

Aus dem Adolf-Butenandt-Institut der Ludwig-Maximilians-Universität  
München

Lehrstuhl: Molekularbiologie Direktor: Prof. Dr. Peter B. Becker  
Arbeitsgruppe: Prof. Dr. Axel Imhof

---

## Characterization of histone lysine demethylation

---



Dissertation zum Erwerb des Doktorgrades der Naturwissenschaften

(Dr. rer. Nat.) an der Medizinischen Fakultät

Der Ludwig-Maximilians-Universität München

Vorgelegt von

**Jianhua Li**

Aus Shanxi, P.R. China

München, 2015



---

**Mit Genehmigung der Medizinischen Fakultät der Ludwig  
Maximilians-Universität München**

Betreuer: Prof. Dr. Axel Imhof

Zweitgutachterin: Priv. Doz. Dr. Dejana Mokranjac

Dekan: Prof. Dr. med. denf. Reinhard Hickel

Tag der mündlichen Prüfung: 16.02.2016

---

## **Eidesstattliche Versicherung**

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema

### **“Characterization of histone lysine demethylation”**

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum

Unterschrift Jianhua Li



<b>1. SUMMARY</b>	<b>6</b>
<b>2. INTRODUCTION</b>	<b>8</b>
<b>2.1 Histone lysine methylation</b>	<b>10</b>
2.1.1 Dynamics of histone lysine methylation	12
2.1.2 Histone lysine demethylases	14
<b>2.2 Previous work on KDM4a</b>	<b>16</b>
2.2.1 Enzymatic related KDM4a function	16
2.2.2 Interacting factors of KDM4a, H3K36me3 and H3K9me3	19
<b>2.3 Therapeutic strategy targeting epigenetic modifiers</b>	<b>20</b>
2.3.1 Disease related KDM4 epigenetics	20
2.3.2 Epigenetic therapies	22
<b>2.4 Application of Mass Spectrometry</b>	<b>23</b>
2.4.1 Mass Spectrometry based proteomics	23
2.4.2 Analysis of PTM's by Mass Spectrometry	24
<b>2.5 Metabolomics and epigenetics</b>	<b>24</b>
2.5.1 Influence of metabolites on epigenetic marks	25
2.5.2 Regulation of metabolic enzyme	25
2.5.3 Characterization of IDH	25
<b>2.6 Aim of the work</b>	<b>27</b>
<b>3. RESULTS</b>	<b>29</b>
<b>3.1 KDM4a inhibitor screening</b>	<b>29</b>
3.1.1 Selection of small molecular fragments	29
3.1.2 Potential targets of selected fragments: Jumonji domain containing proteins in human and fruit fly	30
3.1.3 Establish demethylase enzyme activity assay <i>in vitro</i>	31
3.1.4 KDM4 inhibitors <i>in vitro</i> investigation	33
3.1.5 dKDM4a demethylase activity towards H3K36me3 <i>in vivo</i>	36
3.1.6 Inhibitor JMJ-1, JMJ-4 and JMJ-6 inhibited dKDM4a demethylase activity towards H3K36me3 <i>in vivo</i> verified by antibody based substrate detection	37
3.1.7 Dosage response of L2-4 cells growth curve to inhibitors	40
3.1.8 Inhibitor JMJ-1, JMJ-4 and JMJ-6 inhibited dKDM4a demethylase activity towards H3K36me3 <i>in vivo</i> verified by targeted proteomics based substrate detection	41
3.1.8.1 Identification of H3K36me3 among H3.27.40 peptides by LC-MS/MS	43
3.1.8.2 Quantitation of dKDM4a substrates H3K36me3/me2 and H3K9me3/me2 by LC-MS/MS	45
<b>3.2 Functional analysis of dKDM4a</b>	<b>46</b>
<b>3.3 Single nuclear staining of H3K36me3 in bi-nucleated SL2 and L2-4 cells</b>	<b>49</b>
<b>3.4 Over-expression of IDH led to a reduction of H3K36me3</b>	<b>49</b>
3.4.1 Over-expression of IDH led to reduction of dKDM4a substrate H3K36me3	50
3.4.2 Global histone methylation analysis upon over-expression of IDH	50
3.4.3 Generating IDH mutants cell line	53
<b>4. DISCUSSION</b>	<b>54</b>

<b>4.1 KDM4a inhibitor screening</b>	<b>54</b>
<b>4.2 Function of KDM4a in human and <i>Drosophila melanogaster</i></b>	<b>59</b>
<b>4.3 Human disease related genes therapeutic study using fly cell</b>	<b>61</b>
<b>4.4 Mass spectrometry application</b>	<b>62</b>
<b>4.5 H3K36me3 quantitation and its biological function</b>	<b>64</b>
4.5.1 Quantitation of H3K36me3 methylation by WB and IF	64
4.5.2 Quantitation of H3K36me3 methylation by targeted LC-MS/MS analysis	64
4.5.3 Biological function of H3K36me3	65
<b>4.6 Metabolomics and epigenetics</b>	<b>66</b>
<b>5. MATERIAL AND METHODS</b>	<b>67</b>
<b>5.1 Material</b>	<b>67</b>
5.1.1 Inhibitors investigated in this study	67
5.1.2 Chemicals	68
5.1.3 DNA and Protein Markers	68
5.1.4 Enzymes and Kits	68
5.1.5 Primers	68
5.1.6 Buffers	70
5.1.6.1 Buffer composition for DNA samples	70
5.1.6.2 Buffer composition for protein samples	70
5.1.7 Vectors	71
5.1.8 Plasmid	71
5.1.9 Bacteria strains <i>E. coli</i>	71
5.1.10 Cell culture material	72
5.1.11 Chromatography material	72
5.1.12 Antibodies	73
5.1.13 Synthetic peptides	73
<b>5.2 Methods</b>	<b>76</b>
5.2.1 General DNA and RNA sample methods	76
5.2.2 Cell culture	76
5.2.3 Immune Fluorescence staining for microscope detection	78
5.2.4 General protein sample methods:	79
5.2.5 lysine demethylation <i>in vitro</i> assay	81
5.2.6 Proteomics sample preparation	83
5.2.7 Histone sample preparation	85
5.2.7.1 Acid Extraction of Histones from <i>Drosophila</i> SL2 or L2-4 cells	85
5.2.7.2 Histone in-gel tryptic digestion	85
5.2.7.3 Peptide desalting by C 18 stage tips and carbon tips	86
5.2.7.4 Histone sample LC/MS/MS measurement	86
5.2.7.5 Histone modification analysis	87
<b>6. ABBREVIATION AND APPENDIX</b>	<b>88</b>
<b>6.1 Abbreviation</b>	<b>88</b>
<b>6.2 Appendix 1</b>	<b>90</b>
<b>6.3 Appendix 2</b>	<b>91</b>
<b>6.4 Appendix 3</b>	<b>92</b>

---

<b>7. REFERENCES</b>	<b>95</b>
<b>8. ACKNOWLEDGEMENT</b>	<b>107</b>
<b>9. PUBLICATIONS</b>	<b>108</b>
<b>CURRICULUM VITAE</b>	<b>109</b>

## 1. Summary

Histone lysine methylation plays crucial role in regulating gene activation and repression. Tri-methylation on histone H3K36 is considered as an active mark and enriched in the 3' end of actively transcribed gene body. KDM4a is a Jumonji domain containing lysine specific demethylase. It selectively demethylates H3K36me3/me2 to H3K36me2/me1 and H3K9me3/me2 to H3K9me2/me1. Apart from its demethylase activity, so far it is not clear about its biological function and the regulation mechanism of KDM4a.

In this thesis work, *in vitro* and *in vivo* demethylation assay have been established to validate demethylase activity of *Drosophila* KDM4a towards H3K36me3. Inhibitors targeting KDM4a have been investigated using antibody based substrate detection and mass spectra based substrate detection. I have shown that inhibitor JMJ-1, JMJ-4 and JMJ-6 inhibited demethylase activity of dKDM4a towards H3K36me3 *in vitro* and *in vivo*. Inhibition assay results of JMJ-1, JMJ-4 and JMJ-6 from western blotting, immunofluorescence imaging and targeted proteomics study agreed with each other.

LC-MS based proteomics approach has been applied to study the interactor partners of dKDM4a. MSL1, MSL3 and HP1a have been validated by both mass spectrometry analysis and western blotting detection.

We observed single nuclear staining of H3K36me3 in bi-nucleated SL2 and L2-4 cells.

Regulation of dKDM4a demethylase activity can be from the enzyme itself or the cofactors it requires. KDM4a requires  $\alpha$ -ketoglutarate,  $\text{Fe}^{2+}$  and  $\text{O}_2$  as co-factors to perform demethylation.  $\alpha$ -ketoglutarate is produced by IDH. In the last part of my thesis, I established stable cell line with IDH over-expression. Over-expression of IDH generated more  $\alpha$ -ketoglutarate. Interestingly, we observed reduction of H3K36me3 upon IDH over-expression suggesting stimulation of demethylation by Jmj-C containing KDMs.

## Zusammenfassung

Die Methylierung von Lysinen in Histonen spielt eine entscheidende Rolle bei der Regulation der Gen-Aktivierung und Hemmung. Eine dreifache Methylierung von Lysin 36 am Histone H3 ist ein Zeichen aktiver Transkription und markiert das 3'-Ende des aktiv transkribierten Gens. Die Lysin spezifische Demethylase KDM4a enthält eine sogenannte Jumonji und demethyliert H3 Moleküle die an Lysin beziehungsweise an Lysin 9 methyliert sind. Zu Beginn meiner Doktorarbeit war die biologische Funktion sowie die Regulation dieses wichtigen Enzyms weitestgehend unverstanden.

In meiner Doktorarbeit habe ich ein Testverfahren zur Messung der Demethylase Aktivität von KDM4a *in vitro* etabliert und damit neuartige Inhibitoren untersucht und validiert. Dabei konnte ich zeigen dass die Inhibitoren JMJ-1, JMJ-4 und JMJ-6 in der Lage sind die Demethylase-Aktivität von dKDM4a an H3K36me3 *in vitro* und *in vivo* effizient zu hemmen. Die Ergebnisse wurden anschliessend mit Hilfe unterschiedlicher Methoden wie dem Western-Blotting, der Immunfluoreszenz und der gerichteten Proteomik bestätigt.

Neben der Untersuchung der Inhibitoren konnte ich mit Hilfe der proteomischen Analyse spezifische Interaktoren von KDM4a identifizieren. Die interessanten Kandidaten MSL1, MSL3 und HP1a konnte ich im Western Blotting bestätigen.

Um die Regulation des KDM4a Enzyms durch die notwendigen Kofaktoren  $\alpha$ -Ketoglutarat,  $\text{Fe}^{2+}$  und  $\text{O}_2$  besser zu verstehen, habe ich im letzten Teil der Arbeit stabile Zelllinien generiert, die das Enzym Isocitrat Dehydrogenase (IDH) überexprimieren. Interessanterweise führte eine Überexpression des IDH Enzyms zu einer Reduktion der intrazellulären H3K36me3 Spiegel was auf eine Stimulierung der Demthylasen hindeutet. Diese Regulation der Enzymaktivität durch Metabolite deutet auf eine interessante Verbindung zwischen der epigenetischen Information und dem metabolischen Zustand einer Zelle hin.

## 2. Introduction

The word ‘epigenetics’ was introduced by Waddington to combine the fields of developmental biology and genetics in 1950s (Waddington, 1952). It mainly indicates heritable information that is independent of DNA sequence (Spannhoff et al., 2009). Epigenetic mechanisms involve DNA methylation, chromatin structure, nucleosome remodeling, non-coding regulatory RNA, histone variants and histone modifications (Katada et al., 2012).

The major difference between genetics and epigenetics is whether changes are reversible or not. Genetic changes are stable and irreversible, while epigenetic changes are generally not stable. The epigenotype is defined as actual gene expression pattern of a specific cell lineage. All cell lineages derived from an organism share the expression profile of common housekeeping genes, which are necessary for basic cell metabolism. Specified cell lineages harbor specialized functions, which are performed by cell type specific proteins. The epigenotype of a given cell lineage indicates not only basic household genes, specific genes but also the repressed genes (Holliday et al., 2006).

Among all of the epigenetic mechanisms, DNA methylation was the earliest reported epigenetic event. It has been described by Griffith and Mahler in 1969. Isoschizomers are a pair of restriction enzymes which could recognize the same DNA sequences; one digests un-methylated version and the other one cuts methylated version (Holliday et al., 2006). Discovery of DNA methylation was based on characterization of these two enzymes reading DNA methylation mark.

Genetic information is highly organized in chromatin. Chromatin is a structure only found in eukaryotic organism. Chromatin can be categorized into two major types such as euchromatin and heterochromatin according to their degree of condensation (Sansoni et al., 2014).

Chromatin forms a compact structure, which intrinsically has a repressive function for gene expression. Basic unit of chromatin is nucleosome. Nucleosome remodeling can either tighten or loosen the DNA wrapped on histone octamer (Becker & Hörz, 2002). This can compact or open chromatin to regulate access of transcription factors to underlying DNA sequence (Zhang et al., 2011).

Regulation of gene expression can be manipulated not only by transcription factors but also regulatory non-coding RNAs on chromatin. For instance, Maenner et al. described roX RNA served as scaffold to recruit and activate dosage compensation complex (DCC). This has been done by changing secondary structure of roX RNA stem loop via ATP dependent RNA helicase MLE. MLE is one of the subunit of DCC complex (Maenner et al., 2013).

Presence of different histone variants can change the structure of chromatin and alter gene expression profile by recruiting different transcription factors or a specific chaperone (Banaszynski et al., 2010). For instance, apart from canonical histone H3.1 and H3.2, there are two major variants for histone H3: H3.3 and CENP-A. H3.1 and H3.2 RNA expression peak during S phase in a replication dependent manner. H3.3 RNA is found during the whole cell cycle. H3.3 protein gets incorporated into actively transcribed chromatin regions (Hake et al., 2006; Biterge et al., 2014). H3.3 variant plays an essential role with chromosome decondensation during spermatogenesis both in *Mus* and *Drosophila* (Heijden et al., 2007; Rogers et al., 2004; Wen et al., 2014; Loyola et al., 2007). CENP-A is also called centromere identifier (CID) in *Drosophila*. It is the H3 variant specifically localizing in centromeric chromatin region together with constitutive centromere associated network proteins (Barth et al., 2014). The phosphorylated  $\gamma$ H2A.X is a H2A variant specifically recruited to the DNA damage site by activated DNA damage repair machinery (Panier et al., 2014).

The entire genome is duplicated in S phase before cell division for both mitosis and meiosis. Therefore duplication of histone protein is required in order to assemble and form new chromatin. Heritable epigenetic information is highly controlled by post-translational modifications (PTMs) maintained on histones. And it is conserved during histone protein duplication (Scharf et al., 2009). Temporal and spatial gene expression profile is achieved by activating defined set of genes and meanwhile repressing another set of genes via dynamic PTMs on histones (Imhof, 2004).

Histone lysine methylation has been intensively reported for residue sites including K4, K9, K27, K36 and K79 on histone H3 (Labbé et al., 2014), and K20 on histone H4. Meanwhile histone arginine methylation has also been characterized on residue sites R2, R8, R17, R26 on histone H3 and R3 on histone H4. There are 18 acetylation sites assigned to histone lysine residue including K9, K14, K18, K23, K27, K36 and K56 on histone H3; and K5, K8, K12, K16 on histone H4; K5, K9, K13 on histone H2A; K5, K12, K15, K20 on histone

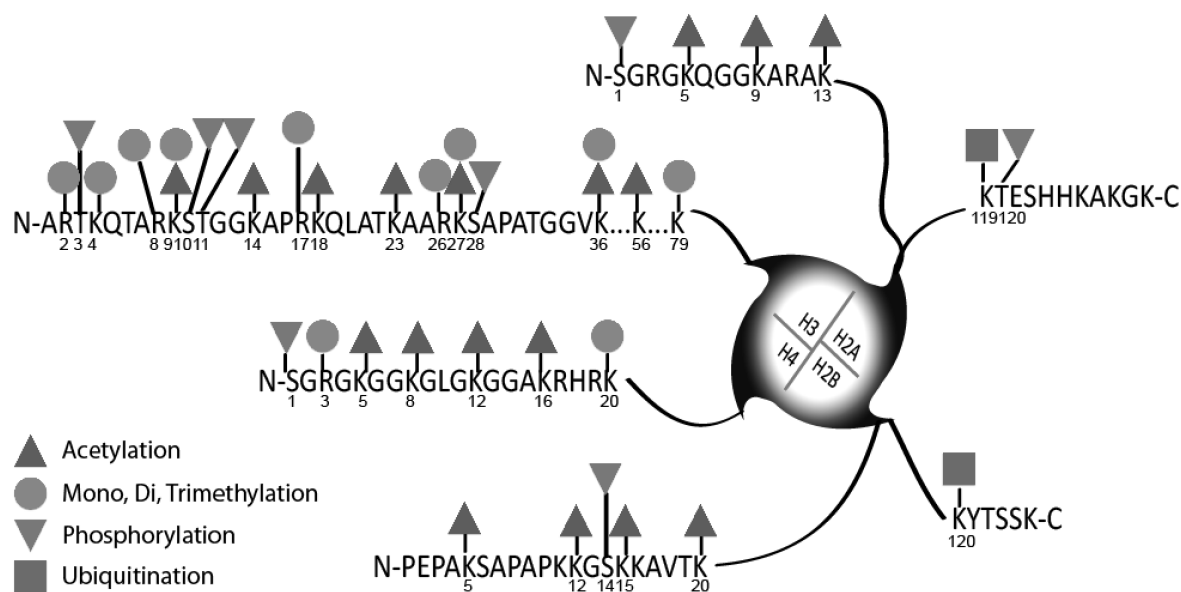
H2B (Bhaumik et al., 2007). Eight known phosphorylation sites on histones are T3, S10, T11 and S28 on histone H3; S1 on histone H4; S1 and T120 on histone H2A; S14 on histone H2B. Meanwhile, K119 on histone H2A and K120 on histone H2B are found to be ubiquitinated (Portela & Esteller, 2010; Figure 1.1).

## 2.1 Histone lysine methylation

Genetic information is stored in chromatin. The nucleosome is formed by 147 base pairs of DNA wrapping around a histone octamer and locked by linker histone protein H1 (Luger et al., 1997). Histones are small basic proteins conserved among all eukaryotic species. They have very high binding affinity to DNA. They share similar structure feature with a hydrophobic globular folding domain and a stretched out flexible N-terminus and C-terminus region (Mosammaparast et al., 2010). Histones can be covalently modified. Histone modifications take place on different residue sites, different states and in a manner of different combinations. Posttranslational histone modifications include methylation, acetylation, phosphorylation, ubiquitination, glycosylation, sumoylation, ADP-ribosylation and so on (Holliday et al., 2006). Regulation of chromatin structure is achieved via histone variants and different PTMs on histone flexible tails to recruit variety of functional proteins.

Histone lysine methylation on different residues and different states (mono-, di-, tri-) can serve as repressive mark or active mark. It reflects the chromatin structure to be closed or open. For instance, H3K4me3 is found in promoter region. It is characterized as a transcription activation mark together with initiating RNA Pol II occupancy for protein encoding genes (Guenther et al., 2007). H3K9me3 is a hallmark for heterochromatin region. It is recognized by the chromo-domain of Heterochromatin Protein HP1. And it acts as a repressive chromatin signature. H4K20me3 indicates silenced genes and is localized together with H3K9me3 in pericentric heterochromatin region (Schotta et al., 2004). H3K27me3 is described as a repressive mark maintaining the HOX genes in silent state. It is associated with Polycomb Repressive Complexes PRC2 and PRC1 (Cao et al., 2002; Eskeland et al., 2010). H3K36me3 is an active mark localized in actively transcribed gene body together with elongating RNA Pol II. It plays a crucial role in transcriptional elongation (Bell et al., 2007). H4K20me1 is set by KMT5a (also called PR-Set7 or Set8) and maintains silent chromatin (Nishioka et al., 2002). H4K20me2 catalyzed by Suv4-20h1/h2, can recruit Origin Recognition Complex (ORC) which leads to initiation of replication (Beck et al., 2012).





**Figure 1.1 | Common PTMs on canonical core histone H2A, H2B, H3 and H4.**

Histones are small basic proteins modified by post-translational modifications on N-terminus and C-terminus tails. The canonical core histones of human consist of H2A (130 aa), H2B (126 aa), H3 (136 aa) and H4 (103 aa). They are highly conserved across eukaryotic species. The most abundant PTMs assigned to histone lysine are acetylation and methylation. Lysine methylation marks are intensively reported on H3-K4, H3-K9, H3-K27, H3-K36 and H4-K20 (Guide to epigenetic marks card, Abcam). They are associated with most known functional role. Amino acid lysine, arginine, serine and threonine are present in histone N-terminus tail in high frequency. Histone PTMs not only functions as single modification but also as combinatorial form within the same peptide. They also function as combination of different histone tails from the same nucleosome (Feller et al., 2015).

With the development of analytical capability, there are new types of PTMs discovered (Sidoli et al., 2015). However, the speed to assign corresponding catalyzing enzymes is far behind discovery. The major limitation is the lack of faithful assay methods. For example, Suv4-20 has been reported as the methyl transferase which methylates H4K20 (Schotta et al., 2004). But it is still not clear so far which is the demethylase for H4K20. The same issue also exists for acetylation, phosphorylation, ubiquitination and also other known covalent modifications on histones. *In vitro* enzyme activity assay can be ideal tool to clarify enzyme and substrate relationship.

Regulation and function of PTMs on histones are intensive ongoing studies in epigenetic field. Histone modifiers are associated to cellular processes such as DNA replication, DNA repair, transcription (Carrozza et al., 2005), chromatin condensation (Schotta et al., 2004), cell division and so on. Therefore it is important to assign the corresponding enzymes for histone PTMs substrate. This is especially the case for lysine methylation with most of the

sites already having functional characterization. Both high throughput fast enzyme activity assay methods and low throughput confident assay approaches are required for *in vitro* study.

### 2.1.1 Dynamics of histone lysine methylation

Histone lysine methylations are covalent chemical modifications added and removed by enzymatic reactions (Figure 1.2).

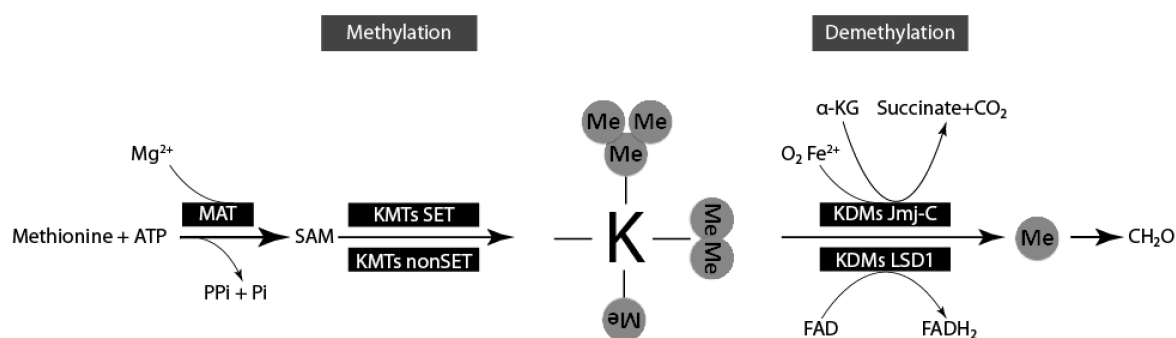
Methylation of histones is catalyzed by histone methyl-transferases (HMTs). There are two types of lysine specific HMTs. One is characterized by SET domain containing proteins such as SetDB1 (Rivera et al., 2015). Another is non-SET domain containing protein such as DOT1 (Ng et al., 2002). Histone methylation can occur on arginine as well. Histone arginine is methylated by protein arginine methyltransferases (PRMTs). Rme1 and Rme2 can be formed. The methyl group is added to arginine symmetrically or asymmetrically for Rme2 (Mosammaparast et al., 2010). All of the methyltransferases require S-adenosyl methionine (SAM) as methyl group donor (Smith & Denu, 2009). SAM is synthesized by methionine adenosyltransferases (MATs). MAT catalyzed the reaction of L-methionine with ATP in the presence of  $Mg^{2+}$ . SAM is generated as product. Pyrophosphate and inorganic phosphate are generated as byproducts (Markham et al., 2008).

SET domain containing HMTs are categorized into two classes: a) mono-methyl specific methyl-transferases; b) di- and tri- methyl specific methyl-transferases. Mono-methyl specific HMTs are represented by KMT7 (Set7/9) catalyzing methylation for H3K4me1. It is specified by releasing the methylated lysine after de-protonation of lysine  $\epsilon$ -carbon-nitrogen bond. Di- and tri- methyl specific HMTs are represented by G9a catalyzing di-methylation and tri-methylation for H3K9. It is specified by not releasing the methylated lysine after de-protonation of lysine  $\epsilon$ -carbon-nitrogen bond (Smith et al., 2009). Most of the histone lysine specific methyl-transferases so far characterized contain SET domain.

KMT4 (DOT1) has been characterized to methylate nucleosomal histone H3 in core globular domain for K79. Its catalytic structure folding is more similar to glycine N-methylase. It harbors an AdoMet binding motif more similar to PRMTs without carrying a typical SET domain (Smith & Denu, 2009; Ng et al., 2002).

Lysine can be unmodified, mono-methylated, di-methylated or tri-methylated. The turnover rate of histone lysine methylation is the same as turnover of histones. For long time being,

lysine methylation was considered as irreversible modification (Byvoet et al., 1972; Duerre et al., 1974; Allis et al., 2001). This view changed with the discovery of lysine specific demethylases. Demethylation of lysine is catalyzed by two types of KDMs: a) the majority of KDMs contain a catalytic Jumonji-domain and require presence of  $O_2$ . They make use of  $\alpha$ -ketoglutarate and  $Fe^{2+}$  as co-factors to perform demethylation. They generate succinate and  $CO_2$  as byproducts. One methyl group is removed each reaction; b) LSD1 belongs to a minor group of KDMs without Jumonji-domain. It makes use of FAD as cofactor and



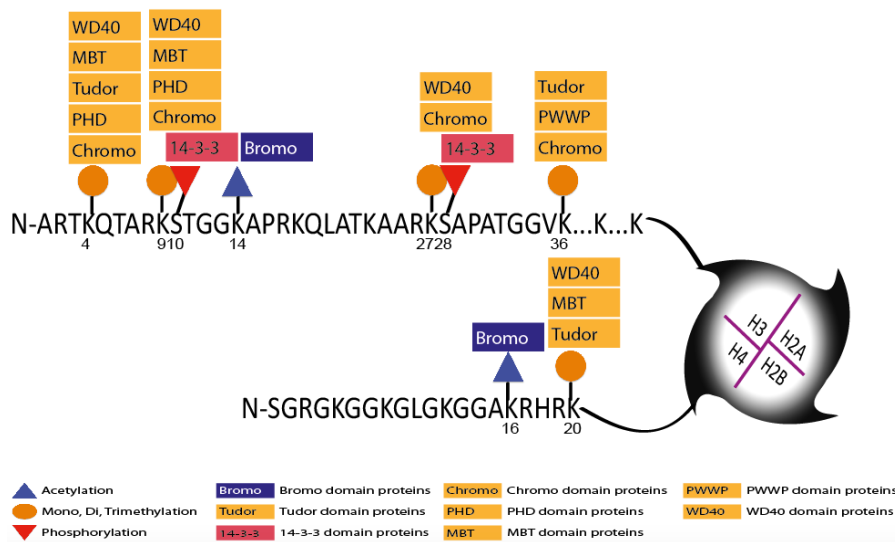
**Figure 1.2 | Lysine methylation is a reversible modification.**

MATs catalyze the reaction of methionine and ATP in the presence of  $Mg^{2+}$  producing SAM. SAM is the general methyl group donor for methylation (Markham et al., 2008). Two categories of KMTs are characterized including the SET domain containing proteins as major group (Rivera et al., 2015) and the minor group DOT1 protein without SET domain (Ng et al., 2002). Histone lysine can be unmodified, mono-methylated, di-methylated or tri-methylated. Demethylation of lysine is catalyzed by two types of KDMs: a) the Jumonji-domain containing proteins; They make use of  $\alpha$ -ketoglutarate,  $Fe^{2+}$  and  $O_2$  as cofactors to perform demethylation and generate succinate and  $CO_2$  as byproducts; b) LSD1 protein which makes use of FAD as cofactor to perform demethylation and generates  $FADH_2$  as byproduct. One methyl group is removed by each demethylation reaction (Mosammaparast et al., 2010).

generates  $FADH_2$  as byproduct. One methyl group is removed by each demethylation reaction as well (Mosammaparast et al., 2010).

Different methylated states on different lysine residues can be read by domain specific histone binding proteins (Figure 1.3). And HMT or KDM catalytic domain can also be brought to the target substrate site by having a binding domain present in the same protein which recognizes proximal PTMs. The Chromo domain from HP1 specifically binds to H3K9me3 and H3K9me2 (Kouzarides et al., 2007). The WD40 repeat domain from EED, one subunit of PRC2 complex, binds to H3K27me3 (Suganuma et al., 2011). The Tudor domain from human KDM4a can bind to H3K4me3 or H4K20me3 (Labbé et al., 2013). The MBT domain repeats from L3MBTL1 recognizes H4K20me2 (Min et al., 2007). The PWWP domain from PSIP1 preferentially binds to H3K36me3 (Pradeepa et al., 2012).

Bromo-PWWP domain from ZMYND11 selectively binds to H3.3K36me3 (Wen et al., 2014). The PHD (Plant Homeo Domain) from C-terminus KDM5a (also called JARID1A) binds to H3K4me3 (Cloos et al., 2008). The lysine adjacent or proximal residue can also be modified. Recruitment of specified domain containing proteins to lysine proximal residues may also play a role in regulating lysine methylation. For instance, Bromo domain from human BRG1 binds to H3K14ac and recruits BAF complex to promoter region (Vicent et al., 2009). Binding of 14-3-3 protein to phosphorylated H3S10 is dependent on the presence of adjacent H3K14ac (Walter et al., 2008).



**Figure 1.3 | Histone PTMs are recognized by specified binding domains.**

Different methylation states can be recognized by different binding proteins. Chromo domain containing proteins bind to tri- or di- methylated histone lysine on H3K4, H3K9, H3K27 and H3K36. MBT domain containing proteins bind to mono- or di- methylated histone lysine on H3K4, H3K9 and H4K20. PHD domain containing proteins bind to tri- or di- methylated histone lysine on H3K4 and H3K9. Tudor domain from human KDM4a can bind to H3K4me3 and H4K20me3. WD40 repeat domain containing proteins bind to repressive trimethylated histone lysine. PWWP domain containing proteins specifically bind to trimethylated lysine. Lysine methylation binding proteins are also regulated by proximal amino acids' modification. 14-3-3 domain containing proteins bind to phosphorylated serine in histone. (Details in Table 2.1).

### 2.1.2 Histone lysine demethylases

Human LSD1, which is a component of the CoREST complex, is the first characterized lysine specific demethylase with experiment evidence. It removes the methyl group from H3K4me2 and H3K4me1 (You et al., 2001). LSD1 homologs are conserved among *S. pombe* (Lan et al., 2007), *C. elegans* (Katz et al., 2009), *Arabidopsis* (Liu et al., 2007), *Drosophila* (Rudolph et al., 2007), *Mus* and *Homo Sapiens* (Shi et al., 2004).

Demethylation by LSD1 involves amine oxidation requiring FAD as cofactor. FAD is reduced to FADH<sub>2</sub> (Shi et al., 2004).

The Jumonji-C domain containing demethylases are dependent on  $\alpha$ -ketoglutarate as cofactor. They belong to a super family of dioxygenases. For these dioxygenases,  $\alpha$ -ketoglutarate oxidative decarboxylation and methyl group hydroxylation are coupled together. Methyl group is released as formaldehyde and meanwhile  $\alpha$ -ketoglutarate is converted to succinate. First report of KDM activity for Jumonji-C domain proteins was based on detection of radioactive labeled formaldehyde release in demethylation reaction by Zhang's Lab. And Jumonji-C domain containing KDMs requires presence of Fe<sup>2+</sup> and O<sub>2</sub> in catalytic center (Tsukada et al., 2006). Comparing to LSD1, which can only demethylate Kme2 and Kme1, the Jumonji-C KDMs are able to demethylate all methylation state Kme3, Kme2 and Kme1 (Mosammaparast & Shi, 2010).

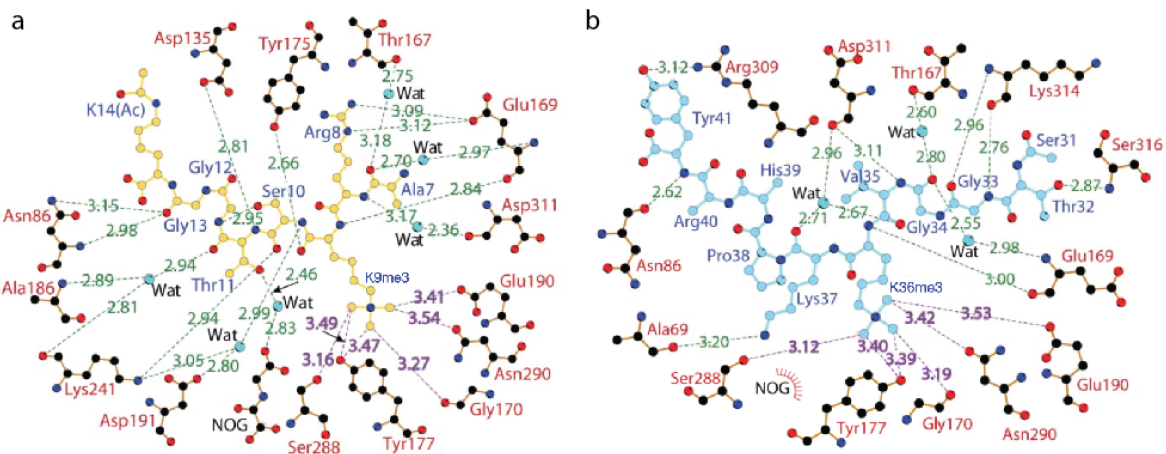
Jumonji-C domain containing KDMs show selectivity towards their substrates. This is achieved by the specific structure of catalytic domain. The conformation is orientated by a) the amino acid residues from catalytic active sites; b) the amino acid residues of substrate; c) the proximal amino acid residues from substrate (Figure 1.4; Ng et al., 2007; Wen et al., 2014).

In order to study the function of these demethylases and clarify substrate specificity for each KDM, it is also crucial to know the complex composition, which these KDMs may form with other components. Knowledge about specific substrate for KDMs can help to predict the role for corresponding KDM.

KDM4a is characterized to demethylate H3K36 and H3K9 from tri-methyl to di-methyl, and from di-methyl to mono-methyl (Allis et al., 2007). H3K36me3 is considered as an active mark, enriched in 3' end of gene body where active transcription is taking place. It is involved in transcriptional elongation by RNA Pol II complex (Bell et al., 2007). H3K36me3 mark is added by SET2. SET2 binds to RNA Pol II complex large subunit Rpb1 CTD region when the serine 2 from heptad repeats is phosphorylated (Eick & Geyer, 2013). The Chromo domain of Eaf3, one component of Rpd3S in yeast, binds to H3K36me3 and recruits Rpd3S complex. Rpd3 complex performs deacetylation to stop intragenic transcription initiation (Carrozza et al., 2005). H3K36me3 is a global mark found in expressed exons' 5' end. It is also involved in alternative splicing (Hon et al., 2009). Determination of alternative splicing site is dependent on transcription elongation by RNA

Pol II recruitment. RNA Pol II binds to SET2 to set H3K36me3 mark. H3K36me3 is closely associated with tissue and cell type specific alternative splicing. Dysfunction of SET2 has shown strong effect in selecting alternative splicing sites (Luco et al., 2010). However up to date, it is not clear how the methylation on H3K36 is reversed after RNA Pol II elongation. Is there demethylase present together with SET2 constantly to switch off this mark? Or can the demethylase be recruited to defined loci by yet unknown mechanisms?

H3K9me3 is thought to be a repressive mark localized in heterochromatin region including pericentromere and telomere. Chromo domain from HP1 recognizes and binds to H3K9me3 and H3K9me2. In addition, HP1 recruits more SUV39H to the region and further enforces H3K9 methylation. Therefore heterochromatin region is spread and maintained (Maison & Almouzni, 2004). H3K9me3 is also found in promoter region and gene body of transcriptionally repressed genes (Sims & Reinberg, 2009). This also supports the hypothesis that H3K9me3 is a repressive mark. In agreement with that H3K4me3 and H3K9ac are active marks, methylation of H3K9 is associated with H3K4 hypomethylation and H3K9 hypoacetylation during X chromosome inactivation (Heard et al., 2001). KDM4a may reverse the functions of H3K36me3 and H3K9me3 by changing the methylation state in these two sites.



**Figure 1.4 | Binding of peptide substrate to hKDM4a catalytic domain**

- a Peptide H3.3\_7.14 with H3K9me3 binding to catalytic pocket formed by JmjN and JmjC;
  - b Peptide H3.3\_31.41 with H3K36me3 binding to catalytic pocket formed by JmjN and JmjC.
- (Figure adapted from Ng et al., 2007)

## 2.2 Previous work on KDM4a

### 2.2.1 Enzymatic related KDM4a function

Human KDM4a (EC 1.14.11) selectively catalyzes demethylation of H3K36me3/me2 to H3K36me2/me1 and H3K9me3/me2 to H3K9me2/me1, but not H3K36me1 to H3K36me0 and H3K9me1 to H3K9me0 (Couture et al., 2007). And hKDM4a prefers H3K9 to H3K36 as substrate (Black et al., 2012). Full length hKDM4a has 1064 amino acids. It includes catalytic region Jmj-N and Jmj-C domain, linker region, one PHD domain, one Zinc finger domain and two Tudor domains (Figure 1.5).

Crystal structure of hKDM4a catalytic region 1-359 aa in presence of  $\alpha$ -ketoglutarate analog (Ng et al., 2007) has identified residue His188, Glu190 and His276 as iron binding sites; Tyr 132, Asn198 and Lys206 are  $\alpha$ -ketoglutarate binding sites. Ng et al. have also co-crystallized hKDM4a with peptide including H3K9me3 (H3.3\_7.14: peptide of histone H3 Ala7 to Lys14) and peptide including H3K36me3 (H3.3\_31.41: peptide of histone H3 Ser31 to Tyr41) (Figure 1.4). The H3K9me3 peptide binds to substrate pocket in a W-shape conformation with downstream residues of H3 Ser10\_Thr11\_Gly12\_Gly13 also interact with the pocket. The H3K36me3 peptide binds to substrate pocket in a U-shape conformation. In this paper, they further showed the evidence that phosphorylation of H3 at position Ser10 abolished KDM4a demethylation activity at Lys9.

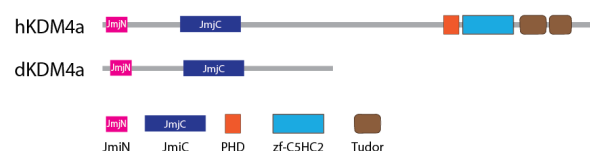
In 2013, Black et al. demonstrated that over-expression of hKDM4a gives rise to extra DNA copy of 1q12, 1q21 and Xq13.1 locus based on cytogenetic band analysis. While the global chromosome stability is not disturbed. Active demethylase activity and intact Tudor domain are required to generate these copy gains. The same regions are observed to gain extra copies in KDM4a up-regulated tumors. Regions with extra copies have re-replication and occupancy of DNA polymerase. MCM and KDM4a are also increased in these regions. In contrast, occupancy of HP1 $\gamma$  is decreased. However, over-expression of Suv39h1 or HP1 $\gamma$  antagonizes these copy gains. Hypothetical mechanisms proposed for the origin of certain genome region copy number change include: a) stalled replication forks during DNA amplification. This generates tandem duplication; b) DNA breaks and repair intermediates are recognized as primers. This initiates re-replication and leads to duplication of a genome region or deletion; c) DNA polymerases collide between head and tail during elongation. The copy gain caused by KDM4a can accumulate within 24hrs during S phase. This is dependent on the demethylase activity of it (Black et al., 2013). Instable chromosome and altered copy number of certain genome region associated with disease have also been observed in Sotos syndrome, breast cancer and ovarian cancer

(Berdasco et al., 2009; Kelly et al., 2010). But it is not clear what the driving force to promote site-specific copy number alteration is.

Apart from demethylase activity towards H3K36me3/me2 and H3K9me3/me2, hKDM4a has been found binding to methylated H4K20 via its Tudor domain. This binding occurs before recruitment of 53BP1 protein to DNA damaged site in double-strand break response (Panier et al., 2014). Meanwhile, hKDM4a has been found to catalyze demethylation of linker histone H1.4K26 (Trojer et al., 2009).

The major KDM enzymes in fruit fly are highly conserved from human (Lloret et al., 2008). In *D. melanogaster*, most of them have single gene encoding each enzyme. In comparison, multiple genes and multiple enzymes are involved in single lysine residue demethylation catalysis in human. In fact, fruit fly and human share highly conserved signaling pathways, regulation of cell proliferation and molecular functions of homologs. Among all known human disease related genes, one third of them have counterparts in *D. melanogaster* (Bier et al., 2005). All of these conservation and simplicity features make fruit fly a promising model organism to study the regulation of human disease relevant genes.

Core catalytic domains of hKDM4a and dKDM4a homologs share 63 % identity and 80 % similarity (Figure 1.5). dKDM4a selectively catalyzes demethylation of H3K36me3/me2 to H3K36me2/me1 both *in vitro* and *in vivo*, but not H3K36me1 to H3K36me0. dKDM4a has shown preference to methylated H3K36 as a substrate in comparison to methylated H3K9. This is based on the observation that over-expression of dKDM4a leads to reduction of H3K36me3 signal while the H3K9me3 signal has no variation. Knock down of dKDM4a has shown dramatic increase for H3K36me3 but only slightly increase for H3K9me3. However *in vitro* demethylation assay with recombinant dKDM4b has shown stronger demethylation of H3K9me3 instead of dKDM4a (Lin et al., 2008).



**Figure 1.5 | Conserved core catalytic domain of KDM4a in *H. sapiens* and *D. melanogaster***

hKDM4a core catalytic domain region 1-359 aa and dKDM4a 1-372 aa share 63 % identity and 80 % similarity. See alignment of protein sequence in appendix 2 (Figure 6.2).



dKDM4a is observed to have RNA peak expression within embryonic stage 0 to 6 hours according to modENCODE Temporal Expression Profile. dKDM4a mutant flies showed shorter life span for male. Phenotype of showing extension-twitching wing and courtship song inter male ones are observed in these mutant flies (Lorbeck et al., 2009).

### 2.2.2 Interacting factors of KDM4a, H3K36me3 and H3K9me3

It is known about the demethylation enzymatic activity of hKDM4a towards H3K36me3 and H3K9me3. But the biological function of KDM4a is more dependent on the macromolecule complexes formed by KDM4a and its interactors. It is also dependent on the complexes formed by H3K36me3 or H3K9me3 together with its individual interactors before KDM4a demethylation activity occurs. Recent characterized interaction partners for KDM4a, H3K36me3 and H3K9me3 are listed in Table 2.1.

**Table 2.1| Interaction partners of KDM4a, H3K36me3 and H3K9me3**

	Interactor name	Species	Binding information	Reference
KDM4a	CG16972	<i>D. mel</i>	Co-IP	Rhee et al., 2014
	CG34163	<i>D. mel</i>	Co-IP	Rhee et al., 2014
	CG10911	<i>D. mel</i>	Two-hybrid	Giot et al., 2003
	EcR	<i>D. mel</i>	Co-IP,WB	Tsurumi et al., 2013
	HP1	<i>D. mel</i>	Co-IP	Alekseyenko et al., 2014
	Ino80	<i>D. mel</i>	Co-IP	Lowe et al., 2014
	Df31	<i>D. mel</i>	Co-IP	Guruharsha et al., 2011; Guruharsha et al., 2012
	miR-1005	<i>D. mel</i>	RRI	Schnall-Levin et al., 2010
	miR-1014	<i>D. mel</i>	RRI	
	miR-14	<i>D. mel</i>	RRI	
	miR-275	<i>D. mel</i>	RRI	
H3K36me3	MSL3	<i>D. mel</i>	Chromo domain	Larschan et al., 2007; Sural et al., 2008
	Eaf3	<i>S. cerevisiae</i>	Chromo domain	Carrozza et al., 2005
	PHF19	<i>Mus</i>	Tudor domain	Brien et al., 2012
	PSIP1	<i>Mus</i>	PWWP domain	Pradeepa et al., 2012
	Brpf1	<i>Mus</i>	PWWP domain	Vezzoli et al., 2010
	Dnmt3a	<i>Mus</i>	PWWP domain	Dhayalan et al., 2010
	MSH-6	<i>H. sapiens</i>	PWWP domain	Vermeulen et al., 2010
	NSD1	<i>H. sapiens</i>	PWWP domain	
	NSD2	<i>H. sapiens</i>	PWWP domain	
	N-PAC	<i>H. sapiens</i>	PWWP domain	
	MRG15	<i>H. sapiens</i>	Chromo domain	Zhang et al., 2006; Luco et al., 2010
	HIRA	<i>Mus</i>	Chaperon	Lin et al., 2014
	HiRA	<i>D. mel</i>	Chaperon	Loppin et al., 2001;

				Loppin et al., 2005
	ZMYND11	<i>Mus</i>	Bromo-PWWP domain	Wen et al., 2014
	HP1	<i>D. mel</i>	Chromo domain	Schotta et al., 2002; Brehm et al., 2004
H3K9me3	Cbx5	<i>H. sapiens</i>	Chromo domain	Vermeulen et al., 2010
	Cbx3	<i>H. sapiens</i>	Chromo domain	
	Cbx1	<i>H. sapiens</i>	Chromo domain	
	CDYL2	<i>H. sapiens</i>	Chromo domain	
	CDYL	<i>H. sapiens</i>	Chromo domain	

Co-IP: Co-immune precipitation. RRI: RNA-RNA interaction. WB: western blotting.

Despite of demethylase enzymatic activity, some of dKDM4a regulated genes do not have H3K36me3 mark. This has raised up the question that dKDM4a may play functions unrelated to demethylase activity (Crona et al., 2013). KDM4a protein contains a PxVxL motif and binds to HP1. HP1 recognizes H3K9me3 (Schotta et al., 2002). PxVxL motif is found in also other HP1 binding factors (Thiru et al., 2004).

Although the protein interactors of KDM4a, H3K36me3 and H3K9me3 can give a hint about the biological functions that KDM4a plays, regulation of KDM4a and especially function beyond demethylation enzymatic activity still need to be addressed.

### 2.3 Therapeutic strategy targeting epigenetic modifiers

Epigenetic modifiers program the covalent modifications on histones in a dynamic manner. Different modification marks and replacement of histone variants have been associated with diseases and cancer (Vardabasso et al., 2014). For instance, methylation, acetylation and ubiquitination have been reported to regulate transcription events. Phosphorylation is mainly involved in regulating cell cycle (Bhaumik et al., 2007). Dysfunction of epigenetic modifiers is frequently observed in disease progression of patients (Table 2.2). Plasticity of histone modifications programmed by epigenetic modifiers has triggered the inspiration to target them as therapeutic strategy for related diseases.

#### 2.3.1 Disease related KDM4 epigenetics

Dysfunction of Jmj-C containing demethylases is commonly involved in human diseases and cancers due to the critical functions they play in DNA replication, encoding gene transcription, cell programming, differentiation and development (Cloos et al., 2008). *KDM4* subfamilies of the Jmj-C domain containing encoding genes include *KDM4a*, *KDM4b*, *KDM4c*, *KDM4d*, *KDM4e* and *KDM4f* in human. Meta-analysis of gene RNA

expression levels of *KDM4* members among human normal tissues demonstrates the regulation of *KDM4* members may follow different pathways (Labbé et al., 2013). *KDM4c* has been found amplified in cerebellar tumors accompanied by amplification of several genome regions including 1p22 (Ehrbrecht et al., 2006). *KDM4c* has also been found up-regulated in oesophageal squamous carcinomas. Knockdown of *KDM4c* by RNAi delayed cell proliferation (Cloos et al., 2006). Interaction between *KDM4c* and androgen receptor (AR) allows demethylation of H3K9me3 by *KDM4c* on AR targeted genes in an androgen dependent manner (Wissmann et al., 2007).

Over-expression of *KDM4a* has been reported in lung cancer, breast cancer and ovarian cancer (Labbé et al., 2013). Up regulation of *KDM4a* in ovarian cancer patients based on RNA-Seq data is significantly correlated with shorter mean time to death. For *KDM4a* over-expression patients, it is 23 months, comparing with 35 months for the patients without *KDM4a* changes (Berry et al., 2012; Black et al., 2013).

**Table 2.2 | Disease associated epigenetic dysfunction**

Epigenetic aberration	Responsible enzyme	Disease	Epigenetic alteration	Comments	Reference
DNA methylation	DNMT1, DNMT3A, DNMT3B and DNMT3L	Rett syndrome	Inability to 'read' DNA methylation	MECP2 mutation	Egger et al., 2004; Urduingio et al., 2009; Feng et al., 2009
		Diabetes	Hypermethylation of <i>PPARGC1A</i> promoter		Villeneuve et al., 2010
		Cancer	Global hypomethylation, hypermethylation of some CpG island promoters, including CIMP		Jones et al., 2007; Fouse et al., 2009; Egger et al., 2004
		Systemic lupus erythematosus	Hypomethylation of CpG islands at specific promoter regions	Decreased DNMT1 and 9 DNMT3B expression	Javierre et al., 2010
		ICF syndrome	Hypomethylation at specific sites	DNMT3B mutation	Egger et al., 2004; Urduingio et al., 2009; Feng et al., 2009
		ATR-X syndrome	Hypomethylation of specific repeat and satellite sequences	ATRX mutation	Egger et al., 2004; Urduingio et al., 2009
Histone acetylation	HATs and HDACs	Rubinstein-Taybi syndrome	Hypoacetylation	Mutation in gene encoding CBP, a known HAT	Egger et al., 2004; Urduingio et al., 2009; Feng et al., 2009
		Diabetes	Hyperacetylation at promoters of inflammatory genes		Villeneuve et al., 2010
		Asthma	Hyperacetylation	Increased HAT	Adcock et al.,

				activity and decreased HDAC activity	2005
		Cancer	H4K16 acetylation loss	Hypomethylation of DNA repetitive sequences	Jones et al., 2007
<b>Histone methylation</b>	HMTs and HDMs	Cancer	H4K20me3 loss	Hypomethylation of DNA repetitive sequences	Jones et al., 2007
		Sotos syndrome	Decreased H4K20me3 and H3K36me3	Loss of function of NSD1, a HMT	Berdasco et al., 2009
		Huntington's disease	Increased H3K9me3 and possibly increased H3K27me3	Increased expression of the HMT ESET; enhanced PRC2 activity	Urduingio et al., 2009
<b>miRNA expression</b>	N/A	Cancer	Decreased miR-101	Increased EZH2, H3K27me3	Varambally et al., 2008; Friedman et al., 2009
			Decreased miR-143	Increased DNMT3A	Ng et al., 2009
			Decreased miR-29	Increased DNMT3A and DNMT3B	Fabbri et al., 2007
			Increased miR-21	Decreased PTEN	Calin et al., 2006
			Increased miR-155	Lower survival rates	Yanaihara et al., 2006

Table adapted from Kelly et al., 2010. Comments describe the possible reasons for corresponding diseases.

### 2.3.2 Epigenetic therapies

Many of disease pathologies are derived from genomic instability caused by epigenetic aberration. The reversible feature of epigenetic modifications makes them ideal targets for pharmacology (Wang et al., 2014). The switch of epigenetic marks often leads to abnormal transcriptional profile and protein expression profile. Therefore, inhibitors of epigenetic modifiers have a high therapeutic potential. Currently, there are several small molecule drugs approved by American Food and Drug Administration (FDA) for modulating epigenetic marks. They have already proceeded to different phases of clinical trials. For instance, hydralazine is used as DNA demethylating agent in clinical trials. SAHA is a HDAC inhibitor approved by FDA to target histone acetylation as a paradigm (Kelly et al., 2010).

A loss or gain of enzymatic function of KDM4a leads to deregulation of H3K36me3 and H3K9me3. This can directly induce oncogene activation and tumor suppressor gene repression. So far 3 types of KDM4 family inhibitors have been reported. They are either mimics for the two cofactors:  $\alpha$ -ketoglutarate or bivalent iron; or mimics for substrates:

H3K36me3 or H3K9me3 (Labbé et al., 2013). High perspective for tumor and cancer therapy is held by them. But up to date, FDA has not approved any histone methylases and demethylases targeting drugs.

## 2.4 Application of Mass Spectrometry

Mass spectrometers separate gas-phase ions according to their mass-to-charge ratio and detect them in proportion to their abundance. These gas-phase ions can also undergo a controlled fragmentation within the mass spectrometer. The resulting fragment ions together with the fragmented molecule (parent ion) can be then used to identify the molecule of interest (Hoffmann & Stroobant, 2007).

In the case of peptides and proteins, the coupling of separation techniques like liquid chromatography (LC) or capillary electrophoresis (CE) (Busnel et al., 2010) prior to the mass spectrometer allows on-line analysis of the separated molecules. During last two decades, continuous developments in sample preparation (Rappsilber et al., 2007), chromatography, ionization, mass analyzers and detectors have led to the possibility to identify and quantify thousands of proteins in one single analysis (Schubert et al., 2012; Sansoni et al., 2014). Currently, several research fields like drug discovery, clinical analysis, food quality control, metabolomics (Zimmermann et al., 2014) and specially proteomics are taking advantage of technological improvements in the field of mass spectrometry.

### 2.4.1 Mass Spectrometry based proteomics

Systems biology requires not only information from the genome or the transcriptome, but also from proteins in order to understand the biological function of the studied system (Tyers & Mann, 2003). Mass spectrometry is the tool of choice because of a) protein sequencing accuracy based on accurate molecular mass information generated by mass spectrometry (Kinter & Sherman, 2000); b) capability of analyzing low abundant proteins from complex biology samples; c) high throughput measurement and deep sequencing. Two different approaches are currently used to identify and quantify proteome: shotgun and targeted approaches. Shotgun approach is more suitable to discovery driven purpose and targeted proteomics fits to more quantitation application for specified signal (Picotti et al., 2013).

With the help of sequenced whole genome information, *in silico* translation has been performed to predict theoretical proteome database for a given species (Krug et al., 2011;

Tyers & Mann, 2003). Combined with chemical derivatation approach and enzymatic digestion (Bonaldi et al., 2004) as well as database search algorithms (Cox et al., 2011), peptides are mapped to proteins. High resolution and accuracy achieved from different mass spectrometer renders the capability to identify isotopic patterns and complicate isobaric structure. In addition, shotgun proteomics makes use of affinity purification such as immunoprecipitation as enrichment strategy for bait. Identification of protein interactome is achieved based on spectrum library *in silico*. Targeted proteomics with selected ion monitoring (SIM) strategy is applied to measure specified clusters of peptides with complex mixtures. But up to date, only few species such as yeast (Picotti et al., 2013), have been measured with full coverage of proteome using mass spectrometry based proteomics study. And this is mainly due to shortage of analytical capability.

#### **2.4.2 Analysis of PTM's by Mass Spectrometry**

PTMs on histones are chemical modifications added to histone proteins after they are translated. Most of the modifications are edited in the nucleus (Olsen & Mann, 2013). But some are also placed immediately after translation in the cytosol (Rivera et al., 2015). Histone PTMs mainly refer to methylation, acetylation, phosphorylation, ubiquitination as mentioned in Figure 1.1. Histone PTMs do not only function as single mark but also in combination with an extensive crosstalk between each other (Kouzarides, 2007). This leads to recruitment of variable binding proteins which form different macromolecular complexes responding to different signaling pathways (Suganuma et al., 2011). The complexity and flexibility posed by histone PTM identification and quantitation analysis makes mass spectrometry the first strategy to use to systematically study global histone PTM profile upon stimuli (Feller et al., 2015). Mass spectrometry allows a high throughput and high speed analysis of histone PTMs and is also the ideal tool for de novo PTM discovery (Dai et al., 2014).

#### **2.5 Metabolomics and epigenetics**

Recent studies have illustrated the response of reactive oxygen species and mitochondrial integrity (Møller et al., 2010). They regulate chromatin landscape via kinases, methylases and acetylases. They play important role in aging process (Salminen et al., 2014; Chin et al., 2014). Emerging evidence of dysfunction for metabolic enzymes in cytosol have shown to shape the epigenetic profile in cell nucleus. For example, homocitrate synthase catalyzes reaction of acetyl-CoA and  $\alpha$ -ketoglutarate. It has been shown to be a bifunctional enzyme

involved in lysine biosynthesis in cytoplasm. It is also able to acetylate histone in cell nucleus in yeast (Scott et al., 2010). Small molecular metabolite like  $\alpha$ -ketoglutarate ( $\alpha$ -KG) has been reported to increase the longevity in *C. elegans*. It reduces ATP synthase activity and inhibits TOR pathway.  $\alpha$ -KG binds to the ATP synthase subunit  $\beta$  (Chin et al., 2014). Emerging data has shown that epigenetic marks serve as sensors and translators of metabolism to change gene expression profile. This is based on the plasticity feature of histone PTMs and DNA methylation (Katada et al., 2012). On the other hand, even different human populations with different epigenome can be traced back by metabolite profiles (Menni et al., 2013). All the evidence indicates that metabolism can be the driving force to initiate variation on epigenetic mark and further on bring phenotype change.

### 2.5.1 Influence of metabolites on epigenetic marks

Metabolism, consisted of catabolism and anabolism, is an important feature of organism to express its life activity. Most of metabolic reactions are dependent on enzyme catalysis. Major functions of metabolism are to a) acquire nutrients, b) convert nutrients to building blocks for the organism, c) assemble the building blocks to macromolecules such as DNA, protein and lipid. Meanwhile all the energy required for life activity and special functional molecules are also from metabolism process. The tricarboxylic acid cycle (TCA) plays a central role in metabolism. It generates metabolic intermediates. One of the key metabolites is  $\alpha$ -ketoglutarate. It serves as cofactor for dioxygenases such as TET domain containing DNA demethylases, Jumonji-C domain containing histone demethylases and hypoxia inducible factor (Cairns et al., 2013). Another key metabolite acetyl-CoA is the donor for histone acetyl-transferases (Salminen et al., 2014). SAM is a small metabolite produced by methionine adenosine transferase (MAT). Variance of SAM regulates H3K9 methylation (Kato et al., 2011) and H3K4me3 (Mentch et al., 2015).

### 2.5.2 Regulation of metabolic enzyme

Regulation of metabolic enzyme is done by different signaling pathway and PTMs on those enzymes (Lee et al., 2014). Metabolic enzymes are the sensors for nutrients. Epigenetic modifiers are the sensors for metabolites. Metabolic enzymes and histones may share the same PTM's modifiers. Dysfunctions of metabolic enzymes are described in many diseases (MacDonald et al., 2009).

### 2.5.3 Characterization of IDH

Isocitrate dehydrogenase (IDH) is a metabolic enzyme. It converts isocitrate to  $\alpha$ -ketoglutarate via oxidative decarboxylation reaction. There are two categories of IDH, one is NAD dependent (IDH-NAD, EC 1.1.1.41). IDH-NAD corresponds to the human homolog IDH3. It is localized mainly in mitochondrial. The other one is NADP dependent (IDH-NADP EC 1.1.1.42). It is localized both in cytosol and mitochondrial. IDH-NADP corresponds to human homolog IDH1 in cytoplasm and IDH2 in mitochondrial (Labussiere et al., 2010). The reaction from isocitrate to  $\alpha$ -ketoglutarate is one of the rate-limiting steps in TCA cycle and is irreversible.

Deregulation of IDH directly affects the mitochondrial electron transport chain, mitochondrial respiration, oxygen consumption, carbon flux in the TCA cycle (Sweetlove et al., 2010; Williams et al., 2010) and hormone metabolism.

Naturally occurring single amino acid point mutation of IDH1 R132H and IDH2 R172H confers the enzymes new catalysis function. Instead of producing  $\alpha$ -KG, 2-hydroxyglutarate (2-HG) is formed as product (Rohle et al., 2013). 2-HG has similar chemical structure to  $\alpha$ -KG. It is a competitive inhibitor for enzymes using  $\alpha$ -KG as cofactor (Xu et al., 2011). IDH1 and IDH2 mutations are found in >70 % of the patients with lower grade gliomas. IDH1 R132H mutation accounts for >95 % of all mutations for IDH1. Cells with IDH1 R132H mutation have increased 2-HG concentration. IDH1 R132H mutation is closely associated with global DNA hyper methylation (Turcan et al., 2012). IDH1 R132H mutation is considered as a prognostic mark for gliomas. Inhibitors targeting the mutant IDH1 have already been reported. Demethylation of H3K9me3 has been observed upon inhibitor treatment for IDH1 R132H (Rohle et al., 2013). Delimited IDH enzymatic activity reduces the  $\alpha$ -KG level and increases HIF- $\alpha$  level. HIF- $\alpha$  is a transcriptional factor promoting the tumor cell growth with low oxygen content (Zhao et al., 2009).



## 2.6 Aim of the work

Genetic information is highly organized in chromatin. The basic unit of chromatin is nucleosome. Nucleosome is formed by DNA wrapping around histone octamer. Histones are small basic proteins. Histones contain many lysine and arginine residues. Therefore histones can be heavily decorated by covalent chemical modifications.

In our lab, we are interested in histone post translational modifications. Histone PTMs may constitute a histone code. Inheritance of epigenetic information is based on the conservation of histone PTMs during cell division. The aim of this thesis is to characterize histone lysine demethylation. I focused my work on the following 4 points.

### **Part 1: KDM4a inhibitor screening**

The intrinsic dynamic and plasticity feature of epigenetic marks is acquired by functions of epigenetic modifiers. They are closely associated with diseases and cancer. This highlights the demand to develop small molecular drugs to target epigenetic modifiers. Up regulation of KDM4a has been found in more than 40 % of ovarian cancer patients. It is closely related to shorter survival time (Black et al., 2013). Up to date, there are no bona fide KDM4a inhibitors to target this enzyme as therapeutic strategy.

### **Part 2: Functional analysis of dKDM4a**

*Drosophila melanogaster* has less encoding gene redundancy compared to humans. It is a great model organism to study human disease related gene regulation. In this thesis, MS based proteomics experiment has been performed to study the protein interactome of dKDM4a. Investigation of complexes formed together with dKDM4a is performed in order to clarify the biological functions of dKDM4a.

### **Part 3: Characterizing dynamic of H3K36me3 at single cell level**

H3K36me3 mark reflects the defined transcriptional profile at a given physiological condition. It is dependent on the cell type and external stimuli. Bulk analysis of H3K36me3 averages the signal to heterogenous cell population and physiological status. It is not clear about the single cell variation for H3K36me3 mark.

### **Part 4: Whether manipulation of metabolic enzyme IDH brings effect for epigenetic marks in nuclear?**

KDM4a's cofactor  $\alpha$ -ketoglutarate is generated outside the nucleus by the metabolic enzyme isocitrate dehydrogenase (IDH). It is the cofactor for dKDM4a in nuclear to catalyzing demethylation. Over-expression of IDH has been performed in order to study its direct effect to dKDM4a' activity *in vivo*.

## 3. Results

### 3.1 KDM4a inhibitor screening

#### 3.1.1 Selection of small molecular fragments

Alteration of epigenetic marks often results in chromatin architectural changes and genomic instability. Chromatin architecture is regulated by post translational modifications on DNA and histones. The intrinsic dynamic and plasticity feature of epigenetic marks was regulated by functions of epigenetic modifiers. Close association between epigenetics with diseases and cancer highlights the demand to develop small molecular drugs to target epigenetic modifiers. For instance, up regulation of KDM4a has been found in more than 40 % of ovarian cancer patients. And this is closely related to shorter survival time (Black et al., 2013). In this thesis work, we investigated to establish different methods to study the putative inhibitors against KDM4a *in vitro* and *in vivo*.

The starting point of this study is the collaboration of our lab with Chroma THERAPEUTICS in order to discover and verify small molecular drugs of targeted cancer therapy. They have performed a high throughput screening of 20000 fragments library. The library is designed against the crystal structure of recombinant human KDM4a protein catalytic domain residue 1-359 aa. High throughput screening was carried out based on formaldehyde dehydrogenase coupled demethylation assay detection. In this detection method, demethylation was detected by releasing of formaldehyde in a coupled reaction. 420 hits were obtained with  $< 1 \mu\text{M}$  inhibition potency for  $1 \mu\text{M}$  hKDM4a enzyme. 150 fragments were selected to determine IC<sub>50</sub> by LC-MS detection *in vitro*. LC-MS measurement has been applied to detect product formation from demethylation directly. 65 fragments out of 150 investigated have shown  $1 \mu\text{M} < \text{IC}_{50} < 510 \mu\text{M}$  for  $1 \mu\text{M}$  hKDM4a. 30 fragment hits out of 65 with low micromolar IC<sub>50</sub> values were subjected to co-crystallization studies with hKDM4a. These 30 fragments are categorized in 3 structural classes. 130 putative crystals were prepared for X-ray analyze. Co-crystals were not formed despite the fact that those selected hits did inhibit hKDM4a activity *in vitro* with low IC<sub>50</sub> values (Figure 3.1 a).

6 hits from the 30 investigated fragments in co-crystallization studies were selected by us to perform further analysis. They have sub-micromolar IC<sub>50</sub> no more than  $10 \mu\text{M}$ . Chemical structures of the 6 compounds are depicted in Figure 1b. JMJ-1, JMJ-2, JMJ-3, JMJ-4, JMJ-5 and JMJ-6 are used to refer to corresponding structures. IC<sub>50</sub> values in Figure 3.1 b were

determined by *in vitro* demethylation assay using LC-MS analysis. It was performed with recombinant protein from hKDM4a residue 1 to 359 aa covering the core demethylase catalytic domain Jmj-N and Jmj-C. Recombinant hKDM4a protein was fused to N-terminus His tag and expressed in bacteria system. It was purified by Nickel column via 6×His tag affinity binding, size-exclusion chromatography and polished over an anion exchange column.

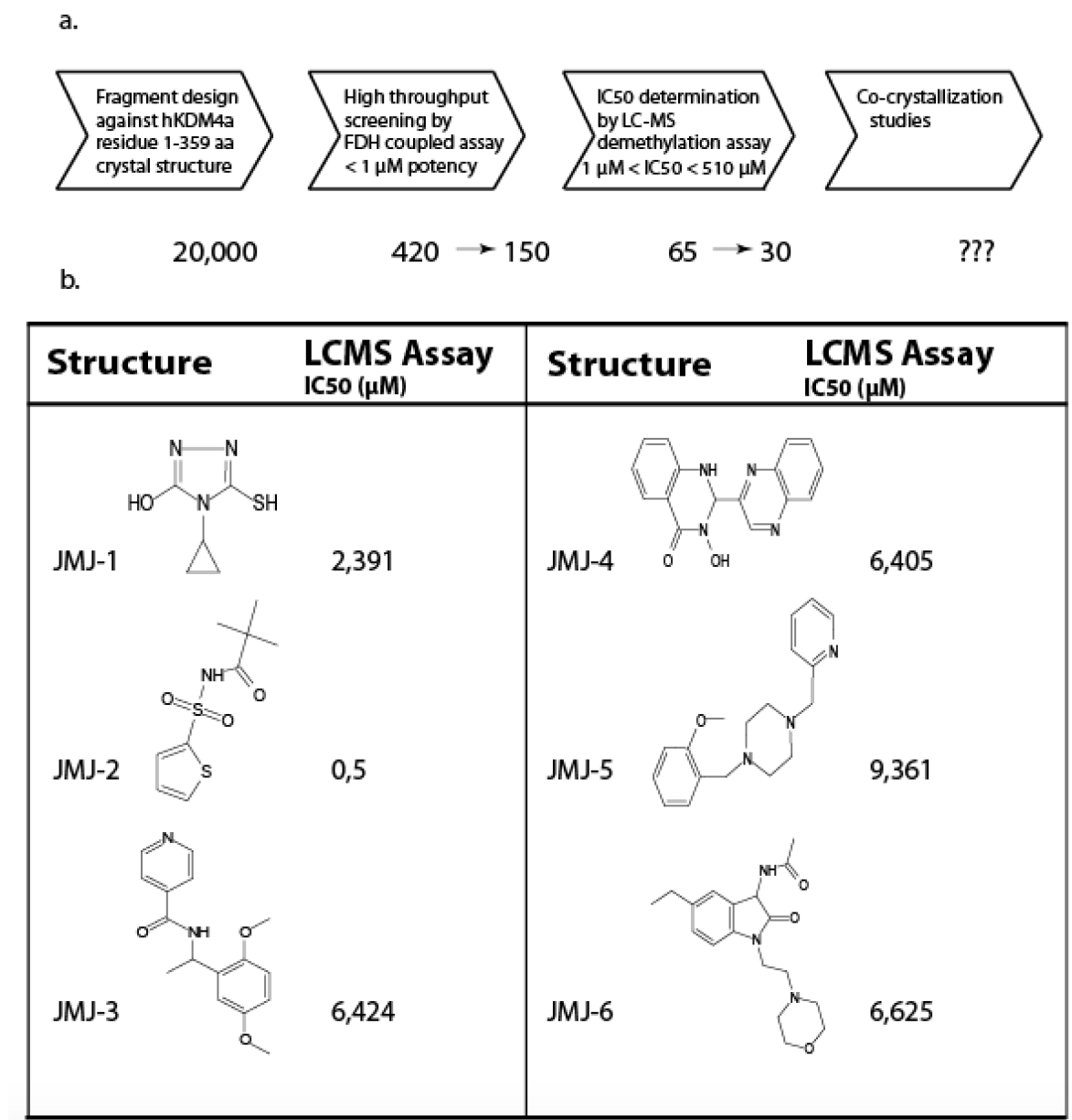
This study focused on investigating the inhibitory activities of selected 6 compounds (Figure 3.1 b) with *Drosophila* KDM4a enzyme *in vitro* and *in vivo*.

### 3.1.2 Potential targets of selected fragments: Jumonji domain containing proteins in human and fruit fly

In order to clarify the potential targets for selected small molecular compounds, *in silico* identification of proteins containing Jumonji domain specific fold were obtained from Conserved Domain Architecture Retrieval Tool (CDART). Filtering has been performed with [cl18224] for Jmj-C superfamily domain and [cl15840] for Jmj-N domain. The taxonomy of INCLSPAN [9605] for *Homo Sapiens* and INCLSPAN [7227] for *Drosophila Melanogaster* were used individually. 259 proteins contain Jmj-C domain and 72 proteins contain Jmj-N domain from human. 20 proteins contain Jmj-C domain and 10 proteins contain Jmj-N domain from fruit fly (see Figure 3.2 a).

20 proteins were identified by retrieving Jmj-C [cl18224] INCLSPAN [7227] from *Drosophila melanogaster* (database 2014 May). Schematic illustration of these 20 proteins is indicated in Figure 3.2 b. 14 proteins were reported from literature which are indicated in bold name in the figure. The remaining 6 proteins were theoretically predicted. All 20 proteins with Jmj-C domain falls in 7 clusters: KDM4 subfamily, UTX subfamily, KDM2 subfamily, Jarid2 subfamily, KDM3 subfamily, PSR subfamily and Lid subfamily. KDM4 subfamily contains both Jmj-N domain [cl15840] and Jmj-C domain [cl18224] in proximity. Apart from these two characterized conserved domain, no other conserved domains have been found in KDM4 subfamily.

There are 6 proteins in the KDM4 subfamily in *Drosophila melanogaster*. It consists of isoform KDM4a transcriptional variant a, isoform KDM4a transcriptional variant b, isoform KDM4b transcriptional variant a, isoform KDM4b transcriptional variant b and other two predicted proteins MIP04757p and AT26080p1. However, there is no experimental evidence for the expression of these predicted isoforms.



**Figure 3.1 | Screening potent hKDM4a inhibitors**

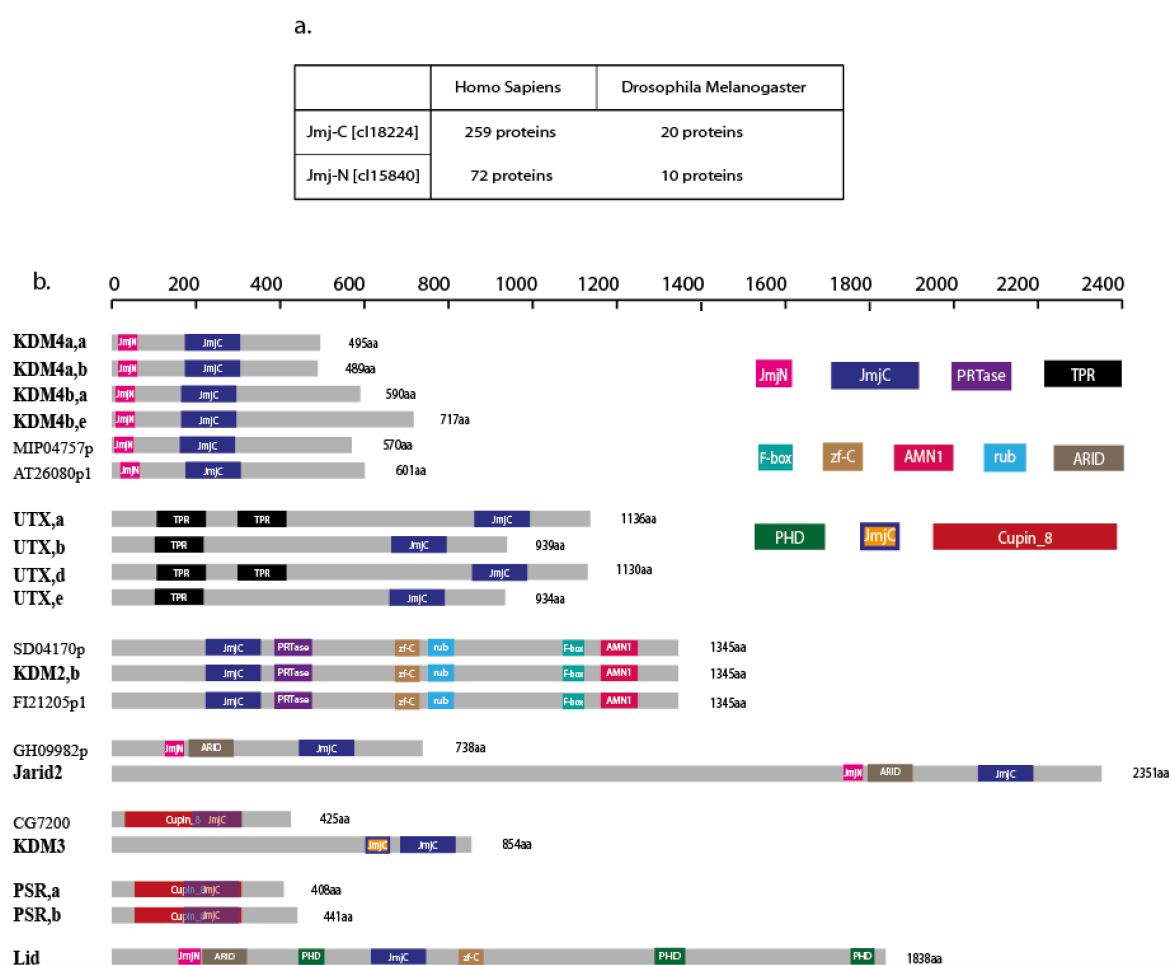
a 20000 fragments against human KDM4a catalytic domain crystal structure obtained with residue 1-359 aa has been selected. High throughput screening has been performed for inhibition assay of demethylation by formaldehyde release detection. 420 fragments shown inhibition activities with  $< 1 \mu\text{M}$  concentration. 150 hits were selected to determine IC<sub>50</sub> by LC-MS detection based demethylation assay. 65 fragments shown  $1 \mu\text{M} < \text{IC}_{50} < 510 \mu\text{M}$ . 30 hits were selected to perform co-crystallization studies with recombinant hKDM4a core catalytic domain. No co-crystals were obtained.

b Chemical structures and IC<sub>50</sub> are indicated for 6 hits selected for analysis with *Drosophila* KDM4a protein.

### 3.1.3 Establish demethylase enzyme activity assay *in vitro*

Recombinant full length *Drosophila* KDM4a protein was fused to N-terminus Strep tag. It was expressed in insect cell Sf21 using bacmid DNA via baculo virus transfection.

Isolation of dKDM4a protein was performed by Strep-tag affinity purification. Strep-tactin sepharose beads were used. Purified dKDM4a was eluted with D-Desthiobiotin. The purified enzyme was in buffer W with 20 % glycerol. Preparation details can be found in method section. To test *in vitro* enzymatic activity of dKDM4a, an increasing amount of dKDM4a enzyme was incubated together with 10  $\mu$ M of peptide substrate in addition of co-factors  $\alpha$ -ketoglutarate and  $\text{Fe}^{2+}$  at 26 °C for 20 min (see Figure 3.3 a). H3.3\_31.41 containing trimethylated lysine in K36 was used as substrate.



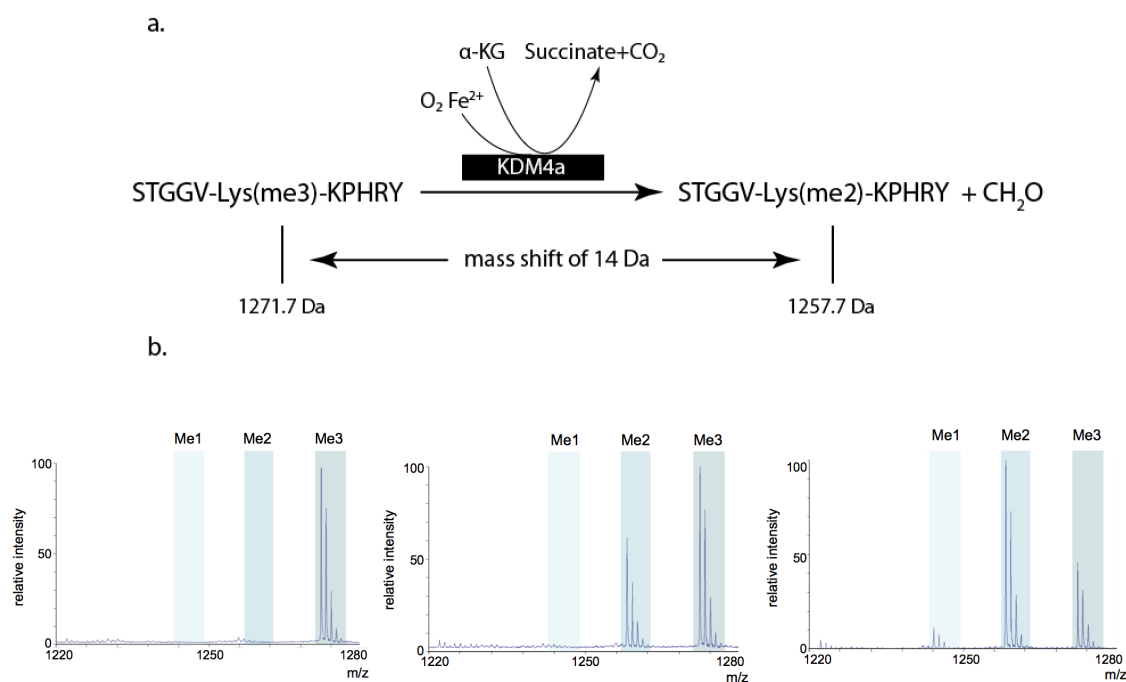
**Figure 3.2 | *In silico* identified Jumonji domain containing proteins**

a 259 proteins in human contain Jmj-C domain specific fold and 72 proteins contain Jmj-N domain specific fold. 20 proteins in fly contain Jmj-C domain specific fold. 10 proteins contain Jmj-N domain specific fold. Bioinformatic identification is performed by CDART from NCBI.

b Schematic illustration of 20 proteins containing Jmj-C domain [c118224] folding in *Drosophila Melanogaster*. Different conserved domains are illustrated in upper right corner.

As shown in Figure 3.3, H3K36me3 was demethylated to H3K36me2 and H3K36me1. Negative control has been done with same reaction setting without adding dKDM4a. Mass-to-charge values of H3K36me3, H3K36me2 and H3K36me1 were measured by MALDI-

ToF Mass Spectrometer Voyager-DE STR (Applied Biosystems). It detects the 14 Dalton mass shift when one methyl group is removed. Mass spectra were indicated in Figure 3.3 b. The most abundant ion species was set as 100 % intensity. Relative intensities of all the other ions species to the most abundant ion are indicated in y axis. Mass-to-charge



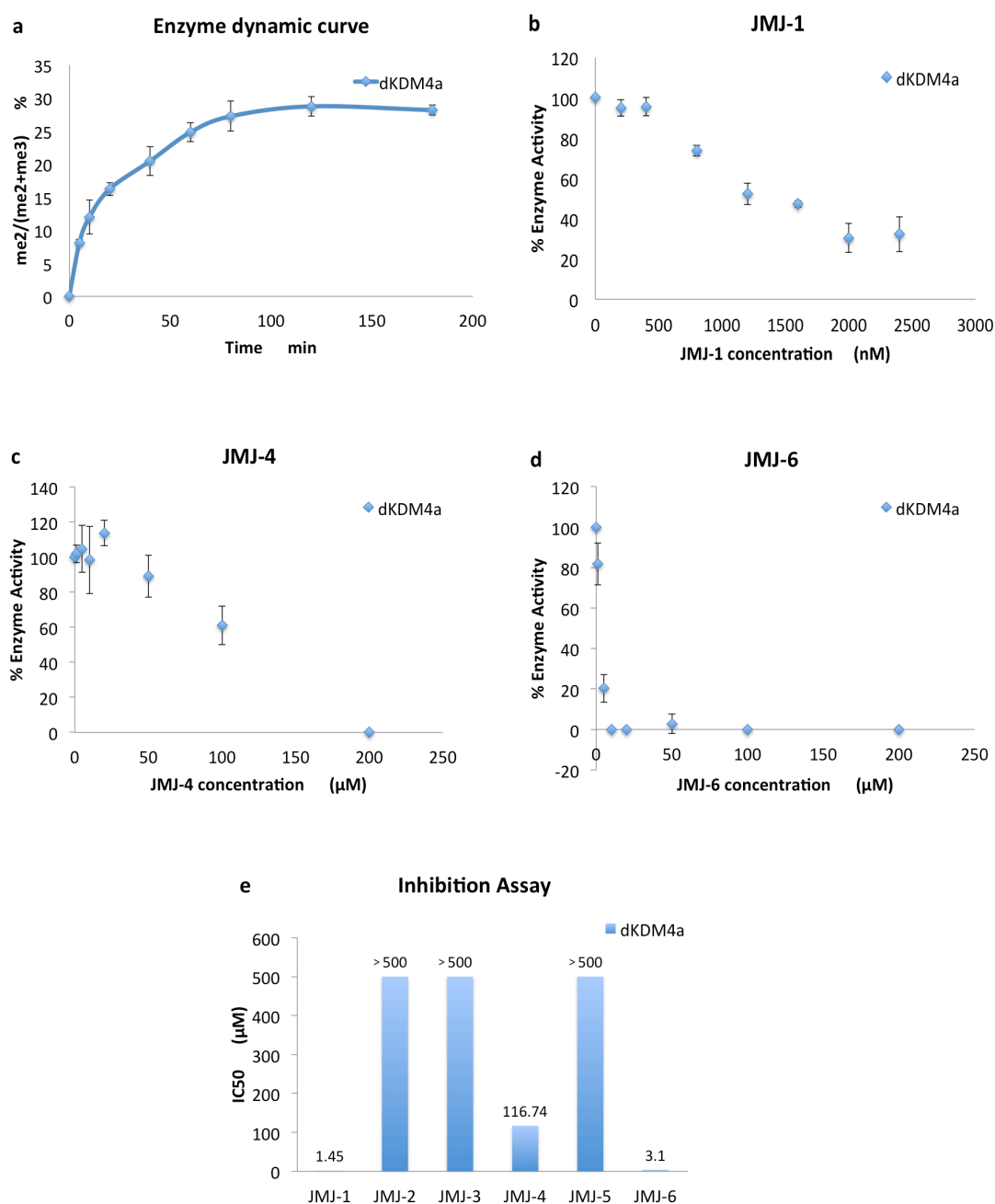
**Figure 3.3 | HDM *in vitro* assay**

a Schematic illustration of histone demethylation reaction *in vitro*. Recombinant full length drosophila KDM4a catalyzed demethylation of H3K36me3 to H3K36me2 in the presence of  $\alpha$ -ketoglutarate, O<sub>2</sub> and Fe<sup>2+</sup>. Meanwhile  $\alpha$ -ketoglutarate was converted to succinate. One molecule of CH<sub>2</sub>O and one molecule of CO<sub>2</sub> were released as byproducts.

b Spectra of the peptide H3.3\_31\_41 with H3K36me3 modification shown here indicating the demethylation of H3K36me3 to H3K36me2 and H3K36me2 to H3K36me1 by dKDM4a *in vitro*.

m/z value is displayed in x axis. The matrix  $\alpha$ -cyano-4-hydroxycinnamic acid was applied to assist peptide to absorb energy from laser and desorb. All the ion species formed by co-crystallization with  $\alpha$ -cyano-4-hydroxycinnamic acid carry one positive charge. Peptide obtained one proton and got ionized as MH<sup>+</sup> which ended up as the m/z value displayed in Figure 3.3 b.

### 3.1.4 KDM4 inhibitors *in vitro* investigation



**Figure 3.4 | dKDM4a inhibition assay in vitro**

a Enzyme dynamic curve of dKDM4a. Time course of 0, 5, 10, 20, 40, 60, 80, 120, 180 min was applied to demethylation reaction with 1  $\mu$ M of dKDM4a enzyme and 10  $\mu$ M of H3K36me3 peptide.

b Inhibitor JMJ-1 concentration was titrated with 0, 200, 400, 800, 1200, 1600, 2000, 2400 nM into 1  $\mu$ M dKDM4a enzyme.

c Inhibitor JMJ-4 concentration was titrated with 0, 1, 5, 10, 20, 50, 100, 200  $\mu$ M into 1  $\mu$ M dKDM4a enzyme.

d Inhibitor JMJ-6 concentration was titrated with 0, 1, 5, 10, 20, 50, 100, 200  $\mu$ M into 1  $\mu$ M dKDM4a enzyme.

e IC50 values of JMJ-1, JMJ-2, JMJ-3, JMJ-4, JMJ-5 and JMJ-6 were indicated in y axis with  $\mu$ M as unit.



In order to determine the linear phase of the *in vitro* demethylation reaction, 1  $\mu\text{M}$  of dKDM4a was incubated with 10  $\mu\text{M}$  of H3K36me3 peptide for 0, 5, 10, 20, 40, 60, 80, 100 and 120 min (Figure 3.4 a). Experiment has been performed in triplicates. Relative intensities read from H3K36me2 and H3K36me3 peptide spectra have been re-integrated as percentage of  $\text{H3K36me2}/(\text{H3K36me2} + \text{H3K36me3})$ . This is depicted as intensity in Figure 3.4 a for y axis. Average values of three replicates are indicated in y axis and standard deviation is displayed as error bars. X axis indicates the time point investigated as minute. The enzymatic reaction was linear for at least 20 min. Reaction was restricted after 40 min and reached plateau phase after 120 min. No more demethylation product has been formed between 120 min and 180 min. Therefore, the reaction products were measured after 20 min in the following inhibition assay in Figure 3.4 b, c, d.

Dosage response curve of inhibitor JMJ-1, JMJ-4 and JMJ-6 are indicated in Figure 3. 4 b, c, d. Average values of three replicates are indicated in the y axis and standard deviation are displayed as error bars. Demethylation activity of 20 min from Figure 3.4 a was calculated as 100 % of enzyme activity. Different inhibitors at different concentration were incubated with 1  $\mu\text{M}$  dKDM4a enzyme for 5 min at room temperature before the reaction was initiated by the addition of substrates. In Figure 3.4 b, inhibitor JMJ-1 concentration was titrated with 0, 200, 400, 800, 1200, 1600, 2000, 2400 nM. Enzyme activity percentage is indicated in y axis. Inhibitor JMJ-1 concentration with nM as unit is indicated in x axis. In Figure 3.4 c, inhibitor JMJ-4 concentration was titrated with 0, 1, 5, 10, 20, 50, 100, 200  $\mu\text{M}$ . Inhibitor JMJ-4 concentration with  $\mu\text{M}$  as unit is indicated in x axis. In Figure 3.4 d, inhibitor JMJ-6 concentration was titrated with 0, 1, 5, 10, 20, 50, 100, 200  $\mu\text{M}$ . Experiment control has been done with 10 % DMSO (v/v) concentration pre-incubation with 1  $\mu\text{M}$  dKDM4a enzyme before demethylation reaction was set up. And the final DMSO (v/v) concentration in the reaction was 1 %. 1 % DMSO (v/v) did not inhibit dKDM4a demethylase activity *in vitro*.

IC<sub>50</sub> values of inhibitor JMJ-1, JMJ-4 and JMJ-6 were calculated from the dosage response curve illustrated in Figure 3.4 b, c, d. Fitting formulas used to calculate IC<sub>50</sub> values were as following:

JMJ-1 :  $y = -0.0392x + 106.86$ ,  $R^2 = 0.9636$ , unit for y is nM;

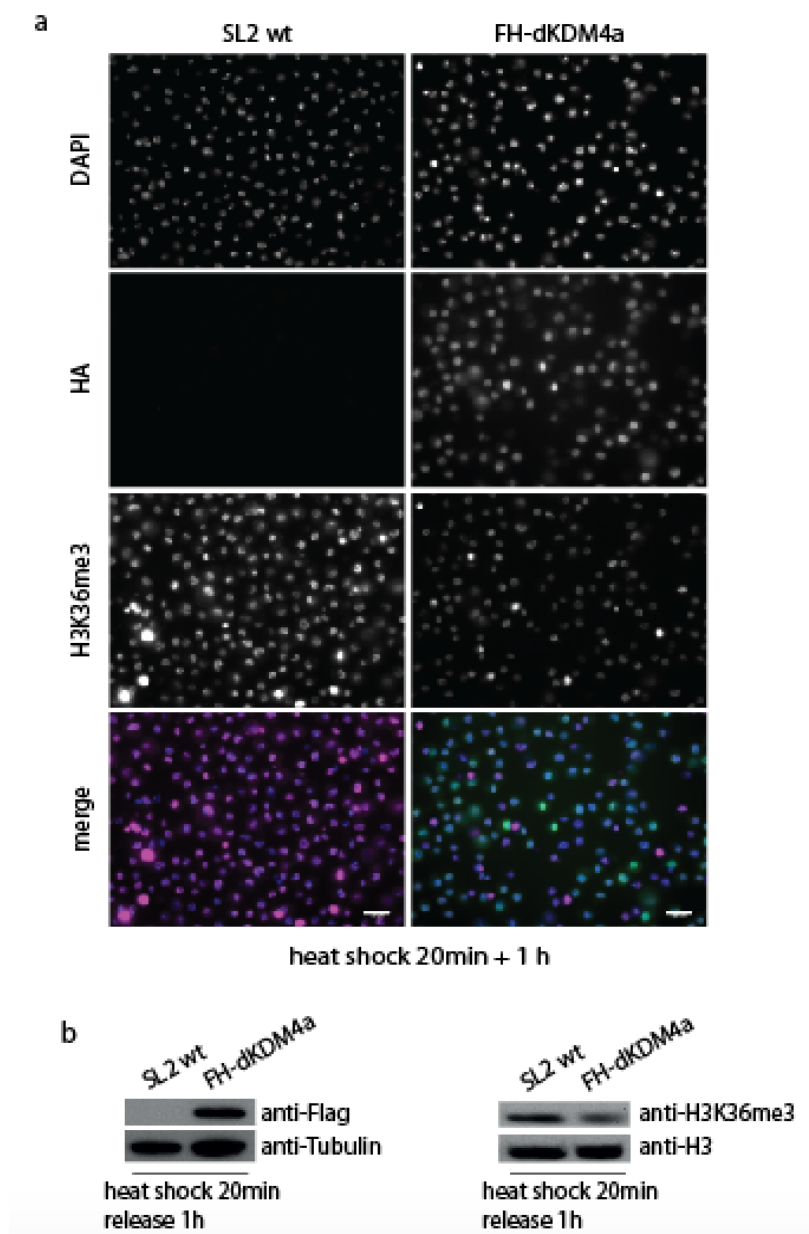
JMJ-4 :  $y = -0.5966x + 119.65$ ,  $R^2 = 0.9996$ , unit for y is  $\mu\text{M}$ ;

JMJ-6 :  $y = -15.782x + 98.925$ ,  $R^2 = 0.9991$ , unit for y is  $\mu\text{M}$ .

For inhibitor JMJ-2, JMJ-3 and JMJ-5, no inhibition effects were observed when incubating 500  $\mu$ M of inhibitors with 1  $\mu$ M dKDM4a enzyme individually. Therefore the IC<sub>50</sub> for those three compounds were considered > 500  $\mu$ M. No determinations of IC<sub>50</sub> value were performed with them.

### 3.1.5 dKDM4a demethylase activity towards H3K36me3 *in vivo*

To establish a dKDM4a over-expressing stable cell line, dKDM4a full length encoding sequence in pHFW vector was transfected into low passage SL2 cells. X-tremeGENE HP DNA Transfection Reagent was used for transfection. dKDM4a was fused to 3 $\times$ Flag-3 $\times$ HA tag at the N-terminus. In Figure 3.5 a, over-expression was induced by heat shock at 37 °C for 20min via Hsp70 heat shock promotor. It was followed by 1 hour release at 26 °C after induction. Cells were de-attached and seeded in coverslides to make them settle down for 30 min. Immunofluorescence staining was performed as described in the methods. Fluorescent signal were recorded immediately after the slides preparation. Staining of HA tag represented dKDM4a over-expression. Monoclonal rat primary antibody  $\alpha$ -HA R001 3F10 from Rothe was used. Donkey anti rat Alexa 488 was used as secondary antibody which gave the green signal in merge channel. Staining of H3K36me3 was performed using polyclonal rabbit primary antibody Ab9050 from Abcam. Donkey anti rabbit Alexa 647 was used as secondary antibody which gave the magenta signal in merge channel. Final concentration of 0.5  $\mu$ g/ml DAPI was used to stain DNA as control to indicate the distance of the cells to the detector. This gave the blue signal in merge channel. Grayscale was used to display single channel image in order to get optimized visualization. White scale bar indicated 20  $\mu$ m. Exposure times for DAPI, HA and H3K36me3 were 15 ms, 120 ms and 400 ms individually. The same display parameters have been applied for the same channel. Western blotting has been performed to detect variation of H3K36me3 signal upon dKDM4a over-expression in SL2 cells (Figure 3.5 b). dKDM4a over-expression was verified by western blotting using an anti Flag antibody and an anti tubulin antibody as loading control. Acid extracted histone proteins were used to detect variation of H3K36me3 upon dKDM4a over-expression. Histone H3 was used as loading control. Over-expression was induced by 20 min heat shock at 37 °C and followed up by 1 hour release at 26 °C.



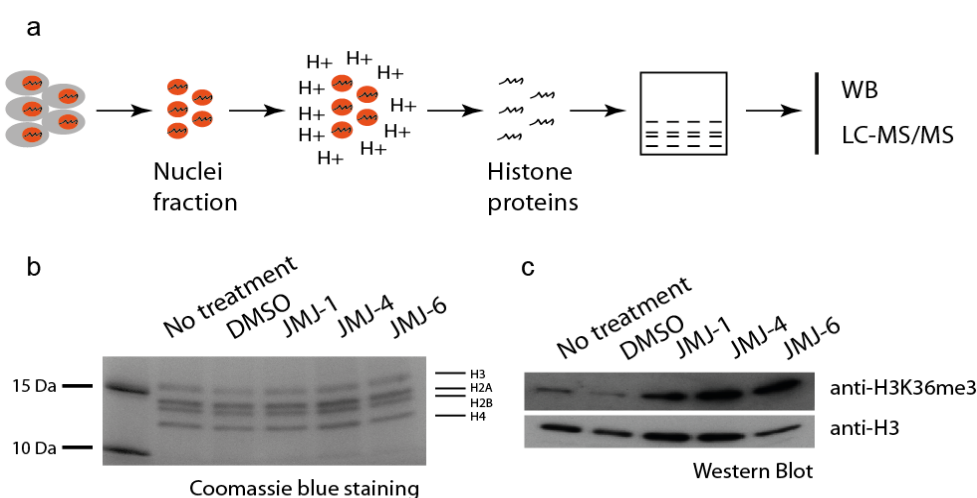
**Figure 3.5 | dKDM4a demethylated H3K36me3 in cell**

a Immunofluorescent detection was performed by staining HA tag to indicate over-expression of dKDM4a and staining H3K36me3 to indicate the variation of dKDM4a substrate. Same exposure time for same channel and same display setting was applied in the fluorescent signal analysis. DAPI staining of DNA was used as control to indicate the distance of the cells to the detector. Grayscale was used to display single channel image. White scale bar represented 20  $\mu$ m.

b Western blotting against Flag tag indicating over-expression of dKDM4a, blotting against Tubulin was used as proteins loading control; acid extracted histone proteins were used to detect variation of H3K36me3 upon over-expression of dKDM4a using histone H3 as loading control. Over-expression was induced by 20 min heat shock at 37 °C and followed up by 1 hour release at 26 °C.

### 3.1.6 Inhibitor JMJ-1, JMJ-4 and JMJ-6 inhibited dKDM4a demethylase activity towards H3K36me3 *in vivo* verified by antibody based substrate detection

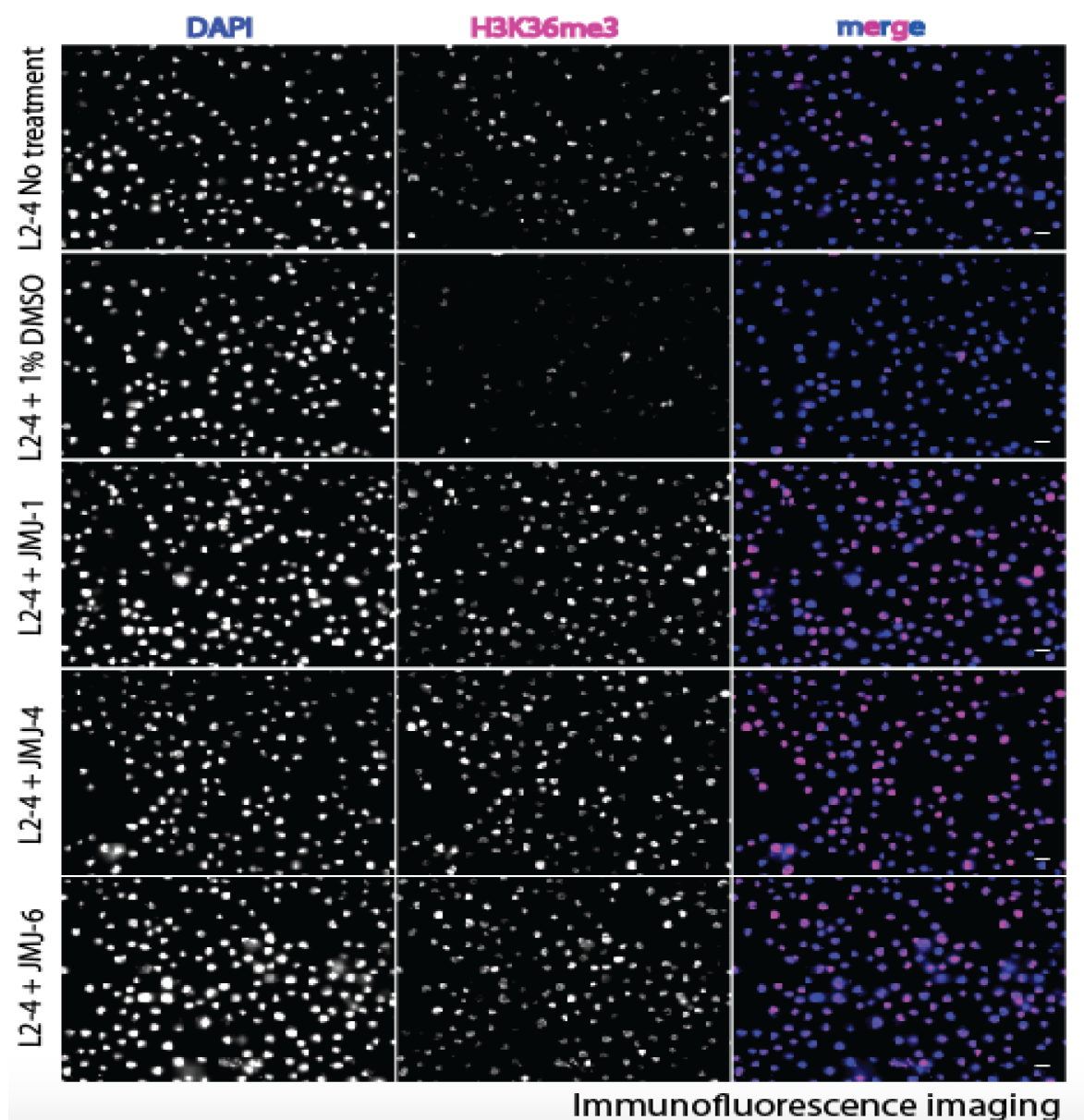
*Drosophila* cell line L2-4 has been used to investigate inhibition effect of dKDM4a inhibitors *in vivo*. 100  $\mu$ M of inhibitor JMJ-1, JMJ-4 and JMJ-6 were added to the culture medium individually and incubated at 26 °C for 48 hours. Equivalent amounts of solvent at final concentration of 1 % DMSO (v/v) and no treatment conditions were used as control groups. Cell pellets were harvested and followed by nuclei isolation. Acid extraction has been performed with nuclei fraction. Extracted histone proteins were separated by SDS-PAGE gel followed by western blotting detection or LC-MS/MS measurement. A schematic illustration of the workflow with histone proteins were indicated in Figure 3.6 a. Histone proteins in SDS-PAGE gel were stained by coomassie G250 as shown in Figure 3.6 b. Variation of dKDM4a substrate H3K36me3 was detected by western blotting using anti-H3K36me3 specific antibody. Comparing to L2-4 cells with no treatment, 1 % DMSO (v/v) treatment led to decrease of H3K36me3. Cells treated with 100  $\mu$ M final concentration of inhibitor JMJ-1, JMJ-4 and JMJ-6 individually shown increased level of H3K36me3 based on antibody detection by western blotting. Total histone protein H3 was used as loading control (see Figure 3.6 c). Chemiluminescence detection was applied using ECL Prime Western Blotting Detection Reagent from Amersham.



**Figure 3.6 | Inhibitor JMJ-1, JMJ-4 and JMJ-6 inhibited dKDM4a demethylase activity towards H3K36me3 *in vivo* detected by WB**

- L2-4 cell nuclei fraction has been isolated from cell pellet, acid extraction has been applied to isolate histone proteins. Histone proteins have been immobilized in SDS-PAGE gel and followed up by western blotting and LC-MS/MS detection.
- SDS-PAGE gel separation of acid extracted histone proteins from L2-4 cells, protein bands were visualized by coomassie blue staining.
- L2-4 cells were cultured with 100  $\mu$ M of inhibitor JMJ-1, JMJ-4 and JMJ-6 individually for 48 hours. Cells with no treatment and 1 % DMSO have been used as control. Blotting against H3K36me3 indicated variation of dKDM4a substrate and blotting against histone H3 total protein was used as loading control.

Immunofluorescence imaging experiment has been performed to investigate the inhibition response at a single cell level as shown in Figure 3.7.



**Figure 3.7 | Inhibitor JMJ-1, JMJ-4 and JMJ-6 inhibited dKDM4a demethylase activity towards H3K36me3 *in vivo* detected by immunofluorescence imaging**

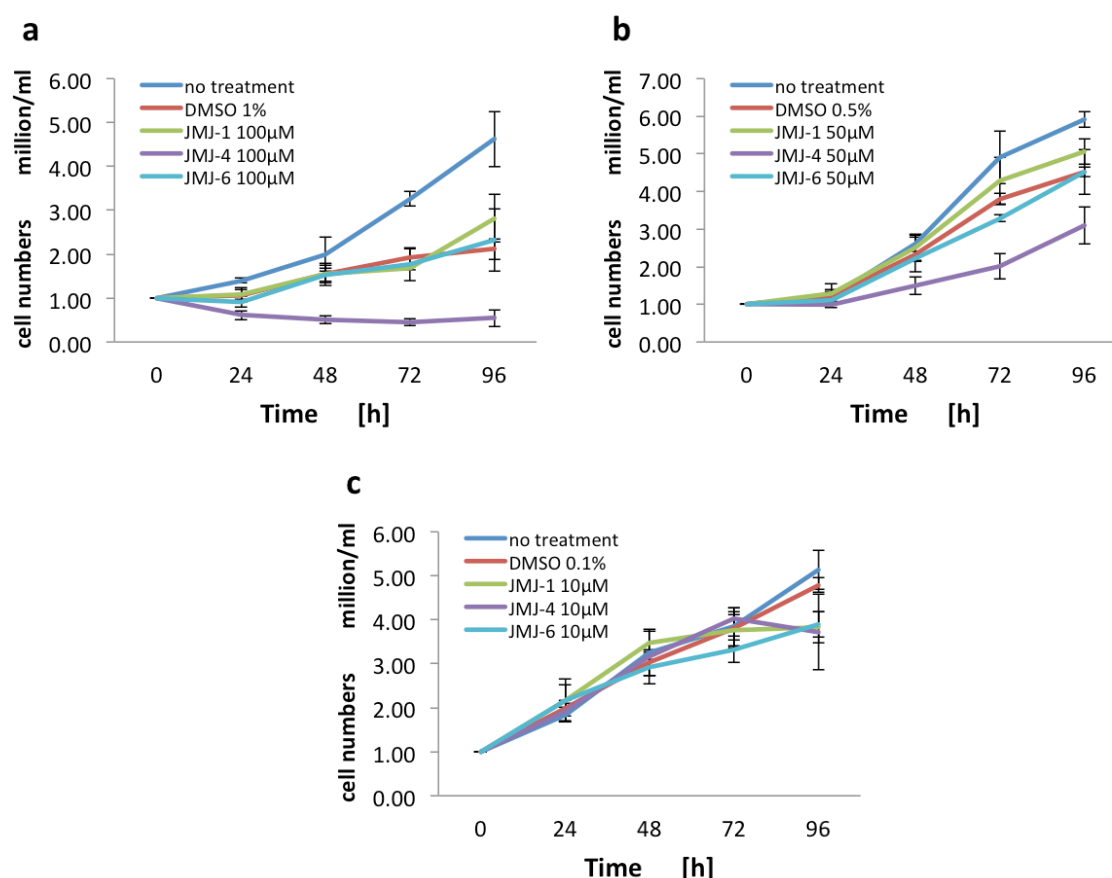
Five experiment conditions have been applied: L2-4 cells with no treatment, L2-4 cells treated with 1 % DMSO (v/v), L2-4 cells treated with 100  $\mu$ M final concentration of inhibitor JMJ-1, JMJ-4 and JMJ-6 individually. Cells were incubated at 26 °C for 48 hours before harvest for immunofluorescence microscopy detection. DAPI staining of DNA was used as control to indicate the distance of the cells to the detector. Staining of H3K36me3 indicated the variation of dKDM4a substrate response upon five conditions applied here. Grayscale was used to display single channel image. Blue staining represented DNA, magenta staining represents H3K36me3 in merge channel. White scale bar represented 10  $\mu$ m.

The same treatments setup as western blotting have been applied. L2-4 cells with no treatment, L2-4 cells treated with final concentration 1 % DMSO (v/v), L2-4 cells treated

with 100  $\mu\text{M}$  final concentration of inhibitor JMJ-1, JMJ-4 and JMJ-6 individually. Cells were incubated at 26  $^{\circ}\text{C}$  for 48 hours before harvest for immunofluorescence microscopy detection. DAPI staining of DNA was used as control to indicate the distance of cells to detector which gave the blue signal in merge channel. H3K36me3 staining was performed using polyclonal rabbit primary antibody Ab9050 and donkey anti rabbit Alexa 647 as secondary antibody which gave the magenta signal in merge channel.

Grayscale was used to display single channel image in order for an optimized visualization. The scale bar indicates 10  $\mu\text{m}$ . Exposure time for DAPI channel was 15 ms and for H3K36me3 channel was 150 ms. The same display parameters have been applied for the same channel.

### 3.1.7 Dosage response of L2-4 cells growth curve to inhibitors



**Figure 3.8 | Inhibitors' dosage response to L2-4 cells growth curve**

Time course of 0, 24, 48, 72, 96 hours has been applied to L2-4 cells treated with 10  $\mu\text{M}$ , 50  $\mu\text{M}$ , 100  $\mu\text{M}$  of inhibitors as three group a, b c. Each group consisted of L2-4 cells with no treatment (dark blue line), L2-4 cells treated with DMSO (red line), L2-4 cells treated with JMJ-1 (green line), JMJ-4 (purple line) and JMJ-6 (light blue line). Mean value of cell density from 4 biological replicates was indicated as cell numbers (million cells per ml) plotted in y axis. The error bar indicated standard derivation.

L2-4 cells were seeded with density of 1 million cells per ml as starting point. Time course of 0, 24, 48, 72, 96 hours has been applied to investigate inhibitor effect on cell growth. Three groups of experiments have been done in 4 replicates. First group: L2-4 cells with no treatment, L2-4 cells treated with 1 % DMSO (v/v), L2-4 cells treated with 100  $\mu$ M final concentration of inhibitor JMJ-1, JMJ-4 and JMJ-6 individually (Figure 3.8 a). Second group: L2-4 cells with no treatment, L2-4 cells treated with 0.5 % DMSO (v/v), L2-4 cells treated with 50  $\mu$ M final concentration of inhibitor JMJ-1, JMJ-4 and JMJ-6 individually (Figure 3.8 b). Third group: L2-4 cells with no treatment, L2-4 cells treated with 0.1 % DMSO (v/v), L2-4 cells treated with 10  $\mu$ M final concentration of inhibitor JMJ-1, JMJ-4 and JMJ-6 individually (Figure 3.8 c).

In Figure 3.8 a, b, c, time course was displayed in x axis, cell density was displayed in y axis. Data points indicate the mean value from 4 biologic replicates and error bar indicated standard deviation. Growth curve of L2-4 cells with no treatment is represented by a dark blue line, L2-4 cells treated with DMSO are represented by a red line, L2-4 cells treated with JMJ-1 JMJ-4 and JMJ-6 are represented by a green line, purple line and light blue line respectively.

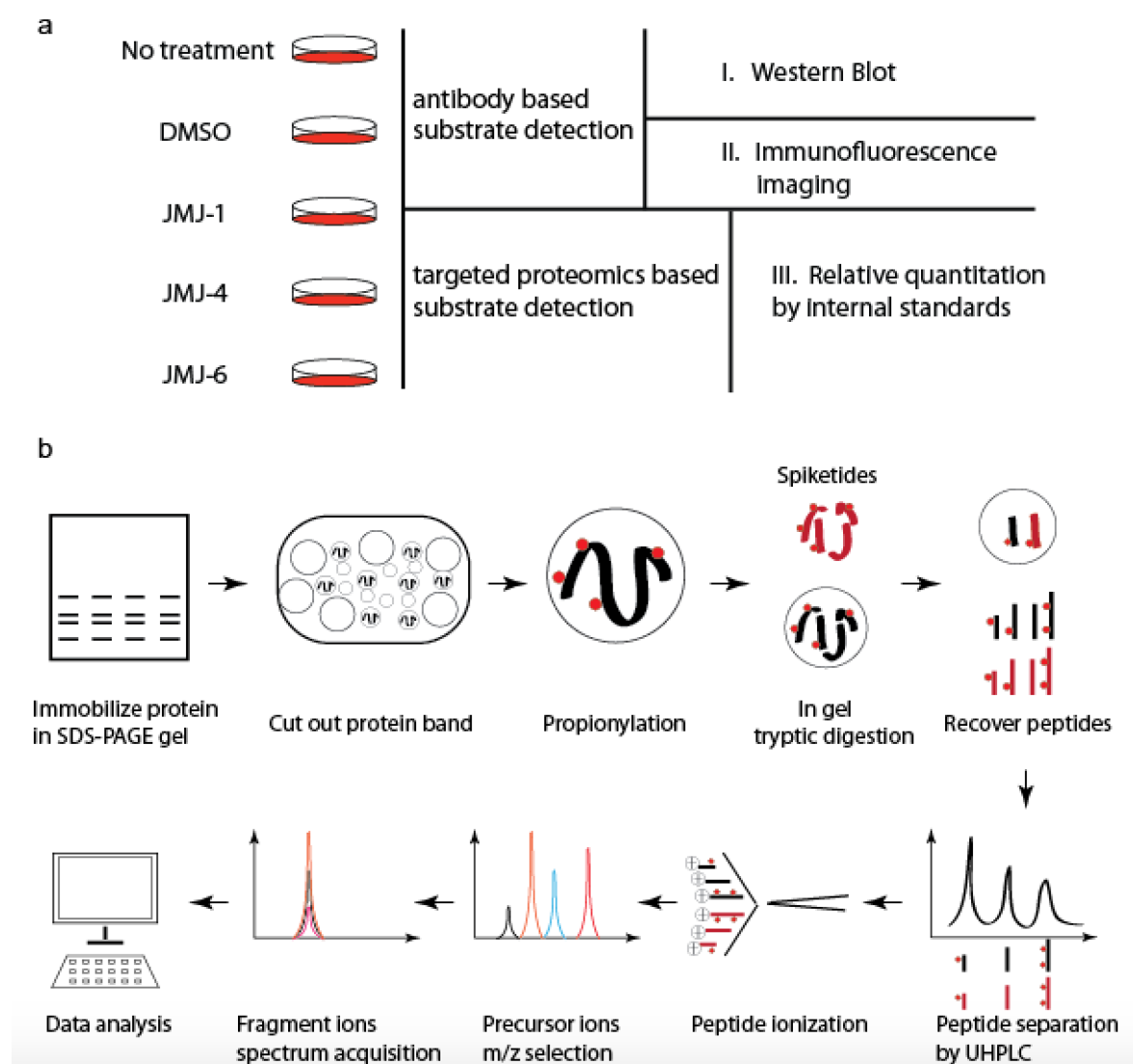
### **3.1.8 Inhibitor JMJ-1, JMJ-4 and JMJ-6 inhibited dKDM4a demethylase activity towards H3K36me3 *in vivo* verified by targeted proteomics based substrate detection**

Conventional detection methods using antibody such as western blotting and immune fluorescence imaging, can be misleading due to a lack of antibody specificity. In order to validate the results from antibody based substrate detection, targeted proteomics based substrate detection has been performed. Spiketides have been applied as internal standard to assign each modification to right ion chromatography elution peak (Figure 3.9 a).

The basic workflow is outlined in Figure 3.9 b. In short, histone protein bands were separated by SDS-PAGE gel, and cut out. Unmodified and mono-methylated lysines of histone were protected from trypsin digestion by propionylation (Villar-Garea et al., 2008). Spiketides with Q-tag were added and digested together with the gel separated histones. After trypsin digestion, light peptide of interest from native protein and heavy peptide of interest from spiketide were generated. The heavy spiketide has exactly same amino acid sequence and modification as light version, but only differs by a C-terminus isotopically labeled arginine. Heavy spiketide was always 10 Da heavier than corresponding light



peptide. Acid extraction has been performed to recover the peptides after tryptic digestion from gel. Peptides were loaded to reversed phase Ultra High Performance Liquid



### Figure 3.9 | Methods development

**a** Biological sample investigated in this study consisted of fly cell line L2-4 cell with no treatment, treated with DMSO, MJM-1, MJM-4 and MJM-6 individually. Two strategies have been applied to study the enzyme and substrate relationship: antibody based substrate detection including western blotting and immune fluorescence imaging and targeted proteomics based substrate detection by including internal standard to do relative quantitation.

**b** Workflow for targeted proteomics: proteins of interest were immobilized in SDS-PAGE gel; protein band was cut out and subjected to propionylation; spiketides library have been added in and digested together with protein in gel; peptides from tryptic digestion have been recovered from gel using acid extraction and loaded to UHPLC column to separate; peptides eluted from UHPLC were ionized by nano electro spray ionization and sprayed into TT6600 mass spectrometer; precursor ions with m/z of interest were selected in the first quadrupole and full mass spectrum of fragmented ions were acquired in the ToF mass analyzer; PeakView® 2.1 and MultiQuant 3.0 software were used for data analysis.



Chromatography (UHPLC) to separate. Peptides eluted from UHPLC were subjected to Electro Spray Ionization and measured by a TT6600 mass spectrometer. TT6600 consists of quadrupole, quadrupole and a ToF mass analyzer. Peptides with  $m/z$  of interest, both light and heavy versions have been selected in the first quadrupole. Collision induced dissociation was performed by the second quadrupole using collision energy which was tailored for the peptides  $m/z$  of interest.

Mass spectra of fragmented ions, together with precursor ions were acquired by Time of Flight mass analyzer. Data analysis has been done with PeakView<sup>®</sup> 2.1 for identification and with MultiQuant 3.0 for relative quantitation.

### 3.1.8.1 Identification of H3K36me3 among H3.27.40 peptides by LC-MS/MS

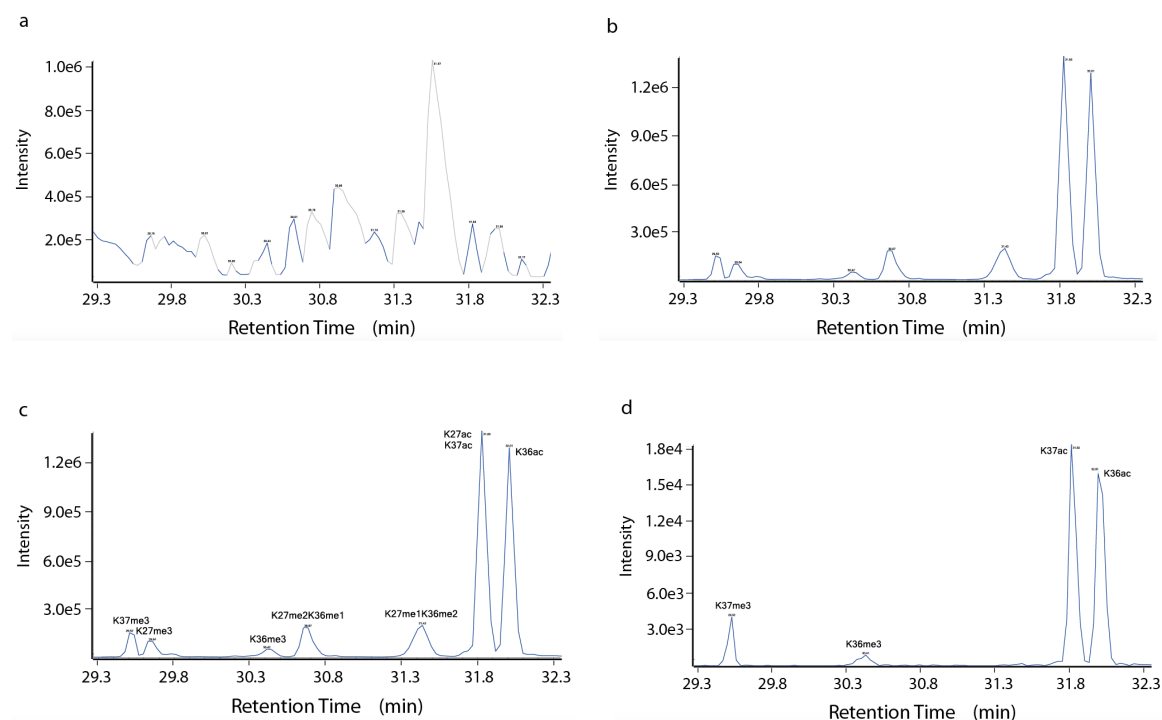
Extracted ion chromatogram from UHPLC separated peptides are shown in Figure 3.10 a. The heavy spiketide peptide generated by trypsin digestion containing H3K36me3 (KSAPATGGV-Lys(Me3)-KPH-R\*) is used as an example to explain how to assign modification to the right peak. This peptide was protonated to carry positive charge. The most abundant ions are triply charged by the uptake of three protons ( $H^+$ ). It accounted for a  $m/z$  value of 533.3425 for H3K36me3. Precursor ion for  $m/z$  of 533.3 was selected to target peptide containing H3K36me3 spiketide and accumulated for 40 milliseconds. This ended up with 6 sec peak width at Full Width Half Maximum Height (FWHM) as indicated in Figure 3.10 b and c. Collision energy used particularly for  $m/z$  of 533.3 was 30.3 %.

**Table 3.1  $m/z$  for 8 modifications in peptide H3.27.40 precursor ions and fragmented ions b : y series from heavy spiketides**

Selected precursor ion $m/z$	Isomeric modification H3.27.40_K27.K36.K37	b1 y14	b2 y13	b3 y12	b4 y11	b5 y10	b6 y9	b7 y8	b8 y7	b9 y6	b10 y5	b11 y4	b12 y3	b13 y2	b14 y1
533.3	H3.27.40_3.0.0	171.150	258.182	329.219	426.272	497.309	598.356	655.378	712.399	811.468	995.589	1179.710	1276.763	1413.822	-
	-	-	1427.870	1340.838	1269.801	1172.748	1101.711	1000.663	943.642	886.620	787.552	603.431	419.309	322.257	185.198
	H3.27.40_0.3.0	185.129	272.161	343.198	440.251	511.288	612.336	669.357	726.379	825.447	995.589	1179.710	1276.763	1413.822	-
	-	-	1413.891	1326.859	1255.821	1158.769	1087.732	986.684	929.662	872.641	773.573	603.431	419.309	322.257	185.198
	H3.27.40_0.0.3	185.129	272.161	343.198	440.251	511.288	612.336	669.357	726.379	825.447	1009.568	1179.710	1276.763	1413.822	-
	-	-	1413.891	1326.859	1255.821	1158.769	1087.732	986.684	929.662	872.641	773.573	589.451	419.309	322.257	185.198
	H3.27.40_1.2.0	199.145	286.177	357.214	454.267	525.304	626.351	683.373	740.394	839.463	995.589	1179.710	1276.763	1413.822	-
	-	-	1399.875	1312.843	1241.806	1144.753	1073.716	972.668	915.647	858.625	759.557	603.431	419.309	322.257	185.198
	H3.27.40_2.1.0	157.134	244.166	315.203	412.256	483.293	584.341	641.362	698.384	797.452	995.589	1179.710	1276.763	1413.822	-
	-	-	1441.885	1354.853	1283.816	1186.764	1115.726	1014.679	957.657	900.636	801.567	603.431	419.309	322.257	185.198
	H3.27.40_a.0.0	171.113	258.145	329.182	426.235	497.272	598.320	655.342	712.363	811.431	995.553	1179.674	1276.727	1413.785	-
	-	-	1427.870	1340.838	1269.801	1172.748	1101.711	1000.663	943.642	886.620	787.552	603.431	419.309	322.257	185.198
	H3.27.40_0.a.0	185.129	272.161	343.198	440.251	511.288	612.336	669.357	726.379	825.447	995.553	1179.674	1276.727	1413.785	-
	-	-	1413.854	1326.822	1255.785	1158.732	1087.695	986.647	929.626	872.605	773.536	603.431	419.309	322.257	185.198
	H3.27.40_0.0.a	185.129	272.161	343.198	440.251	511.288	612.336	669.357	726.379	825.447	1009.568	1179.674	1276.727	1413.785	-
	-	-	1413.854	1326.822	1255.785	1158.732	1087.695	986.647	929.626	872.605	773.536	589.415	419.309	322.257	185.198

Light peptides ion chromatogram peak assignments have been done by using corresponding heavy spiketides. They were co-eluted together with each other. In the spiketides library applied in this study, there were 8 modifications for peptide H3.27.40 having  $m/z$  of 533.3 for triply charged ion state. This included H3K27me3, H3K36me3, H3K37me3,

H3K27me1K36me2, H3K27me2K36me1, H3K27ac, H3K36ac and H3K37ac (Figure 3.10 c). Most prominent collision generating b and y series of fragmented ions were investigated.



**Figure 3.10 Identification of H3K36me3 from peptide H3.27.40 spiketides**

a BPC ion chromatogram indicated peptides separation by UHPLC in retention time range of 29.3 to 32.3 min.

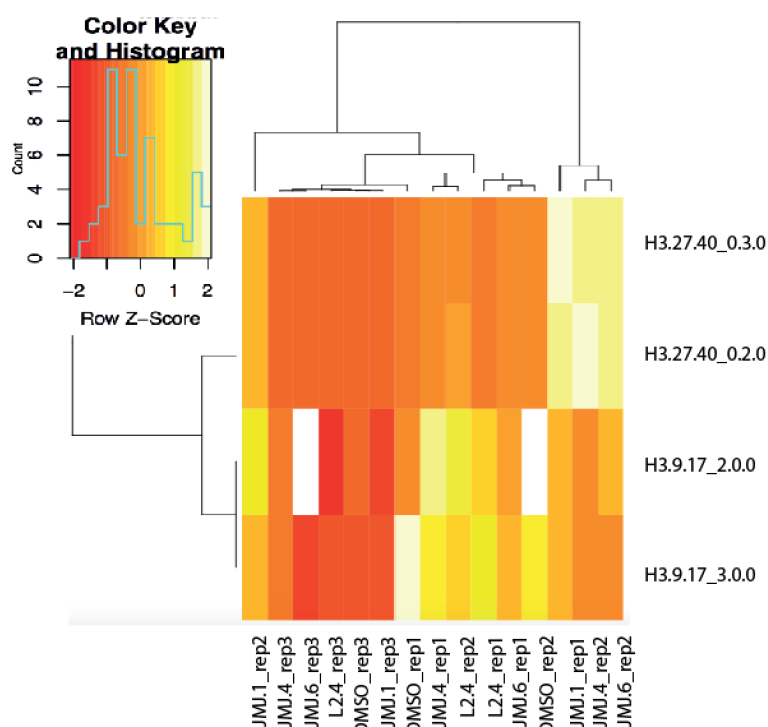
b BPC ion chromatogram indicated selected precursor ions with m/z 533.3 for triply charged peptides in retention time range of 29.3 to 32.3 min.

c Ion chromatogram peak assignment for 8 modifications in H3.27.40 spiketides with m/z 533.3. They assigned to seven peaks separated by UHPLC. From left to right peak identification: 1<sup>st</sup> peak: H3K37me3, 2<sup>nd</sup> peak: H3K27me3, 3<sup>rd</sup> peak: H3K36me3, 4<sup>th</sup> peak: H3K27me2K36me1, 5<sup>th</sup> peak: H3K27me1K36me2, 6<sup>th</sup> peak: H3K27ac and H3K37ac, 7<sup>th</sup> peak: H3K36ac.

d Extracted fragmented ion b3 chromatogram with m/z 343.20 used as diagnostic ion for H3K36me3 in retention time range of 29.3 to 32.3 min. From left to right peak identification: 1<sup>st</sup> peak: b3 from H3K37me3, 2<sup>nd</sup> peak: b3 from H3K36me3, 3<sup>rd</sup> peak: b3 from H3K37ac, 4<sup>th</sup> peak: b3 from H3K36ac.

Identification and assignment of each ion chromatogram peak for particular modification have been validated by the presence of diagnostic fragmented ion. The diagnostic ions that are unique for each modification are shown in Table 3.1 and Figure 3.10 d. Darker blue indicated m/z for fragmented ions shared between different modifications and lighter blue indicated m/z for unique fragmented ions to the specific modification. Quantitation of each ion chromatogram peak has been done with the abundant diagnostic fragmented ions (Figure 3.10 d). Ion peak intensity was extracted by integrating the area under the peak.

### 3.1.8.2 Quantitation of dKDM4a substrates H3K36me3/me2 and H3K9me3/me2 by LC-MS/MS



**Figure 3.11 Relative quantitation of H3K36me3/me2 and H3K9me3/me2 by LC-MS/MS**

Heat map of RQ values of histone H3 H3K36me3/me2 and H3K9me3/me2. Histones were prepared from fly cell line L2-4 cells with no treatment, treated with DMSO, JMJ-1, JMJ-4 and JMJ-6 individually. Three replicates for each treatment have been included. H3K36me3 = H3.27.40\_0.3.0; H3K36me2 = H3.27.40\_0.2.0; H3K9me3 = H3.9.17\_3.0.0; H3K9me2 = H3.9.17\_2.0.0. 0 = unmodified; 2 = dimethylation, 3 = trimethylation. Z-Score 0 = mean RQ value of the row. Color code indicates the distance to the mean. White color indicates missing value.

3 amino acid residues for peptide H3.9.17: H3.9.17\_K9.S10.K14

3 amino acid residues for peptide H3.27.40: H3.27.40\_K27.K36.K37

Ion intensity value by integrating the area under the chromatogram peak has been used as the value for quantitation. This reflected the peptides ion abundance. Relative quantitation of H3K36me3 has been done by taking the RQ (Relative Quantitation) value.  $RQ \text{ value} = \frac{(H3.27.40\_K36me3\_b3 \text{ Light} / H3.27.40\_K36me3\_b3 \text{ Heavy})}{(H3.41.49\_unmodified \text{ Light} * H3.54.63\_unmodified\_y7 \text{ Light} / H3.41.49\_unmodified \text{ Heavy} / H3.54.63\_unmodified\_y7 \text{ Heavy})}$ . H3.27.40\_K36me3\_b3 Heavy spiketide signal has been used to normalize liquid chromatography column peptide response. H3.41.49\_unmodified Light peptide signal versus heavy spiketide signal ratio has been used to normalize the LC loading material for each protein band in SDS-PAGE. H3.54.63\_unmodified\_y7 Light peptide signal versus heavy spiketide signal ratio has been

used to normalize the MS/MS fragmentation response. RQ values from three replicates were indicated in Figure 3.11.

RQ value for H3K36me2 =  $(\text{H3.27.40\_K36me2\_b3 Light} / \text{H3.27.40\_K36me2\_b3 Heavy}) / (\text{H3.41.49\_unmodified Light} * \text{H3.54.63\_unmodified\_y7 Light} / \text{H3.41.49\_unmodified Heavy} / \text{H3.54.63\_unmodified\_y7 Heavy})$ . RQ values from three replicates were indicated in Figure 3.11.

RQ value for H3K9me3 =  $(\text{H3.9.17\_K9me3 Light} / \text{H3.9.17\_K9me3 Heavy}) / (\text{H3.41.49\_unmodified Light} / \text{H3.41.49\_unmodified Heavy})$ . RQ value for H3K9me2 =  $(\text{H3.9.17\_K9me2 Light} / \text{H3.9.17\_K9me2 Heavy}) / (\text{H3.41.49\_unmodified Light} / \text{H3.41.49\_unmodified Heavy})$  (see Figure 3.11).

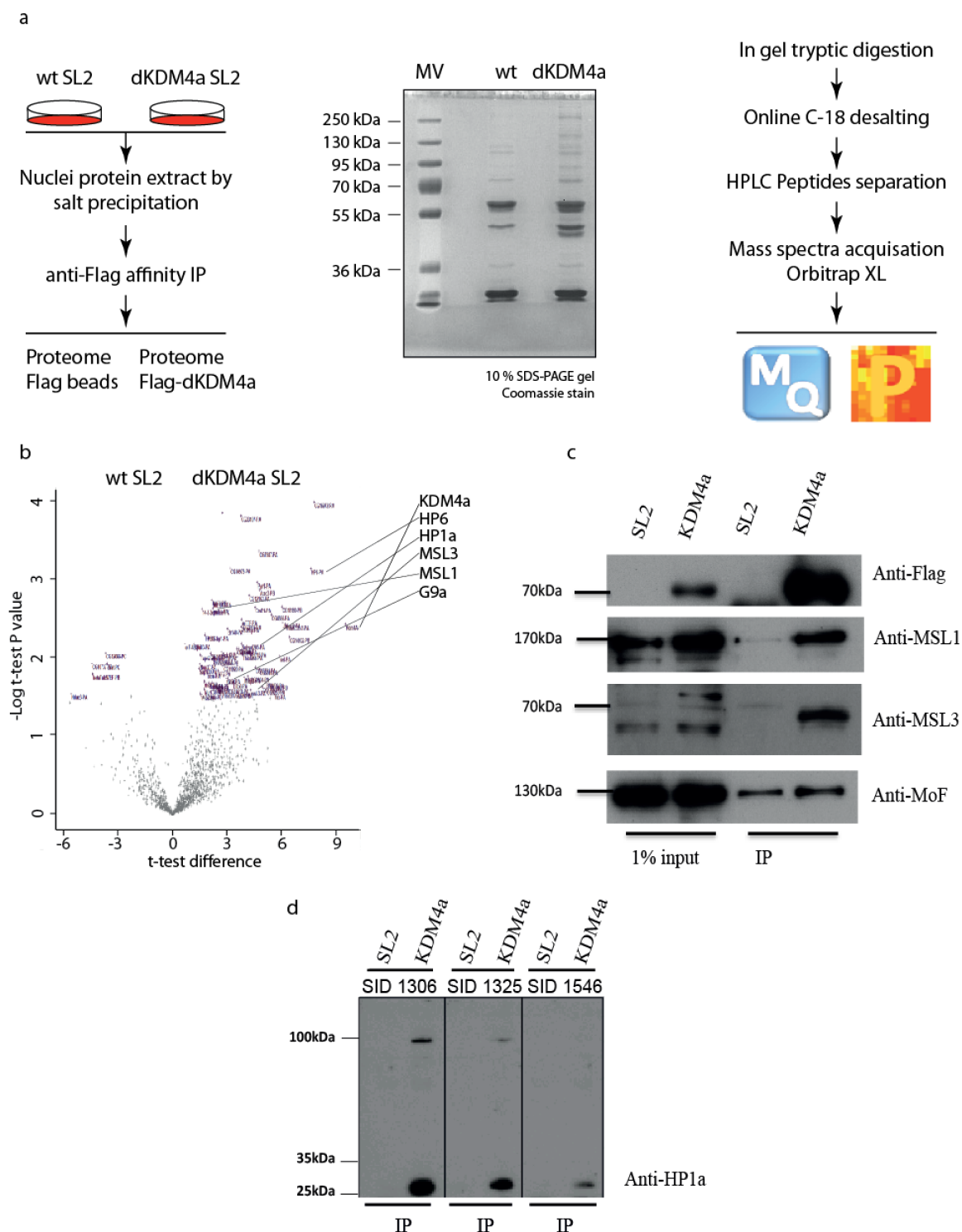
### 3.2 Functional analysis of dKDM4a

Apart from already known demethylase enzyme activity, biologic function of dKDM4a is dependent on the location of complex formed with the presence of dKDM4a. In order to study the complex formed together with dKDM4a, investigation of dKDM4a interactome has been performed by applying shotgun proteomics approach using mass spectrometry.

Full length of dKDM4a was fused to a 3×Flag-3×HA tag at the N-terminus using the expression vector pHFHW. The fly cell line SL2 was used to establish dKDM4a over-expression stable cell line. SL2 wild type cells and SL2 with dKDM4a over-expression cells were cultured in parallel. They were expanded to large scale to prepare nuclei protein extract by salt precipitation using  $(\text{NH}_4)_2\text{SO}_4$ . Nuclei protein extract has been applied to anti-Flag M2 beads to fish Flag tagged dKDM4a and its interactors by affinity immune precipitation. Proteins enriched from SL2 wild type cells and dKDM4a over-expression SL2 cells were loaded in parallel in SDS-PAGE gel. Peptides recovered from cut out bands after trypsin digestion were desalted online before sent for LC-MS/MS measurement. HPLC and Orbitrap XL mass spectrometer were applied. Nano-electrospray was used as ion source. Collision energy used for all collision induced dissociation (CID) was 30 % for all fragmentation to acquire MS/MS spectra. Peptide identification and protein mapping have been done by Andromeda peptide search engine integrated in MaxQuant (Cox et al. 2011). The search was performed against dmel-all-translation-r5.24 fasta database. This database was used to generate theoretical tryptic peptides from whole translation of *Drosophila melanogaster* genome release version 5.24 fasta DNA sequence. The

translation was assembled by 6 possible open reading frames from Watson strand and Crick strand.

Raw mass spectra data acquired from LC-MS/MS were identified by Andromeda algorithm. They were quantified using label free quantification based on iBAQ value by MaxQuant software (see workflow indicated in Figure 3.12 a and protein list identified by mass spectrometer in appendix). Oxidation on methionine, carbamidomethyl on cysteine and acetyl on protein N-terminus residue have been included as fixed modification during alignment with theoretical database. 20 ppm of mass tolerance has been applied. Match between runs of 2 min time window was included. Statistic analysis indicated by volcano plot (Figure 3.12 b) was performed by Perseus software. One experiment set consisted of proteome pulled down by Flag IP from dKDM4a over-expression SL2 cells and wild type SL2 cells as control. Three replicates of each experiment set have been performed. After removing the proteins from contamination and reverse translation, 1337 proteins have been identified in total. Protein IDs (Identities) have been defined by FlyBase (<http://flybase.org>) and converted to gene names. All the proteins identified with P value  $\leq 0.05$  in t-test analysis have been taken into consideration. In Appendix 3, light blue indicated P value  $\leq 0.01$ . Dark blue indicated  $0.01 \leq P \text{ value} \leq 0.05$ . The rest were indicated as black color. Proteins with particular interest of this study have been pointed out in Figure 3.12 b. Star members of Dosage Compensation Complex (DCC) have been validated by western blotting (Figure 3.12 c). Validation of known dKDM4a interactor HP1a was investigated by western blotting with all three replicates shown in Figure 3.12 d. Molecule mass weight of Hp1a is 23.185 kDa. This has been shown as the major band in the blotting. Additional band in position of roughly 100 kDa has shown up in anti-Hp1a blotting after dKDM4a IP.



**Figure 3.12 | Proteome analysis of dKDM4a**

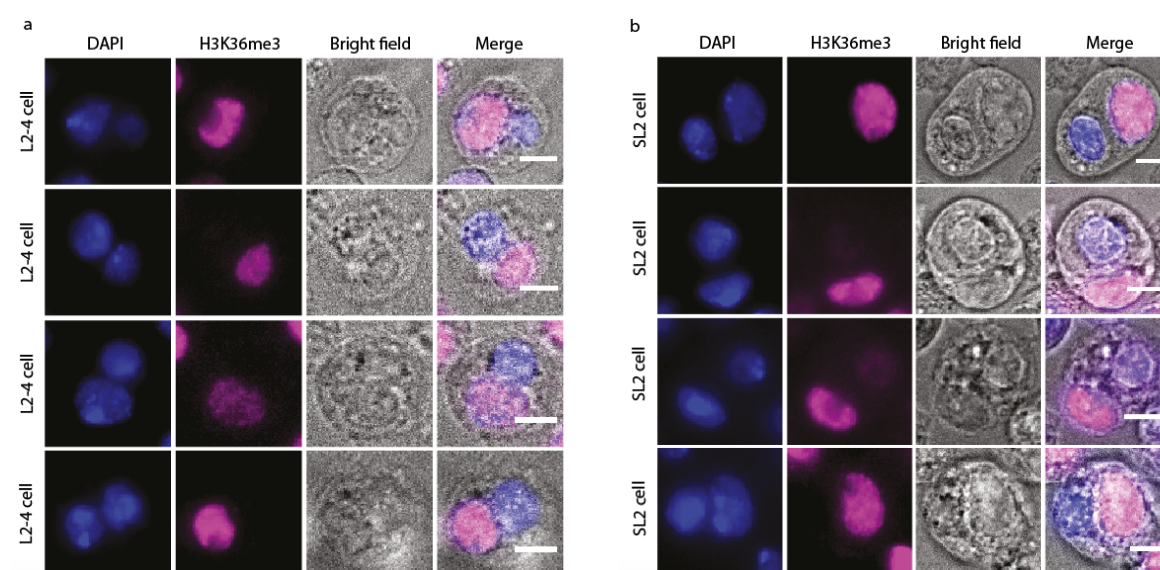
- a Workflow of proteomics analysis for dKDM4a
- b Statistic analysis of proteins identified from anti-Flag IP pull down of nuclei protein extract with wild type SL2 cell and dKDM4a over-expression SL2 cell.
- c Validation of anti-Flag affinity IP enrichment of Flag-tagged dKDM4a, MSL1, MSL3 and MoF by western blotting.
- d Validation of HP1a interaction with dKDM4a by western blotting.



### 3.3 Single nuclear staining of H3K36me3 in bi-nucleated SL2 and L2-4 cells

H3K36me3 is a mark tightly associated with actively transcribed gene body during elongation. Origin of tumor cells may well start from dysfunction of transcriptome profile in single cell. Immune fluorescence imaging has been applied to investigate H3K36me3 variation in single cell level with SL2 cells and L2-4 cells from fly. Single nuclear staining of H3K36me3 has been observed in bi-nucleated cells using antibody detection visualized by immune fluorescence signal shown in Figure 3.13 a and b.

### 3.4 Over-expression of IDH led to a reduction of H3K36me3



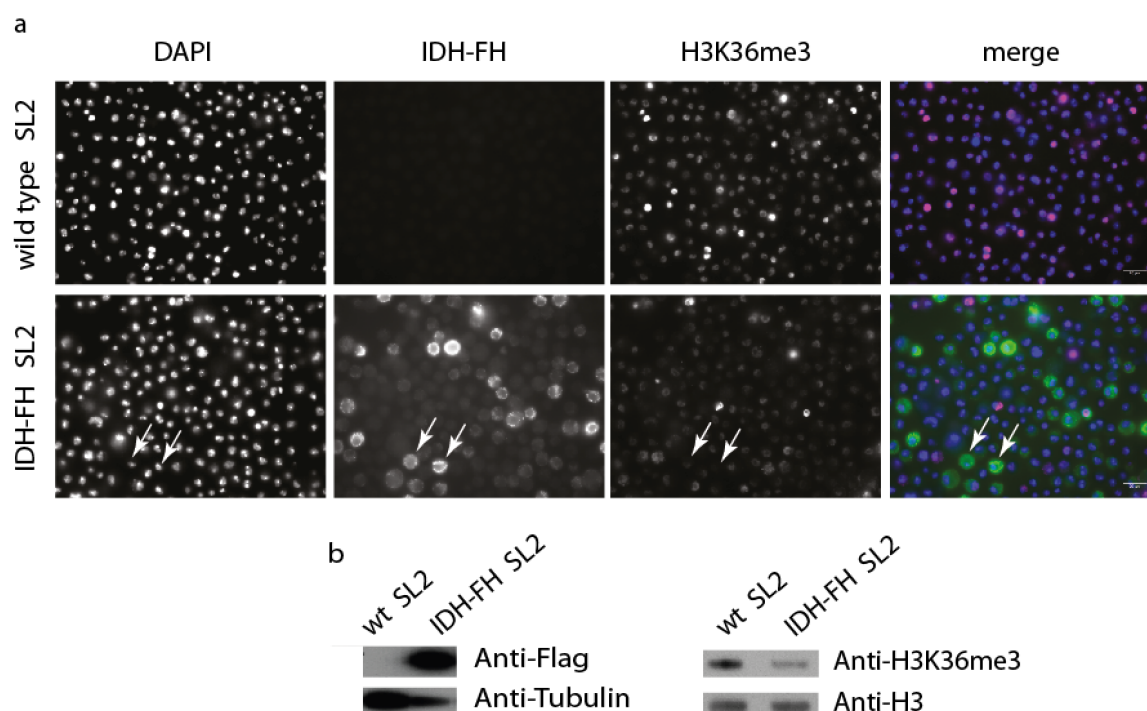
**Figure 3.13 | Immune Fluorescence imaging for H3K36me3 in fly bi-nucleated cell**

DAPI staining of DNA was used as control to indicate the distance of the cells to the detector. Magenta signal represented intensity of H3K36me3. Bright field channel visualized bi-nuclear in single cell. IF staining of L2-4 cells were illustrated in a and IF staining of SL2 cells were illustrated in b. Scale bar represented 5  $\mu$ m.

Demethylase activity of Jumonji domain containing KDM proteins was conferred by the presence of cofactors including bivalent iron, reactive oxygen and  $\alpha$ -ketoglutarate.  $\alpha$ -KG is the intermediate metabolite from tricarboxylic acid cycle. Whether manipulating the cofactors can also regulate Jumonji demethylase enzyme activity, it is still an open question. Here we investigated KDM4a substrates H3K36me3 variation upon IDH over-expression by IF, WB and MS analysis. Stable SL2 cell line with IDH over-expression has been established. IDH (Isocitrate Dehydrogenase) oxidizes isocitrate to  $\alpha$ -ketoglutarate accompanied with reduction of NAD<sup>+</sup> or NADP<sup>+</sup> to NADH or NADPH. Drosophila IDH (CG7176) encoding sequence was constructed in vector pMK33-CFH-HAstop bought from BDGP.

### 3.4.1 Over-expression of IDH led to reduction of dKDM4a substrate H3K36me3

Staining of HA indicates IDH over-expression. Substrate of Jumonji domain containing KDMs was represented by H3K36me3. H3K36me3 signal was reduced upon over-expression of dIDH (CG7176) induced by cooper sulfate. This was detected by immune fluorescence imaging (Figure 3.14 a) and validated by western blotting (Figure 3.14 b). DAPI staining of DNA was included as control to indicate the distance of the cells to the detector. White scale bar represents 20  $\mu$ m. White arrow points to cells showing strongest reduction of H3K36me3 signal upon over-expression of IDH. Exposure time for each channel used here were 20 ms for DAPI, 150 ms for IDH-FH, 600 ms for H3K36me3. Same display parameters were applied in order to compare signal from same channel between different biological samples.



**Figure 3.14 | Over-expression of IDH led to reduction of dKDM4a substrate H3K36me3**

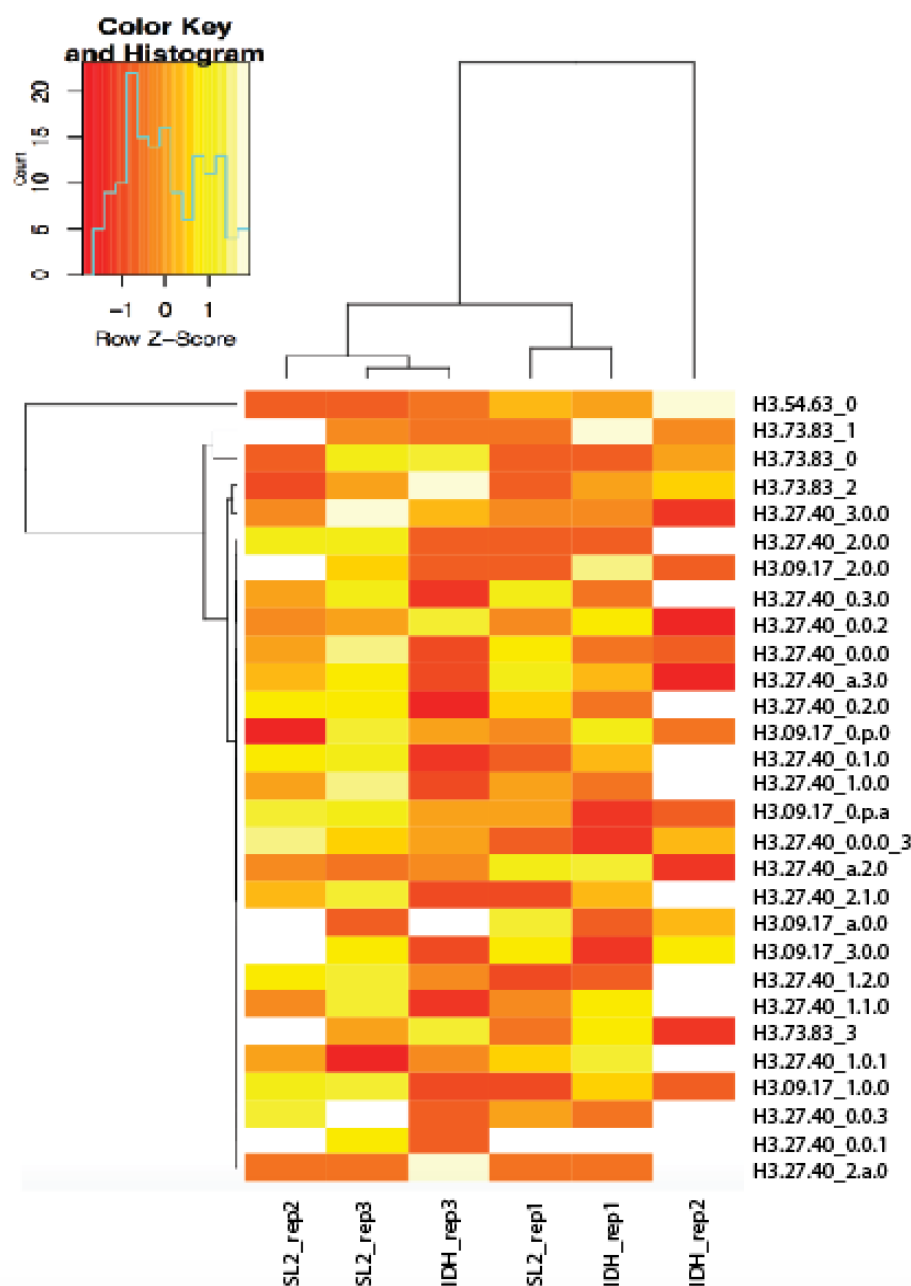
a Immune fluorescence imaging of wild type SL2 cells and IDH-FH transfected SL2 cells. Cells were stained with anti-HA primary antibody representing IDH, anti-H3K36me3 representing substrate of KDM4a.

b Western blotting against Flag tag indicating over-expression of dIDH. Blotting against Tubulin was used as total protein loading control. Blotting against H3K36me3 from wild type SL2 cells and IDH-FH over-expression stable cells were displayed. Blotting of H3 was included as histone protein loading control.

### 3.4.2 Global histone methylation analysis upon over-expression of IDH



Global histone lysine methylation variation upon IDH over-expression in SL2 cells has been investigated by LC-MS/MS analysis. Peptides and amino acid residues investigated in this data set were H3.9.17\_K9.S10.K14, H3.27.40\_K27.K36.K37, H3.54.63\_K56, H3.73.83\_K79. Unmodified peptide H3.41.49\_un was used as normalization signal. Acid extracted histones from three biological replicates of wild type SL2 cells and IDH-FH over-expression SL2 cells were included. 29 modifications were indicated in Figure 3.15. Taken H3K36me3 as example, RQ value = (H3.27.40\_K36me3 Light / H3.27.40\_K36me3 Heavy) / (H3.41.49\_unmodified Light / H3.41.49\_unmodified Heavy). Mean RQ value of the row was set as 0. Color code indicated the distance to the mean value. Missing values from measurement were indicated as white color. Histogram shown in the upper left corner indicated numbers of the sample showing particular Z-score. Targeted proteomics has been applied using UHPLC-TT6600. Synthetic peptides with modifications of interest have been applied to assign each chromatogram peak. Precursor ion ms chromatogram signal has been used for relative quantification.



**Figure 3.15 | Histone methylation analysis of IDH over-expression SL2 stable cell line**

RQ values from precursor ions measured by UHPLC-TT6600\_MRM were indicated as heatmap. RQ values = precursor ion chromatogram Light/Heavy ratio of each modification normalized by precursor ion chromatogram Light/Heavy ratio of H3.41.49\_un. 29 modifications on histone H3 were characterized. Z-Score 0 = mean of the row, color code indicates the distance to the mean. White color indicates missing value.

3 amino acid residues for peptide H3.09.17:

H3.9.17\_K9.S10.K14

3 amino acid residues for peptide H3.27.40:

H3.27.40\_K27.K36.K37

1 amino acid residue for peptide H3.54.63:

H3.54.63\_K56

1 amino acid residue for peptide H3.73.83:

H3.73.83\_K79

0 = unmodified, 1 = mono-methylation, 2 = di-methylation, 3 = tri-methylation, p = phosphorylation, a = acetylation.

### 3.4.3 Generating IDH mutants cell line

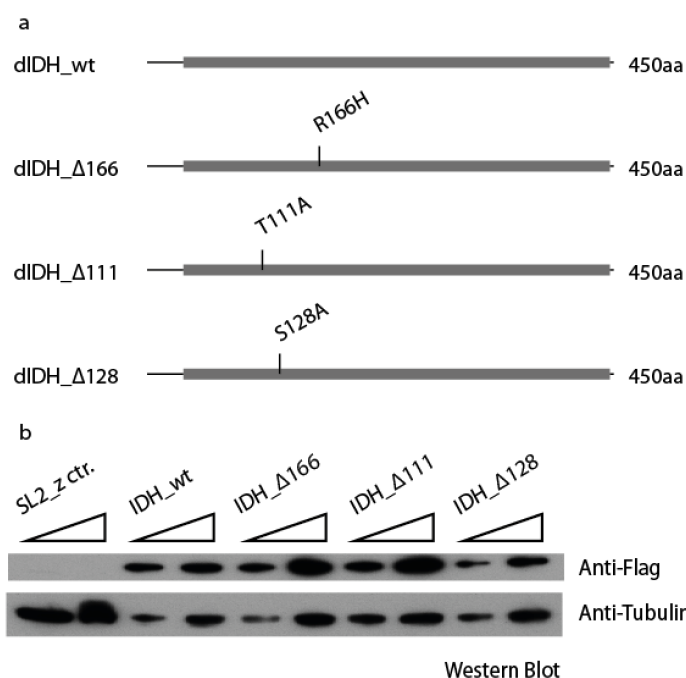
In order to validate the association of reduction effect on histone lysine methylation upon over-expression of IDH with its enzymatic activity, stable cell lines with over-expression of enzymatic null IDH mutants were established. Single nucleic acid point mutation was induced by QuikChange<sup>TM</sup> Site-Directed Mutagenesis Kit using mutagenic primers. IDH (CG7176) mutants constructed in vector pMK33-CFH-HAstop with CuSO<sub>4</sub> promoter have been generated in SL2 cells.

The reference for analysis of *Drosophila melanogaster* IDH enzymatic null mutants is the encoding sequence of Gene bank: NM\_168265 for human IDH1 (Table 3.2).

**Table 3.2 Molecular analysis of *Drosophila* IDH mutants**

Mutant	Molecular alteration			Corresponding to human mutant Yang et al., 2010
	DNA residue in ORF	Codon	Amino acid	Amino acid
IDH 111	331-333	ACT>GCT	T111A	IDH1 T77A
IDH 128	382-384	TCG>GCG	S128A	IDH1 S94A
IDH 166	496-498	CGC>CAC	R166H	IDH1 R132H

Over-expression of IDH protein level has been validated by western blotting using anti-Flag antibody to detect IDH-FH C-terminus tag shown in Figure 3.16.



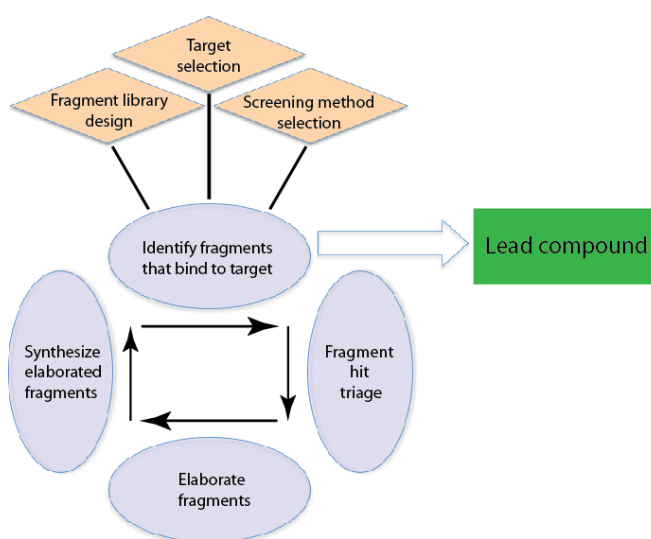
**Figure 3.16 | IDH mutants stable cell line**

Stable cell lines with over-expression of IDH enzymatic null mutants have been established in SL2 cell. a Schematic illustration of point mutations; b Validate of over-expression for wild type IDH and mutated IDH by western blotting

## 4. Discussion

### 4.1 KDM4a inhibitor screening

Fragment based drug discovery starts with selecting the target of interest. In the case of targeting protein enzymes, fragment library design is based on the relevant crystal structure if available. Screening method is selected to determine the binding affinity of the fragments to target. *In silico* docking has been applied to filter out fragments which have lower binding affinity to the target. The top hits with low binding energy and high selectivity towards target of interest can be synthesized directly or combined together to synthesize. The modified molecules are subjected to experimental assay to determine binding affinity. High throughput experimental assay screening has to be applied to filter out redundant structures. 2 or 3 cycles of elaboration are needed before generating lead compound which is qualified to go through optimization steps in medicinal chemistry. (Figure 4.1 adapted from DesJarlais et al., Chapter 6, Methods in ENZYMOLOGY, Volume 493)



**Figure 4.1 | Fragment based drug discovery workflow**

General workflow to develop small molecule drugs against target.

Figure adapted from Methods in Enzymology, 2011.

The 6 small molecule inhibitors investigated in this study have gone through 2 periods of test with human KDM4a by Chroma Therapeutics as shown in Figure 3.1 a. We have performed the 3rd round of *in vitro* testing using the *Drosophila* KDM4a enzyme. dKDM4a shares 63 % identity and 80 % similarity with the core catalytic domain of

hKDM4a (see Appendix 2) used in the *in vitro* demethylase assay.

In this study, bioinformatic analysis identified 259 proteins in human and 20 proteins in *D. mel* containing a Jumonji C domain (Figure 3.2 a). 13 of them have been experimentally validated and are expressed in *D. mel* (Klose et al., 2006). The small molecular fragments were selected against hKDM4a core catalytic region Jmj-N and Jmj-C domains to target the binding sites to substrate or cofactors as illustrated in Figure 1.4. The trial performed by Chroma Therapeutic to construct co-crystal structure with hKDM4a and the compounds investigated in this study were not successful. Therefore, the particular binding behavior for each compound is not clear yet.  $K_m$  for hKDM4a reported by  $O_2$  consumption coupled KDM assay was 31  $\mu M$  and  $K_{cat}$  was 104  $min^{-1}$  (Cascella et al., 2012). In general, Jmj-C containing KDMs have low binding affinity to substrate and low turn over rate which indicating the possibility of low binding affinity of the inhibitors as well. This could well be the reason why the hKDM4a did not form co-crystal structure with inhibitors investigated in this study shown in Figure 3.1 b, despite the fact that all the 6 compounds shown inhibition against hKDM4a with low IC50 values.

**Table 4.1 Proposed targeting sites**

Co-factor binding site (Uniprot)	Binding sites																	
	$\alpha$ -KG				Fe(II) Fe(II)				$\alpha$ -KG $\alpha$ -KG				Fe(II)					
	Substrate (Ng et al. 2007)				H3K9me3 H3K9me3 H3K36me3 H3K36me3				H3K9me3 H3K9me3 H3K36me3 H3K36me3				H3K9me3 H3K9me3 H3K36me3 H3K36me3					
Inhibitor	JMJ-3	JMJ-1	JMJ-3	JMJ-1					JMJ-1	JMJ-1		JMJ-2	JMJ-3	JMJ-3	JMJ-2	JMJ-1	JMJ-2	JMJ-3
Functional group	JMJ-5	JMJ-4	JMJ-5	JMJ-4					JMJ-4	JMJ-4		JMJ-6	JMJ-5	JMJ-5	JMJ-6	JMJ-4	JMJ-6	JMJ-5
Amino acid	-C-O-C-	-OH	-C-O-C-	-OH					-OH	-OH		-C=O	-C-O-C-	-C-O-C-	-C=O	-OH	-C=O	-C-O-C-
AA site in hKDM4A	Asn 86	Tyr 132	Asp 135	Thr 167	Glu 169	Gly 170	Tyr 177	His 188	Glu 190	Asp 191	Asn 198	Lys 206	Lys 241	His 276	Ser 288	Asn 290	Arg 309	Asp 311
AA site in dKDM4A	92	139	142		177	184	195	197	198	205	213	248	283	295	291	316	318	321

As shown in Figure 3.1 b, inhibitor MJM-1 and MJM-4 molecular structures containing hydroxyl residue (-OH) categorize themselves to hydrophilic molecules. They may potentially target acidic amino acid residues such as Asp 135, 311, 191 and Glu 169, 190 in the enzyme catalytic pocket shown in the crystal structure (Figure 1.4 and Table 4.1). Inhibitor MJM-2 molecular structure contains thioacyl residue (O=S=O) and carbonyl residue (-C=O). MJM-6 contains carbonyl residue (-C=O). These residues categorize MJM-2 and MJM-6 with strong negative charge. Therefore they may target to positively charged amino acids such as Arg 309, Lys 241, 314. Inhibitor MJM-3 and MJM-5 structures contain ether residue (-C-O-C-). This categorizes themselves with strong polar. They may target to Asn 86, 290, Ser 288, 316 and Thr 167. Among these 13 proposed targeting amino acid residues, 11 aa are identical for hKDM4a and dKDM4a (Table 4.1).

However, relevant report for cellular concentration of  $\alpha$ -ketoglutarate in *E.coli* was in the range of 30 to 145  $\mu$ M (Zimmermann et al., 2014). From authors' point of view, molecules targeting the binding sites for  $\alpha$ -ketoglutarate may need higher dose to inhibit KDM4a enzyme *in vivo*. This is due to the higher concentration of  $\alpha$ -ketoglutarate present in cells than in recombinant enzyme purified from insect expression system. It contains only the  $\alpha$ -ketoglutarate molecule when it is isolated from other component in the cell.

Lysine demethylation reaction produces equal amount of demethylated product and formaldehyde molecule. Therefore demethylation can be measured by either direct product or byproduct formaldehyde. Lysine demethylation also consumes equivalent amount of  $O_2$  and  $\alpha$ -ketoglutarate.  $O_2$  consumption coupled to FDH assay was also developed to measure demethylation reaction (Cascella et al., 2012). In the high throughput assay from fragment library screening against hKDM4a core catalytic domain, demethylation is detected by coupled FDH reaction. This reaction measures the release of formaldehyde. The byproduct formaldehyde is oxidized to formic acid by dehydrogenase. Meanwhile  $NAD^+$  is reduced to NADH which leads to variation of photo absorbance at 340 nm wavelength (Lizcano et al., 2000).  $IC_{50}$  values of hKDM4a have been determined by measuring demethylated product directly detected by LC-MS. In this study, we applied a MALDI-ToF detection method to directly measure product formation. It overcomes the error prone issue of formaldehyde based detection and time consuming issue of LC-MS detection.

Jmj-C domain alone is enough to confer demethylase activity. But the selectivity of KDM4 subfamily towards H3K36me3 and H3K9me3 is conferred by adjacent Jmj-N domain. Jmj-N domain offers supporting structure for catalytic pocket. And the linking region of hKDM4 regulates its localization in the chromatin via miRNA (Zoabi et al., 2014). dKDM4a consists of Jmj-N (aa 17-59), linking region (aa 60-180), Jmj-C (aa 182-298) (Figure 3.2 b). Linking region of Jmj-N and Jmj-C domain for KDM4a is also highly conserved between human and fly. Therefore, the linker region can be ideal targeting region specifically for KDM4 subfamily shown in Figure 3.2 b. One of the binding sites for  $\alpha$ -ketoglutarate, Tyr 132 indicated in Uniprot database is located in this region.

In general, dKDM4a has low catalysis activity. Demethylation *in vitro* assay with enzyme : substrate ratio 1 : 10 keeps in linear phase within 20 minutes and reaches plateau after 2 hours as shown in Figure 3.4 a. Therefore reaction of enzyme : substrate ratio 1 : 10 at 20 minutes activity has been taken as standard setting for inhibitor test. Dosage response curve

of inhibitors JMJ-1, JMJ-4 and JMJ-6 has been shown in Figure 3.4 b, c, d. The closer for titration concentration range to the true IC<sub>50</sub> value, the more accurate for the IC<sub>50</sub> value read out. IC<sub>50</sub> values calculated from Figure 3.4 b, c, d were shown in Figure 3.4 e. Inhibitor JMJ-2, JMJ-3 and JMJ-5 have not shown inhibition when used at a concentration of 500  $\mu$ M with 1  $\mu$ M enzyme, therefore IC<sub>50</sub> have been indicated as > 500  $\mu$ M. Inhibitors JMJ-1 and JMJ-6 have shown similar IC<sub>50</sub> values for both hKDM4a and dKDM4a. Inhibitor JMJ-4 IC<sub>50</sub> value for dKDM4a is higher than hKDM4a.

The binding affinity between histone modification and different categories of binding domains containing proteins (Figure 1.3) falls in  $\mu$ M to mM range. This is considered to be weak binding (Nikolov et al., 2013). A  $K_d$  > 1 mM would hardly qualify for a specific binding. IC<sub>50</sub> values for inhibitor JMJ-1, JMJ-4 and JMJ-6 for both hKDM4a and dKDM4a investigated in this study show coincidence with this range.

In order to validate these inhibitors in cells, immune fluorescence imaging has been performed to confirm the dKDM4a enzyme activity at the single cell level. As shown in Figure 3.5 a, over-expression of dKDM4a was indicated by HA tag staining using an anti-HA antibody. Corresponding H3K36me<sub>3</sub> signal was decreased upon FH-dKDM4a over-expression. This result indicated dKDM4a shows demethylase activity towards H3K36me<sub>3</sub> when over expressed in a cell line. This supports the data obtained by Lloret-Llinares (Lloret-Llinares et al., 2008). The results were validated by western blotting in Figure 3.5 b.

L2-4 cells with no treatment, treated with 1 % DMSO, 100  $\mu$ M JMJ-1, 100  $\mu$ M JMJ-4 or 100  $\mu$ M JMJ-6 individually have been set as one group of experiment. H3K36me<sub>3</sub> substrate for dKDM4a has been detected after 48 hours treatment (Figure 3.6, Figure 3.7 and Figure 3.9). Figure 3.6 c shows an increase of the H3K36me<sub>3</sub> signal upon dKDM4a inhibitor treatment especially for JMJ-6 as detected by western blotting. A treatment of L2-4 cells with 1 % DMSO has a lower H3K36me<sub>3</sub> signal. The results are consistent with Figure 3.6, Figure 3.7 and Figure 3.11 using different detection method. Accumulation of Fe<sup>2+</sup> has been reported upon DMSO treatment (Friend et al., 1971). This can promote demethylase activity of Jumonji KDMs. This may account for the decrease of H3K36me<sub>3</sub> signal we observed with DMSO treatment. In Figure 3.8, 1 % and 0.5 % DMSO treatment to L2-4 cells shown growth delay. 0.1 % DMSO treatment shown no observed growth delay to L2-4 cells. Therefore, DMSO at 1 % and 0.5 % concentration is not an ideal control. Cells treated with 100  $\mu$ M inhibitor JMJ-4 did not grow. It is not clear whether it has toxic effect

to the cells. In comparison, cells treated with inhibitor JMJ-6 shown strongest increase for H3K36me3, but no severe growth delay. Inhibitor JMJ-1 and JMJ-6 treatment shown similar growth behavior with corresponding DMSO.

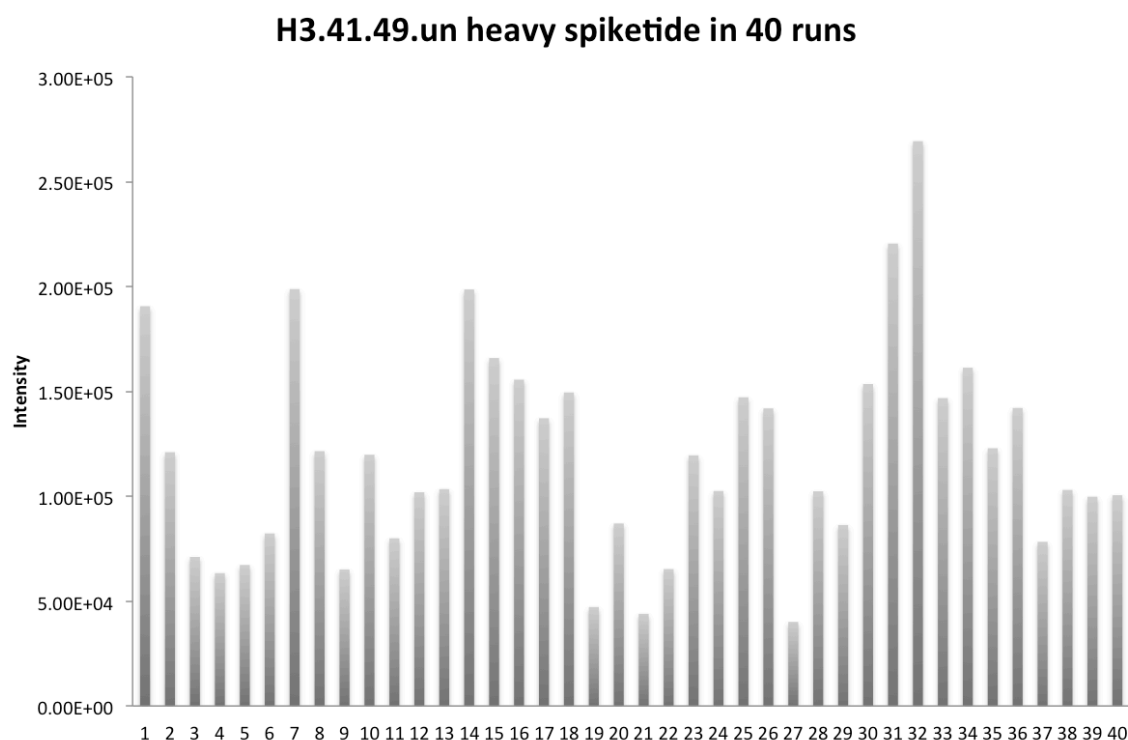
Three methods have been used in this study to measure the inhibitors' effect in cells as shown in Figure 3.9 a. Antibody based substrate detection faces the problem of antibody specificity and sensitivity issue. The detection is particularly limited when PTMs on histones are low abundance. Among all detectable PTMs on histone H3 peptide 27 to 40, H3K36me3 consists of 1.175 % H3K36 sites; H3K36me2 consists of 0.366 %; H3K36me1 consists of 2.05 %; 96.113 % of H3K36 site maintains not modified; H3K36ac consists of 0.278 %. These caused some problems for detecting H3K36me3 by western blotting using H3 protein as loading control. Among all detectable PTMs on histone H3 peptide 9 to 17, H3K9me3 consists of 43.04 %; H3K9me2 consists of 23.7 %; H3K9me1 consists of 4.6 %; 27.78 % of H3K9 site maintains not modified (Feller et al., 2015). Antibody detection detects only one PTM for one site and does not give any information about the PTMs in proximal residues. A targeted proteomics strategy using synthetic spiketides gives the advantage to detect not only one PTM on single site but also combinatorial PTMs on proximity sites. It also overcomes the time consumed to generate a new antibody or the possibility that antibody for certain PTM can not be generated at all. LC-MS/MS sample preparation was illustrated in Figure 3.9 b. Chemical derivatization was applied to protect histone proteins from over-digestion. One of the disadvantages of in gel digestion is low peptides recovery rate from gel. For PTMs lower than 1 % abundance, in solution digestion should be considered. Due to ion suppression effects during ionization for light and heavy peptides with same PTM, heavy spiketides have to be added at a similar concentration as the light peptides. UHPLC gives nice chromatogram peak separation. Targeted precursor ion selection by TT6600 from AB Sciex can sufficiently enrich low abundant PTMs' peptides. Chromatogram peak assignment has been done by observing all the fragmented ions for heavy spiketide shown in Table 3.1 and Figure 3.10 for each PTMs. Precursor ion for H3K36me3 elutes shortly before H3K27me2K36me1 as shown in Figure 3.10 c. Quantitation can be done by either precursor ion (Figure 3.10 c and Figure 3.15) or prognostic fragmented ion such as b3 as shown in Figure 3.10 d and Figure 3.11.

In Figure 3.11, all the experiments from replicate 3 clustered together with low intensity indicating too less protein materials were started with. Low amount of light peptides and high amount of corresponding heavy spiketides lead to mask of signal detection. Clustering



is strongly guided by heavy spiketides. In replicate 2, increases of H3.27.40\_0.3.0 and H3.27.40\_0.2.0 have been validated for JMJ-1, JMJ-4 and JMJ-6 but not H3.9.17\_2.0.0 and H3.9.17\_3.0.0. In replicate 1, increases of H3.27.40\_0.3.0 and H3.27.40\_0.2.0 have been validated for JMJ-1, JMJ-4 and JMJ-6 but not H3.9.17\_2.0.0 and H3.9.17\_3.0.0.

To assess the technical variability of the machine, intensity values of the unmodified spiketide H3.41.49\_un were measured in 40 runs. 200 fmol of this peptide was equally loaded in each run shown in Figure 4.2. The intensity value for the same amount of peptide loaded on the UHPLC column varies up to 7 fold. This indicates that technical variability is still big and signal normalization needs to be applied.



**Figure 4.2 | Running control of unmodified heavy spiketide H3.41.49\_un**

200 fmol spiketide H3.41.49\_un was loaded in the UHPLC column in each run. 40 runs was randomized and run sequentially. Same parameters were applied for data extracting.

## 4.2 Function of KDM4a in human and *Drosophila melanogaster*

Leaky expression without heat shock induction was observed for SL2 that was stably transfected with dKDM4a. The leaky expression of dKDM4a was used to express Flag-KDM4a, anti-Flag IP purification to avoid an overexpression artifact brought by heat shock. Mild IP wash has been performed. Samples were eluted from the beads by boiled rather

than by Flag peptide elution. Therefore enrichment of FH-KDM4a by anti-Flag IP was not very strong as shown in Figure 3.12 a. But it is validated in Figure 3.12 b as strongest protein showing t-test difference in Flag-dKDM4a proteome. In addition, all histone proteins His2A, His2B, His3.3, His4 and His1 have been observed in the protein list enriched by FH-dKDM4a (See appendix Mass Spect list). His2Av was also enriched by FH-dKDM4a.

MSL3 containing Chromo domain which binds to H3K36me3 was validated being enriched in the dKDM4a elution (Figure 3.12 b) and further confirmed by western blotting (Figure 3.12 c). This supports the finding about co-presence of MSL3 and H3K36me3 in actively transcribed X chromosomal gene body (Straub et al., 2013). In addition, DCC component MSL1 was also enriched.

dKDM4a protein contains the binding motif PxVxL to HP1a (Thiru et al., 2004). Enrichment of HP1a by FH-dKDM4a is validated as positive control in Mass Spect protein list and western blotting (Figure 3.12 b and Figure 3.12 d). In addition, there is a higher molecular weight protein band at about 100 kDa observed in the Flag IP enriched FH-dKDM4a proteome detected by anti-HP1a western blotting. It is unclear what this additional protein band is.

Human KDM4a has not only conserved domains for demethylase catalysis activity but also assistant binding domains to bring the demethylase activity to target sites (Figure 1.5). Apart from its demethylation activity of KDM4a, it is also acclaimed recently that KDM4a is localized in heterochromatin region of the genome. This has nothing to do with catalytic activity of demethylation. As it is reported for LYS20 which functions as a homocitrate synthase and meanwhile is also a noncanonical HAT enzyme. It is a critical enzyme in both metabolic pathway and chromatin organization (Scott et al., 2010). This raised up the question whether KDM4a may also be a bifunctional protein. What would be the other role of KDM4a if there is one more?

dKDM4a protein constitution is more similar to protein hKDM4d lacking the targeting domains compared to hKDM4a, hKDM4b and hKDM4c (Lloret et al., 2008). And human KDM4e and KDM4f are considered as pseudogenes for hKDM4d. The encoding region for hKDM4a, hKDM4b and hKDM4c in human genome are 1p34.1, 19p13.3 and 9p24.1 respectively. The encoding region for hKDM4d, hKDM4e, hKDM4f are together in 11q21 which is considered as triplicated retrotransposons of KDM4 family gene (Katoh et al.,

2004). hKDM4a, hKDM4b and hKDM4c are ubiquitously expressed in most of human tissues, while hKDM4d is specifically expressed in testis based on the RNA-seq data (Labbé et al., 2013).

RNAi targeting dKDM4a inhibited cell growth for SL2 and L2-4 cells, we hardly harvested enough material for analysis. This agrees with the finding that hKDM4c knock-down has shown reduced cell proliferation for U2OS and KYSE150 cell line (Cloos et al., 2006).

According to the Drosophila Interactions Database (<http://www.droidb.org>), among dKDM4a interaction network, DF31 and bunch of mir RNA are predicted to interact with dKDM4a. We did not find DF31 enriched by FH-dKDM4a in the list of interactors. This may due to the cell line we used does not contain testis tissue.

In mouse, histone variant H3.3 is deposited into sex chromosomes in meiotic prophase during spermatogenesis (Heijden et al., 2007). This fits to the finding hKDM4d is specifically expressed in testis and demethylase activity shows preference towards H3.3K36me3 variant.

Ubiquitination mediated degradation has been used to regulate the abundance of KDM4 subfamily proteins in cells (Tan et al., 2011; Rechem et al., 2011). hKDM4d localization on chromatin was dependent on RNA binding to N-terminus aa 115 to 236 and C-terminus aa 348 to 523 in U2OS cell line (Zoabi et al., 2014). dKDM4a may follow the same regulation mechanism.

### 4.3 Human disease related genes therapeutic study using fly cell

I used SL2 and L2-4 cell lines, which are derived from embryonic stage of 20-24 hours *Drosophila Melanogaster*. The SL2 cell line displayed many types of morphology: roundish cells, spindle shape cells, macrophage-like cells, epithelial-like cells. The major portion of the cell population is roundish cell with the macrophage-like cells comprised less than 5 %. Roundish cells, spindle shape cells and epithelial-like cells have prominent nucleus and large nucleolus with normally diploid chromosomes (Schneider, 1972). The heterogeneity of SL2 cell population may account for bias of H3K36me3 bulk analysis.

*Drosophila* cell lines are ideal tools to study gene function because of a) the simple and easy way to establish stable transgenic cell lines and b) the possibility to effectively knock down target genes by RNA interference. *Drosophila* also has a lower genetic redundancy

compared to mammals making mutation likely more penetrant. With mutant fly background, defined mutant cell line with more homogeneous population can be acquired compared with stable transformation. Finally, large scale of material can be produced by large scale cultures (Baum & Cherbas, 2007).

The observation of increasing H3K36me3 signal by WB and IF upon treatment of L2-4 cells with compounds CHR-1, CHR-4 and CHR-6 indicated that these three compounds successfully penetrated into the cell and reached dKDM4a in the nucleus.

One of the major challenges of potential epigenetic therapies is the specificity to the targets (Kelly et al., 2010). The small molecules investigated in this study were designed to target KDM4a. But they may also inhibit other Jumonji containing KDMs. Therefore the other histone lysine methylations should also be investigated in order to exclude side effect of these inhibitors.

#### **4.4 Mass spectrometry application**

Conventional biochemical analysis detection method such as western blotting or immune fluorescence imaging relies on specific antibody raised upon immune reaction for the antigen of interest. Usually, only a single protein or PTM is detected in each experiment. Mass spectrometry can potentially detect multiple signals by single run. However, a major challenge of MS for biomolecules is the transfer of the intact molecule into the gas phase during ionization process. Two major ionization techniques have been developed, matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) (Karas & Hillenkamp, 1988; Fenn et al., 1989).

The molecular mass range for analysis by MALDI-ToF is from 100 to 200000 Da. This type of mass spectrometers provides resolution from 400 to 50000 and measurement accuracy of 0.005 % to 0.2 % (Bonaldi et al., 2004). Since predominantly singly charged ions are produced by MALDI ionization, it is possible to analyze low and medium complexity mixtures of sample with MALDI-ToF. It fits well to detect alteration of substrate after enzyme catalysis with mass shift such as methylation, demethylation, acetylation, deacetylation, phosphorylation and dephosphorylation and so on. In this application, the mass of expected products is known and formation of the product can be measured directly.

During ESI ionization process, peptides are protonated to carry different state of charge after ionization. The charge state of the protonated peptide ions can be determined from the isotopic distribution pattern. Difference of 1 mass unit between the two isotopic peaks indicates the charge state being 1+; difference of 0.5 mass unit indicates the charge state being 2+; difference of 0.3 mass unit indicates the charge state being 3+ (Trauger et al., 2002).

Proteomics is a useful tool to study systems biology. Development of ion source ESI makes it possible to analyze complex samples separated by LC directly online coupled to MS such as Orbitrap, Triple ToF or Q exactive. Intensity based absolute quantification (iBAQ) using Adromeda search engine gives the possibility to do statistical analysis for identifying and quantifying complex proteome.

Normally for biomedical or biological researches, the expected output from MS-based measurements is the quantitative difference for targeted proteins between samples. However, there are also cases which require absolute quantification of the protein of interest. (Picotti et al., 2012) To achieve absolute quantification with precision, one possible strategy is to apply AQUA peptides as internal standard. AQUA peptides are isotopically labeled target peptides. They have the identical chemical features as the native peptides. In this study, spiketides have been applied as function of AQUA peptides.

In order to study the stoichiometry of macromolecular complex such as the DCC components enriched by dKDM4a (Figure 3.12 and protein list in Appendix 3), QconCATs strategy can be applied as function of AQUA peptides. QconCATs peptides can be used to quantify heterologous expression of targeted protein with isotopic labeled internal standard. By adding a known amount of QconCATs to the native sample, which serves as internal reference, target proteins can be quantified accordingly. The drawback is one may get differential digestion efficiency due to the context of the sequence where QconCATs peptides derived from. (Picotti et al., 2012)

Targeted proteomics by selecting specific precursor ions of interest has been conferred with the name of next generation of mass spectrometry. It robustly selects precursor ions of interest by quadrupole mass spectrometer via Multiple Reaction Monitoring (MRM). In TToF TT6600, Time of Flight mass analyzer is integrated after CID fragmentation cell which can acquire full spectra of MS/MS ions and precursor ions. MS/MS spectra chromatogram can be extracted via computational tool. Quantitative analysis can be done with MS or MS/MS signal.

## 4.5 H3K36me3 quantitation and its biological function

### 4.5.1 Quantitation of H3K36me3 methylation by WB and IF

H3K36me3 is a highly dynamic modification associated with actively transcribed gene body upon onset of transcription elongation. Therefore H3K36me3 deposition is dependent on active transcription in a temporal and spatial manner. Since each individual cell and cell type can define the expression profile in general. H3K36me3 level can differ from single cell.

Based on the *in vitro* inhibition assay with the inhibitors and the inhibitors treatment in cell data set I have presented here (Figure 3.4, Figure 3.6, Figure 3.7 and Figure 3.11), I acclaim that H3K36me3 signal was increased upon inhibition of dKDM4a enzymatic activity.

Bulk scale analysis approaches such as western blotting can easily mask the active transcription loci for H3K36me3 response. Because it averages H3K36me3 signal to whole chromatin region and whole cell population. It easily gives a pseudo output of not showing any effect, despite the fact of enlarging the native signal by orders of magnitude combining primary antibody, secondary antibody and HRP-ECL detection. The fact of observing several fold enrichment from WB detection reflects the alteration from active transcripts signal averaged by whole chromatin region and whole cell population. So the several fold signal enrichment from WB could be concentrated at active transcription loci for H3K36me3 response up to thousands fold alteration.

This signal masking issue can be partially addressed by immune fluorescent imaging via microscopy detection where the H3K36me3 signal can be measured in a single cell. Therefore the signal masking by heterogenous cell population from bulk analysis is eliminated. But the H3K36me3 signal in IF we observed is still after magnification of several orders of magnitude. The disadvantages coming with IF staining and microscopy detection are the cross-reactions between different secondary antibodies conjugated to fluorophore and signal overlapping. This is due to the close wave length the chosen fluorophores needed for excitation and emission.

### 4.5.2 Quantitation of H3K36me3 methylation by targeted LC-MS/MS analysis

According to Feller et al., single modification H3K36me3 contributes to approximately 1.10 % of the total histone H3 protein. It functions not only as single mark but also in combination with other proximal modification on lysine residues (Feller et al., 2015).

Analysis of low abundant PTM is challenging due to the detection limit of the mass spectrometer. Application of targeted proteomics for ion transition with defined  $m/z$  value enriches precursor ions of interest. Due to ion suppression between light naturally present peptide ions and isotopically labeled spiketide ions, amount of spiketides included must take consideration of naturally occurred stoichiometry for corresponding PTM peptide signal. Here I applied 200 fmol of heavy spiketide for each PTM signal in the peptide library, this led to ion suppression during ionization especially for the low abundant PTMs. 1 to 1 ratio for heavy and light peptide ions with same modification is suggested. Normalization should be included not only for material loading on the LC column such as using unmodified peptide H3.41.49\_un but also for MS/MS fragmentation event. Decision of choosing MS/MS ions to do quantitation is based on: a) prognostic feature using a particular fragmented ion for the precursor ion; b) relative abundance of the chosen MS/MS ion among all fragmented ions. Regarding normalization, histone protein band intensity can also be normalized by coomassie blue staining.

#### 4.5.3 Biological function of H3K36me3

H3K36me3 is an active mark which is found in the 3' region of the gene encoding body and the 5' exons of the active transcription (Bell et al., 2007). Inhibition of demethylases targeting H3K36me3 could induce the accumulation of this mark thereby stimulating expression of the underlying gene. 1.10 % of total histone H3 protein showing H3K36me3 may indicate there is 1.10 % of the genome being actively transcribed at a given time point for SL2 fly cell line. Active transcripts consist of 5.99 % constitutive exons, 65.50 % alternatively spliced exons and 28.51 % alternative skipped exons according to the clustering of exons by Mayer et al. in human cell line (Mayer et al., 2015). It might be that the alternative spliced exons accounting for the most variation of H3K36me3 modification. Therefore, the variation of H3K36me3 could be highly dependent on the cell type under investigation and the homogeneity of the cell population.

During cell division, genetic information must be faithfully transmitted from parental cell to the two daughter cells precisely. The whole genome has to be de-condensed and replicated in order to get the entire cell duplicated. Meanwhile histones also need to be duplicated with high fidelity of PTMs on them. In SL2 and L2-4 fly cells, bi-nucleated cells are frequently detected (Figure 3.13). The origin of these cells is still unclear but interestingly the two nuclei show a dramatically different H3K36me3 staining pattern. If

these cells are resolved into two individual cells, the difference in H3K36me3 would result in two very different epigenetic landscapes similar to what is observed in asymmetric divisions during stem cell differentiation or the development of tumor cells.

#### 4.6 Metabolomics and epigenetics

Mutations of metabolic enzymes such as fumarate hydratase and succinate dehydrogenase deregulate the level of fumarate and succinate. These mutations have shown to inhibit histone demethylation via Jmj-C containing KDMs (Xiao et al., 2012). In this study, we observed a decrease of H3K36me3 in nuclei of cells over-expressing IDH (IDH1-NADP) in cytoplasm (Figure 3.14 a). Over-expression of IDH generates more  $\alpha$ -ketoglutarate, we assume that all Jmj-C domain containing proteins which use  $\alpha$ -ketoglutarate as cofactors could potentially get influenced by it. Therefore we have analyzed all abundant histone methylations upon over-expression of IDH (Figure 3.15). These analyses show that most of the methylation signals on histones are reduced upon IDH over-expression. Considering the clustering is strongly guided by missing value in this data set, further analysis need to be done in order to make conclusion.

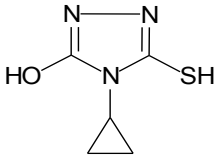
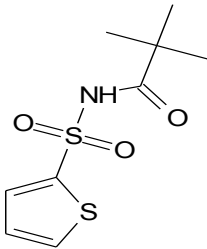
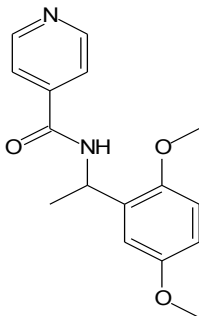
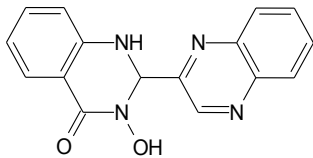
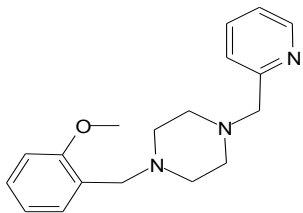
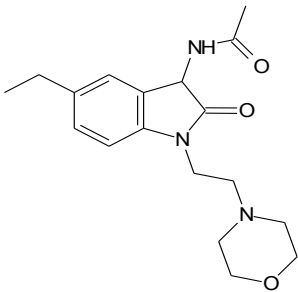
In order to validate the correlation of IDH enzyme activity with the decrease effect on histone methylation, I have generated stable cell lines over-expressing IDH carrying various mutations. Molecular analysis of *Drosophila* IDH mutants are shown in Table 3.2 and validation of stable cell line over-expression are shown in Figure 3.16. IDH enzyme null mutants from human have been taken as reference to generate mutants for dIDH (Yang et al., 2010).



## 5. Material and methods

### 5.1 Material

#### 5.1.1 Inhibitors investigated in this study

Structure	MW: g/mol	Name [Chemical name]
	157.2	<b>JMJ-1</b>  [4-cyclopropyl-5-mercapto-4H-1,2,4-triazol-3-ol]
	247.3	<b>JMJ-2</b>  [N-(thiophen-2-ylsulfonyl)pivalamide]
	286.3	<b>JMJ-3</b>  [N-(1-(2,5-dimethoxyphenyl)ethyl)isonicotinamide]
	292.3	<b>JMJ-4</b>  [4(1H)-Quinazolinone, 2,3-dihydro-3-hydroxy-2-(2-quinoxaliny)-]
	297	<b>JMJ-5</b>  [1-(2-methoxybenzyl)-4-(pyridin-2-ylmethyl)piperazine]
	331.4	<b>JMJ-6</b>  [N-(5-ethyl-1-(2-morpholinoethyl)-2-oxoindolin-3-yl)acetamide]

### 5.1.2 Chemicals

All common chemicals were purchased from Pharmacia, E. Merck, Pierce, Promega, Roche, Roth, Sigma.

Special requirement for this work:

Name	Source
Ferrous Sulfate Heptahydrate	Sigma-Aldrich
$\alpha$ -Ketoglutaric Acid Sodium Salt	Sigma-Aldrich
Sodium ascorbate	Sigma-Aldrich
2-Hydroxyglutaric Acid Disodium Salt	Santa Cruz Biotech
Propionic anhydride	VWR
Dithiothreitol	Roth
Iodoacetamide	Roth
Image-iT FX signal enhancer	Invitrogen
Trifluoroacetic acid	VWR
Acetonitril	Roth
Vectashield mounting medium H-1000	Vector Labs
Normal goat serum	Dianova
$\alpha$ -cyano-4-hydroxy-cinnamic acid	Sigma

### 5.1.3 DNA and Protein Markers

Name	Source
GeneRuler 1kb DNA ladder	New England BioLabs
123bp DNA ladder	Invitrogen
100bp DNA ladder	New England BioLabs
peqGold Prestained Protein Marker IV	peQlab
peqGold Prestained Protein Marker V	peQlab

### 5.1.4 Enzymes and Kits

Product	Company
Taq DNA Polymerase	VWR
Pfu Turbo DNA Polymerase	Agilent
Restriction endonucleases	NEB
Gel extraction Kit	Qiagen
Miniprep Kit	Qiagen
RNeasy Mini Kit	Qiagen
Maxiprep Kit	Qiagen
Clarity <sup>TM</sup> Western ECL Substrate Kit	BIO-RAD
ECL <sup>TM</sup> Prime Western Blotting Detection Reagent	GE Healthcare
ECL Kit (enhanced chemoluminescence)	GE Healthcare
QuikChange <sup>TM</sup> Site-Directed Mutagenesis Kit	Stratagene
T7 MEGAscript <sup>®</sup> Kit	Ambion
Trypsin Gold, Mass Spectrometry Grade	Promega
Bio-Rad Protein Assay	BIO-RAD

### 5.1.5 Primers

	Oligo name	Sequence	Description
1	Jmjd2a-1fw	5'-AAAAAGCAGGCTCCGCCATGTCCACGAGATCTTCAT-3'	pDONR/zeo
2	Jmjd2a-2rev	5'-AGAAAGCTGGGTCTCAATCCTCGTCGTCAAGTG-3'	pDONR/zeo
3	KDM4A-1020rev	5'-AGAAAGCTGGGTCTCACAGCCAATTATCGTATC-3'	pDONR/zeo
4	attB1-fw	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT-3'	pDONR/zeo
5	attB2-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGT-3'	pDONR/zeo
6	FP312	5'- TTAATACGACTCACTATAGGGAGACAATGGATGTGAAC GAAACG-3'	RNA interference
7	RP312	5'- TTAATACGACTCACTATAGGGAGATCCTCGTCGTCAAG TGTGAG-3'	
8	FP400	5'- TTAATACGACTCACTATAGGGAGAAACTCCCAACCATT GCGTCT-3'	RNA interference
9	RP400	5'- TTAATACGACTCACTATAGGGAGACATATTTGTTTGCA CGAATT-3'	
10	GST-FP	5'- TTAATACGACTCACTATAGGGAGAAGTTTGAATTGGGT TTGGAGTTTCC-3'	RNA interference control
11	GST-RP	5'- TTAATACGACTCACTATAGGGAGATCGCCACCACCAAA CGTGG-3'	
12	KDM4A-735	5'-TCGCCATAAGATGACCATGA-3'	For q-PCR
13	KDM4A-843	5'-GCCGAAGGGAAATGTAATCA-3'	
14	KDM4A-34	5'-CAGAACAAAGTGCCCCGTAT-3'	For q-PCR
15	KDM4A-135	5'-TAAGTGTGCTCCCCGAGACT-3'	
16	M13-FP [uni_21]	5'-TGTAACGACGCGCCAGT-3'	For sequencing
17	M13-RP [rev_29]	5'-CAGGAAACAGCTATGACC-3'	
18	T7-FP	5'-TAATACGACTCACTATAGGG-3'	For sequencing
19	T7-RP	5'-CTAGTTATTGCTCAGCGGT-3'	
20	ON130pMT fw	5'-CATCAGTTGTGGTCAGCAGC-3'	For sequencing
21	ON152pMT rev	5'-CAATCCTAAACCCATTTGC-3'	For sequencing
22	pMK33-CFH-HAstop	5'-CTAAGCAGCAGCGTAATCTG-3'	For sequencing
23	CG7176_RA 385-409 R132H fw	5'-GTGATCGGTCACCATGCCCACGCC-3'	For R132H point mutation
24	CG7176_RA 385-409 R132H rew	5'-GGCGTGGGCATGGTGACCGATCACAATAG-3'	For R132H point mutation

25	CG7176_RA 218-244 T77A fw	5'-GCGCCACAATCGCTCCCGACGAGAAGC-3'	For T77A point mutation
26	CG7176_RA 218-244 T77A rew	5'-GCTTCTCGTCGGGAGCGATTGTGGCG-3'	For T77A point mutation
27	CG7176_RA 270-294 S94A fw	5'-GATGTGGAAGGCGCCCAACGGTACC-3'	For S94A point mutation
28	CG7176_RA 270-294 S94A rew	5'-GGTACCGTTGGGCGCCTTCCACATC-3'	For S94A point mutation
29	CG7176-762-rew	5'-GCGTCCATCGTACTTCTTCA-3'	For sequencing
30	CG7176-390-fw	5'-CGGTACCATCCGTAACATCT-3'	For sequencing
31	CG6439-934-rew	5'-TGGGATTAGCCACGTTCTTG-3'	For sequencing
32	CG6439-511-fw	5'-CTAAGCAGCAGCGTAATCTG-3'	For sequencing
33	ON231 GAPDH1 RT_fw	5'-GTGACCTACGCAGAAAGCTAG-3'	For q-PCR
34	ON232 GAPDH1 RT_rev	5'-GCTATTACGACTGCCGCTTTTTC-3'	
35	ON233 tub97E_RT_fw	5'-GAGCAAGAACAGCAGCTACTTTGT-3'	For q-PCR
36	ON234 tub97E_RT_rev	5'-CACCTTGACGTTGTTGGGAAT-3'	

All the primers were synthesized by Eurofins MWG.

All the sequencing was done by GATC.

### 5.1.6 Buffers

#### 5.1.6.1 Buffer composition for DNA samples

Orange G 5 × DNA loading dye: 0.3 % (w/v) Orange G

Ethidiumbromid staining DNA with agarose gel:

Stock solution of EtBr: 10 mg/ml of ethidiumbromid (MW: 394.294 g/mol) in water. Use 1 to 100 dilution in agarose gel.

0.8 % to 1.5 % (w/v) of agarose gel was used to separate DNA fragment.

#### 5.1.6.2 Buffer composition for protein samples

6 × Laemmli buffer: 300 mM Tris pH6.8; 12 % SDS; 60 % Glycerol; 1.2 % Bromphenolblue, 5 % β-mercaptoethanol, store in -20 °C.

Coomassie G250 stainig buffer for SDS PAGE gel:

Working solution: 0.25 % Coomassie G250, 50 % Methanol, 10 % Acetic Acid

### 5.1.7 Vectors

Vector name	Promotor	Tag	Expression system	Bacteria resistance	Selection
pOT2	T7		Bacterial	Chloramphenicol	
pNIC28-Bsa4	T7	N-HIS6	Bacterial	Kanamycin	
pDONR/ZEO	T7		Bacterial	Chloramphenicol	Zeomycin
pDEST17	T7	N-HIS6	E.coli	Ampicillin	
pDEST10-OneStrep	Polyhedrin	C-ONE-Strep/8His/TEV	Baculoviral	Ampicillin	Gentamycin
pHFHW	Hsp70	N-3×Flag-3×HA	Insect	Ampicillin	Puromycin
pMK33-C-TAP-Flag-HA-BD	Metallothionein	C-Flag-HA	SL2/L2-4	Ampicillin	Hygromycin
pGEX-6P-1	T7	GST	Bacterial	Ampicillin	

### 5.1.8 Plasmid

Gene name	Vector	Description
hJMJD2a_1-359	pNIC28-Bsa4	Bacterial expression, purify over Nickel column
dKDM4a	pOT2	Commercial plasmid containing coding sequence
dKDM4a	pDONR/ZEO	For Gateway entry cloning
dKDM4a	pDEST17	Bacterial expression, purify over Nickel column
dKDM4a	pDEST10-OneStrep	Insect expression, purify over Strep-tactin sepharose beads
dKDM4a	pHFHW	For stable cell line over-expressing dKDM4a
dIDH1_wt	pMK33-C-TAP-Flag-HA-BD	For stable cell line over-expressing dIDH1_wt
dIDH1_111	pMK33-C-TAP-Flag-HA-BD	For stable cell line over-expressing dIDH1_111
dIDH1_128	pMK33-C-TAP-Flag-HA-BD	For stable cell line over-expressing dIDH1_128
dIDH1_166	pMK33-C-TAP-Flag-HA-BD	For stable cell line over-expressing dIDH1_166
dIDH1_111_128	pMK33-C-TAP-Flag-HA-BD	For stable cell line over-expressing dIDH1_111_128

### 5.1.9 Bacteria strains *E. coli*

Bacteria strain	Infor.
DH5α from Life Technologies	F-, lacI-, recA1, endA1, hsdR17, (lacZYA-argF), U169, F80dlacZM15, supE44, thi-1,

Rosetta (DE3) pLysS from Novagen	gyrA96, relA1 (Hanahan D, 1985) F- ompT hsdSB(rB-mB-) gal dcm (DE3) pLysSRARE (argU+, argW+, cam+, ileX+, glyT+, leuW+, proL+)
----------------------------------	---

#### 5.1.10 Cell culture material

Cell line	Origin	Infor.	Culture Medium
SL2	<i>Drosophila melanogaster</i>	Semi-adherent, Doubling time 93 hrs, Oregon R embryos, 20 to 24 h	Schneider's + 10 % FCS, Penicillin, final concentration at 100 units/ml Streptomycin, final concentration at 100 µg/ml
L2-4	<i>Drosophila melanogaster</i>	Sub-clone of SL2	Schneider's + 10 % FCS, Penicillin, final concentration at 100 units/ml Streptomycin, final concentration at 100 µg/ml
Sf21			Sf-900 II SFM containing 10 % FCS and 100 µg/ml Gentamycin final concentration

Item	Source
Trypan blue	Sigma
Hemocytometers	Bright-Line Reichert
Plastic-wares :	Greiner Bio-One GmbH
6-well plate, 25-cm <sup>2</sup> T flask, 75-cm <sup>2</sup> T flask, 150-cm <sup>2</sup> T flask	
Cryovial	Roth
Isopropanol cell freezing container	NALGENE™
Incubator maintaining 26 °C	LMS Cooled Incubator
Laminar flow hood	Thermo Scientific HERA Safe
Inverted compound microscope	Leica
X-tremeGENE HP DNA Transfection Reagent	Roth
Hygromycin B	PAA
Puromycin	PAA
Penicillin	GIBCO
Streptomycin	GIBCO
Schneider's Drosophila Medium	GIBCO
Fetal Calf Serum	Sigma

#### 5.1.11 Chromatography material

Beads/ Column	Source
Protino® Ni-NTA Column 1 ml	MACHEREY-NAGEL
Gelfiltration column Superose 6	Amersham
Gelfiltration column Superdex 200	Amersham
Gluthathione-Sepharose-4B	VWR

Anti-Flag M2 Affinity Gel	Sigma
MonoQ™ 5/50 GL	Amersham
Strep-Tactin Sepharose	IBA

### 5.1.12 Antibodies

Primary antibodies	Resource	Dilution for western blotting	Secondary antibodies	2 <sup>nd</sup> Ab Dilution for ECL
$\alpha$ -H3K4me3	Diagenode	1:1000	$\alpha$ -rabbit	1:5000
$\alpha$ -H3K36me3	Abcam	1:2000	$\alpha$ -rabbit	1:5000
$\alpha$ -H3K36me2		1:5000	$\alpha$ -rabbit	1:5000
$\alpha$ -H3K9me3	Diagenode	1:1000	$\alpha$ -rabbit	1:5000
$\alpha$ -H3K9me2	Millipore		$\alpha$ -rabbit	1:5000
$\alpha$ -H3K9me2	ActiveMotif		$\alpha$ -rabbit	1:5000
$\alpha$ -H3	Abcam	1:1000	$\alpha$ -rabbit	1:5000
$\alpha$ -H2B 101-11		1:50	$\alpha$ -rat	1:5000
$\alpha$ -Flag	Sigma	1:5000	$\alpha$ -mouse	1:5000
$\alpha$ -HA-Biotin	E.Kremmer	1:3000	$\alpha$ -rat	1:5000
A058				
$\alpha$ -HA R001 3F10	Rothe	1:50	$\alpha$ -rat	1:5000
$\alpha$ -Tubulin	Sigma	1:5000	$\alpha$ -mouse	1:5000
$\alpha$ -HP1 A049		1:1000	$\alpha$ -rabbit	1:5000
$\alpha$ -HP6 (umbrea)		1:500	$\alpha$ -rabbit	1:5000
$\alpha$ -MOF	P. Becker	1:2000	$\alpha$ -rabbit	1:1000
$\alpha$ -MSL1	P. Becker	1:1000	$\alpha$ -rabbit	1:1000
$\alpha$ -MSL2	P. Becker	1:500	$\alpha$ -rat	1:2000
$\alpha$ -MSL3 1C9-5	P. Becker	1:50	$\alpha$ -rat	1:2000
$\alpha$ -MSL3 Rb74	P. Becker	1:1000	$\alpha$ -rabbit	1:2000
$\alpha$ -H4K16ac	Santa Cruz	1:500	$\alpha$ -rabbit	
$\alpha$ -streptavidin-HRP	Biolegend	1:2000		
$\alpha$ -mouse-HRP	Amersham	1:10000		
$\alpha$ -rabbit-HRP	Amersham	1:10000		
$\alpha$ -rat-HRP	Amersham	1:10000		
$\alpha$ -mouse-IRDye 800	Biomol	1:10000		
$\alpha$ -rabbit-IRDye 800	Biomol	1:10000		
$\alpha$ -rat-IRDye 800	Biomol	1:10000		

### 5.1.13 Synthetic peptides

For peptide H3.27.40

Peptide sequence	Modification K27_K36_K37	ID
STGGV-Lys(Me3)-KPHRY	- 3 -	Peptide 31-41 H3.3

ARKSAPATGGVKKPH-R*--Qtag	- - -	Peptide 25-40
ARKSAPSTGGVKKPH-R*--Qtag	- - -	Peptide 25-40 H3.3
ARLys(Me)-SAPATGGVKKPH-R*--Qtag	1 - -	Peptide 25-40
ARKSAPATGGV-Lys(Me)-KPH-R*--Qtag	- 1 -	Peptide 25-40
ARKSAPATGGVK-Lys(Me)-PH-R*--Qtag	- - 1	Peptide 25-40
ARLys(Me2)-SAPATGGVKKPH-R*--Qtag	2 - -	Peptide 25-40
ARKSAPATGGV-Lys(Me2)-KPH-R*--Qtag	- 2 -	Peptide 25-40
ARKSAPATGGVK-Lys(Me2)-PH-R*--Qtag	- - 2	Peptide 25-40
ARLys(Me3)-SAPATGGVKKPH-R*--Qtag	3 - -	Peptide 25-40
ARKSAPATGGV-Lys(Me3)-KPH-R*--Qtag	- 3 -	Peptide 25-40
ARKSAPATGGVK-Lys(Me3)-PH-R*--Qtag	- - 3	Peptide 25-40
ARLys(Ac)-SAPATGGVKKPH-R*--Qtag	a - -	Peptide 25-40
ARKSAPATGGV-Lys(ac)-KPH-R*--Qtag	- a -	Peptide 25-40
ARKSAPATGGVK-Lys(ac)-PH-R*--Qtag	- - a	Peptide 25-40
ARLys(Me)-SAPATGGVK-Lys(Me)-PH-R*--Qtag	1 - 1	Peptide 25-40
ARLys(Me)-SAPATGGV-Lys(Me)-KPHR*--Qtag	1 1 -	Peptide 25-40
ARLys(Me)-SAPATGGV-Lys(Me2)-KPHR*--Qtag	1 2 -	Peptide 25-40
ARLys(Me2)-SAPATGGV-Lys(Me)-KPH-R*--Qtag	2 1 -	Peptide 25-40
ARLys(Me2)-SAPATGGV-Lys(ac)-KPH-R*--Qtag	2 a -	Peptide 25-40
ARLys(ac)-SAPATGGV-Lys(Me2)-KPH-R*--Qtag	a 2 -	Peptide 25-40
ARLys(ac)-SAPATGGV-Lys(Me3)-KPH-R*--Qtag	a 3 -	Peptide 25-40

Synthetic peptide for histone protein H3.3 amino acid residues 31\_41 was used as substrate for HDM assay *in vitro*.

Synthetic peptides for histone protein H3 amino acid residues 25\_40 with an isotopically labeled arginine at the C-terminus were used as spike tides to identify the correct retention time for corresponding natural peptides and as internal standard for relative quantitation of histone modifications.

Annotation: “-”: unmodified residue; “1”: lysine mono-methylation; “2”: lysine di-methylation; “3”: lysine tri-methylation; “a”: lysine acetylation.

For peptide H3.9.17

Peptide sequence	Modification K9_S10_K14	ID
ARKSTGGKAP-R*--Qtag	- - -	Peptide 7-17
AR-Lys(Me)-STGGKAP-R*--Qtag	1 - -	Peptide 7-17
AR-Lys(Me2)-STGGKAP-R*--Qtag	2 - -	Peptide 7-17
AR-Lys(Me3)-STGGKAP-R*--Qtag	3 - -	Peptide 7-17
ARKSTGG-Lys(Ac)-AP-R*--Qtag	- - a	Peptide 7-17
ARK-pS-TGGKAP-R*--Qtag	- p -	Peptide 7-17
ARK-pS-TGG-Lys(Ac)-AP-R*--Qtag	- p a	Peptide 7-17
AR-Lys(Ac)-STGGKAP-R*--Qtag	a - -	Peptide 7-17
AR-Lys(Ac)-STGG-Lys(Ac)-AP-R*--Qtag	a - a	Peptide 7-17



AR-Lys(Me)-STGG-Lys(ac)-AP-R*-Qtag	1 - a	Peptide 7-17
AR-Lys(Me <sub>2</sub> )-STGG-Lys(ac)-AP-R*-Qtag	2 - a	Peptide 7-17
AR-Lys(Me <sub>3</sub> )-STGG-Lys(ac)-AP-R*-Qtag	3 - a	Peptide 7-17

Synthetic peptides for histone protein H3 amino acid residues 7\_17 with isotopically labeled arginine at the C-terminus were used as spike tides to identify correct retention time for corresponding natural peptides and also as internal standard for relative quantitation of histone modifications.

Annotation: “-”: unmodified residue; “1”: lysine mono-methylation; “2”: lysine dimethylation; “3”: lysine trimethylation; “a”: lysine acetylation; “p”: serine phosphorylation.

For peptide H3.41.49

Peptide sequence	ID
RYPGTVLR*-Qtag	No modification Peptide 40-49

Peptide H3.41.49 was used as loading control of H3 protein amount.

## 5.2 Methods

### 5.2.1 General DNA and RNA sample methods

#### **PCR setup and cycle**

10 × reaction buffer, 50 ng of dsDNA template, 2.5 pmol/μl forward primer, 2.5 pmol/μl reverse primer, 200 μM of each dNTP, 0.05 U/μl of DNA polymerase, fill up to final volume with ddH<sub>2</sub>O.

95 °C 4 min, 25 cycle of [94 °C 30 sec, 62 °C 30 sec, 72 °C 1 min/kb of DNA fragment], 72 °C 10 min, 4 °C 5 min.

#### **Restriction enzyme digestion**

1 μg of DNA template, 10 × reaction buffer, 1 U of restriction enzyme, fill up to final volume with ddH<sub>2</sub>O. Digestion was performed at 37 °C for 3 hours and product was analyzed by running agarose gel electrophoresis.

#### **Single point mutation generation**

Assign target mutation site and corresponding DNA site. Mutagenic primers were designed to make sure both of primer should contain the mutation. In the case of having target gene ORF in a plasmid, single point mutation was generated by QuikChange™ Site-Directed Mutagenesis Kit from Stratagene according to its manuscript. Plasmid after including target mutation site was sequenced to confirm mutagenesis.

#### **Transformation of *DH5α* competent cells**

No less than 200 ng of plasmid was transformed to 100 μl of *DH5α* competent cells. The mixture was incubated on ice for 30 min. Heat shock at 42 °C for 90 sec was performed. The cells were stabilized on ice for 5 min. Add 500 μl of LB and shake at 37 °C, 350 rpm for 1 h. The cells were plated on LB plate with antibiotic for selection. After incubation at 37 °C overnight, check the single colony the next day.

Plasmid amplified from bacteria *E. coli* was extracted by Miniprep Kit from Qiagen according to its handbook. Concentration of the plasmid was determined by NanoDrop ND-1000 Spectrophotometer. DNA sequencing and oligo nucleotide synthesis were done by the company Eurofins Genomics in this study if not mentioned else particularly.

### 5.2.2 Cell culture

For fly cell line SL2 and L2-4 maintenance, Schneider's media was complemented with 10 % heat inactivated fetal calf serum and 100 U/ml Penicillin-G and 100 µg/ml streptomycin. Cells were split every 3 or 4 days down to 20 % of confluence.

To prepare a frozen cell stock, a fully confluent T 75 flask was used. Cells were detached from the flask by pipetting, centrifuged at 1200 rpm for 5 min, resuspended in freezing medium containing 10 % DMSO, 40 % Schneider Media and 50 % heat inactivated FCS. Resuspended cells were frozen in aliquots of 1 ml in cryo vial.

To thaw frozen cell stocks, cells were quickly warmed up to thaw and transferred to 10 ml of fresh complete medium, centrifuged at 1200 rpm for 5 min, resuspended in fresh complete medium.

### **Loss of function and gain of function experiment**

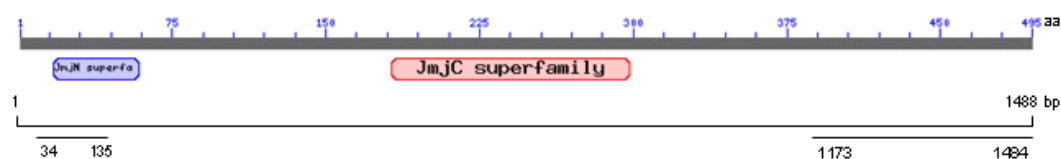
#### **Knockdown procedure**

A plasmid containing gene ORF of interest was taken as PCR template. Primers were applied to include T7 promoter to flank both side of linear DNA product. PCR was performed by using Taq DNA polymerase and cycling at 95 °C 4 min, 25 cycle of [94 °C 30 sec, 62 °C 30 sec, 72 °C 30 sec], 72 °C 10 min. Concentration measurement and quality control of PCR product were done by NanoDrop 2000 spectrophotometer. PCR product was run in 1 % agarose gel to check the fragment size and later on purified by QIAquick Gel Extraction Kit according to its manuscript. Amplicon of GST was used as negative control for knock down.

*In vitro* transcription was done by MEGAscript T7 Kit from Ambion according to its manuscript using 1 µg of DNA template and yield should be around 100 µg per reaction. Concentration measurement and quality control of dsRNA was done by NanoDrop ND-1000 spectrophotometer.

L2-4 cells in log growth phase were used. Culture medium was changed from containing FCS to without FCS but with Penicillin and streptomycin. 5 ml of cells were seeded at density of 2.0 million cells per ml in T75 flask. Adding 50 µg of dsRNA, the mixture was gently shaken for 10 min at room temperature. Then incubation was done for 50 min at 26 °C, followed by adding 10 ml culture medium containing FCS with Penicillin and streptomycin. After incubation for 7 days, cells were harvested and subjected to extract the total RNA according to manuscript of RNeasy<sup>®</sup> Mini. First-strand cDNA was synthesized

by SuperScript<sup>®</sup> III First-Strand Synthesis System for RT-PCR Kit. qPCR was performed using Fast SYBR<sup>®</sup> Green Kit by Light Cycler 480.



### Establishment of stable cell line over-expressing gene of interest

#### Figure 5.1 | Scheme of RNAi targeting region for dKDM4a

dsRNA was transcribed in vitro from dKDM4a coding sequence region 1173 to 1484 nt. qPCR primers for amplification were designed against dKDM4a coding sequence region 34 to 135 nt.

Low passage cells were seeded at 2.5 million cells in 3 ml complete culture medium in 6-well plate one day before transfection. In this work all the transfection were done by X-tremeGENE HP DNA Transfection Reagent according to its product protocol for insect cells. 2 µg of plasmid DNA in diluent was mixed up with 4 µl of transfection reagent. The mixture was allowed to stay at room temperature to form complex. The mixture of HP reagent and DNA complex was added dropwise to the cells. After 48 hours of incubation, one third of the cells were brought in selection medium, and the rest two thirds of the cells were induced to check transfection efficiency.

In this work, dKDM4a SL2 stable cell line was established by co-transfection of pHFW-dKDM4a with puromycin resistant plasmid in 10 to 1 ratio. Selection was done by applying 10 µg/ml of puromycin in the culture medium and 20 min of 37 °C heatshock was performed to induce the expression.

All the IDH cell line and IDH mutant cell lines were transfected by either pMK33-IDH-CFH or pMK33-IDH\_mutant-CFH plasmid which contains hygromycin resistant gene in the plasmid. Selection was done by applying 250 µg/ml of hygromycin B in the culture medium and 250 µM of CuSO<sub>4</sub> was included to induce the expression.

#### 5.2.3 Immune Fluorescence staining for microscope detection

The cells were de-attached, transferred to coverslip and allowed to settle down for 30 min at room temperature. The coverslip was transferred into 12-well plate and washed with PBS for 10 min. Cells were immersed in 3.7 % formaldehyde in PBS for fixation 10 min at room temperature. They were washed twice with PBS for 10 min each and permeabilized with 0.25 % Triton X-100 in PBS on ice for 6 min. Cells were washed twice with PBS for 10

min each, transferred to parafilm and blocked with Image-IT<sup>®</sup> FX signal enhancer for 1 hour. This was followed by overnight incubation at 4 °C with primary antibody diluted in 5 % normal goat serum in PBS. The cells were washed in 12-well plate with 0.1 % Triton X-100 in PBS for 10 min twice and incubated typically 1 hour at room temperature with corresponding secondary antibody diluted in 5 % normal goat serum in PBS. Cells were washed in 12-well plate with 0.1 % Triton X-100 in PBS for 10 min twice. Mount was done with DAPI in Vectashield to stain DNA and the coverslip was sealed on microscope slide. Fluorescence signal picture was acquired by Axiovert 200 and analyzed by Fiji software.

#### 5.2.4 General protein sample methods:

##### **Isolate proteins by SDS-PAGE electrophoresis.**

Protein samples were mixed together with Laemmli loading buffer, heat-denatured at 95 °C for 5 min, loaded on SDS-PAGE gel. The gel was run in Invitrogen XCell SureLock<sup>™</sup> Electrophoresis cell chamber at 200 V voltage until the dye front reaches the gel edge.

##### **Coomassie stain and de-stain**

SDS-PAGE gel was stained for 25 min in coomassie staining buffer containing 0.25 % Coomassie G250, 50 % methanol and 10 % acetic acid. It was de-stained in 10 % acetic acid overnight.

##### **Western Blotting**

The gel was de-assembled from electrophoresis device. In order to transfer the protein to PVDF membrane, it was re-assembled from negative electrode to positive electrode in the order with wet sponge - Whatman paper - gel - PVDF membrane - Whatman paper - wet sponge. Protein was transferred from gel to PVDF membrane in Bio-Rad Mini Trans-Blotting Cell chamber.

Transferring was done in cold room 400 mA 1 h for small histones or 300 mA 2h for general proteins or 50 mA overnight for big size proteins.

The membrane was de-assembled and marked with the protein side. It was blocked in 5 % (w/v) milk/PBS or 3 % BSA in PBS/0.1 % Tween for at least 1 h at RT on a shaker in order to get rid of unspecific background. The membrane was blotted with primary antibody against the protein of interest with a proper dilution in 1 % (w/v) milk/PBS for overnight in cold room or 3 h at RT. Washing was done with 0.1 % Tween/PBS for 3×10 min at RT. The membrane was blotted with secondary antibody conjugated with HRP for 1 min to 1 h

which is depended on the intensity of the signal. It was washed with 0.1 % Tween/PBS for 4×10 min at RT.

Detection of signal was done by chemiluminescence using Amersham ECL Prime Western Blotting Detection Reagent RPN2232. Detection solution A and B were mixed with 1:1 ratio and 500 µl of this mixture was used for one 6×8 cm<sup>2</sup> membrane. The mixture was loaded on a Saran film. The membrane was placed with protein side down. Incubate the reaction for 5 min at RT.

The solution was drained off and the membrane was wrapped with Saran film. The signal was exposed from membrane to a sheet of autoradiography film.

#### **Reversible protein staining with ponceau S**

After proteins were transferred onto the PVDF membrane, the membrane was incubated in Ponceau S Staining Solution at room temperature for 5 minutes. Protein bands were visualized. The membrane was washed with distilled water and shortly immersed in 0.1 M NaOH solution. Protein bands started to disappear after 10-30 seconds. The membrane was washed with distilled water for 2-3 minutes.

#### **Strip western blotting membrane**

The membrane was incubated in 0.2 M Glycine pH2.4 for 15 min at RT on a shaker. Washing was done with 0.1 % Tween/PBS for 3×10 min at RT.

#### **Whole cell extract by RIPA buffer**

Cell pellet was resuspended in RIPA buffer containing 0.1 % SDS, 0.5 % Deoxycholate, 0.5 % NP40, 1mM EDTA, 50 mM Tris pH7.5 and 150mM NaCl, in addition of proteinase inhibitors. Samples were vortexed shortly and incubated at 4 °C for 10 min. Laemmli buffer was added and sample mixture was incubated at 95 °C, 5 min for denaturing protein before loading the sample into SDS-PAGE Gel.

#### **Nuclear extract preparation**

I collected up to 10 billion cells and they were centrifuged at 1000 g for 20 min at 4 °C. Cell pellet was resuspended in PBS and separated into two 50ml-falcon tubes. Centrifugation and wash were repeated. Then cells were resuspended in ice cold Buffer A consisting of 10 mM HEPES pH7.6, 15 mM KCl, 2 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 1 mM DTT and supplied with complete protease inhibitors. The suspension was placed on ice for 30 min to break cell membrane by swelling followed by homogenization with B. Braun S fit pestle in cold room. The homogenized sample was mixed with Buffer B in 10 to 1

volume ratio. Buffer B consisted of 50 mM HEPES pH7.6, 1 M KCl, 30 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 1 mM DTT and was supplied with complete protease inhibitors. Centrifugation was done at 8000 g for 25 min at 4 °C and the pellet was nuclei fraction. Mixture of Buffer A and Buffer B in 9 to 1 volume ratio was added to the nuclei pellet in addition with 0.4 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> for final concentration. The mixture was rotated in cold room for 25 min and followed by Ultracentrifugation with Ti 70 at 40000 rpm for 1.5 hours at 4 °C. Supernatant containing the nuclei protein was transferred into a Ti 45 tube and solid (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was added to reach final concentration of 0.3 g/ml. Nuclei protein was precipitated and centrifuged at 15000 rpm for 35 min at 4 °C. The pellet was dissolved in Buffer C consisting of 25 mM HEPES pH7.6, 150 mM KCl, 12.5 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 1 mM DTT, 10 % glycerol (v/v) and supplied with complete protease inhibitors.

Extracted nuclei protein was dialyzed against Buffer C with 3 changes for 1 liter 1 hour each at 4 °C. Dialyzed proteins were centrifuged, aliquoted and snap chilled in liquid nitrogen. Protein concentration was determined by Bradford using BSA as standard with known concentration and detecting the readout from spectrophotometer at 595 nm wave length. Extracted nuclei protein could be stored in - 80 °C for months before use.

#### **Protein concentration measurement**

200 µl of protein dye was taken from Bio-Rad Protein Assay, known concentration of BSA dilution series were diluted in the same buffer as protein sample. Absorbance value was detected by Spectrophotometer at wave length 595 nm. Standard curve of absorbance value against concentration was made. Protein sample concentration was calculated according to the standard curve by reading absorbance value.

#### **5.2.5 lysine demethylation *in vitro* assay**

##### **Generation of recombinant dKDM4a enzyme by Baculo virus-insect cell expression system**

Low passage Sf21 cells were seeded with 0.8 million cells in 2 ml complete culture medium, namely Sf900 II SFM complemented with 10 % FCS and 100 µg/ml Gentamycin as final concentration, in 6-well plate. Cells were allowed to attach to the bottom for 1 hour at 27 °C. Transfection was done by X-tremeGENE HP DNA Transfection Reagent according to its product protocol for insect cells. Here 2 µg of bacmid DNA was used. Transfected cells were incubated at 27 °C for 3 days, take the supernatant of culture and label it as P1. SF21 cells were split down to 0.5 million cells per ml density, and P1 was diluted in 1 to 100 ratio. 20 ml cells were infected with 10 µl dilution to amplify

baculovirus. Supernatant of culture was harvested after 7 days and labeled as P2. P2 was used to infect Sf21 cells to express protein.

#### **Strep-tagged dKDM4a protein purification**

Strep-tagged dKDM4a over-expression Sf21 cells were lysed in buffer W containing 100 mM Tris-Cl pH8.0, 150 mM NaCl in addition of complete protease inhibitors. Sonification was done to break cell membrane and high speed centrifugation was performed to clarify the lysate. Supernatant was applied to Strep-tactin sepharose beads equilibrated by buffer W and packaged in flow through column. The beads were washed with 5 column volume of buffer W. Strep-tagged dKDM4a was eluted by 3 column volume of 2.5 mM D-Desthiobiotin in buffer W. Glycerol was added to 20 % of the total volume. Samples were aliquoted in small fraction, chilled in liquid nitrogen and stored at - 80 °C.

Protein concentration was determined by LI-COR Odyssey with known concentration of BSA providing standard curve loaded in SDS-PAGE gel together with sample protein.

#### **His-tagged hKDM4a protein purification**

Bacteria culture expressing recombinant hKDM4a residues 1-359 amino acids was resuspended in buffer A consisting of 50 mM HEPES pH7.6, 500 mM NaCl, 20 mM Imidazole, 0.5 mM DTT, in addition of complete proteinase inhibitor including PMSF, Leupeptin, Aprotinin, Pepstain. The cell membrane was broken by sonication. The sample mixture was added with Triton X-100 to final concentration of 0.5 %, rotated at 4 °C for 30 min. Centrifugation at 18000 rpm for 30 min at 4 °C. Supernatant was loaded to the super loop and assembled in AKTA-UPC 900 purification system. Nickel column was equilibrated with 5 fold column volume of buffer A until the conductivity reached a new plateau. The sample was injected from super loop and flowed with buffer A at 1 ml/min for 10 min. Buffer B consisted of 50 mM HEPES pH7.6, 500 mM NaCl, 300 mM Imidazole, 0.5 mM DTT, in addition of complete proteinase. 25 min 7 % of buffer B, 93 % of buffer A was applied to get mixture of 40 mM Imidazole to wash away the unspecific binding materials. His-tagged hKDM4a proteins was eluted 10 min by 250 mM Imidazole by mixing 80 % buffer B and 20 % buffer A. Fractions from elution were collected and loaded on SDS-PAGE gel.

#### **HDM assay**

1  $\mu$ M of KDM4a enzyme and 10  $\mu$ M H3K36me3 peptide as substrate were incubated together with 100  $\mu$ M FeSO<sub>4</sub>, 100  $\mu$ M  $\alpha$ -ketoglutarate, 500  $\mu$ M Ascorbic acid, 50 mM HEPES pH 7.6, 50 mM NaCl at 26 °C for 20 min.



**KDM4a kinetics**

The demethylation reaction was incubated at 26 °C from 0 to 180 min. At each time point of 0 min, 5 min, 10 min, 20 min, 40 min, 60 min, 80 min, 100 min, 120 min, a aliquot of reaction was taken out and quenched by 0.2 % TFA.

**MALDI-ToF analysis of demethylation *in vitro***

HDM assay after reaction was concentrated and de-salted by C 18 resin column using Zip Tips (Millipore). Purified substrate and product peptides were spotted to target plate together with saturated  $\alpha$ -cyano-4-hydroxy-cinnamic acid in 0.1 % TFA / 80 % ACN.

Mass spectra were acquired by Applied Biosystems (Framingham, CA, USA) Voyager DE STR with mass range from 800 to 1500 amu according to the method described by (Bonaldi et al., 2004).

**KDM4a inhibition assay by inhibitor *in vitro***

1  $\mu$ M of enzyme and inhibitor dilution series were pre-incubated for 5 min at room temperature. The rest of HDM assay components were added before measuring the enzyme activity with 20 min incubation at 26 °C by MALDI-ToF.

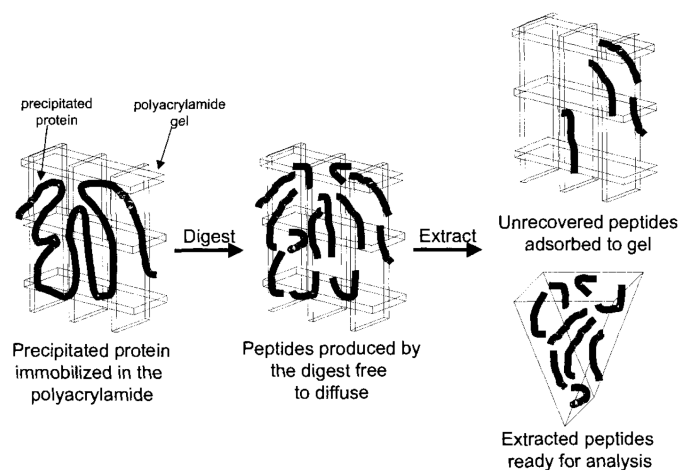
**5.2.6 Proteomics sample preparation****Flag tag affinity purification**

Equal amount of Anti-Flag<sup>®</sup> M2 Affinity Gel slurry was taken for each sample. It was shortly spun down to settle the beads. The beads were equilibrated with 10 fold of beads volume of the same buffer as the sample dissolved in, here for instance, Buffer C as described in Nuclear extract preparation was used. Flag-tagged protein extract was applied to the beads. The mixture was rotated at 4 °C for 3 hours, washed with Buffer C for 10 min at 4 °C for 3 times. Elution of purified Flag-tagged protein was done by adding laemmli buffer to the beads and boiling at 95 °C for 10 min.

**Proteomics sample in-gel tryptic digestion**

Immuno precipitated proteins from Flag tag affinity purification using nuclear extract of dKDM4a over-expressing SL2 cell line and wild type SL2 cell line as control were separated by SDS-PAGE gel and cut into 8 fractions. De-staining was done with 50 % acetonitrile (ACN) / 20 mM  $\text{NH}_4\text{HCO}_3$  buffer at 37 °C for 60 min. After washing with water for two times, the gel pieces were dehydrated with 100 % ACN for  $3 \times 10$  min. Reduction of cysteine bisulfate bond was applied by rehydrating the gel pieces with 10 mM DTT (1,4-Dithiothreitol) in 20 mM  $\text{NH}_4\text{HCO}_3$  buffer and incubating 1 hour at room

temperature. Alkylation was performed by adding 55 mM Iodoacetamide (IAA) in 20 mM  $\text{NH}_4\text{HCO}_3$  buffer. The gel pieces were washed with 20 mM  $\text{NH}_4\text{HCO}_3$  buffer once and dehydrated with 100 % ACN for three times. Trypsin digestion was performed at 37 °C overnight by rehydrating the gel pieces with trypsin in 20 mM  $\text{NH}_4\text{HCO}_3$  buffer. 1 µg of trypsin was applied per 50 µg proteins. Acid extraction was followed the next day. 50 µl of 50 % ACN / 0.25 % TFA was applied to the gel pieces and incubation was done at 37 °C for 10 min repeat once. 50 µl of 100 % ACN was applied to the gel pieces and repeated once. Solutions were pooled from each step for the same sample together and dried in Speed vacuum. The resulting peptides were dissolved in 0.1 % FA to acidify before loading on the LC column.



**Figure 5.2 | Recovery tryptic digested peptides from immobilized protein in SDS-PAGE gel.**

The proteins were separated by SDS-PAGE gel according to their size. They were immobilized in the gel by methanol fixation in the staining steps to prevent from diffusion out from the gel during tedious buffer changing steps. Peptides generated from tryptic digestion were much smaller in size comparing to proteins, therefore could diffuse out from gel and extracted by acid. The strong hydrophobic peptides may bind to the polyacrylamide gel and hard to be recovered (adapted from Kinter & Sherman, 2000).

### Proteomics sample LC/MS/MS measurement

Samples were prepared by in gel tryptic digestion before LC-MS/MS runs. Three IP pull down replicates ended up with 48 samples with each 90 min LC-MS/MS runs. For proteomic analysis, reversed-phase UPLC system DIONEX Ultimate 3000 RSLC (75 µm × 15 cm C 18 Reprosil pure 2.4 µm Dr. Maisch) from Thermo Fisher Scientific was coupled on-line with nano-electrospray ionization and LTQ Orbitrap XL mass spectrometer from Thermo Fisher Scientific.

### Proteomics sample data analysis

Raw file data was analysed using MaxQuant Version 1.2.2.5 for protein identification and quantification using Andromeda search engine against dmel-all-translation-r5.24 fasta database. Quantification was performed by integrating the area under the extracted ion chromatogram peak of the corresponding ion. Protein IDs identified from contamination database and reverse transcription database were excluded. Missing values were imputed by shifting the Gaussian distribution of validated data values with 0.3 of the width and 1.8 of standard deviation downwards. Output file from MaxQuant was subjected to statistical analysis by Perseus based on iBAQ value.

## 5.2.7 Histone sample preparation

### 5.2.7.1 Acid Extraction of Histones from *Drosophila* SL2 or L2-4 cells

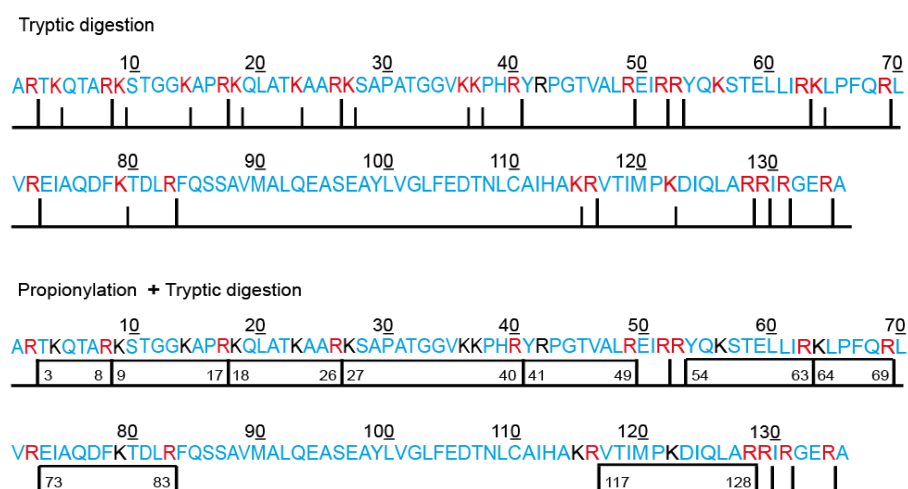
Cells were centrifuged at 1500 rpm for 10 min at 4 °C, washed once with PBS and centrifuged again. Cell pellet was resuspended in ice cold PBS containing 0.3 % of Triton X 100 and complete protease inhibitors. Samples were rotated for 10 min at 4 °C and centrifuged at 3500 rpm for 10 min at 4 °C. The pellet was nuclei fraction. The nuclei fraction was centrifuged and resuspended in ice cold 0.4 M HCl. Acid extraction was performed overnight at 4 °C. Supernatant was transferred to molecularporous membrane tubing with MWCO of 6-8000 after centrifugation. Samples were dialyzed against 3 changes of 1 liter of 100 mM acetic acid for 1 hour each at 4 °C. Finally, the samples were transferred to 1.5 ml cold Eppendorf tube and freezed in -80 °C for 10 min before lyophilization. Resuspended histones in loading dye Laemmli buffer were boiled for 5 min at 95 °C.

### 5.2.7.2 Histone in-gel tryptic digestion

Histone protein bands were separated in SDS-PAGE gel, visualized by Coomassie G250 dye staining, cut into pieces, de-stained with 50 % ACN / 50 mM ammonium bicarbonate at 37 °C for 60 min. After washing with water for two times, acylation step was performed by adding 1 µl of propionic anhydride together with 1 M ammonium bicarbonate in 50 µl volume to the sample and incubated at room temperature for 30 min. Washing step was repeated for three times. 50 % ACN was added to the sample and incubated at 37 °C for 15 min. Dehydration was performed by adding 50 µl of 100 % ACN to the gel pieces. Proteolytic enzyme trypsin was incorporated into the gel piece by rehydrating with solution containing trypsin in 50 mM ammonium bicarbonate in addition of 0.5 pmol for each of the spiketides. Digestion was incubated at 37 °C overnight. In gel acid extraction was

performed the next day in six steps. Step 1, collect the solution from trypsin digestion for each sample. Step 2, wash gel pieces with 50 mM of ammonium bicarbonate. Step 3, dehydrate with 50 % ACN / 25 mM ammonium bicarbonate. Step 4, acid extraction with 5 % formic acid. Step 5, dehydrate with 50 % ACN / 2.5 % formic acid. Step 6, dehydrate with 100 % ACN. Solutions from step 1 to 6 were pooled for the same sample together and dried by speed vacuum. Peptide dry powder was resuspended in 0.1 % trifluoroacetic acid (TFA) to acidify.

All reagents used here for mass spectrometry work were MS and HPLC grade quality.



**Figure 5.3 | Schematic illustration of propionylation strategy application.**

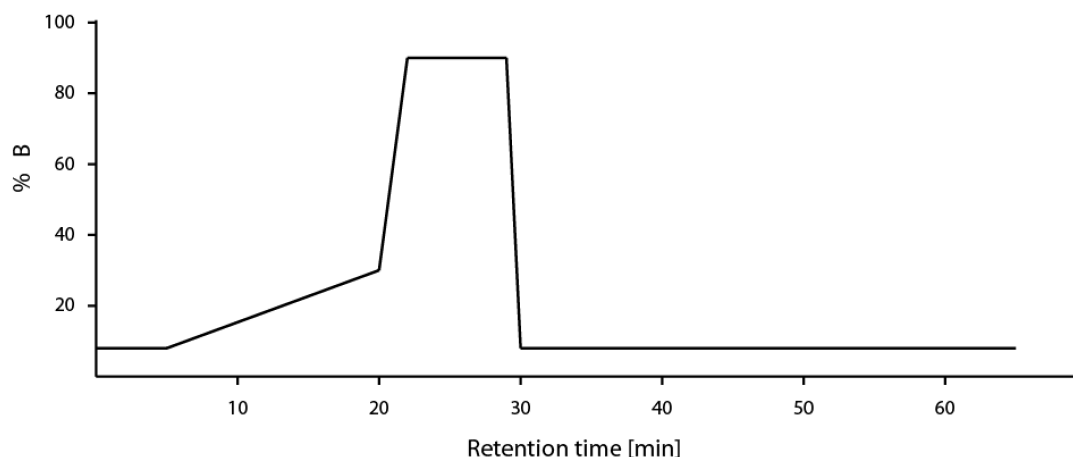
Propionylation was applied to block trypsin cleavage after unmodified and mono-methylated lysine to make trypsin digestion behave as Arginine-C cleavage enzyme.

#### 5.2.7.3 Peptide desalting by C 18 stage tips and carbon tips

C 18 stage tips and carbon tips were conditioned by 100 % methanol, wetted by 80 % ACN / 0.1 % TFA and equilibrated by 0.1 % TFA. Peptide solution was loaded twice on C 18 stage tips and the flow through was loaded on carbon tips twice. The tips were washed three times with 0.1 % TFA and peptide was eluted in 80 % ACN / 0.25 % TFA. The elution of the same sample from C 18 stage tips and carbon tips were pooled together in LC tube. The sample was dried in speed vacuum and resuspended in 0.1 % TFA before load on the UHPLC column.

#### 5.2.7.4 Histone sample LC/MS/MS measurement

For histone analysis, reversed-phase UHPLC system DIONEX Ultimate 3000 RSLC (75  $\mu\text{m}$  x 15 cm C 18 Reprosil pure 2.4  $\mu\text{m}$  Dr. Maisch) was coupled on-line with nano-electrospray ionization and Triple ToF TT6600 mass spectrometry from AB Sciex.



**Figure 5.4 | UHPLC data acquisition program**

NC pump flowed at 0.285  $\mu\text{L}/\text{min}$ . Nano spray speed was 50 nl/sec. Samples were loaded to the column from 0<sup>th</sup> to 5<sup>th</sup> min with 4 %B after injection, gradient was applied from 4 % to 35 % of B starting from 5<sup>th</sup> min to 20<sup>th</sup> min; 35 % to 90 % of B starting from 20<sup>th</sup> min to 22<sup>th</sup> min; 90 % of B was kept from 22<sup>th</sup> min to 29<sup>th</sup> min; gradient dropped from 90 % to 4 % of B starting from 29<sup>th</sup> min to 30<sup>th</sup> min; 4 % of B was kept from 30<sup>th</sup> min to 65<sup>th</sup> min to regenerate the column. A = 0.1 % TFA (v/v), B = 100 % ACN (v/v), %A + %B = 100 %.

Survey scan was performed first for each LC/MS/MS run. Multiple reaction monitoring (MRM) was applied to target the parental ion and collision energy were calculated according to the mass over charge and peptide bond feature for each peptide of interest.

#### 5.2.7.5 Histone modification analysis

Raw files with “.wiff” format acquired from Triple ToF mass spectrometer TT6600 were analyzed by PeakView V 2.1 and MultiQuant 3.0. Quantitation of the peptide was done by integrating the area under the LC chromatogram peak. This was defined as MS1 quantitation. In the case of different modifications with same mass-to-charge value within same peptide sequence, and the peaks were not distinguishable from each other by MS1, quantitation was done by integrating the area under the LC chromatogram peak of extracted diagnostic fragmented ion. This was defined as MS2 quantitation. Identification of different modification on peptide was done by PeakView V 2.1. Single modification quantitation among different samples was done by PeakView V 2.1. Multiple modification signal quantitation among multiple samples were done by MultiQuant 3.0 using MQ4 peak integration package.

## 6. Abbreviation and appendix

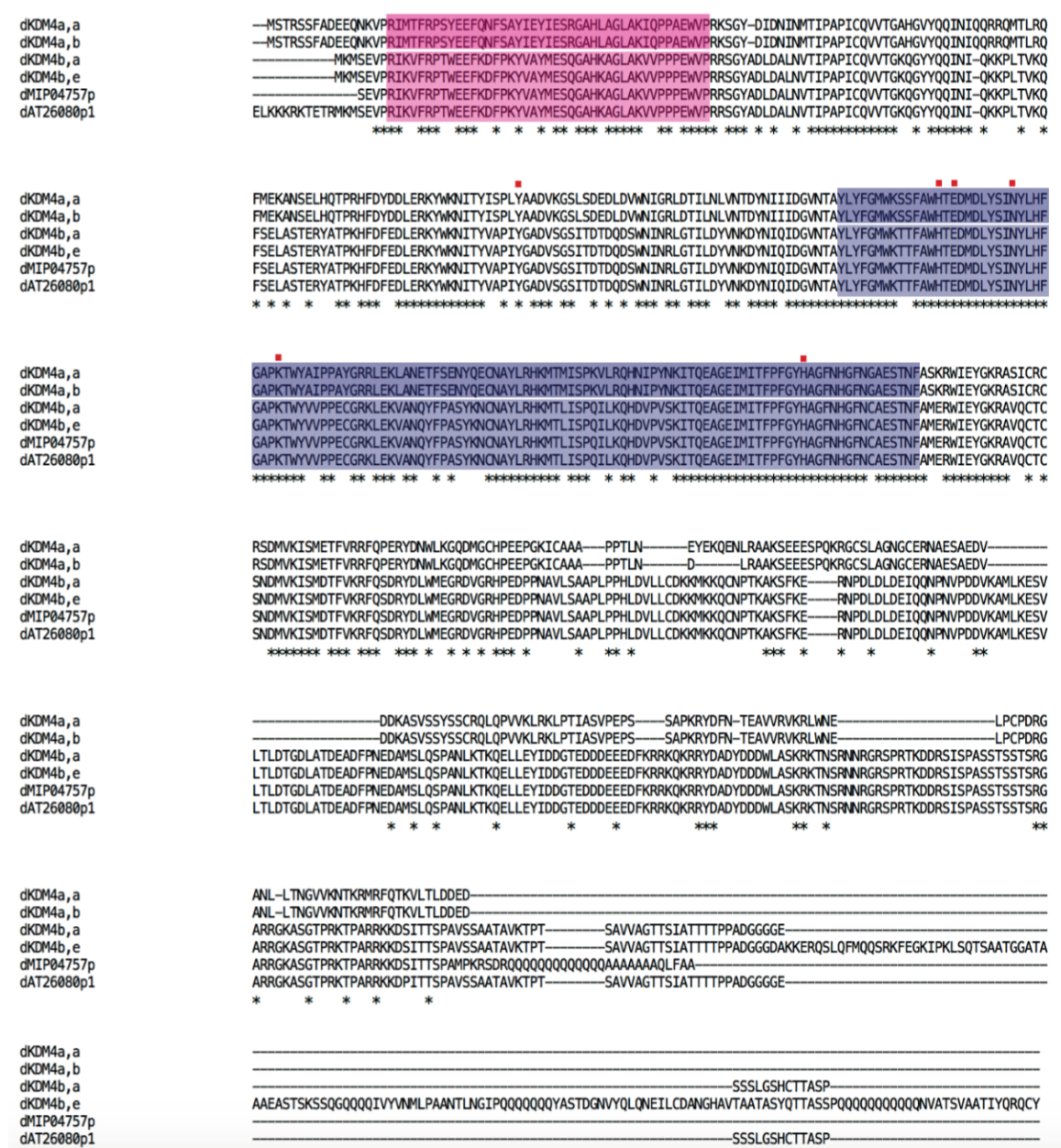
### 6.1 Abbreviation

ADP	Adenosindiphosphate
Ala	Alanine
AR	Androgen receptor
Asn	Asparagine
ATP	Adenosine-5'-triphosphate
ATR-X	Alpha-thalassemia X-linked
BRG1	Transcription activator
Cbx1, 3, 5	Chromobox homolog 1, 3, 5
CDYL	Chromodomain protein, Y- like
CENP-A	Centromeric protein A
CIMP	CpG island methylator phenotype
Co-IP	Co- Immunoprecipitation
CoREST	Co-repressor to REST
CTD	C-terminus domain
Da	Daltons
DCC	Dosage compensation complex
Df31	Decondensation factor 31
dKDM4a	Drosophila Lysine specific demethylase 4 A
dKDM4B	Drosophila Lysine specific demethylase 4 B
DNA	Deoxyribonucleic acid
Dnmt3a	DNA methyl transferase
DOT1	Disrupter of telomeric silencing
Eaf3	Component of the Rpd3S histone deacetylase complex
EcR	Ecdysone receptor
EED	Polycomb group (PcG) protein
FAD	Flavin adenine dinucleotide
FADH2	Reducted FADH2
G9a	Lysine specific methyltransferase
Glu	Glutaminic Acid
Gly	Glycine
H1, H2A, H2B, H3, H4	Histone protein
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDM	Histone demethylase
HIRA	Histone cell cycle regulator
His	Histidine
hKDM4a	Human Lysine specific demethylase 4 a
HMT	histone methyltransferase
HP1	Heterochromatin protein 1
HP1 $\gamma$	Heterochromatin protein 1 $\gamma$
ICF	immunodeficiency, centromere instability and facial anomalies
Ino80	A member of the SNF2 family of ATPases
IP	Immunoprecipitation
JARID1A	Jumonji, AT rich interactive domain 1A
Jmj-C	Jumonji C
KDM4A	Lysine specific demethylase 4 A
KDM4B	Lysine specific demethylase 4 B
KDM4C	Lysine specific demethylase 4 C
KDM4D	Lysine specific demethylase 4 D
KDM4E	Lysine specific demethylase 4 E
KDM4F	Lysine specific demethylase 4 F
KMT5A	Lysine specific methyltransferase 5 a
L3MBTL1	Lethal (3) malignant brain tumor-like 3
LC-MS	Liquid chromatography mass spectrometry
LSD1	Lysine specific demethylase 1

Lys	Lysine
MALDI-ToF	Matrix Assisted Laser Desorption Ionization – Time of Flight
MAT	Methionine adenosyltransferases
MLE	Maleless
MSL3	Male-specific lethal 3
N-PAC	Cytokine-like nuclear factor
NSD1	Nuclear receptor-binding, su(var), enhancer-of-zeste and trithorax domain-containing protein 1
NSD2	Nuclear receptor-binding, su(var), enhancer-of-zeste and trithorax domain-containing protein 2
ORC	Origin recognition complex
PHD	Plant homeodomain
PHF19	PhD finger protein 19
PR-Set7	PR-SET domain containing protein 7
PRC1/2	Polycomb repressive complex 1/2
PRMT	Protein arginine methyltransferase
PSIP1	PC4 and SFRS1 interacting protein 1
PTM	Post translational modification
PWWP	Proline-Tryptophan- Proline-Tryptophan motif
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RNA-Seq	RNA-Sequencing
RNAi	RNA interference
Rpd3S	Reduced potassium dependency 3 S
RRI	RNA RNA interaction
SAHA	Suberoylanilide hydroxamic acid
SAM	S-adenosyl methionine
Ser	Serine
SET	Suppressor of position effect variegation 3-9, SU(VAR)3-9; Enhancer of zeste, E(Z) and Trithorax, Trx
Set8	SET domain containing 8
SRM	Selected reaction monitor
SUV39H	Suppressor of position effect variegation 3-9 homolog
Suv4-20h1/h2	Suppressor of variegation 4-20 homolog 1 / homolog 2
TCA	Tricarboxylic acid
Thr	Threonine
TOR	Target Of Rapamycin
Tyr	Tyrosine
ZMYND11	Zinc finger, MYND- type containing 11
$\alpha$ -KG	$\alpha$ - ketoglutarate



## 6.2 Appendix 1



**Figure 6.1 | Protein sequence alignment of KDM4 family in *Drosophila melanogaster***

Pink box indicates Jmj-N domain; blue box indicates Jmj-C domain.

“\*” illustrates the identical amino acids present in both hKDM4a and dKDM4a.



### 6.3 Appendix 2

hKDM4a	MASES----ETLNPSA	RIMTFYPTMEEFRNFSRYIAYIESQGAHRAGLAKVVPKEWKPR
dKDM4a, a	MSTRSSFADDEEQNKVP	RIMTFRPSYEEFQNFSAIYIEYIESRGAGHLAKIQPPAEWVPR
	* * * * *	***** * *** ** * * * * * * * * * * * * * *
		■ ■
hKDM4a	ASYDDIDDL--VIPAPIQQLVTGQSGFLTQYNIQ-KKAMTVREFRKIANSDKYCTPRYSE	
dKDM4a, a	KSGYDIDNINMTIPAPICQVVTGAHGVYQQINIQRQMTLRQFMKANSELHQTPRHFD	
	* ***	***** * *** * * *** * * * * * * * * * *
hKDM4a	FEELERKYWKNLTFNPPIYGADVNGTLYEKHVDENIGRLRLTILDLVEKESGITIEGVNT	
dKDM4a, a	YDDLERYWKNITYISPLYAADVKGSLSDLDVWNIGRLDTILNLVNTDYNIIIDGVNT	
	***** * * * * * *	* ***** ** * * * * * * * * * *
hKDM4a	PYLYFGMWKTSFAWHTEDMDLYSINYLHFGEPKSWYSVPPHEGKRRLERLAKGFFPGSAQS	
dKDM4a, a	AYLYFGMWKSSFAWHTEDMDLYSINYLHFGAPKTWYAIPPAYGRREKLANETFSYNYQE	
	*****	***** * * * * * * * * * * * * * * * * * *
hKDM4a	CEAFLRHKMTLISPLMLKKYGIPFDKVTQEAGEFMITFPYGYHAGFNHGFNCAESTNFAT	
dKDM4a, a	CNAYLRHKMTMISPKVLRQHNIPYNKITQEAGEIMITFPFGYHAGFNHGFNGAESTNFAS	
	* * *****	*** * * * * * * * * * * * * * * * * * *
hKDM4a	RRWIEYGKQAVLCSCRKDMVKISMDFVRKFQPERYKLWKAGKDNVIDH-----TLPT	
dKDM4a, a	KRWIEYGKRASICRCRSDMKISMETFVRRFQPERYDNWLKGQDMGCHPEEPGKICAAAP	
	*****	* * * * * * * * * * * * * * * * * *
hKDM4a	PEAAEFLKESEL	
dKDM4a, a	PTLNEYEKQENL	
	* * * *	

**Figure 6.2 | Protein sequence alignment of hKDM4a residue 1-359 aa and dKDM4a residue 1-372 aa**

Core catalytic region of hKDM4a 1-359 aa and dKDM4a residue 1-372 aa shares 63 % identities and 80 % similarities. On top of the sequence, red dots depict the binding sites of cofactor  $\alpha$ -ketoglutarate and  $\text{Fe}^{2+}$ ; Blue dots depict the key structure sites apart from binding sites. “\*” illustrates the identical amino acids present in both hKDM4a and dKDM4a.

## 6.4 Appendix 3

iBAQ 1306 wt	iBAQ 1325 wt	iBAQ 1546 wt	iBAQ 1306 KDM4a	iBAQ 1325 KDM4a	iBAQ 1546 KDM4a	-Log t-test p value	t-test Difference	Current symbol
15.725	16.550	16.476	23.836	23.398	24.882	3.960	7.788	CG16972
12.741	12.323	12.392	15.508	15.023	15.136	3.822	2.737	-
12.559	12.050	12.159	14.388	14.691	14.876	3.471	2.395	CG33107
11.512	11.141	11.444	16.694	17.590	19.325	2.946	6.504	CG18190
11.117	11.179	10.391	15.041	17.111	16.217	2.885	5.227	Atac3
11.768	11.212	12.762	17.141	18.709	19.715	2.782	6.607	VhaAC39-1
12.133	11.671	13.205	16.215	15.918	16.900	2.756	4.008	CG7987
11.704	11.535	11.938	14.304	13.406	14.240	2.714	2.257	RAF2
11.949	10.328	11.557	16.850	17.314	15.451	2.677	5.260	Ppt1
13.910	14.784	14.650	16.843	16.848	16.308	2.616	2.218	msl-1
22.121	21.666	21.544	23.085	23.631	23.430	2.604	1.605	14-3-3epsilon
11.845	9.738	12.978	18.838	20.799	18.599	2.588	7.892	HP6
11.483	11.894	11.893	15.138	14.392	16.200	2.522	3.486	SdhB
17.115	15.951	15.148	20.147	19.435	20.009	2.464	3.792	mTTF
10.614	11.806	11.820	15.884	17.182	14.996	2.448	4.607	CtsB1
17.022	15.134	16.294	19.431	19.998	20.580	2.411	3.853	CG3902
14.418	17.604	15.565	23.800	27.933	24.412	2.408	9.519	KDM4A
15.513	15.183	15.167	19.033	20.181	17.981	2.372	3.777	26-29-p
9.474	10.515	11.879	15.644	18.789	18.354	2.353	6.973	Jafrac2
11.945	10.806	11.725	14.670	13.924	13.639	2.291	2.586	MESR4
22.462	21.904	22.573	23.598	24.397	24.278	2.272	1.778	FK506-bp1
11.620	10.445	10.386	15.482	16.731	14.066	2.218	4.610	CG4866
11.319	13.901	10.776	17.038	19.593	19.001	2.215	6.545	CG14932
17.471	16.776	17.088	18.446	18.323	18.077	2.164	1.170	Rpn13
19.130	19.791	18.647	23.187	23.986	21.734	2.163	3.779	Su(var)205
19.857	19.701	19.460	20.350	20.462	20.238	2.161	0.677	eIF-4E
13.758	13.094	12.731	16.705	15.601	17.864	2.097	3.529	CG14200
10.873	10.866	10.335	12.959	15.444	14.906	2.072	3.745	Cwc25
12.341	11.156	10.216	14.920	14.145	14.217	2.067	3.189	CG10979
17.227	15.481	16.276	18.604	19.009	19.461	2.063	2.697	RpL12
14.609	15.281	16.059	17.164	17.398	17.657	2.039	2.090	CG11982
14.340	16.229	15.357	10.795	12.518	10.184	2.022	-4.143	Drat
18.368	18.482	18.021	19.390	20.359	19.632	2.019	1.504	Tap42
20.204	18.841	19.344	21.252	21.491	22.014	2.013	2.123	epsilonCOP
18.780	19.111	19.481	21.555	20.483	21.998	1.967	2.221	zetaCOP
10.826	11.967	14.107	16.154	17.569	17.450	1.958	4.758	RpL15
11.228	12.917	13.006	14.816	15.040	15.114	1.950	2.606	CG9044
17.031	16.688	17.056	17.921	19.036	18.318	1.905	1.500	Nup37
10.824	11.006	12.841	14.902	14.104	15.531	1.898	3.289	G9a
15.594	15.716	14.273	17.566	18.899	17.463	1.883	2.782	CG11920
14.276	11.390	11.351	16.522	17.083	16.106	1.861	4.232	Thiolase
13.419	13.811	13.684	15.705	14.561	15.021	1.843	1.458	Sap130

12.089	12.068	12.360	16.730	18.283	14.618	1.828	4.371	CG16721
20.186	20.486	19.356	21.480	23.003	23.037	1.815	2.497	SF2
21.311	23.364	23.986	16.950	19.383	19.097	1.777	-4.410	betaTub97EF
11.720	11.973	9.393	14.062	14.670	14.273	1.769	3.306	-
16.558	19.009	19.856	22.356	22.071	23.406	1.745	4.137	l(1)G0004
10.989	13.744	12.260	15.015	17.093	16.805	1.743	3.973	Mfap1
16.155	17.746	17.505	12.649	14.619	11.058	1.725	-4.360	CG32066
12.832	13.612	14.195	15.197	14.890	15.745	1.683	1.731	tho2
11.326	11.990	8.900	17.341	13.893	16.356	1.676	5.125	BRWD3
10.823	10.852	12.689	14.657	13.684	15.954	1.670	3.310	E(bx)
20.466	21.213	21.543	22.372	23.278	24.128	1.662	2.186	Chd64
15.568	14.961	16.752	17.560	17.990	17.660	1.662	1.976	Surf6
15.362	15.493	16.610	17.092	18.531	18.112	1.638	2.090	Brms1
12.690	11.018	11.998	13.988	17.282	17.231	1.633	4.265	CG6018
11.083	14.700	14.620	17.637	17.710	19.840	1.617	4.928	LRR
15.734	14.920	16.342	18.508	18.504	16.960	1.614	2.326	DCP1
12.456	11.901	12.374	15.731	18.542	14.687	1.604	4.076	sel
17.593	19.554	19.646	21.890	21.338	21.038	1.598	2.491	His2A:CG31618
17.036	15.978	18.994	12.752	14.577	11.372	1.589	-4.436	CG32086
18.625	19.030	18.166	19.889	21.207	19.923	1.589	1.732	Nurf-38
11.536	11.407	10.203	13.137	15.226	13.180	1.587	2.799	mab-21
10.544	14.015	11.007	15.142	17.641	16.315	1.586	4.511	msl-3
11.434	11.317	12.710	15.030	13.802	13.418	1.576	2.263	Br140
12.165	11.987	10.606	13.738	13.458	15.341	1.558	2.593	tlk
17.383	19.104	20.059	21.349	21.406	21.972	1.554	2.727	His2B:CG17949
14.229	16.028	14.162	17.319	16.532	17.238	1.552	2.223	Prp38
8.123	12.276	11.158	14.437	17.561	15.212	1.545	5.218	MCPH1
13.978	16.780	16.856	18.603	19.390	20.246	1.542	3.542	Ge-1
20.993	21.222	22.199	22.479	23.380	23.078	1.530	1.508	14-3-3zeta
12.549	14.127	13.914	16.030	19.880	21.504	1.526	5.608	fax
13.810	14.641	13.823	15.891	14.921	16.267	1.522	1.602	mu2
16.449	19.070	16.759	14.640	11.577	9.651	1.513	-5.470	Hdac3
12.297	15.851	12.859	16.509	18.598	18.183	1.488	4.094	Pdp1
15.598	10.532	12.826	17.154	17.897	19.966	1.470	5.354	tmod
12.121	16.584	15.313	10.415	10.930	8.866	1.458	-4.603	CG10737
14.994	14.535	16.051	12.564	13.981	11.921	1.456	-2.371	FKBP59
16.918	15.619	12.858	18.442	19.027	19.562	1.453	3.879	eIF-5A
11.622	15.026	14.829	16.578	18.653	19.689	1.452	4.481	su(sable)
18.193	18.192	17.243	20.562	20.728	18.805	1.448	2.156	CG7945
12.305	13.126	14.230	15.812	15.280	18.104	1.434	3.178	Lk6
18.960	19.698	18.169	20.991	22.284	20.298	1.432	2.249	DnaJ-1
20.756	19.023	22.330	13.714	13.184	18.239	1.412	-5.657	regucalcin
18.210	17.965	18.711	19.815	20.499	22.715	1.402	2.714	lva
13.804	11.979	10.649	14.196	16.121	16.095	1.392	3.327	CG1024
13.513	13.891	14.239	14.612	15.660	16.558	1.347	1.729	dbr

15.739	10.217	12.026	16.801	17.131	18.156	1.314	4.702	CG12262
18.172	17.981	16.635	19.059	19.102	18.771	1.308	1.382	eIF4AIII
15.418	17.514	17.041	18.252	20.400	18.858	1.306	2.512	CG11577
20.783	20.742	21.597	21.641	22.949	23.238	1.302	1.569	Gie
10.476	10.650	12.173	12.362	16.010	16.004	1.302	3.692	CG42748
17.606	19.389	17.877	16.819	16.779	16.633	1.300	-1.547	srp

## 7. References

- Adcock I M, Ito K and Barnes P J. 2005. Histone deacetylation: an important mechanism in inflammatory lung diseases. *COPD* 2:445-455.
- Alekseyenko A A, Gorchakov A A, Zee B M, Fuchs S M, Kharchenko P V, Kuroda M I. 2014. Heterochromatin-associated interactions of *Drosophila* HP1a with dADD1, HIPPI, and repetitive RNAs. *Genes Dev* 28:1445-1460.
- Allis C D, Berger S L, Cote J, Dent S, Jenuwien T, Kouzarides T, Pillus L, Reinberg D, Shi Y, Shiekhhattar R, Shilatifard A, Workman J and Zhang Y. 2007. New nomenclature for chromatin-modifying enzymes. *Cell* 131(4):633-636.
- Banaszynski L A, Allis C D, Lewis P W. 2010. Histone variants in metazoan development. *Dev Cell* 19:662-674.
- Barth T K, Schade G O M, Schmidt A, Vetter I, Wirth M, Heun P, Thomae A W and Imhof A. 2014. Identification of novel *Drosophila* centromere associated proteins. *Proteomics* 14:2167-2178.
- Baum B & Cherbas L. 2007. *Drosophila : Methods and Protocols*. Methods in Molecular Biology 420:391-424.
- Beck D B, Oda H, Shen S S and Reinberg D. 2012. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev* 26:325-337.
- Becker B P and Hörz W. 2002. ATP-dependent nucleosome remodeling. *Annu Rev Biochem* 71(1):247-273.
- Bell O, Wirbelauer C, Hild M, Scharf A N D, Schwaiger M, MacAlpine D M, Zilbermann F, Leeuwen F, Bell S P, Imhof A, Garza D, Peters A HFM and Schübeler D. 2007. Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in *Drosophila*. *EMBO* 26:4974-4984.
- Berdasco M, Ropero S, Setien F, Fraga M F, Lapunzina P, Losson R, Alaminos M, Cheung N-K, Rahman N and Esteller M. 2009. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc Natl Acad Sci USA* 106:21830-21835.
- Berry W L, Shin S, Lightfoot S A and Janknecht R. 2012. Oncogenic features of the JMJD2A histone demethylase in breast cancer. *Int J Oncol* 41:1701-1706.
- Bhaumik S R, Smith E and Shilatifard A. 2007. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol* 14:1008-1016.
- Biterge B and Schneider R. 2014. Histone variants: key players of chromatin. *Cell Tissue Res* 356:457-466.
- Black J C, Manning A L, Rechem C V, Kim J, Ladd B, Cho J, Pineda C M, Murphy N, Daniels D L, Montagna C, Lewis P W, Glass K, Allis C D, Dyson N J, Getz G and Whetstine J R. KDM4A Lysine Demethylase Induces Site-Specific Copy Gain and Rereplication of Regions Amplified in

Tumors. *Cell* 154:1-15.

Black J C, Rechem C V and Whetstine J R. 2012. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell* 48:491-507.

Bonaldi T, Imhof A and Regula J T. 2004. A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. *Proteomics* 4(5):1382-1396.

Brehm A, Tufteland K R, Aasland R and Becker P B. 2004. The many colours of chromodomains. *Bioessays* 26:133-140.

Brien G L, Gambero G, O'Connell D J, Jerman E, Turner S A, Egan C M, Dunne E J, Jurgens M C, Wynne K, Piao L, Lohan J, Ferguson N, Shi X, Sinha K M, Loftus B J, Cagney G and Bracken A P. 2012. Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat Struct Mol Biol* 19:1273-1281.

Busnel J-M, Schoenmaker B, Ramautar R, Carrasco-Pancorbo A, Ratnayake C, Feitelson J S, Chapman J D, Deelder A M and Mayboroda O A. 2010. High capacity capillary electrophoresis-electrospray ionization mass spectrometry: coupling a porous sheathless interface with transient-isotachopheresis. *Anal Chem* 82(22):9476-9483.

Byvoet P, Shepherd G R, Hardin J M, Noland B J. 1972. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch Biochem Biophys* 148:558-567.

Cairns R A and Mak T W. 2013. Oncogenic isocitrate dehydrogenase mutations: mechanisms, models, and clinical opportunities. *Cancer Discov* 3:730-741.

Calin G A and Croce C M. 2006. MicroRNA signatures in human cancers. *Nat Rev Cancer* 6:857-866.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones R S, Zhang Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298:1039-1043.

Carrozza M J, Li B, Florens L, Suganuma T, Swanson S K, Lee K K, Shia W, Anderson S, Yates J, Washburn M P and Workman J L. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123:581-592.

Chin R M, Fu X, Pai M Y, Vergnes L, Hwang H, Deng G, Diep S, Lomenick B, Meli V S, Monsalve G C, Hu E, Whelan S A, Wang J X, Jung G, Solis G M, Fazlollahi F, Kaweeteerawat C, Quach A, Nili M, Krall A S, Godwin H A, Chang H R, Faull K F, Guo F, Jiang M, Trauger S A, Saghatelian A, Braas D, Christofk H R, Clarke C F, Teitell M A, Petrascheck M, Reue K, Jung M E, Frand A R and Huang J. 2014. The metabolite  $\alpha$ -ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. *Nature* 509:397-401.

Cloos P A C, Christensen J, Agger K and Helin K. 2008. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev* 22:1115-1140.

Cloos P A C, Christensen J, Agger K, Maiolica A, Rappsilber J, Antal T, Hansen K H and Helin K. 2006. The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature* 442:307-311.

- Couture J F, Collazo E, Ortiz-Tello P A, Brunzelle J S and Trievel R C. 2007. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat Struct Mol Biol* 14:689-695.
- Cox J and Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* 26:1367-1372.
- Cox J, Neuhauser N, Michalski A, Scheltema R A, Olsen J V, and Mann M. 2011. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J Proteome Res* 10:1794-1805.
- Crona F, Dahlberg O, Lundberg L E, Larsson J and Mannervik M. 2013. Gene regulation by the lysine demethylase KDM4A in *Drosophila*. *Dev Biol* 373:453-463.
- Dai L, Peng C, Montellier E, Lu Z, Chen Y, Ishii H, Debernardi A, Buchou T, Rousseaux S, Jin F, Sabari B R, Deng Z, Allis D C, Ren B, Khochbin S and Zhao Y. 2014. Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat Chem Biol* 10(5):365-370.
- Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, and Jeltsch A. 2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* 285:26114-26120.
- Duerre J A and Lee C T. 1974. In vivo methylation and turnover of rat brain histones. *J Neurochem* 23:541-547.
- Egger G, Liang G, Aparicio A and Jones PA. 2004. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429:457-463.
- Ehrbrecht A, Müller U, Wolter M, Hoischen A, Koch A, Radlwimmer B, Actor B, Mincheva A, Pietsch T, Lichter P, Reifemberger G and Weber R G. 2006. Comprehensive genomic analysis of desmoplastic medulloblastomas: identification of novel amplified genes and separate evaluation of the different histological components. *J Pathol* 208:554-563.
- Eick D and Geyer M. 2013. The RNA Polymerase II Carboxy-Terminus Domain (CTD) Code. *Chem Rev* 113:8456-8490.
- Eskeland R, Leeb M, Grimes G R, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultschi A I, Wutz A and Bickmore W A. 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* 38(3):452-464.
- Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, Liu S, Alder H, Costinean S, Fernandez-Cymering C, Volinia S, Guler G, Morrison C D, Chan K K, Marcucci G, Calin G A, Huebner K and Croce C M. 2007. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc Natl Acad Sci USA* 104:15805-15810.
- Feller C, Forne I, Imhof A and Becker P B. 2015. Global and Specific Responses of the Histone Acetylome to Systematic Perturbation. *Mol Cell* 57:559-571.
- Feng J and Fan G. 2009. The role of DNA methylation in the central nervous system and neuropsychiatric disorders. *Int Rev Neurobiol* 89:67-84.
- Fenn J B, Mann M, Meng C K, Wong S F and Whitehouse C M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246:64-71.

- Fouse S D and Costello J F. 2009. Epigenetics of neurological cancers. *Future Oncol* 5:1615-1629.
- Friedman J M, Liang G, Liu C-C, Wolff E M, Tsai Y C, Ye W, Zhou X and Jones P A. 2009. The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. *Cancer Res* 69:2623-2629.
- Friend C, Scher W, Holland JG, Sato T. 1971. Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethyl sulfoxide. *Proceedings of the National Academy of Sciences of the United States of America*, 68:378-382.
- Gräff J and Mansuy M I. 2008. Epigenetic codes in cognition and behavior. *Behav Brain Res* 192:70-87.
- Greer E L and Shi Y. 2012. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* 13:343-357.
- Guenther M G, Levine S S, Boyer L A, Jaenisch R and Young R A. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130:77-88.
- Guruharsha K G, Obar R A, Mintseris J, Aishwarya K, Krishnan R T, Vijayraghavan K and Artavanis-Tsakonas S. 2012. Drosophila protein interaction map (DPiM): a paradigm for metazoan protein complex interactions. *Fly (Austin)* 6:246-253.
- Guruharsha K G, Rual J-F, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee D Y, Cenaj O, McKillip E, Shah S, Stapleton M, Wan K H, Yu C, Parsa B, Carlson J W, Chen X, Kapadia B, Vijayraghavan K, Gygi S P, Celniker S E, Obar R A and Artavanis-Tsakonas S. 2011. A protein complex network of Drosophila melanogaster. *Cell* 147:690-703.
- Hake S B, Garcia B A, Duncan E M, Kauer M, Dellaire G, Shabanowitz J, Bazett-Jones D P, Allis C D and Hunt D F. 2006. Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J Biol Chem* 281:559-568.
- Heard E, Rougeulle C, Arnaud D, Avner P, Allis C D and Spector D L. 2001. Methylation of Histone H3 at Lys-9 Is an Early Mark on the X Chromosome during X Inactivation. *Cell* 107:727-738.
- Heijden G W, Derijck A A, Posfai e, Giele M, Pelczar P, Ramos L, wansink D G, van der vlag J, Peters A H, de Boer P. 2007. Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation. *Nat Genet* 39:251-258.
- Hoffmann E. and Stroobant V. 2007. *Mass Spectrometry Principles and Applications*. (Third Edition) ISBN 978-0-470-03310-4.
- Holliday R. 2005. DNA methylation and epigenotypes. *Biochemistry* 70:612-617.
- Holliday R. 2006. A historical overview. *Epigenetics* 1:76-80.
- Hon G, Wang W and Ren B. 2009. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput Biol* 5:e1000566.
- Imhof A. 2004. *Histone Modifications-Marks for Gene Expression?* ISBN 978-1-4419-9072-3.



- Javierre B M, Fernandez A F, Richter J, Al-Shahrour F, Martin-Subero J I, Rodriguez-Ubreva J, Berdasco M, Fraga M F, O'Hanlon T P, Rider L G, Jacinto F V, Lopez-Longo F J, Dopazo J, Forn M, Peinado M A, Carreño L, Sawalha A H, Harley J B, Siebert R, Esteller M, Miller F W and Ballestar E. 2010. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res* 20:170-179.
- Jones P A and Baylin S B. 2007. The epigenomics of cancer. *Cell* 128:683-692.
- Karas M & hillenkamp F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10, 000 daltons. *Anal Chem* 60:2299-2301.
- Katada S, Imhof A and Sassone-Corsi P. 2012. Connecting threads: epigenetics and metabolism. *Cell* 148:24-28.
- Katoh M and Katoh M. 2004. Identification and characterization of JMJD2 family genes in silico. *Int J Oncol* 24:1623-1628.
- Katoh Y, Ikura T, Hoshikawa Y, Tashiro S, Ito T, Ohta M, Kera Y, Noda T and Igarashi K. 2011. Methionine adenosyltransferase II serves as a transcriptional corepressor of Maf oncoprotein. *Mol Cell* 41:554-566.
- Katz D J, Edwards T M, Reinke V, Kelly W G. 2009. A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* 137:308-320.
- Kelly T K, Carvalho D D and Jones P A. 2010. Epigenetic modifications as therapeutic targets. *Nat Biotechnol* 28:1069-1078.
- Kinter M and Sherman N E. 2000. Protein sequencing and identification using tandem mass spectrometry.
- Klose R J, Kallin E M and Zhang Y. 2006. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* 7:715-727.
- Kouzarides T. 2007. Chromatin modifications and their function. *Cell* 128(4):693-705.
- Kuo L C. 2011. Fragment-Based Drug Design: Tools, Practical Approaches, and Examples. *Methods in ENZYMOLOGY* 493:137-138.
- Labbé R M, Holowatyj A and Yang Z. 2013. Histone lysine demethylase (KDM) subfamily 4: structures, functions and therapeutic potential. *Am J Transl Res* 6:1-15.
- Labussiere M, Sanson M, Idbaih A and Delattre J-Y. 2010. IDH1 gene mutations: a new paradigm in glioma prognosis and therapy? *Oncologist* 15:196-199.
- Lan F, Zaratiegui M, Villen J, Vaughn M W, Verdel A, Huarte M, Shi Y, Gygi S P, Moazed D, Martienssen R A and Shi Y. 2007. *S. pombe* LSD1 homologs regulate heterochromatin propagation and euchromatic gene transcription. *Mol Cell* 26:89-101.
- Larschan E, Alekseyenko A A, Gortchakov A A, Peng S, Li B, Yang P, Workman J L, Park P J, Kuroda M I. 2007. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* 28:121-133.
- Lee J V, Carrer A, Shah S, Snyder N W, Wei S, Venneti S, Worth A J, Yuan Z F, Lim H-W, Liu S, Jackson E, Aiello N M, Haas N B, Rebbeck T R, Judkins A, Won K-J, Chodosh L A, Garcia B A,

- Stanger B Z, Feldman M D, Blair I A and Wellen K E. 2014. Akt-dependent metabolic reprogramming regulates tumor cell histone acetylation. *Cell metabolism* 20:306-319.
- Lin C-J, Koh F M, Wong P, Conti M and Ramalho-Santos M. 2014. Hira-mediated H3.3 incorporation is required for DNA replication and ribosomal RNA transcription in the mouse zygote. *Dev Cell* 30:268-279.
- Lin C, Li B, Swanson S, Zhang Y, Florens L, Washburn M P, Abmayr S M, and Workman J L. 2008. Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the *Drosophila* KDM4A demethylase. *Mol Cell* 32:696-706.
- Liu F, Quesada V, Crevillen P, Baurle I, Swiezewski S, Dean C. 2007. The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to down regulate FLC *Mol Cell* 28:398-407.
- Lizcano J M, Unzeta M and Tipton K F. 2000. A spectrophotometric method for determining the oxidative deamination of methylamine by the amine oxidases. *Anal Biochem* 286:75-79.
- Lloret-Llinares M, Carré C, Vaquero A, Olano N and Azorín F. 2008. Characterization of *Drosophila melanogaster* JmjC+N histone demethylases. *Nucleic Acids Res* 36:2852-2863.
- Lohse B, Kristensen J L, Kristensen L H, Agger K, Helin K, Gajhede M, Clausen R P, 2011. Inhibitors of histone demethylases. *Bioorg Med Chem* 19:3625-3636.
- Loppin B, Berger F and Couble P. 2001. The *Drosophila* maternal gene sesame is required for sperm chromatin remodeling at fertilization. *Chromosoma* 110:430-440.
- Loppin B, Bonnefoy E, Anselme C, Laurençon A, Karr T L and Couble P. 2005. The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. *Nature* 437:1386-1390.
- Lorbeck M T, Singh N, Zervos A, Dhatta M, Lapchenko M, Yang C and Elefant F. 2010. The histone demethylase Dmel\KDM4A controls genes required for life span and male-specific sex determination in *Drosophila*. *Gene* 450:8-17.
- Lowe N, Rees J S, Roote J, Ryder E, Armean I M, Johnson G, Drummond E, Spriggs H, Drummond J, Magbanua J P, Naylor H, Sanson B, Bastock R, Huelsmann S, Trovisco V, Landgraf M, Knowles-Barley S, Armstrong J D, White-Cooper H, Hansen C, Phillips R G, UK *Drosophila* Protein Trap Screening Consortium, Lilley K S, Russell S, and Johnston D. 2014. Analysis of the expression patterns, subcellular localisations and interaction partners of *Drosophila* proteins using a pigP protein trap library. *Development* 141:3994-4005.
- Loyola A and Almouzni G. 2007. Marking histone H3 variants: How, when and why? *Trends Biochem Sci* 32:425-433.
- Luco R F, Pan Q, Tominaga K, Blencowe B J, Pereira-Smith O M and Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* 327:996-1000.
- Luger K, Mäder A W, Richmond R K, Sargent D F, and Richmond T J. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251-260.
- MacDonald M J, Longacre M J, Langberg E-C, Tibell A, Kendrick M A, Fukao T and Ostenson C-G. 2009. Decreased levels of metabolic enzymes in pancreatic islets of patients with type 2 diabetes. *Diabetologia* 52:1087-1091.

- Maenner S, Müller M, Fröhlich J, Langer D and Becker P B. 2013. ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol Cell* 51:174-184.
- Maison C and Almouzni G. 2004. HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* 5:296-304.
- Markham G D and Pajares M A. 2008. Structure-function relationships in methionine adenosyltransferases. *Cell Mol Life Sci* 66:636-648.
- Menni C, Kastenmüller G, Petersen A K, Bell J T, Psatha M, Tsai P C, Gieger C, Schulz H, Erte I, John S, Brosnan M J, Wilson S G, Tsaprouni L, Lim E M, Stuckey B, Deloukas P, Mohny R, Suhre K, Spector T D and Valdes A M. 2013. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol* 42:1111-1119.
- Mentch S J, Mehrmohamadi M, Huang L, Liu X, Gupta D, Mattocks D, Padilla P G, Ables G, Bamman M M, Thalacker-Mercer A E, Nichenametla S N and Locasale J W. 2015. Histone Methylation Dynamics and Gene Regulation Occur through the Sensing of One-Carbon Metabolism. *Cell Metabolism* 22:1-13.
- Min J, Allali-Hassani A, Nady N, Qi C, Ouyang H, Liu Y, MacKenzie F, Vedadi M and Arrowsmith C H. 2007. L3MBTL1 recognition of mono- and dimethylated histones. *Nat Struct Mol Biol* 14:1229-1230.
- Møller I M and Sweetlove L J. 2010. ROS signalling – specificity is required. *Trends in plant science* 15:370-374.
- Mosammaparast N and Shi Y. 2010. Reversal of Histone Methylation: Biochemical and Molecular Mechanisms of Histone Demethylases. *Annu. Rev. Biochem.* 79:155-179.
- Musselman C A, Gibson M D, Hartwick E W, North J A, Gatchalian J, Poirier M G and Kutateladze T G. 2013. Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nat Commun* 4:2969-2967.
- Ng E K O, Tsang W P, Ng S S M, Jin H C, Yu J, Li J J, Röcken C, Ebert M P A, Kwok T T and Sung J J Y. 2009. MicroRNA-143 targets DNA methyltransferases 3A in colorectal cancer. *Br J Cancer* 101:699-706.
- Ng H H, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y and Struhl K. 2002. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev* 16:1518-1527.
- Ng S S, Kavanagh K L, McDonough M A, Butler D, Pilka E S, Lienard B M R, Bray J E, Savitsky P, Gileadi O, Delft F, Rose N R, Offer J, Scheinost J C, Borowski T, Sundstrom M, Schofield C J and Oppermann U. 2007. Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* 448:87-91.
- Nikolov M and Fischle W. 2013. Systematic analysis of histone modification readout. *Mol BioSystems* 9:182-194.
- Nishioka K, Rice J C, Sarma K, Erdjument-Bromage H, Werner J, Wang Y, Chuikov S, Valenzuela P, Tempst P, Steward R, Lis J T, Allis C D and Reinberg Danny. 2002. PR-Set7 is a nucleosome-

specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol Cell* 9: 1201-1213.

Opel M, Lando D, Bonilla C, Trewick S C, Boukaba A, Walfridsson J, Cauwood J, Werler P J H, Carr A M, Kouzarides T, Murzina N V, Allshire R C, Ekwall K and Laue E D. 2007. Genome-wide studies of histone demethylation catalysed by the fission yeast homologues of mammalian LSD1. *PLoS ONE* 2:e386.

Panier S and Boulton S J. 2014. Double-strand break repair: 53BP1 comes into focus. *Nat Rev Mol Cell Biol* 15:7-18.

Pedersen M T and Helin K. 2010. Histone demethylases in development and disease. *Trends Cell Biol* 20:662-671.

Picotti P, Clément-Ziza M, Lam H, Campbell D S, Schmidt A, Deutsch E W, Röst H, Sun Z, Rinner O, Reiter L, Shen Q, Michaelson J J, Frei A, Alberti S, Kusebauch U, Wollscheid B, Moritz R L, Beyer A and Aebersold R. 2013. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* 494:266-270.

Picotti P, Clément-Ziza M, Lam H, Campbell S D, Schmidt A, Deutsch W E, Röst H, Sun Z, Rinner O, Reiter L, Shen Q, Michaelson J J, Frei A, Alberti S, Kusebauch U, Wollscheid B, Moritz L R, Beyer A and Aebersold R. 2013. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* 494(7436):266-270.

Portela A, Esteller M. 2010. Epigenetic modifications and human disease. *Nature Biotechnology* 28:1057-1068.

Pradeepa M M, Sutherland H G, Ule J, Grimes G R, Bickmore W A and Reik W. 2012. Psip1/Ledgfp52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing. *PLoS Genet* 8:e1002717.

Rappsilber J, Mann M & Ishihama Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2(8):1896-1906.

Recham C, Black J C, Abbas T, Allen A, Rinehart C A, Yuan G-C, Dutta A and Whetstine J R. 2011. The SKP1-Cul1-F-box and leucine-rich repeat protein 4 (SCF-FbxL4) ubiquitin ligase regulates lysine demethylase 4A (KDM4A)/Jumonji domain-containing 2A (JMJD2A) protein. *J Biol Chem* 286:30462-30470.

Rhee D Y, Cho, D Y, Zhai B, Slattery M, Ma L, Mintseris J, Wong C Y, White K P, Celniker S E, Przytycka T M, Gygi S P, Obar R A, Artavanis-Tsakonas S. 2014. Transcription factor networks in *Drosophila melanogaster*. *Cell Rep* 8:2031-2043.

Rodriguez-Paredes M, Esteller M. 2011. Cancer epigenetics reaches mainstream oncology. *Nature Medicine* 17:330-339.

Rogers R S, Inselman A, Handel M A, Matunis M J. 2004. SUMO modified proteins localize to the XY body of pachytene spermatocytes. *Chromosoma* 113:233-243.

Rohle D, Popovici-Muller J, Palaskas N, Turcan S, Grommes C, Campos C, Tsoi J, Clark O, Oldrini B, Komisopoulou E, Kunii K, Pedraza A, Schalm S, Silverman L, Miller A, Wang F, Yang H, Chen Y, Kernytzky A, Rosenblum M K, Liu W, Biller S A, Su S M, Brennan C W, Chan T A,

- Graeber T G, Yen K E and Mellinghoff I K. 2013. An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* 340:626-630.
- Rudolph T, Yonezawa M, Lein S, Heidrich K, Kubicek S, Schäfer C, Phalke S, Walther M, Schmidt A, Jenuwein T and Reuter G. 2007. Heterochromatin formation in *Drosophila* is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol Cell* 26:103-115.
- Salminen A, Kaarniranta K, Hiltunen M and Kauppinen A. 2014. Krebs cycle dysfunction shapes epigenetic landscape of chromatin: Novel insights into mitochondrial regulation of aging process. *Cell Signal* 26:1598-1603.
- Sansoni V, Casas-Delucchi C S, Rajan M, Schmidt A, Bönisch C, Thomae A W, Staeger M S, Hake S B, Cardoso M C and Imhof A. 2014. The histone variant H2A.Bbd is enriched at sites of DNA synthesis. *Nucleic Acids Res* 42(10):6405-6420.
- Scharf A N D, Barth T K and Imhof A. 2009. Establishment of histone modifications after chromatin assembly. *Nucleic Acids Res* 37:5032-5040.
- Schnall-Levin M, Zhao Y, Perrimon N and Berger B. 2010. Conserved microRNA targeting in *Drosophila* is as widespread in coding regions as in 3'UTRs. *Proc Natl Acad Sci USA* 107:15751-15756.
- Schneider I. 1972. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J Embryol exp Morph* 27:353-365.
- Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D and Jenuwein T. 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *GENES & DEVELOPMENT* 18:1251-1262.
- Schubert T, Pusch M C, Diermeier S, Benes V, Kremmer E, Imhof A and Längst G. 2012. Df31 protein and snoRNAs maintain accessible higher order structures of chromatin. *Mol Cell* 48(3):434-444.
- Scott E M and Pillus L. 2010. Homocitrate synthase connects amino acid metabolism to chromatin functions through Esa1 and DNA damage. *Genes Dev* 24:1903-1913.
- Semenza G L. 2010. HIF-1: upstream and downstream of cancer metabolism. *Curr Opin Genet Dev* 20:51-56.
- Shi Y, Lan F, Matson C, Mulligan P, Whetstone J R, Cole P A, Casero R A and Shi Y. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119:941-953.
- Shi Y. 2007. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nat Rev Genet* 8:829-833.
- Sidoli S, Lin S, Xiong L, Bhanu N V, Karch K R, Johansen E, Hunter C, Mollah S and Garcia B A. 2015. SWATH Analysis for Characterization and Quantification of Histone Post-translational Modifications. *Molecular & Cellular Proteomics* 14(9):2420-2428.
- Sims III R J and Reinberg D. 2009. Processing the H3K36me3 signature. *Nat Genet* 41:270-271.
- Smith B C and Denu J M. 2009. Chemical mechanisms of histone lysine and arginine modifications. *Biochim Biophys Acta* 1789:45-47.

- Spannhoff A, Hauser A-T, Heinke R, Sippl W and Jung M. 2009. The emerging therapeutic potential of histone methyltransferase and demethylase inhibitors. *ChemMedChem* 4:1568-1582.
- Straub T, Zabel A, Gilfillan G D, Feller C and Becker P B. 2013. Different chromatin interfaces of the *Drosophila* dosage compensation complex revealed by high-shear ChIP-seq. *Genome Res* 23:473-485.
- Suganuma T and Workman J L. 2011. Signals and combinatorial functions of histone modifications. *Annu Rev Biochem* 80:473-499.
- Sural T H, Peng S, Li B, Workman J L, Park P J, Kuroda M I. 2008. The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. *Nat Struct Mol Biol* 15:1318-1325.
- Sweetlove L J, Beard K F M, Nunes-Nesi A, Fernie A R and Ratcliffe R G. 2010. Not just a circle: flux modes in the plant TCA cycle. *Trends Plant Sci* 15:462-470.
- Tan M K, Lim H J and Harper J W. 2011. SCF(FBXO22) regulates histone H3 lysine 9 and 36 methylation levels by targeting histone demethylase KDM4A for ubiquitin-mediated proteasomal degradation. *Mol Cell Biol* 31:3687-3699.
- Tan M, Luo H, Lee S, Jin F, Yang J S, Montellier E, Buchou T, Cheng Z, Rousseaux S, Rajagopal N, Lu Z, Ye Z, Zhu Q, Wysocka J, Ye Y, Khochbin S, Ren B and Zhao Y. 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146:1016-1028.
- Thiru A, Nietlispach D, Mott H R, Okuwaki M, Lyon D, Nielsen P R, Hirshberg M, Verreault A, Murzina A V and Laue E D. 2004. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *EMBO J* 23:489-499.
- Trauger S A, Webb W and Siuzdak G. 2002. Peptide and protein analysis with mass spectrometry. *Spectroscopy* 16:15-28.
- Trojer P, Zhang J, Yonezawa M, Schmidt A, Zheng H, Jenuwein T and Reinberg D. 2009. Dynamic Histone H1 Isotype 4 Methylation and Demethylation by Histone Lysine Methyltransferase G9a/KMT1C and the Jumonji Domain-containing JMJD2/KDM4 Proteins. *J Biol Chem* 284:8395-8405.
- Tsukada Y, Fang J, Erdjument-Bromage H, Warren M E, Borchers C H, Tempst P and Zhang Y. 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439:811-816.
- Tsurumi A, Dutta P, Yan S J, Shang R, Li W X. 2013. *Drosophila* Kdm4 demethylases in histone H3 lysine 9 demethylation and ecdysteroid signaling. *Sci Rep* 3:2894-2903.
- Turcan S, Rohle D, Goenka A, Walsh L A, Fang F, Yilmaz E, Campos C, Fabius A W M, Lu C, Ward P S, Thompson C B, Kaufman A, Guryanova O, Levine R, Heguy A, Viale A, Morris L G T, Huse J T, Mellinghoff I K and Chan T A. 2012. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* 483:479-483.
- Urdinguio RG, Sanchez-Mut J V and Esteller M. 2009. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* 8:1056-1072.
- Varambally S, Cao Q, Mani R-S, Shankar S, Wang X, Ateeq B, Laxman B, Cao X, Jing X,

- Ramnarayanan K, Brenner J C, Yu J, Kim J H, Han B, Tan P, Kumar-Sinha C, Lonigro R J, Palanisamy N, Maher C A and Chinnaiyan A M. 2008. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science* 322:1695-1699.
- Vardabasso C, Hasson D, Ratnakumar K, Chung C, Duarte L F and Bernstein E. 2014. Histone variants: emerging players in cancer biology. *Cell Mol Life Sci* 71:379-404.
- Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, et al. 2010. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 142: 967-980.
- Vezzoli A, Bonadies N, Allen M D, Freund S M, Santiveri C M, Kvinlaug B T, Huntly B J P, Göttgens B, and Bycroft M. 2010. Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nat Struct Mol Biol* 17:617-619.
- Vicent G P, Zaurin R, Nacht A S, Li A, Font-Mateu J, Dily F, Vermeulen M, Mann M and Beato M. 2009. Two chromatin remodeling activities cooperate during activation of hormone responsive promoters. *PLoS Genet* 5:e1000567.
- Villar-Garea A, Israel L and Imhof A. 2008. Analysis of histone modifications by mass spectrometry. *Curr Protoc Protein Sci* 14:10.
- Villeneuve L M and Natarajan R. 2010. The role of epigenetics in the pathology of diabetic complications. *Am J Physiol Renal Physiol* 299:F14-F25.
- Waddington CH. 1952. *Epigenetics of Birds*. Cambridge.
- Wagner T, Robaa D, Sippl W and Jung M. 2014. Mind the Methyl: Methyllysine Binding Proteins in Epigenetic Regulation. *ChemMedChem* 9:466-483.
- Walter W, Clynes D, Tang Y, Marmorstein R, Mellor J and Berger S L. 2008. 14-3-3 Interaction with Histone H3 Involves a Dual Modification Pattern of Phosphoacetylation. *Mol Cell Biol* 28:2840-2849.
- Wang F, Fischhaber P L, Guo C and Tang T S. 2014. Epigenetic modifications as novel therapeutic targets for Huntington's disease. *Epigenomics* 6:287-297.
- Wen H, Li Y Y, Xi Y X, Jiang S M, Stratton S, Peng D, Tanaka K, Ren Y F, Xia Z, Wu J, Li B, Barton M C, Li W, Li H T, and Shi X B. 2014. ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* 508:263-268.
- Williams T C R, Poolman M G, Howden A J M, Schwarzlander M, Fell D A, Ratcliffe R G and Sweetlove L J. 2010. A Genome-Scale Metabolic Model Accurately Predicts Fluxes in Central Carbon Metabolism under Stress Conditions. *Plant Physiology* 154:311-323.
- Wissmann M, Yin N, Müller J M, Greschik H, Fodor B D, Jenuwein T, Vogler C, Schneider R, Günther T, Buettner R, Metzger E and Schüle R. 2007. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat Cell Biol* 9:347-353.
- Xiao M, Yang H, Xu W, Ma S, Lin H, Zhu H, Liu L, Liu Y, Yang C, Xu Y, Zhao S, Ye D, Xiong Y and Guan K L. 2012. Inhibition of  $\alpha$ -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev* 26:1326-1338.

Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim S H, Ito S, Yang C, Wang P, Xiao M-T, Liu L-X, Jiang W-Q, Liu J, Zhang J-Y, Wang B, Frye S, Zhang Y, Xu Y-H, Lei Q-Y, Guan K-L, Zhao S-M and Xiong Y. 2011. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases. *Cancer cell* 19:17-30.

Yanaihara, N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens R M, Okamoto A, Yokota J, Tanaka T, Calin G A, Liu C G, Croce C M and Harris C C. 2006. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 9:189-198.

You A, Tong J K, Grozinger C M, Schreiber S L. 2001. CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proc Natl Acad Sci* 98:1454-1458.

Zhang P, Du J, Sun B, Dong X, Xu G, Zhou J, Huang Q, Liu Q, Hao Q and Ding J. 2006. Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res* 34:6621–6628.

Zhao S, Lin Y, Xu W, Jiang W, Zha Z, Wang P, Yu W, Li Z, Gong L, Peng Y, Ding J, Lei Q, Guan K-L and Xiong Y. 2009. Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1 $\alpha$ . *Science* 324:261-265.

Zoabi M, Nadar-Ponniah P T, Khoury-Haddad H, Usaj M, Budowski-Tal I, Haran T, Henn A, Mandel-Gutfreund Y and Ayoub N. 2014. RNA-dependent chromatin localization of KDM4D lysine demethylase promotes H3K9me3 demethylation. *Nucleic Acids Res* 42:13026-13038.



## 8. Acknowledgement

First I would like to thank my supervisor Prof. Axel Imhof for giving the opportunity to work in his lab and providing these interesting projects for my Ph.D work. Thank you so much for spending so much time and being patient to teach me everything. I appreciate the effort you made to setup collaborations and to improve the projects.

Thanks to all the former and current Imhof members for being nice to me and creating the great working atmosphere. Thanks to Irene Vetter to organize the lab and share nice protocols from the lab. Thanks to Dr. Andreas Thomae to teach me working with microscopy and sharing precious experience. I appreciate the time and thoughtful discussion you had for me. Many thanks to former colleague Simone Vollmer to help me setup in the lab and being always willing to help. Thanks to Teresa Barth, Thomas Gerland, Moritz Völker-Albert, Shibojyoti Lahiri, Natalia Kochanova, Andrea Lukacs, Victor Solis and Andreas Fedisch for the nice coffee time and chatting time.

I am really thankful to my thesis advisory committee members: Prof. Axel Imhof, Prof. Gunnar Schotta, Prof. Christoph Turk and Dr. Shahaf Peleg. Thank you all for providing thoughtful and constructive suggestions in the TAC meetings.

Thanks to the ZfP mass spectrometry team: Ignasi, Andreas, Pierre and Marc for offering spiketides, developing methods and running the samples for me.

Thanks to China Scholarship Council – LMU program and DFG offered the funding to support my Ph.D work.

Thanks to Becker department for the collaborating environment and inspiring discussions in the lab meeting.

My apologies and heartfelt thanks go to my family. Thank you all for standing by me with patient understanding.

Finally I owe a deep debt of gratitude to my husband Dr. Rui Fan for being always patient to me and unwavering support.

## 9. Publications

1. Rivera, C., Saavedra, F., Alvarez, F., Díaz-Celis, C., Ugalde, V., Li, J., et al. (2015). Methylation of histone H3 lysine 9 occurs during translation. *Nucleic Acids Res.* 43: 9097-9106. doi: 10.1093/nar/gkv929
2. Zhu J., Sun Y., Luo J., Wu M., Li J., Cao Y., et al. (2015). Specificity protein 1 regulates gene expression related to fatty acid metabolism in goat mammary epithelial cells. *Int J Mol Sci.* 16: 1806-1820. doi: 10.3390/ijms16011806 PMID: 25594872
3. Li B., Sheng M., Li J., Yan G., Lin A., Li M., et al. (2014). Tear proteomic analysis of Sjogren syndrome patients with dry eye syndrome by two-dimensional-nano-liquid chromatography coupled with tandem mass spectrometry, *Scientific Reports.* 4: 5772-5777. doi: 10.1038/srep05772

## Curriculum Vitae

### Jianhua Li

Born on July 27<sup>th</sup>, 1983 in Taiyuan, Shanxi province. P.R.China

Chinese Citizen

Married

Language: Chinese, English, German

### Address

Bio-Medical Center

Ludwig-Maximilians-University of Munich

Großhaderner Str.9

82152 Planegg-Martinsried Munich

Germany

[http://www.molekularbiologie.abi.med.uni-muenchen.de/personen/imhof\\_group/li/index.html](http://www.molekularbiologie.abi.med.uni-muenchen.de/personen/imhof_group/li/index.html)

Tel.: --49 17685653231

E-mail: [lijianhua991@gmail.com](mailto:lijianhua991@gmail.com)

### Education

**2010.9 – 2016.2** Ph.D specialized in Medicine, Medical Faculty, University of Munich (LMU), Germany

**2009.9 – 2010.7** Ph.D work specialized in Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, Northwest A&F University, P.R.China, declined on my part

**2006.9 – 2009.7** Master specialized in Genetics, College of Animal Science and Technology, Northwest A&F University, P.R.China

**2003.9 – 2006.7** Bachelor specialized in Bio-technology, College of Life Sciences, Northwest A&F University, P.R.China

### Degrees

**Feb 2016** Ph.D degree specialized in Medicine

**July 2009** Master degree specialized in Genetics

**July 2007** Bachelor degree specialized in Bio-technology

### Honors and Awards

**09, 2010** China Scholarship Council - University of Munich (LMU) scholarship

**11, 2008** HEFENG scholarship

**07, 2008** Participated the summer school hold by Chinese Academy of Science for postgraduate students

**04, 2005** Qualified with operation skill for Electronic Microscopy, HPLC, Photo Spectrometer.

**Training and meetings**

2010 October 2<sup>nd</sup> Chromatin symposium, Munich, Germany

2011 November EpiGeneSys 1<sup>st</sup> Annual Meeting, Vienna, Austria  
Poster: Characterization of histone demethylase KDM4A in *Drosophila melanogaster*

2013 July Science Camp, Regensburg, Germany  
Oral presentation: Functional Characterization of KDM4A in *Drosophila melanogaster*

2013 September EpiGeneSys-IMB Workshop, Mainz, Germany

2014 May Nuclear Proteomics conference, Kos, Greece  
Oral presentation: Screen for novel KDM4A inhibitors

2014 July Science Camp, Regensburg, Germany  
Poster: Screen for novel KDM4A inhibitors

2014 November EpiGeneSys 4<sup>th</sup> Annual Meeting, Barcelona, Spain

2015 January Doktorandenseminar 2015, Hohenroda, Germany  
Oral presentation: Analysis of H3K36 methylation using LC-MS and CESI-MS

2015 June STATegra meeting, Munich

2015 November LMU-CHINA academic network scientific forum  
Poster: Characterization of histone lysine demethylation in *Drosophila melanogaster*