y and b ions with color labels were identified with high confidence.

2.2.5 Histone mutant H2BK43 to R showed deleterious effects on mESCs survival

In order to examine the biological significance of H2BK43, we created two mutants H2BK43 to A and H2BK43 to R, to abolish and mimic the amino acid nature of lysine respectively. Our results indicate a relatively quick nucleosome integration of histone H2BK43 to A, comparing to WT H2B and H2BK43 to R (**Figure 2.6 A**). In addition, after single cell clone selection, expression of 3X FLAG WT H2B increased significantly, indicating that 3X FLAG WT H2B had also integrated into nucleosomes (**Figure 2.6 A**). However, while neomycin resistance was consecutively maintained in the H2BK43 to R mutant clones, anti-FLAG antibody failed to detect expression of FLAG tags, suggesting that no H2BK43 to R expression could be found in the cells (**Figure 2.6 B**).

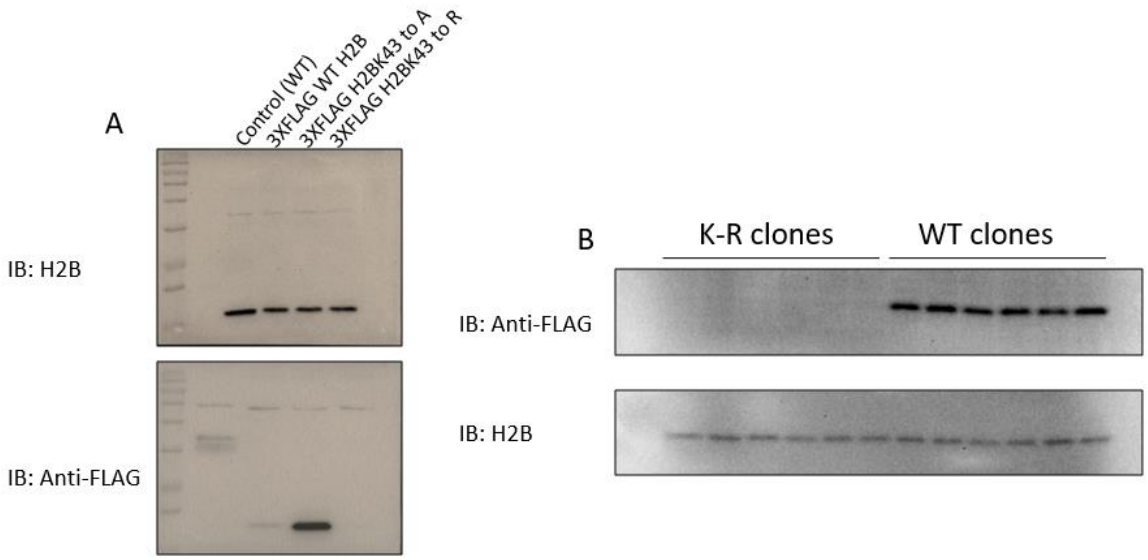


Figure 2.6: Histone H2BK43 to R was a potential deleterious trait for cell survival. A) mESCs stably expressing 3X FLAG H2BK43 to A mutant histones 7 days after transfection. Upper panel and lower panel were two different membranes. 5 μg of each whole cell lysate samples was loaded on each lane. **B)** Selected single cell clones stably expressing 3X FLAG WT H2B histones but not H2BK43 to R mutants. Lower panel was a stripped membrane after probed with anti-FLAG.

2.2.6 KDM5B was downregulated at differentiation day 5, but not KDM5A and KDM5C

Although complete knockout of KDM5B had shown embryonic lethality at E4.5, knockdown KDM5B in mESCs was successfully created and maintained. Among all KDM5B stable knockdowns we obtained, PLKO-sh91 had the lowest expression and displayed normal stem cell morphology compared to control (**Figure 2.7 A and B**). As expected, in PLKO-sh91, KDM5A expression was unaffected. Interestingly, from our RT-qPCR results, KDM5C was upregulated in about 2-fold while KDM5B was undetected (**Figure 2.7 C**). For the stem cell pluripotency markers, no significant increase of Nanog was detected and there was about 1.8-fold increase of Oct4 detected (**Figure 2.7 C**).

We also examined the dynamics of KDM5B during neuronal differentiation. In line with our previous finding, both western blot and RT-qPCR results verified that KDM5B expression declined after withdrawal of LIF and neuronal induction with B27 supplement (**Figure 2.8 A and B**). KDM5B expression declined by 20% at differentiation day 5, then further 40% at differentiation day 7. However, KDM5A levels remained relatively stable until differentiation day 7 and KDM5C showed upregulation from differentiation day 3 and downregulation at day 7 (**Figure 2.8 A**). Both Oct4 and Nanog were downregulated immediately upon LIF removal (**Figure 2.8 A**).

In addition, important neurogenesis marks such as AscI, Nestin, Pax6 and TrkB were immediately upregulated after the withdrawal of LIF and neuronal induction with B27 supplement. Pax6 declined at day 7 and TrkB was downregulated after day 3 (**Figure 2.8 C**). Notably, the DNA methyltransferase Dmnt3b, which negatively regulates stem cell pluripotency, increased gradually upon LIF removal (**Figure 2.8 C**).

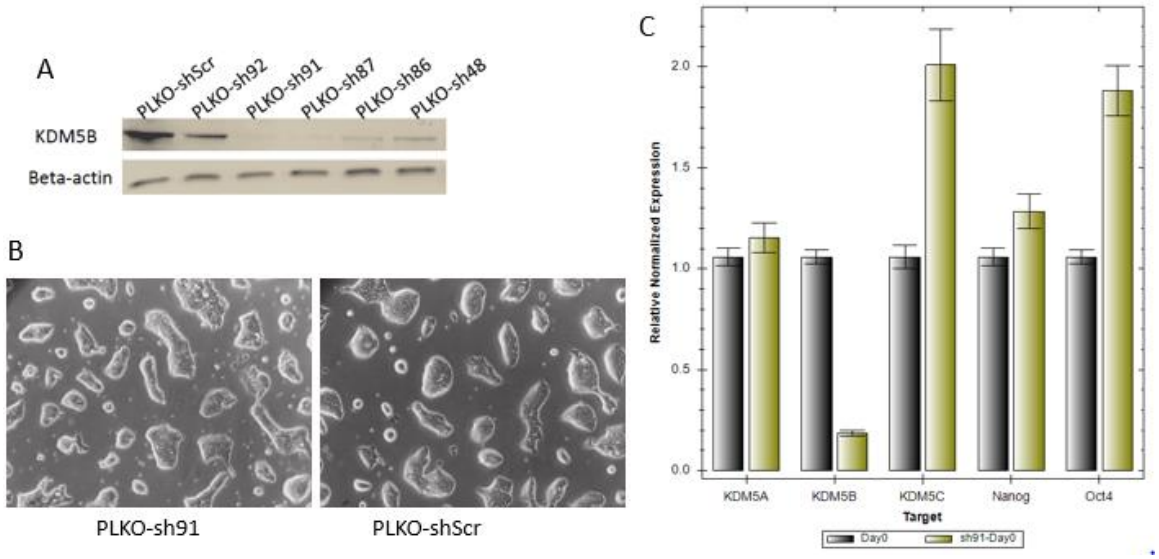


Figure 2.7: KDM5B knockdown in mESCs displayed normal morphology but caused compensational upregulation of KDM5C. A) Stable KDM5B knockdown mESCs. Whole cell lysate was collected 5 days after viral infection and puromycin selection; 5 µg of protein was loaded on each lane. **B)** KDM5B knockdown mESCs had same morphology of control mESCs. Both cells were infected by same MOI of viral particles. Images were taken at 100X total magnification with phrase-contrast filter. **C)** KDM5C increased in KDM5B knockdown cells. cDNA samples were prepared on same day as whole cell lysates were collected. Sample contained no biological repeats; error-bar was generated by four technical repeats.

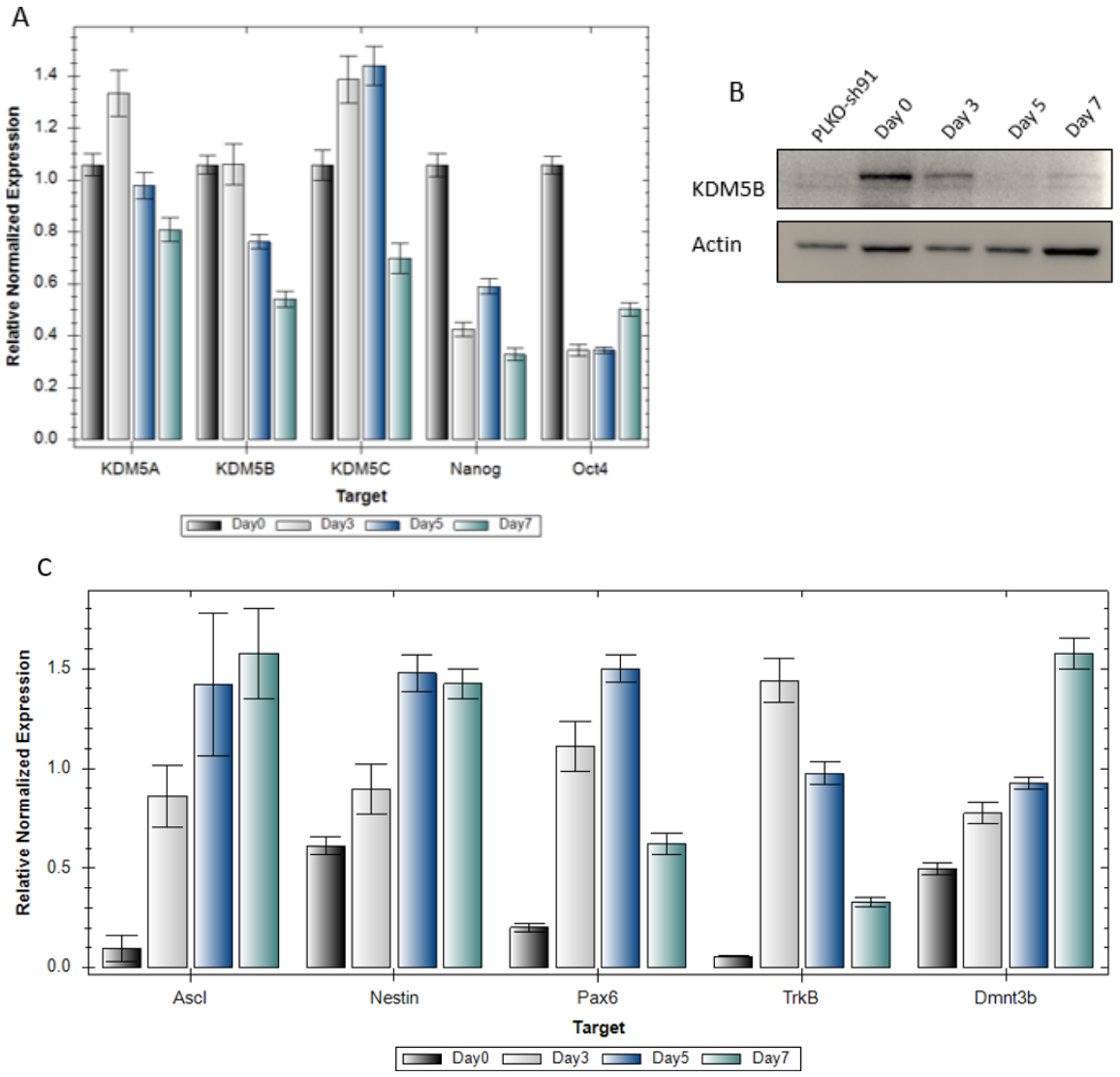


Figure 2.8: Downregulation of KDM5B during neuronal differentiation. A) Dynamics of KDM5A, KDM5B and KDM5B during neuronal differentiation. Data showed here was compared to day 0 (set to 1) in relative expression. Beta-actin and ribosomal subunit S18 were the housekeeping genes used for normalization. **B)** KDM5B expression in western blot showed same trend as RT-qPCR. 8% Tris-glycine SDS-PAGE in 1X Towbin running buffer was used for protein separation**. C)** Expression of differentiation markers in neurogenesis. Data showed here were relative quantifications using same set of housekeeping genes as section A for normalization (IF images of differentiated cells showed in **(appendix H**)).

2.2.7 KDM5B knockdown in breast cancer MCF7 cells suppressed cell proliferation

We also conducted siRNA knockdown in MCF7 cells since the lentiviral shRNAs were specific to mouse genes only. KDM5B knockdown MCF7 showed lower proliferation rate compared to WT. KDM5B overexpression in MCF7 was not detected after transfection with a construct that has not been sequenced (**Figure 2.9**).

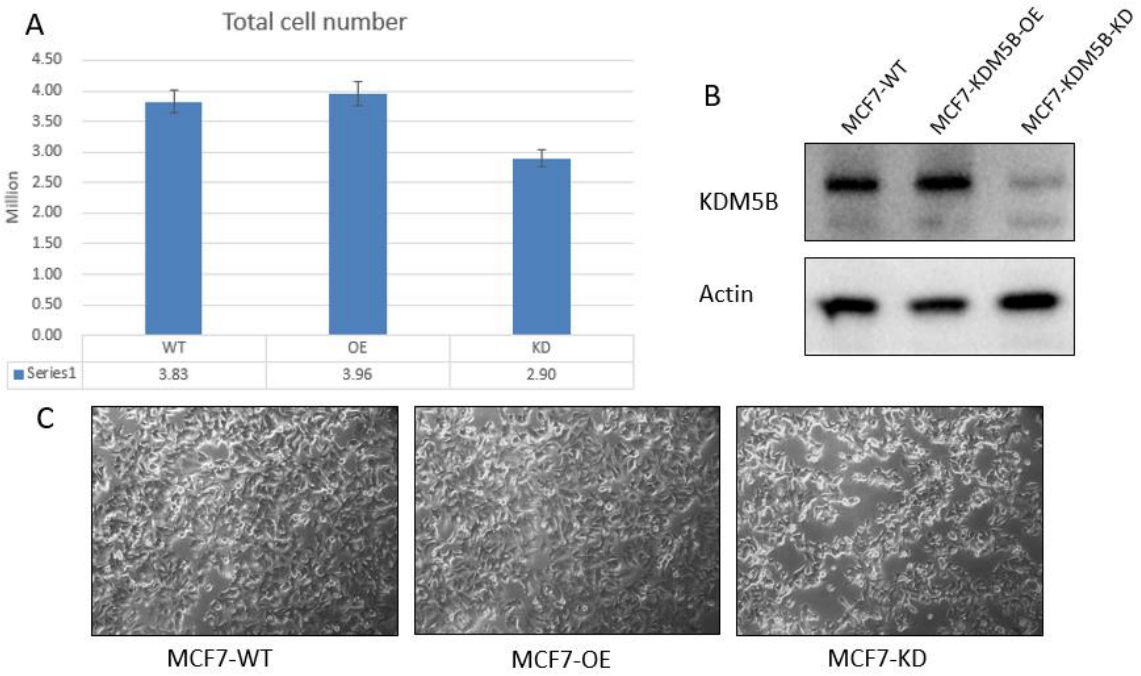


Figure 2.9: KDM5B knockdown impaired cell proliferation in MCF7 cells. A) Total cell counts of MCF7 3 days after transfection. Cell counts were performed using two biological repeats and actual counting included four technical repeats. **B)** KDM5B expression significantly decreased after siRNA transfection. 8% Bis-Tris SDS-PAGE in 1X MOPS running buffer was used for protein separation**. C)** KDM5B knockdown displayed lower culture confluency. Images were taken 3 days after transfection at 100X total magnification with phrase-contrast filter.

2.2.8 H2BK43me2 was a potential substrate of KDM5B in MCF7 cells but not mESCs

To confirm that H2BK43me2 was a substrate of KDM5B under cell physiological conditions, we used both differentiated mESCs and MCF7 cells. In mESCs, KDM5B knockdown induced accumulation of H3K4me3 but not H2BK43me2. During differentiation, KDM5B was downregulated and H3K4 trimethylation was increased after Day 3. No significant upregulation of H2BK43me2 was observed in western blots (**Figure 2.10 A**). In MCF7 cells, KDM5B knockdown also induced a slight increase of H3K4me3. Notably, unlike in mESCs, H2BK43me2 accumulation was observed in KDM5B knockdown MCF7 cells (**Figure 2.10 B**).

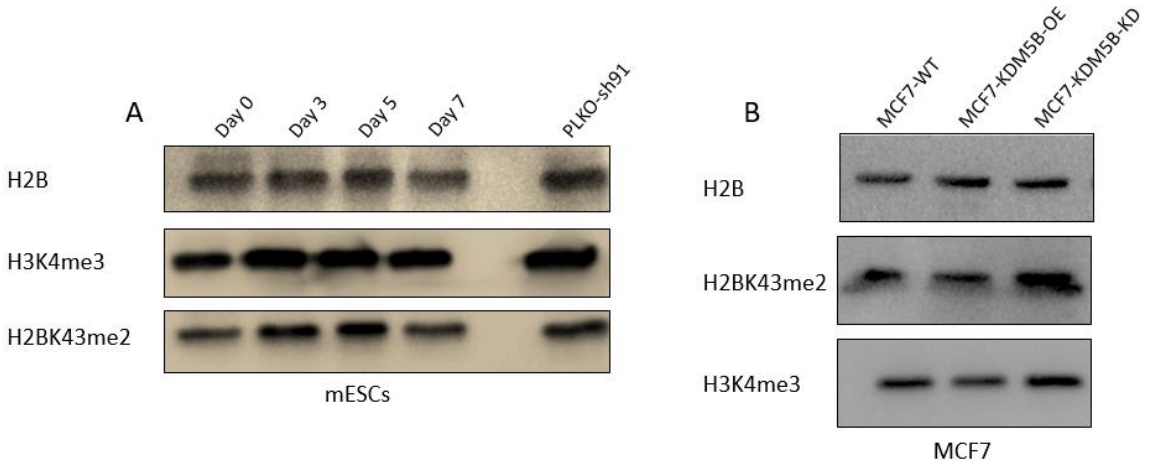


Figure 2.10: KDM5B expression affected H3K4me3 and H2BK43me2 levels in MCF7 cells. A) KDM5B expression affected H3K4me3 but not H2BK43me2 in mESCs. Samples were collected at specific differentiation days and histone methylation was protected by adding boiling lysis buffer that contained 2% SDS. 15% Tris-glycine SDS-PAGE in 1X Towbin running buffer was used for protein separation. **B)** KDM5B expression affected both H3K4me3 and H2BK43me2 in MCF7 cells. Specificity of antibodies against histone modification (K4me3 and K43me2) was tested before used in western blots (**Appendix E**).

Chapter 3

General discussion and future study

3.1 General discussion and conclusions

Although various studies have been done on KDM5B and its binding partners in recent years, the focus remained on histone H3 N-terminal sequences. The recruitment of KDM5B by H3K4me3 was unnatural as it binds considerably tight to the PHD3 domain (0.5 ± 0.2 μM). Nevertheless, all three stages of H3K4, me2, me1 and me0, showed very strong binding to PHD3 domain (1.3 ± 0.4 μM, 2.0 ± 0.9 μM and 4.0 ± 1.7 μM respectively). Knowing that H3K4me3/2 were substrates of KDM5B, such a tight binding may constrain enzymatic reactions. The author reasoned that demethylation mediated by KDM5B did not take place at the H3K4me3 it anchored; instead, KDM5B was recruited by H3K4me3 and demethylated the surrounding nucleosomes. The hypothetical model they suggested did not come with sufficient evidence (Klein *et al.*, 2014). In addition, study from Xie *et al.* suggested that KDM5B could be recruited by H3K36me3 directly. In line with this finding, we confirmed that H3K36me3 could be recognized by KDM5B in histone excluded cell lysates while it could not be recognized by recombinant KDM5B proteins (**figure 2.3 and 2.4**). This suggests that H3K36me3 might recruit KDM5B indirectly in cells. Notably, it was reported that KDM5B did not overlap with H3K36me3 at the TSS in mESCs ChIP-seq data (Schmitz *et al.*, 2011).

H2BK43me0 and H2BK43me2 are potential binding partners of KDM5B

No KDM5B PHD2 binding partner has been identified so far. However, in our histone peptide pull-down assay (**Figure 2.3**), unmodified H2BK43 displayed moderate interaction with PHD2; FP analysis have also shown that, though with low affinity, unmodified H2BK43 was a potential binding partner to PHD2. Binding between PHD domains and methyllysine requires complete coverage of the site with the hydrophobic cage of PHD domains. Prolonged binding may constrain potential enzymatic reactions taking place on the site. Therefore, having a weak binding between histones and chromatin binding proteins is more rational than a nanomolar scale binding which may

result in a tight binding and reduce the flexibility of chromatin. Interestingly, both H2BK43 unmodified and H2BK43me2 showed interactions with PHD1, indicating that KDM5B was capable of recognizing multiple stages of H2BK43 methylation. It is not surprising that H2BK43me2 failed to pull down KDM5B in histone excluded cell lysates. Note H2BK43me0 should have been tested instead of H2BK43me2 since it is a better binder.

Like many methyl binding domains, PHD domains of KDM5B has a broad range of binding partners (Klein *et al.*, 2014). Most of binding partners of KDM5B, or even the whole KDM5 family, is reported to be located on histone H3, which is a core histone that has already been well studied. However, the potential binding partners that we identified, namely H2B, has not yet been well studied. Elucidating the interactions between H2BK43me2 and KDM5B or its other family members can add a further step in the understanding of how these demethylases contribute to transcriptional regulations. In our study, we showed that H2BK43 and H2BK43me2 peptides could pull down GST-tagged PHD1 and PHD2 domains, although interactions were not very strong (**Figure 2.3**). As mentioned above, KDM5 family is not only acting as demethylases but also being epigenetic recruiters with their DNA and nucleosome binding domains. It is also possible that KDM5B recruits other proteins upon binding to H2BK43me2/0. Our findings were mainly based on *in vitro* experiments, but it is rational that there should be binding partners for the PHD2 domains, as all KDM5 members contain this domain. Together, we believe that histone H2BK43 and H2BK43me2 are both potential binding partners of KDM5B.

H2BK43 modification can be variant specific

We successfully confirmed that histone H2BK43me2 exists under cell physiological conditions. However, the typical variants, H2B Type2-E that we identified K43me2 on, was reported rarely expressing in somatic cells (Molden *et al.*, 2015). Note that Molden *et al.* identified most histone H2B PTMs by lysine propionylation and Arginine-C enzyme digestion, we applied a different digestion scheme, and therefore our identifications may have shown discrepancy compared to their results. Nevertheless, our

western blot results verified the presence of H2BK43me2 *in vitro*. Knowing the low expression of H2B type-E in somatic cells, it is necessary to assess the expression of histone H2B variants in mESCs for further understanding.

H2BK43 is critical for cell survival

In our histone H2B mutant experiments, only H2BK43 to A was created successfully; H2BK43 to R cells were selected out in the mESC medium containing neomycin (**Figure 2.6 B**). Since non-viral transfection was achieved by homologous recombination, failure to produce H2BK43 to R mutant suggested that this trait was deleterious in natural selection. However, compared to H2BK43 to A, we expect a milder phenotype from the mimetic mutation to arginine. To fully investigate this phenomenon, lentiviral infection of H2BK43 to R and H2BK43 to A could be used. If H2BK43 to R was a true deleterious trait, infected cells would not be able to survive in the neomycin selection medium.

H2BK43me2 is a potential substrate in MCF7 cell

KDM5B might not be the only enzyme acting on H3K4me3 and H2BK43me2, as KDM5C and KDM5A also share the same architecture. Indeed, it was reported that KDM5A could also remove methyl groups from H3K4me3/2 (Torres *et al.*, 2015). Although most studies focused on the epigenetic reader and DNA binding properties of KDM5C (Iwase *et al*, 2016; Putchkourov *et al.*, 2013), KDM5C was suggested to be a H3K4me3 demethylase (Rondinelli *et al.*, 2015). Therefore, the compensatory elevation of KDM5C might mask the effect of KDM5B knockdown, and therefore H3K4me3 and H2BK43me2 were maintained on stable levels (**Figure 2.7 A** and **Figure 2.8**).

Due to the atypical positioning of H2BK43, identification of H2BK43me2 under cell physiological conditions can be elusive. As mentioned above, histone H2B has a relatively short N-terminal tail compared to H3 and H4. H2BK43 is covered by DNA double helix and therefore its exposure is limited (**Figure 1.11**). However, different conditions such as cell types, cell cycles and extracellular stimulations, may expose this site to nucleoplasm by sliding the DNA double helix (Carins, 2007). Therefore, the recognition mechanism between H2BK43me2 and KDM5B can also be contingent on the

genetic background of cells, which possibly explains why KDM5B knockdowns behaved differently in MCF7 than mESCs and differentiated mESCs. In addition, KDM5B shows no demethylation activity to H3K4me3 and has no effects on brain development beyond E12.5, indicating that cell type is a critical factor to consider when studying epigenetic events (Schmitz *et al.*, 2011).

While our biochemical studies of enzymatic properties of KDM5B have been established, investigations in the enzymatic activities under cell physiological conditions are still necessary. We observed that accumulation of H3K4me3 occurred after the knockdown of KDM5B in both mESCs and MCF7 cells (**Figure 2.10**). In KDM5B knockdown cells, H2BK43me2 levels remained stable in mESCs but increased in MCF7. As mentioned earlier, chromatin architectures could be cell type dependent. Without the complete profiling of the expression of KDM5 family in MCF7, it is difficult to explain why mESCs did not show changes in H2BK43me2 levels. Changes in H3K4me3 levels were not only mediated by KDM5B, but other family members as well. The same theory could account for H2BK43me2 while it is possible that compensatory effects did not occur in MCF7. Note that studies of histone modifications are highly cell status dependent. For example, H3S10 phosphorylation only occurs during transcriptional activation of immediate-early genes upon mitogen induction (Thomson *et al.*, 2001). However, H3S10 phosphorylation only occurs spontaneously during G2 phase within pericentromeric heterochromatin (Hendzel *et al.*, 1997). During stem cell differentiations, the system quickly increases its complexity. Multiple variables are introduced, such as rapid changes of chromatin architecture, other competitive nucleosome binding proteins, and dynamics of other histone modifications, making the search of right timing that showed maximum correlation of KDM5B and H2BK43me2 more difficult and further limited our conclusion. Therefore, a fair conclusion to draw at this point is that H2BK43me2 can be a potential substrate of KDM5B in MCF7 cells.

3.2 Future studies

To assay whether KDM5B binds to H2BK43me2 under cell physiological conditions, demonstration of actual ligand interaction is necessary. One approach is to use fusion protein A- mutant BirA (Roux *et al.*, 2012). In brief, cells are first fixed and stained with H2BK43me2 specific antibody then biotinylated by protein A- mutant BirA to identify its interaction partners. Again, multiple cell lines and different phases of cell cycle need to be tested before the best condition is found. This experiment retains a native condition of nucleosome and therefore reveals the natural interactions between these proteins. Also, comparing the KDM5B ChIP data with H2BK43me2 may help in understanding the epigenetic significances behind this interaction.

To confirm that KDM5B is the enzyme involved in H2BK43me2 demethylation, an overexpression of KDM5B can be generated. Unlike the knockdowns in this study, overexpression of KDM5B may bypass problems of compensatory effects from other KDM5 family members. The KDM5B overexpression construct used in this study was not sequenced, which may explain why KDM5B overexpression was not observed. In addition, the size of KDM5B is about 160kDa, thus lentiviral infection may have provided better results compared to transfection.

Demethylation and methylation always come in pairs, and therefore it is important to study which histone lysine methyltransferase (KMTs) regulates H2BK43 methylations after we confirmed that KDM5B catalyzes the demethylations under cell physiological conditions. Two possible approaches include using RT-qPCR to screen KMTs that are affected in KDM5B knockdown cells and H2BK43me0 peptide pull-down assays. The first approach is based on the compensatory effect that may have been caused by KDM5B knockdown. Since global histone methylation is likely unchanged, a corresponding KMT may need to be downregulated. The second approach relays on a potential H2BK43me0 specific KMT that can both read and modify this site. If a specific KMT is not directly recruited by H2BK43me0 but it is recruited by other reader proteins, then the pull-down assay may not be a suitable method.

It is possible that during differentiation, H2BK43me2 level will not show differences in global level. However, histone modification can act on specific loci to regulate gene expression. In order to explore which genes H2BK43me2 regulates, ChIP-seq analysis can be conducted. Based on the histone code theory, it is very likely that H2BK43me2 is not acting along but in concert with other histone marks. Therefore, in addition to regular ChIP-seq using only H2BK43me2 antibody, Co-ChIP-seq combining other histone modification antibodies – such as H3K4me3, H3k9me3 and H3K36me3 – can also be used. This approach will not only allow us to elucidate the regulatory regions that H2BK43me2 governs, but may also deepen our understanding of how this novel histone modification can cross-talk with other known histone marks to regulate gene expression.

References

Abarrategui, I and Helin, K. (2011). Jarid1b targets genes regulating development and is involved in neural differentiation. *EMBO J.* 30:4586-4600.

Allfrey, V.G., Faulkner, R. and Mirsky, A.E. (1964). Acetylation and methylation of Histone and Their Possible Role in the regulation of Rna Synthesis. *Proc Natl Acad Sci.* 51(5): 786-794.

Bamodu, O.A., Huang, WC., Lee, WH., Wu, A., Wang, L.S., Hsiao, M., Yeh, CT and Chao, TY. (2016). Aberrant KDm5B expression promotes aggressive breast cancer through MALAT1 overexpression and downregulation of has-miR-448. *BMC Cancer*. 16:160.

Bhaumik S.R., Smith, E. and Shilatifard A. (2007). Covalent modifications of histones during development and disease pathogenesis. *Nature Structural and Molecular Biology*. 14: 1008-1016.

Biggar, K.K. and Li, S-C. S. (2015). Non-histone protein methylation as a regulator of cellular signal transduction. *Nat Rev Mol Cell Biol.* 16(1):5-17.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y. and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*. 84(6): 843-851.

Cairns, B.R. (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. *Nature Structural & Molecular Biology*. 14(11): 989–996.

Campos E.I. and Reinberg D. (2009). Histones: Annotating Chromatin. *Annu Rev Genet.* 43:559-599.

Catchpole, S., Spencer-Dene, B., Hall, D., Santangelo, S., Rosewell, I., Guenatri, M., Beatson, R., Scibetta, A.G., Burchell, J.M. and Taylor-Papadimitriou, J. (2011) PLU-1/JARID1B/KDM5B is required for embryonic survival and contributes to cell proliferation in the mammary gland and in ER+ breast cancer cells. *Int J Oncol*. 38: 1267–1277.

Carpino, L.A. and Han, G.Y. (1970). 9-Fluorenylmethoxycarbonyl function, a new base-sensitive amino protecting group. *JACS*. 92(19): 5748-5749.

Chen, D., Ma, H., Hong, H., Koh, S.S., Huang, S.M., Schurter, B.T., Aswad, D.W., and Stallcup, M.R. (1999). Regulation of transcription by a protein methyltransferase. *Science.* 284(5423):2174-2177.

Condic, M.L. (2014). Totipotency: What it is and what it is not. *Stem Cells and Development.* 23(8): 796-812.

Cheung, P., Tanner, K.G., Chueng, W.L., Sassone-Corsi, P., Denu, J.M. and Allis, C.D. (2000), Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol Cell.* 5(6): 905-915.

Clayton, A.L., Rose, S., Barratt, M.J. and Mahadevan, L.C. (2000) Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosome upon gene activation. *EMBO journal.* 19(14): 3714-3726.

Dey, B.K,, Stalker, L., Schnerch, A., Bhatia, M., Taylor-Papidimitriou, J. and Wynder, C. (2008.). The Histone Demethylase KDM5b/JARID1b plays a role in cell fate decisions by blocking terminal differentiation. *Mol Cell Biol*. 28:5312–5327.

Evans, M.J. and Kaufman M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*. 292:154–156.

Falkenberg, K.J. and Johnstone, R.W. (2014). Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat Rev Drug Discovery.* 13: 673-691.

Hayami, S., Yoshimatsu, M., Veerakumarasivam, A., Unoki, M., Iwai, Y., Tsunoda, T., Field, H.I., Kelly, J.D., Neal, D.E., Yamaue, H., Ponder, B.A., Nakamura, Y. and Hamamoto, R. (2010). Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: involvement in the proliferation of cancer cells through the E2F/Rb pathway. *Mol Cancer.* 9:59.

Hendzel, M.J., Wei, Y., Manciin, M.A., Van Hooser, A., Ranalli, T., Brinkley, B.R., Bazett-Jones B.R. and Allis, C.D. (1997). Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma.* 106(6): 348-360.

Iwase, S., Brookes, E., Agarwal, S., Badeaux, A.I., Ito, H., Vallianatos, C.N., Tomassy, G.S., Kasza, T., Lin, G., Thompson, A., Gu, L., Kwan, KY. Chen, C. Sartor, M.A., Egan, B., Xu, J. and Shi, Y. (2016). A mouse model of X-linked intellectual disability associated with impaired removal of histone methylation. *Cell Reports.* 14(5): 1000–1009.

Jin, Y., Wang, Y., Walker, D.L., Dong, H., Conley, C., Johansen, J. and Johansen, K.M. (1999). JIL-1: a novel chromosomal tandem kinase implicated in transcriptional regulation in Drosophila. *Mol Cell.* 4(1): 129-135.

Klein, B.J., Piao, L., Xi. Y., Rincon-Arano, H., Rothbart, S.B., Peng, D., Wen, H., Larson, C., Zhang, X., Zheng, X., Cortazar, M.A., Pena, P.V., Mangan, A., Bentley, D.L., Strahl, B.D., Groudine, M., Li, W., Shi, X. and Kutateladze, T.G. (2014). The histone-H3K4-Specific Demethylase KDM5B binds to Its Substrate and Product Through Distinct PHD Fingers. *Cell Reports.* 6: 325-335.

Kidder, B.L., Hu, G. and Zhao, K. (2014). KDM5B focuses H3K4 methylation near promoters and enhancers during embryonic stem cell self-renewal and differentiation. *Genome Biology.* 15: R32

Kornberg, R.D. (1974). Chromatin structure: A repeating unit of histones and DNA. *Science.*184:868–871.

Kim, Y.Z. (2014). Altered Histone modifications in Gliomas. *Brain Tumor Res Treat*. 2(1): 7-21.

Liu, H., Galka, M., Iberg, A., Wang, Z., Li, L., Voss, C., Jiang, X., Lajoie, G., Huang, Z., Bedford, M.T.and Li, S.SC. (2010). Systematic Identification of Methyllysine-Driven Interactions for Histone and Nonhistone Targets. *J. Proteome Res.* 29(11): 5827-5836.

Liu, Z., Li, F., Zhang, B., Li, S., Wu, J. and Shi, Y. (2015). Structural basis of plant homeodomain finger 6 (PHF6) recognition by the retinoblastoma binding protein 4 (RBBP4) component of the nucleosome remodeling and deacetylase (NuRD) complex.*J Biol Chem.* 290(10):6630-6638

Lu, F. and Yi, Z. (2015). Cell totipotency: molecular features, induction, and maintenance. *Natl Sci Rev.* 2(2): 217-225.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci*. 78:7634–7638.

Molden, R.C., Bhanu, N.V., Leroy, G., Arnaudo, A.M. and Garcia, B.A. (2015) Multi-faceted quantitative proteomics analysis of histone H2B isoforms and their modifications. *Epigenetics & Chromatin.* 8:15.

Musselman, C.A., Lalonde, ME., Cote, J. and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. *Nature Structural and Molecular Biology*. 19: 1218-1227. Lin, C.S., Lin, Y.C., Adebayo, B.O., Wu, A., Chen, J.H., Peng, Y.J., Cheng, M.F., Lee, W.H., Hsiao, M., Chao, T.Y. and Yeh, C.T. (2015). Silencing

JARID1B suppresses oncogenicity, stemness and increases radiation sensitivity in human oral carcinoma. *Cancer lett.* 368(1):36-45.

Ohtsuka, S., Nakai-Futatsugi, Y., and Niwa, H. (2015). LIF signal in mouse embryonic stem cells. *JAK-STAT.* 4(2): 1-19.

Moreau JF, Donaldson DD, Bennett F, Witek-Giannotti J, Clark SC, Wong GG. (1988). Leukemia inhibitory factor is identical to the myeloid growth factor human interleukin for DA cells. *Nature*. 336:690-2.

Outchkourov, N.S., Muiño, J.M., Kaufmann, K., van Ijcken, W.F., Groot Koerkamp, M.J., van Leenen, D., de Graaf, P., Holstege, F.C., Grosveld, F.G. and Timmers, H.T. (2013). Balancing of histone H3K4 methylation states by the Kdm5c/SMCX histone demethylase modulates promoter and enhancer function. *Cell Reports*. 3(4): 1071-1079

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D. and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*. 406(6796): 593-599.

Robzyk, K., Recht, J. and Osley, M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. *Science.* 287(5452):501-504.

Rondinelli, B., Schwerer, H., Antonini, E., Gaviraghi, M., Lupi, A., Frenquelli, M., Cittaro, D., Segalla, S., Lemaitre, JM. And Tonon, G. (2015). H3K4me3 demethylation by the histone demethylase KDM5C/JARID1C promotes DNA replication origin firing. *Nucleic Acids Research*. 43(5): 2560-2574.

Roux, K.J., Kim, D.I., Raida, M. and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. The Journal of Cell Biology. 196(6): 801-810.

Schmitz, S., Albert, M., Malatesta, M., Morey, L., Johansen, J., Bak, M., Tommerup, N., Abarrategui, I and Helin, K. (2011). Jarid1b targets genes regulating development and is involved in neural differentiation. *EMBO J.* 30:4586-4600.

Shechter, D., Dormann, H.L., Allis, C.D. and Hake, S.B. (2007). Extraction, purification and analysis of histones. *Nat Protoc.* 2(6): 1445-1457.

Smith, A.G. and Hooper, M.L. (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev Biol*. 121:1-9.

Schmitz, S., Albert, M., Malatesta, M., Morey, L., Johansen, J., Bak, M., Tommerup, N., Abarrategui, I and Helin, K. (2011). Jarid1b targets genes regulating development and is involved in neural differentiation. *EMBO J.* 30:4586-4600.

Sone, M., Morone, N., Nakamura, T., Tanaka, A., Okita, K., Woltjen, K., Nakagawa, M., Heuser, J.E., Yamada, Y., Yamanaka, S. and Yamamoto, T. (2017). Hybrid cellular metabolism coordinated by Zic3 and Esrrb synergistically enhances induction of naive pluripotency. *Cell Metabolism.* 25(5): 1103-1117.

Strahl, B.D. and Allis, C.D. (2000). The language of covalent histone modifications. *Nature*. 403(6765): 41-45.

Salminen, A., Kaarniranta, K. and Kauppinen, A. (2016). Hypoxia-iducible histone lysine demethylases: impact on the aging process and age-related diseases. *Aging Dis.* 7(2):180-200

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.*126(4): 663–76.

Tanuton, J., Hassig, C.A. and Schreiber, S.L. (1996). A Mammalian Histone Deacetylase Related to the Yeast Transcriptional Regulator Rpd3p. *Science.* 272(5276): 408-411.

Thomson, S., Clayton, A.L. and Mahadevan, L.C. (2001). Independent dynamic regulation of histone phosphorylation and acetylation during immediate-early Gene Induction. *Mol Cell.* 8(6): 1231-1241.

Torres, I.O., Kuchenbecker, K.M., Nandi, C.I., Fletterick, R.J., Kelly, MJ. S. and Fujimori, D.G. (2015). Histone demethylase KDM5A is regulated by its reader domain through a positive-feedback mechanism. *Nat Communications*. 6:6204.

Van Holde, K.E., Sahasrabuddhe, C.G. and Shaw, B.R. (1974). A model for particulate structure in chromatin. *Nucleic Acids Res.* 1:1579–1586

Woodcock, C.L. (2005). A milestone in the odyssey of higher-order chromatin structure. *Nature Structural and Molecular Biology.* 12:639–640.

Wolffe, A.P. and Hayes, J.J. (1999). Chromatin disruption and modification. *Nucleic Acid Res.* 27(3): 711-720.

Wolffe, A.P. and Matzke, M.A. (1999). Epigenetics: Regulation Through Repression. *Science.* 286(5439): 481-486.

Welham, M.J., Kingham, E., Sanchez-Ripoll, Y., Kumpfmueller, B., Storm, M. and Bone, H. (2011). Controlling embryonic stem cell proliferation and pluripotency: the role of PI3K- and GSK-3-dependent signalling. *Biochem Soc Trans.* 39(2): 674-678.

Wysocka, J. (2006). Identifying novel proteins recognizing histone modifications using peptide pull-down assay. Methods. 40(4): 339-343.

Xie, L., Pelz, C., Wang, W., Bashar, A., Varlamova, O., Shadle, S. and Impey, S. (2011). KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO J*. 30:1473-1484.

Yamane, K., Tateishi, K., Klose, R.J., Fang, J., Fabrizio, L.A., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P., and Zhang Y. (2007). PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol Cell.* 25:801-812.

Yamanaka, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. and Yamanaka, S. (2008). Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors. *Science.* 322(5903): 949–53.

Tables

Table 1: List of peptides used in this study

Name	Sequence	Special labels
Histone H3K4me0	Biotin-ahx-MARTK(me0)QTARKS	Biotin
Histone H3K4me1	Biotin-ahx-MARTK(me1)QTARKS	Biotin
Histone H3K4me2	Biotin-ahx-MARTK(me2)QTARKS	Biotin
Histone H3K4me3	Biotin-ahx-MARTK(me3)QTARKS	Biotin
Histone H3K9me0	Biotin-ahx-TKQTARK(me0)STGGKA	Biotin
Histone H3K9me1	Biotin-ahx-TKQTARK(me1)STGGKA	Biotin
Histone H3K9me2	Biotin-ahx-TKQTARK(me2)STGGKA	Biotin
Histone H3K9me3	Biotin-ahx-TKQTARK(me3)STGGKA	Biotin
Histone H3K27me0	Biotin-ahx-ATKAARK(me0)SAPATG	Biotin
Histone H3K27me1	Biotin-ahx-ATKAARK(me1)SAPATG	Biotin
Histone H3K27me2	Biotin-ahx-ATKAARK(me2)SAPATG	Biotin
Histone H3K27me3	Biotin-ahx-ATKAARK(me3)SAPATG	Biotin
Histone H3K36me0	Biotin-ahx-PATGGVK(me0)KPHRYR	Biotin
Histone H3K36me1	Biotin-ahx-PATGGVK(me1)KPHRYR	Biotin
Histone H3K36me2	Biotin-ahx-PATGGVK(me2)KPHRYR	Biotin
Histone H3K36me3	Biotin-ahx-PATGGVK(me3)KPHRYR	Biotin
Histone H3K79me0	Biotin-ahx-EIAQDFK(me0)TDLRFQ	Biotin
Histone H3K79me1	Biotin-ahx-EIAQDFK(me1)TDLRFQ	Biotin
Histone H3K79me2	Biotin-ahx-EIAQDFK(me2)TDLRFQ	Biotin
Histone H3K79me3	Biotin-ahx-EIAQDFK(me3)TDLRFQ	Biotin
Histone H4K20me0	Biotin-ahx-GAKRHRK(me0)VLRDNI	Biotin
Histone H4K20me1	Biotin-ahx-GAKRHRK(me1)VLRDNI	Biotin
Histone H4K20me2	Biotin-ahx-GAKRHRK(me2)VLRDNI	Biotin
Histone H4K20me3	Biotin-ahx-GAKRHRK(me3)VLRDNI	Biotin
Histone H2BK43me0	Biotin-ahx-YVYKVLK(me0)QVHPDT	Biotin
Histone H2BK43me1	Biotin-ahx-YVYKVLK(me1)QVHPDT	Biotin
Histone H2BK43me2	Biotin-ahx-YVYKVLK(me2)QVHPDT	Biotin
Histone H2BK43me3	Biotin-ahx-YVYKVLK(me3)QVHPDT	Biotin
Histone H3K4me0	6FAM-ahx-MARTK(me0)QTARKS	Fluorescence
Histone H3K4me1	6FAM-ahx-MARTK(me1)QTARKS	Fluorescence
Histone H3K4me2	6FAM-ahx-MARTK(me2)QTARKS	Fluorescence
Histone H3K4me3	6FAM-ahx-MARTK(me3)QTARKS	Fluorescence
Histone H2BK43me0	6FAM-ahx-YVYKVLK(me0)QVHPDT	Fluorescence
Histone H2BK43me1	6FAM-ahx-YVYKVLK(me1)QVHPDT	Fluorescence
Histone H2BK43me2	6FAM-ahx-YVYKVLK(me2)QVHPDT	Fluorescence
Histone H2BK43me3	6FAM-ahx-YVYKVLK(me3)QVHPDT	Fluorescence

Table 2: List of shRNAs and siRNAs used in this study

Name	Sequence/ID	Type
sh-Scramble	CCTAAGGTTAAGTCGCCCTCG	shRNA

KDM5B sh-48 (Mouse)	TTCGCTTGTGATGTCGATAAA	shRNA
KDM5B sh-86(Mouse)	ATCGCTTGCTGCACCGTTATT	shRNA
KDM5B sh-87(Mouse)	TCTACTGATTACGAAGGTATT	shRNA
KDM5B sh-91(Mouse)	GCCTACATCATGTGAAAGAAT	shRNA
KDM5B sh-92(Mouse)	CCTGAAATTCAGGAGCTTTAT	shRNA
si-Scramble	Sigma Universal siRNA control	siRNA
KDM5B-si1(Human)	SASI_Hs01_00044944	siRNA
KDM5B-si2(Human)	SASI_Hs01_00044945	siRNA

Table 3: List of plasmids used in this study

Name	Expression	Purpose
pLKO.1-Scramble	Mammalian	KDM5B knockdown control
pLKO.1-sh48	Mammalian	KDM5B knockdown
pLKO.1-sh86	Mammalian	KDM5B knockdown
pLKO.1-sh87	Mammalian	KDM5B knockdown
pLKO.1-sh91	Mammalian	KDM5B knockdown
pLKO.1-sh92	Mammalian	KDM5B knockdown
pCMV-dR8.2	Mammalian	Lentiviral packaging plasmid
pCMV-VSVG	Mammalian	Lentiviral envelop plasmid
pENTR3C-KDM5B	Bacterial	Donor Plasmid
pGEX-6P-1-PHD1	Bacterial	Protein purification
pGEX-6P-1-PHD2	Bacterial	Protein purification
pGEX-6P-1-PHD3	Bacterial	Protein purification
pMAX-DEST	Mammalian	KDM5B overexpression
pLenti-CMV-Puro-DEST	Mammalian	KDM5B overexpression
pEGFP-H2BK43WT-3xFlag	Mammalian	H2B expression
pEGFP-H2BK43-A-3xFlag	Mammalian	H2B mutant expression
pEGFP-H2BK43-R-3xFlag	Mammalian	H2B mutant expression

Table 4: List of primers used in this study

Plasmid	Forward primer	Reverse primer
pGEX-6P-1	GGGCTGGCAAGCCACGTTTGGTG	
pCMV	CGCAAATGGGCGGTAGGCGTG	
pLKO.1	GACTATCATATGCTTACCGT	

Gene	Forward primer	Reverse primer
Oct4	TAGGTGAGCCGTCTTTCCAC	GCTTAGCCAGGTTCGAGGAT
Sox2	AAGAAAGGAGAGAAGTTTGGAGCC	GAGATCTGGCGGAGAATAGTTGG
Nanog	TCTTCCTGGTCCCCACAGTTT	GCAAGAATAGTTCTCGGGATGAA
18S	CGGCTACCACATCCAAGGAA	GCTGGAATTACCGCGGCT
GAPDH	TCACCACCATGGAGAAGGC	GCTAAGCAGTTGGTGGTGCA
Pax6	TACCAGTGTCTACCAGCCAAT	TGCACGAGTATGAGGAGGTCT
Nestin	CCCTGAAGTCGAGGAGCTG	CTGCTGCACCTCTAAGCGA
TrkB	CTGGGGCTTATGCCTGCTG	AGGCTCAGTACACCAAATCCTA
GFAP	CGGAGACGCATCACCTCTG	AGGGAGTGGAGGAGTCATTCG
Ascl	GCAACCGGGTCAAGTTGGT	GTCGTTGGAGTAGTTGGGGG
Actb	TCCTAGCACCATGAAGATC	AAACGCAGCTCAGTAACAG
NeuroD1	ACAGACGCTCTGCAAAGGTTT	GGACTGGTAGGAGTAGGGATG
SYP	CAGTTCCGGGTGGTCAAGG	ACTCTCCGTCTTGTTGGCAC
Dmnt3B	AGCGGGTATGAGGAGTGCAT	GGGAGCATCCTTCGTGTCTG
KDM5B (mouse)	CTGGGAAGAGTTCGCGGAC	CGCGGGGTGAAATGAAGTTTAT
KDM5B (human)	CAATGCTGTGGACCTGTATGT	TACGGAGGGTATAGTCCCTGG

Table 5: List of antibodies used in this study

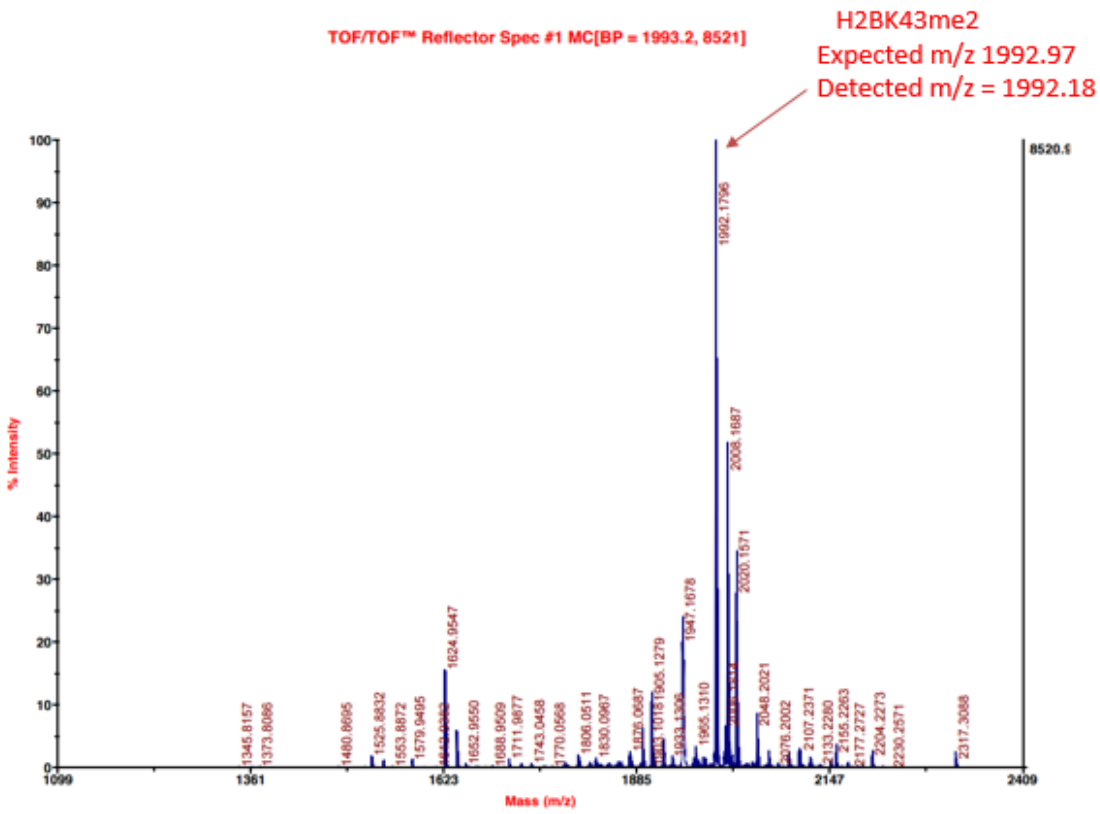
Target	Company	Dilution
KDM5B/JARID1B	Abcam	1:5000
Beta-actin	Santa-Cruz	1:2000
Histone H2B	EMD	1:5000
Histone H3K4me3	Active motif	1:5000
Histone H3K9me3	Active motif	1:5000
Histone H2BK43me2	Abcam	1:2000
Nestin	Santa-Cruz	1:100
GFAP	Sigma	1:100
ND-160	Sigma	1:100
Beta-III tubulin	Santa-Cruz	1: 100
FLAG	Sigma	1:5000
GST	Santa-Cruz	1:2000

Table 6: Computer-predicted new histone methylation sites

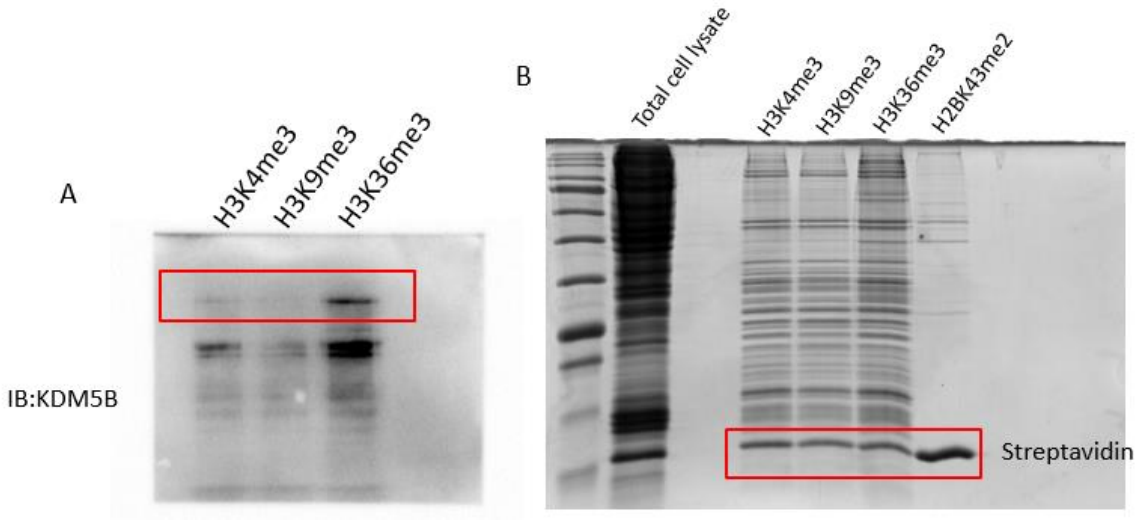
Prediction	ID	Name	Site	Prediction	ID	Name	Site
0.95	P07305	H1F0	147	0.954	P16403	HIST1H1C	148
0.948	P07305	H1F0	172	0.95	P16403	HIST1H1C	194
0.944	P07305	H1F0	193	0.949	P16403	HIST1H1C	183
0.94	P07305	H1F0	136	0.945	P16403	HIST1H1C	206
0.937	P07305	H1F0	159	0.944	P16403	HIST1H1C	172
0.909	P07305	H1F0	122	0.935	P16403	HIST1H1C	137
0.85	P07305	H1F0	111	0.895	P16403	HIST1H1C	110
0.932	Q8IZA3	H1FOO	189	0.891	P16403	HIST1H1C	122

0.924	Q8IZA3	H1FOO	175	0.858	P16403	HIST1H1C	97
0.921	Q8IZA3	H1FOO	164	0.848	P16403	HIST1H1C	23
0.9	Q8IZA3	H1FOO	216	0.942	P16402	HIST1H1D	207
0.894	Q8IZA3	H1FOO	264	0.94	P16402	HIST1H1D	220
0.883	Q8IZA3	H1FOO	237	0.938	P16402	HIST1H1D	184
0.88	Q8IZA3	H1FOO	139	0.931	P16402	HIST1H1D	150
0.872	Q8IZA3	H1FOO	153	0.927	P16402	HIST1H1D	161
0.871	Q8IZA3	H1FOO	280	0.919	P16402	HIST1H1D	138
0.847	Q8IZA3	H1FOO	298	0.91	P16402	HIST1H1D	173
0.832	Q8IZA3	H1FOO	327	0.899	P16402	HIST1H1D	120
0.927	Q92522	H1FX	195	0.827	P16402	HIST1H1D	26
0.912	Q92522	H1FX	166	0.952	P10412	HIST1H1E	130
0.909	Q92522	H1FX	213	0.95	P10412	HIST1H1E	149
0.907	Q92522	H1FX	180	0.947	P10412	HIST1H1E	182
0.875	Q92522	H1FX	146	0.945	P10412	HIST1H1E	197
0.863	O75367	H2AFY	8	0.944	P10412	HIST1H1E	218
0.961	P16401	HIST1H1B	150	0.943	P10412	HIST1H1E	169
0.96	P16401	HIST1H1B	161	0.909	P10412	HIST1H1E	117
0.956	P16401	HIST1H1B	133	0.871	P10412	HIST1H1E	90
0.955	P16401	HIST1H1B	207	0.85	P10412	HIST1H1E	34
0.954	P16401	HIST1H1B	182	0.87	Q96A08	HIST1H2BA	25
0.951	P16401	HIST1H1B	194	0.845	Q96A08	HIST1H2BA	14
0.95	P16401	HIST1H1B	220	0.83	Q96A08	HIST1H2BA	36

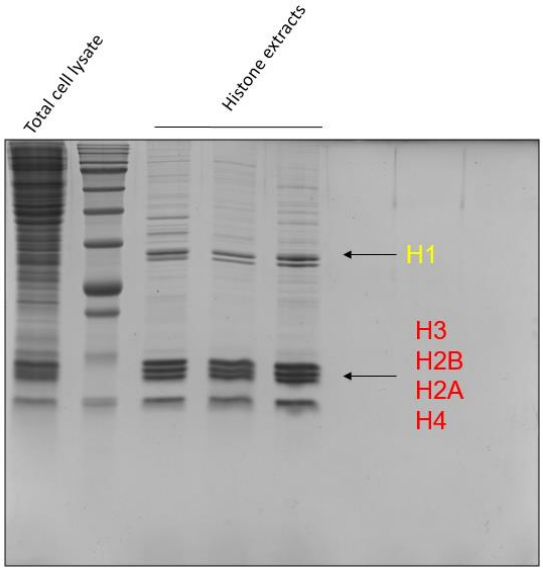
Appendices



Appendix A. MALDI spectrum example of synthetic histone peptide H2BK43me2. The major peak should have mass (m/z) value within 1 Dalton difference compared to theoretical mass. Only 0.79 Daltons difference was observed in the spectrum above, confirmed the identity of the H2BK43me2 peptide.



Appendix B. A) Uncropped WB membrane of KDM5B histone peptide pull-downs. KDM5B (red box) is about 180 kDa. B) Double amount of H2BK43me2 beads yielded similar pull-down result. Coomassie stained SDS-PAGE showed the amount of streptavidin (red box) used in the pull-down. Streptavidin beads were boiled in 100ul of 1x loading buffer to release any bound proteins. Each lane was loaded with 20ul boiled sample.



Appendix C. Coomassie stained SDS-PAGE of acid extract histones. Approximately 100ug of histone proteins could be extracted from 1 x 10^7 cells. Each line was loaded with about 2ug of histone extracts. Histone extracts should display very low background of non-histone proteins.

(1) NheI_Kozak_Q64475_3xFlag_Stop_BamHI

GCTAGCGCCACCATGGCACCTGAGCCCTCCAAGTCCGCTCCAGCACCTA
AGAAGGGAAGCAAAAAAGCCATCAGCAAGGCGCAGAAGAAAGACGGC
AAGAAACGCAAGAGGAGCCGTAAGGAGAGTTATAGCGTCTACGTCTATA
AGGTGTTGAAGCAAGTGCACCCTGACACCGGGATCAGTTCCAAGGCCAT
GGGCATCATGAACTCTTTCGTGAACGATATTTTTGAGAGAATCGCGTCTG
AGGCGTCTAGACTGGCCCACTATAATAAGAGATCCACCATAACCAGCCGG
GAGATCCAGACTGCCGTGAGGCTGCTGCTTCCCGGAGAACTAGCAAAGC
ATGCCGTGAGTGAGGGCACAAAAGCAGTGACTAAGTACACAAGTAGCA
AGGATTATAAGGATCATGATGGGGATTATAAGGATCATGATATAGATTATA
AGGATGATGATGATAAGTGAGGATCC

Q64475 (H2B1B_MOUSE) H2BWT + 3X FLAG
MAPEPSKSAPAPKKGSKKAISKAQKKDGKKRKRSRKESYSVYVYKVLKQV
HPDTGISSKAMGIMNSFVNDIFERIASEASRLAHYNKRSTITSREIQTAVRLL
LPGELAKHAVSEGTKAVTKYTSSKDYKDHDGDYKDHDIDYKDDDDK

(2) NheI_Kozak_Q64475(K43R)_3xFlag_Stop_BamHI

GCTAGCGCCACCATGGCTCCTGAGCCAAGTAAAAGCGCCCCTGCTCCTA
AGAAGGGAAGTAAGAAGGCAATTAGCAAGGCACAGAAAAAGGATGGA
AAAAAGCGGAAGAGATCTCGTAAAGAGTCCTACAGTGTGTATGTGTACC
GGGTGCTAAAACAAGTGCACCCGGATACTGGCATCAGCAGTAAAGCGAT
GGGCATTATGAACAGCTTTGTCAACGACATATTCGAGCGCATCGCCAGTG
AGGCTAGTAGGTTAGCCCATTACAACAAGCGTAGCACGATTACATCTAGA
GAGATCCAGACGGCAGTAAGGCTCCTTCTTCCCGGCGAGCTGGCTAAGC
ACGCCGTCAGCGAGGGGACCAAAGCCGTGACGAAGTACACCTCTTCCA
AGGATTATAAGGATCATGATGGGGATTATAAGGATCATGATATAGATTATA
AGGATGATGATGATAAGTGAGGATCC

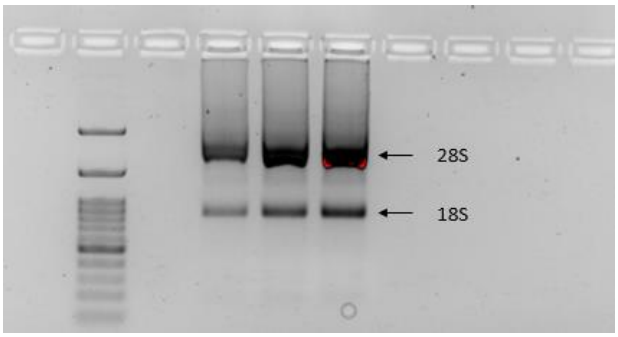
Q64475 (H2B1B_MOUSE) H2B43K-R + 3X FLAG
MAPEPSKSAPAPKKGSKKAISKAQKKDGKKRKRSRKESYSVYVYRVLKQV
HPDTGISSKAMGIMNSFVNDIFERIASEASRLAHYNKRSTITSREIQTAVRLL
LPGELAKHAVSEGTKAVTKYTSSKDYKDHDGDYKDHDIDYKDDDDK

(3) NheI_Kozak_Q64475(K43A)_3xFlag_Stop_BamHI

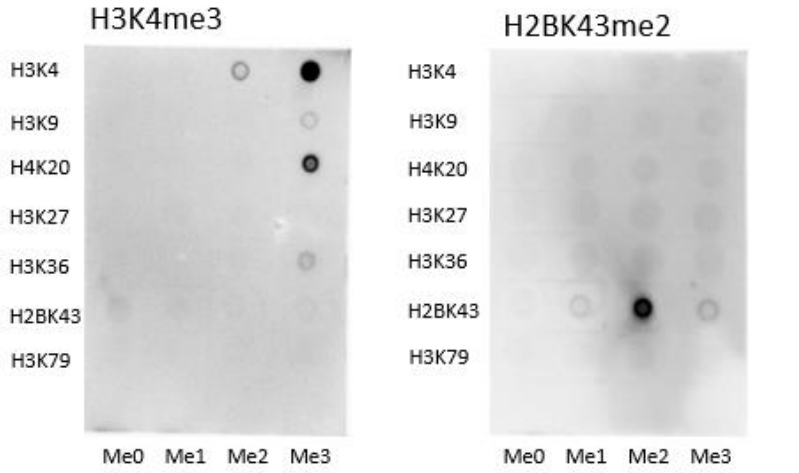
GCTAGCGCCACCATGGCCCCCGAGCCCAGCAAGAGCGCCCCCGCCCCCA
AGAAGGGCAGCAAGAAGGCCATCAGCAAGGCCCAGAAAAAGGATGGG
AAGAAGCGGAAGAGGAGCAGGAAGGAGAGCTACTCCGTCTACGTGTAC
GCCGTGCTGAAGCAGGTGCACCCCGATACCGGCATCAGCAGCAAGGCC
ATGGGCATCATGAACAGCTTCGTGAACGACATCTTCGAGCGGATCGCCA
GCGAGGCCTCAAGGCTGGCCCACTACAACAAGAGGAGCACAATCACAA
GCCGGGAGATCCAGACTGCCGTGAGACTGCTGCTGCCAGGAGAGCTGG
CTAAGCACGCCGTGAGTGAGGGGACAAAGGCCGTGACAAAGTACACTA
GCAGTAAGGATTATAAGGATCATGATGGGGATTATAAGGATCATGATATAG
ATTATAAGGATGATGATGATAAGTGAGGATCC

Q64475 (H2B1B_MOUSE) H2B43K-A + 3X FLAG
MAPEPSKSAPAPKKGSKKAISKAQKKDGKKRKRSRKESYSVYVYAVLKQV
HPDTGISSKAMGIMNSFVNDIFERIASEASRLAHYNKRSTITSREIQTAVRLL
LPGELAKHAVSEGTKAVTKYTSSKDYKDHDGDYKDHDIDYKDDDDK

Appendix D. Sequencing results of histone H2B WT and mutant constructs. Restriction sites were highlighted in yellow. 3X FLAG sequences (both DNA and protein) were highlighted in green. Specific mutations were highlighted in purple. Note that the cDNA sequences were optimized by the supplier to enhance stability and expression efficiency.



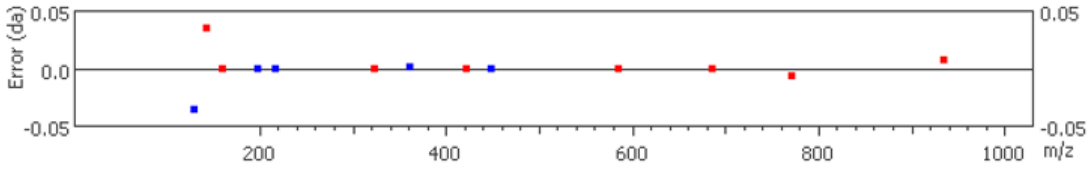
Appendix E. Example of 1% bleach gel to confirm mRNA integrity. Intact mRNA have the 28S band about twice brighter than 18S. Also, no bands/smears were detected below 18S indicated that degradation did not occur.



Appendix F. Validation of antibodies via dot-blot assay. H3K4me3 antibody used in this study showed strong cross reaction with H4K20me3. However, considerable molecular weight difference between histone H3 and histone H4 allowed us to distinguish them. H2BK43me2 antibody showed high specificity in this dot-blot assay. Peptide amount was normalized by Ponceau S. staining.

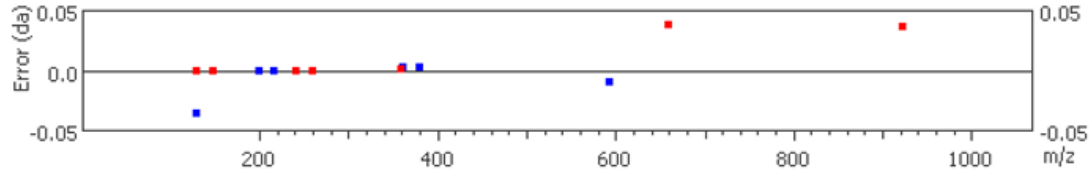
H2BK43me1

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	130.09	112.04	113.02	65.53	E					9
2	217.08	199.07	200.06	109.04	S	1022.52	1004.51	1005.49	511.76	8
3	380.15	362.13	363.12	190.57	Y	935.48	917.48	918.46	468.24	7
4	467.18	449.17	450.15	234.09	S	772.43	754.41	755.40	386.71	6
5	566.25	548.24	549.22	283.62	V	685.39	667.38	668.36	343.20	5
6	729.31	711.30	712.28	365.15	Y	586.32	568.31	569.30	293.66	4
7	828.38	810.37	811.35	414.69	V	423.26	405.25	406.23	212.13	3
8	991.44	973.43	974.41	496.22	Y	324.19	306.18	307.16	162.60	2
9					K(+14.02)	161.13	143.12	144.07	81.06	1



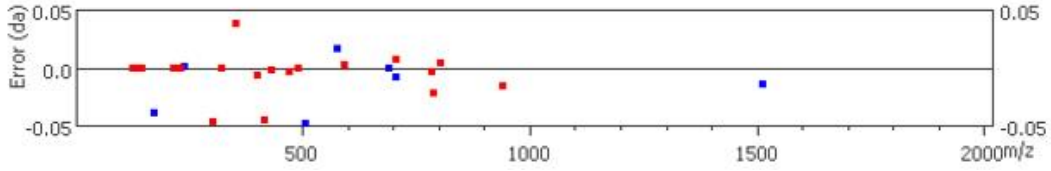
H2BK43me2

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	130.09	112.04	113.02	65.53	E					12
2	217.08	199.07	200.06	109.04	S	1404.81	1386.80	1387.79	702.91	11
3	380.14	362.13	363.12	190.57	Y	1317.78	1299.77	1300.75	659.39	10
4	481.19	463.18	464.17	241.10	S(+14.02)	1154.72	1136.71	1137.69	577.86	9
5	594.29	576.27	577.25	297.64	I	1053.67	1035.66	1036.64	527.34	8
6	757.34	739.33	740.31	379.17	Y	940.59	922.54	923.56	470.79	7
7	856.41	838.40	839.38	428.70	V	777.52	759.51	760.50	389.26	6
8	1019.47	1001.46	1002.45	510.24	Y	678.45	660.40	661.43	339.73	5
9	1175.60	1157.59	1158.57	588.30	K(+28.03)	515.39	497.38	498.36	258.20	4
10	1274.67	1256.66	1257.64	637.83	V	359.26	341.25	342.24	180.13	3
11	1387.75	1369.74	1370.72	694.38	L	260.20	242.19	243.17	130.60	2
12					K	147.11	129.10	130.09	74.06	1

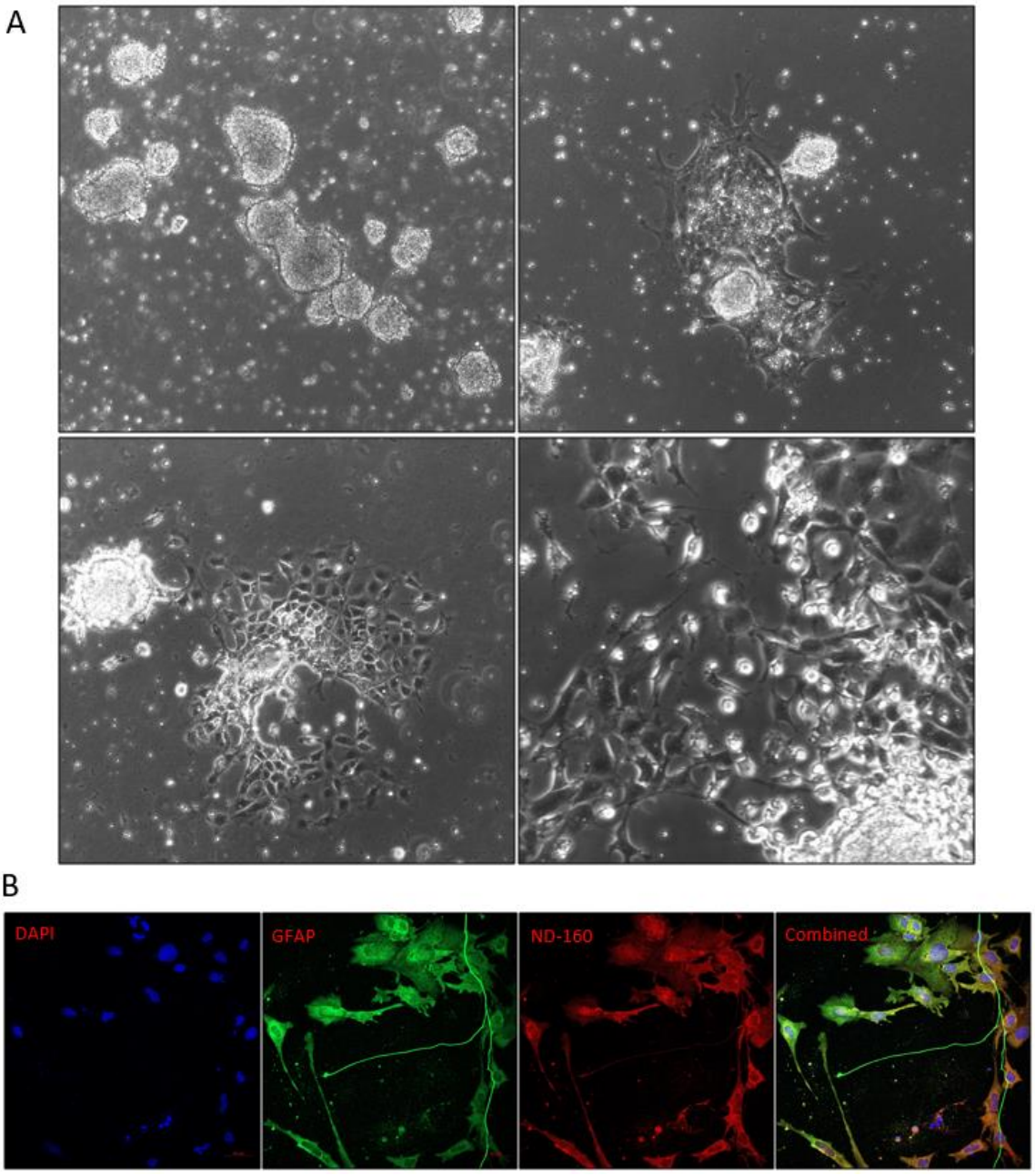


H2BK43me3

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.10	111.09	112.08	65.05	K					24
2	258.15	240.13	241.12	129.57	E	2739.50	2721.49	2722.47	1370.25	23
3	345.18	327.17	328.15	173.13	S	2610.45	2592.44	2593.43	1305.73	22
4	508.29	490.23	491.21	254.62	Y	2523.42	2505.41	2506.40	1262.21	21
5	595.27	577.24	578.25	298.14	S	2360.36	2342.35	2343.33	1180.68	20
6	708.37	690.35	691.33	354.68	I	2273.33	2255.32	2256.30	1137.16	19
7	871.42	853.41	854.39	436.21	Y	2160.24	2142.23	2143.22	1080.62	18
8	984.50	966.49	967.48	492.75	I	1997.18	1979.17	1980.15	999.09	17
9	1147.57	1129.56	1130.54	574.28	Y	1884.10	1866.08	1867.07	942.55	16
10	1317.71	1299.70	1300.68	659.35	K(+42.05)	1721.03	1703.02	1704.01	861.02	15
11	1416.78	1398.77	1399.75	708.89	V	1550.89	1532.88	1533.86	775.95	14
12	1529.86	1511.85	1512.85	765.43	L	1451.82	1433.81	1434.79	726.41	13
13	1700.00	1681.99	1682.98	850.50	K(+42.05)	1338.74	1320.73	1321.71	669.87	12
14	1828.06	1810.05	1811.04	914.53	Q	1168.60	1150.59	1151.57	584.80	11
15	1927.13	1909.12	1910.10	964.07	V	1040.54	1022.53	1023.51	520.77	10
16	2064.19	2046.18	2047.16	1032.59	H	941.48	923.46	924.44	471.23	9
17	2161.24	2143.23	2144.22	1081.12	P	804.40	786.40	787.41	402.71	8
18	2276.27	2258.26	2259.24	1138.63	D	707.35	689.35	690.33	354.14	7
19	2377.32	2359.31	2360.29	1189.16	T	592.33	574.32	575.30	296.67	6
20	2434.34	2416.33	2417.31	1217.67	G	491.28	473.28	474.26	246.14	5
21	2547.42	2529.41	2530.40	1274.21	I	434.26	416.30	417.23	217.63	4
22	2634.45	2616.44	2617.43	1317.73	S	321.18	303.21	304.15	161.09	3
23	2721.49	2703.48	2704.46	1361.24	S	234.14	216.13	217.12	117.57	2
24					K	147.11	129.10	130.09	74.06	1



Appendix G. Raw data of H2BK43me1/2/3 daughter ions detected by MS in histone extracts. From left to right, first five columns indicated positions and detected molecular weights of b cleavages (ie: b1 of H2BK43me3 contained only K). b, b-H_2O, b-NH_3 and b (2+) were different status of b ions. For example, b1 ion had same molecular weight equals to K; b-H_2O ion had molecular weight equals to K minus H_2O; b-NH_3 ion had molecular weight equals to K minus NH_3; b (2+) ion had half molecular weight of K. Sample principle applied to y ions in the last five columns. Red and blue highlighted numbers indicated detections with high confidence.



Appendix H. Phase-contrast and confocal microscopy of differentiated mESCs in the neurogenesis axis. **A)** From upper left to lower right were day 3, day 5, day 7 and day 9 cells in differentiation media. Day 9 picture was taken under 200X total magnifications while all other pictures were taken under 100X total magnifications. Neuronal networks appeared clearly at Day 9. **B)** Immunostaining confocal microscopy of Day 12 differentiated cells. Cells that displayed strong signal of GFAP showed significantly weaker signal of ND-160, suggested there was a mixed population in the culture.

Curriculum Vitae

Name:	Qi Fang
Post-secondary Education and Degrees:	Seneca College Toronto, Ontario, Canada 2007-2010 Advanced Diploma
	York University Toronto, Ontario, Canada 2012-2015 B.Sc.
	The University of Western Ontario London, Ontario, Canada 2015-2017 M.Sc.
Honours and Awards:	Dean's Honor Roll 2015
	Western Graduate Research Scholarship 2015-2017