# Characterization of lysine demethylase KDM5 sub-family:
## Substrate specificity and identification of potential novel non-histone substrates

By

Matthew Hoekstra

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Biology with specialization in Biochemistry

Carleton University
Ottawa, Ontario

## Abstract

A major regulatory influence over cell biology is lysine methylation and demethylation within histone proteins. The KDM5/JARID1 sub-family are 2-oxoglutarate and Fe(II)-dependent lysine-specific histone demethylases that are characterized by their Jumonji catalytic domains. This enzyme family is known to facilitate the removal of tri-/di-methyl modifications from lysine 4 of histone H3 (i.e., H3-K4me3/2), a mark associated with active gene expression. As a result, studies to date have revolved around KDM5's influence on disease through their ability to regulate H3-K4me2/3. Recently, evidence has demonstrated that KDM5's may influence disease beyond H3-K4 demethylation, making it critical to further investigate KDM5 demethylation of non-histone proteins. In efforts to help identify potential non-histone substrates for the KDM5 family, we developed a library of 180 permutated peptide substrates (PPS), with sequences that are systematically altered from the WT H3-K4me3 sequence. From this library, we characterized recombinant KDM5A/B/C/D substrate preference. Subsequently we developed recognition motifs for each KDM5 demethylase and used them to predict potential substrates for KDM5A/B/C/D. Demethylation activity was then profiled to generate a list of high/medium/low-ranking substrates for further *in vitro* validation for each of KDM5A/B/C/D. Through this approach, we analyzed prediction success rate and identified 66 high-ranking substrates in which KDM5 demethylases displayed significant *in vitro* activity towards. We further shown the ability to monitor changes in cellular methylation in a handful of the 66 high ranking candidate substrates in response to KDM5 inhibition. Specifically, we

focused validation efforts on a high-ranking KDM5A novel substrate: p53-K370me3. We demonstrated significant recombinant $KDM5A_{1\text{-}588\Delta AP}$ and $KDM5A_{1\text{-}801}$ activity towards the p53-K370me3 substrate *in vitro*. We then monitored KDM5A-mediated demethylation of the p53-K370me3/2 substrate in HCT 116 cells using a combination of wild-type KDM5A and inactive-mutant $KDM5A_{H483A}$ overexpression plasmids, along with immunoblotting, (co-) immunoprecipitation and mass spectrometry analysis. Furthermore, we have shown that KDM5A expression influences the established p53-53BP1 interaction. Finally, we identified a novel p53-TAF5 interaction dominated through the p53-K370me3 state and how KDM5A activity might affect this interaction. Ultimately, we have provided the first evidence of a KDM5 demethylase targeting a non-histone substrate for demethylation, via the novel KDM5A demethylation of the p53-K370me3 substrate.

## Preface

The following articles included in this thesis were either published or submitted to a scientific journal for means of publishing at the time the thesis was completed.

Chapter 2 contains the following paper that was published in *Star Protocols- Cell Press* that was reproduced in part:

Hoekstra, M., Chopra, A., Willmore, W.G., Biggar, K.K. (2022). Evaluation of jumonji C lysine demethylase substrate preference to guide identification of in vitro substrates. ***Star Protocols*** 3(2), 101271.



The article was reproduced in part because as per section 12.4A of the general regulations of graduate studies, I aimed to maintain a coherent account of a unified research project. I eliminated the last figure from the original publication as the changes to the described protocol depicted in that figure was not applied to any experiments done in this thesis. Moreover, I eliminated one half of the previous figure as to further maintain the coherent account of a unified research project, as it would appear in Chapter 3 in a more relevant context.

Chapter 3 contains the following paper that was published in *Analytical Biochemistry* that was reproduced in whole:

Hoekstra, M., Biggar, K.K. (2021). Identification of in vitro JMJD lysine demethylase candidate substrates via systematic determination of substrate preference. ***Anal. Biochem.*** 633, 114429.

Chapter 4 contains the following paper that was submitted and in review to the *Journal of Biochemistry* at the time this thesis was submitted and was reproduced in whole:

Hoekstra, M., Ridgeway, N.H., Biggar, K.K. (2022). Characterization of KDM5 lysine demethylase family substrate preference and identification of novel substrates. ***J. Biochem.,*** *In review at time of thesis submission*.

**From:** Kyle Biggar
**Date:** 04/07/2022 01.34 PM

**Submission ID:** 1281755
Date: 04 Jul 2022 2:34pm

**Name:** Dr. Kyle Biggar
Institute/company: Carleton University
Address: 1125 Colonel By Drive
Post/Zip Code: K1S 5B6
City: Ottawa
State/Territory: ON
Country: Canada
Telephone: 6138759176
Email: kyle_biggar@carleton.ca

**Type of Publication:** Journal

Title: Analytical Biochemistry
Auhtors: Matthew Hoekstra and Kyle K. Biggar
Year: 2021
From page: 114429
To page: 114429
ISSN: 0003-2697
Volume: 633
Article title: Identification of in vitro JMJD lysine demethylase candidate substrates via systematic determination of substrate preference.

**I would like to use:** Full article / chapter

I am the author of the Elsevier material: Yes

In what format will you use the material: Print and Electronic
Translation: No

**Proposed use:** Reuse in a thesis/dissertation

Material can be extracted: No

**Additional Comments / Information:** Permission is for inclusion in the Ph.D. thesis of Matthew Hoekstra.

---

**Rights and Permissions (ELS)** <Permissions@elsevier.com> Tue, Jul 5, 2022 at 12:03 PM
Reply-To: "Rights and Permissions (ELS)" <Permissions@elsevier.com>
To: kyle_biggar@carleton.ca

[External Email]

Dear Dr. Kyle Biggar,

We hereby grant you permission to reproduce the material detailed below in print and electronic format at no charge subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication as follows:

"This article was published in Publication title, Vol number, Author(s), Title of article, Page Nos, Copyright Elsevier (or appropriate Society name) (Year)."

3. This permission is granted for non-exclusive world rights in all languages.

4. Reproduction of this material is granted for the purpose for which permission is hereby given, and includes use in any future editions.

Kind regards,

**Roopa Lingayath**
Senior Copyrights Coordinator
**ELSEVIER** | HCM - Health Content Management

Visit Elsevier Permissions

Carleton UNIVERSITY
Canada's Capital University

Dr. Kyle K Biggar, PhD
Institute of Biochemistry
Carleton University
1125 Colonel By Drive
Ottawa, Ontario, Canada K1S 5B6
Tel: (613) 520-2600 (ext. 4487)
Email: kyle_biggar@carleton.ca

Sept 6th, 2022

This letter acknowledges that all listed original student contributions to material included in the thesis are accurate. Complete authorship details are listed at the beginning of each thesis chapter. Matthew Hoekstra is the primary contributor to original work presented herein.

Sincerely,

Kyle K. Biggar
Carleton University

## Acknowledgements

First and foremost, I would like to thank my thesis supervisor Dr. Kyle K Biggar for allowing me to join his lab when I had no prior wet-lab experience. He took the time to show me how to be a proper scientist including how to conduct myself in a laboratory, to planning and executing experiments, as well as data analysis and manuscript writing. I am thankful for all his support, scientific advice and allowing me branch out and apply myself to many areas of biochemistry and molecular biology. He has been a fantastic mentor and will continue to be a scientific role model for myself as I progress in the field of science throughout my career. I would also like to thank Dr. Hemanta Adhikary and Dr. Valentina Lukinović for the countless hours of technical and practical expertise as I completed my thesis research.

I would like to thank all members of the Biggar lab from past and present, there are simply too many to address but I will always cherish my time in this lab. It was always a joy coming to do research in such a fun, caring and inclusive environment with the nicest and joyful researchers and fellow students. It truly was a lab based on team camaraderie and support to help each other achieve our goals and ambitions. I will never forget the time I spent in the Biggar lab, the friends I have made for life and the aspirations I achieved there.

Most of all, I would like to thank my father Jim, mother Lisa and the rest of my family and friends for all the continuous support you have provided throughout the years. You were all a rock for me when ever needed and I don’t know where I would be without all of you today.

## Table of Contents

Title Page……i
Abstract……ii
Preface……iv
Acknowledgements……ixx
Table of Contents……x
List of Tables……xiiii
List of Illustrations……xiiiii
List of Appendices……xv
Chapter 1: Introduction……1
Chromatin Regulation……2
Histone post-translational modifications……4
Histone methyl regulatory proteins……13
Methods to study demethylase activity……31
Thesis Hypothesis……35
Chapter 2: Method Development……36
Chapter Summary……37
Manuscript Summary……41
Graphical abstract……41
Before you begin……42
Design permutated peptide substrate (PPS) library sequences……43
Key resources table……47
Materials and equipment (optional)……49
Step-by-step method details……51
Expected outcomes……68
Limitations……71
Troubleshooting……72
Acknowledgments……74
Author contributions……75
Chapter 3……76
Chapter Summary……77
Manuscript Summary……78

Introduction ..... 79
Materials and Methods ..... 83
Results ..... 87
Discussion ..... 97
Conclusion ..... 100
Chapter 4 ..... 102
Chapter Summary ..... 103
Manuscript Summary ..... 104
Introduction ..... 105
Materials and Methods ..... 108
Results ..... 117
Discussion ..... 132
Author Contributions ..... 136
Acknowledgements ..... 137
Funding source ..... 137
Chapter 5 ..... 138
Chapter summary ..... 139
Introduction ..... 140
Materials and Methods ..... 142
Results ..... 158
Discussion ..... 180
Chapter 6 ..... 185
References ..... 193
Appendices ..... 215

## List of Tables

**Table 2.1.** Example of permutation of the H3-K4me3 peptide sequence at the -3 position (underlined).

**Table 2.2.** Example of plate layout for JmjC 2x dilution series.

**Table 2.3.** Example of plate layout for testing JmjC KDM activity

**Table 2.4.** Example of plate layout for PPS library experiments.

**Table 2.5.** Example of motif-based scoring.

**Table 3.1.** Candidate protein substrates of KDM5A

**Table 4.1**. KDM5 full sequence homology and Jumonji domain homology analysis.

**Table 5.1.** Results of p53-K370me3 peptide interactomics through p53-K370me3 peptide pull down coupled with reductive dimethylation.

## List of Illustrations

**Figure 1.1.** Post-translational modifications on histone H3 protein.

**Figure 1.2.** Dynamic regulation of lysine methylation.

**Figure 1.3.** Biochemical mechanisms of JmJC-catalyzed and (FAD)-dependent amine oxidase-catalyzed lysine demethylation.

**Figure 1.4.** KDM5 sub-family sequence homology.

**Figure 1.5.** Domain architecture of KDM5 demethylase enzymes.

**Figure 2.1.** Determining non-saturating JmjC KDM concentration.

**Figure 2.2.** Validation of JmjC KDM methyl-state preference.

**Figure 2.3.** Representation of KDM substrate preference as a heat-map.

**Figure 2.4**. Representation of the screening data as PeSA generated sequence motif.

**Figure 2.5**. Distribution of peptide scores for predicted KDM5A substrates.

**Figure 3.1.** KDM5A activity toward H3K4 peptides.

**Figure 3.2.** Predicted KDM5A$_{1\text{-}801}$ specificity by permutated peptide substrate library.

**Figure 3.3.** Quantification of relative KDM5A$_{1\text{-}801}$ activity towards high, medium and low ranked substrates.

**Figure 3.4.** Insight into biological relevance of predicted KDM5A substrates.

**Figure 4.1.** Setup of PPS System.

**Figure 4.2.** Assessing KDM5A/B/C/D substrate specificity through a permutated peptide substrate (PPS) library.

**Figure 4.3.** Identifying high-ranking potential substrates for KDM5A/B/C/D.

**Figure 4.4.** Quantifying relative KDM5A/B/C/D activity towards high, medium and low ranked substrates.

**Figure 4.5.** Detecting changes in lysine methylation as a response to CPI-455 inhibition.

**Figure 5.1.** Determining KDM5A$_{1\text{-}588\Delta AP}$ specificity by permutated peptide substrate library.

**Figure 5.2**. Monitoring KDM5A demethylation of p53-K370me3 substrate.

**Figure 5.3.** Analysis of KDM5A mediated demethylation of p53-K370me3 substrate in HCT116 cells.

**Figure 5.4.** Identification of new TAF5 interaction with p53 K370me3 site through p53-K370me3 peptide pull down coupled with reductive dimethylation.

**Figure 5.5.** Investigation of KDM5A influence on novel TAF5-p53 interaction.

**Figure 5.6.** Schematic model for the dynamic regulation of p53-K370 methylation and the role in regulating p53 activity and p53 protein interactors.

## List of Appendices

**Appendix A.** Detection of JMJD-type KDM demethylation *in vitro.*

**Appendix B.** Quantifying peptide concentration and proposed KDM5A-mediated regulation of MTA1-NuRD co-repressor complex.

**Appendix C.** Proposed KDM5A-mediated regulation of MTA1-NuRD co-repressor complex.

**Appendix D.** KDM5 *in vitro* activity towards PPS library assessment of activity to predicted ranked substrates.

**Appendix E.** KDM5A *in vitro* activity towards p53 and assessment of to various methylated lysine sites in p53 C-terminal regulatory domain.

**Appendix F.** Development of WT and Mutant KDM5A overexpression plasmids and assessment of KDM5A-mediated demethylation in HEK 293 cells.

**Appendix G.** Nutlin 3A optimization in HCT 116 cells.

**Appendix H.** Analysis of p53-K370me2/1 peptide interactomics.

**Appendix I.** List of publications and patents

**Appendix J.** Communication at scientific meetings.

# Chapter 1

# Introduction to the methyllysine proteome and methyllysine-modifying enzymes

**Matthew Hoekstra[1,2], Kyle Biggar[1,2]**

[1]Institute of Biochemistry, Carleton University, Ottawa, ON, K1S 5B6, Canada

[2]Department of Biology, Carleton University, Ottawa, ON, K1S 5B6, Canada

## Chromatin Regulation

### *Epigenetics*

Despite the Human Genome Project (HGP) providing the blueprint of the genes encoding the entire human protein set, we are still left with limited knowledge regarding the functional regulation of the approximately 20,000 protein-coding human genes. For example, there remains approximately 30% (i.e., ~6,000) protein-encoding genes that have yet to be functionally validated in any way[1]. As well, many of the proteins which have been identified lack information in protein abundance, function, sub-cellular localization, and distribution[2]. To further add to the complexity of understanding the regulatory intimacies of protein function, many of the proteins have splice variants that form different isoforms and possess post-translational modifications (PTMs) that collectively can influence protein function[2]. Together there remains the complex task to identify and decipher the roles of each protein in the proteome; a technically demanding feat. By understanding the regulation and function of the proteome, we will better the understanding of human biology at the molecular level. This will provide the knowledge needed for development of new medical, therapeutic, diagnostic, prognostic and preventative measures and applications[2]. Genetic and epigenetic causes for human disease and disorders are constantly being identified. Epigenetic modifications are considered specific, stable, and potentially heritable modifications in gene expression that can occur in a biological organism, without alteration of the organisms genetic code[3].

*Processes of chromatin regulation*

Chromatin regulation provides a level of control over the timely expression of genes that are necessary for cellular function and homeostasis[4]. In this way, chromatin is regulated by several processes and modifications. These modifications can be narrowed down to three major mechanisms which include: (1) modification of DNA such as DNA methylation, (2) non-protein-coding RNA (ncRNAs) formation, and (3) protein and protein complexes that remodel chromatin structure, such as histone modifications[3,4]. DNA methylation to date has been the best-studied epigenetic modification. Generally, DNA methylation represses gene expression and occurs through DNA-specific methyltransferase enzymes at CpG sites (i.e., cytosine (C) and guanine (G) dinucleotides separated by a phosphate (p))[3]. DNA methylation can also serve as "epigenetic memory" that can persist through cell division. Secondly, ncRNA formation has been a more recent factor identified to influence cells epigenetics. ncRNAs such as microRNA (miRNA), transfer RNA (tRNA) and small interfering RNA (siRNA) have all been identified to regulate chromatin modifications, genetic imprinting, DNA methylation and transcriptional silencing[3,5]. Specifically, miRNAs have been reported to regulate over 50% of protein-coding genes in humans. Furthermore, ncRNAs are regulated by methylation of the CpG islands and/or histone modifications[5]. Finally, histone modification regulates gene expression through chromatin remodeling. Specific post translational modifications (PTMs) occur commonly on specific amino acid sites on histone tails. PTMs currently recognized to occur on histone tails include acetylation, phosphorylation,

ubiquitylation, methylation, ribosylation, and sumoylation [3,5]. These PTMs on N-terminal histone tails can play key roles in the regulation of DNA replication, recombination, repair, and transcription[6]. As complex and divergent epigenetic modifications are, the cellular epigenome can be dynamic and further influenced by genetic and environmental factors[5]. Likely a reflection of their critical regulatory roles in cell biology, an increasing number of epigenetic factors have been identified as genetic drivers of many human disorders and diseases[4]. Through the documentation and discovery of these epigenetic modifications and their mechanisms, it is possible that we may be able to identify molecular causes for specific diseases and disorders.

## Histone post-translational modifications

At the base of chromatin structure and regulation are histone proteins. There are five highly conserved histone proteins, H1, H2A, H2B, H3 and H4, with the latter 4 making up the core histone octamer[7–9]. These histones function as the anchoring blocks to wrap and package eukaryotic DNA into nucleosomes. These nucleosomes are the subunits that together are folded into chromatin structures[7–9]. Histones are dynamic structural elements in chromatin that are responsible for genetic transcription regulation[7–9]. Like most proteins, histone proteins are subjected PTMs, and these PTMs are important for regulation of both chromatin structure, chromatin remodelling, and genetic transcription[7,8]. Histone's are subjected to many PTMs including acetylation, ubiquitination, phosphorylation and methylation[7,8].. Histone proteins can have a complex

dynamic array of these four key PTMs that can exist on dozens of histone residues throughout the histone tails[7,8,10] (Fig. 1.1.). These PTMs take place on the tails of each individual histone and each PTM can have drastic effects on subsequent chromatin structure and remodelling, and genetic transactivation[7,8]. Moreover, histone PTMs can have affects on factors such as chromatin structure by manipulating histone-histone, protein-histone and DNA-histone interactions[7,8]. As a result, histone PTM's on specific amino acid residue(s) can have dramatic effects on chromatin and genetic regulation[7,9]. For example, its well studied how acetylation and methylation can both occur on the H3K9 substrate, resulting in opposite regulation of genetic transcription[10,11]. Furthermore, PTMs of histone proteins can effect the recruitment of other proteins or protein scaffolds that can mediated various other molecular function(s)[7–9]. As a result, the collective interplay of these various PTMs is referred to as the "histone code"[7–9]. As the four core histone proteins are also rich in positive amino acid residues they are generally, overall positively charged[8]. This relative positive charge is what allows them to associate with the relatively negatively charged phosphate backbone of DNA. As a result, the positively charged lysine and arginine residues are some of the most common targeted for PTM-mediated regulation of histone proteins[8].

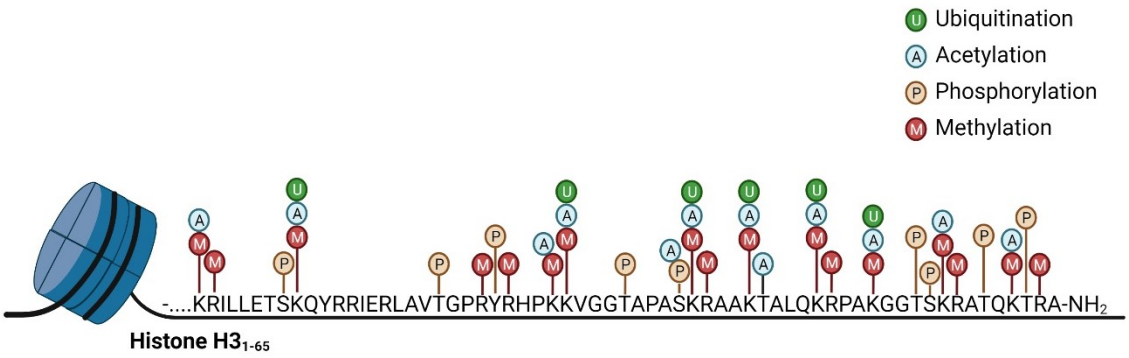



**Figure 1.1. Post-translational modifications on histone H3 protein.**

Demonstrating the complex dynamic array of the key four post-translational modifications that exist on histone proteins, using histone H3 as the prototypic example. Data is summarized from UniProtKB and literature[10,12]. Created with BioRender.com.

*Histone acetylation*

Acetylation of histone proteins was one of the first histone PTMs identified[13]. It involves the addition of an acetyl group ($CH_3CO$) to the positively charged lysine residues to the histone tails[13]. This histone PTM is reversible and therefore, is regulated by histone acetyl transferase (HATs) enzymes that add acetyl groups and histone deacetylases (HDACs) enzymes that remove the acetyl groups[13–15]. Specifically, it is the ε-amino group of the lysine residues, which through an acetyl CoA cofactor, is acetylated[13]. Generally, acetylation is a PTM mark associated with transcriptional activation[8,13,14]. This is due to the acetylation on lysine residues causing a decrease to the positive charge associated with histone proteins and therefore, can reduce their interaction with the negatively charged DNA[8,13,14]. As a result, nucleosomes become more flexible, permissive, and therefore can be accessed by other protein families resulting in gene transcription [8,13,14]. Both the HATs and HDACs families are diverse and even vary in subcellular localization. To date, in humans there have been about 30 HATs identified and classified into two classes, type A for nuclear localization and type B for cytoplasmic localization[13]. Opposingly, there are 18 HDACs identified and are classified into four classes based on sequence homology[16]. Generally for both HATs and HDACs, multi-protein complexes influence the substrate preference, substrate targeting and subsequent functions of the these acetyl-regulating enzyme families[14,15]. The interchange of acetylation mediated by HATs and HDACs influence changes in chromatin structure and states between genetic transcription and repression[15]. As a result, regulation of

histone acetylation has shown to be critical in general cellular development and homeostasis[13–15]. Dysregulation of acetylation has been linked to disease and disorder development such as cancer and X-chromosome related disorders[17].

*Histone ubiquitination*

Histone ubiquitination is one of the lesser abundant histone PTMs, but plays a critical role in signaling the DNA damage response[18,19]. Ubiquitin is itself a protein (76 amino acids) that is ubiquitously expressed and conserved throughout eukaryotes[20]. The ubiquitination of histone proteins is carried out by ubiquitin-conjugating enzyme (E2 class) and ubiquitin ligases (E3 class)[18,19]. Similar to other PTMs, ubiquitination of histone proteins is a reversible PTM. Ubiquitin-specific peptidases, called deubiquitinating enzymes (DUBs), are responsible for the removal of these PTMs[18,19]. Although all of the four core histone proteins have been shown to be ubiquitinated, H2A and H2B are the most highly ubiquitinated histones, with H2A being the first histone identified to be targeted[20]; approximately 5-15% of total H2A protein is known to undergo ubiquitination[18,19]. Interestingly, monoubiquitinated H2A has been related to gene transcription repression and monoubiquitination at K119 specifically on H2A has been associated with sites of DNA double-stranded breaks (DSBs)[18,19]. Specifically, it was shown that monoubiquitination at this site represses gene transcription via inhibition of RNA polymerase II transcriptional elongation[21]. Opposingly, monoubiquitination at K123 on H2B has been related to gene transcription activation and this site is also been shown to be with sites of

DSBs[18,19]. Apart from these residues, there is increasing evidence that other lysine residues in histone proteins can also be targeted by ubiquitin ligases. In H2A its been shown that K13 and K15 can be ubiquitinated, although its been demonstrated that there might be association of these sites with DNA damage response, the impact of genetic transcription activation or repression is still unknown[22].

*Histone phosphorylation*

Histone phosphorylation is one of the most abundant and researched of the histone PTMs. This primarily stems from the tails of each histone protein have various tyrosine, serine and threonine sites that are subject to phosphorylation[23]. Histone phosphorylation is facilitated through a variety of protein kinases and opposing phosphatases that catalyze the dephosphorylation of these hydroxyl-containing amino acids[23]. Histone phosphorylation is involved in many molecular functions surrounding histones such as DNA damage repair, gene transcription regulation and chromatin compaction[23–25]. For example, phosphorylation of Ser10/28 on histone H3 has been shown to be involved in chromatin structure regulation and chromatin compaction during cell division[23]. Although a variant, phosphorylation of H2AX-S139 has been well documented as a key facilitator in recruitment of proteins involved in DNA damage repair[23,25]. Similar to ubiquitination, many histone phosphorylation events can have an influence on histone interaction with protein complexes and effector proteins[23].

For example, H3-S10/T11/S28 have been shown to be linked with acetylation of H3, through a HAT (Gcn5) dependent acetylation[23].

*Histone methylation*

Chromatin modifications facilitated through dynamic changes in histone PTMs permit specific genes to be locally accessed and transcribed, as well as silenced and/or buried. In this way, histone methylation has become a primary focal point of modern histone and epigenetics research. Interestingly, both the positively charged lysine and arginine residues on histone tails can be targeted for methylation[26]. The addition of methyl groups to these residues do not themselves manipulate the overall charge of the histone proteins (i.e., charge shielding) to the degree that other histone PTMs do. Instead, they typically form dynamic interaction sites for methyl-binding domain containing proteins (MBD) to facilitate functional protein:protein interactions (more information in the following section(s)). In this way these MBD proteins are “readers” of histone methylation[27]. As such, lysine methyltransferase enzymes (“writers”) add methyl groups and lysine demethylase enzymes (“erasers”) remove set methyl groups[27]. These three enzyme families work cooperatively to regulate dynamic lysine methylation (Fig. 1.2). Although methylation is dynamic like other histone PTMs, such as acetylation, a recent increase in research revolving around histone methylation is due to the significance of associated mechanisms of histone methylation in human disease[28]. Primarily, lysine residues on histone tails can be methylated to either a mono-methylated, di-methylated or tri-methylated state. In

contrast, arginine residues can also be methylated, but towards a mono-or di-methyl state through a symmetric or asymmetric methylation of either terminal guanidino nitrogens[28]. Histone lysine methylation is unique compared to other histone PTMs as it has been associated with transcriptional activation or repression. Specifically, the level of activation or repression depends on the degree of methylation (mono-, di- or tri-methylation) on a specific lysine residue, out of a number of possible lysine residues residing within the histone tail[29]. For example, methylation on H3K9 (i.e., histone 3, lysine 9) and H3K27 have been frequently observed to be linked to gene silencing. The opposite has been observed for H3K4 and H3K36 methylation, which is commonly associated with active gene expression[29]. H3K4me3/2 (histone H3 lysine 4 di- or tri-methylated) states specifically have been identified as transcriptional start sites of actively transcribed genes[28]. This critical role of dynamic H3K4me in gene activation has made this mark critical to investigate in terms of disease development[29]. The potential significance of the H3K4 residue has encouraged researchers to investigate the role that all lysine methylation/demethylation facilitates in normal and diseased cells/tissues[28].

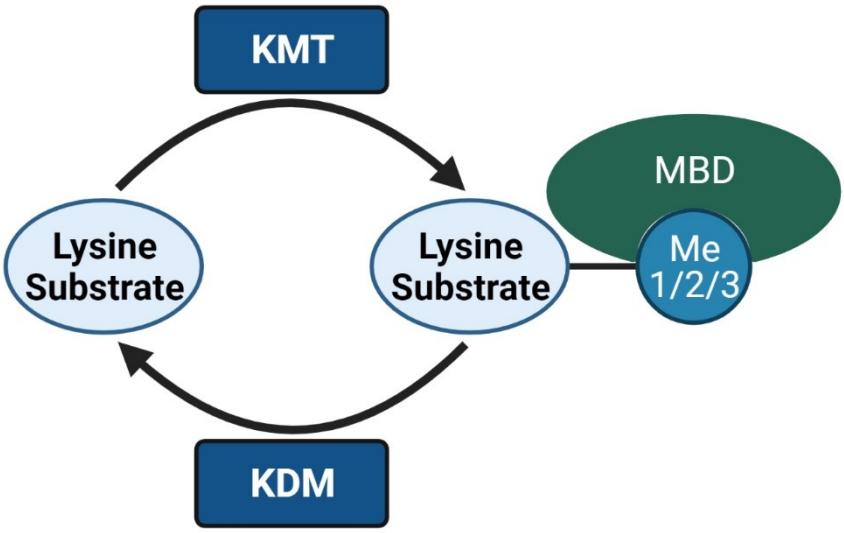



**Figure 1.2. Dynamic regulation of lysine methylation.** Visual representation of methyllysine regulation through lysine methyltransferase (KMTs), demethylase (KDMs) and methyl-binding domains (MBDs). Created with BioRender.com.

## Histone methyl regulatory proteins

*Histone methylbinding domains (MBDs)*

Beyond the simple addition of one (or more) methyl-modifications to a specific lysine, each methylation event can also constitute a specific signal that is recognized by highly evolved and conserved methyl-lysine binding domains (MBDs)[30]. Interestingly, these MBDs have been shown in many cases to not only recognize the methylated lysine residue, but also the surrounding amino acid sequence as to recruit MBDs to a specific methylation site[28,30]. There are four subdomain classes in MBD proteins that have been identified. These four classes include: (1) Tudor/Royal class, (2) Ankyrin repeat motifs, (3) WD40 repeat and (4) Plant homeodomain (PHD) finger domain [30,31]. The Tudor/Royal class consists of further sub-class members including the Tudor domains, chromatin organization modifier domains (chromodomain; CD), proline-tryptophan-tryptophan-proline (PWWP) domain and malignant brain tumor (MBT) methyl-binding proteins[30,31]. Collectively, these domains are commonly referred to as the methyl “readers” of the histone code[29,30].

Each MBD family is unique as to which residues and degrees of methylation they target[8,30]. The Tudor/Royal family are classified as so due to their conserved β-barrel secondary structure, also referred to as a Tudor barrel[30]. Generally, Tudor domains interact with Kme3/2 sites and are usually associated with catalytic active domains, as well there can be two of them forming Tandem domains[30]. Chromodomains are small 40-60 amino acid residue domains that are well attributed with a variety of roles surrounding genetic transcription activation,

repression, as well as cellular differentiation and cancer progression[30]. Interestingly, there are three sub-classes of chromodomains that exist: polycomb group (PcG) family, chromatin helicase DNA-binding (CHD) family and heterochromatin-associated protein 1 (HP1) family[30]. All three sub-families use a aromatic residue-dominant binding pocket in order to bind Kme1/2/3 methyl marks on H3K4/9/27 substrates[30]. As the PWWP family name suggests, methyl binding is dominated by the four aromatic-containing residues that exist in these MBD domains[30]. PWWP domains are usually associated with catalytic methyltransferase or PHD domains[30]. Finally, the MBT family has been shown to interact with Kme1/2 methylation degrees on histone H3 and H4 proteins[30,31]. Generally, binding of the MBT family drives repression of genetic transcription, which is how it was first identified in brain malignancies via supressing the tumor suppressing gene[32]. The MBT family is unique as typically these proteins contain 2-4 MBT domains, surrounded by other functional or substrate binding domains[30,31]. Although many domains may be present, the recognition and binding to a methyl group is typically dominated through only one domain's respective aromatic binding ring[30,31].

The ankyrin repeat motifs are one of the most abundant MBD domains that exist and is observed in >70,000 ankyrin repeat containing proteins in eukaryotes[30,33]. The most known ankyrin motif containing protein is the cytoskeletal protein named ankyrin itself, which contains 24 copies of this motif[34]. Generally, most ankyrin motif containing proteins possess 6 or less motifs and are characterized by a 30-50 amino acid residue repeat containing α-helices,

turns and β-turns[30,33]. These motifs have been observed to be involved in many molecular functions including genetic transcription activators, cell-cycle regulation, signal and ion transporters/transducers and as a result, are frequently altered in disease phenotypes[30,33]. This affect in so many biological functions is most likely a result of the observation that these motifs are present at the boundary of protein-protein interactions[30,33]. Similar as other MBDs, ankyrin motifs are commonly associated with other methyl-modifying enzymes. For example, it was shown that the ankyrin motif in G9a methyltransferase specifically binds the H3K9me1/2 substrate and helps facilitate its catalytic activity[35].

The WD40 repeat family of MBDs is one of the largest MDB families (<90,000 WD40 containing proteins), although only approximately 18 of these MBDs are associated with chromatin binding and regulation[30,36]. Generally, proteins that possess WD40 domains usually contain 4 or more repeats of the WD40 domains which are individually characterized by approximately 40-60 amino acids which are flanked with highly conserved tryptophan-aspartic acid (W-D) residues[36,37]. The WD40 repeats are unique as each WD40 unit forms an unusual β-propeller “blade” structure where each blade contains four anti-parallel β-sheets built into a toroidal micro-structure[36,37]. This structure is critical for histone H3 binding as binding occurs in the center of the repeated β-propeller[36,37]. This is not consistent in all substrate binding, as interaction with histone H4 occurs in more proximal regions of the WD40 domain[38]. WD40 repeat proteins have been shown to be involved in a variety of biological processes

including genetic transcription regulation, vesicle formation and trafficking regulation, cell division[36,37]. The WD40 proteins were some of the first MBDs identified to be sub-components of larger histone-modifying complexes[36]. Consistently with other MBDs, there are many histone methyl-modifying enzymes that contain WD40 repeats such as Set1 and MLL2/3/4 methyl transferase enzymes[39,40].

The PHD finger family are dynamic readers of the methyllysine histone code, with over 100 PHD finger domain-containing proteins in humans[30,41]. Structurally, PHD finger domains contain 50-80 residues that consist of a conserved zinc-binding motif, characterised by two antiparallel β-sheets which is anchored by the zinc ion[41,42]. Many PHD domain-containing proteins play roles in genetic transcription regulation and are present in many histone methyl modifying enzymes[41,42]. PHD domains are unique compared to other MBD's, as they have been shown to interact with methyllysine residues of all modification states (i.e Kme3/2/1), but also the non-methylated lysine residues[30,41,43]. One of the most widely investigated sub-classes of PHD finger domains was described to specifically bind to the H3K4me3 mark, whereas another PHD finger family specifically interacts with the H3K4me0 mark[30,41,43]. Recently, there has been evidence demonstrating other PHD finger domain-containing proteins do not bind only to the H3K4 substrate, but as well to H3K9me and, in yeast, to H3K36me[44,45]. Although there has been progress in defining the binding preference(s) of this MBD family, there are still many PHD fingers that have non methyl-lysine binding preference or unknown binding functions[42]. Dysregulation

of PHD finger domain functions have been implicated in a wide array of human disease and disorders[42,46,47]. Typically, mutations in the PHD fingers drive disease phenotypes as a result of dysregulation in cellular development and changes in histone methylation that effect gene transcription [42,46]. This paradigm becomes increasingly complex as some proteins contain 2 or more PHD finger domains; this includes methyllysine-modifying enzymes[42,46,48].

*Histone lysine methyltransferases (KMTs)*

Just as MBDs can be referred to as the “readers” of the histone methyl code, lysine methyltransferases (KMTs) can are the “writers” of histone methylation[27,49,50]. The histone KMT family of enzymes facilitate the addition of methyl group(s) to lysine residues. Currently, there have been more then 50 KMTs identified in humans and, similarly to MBDs, have demonstrated explicit substrate specificity[27,49,51]. The specific lysine sites and degree of methylation (e.g., mono vs. di- or tri-methylation) varies drastically between varying members of the KMT family[27,49]. There 2 large and diverse families of KMTs that are classified based on their conserved catalytic domain. These families are the *SU(VAR)3-9, enhancer-of-Zeste and Trihorax* (SET) domain family and the seven-β-strand (7βS) / DOT1-like domain family[49,50]. Generally speaking, most SET-domain KMTs function in tandem with an accessory domain, such as a MDB that guides in substrate specificity and preference[50,52]. Furthermore, many function in complexes with other histone-modifying enzymes[49,50].

In humans, there are currently 55 defined proteins that contain a SET domain, half of which are known active KMTs that target histone lysine residues for methylation[49,50]. There are sub-families of the SET domain containing KMTs including SET1, SET2 SUV39, SMYD, EZH, and RIZ sub-families[51,52], although some classification of sub-families have varied in literature[49,50]. All SET domain containing KMTs facilitates their catalytic activity towards their substrates by methylating the amino group of a lysine residue using the same co-substrate: *S*-adenosyl-L-methionine (SAM)[52]. Members of the same sub-family of KMT's generally target the same or a variety of specific histone lysine substrates[49–52]. For example, the SET1 family primarily targets the histone H3K4 residue, a histone mark commonly observed to be associated with transcriptional activation[49,52]. Other sub-families such as EZH family, all target the H3K27 substrate which is a mark associated with euchromatic silencing[49,50,52].

The 7βS family of KMTs is technically larger and more diverse then the SET family of KMTs; currently known to contain approximately 150 members in humans[49]. Unlike the SET family which primarily, but not exclusively, targets lysine residues, a variety of 7βS-containing proteins have been shown to methylate many substrates containing lysine, arginine and histidine residues as well as DNA, RNA and other biomolecules[49,50]. The 7βS family has diverged from the SET family to such a degree that the human DOT1-like protein does not even contain a SET domain[50,53]. It is characterized more similar to that of a arginine methyltransferase and facilitates its methylation via a similar mechanism[50]. Given this insight, the DOT1L is the only KMT that targets the H3K79 substrate; a

histone mark associated with gene activation[53]. Furthermore, DOT1L is uniquely capable of methylating this substrate to all three states of methylation[50,53]. As a result of being implicated in a variety of cellular processes, it has been well documented that this KMT plays a role in disease/cancer progression[54].

KMTs have been well documented to be regulators of cellular function, as well as drivers of human disorders and disease when dysregulated[27,49]. This is of course partially due to the fact that methylation of these key histone lysine residues have dramatic influence on activation or repression of key gene expression[49,50]. In recent years, there has been increasing evidence that both SET and 7βS families of KMTs can also methylate non-histone proteins[55–57]. This is critical knowledge and should be further investigated as targeting non-histone proteins for methylation may help us further understand the molecule mechanises driving human disorders and diseases. This becomes increasingly apparent as p53, retinoblastoma protein 1 and STAT3 have all been demonstrated to be functionally regulated through dynamic lysine methylation[55,57].

*Histone lysine demethylases (KDMs)*

Just as histone KMTs are the “writers” of the histone code, correspondingly the histone lysine demethylase (KDMs) family of enzymes are the “erasers” of histone methylation. These enzymes demethylate through amine oxidation, deamination and hydroxylation[29]. There are currently two known classes of KDM enzymes: flavine adenine dinucleotide (FAD)-dependent amine

oxidases and the Jumonji C-terminal domain (JmjC) containing family of demethylases, which contain over 30 members[58–61]. The first of these classes discovered was the FAD-dependent amine oxidases, with lysine-specific demethylase 1 (LSD1, KDM1A) being the first histone demethylase identified[61]. The LSD1 family specifically targets mono- and di-methylated lysine residues and LSD1 was shown to target the histone H3K4me1/2 substrate for demethylation[61]. It catalyzes these reactions using cellular FAD and $O_2$ to help facilitate the electron transfer in the FAD-dependent amine oxidase reaction[61] (Fig. 1.3). The second histone KDM class is the JmjC-containing family of demethylases, which is the larger of the two classes with over 7 sub-families: KDM2 (JHDM1), KDM3 (JHDM2), KDM4 (JHDM3), KDM5 (JARID), KDM6 (UTX/UTY), KDM7 (JHDM1D), KDM9 (dpy-21), PHF2/PHF8 and JmjC domain only[58,62]. The JmjC domain belongs to the superfamily of $Fe^{2+}$-dependent dioxygenases, conserved from yeast to humans[58,62].

Identification of these JmjC-containing sub-families began in 2006 with identification of KDM2A/JHDM1A which was shown to target the H3K36me1/2 substrate[63]. Since then, a wide array of JmjC-containing KDMs have been identified and all KDM2-9 sub-families target all three methylation degrees on a variety of histone lysine residues[58,62]. All JmjC-containing KDMs use $Fe^{2+}$ and α-ketoglutarate and as cofactors in the presence of $O_2$[58,62]. Through this reaction, the methyl group is converted to a hydroxymethyl group, which upon completion of the electron transfer produced the products of formaldehyde, succinate and $CO_2$[58,62] (Fig. 1.3). Specifically, the $Fe^{2+}$ ion helps facilitate the hydroxylation of

the protein substrates[51,58,62]. Similar to that of the histone KMTs, majority of the JmjC-containing KDMs possess MBDs that aid in substrate recognition, binding and regulation of the KDMs activity[43,58,62,64,65]. As a result of regulation of methylation of key histone marks associated with genetic transcription activation and repression, it is to no surprise that this class of KDMs have been well documented to be critical in general cellular homeostasis[58,62]. Parallel to histone KMTs, studies recently have suggested histone KDMs could serve as oncogenic or tumorigenic drivers due to their adherent ability to manipulate and sculpt histone methylation[28,29,51]. As this becomes increasingly apparent as more research is being done, that it is critical to investigate key histone marks and the KMT/KDM enzymes that target it[66].

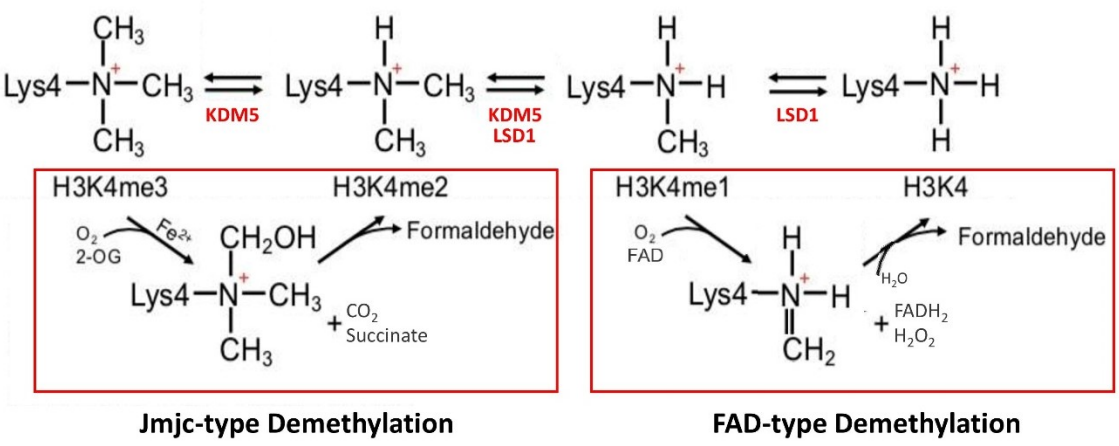



**Figure 1.3. Biochemical mechanisms of JmJC-catalyzed and (FAD)-dependent amine oxidase-catalyzed lysine demethylation.** H3K4 substrate shown as prototypic example of histone substrate. KDM5 family used as example of JmjC-containing lysine demethylase and LSD1 used as example of (FAD)-dependent amine oxidase lysine demethylase.

*KDM demethylase enzymes and non-histone substrates*

As KDM enzymes have been identified to influence disease progression, there also growing evidence that their disease-related function(s) cannot be solely described through histone (de)methylation. This highlights the growing realization that KDM enzymes are also capable of demethylating non-histone proteins, and that the collection of these substrates help to fully describe the function of the KDM. KDM1A/LSD1 was the first lysine-specific demethylase to be identified to target non-histone proteins[67]. It was demonstrated that KDM1A removes both the mono- and di-methylation states on p53 at lysine 370 (i.e., p53(K370me1/2))[67]. This finding was critical as it not only demonstrated that lysine-specific demethylases can target ubiquitously expressed proteins, such as histones, but as well they can target proteins such as p53. Furthermore, this study demonstrates that that the removal of the methyl groups on p53 K370 dynamically regulated p53 activity. This ultimately leads to critical changes in downstream p53 pathways such as DNA damage repair and p53 induced apoptosis[67]. Although KDM1A belongs to the LSD family of lysine demethylase, there is new evidence to support that members from the JMJD-containing family of lysine demethylases may also target non-histone proteins. KDM3A was shown to demethylate p53(K372me1), which has been hypothesized to lead to chemoresistance through an enhancement of the DNA repair response[68]. This is critical as these study demonstrates the not only ability of both classes of KDMs to target non-histone protein, but to dynamically effect the activity of these non-histone proteins leading to the further progression of cancer[6].

*KDM5 demethylase sub-family*

Recently there has been a large increase in evidence obtained through both primary tumors and cancer cell models that suggests a role for the KDM5 sub-family as oncogenic drivers[69]. This revelation has made it clear that it is now critical to investigate the basic biochemistry of this family to provide insight into how it achieves its disease function. As this sub-family was only discovered in the last 15 years, there has been minimal research completed outside the scope of histone H3K4me3/2 demethylation[48,70]. Only recently, has a second substrate for KDM5 been reported, the KDM5B substrate histone H2BK43me2[71]. Although its conservation within the KDM5 family is not yet clear, this provides a clear indication that KDM5 is capable of demethylating substrates that deviate from the histone H3K4 sequence.

The KDM5 sub-family is well documented to catalyze the removal of methyl groups from tri- and di-methylated lysine 4 of histone 3 (H3K4me3/2)[48]. As transcriptional activity is correlated with H3K4me3 and possibly H3K4me2, removal of the methyl groups by KDM5 is believed to trigger transcriptional repression[69]. The KDM5 sub-family has four members: KDM5A, KDM5B, KDM5C and KDM5D. The four members have a high degree of homology in domain organization and possess common homology in sequence[48,72]. Phylogenetically speaking KDM5A/KDM5B and KDM5C/KDM5D are closely related[73] (Fig. 1.4). The KDM5 sub-family of demethylases is a unique family as the catalytic Jumonji domain is separated by an atypical insertion of a DNA-binding AT-rich interacting domain (ARID) and a PHD MBD (PHD1). This has resulted in the Jumonji

domain being separated into two sub-domains, an N-terminal (JmjN) and C-terminal (JmjC). KDM5 further possesses two (KDM5C + KDM5D) or three (KDM5A + KDM5B) PHD domains, a C5CH2-Type zinc finger domain and a conserved (but uncharacterized) PLU-1 domain[48]. A schematic representation of the domain architecture of KDM5A, KDM5B, KDM5C and KDM5D can be seen in Figure 1.5. It has been demonstrated that removal of the ARID and PHD1 domains is required for bacterial expression and are expendable for *in vitro* enzymatic activity[48,72]. Generally, KDM5 substrate recognition has been shown to be determined through the PHD1 and PHD3 (KDM5A/B) MBDs, whereas the JmjC domain solely facilitates the catalytic demethylation of substrate(s)[43,48,64,65,74–76]. Although the presence of a methylated lysine residue is a key factor in PHD3 binding and subsequent JMJD demethylation, previous studies analyzing PHD1 binding have demonstrated that KDM5A is able to accommodate substrates with sequences that vary from histone H3K4[74]. This type of study has not been performed on any other KDM5, although its been well documented that KDM5B binds to its substrates distinctly through its PHD1 and PHD3 domains[65].

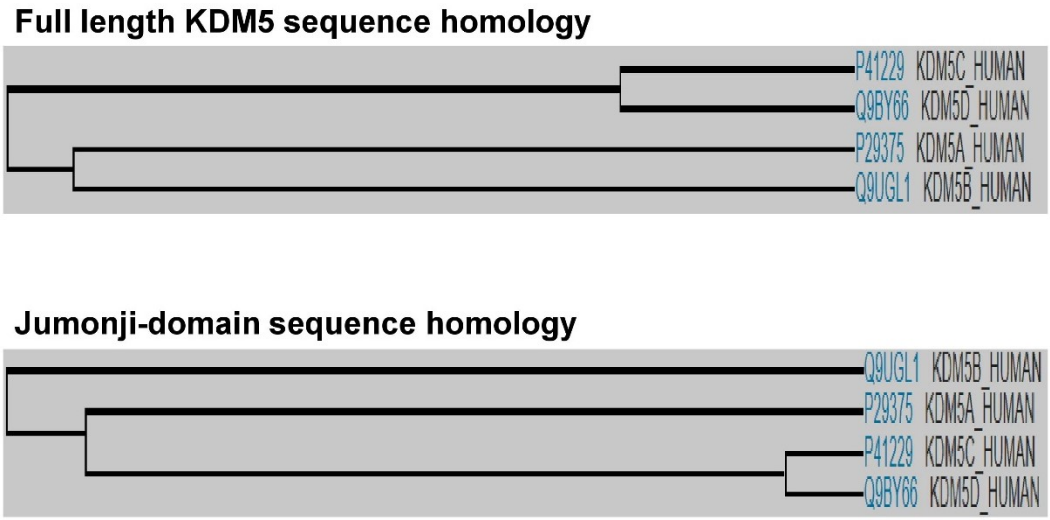



**Figure 1.4. KDM5 sub-family sequence homology.** Dendrogram representing full length KDM5 sequence homology and Jumonji-domain sequence homology. Analysis completed through *ClustalΩ* multiple sequence alignment[77].

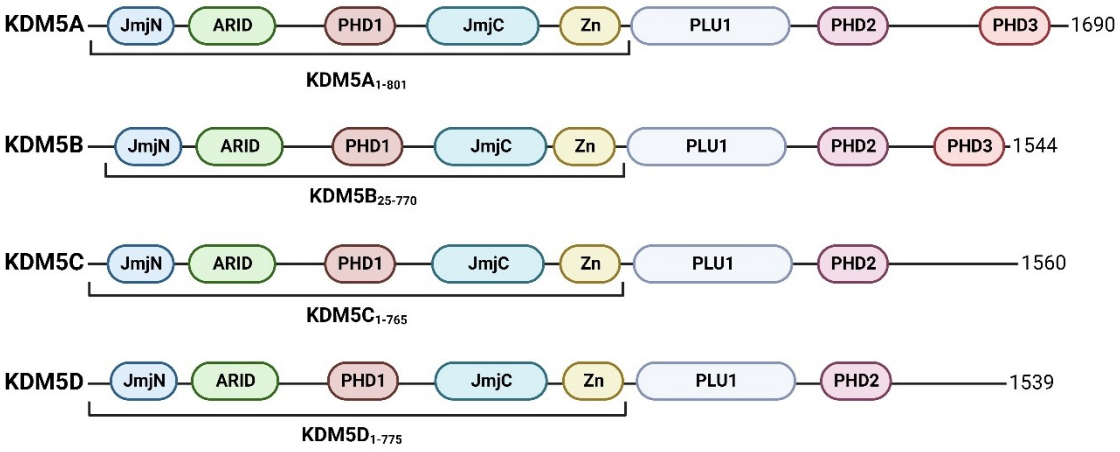



**Figure 1.5. Domain architecture of KDM5 demethylase enzymes.** Schematic representation of domain architecture of KDM5 sub-family of demethylase enzymes. JmjN and JmjC represent the two Jumonji catalytic sub-domains; ARID (AT-rich interacting domain) is a DNA binding domain shown to interact with CCGCCC sequence; PHD1-3 (plant homeodomain) are methyl binding domains, KDM5A/B have 3 PHD's whereas KDM5C/D only have 2 PHD domains; Zn (zinc finger domain) is a C5CH2-Type zinc finger domain; PLU-1 is a conserved but uncharacterized domain. Brackets represents the constructs used in this thesis[48]. Created with BioRender.com.

*KDM5A*

KDM5A (also known as retinoblastoma binding protein-2 (RBP2)) is an important regulator of cell cycle and has been hypothesized to be a key player in development of tumors[48,78]. KDM5A is over-expressed in a variety of cancers including gastric, lung, hepatocellular, colorectal and cervical cancers[78]. KDM5B as well, regulates cell cycle progression but through repression of cell cycle checkpoint genes[78]. KDM5B has also been demonstrated to repress anti-proliferative and tumor suppressor genes[48] and has been shown to over expressed in breast, prostate, bladder, gastric, lung, cervical and some form of leukemias[78]. KDM5C has been shown to be expressed the most in neuronal tissue[48]. KDM5C mutations have been shown to be associated with Huntington's disease and X-linked mental retardation and autism[48,69]. KDM5C overexpression was also observed in renal cell carcinoma as well in hypoxic tumors[29,69]. KDM5D is 85% identical to KDM5C and has been identified to be implicated in prostate cancer[69,72]. The clear connection between the KDM5 family of enzymes and their frequently observed implications in a variety of disease and disorders, including cancer progression, has demonstrated that it is critical to investigate their cellular mechanisms.

KDM5A was first identified to have an effect on tumorigenesis as the tumor-suppressive activity of retinoblastoma (RB) protein was dependent on KDM5A isolation[79]. RB inhibits cell cycle progression by inhibiting S-phase entry as well as promotes differentiation and senescence[79]. It is the ability of the RB protein to bind with KDM5A that was shown to inhibit differentiation and

senescence, as inhibiting KDM5A with siRNA was shown to re-initiate these processes[79]. Since this discovery, the majority of research has gone into epigenetic regulation by KDM5A in cancers. In breast cancer, KDM5A has been shown to promote metastasis as a positive regulator of metastasis genes in estrogen receptor negative (ER-) breast cancers[80]. Studies have also shown that KDM5A is overexpressed in gastric and cervical cancer cells. In these cells, KDM5A triggers cellular senescence through the binding of gene promoters of cyclin-dependent kinase inhibitors p16, p21 and p27 through demethylating H3K4me3 at these promoter locations[81]. Furthermore, KDM5A has bene demonstrated to play an epigenetic role in acute myeloid leukemia (AML) as well as lung, hepatocellular, colorectal cancer[29,48,78]. Generally, KDM5A mechanistically promotes cell growth and inhibits senescence and differentiation through it's ability to demethylate H3K4me3 back to a mono-methyl state[29]. As cancer genomics have demonstrated, a decrease in global gene expression in cancer cells due to overexpression of KDM5A can result in cancer progression by a variety of pathways.

Although there has been progress in understanding KDM5A regulation in cancers, there has been very little research examining potential oncogenic substrates in which KDM5A may act upon. This is surprising due to the knowledge that JMJD-containing KDM enzymes have been demonstrated to have activity on non-histone proteins[68]. KDM5A is generally shown to be regulated by its PHD1 reader domain through a positive-feedback mechanism[43]. The PHD1 domain, separating the JmjN/JmjC catalytic domains, has been

demonstrated to bind to unmodified H3 peptide[43]. The PHD1 binding to the histone H3 tail stimulates the catalytic JmjN/JmjC removal of methyl marks from the H3K4me3 substrate[43]. This study demonstrates how KDM5A appears to function through coupling activity between the reader (PHD1) domain and catalytic (JmjN/JmjC) domain. In murine models studying AML progression, KDM5A was identified to form a fusion protein with a nuclear pore complex protein NUP98[82]. This fusion protein was shown to occur through the PHD3 domain of KDM5A, similar to what has been observed to occur for KDM5A and H3K4me3 substrate[82]. Although this doesn't demonstrate the ability of KDM5A to be catalytically active on a non-histone lysine-containing protein substrate, it validates the potential of reader (PHD) domains to bind and interact with non-histone, lysine-methylated proteins. Such documented KDM5A interactions with non-histone proteins, as well as other proposed interactions with proteins involved in oncogenesis, such as Myc proteins (proto-oncogene proteins), makes KDM5A activity on non-histone proteins timely to investigate[83]. These interactions, in addition to the previously identified KDM1A/KDM3A-p53 interactions, demonstrates the high likelihood of KDM5A being able to demethylate non-histone proteins. Whether the KDM5A-mediated removal of methyl groups on non-histone proteins lead to oncogenesis or further tumorigenesis is yet to be determined.

## Methods to study demethylase activity

In recent years there has been an increasing number of methodologies developed that examine and characterize how histone modifying enzymes molecularly target, bind, and have catalytic activity towards histone and non-histone proteins. These include immuno-based techniques, radioactive and coupled enzyme assays, peptide binding arrays, and mass spectrometry-coupled reactions.

*Immuno-based techniques*

As reliable PTM-specific antibodies have become increasingly available commercially, a variety of immuno-based techniques to study histone methylation has been developed relying on a range of biophysical techniques. More specifically, site-specific and pan-methyl antibodies are available and allow for the analysis of lysine methylation[84]. As some antibodies are very specific and can discriminate between varying degree's of methylation[84], the reality is that many methyl-specific antibodies have been demonstrated to show significant issues in specific epitope binding and specificity between methyl-states (e.g., Kme2 vs. Kme3)[84,85]. These binding discrepancies are usually caused by relatively similar $K_d$ values between methyl-states or the presence of neighbouring PTMs that can negatively affect binding[84]. Recently, there has been an increase in the use of homogenous antibody-based screens to study KDM reactions. Bead-based techniques such as amplified luminescent proximity homogenous assay (ALPHA) and time-resolved fluorescence resonance energy transfer (TR-FRET) are two

common assays used in KDM activity studies[86,87]. These bead-based approaches can be advantageous as they can be signal sensitive, provide a large library size which allows for high-throughput screening[88,89]. The price of beads, a relatively large sample size, and labor-intensive initial screening and preparation are significant disadvantages to these techniques[87,90]. Furthermore, compounds that bind singlet oxygen can in interfere with ALPHA results[87].

*Radioactive and enzyme-coupled assays*

Radioactive and enzyme-coupled assays have been developed that all work on the principles of measuring either the formation of reaction products or depletion of substrates and co-substrates. Depletion and conversion of radioactively labelled $^{3}$H-SAM or $^{3}$H-labeled methylated histone substrates has been a well-documented method and the gold standard to detect active histone methylation or demethylation towards specific substrates[84,85,87]. However, radiolabeled substrate preparation, logistical difficulty in detecting weak $^{3}$H beta emission, and long reaction and data collection times are major limitations of these assays[84,85,87]. Given the commercial availability, enzyme coupled assays have been another assay system used to study methyl-related enzyme characteristics. Most of these types of assays are based on coupling reaction products or substrates into a detectable fluorescent or luminescent signal[91,92]. Formaldehyde dehydrogenase (FDH) coupled reactions are a common assay used to study lysine demethylation reactions, as formaldehyde is a by product of the reaction[84,92]. This type of coupled assay is ideal for individuals looking to

derive enzyme kinetics, as these assays are both sensitive and permit real-time measurements of activity[92,93]. Disadvantages to these assays include the presence of reaction artifacts and absorption/fluorescence interference and quenching, all of which can lead to false positive and negative results[87,94].

*Peptide Arrays*

The use of peptide arrays has been a well-established and high-throughput method to study enzyme-substrate binding or protein-protein interactions[90,95]. Peptide arrays are advantageous as they provide relatively quick and robust way to study a large number of peptide substrates[90,95]. This is ideal when looking to study a specific binding region or to identify new interactions for a specific protein of interest[90]. Furthermore, when systemically designed to iteratively explore tolerable peptide mutations, peptide arrays can elucidate amino acid patterns of enzyme-substrate or protein-protein interactions[96]. This has been shown to be useful when studying enzymes within a common family or class, as binding and catalytic domains are commonly observed to be evolutionarily conserved[75,97]. If a substrate is well-established, peptide arrays can also be semi-quantitative as relative binding of an unknown substrate when compared to a known substrate interaction can be relatively determined[96]. The largest disadvantage of these arrays is that the peptides printed on these arrays are only able to represent  a small region of a protein[95,96]. As a result, the interactions discovered need much further validation as protein folding, protein charge and displacement of key residues that facilitate appropriate binding may

not be encompassed in the peptide immobilized on the peptide array[95]. Furthermore, an enzyme interaction with a peptide substrate through only binding is only an approximation. As enzyme binding is not synonymous with an enzyme catalytically targeting a substrate, results can suffer and include high false positive rates[96].

*Mass spectrometry*

Mass spectrometry (MS) coupled assays have been a commonly used method to derive quantitative activity data for KDM and KMT enzymes. This is possible as a direct detection of a 14 Da change in product mass is attributed to a gain or loss of a single methyl group[87]. As reactions and their products can be directly input into a MS quadrupole, or via an initial chromatography system, this technique has been proven to be highly specific and sensitive to changes in KDM/KMT product mass resulting from changes in methylation[98,99]. Advantages to MS include the evasion of common signal quenching and non-specific interactions that other assay techniques contain, as well as it does not require methyl-specific antibodies[87,98]. When screening substrates or inhibitors, the low throughput and specialized instruments are major disadvantages to MS-coupled techniques[100]. However, this problem can be somewhat mitigated upon by using multiplexing strategies, which analyze samples from several pooled reactions simultaneously using “mass-tagged” substrates[100].

## Thesis Hypothesis

My research is aimed towards the investigation of how KDM5 demethylases actively target methylated proteins, beyond histone H3K4me3/2. Specifically, I am interested in how KDM5 demethylases target non-histone proteins, and what influence these KDM5-regulated demethylation events have on non-histone protein activity to provide insights into cell function and cellular homeostasis. **My hypothesis is that the lysine-specific demethylase KDM5 family is capable of actively demethylating non-histone protein substrates, including substrates that help facilitate the function(s) of KDM5 beyond chromatin regulation.** To address this hypothesis, I have outlined four specific aims to my research:

**Aim 1)** Develop a systematic approach to predict KDM substrates *in vitro.*

**Aim 2)** Characterize substrate selection preference of the KDM5 family and identify potential non-histone substrates.

**Aim 3)** Annotate *in vitro* KDM5 demethylation activity of select non-histone protein substrate(s) to identify unique and conserved demethylase activity within the KDM5 family.

**Aim 4)** Determine how KDM5 demethylation activity influences cellular function of select non-histone protein substrates.

# Chapter 2

# Evaluation of Jumonji C lysine demethylase substrate preference to guide identification of *in vitro* substrates

**Matthew Hoekstra[1,2], Anand Chopra[1,2], William G. Willmore[1,2], Kyle Biggar[1,2]**

[1]Institute of Biochemistry, Carleton University, Ottawa, ON, K1S 5B6, Canada

[2]Department of Biology, Carleton University, Ottawa, ON, K1S 5B6, Canada

## Published manuscript

Hoekstra, M., Chopra, A., Willmore, W.G., Biggar, K.K. (2022). Evaluation of jumonji C lysine demethylase substrate preference to guide identification of in vitro substrates. ***Star Protocols*** 3(2), 101271.

## Chapter Summary

As previously mentioned in Chapter 1, despite the apparent influence the KDM5 family has on a variety of human disorders and disease the only known substrate for the KDM5 family is H3K4me3/2, while H2BK43me2 has only yet been validated for KDM5B[48,71]. As a result, studies have primarily revolved around the ability of KDM5 to regulate gene expression through removal regulation of the histone H3K4me3/2 modification[69]. Although thoroughly investigated, the defined mechanism through which KDM5 family members specifically promote tumorigenesis continues to elude scientists. One hypothesis that is supported experimentally is that the KDM5 represses important tumor-suppressing genes by reducing active histone H3K4me3 marks at key promoters[69,78,101–104]. Although this does shed light on a potential mechanism, there remains little evidence that the observed oncogenic influence of KDM5 is attributable solely to changes in histone methylation[6]. As there is growing evidence that KDM5 demethylases are capable of demethylating non-histone proteins, our overarching hypothesis is that this KDM family is demethylating critical non-histone proteins that are essential for general cellular homeostasis.[6,56,69]. If true, it is critical to further investigate how new substrates could help reveal novel cellular function(s) of KDM5 in healthy and diseased cells[69,93].

To discover novel potential non-histone substrates for the KDM5 family, we first had to develop a method to predict or identify new substrates. We determined the optimal way to accurately predict novel substrates was to fully

understand KDM5 substrate preference. At the initiation of this thesis, a method to monitor KDM substrate preference had yet to be developed. Currently, most methods regarding KMT (but not KDM) substrate preference or specificity has been surrounded around the use of peptide array-based methods[84,90]. Although these techniques can offer high-throughput screening, these arrays require an added label to be able to visualize signal. This is typically a $^{3}H$ radioisotope label donated by $^{3}H$-SAM for KMT assays[84]. A similar approach would not be possible for KDMs as these enzymes would need to remove a 'label' (i.e., a negative signal) introducing logistical difficulty in detection and analysis. Herein, we propose an approach to study substrate preference that is evolved from a theoretical combination of peptide array-based permutations assays, as well as in solution succinate-based detection of KDM activity. Our method utilizes a library of mutated peptides that are systematically mutated from an established substrate peptide, histone H3K4me3. Using this highly specialized peptide library, we can assess position-specific amino acid preferences displayed by any KDM. The result is a semi-quantification method that can measure relative demethylation activity towards peptide substrates. Through observing their activity towards each of the systematically mutated substrates, in relative comparison to the established WT substrate peptide, we can conclude which amino acid mutations resulted in a relative increase or decrease in KDM activity. Ultimately from this data, a KDM recognition motif is produced that be used to determine how likely a methyllysine-containing protein in the methyl proteome is to a be a substrate of the KDM enzyme in question.

*Investigation of formaldehyde-based detection assays*

To start investigating a detection assay to monitor JmjC-KDM activity, we began investigating the feasibility of directly measuring demethylation-coupled formaldehyde production. Indeed, other groups have had success measuring radiolabeled formaldehyde production[85]. We first evaluated this approach using various organic chemicals that would readily react with formaldehyde to produce a product that could be spectrophotometrically detectable. We first investigated (1) 4-amino-3-penten-2-one (i.e., Fluoral-P; Appendix A, Fig. A1) and (2) methyl acetoacetate in the presence of ammonium acetate (Appendix A, Fig. A2), as both have been shown to readily react with formaldehyde to form a detectable product[105–107]. Although both these methods work fine in large volumes and for high [formaldehyde], when applied to our system the product signal was problematically low. As a result, this led to large error and both assays and its components did not translate well to a low-volume microplate assay format that is necessary for screening.

The next chemical-based detection assay we attempted was 2-methyl acetoacetanilide. 2-methyl acetoacetanilide in the presence of ammonium acetate will react with formaldehyde to yield a fluorescent compound (excitation and emission of 360nm and 460nm, respectively)[108] (Appendix A, Fig. A3). We first began with establishing a formaldehyde standard curve (Appendix A, Fig. A4A) to determine detection limits, followed by assessing KDM5A$_{1\text{-}588\Delta AP}$ activity towards the histone H3K4me3 peptide substrate via a 2x KDM5A$_{1\text{-}588\Delta AP}$ dilution curve (Appendix A, Fig. A4B). Although we do see a dose-responsive signal from

this assay, it consistently appeared to suffer from limitations of sensitivity, large error, and a general lack of robustness.

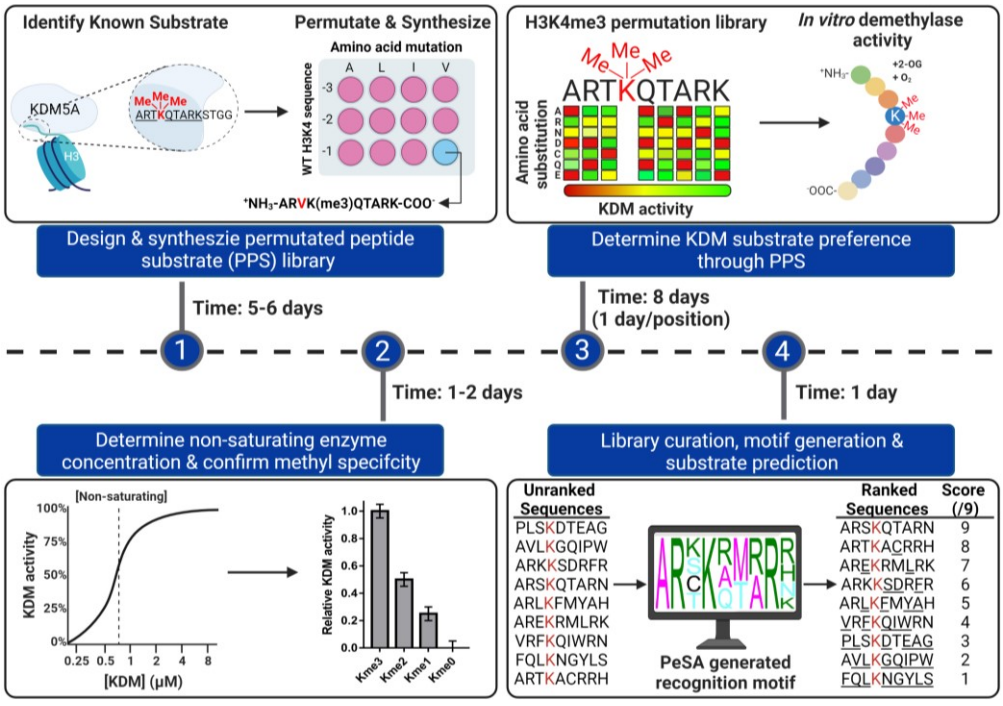
*Investigation of a succinate-based detection assay*

In our efforts to find a robust and sensitive assay to use to study KDM substrate preference, we discovered the commercially available *Promega Succinate-Glo*™ assay that monitors succinate production (a product of JMJD/KDM demethylation)[109] (Appendix A, Fig. A5). We performed a succinate-standard curve to determine the optimal reaction conditions to detect demethylation (Appendix A, Fig. A6A). Following this, we then assessed $KDM5A_{1\text{-}588\Delta AP}$ activity via a 2x $KDM5A_{1\text{-}588\Delta AP}$ dilution curve (Appendix A, Fig. A6B). Once we confirmed the ability of this luminescent assay to accurately quantify relative $KDM5A_{1\text{-}588\Delta AP}$ activity, we then sought to combine this detection assay with our specialized permutated peptide substrate (PPS) library to a fuller-length recombinant $KDM5A_{1\text{-}801}$ enzyme. Although predominantly interested in applying our method to the KDM5 family, we wanted to investigate whether our method could be applied to any JmjC-containing KDM. As a result, we also applied this method to $KDM3A_{515\text{-}1317}$. **The following sections are from my published methods manuscript in *Star Protocols- Cell Press* that describes the full method and pipeline we applied to assess $KDM5A_{1\text{-}801}$ and $KDM3A_{515\text{-}1317}$ substrate preference and to predict substrates. Importantly, this is the first systematic exploration of substrate preference for any KDM enzyme.**

## Manuscript Summary

Jumonji C (JmjC) domain-containing lysine demethylases (KDMs) catalyze the removal of methyl groups from the $\varepsilon$-amino group of lysine residues. Within the realm of lysine methylation, the discovery of lysine methyltransferase (KMTs) substrates has been burgeoning due to established systematic substrate screening protocols. Here, we describe the first protocol enabling the systematic identification of JmjC KDM substrate preference and *in vitro* substrates. This protocol provides a gateway for deepening our understanding of how JmjC KDMs impart their biological function(s).

## Graphical abstract




This graphical abstract was reprinted with permission of *Star Protocols – Cell Press.*

## Before you begin

The protocol below describes the specific steps for determining a substrate recognition motif for the KDM5A demethylase, as well as how to leverage this information for *in vitro* substrate discovery. Results for KDM3A are shown alongside KMD5A to highlight (1) how KDM substrate preference deviates between enzymes and (2) that the method is not solely applicable to a single enzyme.

### *Theoretical considerations*

The procedure for determining JmjC KDM substrate recognition motifs involves monitoring enzyme activity towards a library of peptides representing systematic mutations from the canonical substrate sequence (i.e., peptide permutation library). Analogously, this approach has been established for KMTs [110–114]. Thus, this methodology is limited to those JmjC KDMs with a known peptide substrate. Many methyl-modifying enzymes have established substrates within the histone code [115]. For example, the canonical histone substrate for the KDM5 family of enzymes is histone H3 trimethylated at lysine-4 (i.e., H3-K4me3).

### *Practical considerations*

To note, the methodology begins assuming a pure source of recombinant enzyme is available. Recombinant KDM5A and KDM3A enzymes were expressed and purified as described [92,116]. Furthermore, although our example monitors JmjC KDM activity via detection of succinate (i.e., turnover of the 2-

oxoglutarate cofactor), there are other methods for detecting *in vitro* JmjC KDM activity. This includes other techniques monitoring cofactor turnover and formaldehyde formation that would also be viable alternatives for detecting enzyme activity [109,117–119].

## Design permutated peptide substrate (PPS) library sequences

Permutation libraries consist of peptide sequences whereby a given residue position is mutated to all other naturally occurring amino acids while leaving the remainder of the sequence unaltered. Systematically performing this for multiple residue positions enables the assessment of amino acid specificity of peptide-protein interactions over a defined window. The specificity of numerous KMTs has been mapped via this approach and the epitope is generally defined by the residues occurring directly proximal to the modification site. Structural analysis of JMJD14, a plant KDM5 enzyme, found that residues directly proximal to the H3-K4me3 target site are important for substrate recognition [120]. Thus, the permutated window consisted of residues -3 to +5 amino acids relative to the H3-K4me3 methylation site (ART**K**QTARKSTGGKA; K4 position bold, permutated window underlined).

**Timing: 1-2 h**

1. Retrieve peptide sequence known to permit demethylase activity (e.g., the H3-K4 sequence [ART(Kme3)QTARKSTGGKA] was used for KDM5A).

2. At each residue position within the desired window (e.g., -3 to +5 relative to the methylation site), generate 19 other sequences whereby the wild-type residue is exchanged to another naturally occurring amino acid (Table 2.1 shows an example for the -3 position only).

**Table 2.1. Example of permutation of the H3-K4me3 peptide sequence at the -3 position (underlined).**

| Mutation | Peptide sequence |
|---|---|
| A(WT) | ART(Kme3)QTARKSTGGKA |
| R | RRT(Kme3)QTARKSTGGKA |
| N | NRT(Kme3)QTARKSTGGKA |
| D | DRT(Kme3)QTARKSTGGKA |
| C | CRT(Kme3)QTARKSTGGKA |
| Q | QRT(Kme3)QTARKSTGGKA |
| E | ERT(Kme3)QTARKSTGGKA |
| G | GRT(Kme3)QTARKSTGGKA |
| H | HRT(Kme3)QTARKSTGGKA |
| I | IRT(Kme3)QTARKSTGGKA |
| L | LRT(Kme3)QTARKSTGGKA |
| K | KRT(Kme3)QTARKSTGGKA |
| M | MRT(Kme3)QTARKSTGGKA |
| F | FRT(Kme3)QTARKSTGGKA |
| P | PRT(Kme3)QTARKSTGGKA |
| S | SRT(Kme3)QTARKSTGGKA |
| T | TRT(Kme3)QTARKSTGGKA |
| W | WRT(Kme3)QTARKSTGGKA |

| | |
|---|---|
| Y | YRT(Kme3)QTARKSTGGKA |
| V | VRT(Kme3)QTARKSTGGKA |

***Optional:*** *Add a tryptophan residue to the C-terminal end of the peptide sequence, separated from the peptide sequence by a flexible linker (e.g., 6-aminohexanoic acid (ahx)), to enable quantification of peptide concentration through tryptophan florescence.*

*Preparation of peptides, cofactors, and buffer stocks*

Due to the focus of the protocol described herein, the nonessential need to obtain peptides via the same methodology used, and the broad availability of peptide synthesis protocols, a detailed synthesis protocol is not described. To note, the peptides were synthesized following standard Fmoc (N-(9-fluorenyl)methoxycarbonyl) chemistry, on an automated ResPep SL peptide synthesizer (Intavis), at a scale of 2 µmol following procedures previously described [121,122].

**Timing: 1 day**

3. Dissolve dry peptides in 1X phosphate-buffered saline (PBS; pH 7), or another activity-compatible buffer), to a final concentration of 10 mM.

***Note:*** *Given that peptides were synthesized at a scale of 2 µmol, dissolve the peptides in 200 µL of 1X PBS to obtain a concentration of 10 mM.*

***Note****: Peptides synthesized in-house may be highly acidic due to residue trifluoracetic acid and may require pH to be adjusted to 7 by gradual addition of*

*sodium hydroxide (a 5M NaOH solution was used to add low volumes of 5-10 µL to peptides).*

***Critical:*** *all buffers and reagents should be prepared in ATP-free water. This includes cofactors and buffers described below.*

***Optional:*** *If peptides are synthesized in-house, it is highly recommended to quantify the peptides to ensure proper concentration. If an ahx-Trp is added to the C-terminal ends of peptide sequences, concentration may be determined by absorbance at 280 nm and the extinction coefficient of the peptide (available using the Expasy ProtParam tool; https://web.expasy.org/protparam/).*

4. Prepare 4X peptide stocks (40 µM) by diluting peptides in 1X PBS. Next, prepare a 1 mM peptide stock solution for the wild-type substrate peptide in 1X PBS.

***Pause Point:*** *Peptide stocks should be aliquoted and stored at -20°C until use.*

5. Make 10mM solutions of ascorbic acid, α-ketoglutarate, and $Fe(II)SO_4$ in a volume of 1 mL using ATP-free water, in separate 1.5 mL Eppendorf tubes.
    a. Dilute 10 mM cofactor stocks to 1 mM with ATP-free water.
6. Make a 40 mL solution of 0.5M HEPES in a 50 mL conical tube. Dissolve HEPES in 30 mL of ATP-free water and adjust pH to 7.5 using 5M NaOH. Top up volume to 40 mL to yield a final concentration of 0.5 M HEPES.

***Note:*** *Stores HEPES solution at 4°C protected from light.*

# Key resources table

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Chemicals, peptides, and recombinant proteins** | | |
| KDM5A$_{1-801}$-His (*Homo sapiens)* | SGC Oxford | n/a |
| KDM3A$_{515-1317}$-His-Flag (*Homo sapiens)* | [116] | n/a |
| Fmoc-Ala-OH | P3 Biosystems | Cat#41004 |
| Fmoc-Arg(Pbf)-OH | P3 Biosystems | Cat#41002 |
| Fmoc-Asn(Trt)-OH | P3 Biosystems | Cat#41007 |
| Fmoc-Asp(OtBu)-OH | P3 Biosystems | Cat#41019 |
| Fmoc-Cys(Trt)-OH | P3 Biosystems | Cat#41008 |
| Fmoc-Glu(OtBu)-OH | P3 Biosystems | Cat#41005 |
| Fmoc-Gln(Trt)-OH | P3 Biosystems | Cat#41011 |
| Fmoc-Gly-OH | P3 Biosystems | Cat#41010 |
| Fmoc-His(Trt)-OH | P3 Biosystems | Cat#41017 |
| Fmoc-Ile-OH | P3 Biosystems | Cat#41018 |
| Fmoc-Leu-OH | P3 Biosystems | Cat#41003 |
| Fmoc-Lys(Boc)-OH | P3 Biosystems | Cat#41001 |
| Fmoc-Met-OH | P3 Biosystems | Cat#41020 |
| Fmoc-Phe-OH | P3 Biosystems | Cat#41013 |
| Fmoc-Pro-OH | P3 Biosystems | Cat#41009 |
| Fmoc-Ser(tBu)-OH | P3 Biosystems | Cat#41006 |
| Fmoc-Thr(tBu)-OH | P3 Biosystems | Cat#41016 |
| Fmoc-Trp(Boc)-OH | P3 Biosystems | Cat#41012 |
| Fmoc-Tyr(tBu)-OH | P3 Biosystems | Cat#41014 |
| Fmoc-Val-OH | P3 Biosystems | Cat#41015 |
| Fmoc-Lys(Boc,Me)-OH | P3 Biosystems | Cat#47238 |
| Fmoc-Lys(Me)2-OH.HCl | P3 Biosystems | Cat#47237 |
| Fmoc-Lys(Me)3-OH chloride | Sigma | Cat#F5062 |
| HBTU | P3 Biosystems | Cat#31001 |
| NMM | Alfa Aesar | Cat#A12158 |

| | | |
|---|---|---|
| Rink Amide MBHA Resin | P3 Biosystems | Cat#52002 |
| Acetic anhydride | Fisher Scientific | Cat#108-24-7 |
| Piperidine | Sigma-Aldrich | Cat#104094 |
| DMF | Fisher Scientific | Cat#D119-20 |
| Ethanol | Commercial Alcohols | Cat#P016EAAN |
| DCM | Acros Organics | Cat#354800025 |
| Trifluoroacetic acid | Fisher Scientific | Cat#L06374 |
| TIPS | Acros Organics | Cat#214922500 |
| Ethyl Ether | Fisher Scientific | Cat#E138-4 |
| HEPES | BioShop | Cat#HEP005.1 |
| $Fe(II)SO_4$ | BDH Chemicals | Cat#B28400 |
| α-ketoglutarate | Sigma | Cat#K-3752 |
| Ascorbic acid | J.T. Baker Chemical Co. | Cat#B581.5 |
| Tris(2-carboxyethyl)phosphine (TCEP) HCl salt | Sigma-Aldrich | Cat#646547 |
| Bovine serum albumin | BioShop | Cat#ALB001.10 |
| Dimethyl sulfoxide | Bio Basic Canada | Cat#D0231 |
| NaCl | BioShop | Cat#SOD002.1 |
| KCl | BioShop | Cat#POC308.50 |
| Sodium phosphate monobasic monohydrate | Sigma-Aldrich | Cat#S9638 |
| Sodium phosphate dibasic anhydrous | BioShop | Cat#SPD307.50 |
| Potassium phosphate monobasic | BioShop | Cat#PPM666.1 |
| Acetic acid | Anachemia | Cat#00598-463 |
| **Critical commercial assays** | | |
| Succinate-Glo JmjC Demethylase/Hydroxylase Assay | Promega Corporation[123] | Cat#V7990 |
| **Software and algorithms** | | |
| Peptide Specificity Analyst (PeSA) | (Topcu & Biggar, 2019) | |
| **Other** | | |
| BioTek Cytation 5 microplate reader | BioTek | Cat#BTCYT5M |
| Grenier 384-well plate, white | Grenier | Cat#781075 |

## Materials and equipment (optional)

### 2X Reaction Buffer/Peptide Mix

| Reagent | Final concentration | Amount |
|---|---|---|
| 0.5 M HEPES (pH 7.5) | 50 mM | 50 µL |
| 1mM ascorbic acid | 200 µM | 100 µL |
| 1mM Fe(II)$SO_4$ | 20 µM | 10 µL |
| 1mM 2-oxoglutarate | 20 µM | 10 µL |
| DMSO | 2% | 10 µL |
| ATP-free water | n/a | 310 µL |
| 1 mM peptide | 20 µM | 10 µL |
| **Total** | **n/a** | **500 µL** |

### 2X Reaction Buffer

| Reagent | Final concentration | Amount |
|---|---|---|
| 0.5 M HEPES (pH 7.5) | 50 mM | 50 µL |
| 1mM ascorbic acid | 200 µM | 100 µL |
| 1mM Fe(II)$SO_4$ | 20 µM | 10 µL |
| 1mM 2-oxoglutarate | 20 µM | 10 µL |
| DMSO | 2% | 10 µL |
| ATP-free water | n/a | 320 µL |
| **Total** | **n/a** | **500 µL** |

**Succinate Detection Reagent I**

| Reagent | Final concentration | Amount |
|---|---|---|
| Succinate-Glo™ Solution | n/a | 5 µL |
| Acetoacetyl-CoA 100X | 1X | 5 µL |
| Succinate-Glo™ Buffer | n/a | 500 µL |
| **Total** | **n/a** | **510 µL** |

*[Commercially available Promega Succinate-Glo™ assay reagents are stored at -80 °C. Begin thawing reagent components on ice ~1 hour before use. For Acetoacetyl-CoA 100X, specifically, thaw at 22-25 °C for 5 minutes before addition to Succinate-Glo™ Buffer.]*

***Alternatives:*** *Here a BioTek Cytation 5 microplate reader was used to read luminescence, using a gain of 175 and plate height of 1.25 mm. Also, an Eppendorf centrifuge 5304 R with an A-2-DWP rotor was used. Other plate readers and centrifuges may be used.*

***Alternatives:*** *Here a 5 µL enzyme reaction is performed in each well of a 384-well plate. Alternative reaction volumes may be used; however, volumes of downstream commercially available Promega Succinate-Glo™ assay kit components must be scaled accordingly to maintain the ratio of 1:1:2 (reaction: Succinate Detection Reagent I: Succinate Detection Reagent II).*

***Alternatives:*** *Here a peptide list comprising the known methylproteome was scored using JmjC KDM recognition motifs.*

## Step-by-step method details

*Determine non-saturating enzyme concentration*

**Timing: 4 hours**

Given a fixed set of reaction parameters (e.g., cofactor concentration, pH, temperature, etc.), a user can determine the optimal concentration range of JmjC KDM yielding a non-saturated signal. Establishing a non-saturating enzyme concentration is key for downstream steps when comparing the effect of different peptides on JmjC KDM activity. This is achieved by dilution series of the given KDM from the highest concentration possible to a low or sub nanomolar concentration (e.g., <1-10 nM). In our prototypic example of KDM5A and KDM3A, by dilution series we determined enzyme concentrations yielding non-saturating signal (Fig. 2.1).

***Note:*** *To maximize the signal-to-background ratio, we recommend users of this methodology establish optimal conditions (e.g., pH, temperature, etc.) prior to beginning this workflow. This may be done experimentally or by literature search.*

1. Place a white 384-well microplate on a Peltier device and set temperature to 4°C.
2. Prepare 2X Reaction Buffer/Peptide Mix.
    a. Keep on ice or at 4°C for same day use.
3. Add 2.5 µL of 2X Reaction Buffer/Peptide Mix to all experimental wells in the 384-well microplate.
4. Thaw JmjC KDM stock protein on ice.

5. Perform a 20 µL dilution series (e.g., 4X or 2X) of the JmjC KDM protein in 1.5mL Eppendorf tubes on ice in the same buffer used initially for storage (see Table 2 for an example dilution range).

***Note:*** *Save at least a 20 µL aliquot of storage buffer for the "no enzyme" control.*

***Alternative:*** *Dilution series may be performed in higher or lower volumes depending on the number of replicate reactions performed.*

6. Set Peltier device to 23°C.

***Alternative:*** *Other temperatures for the JmjC KDM reaction may be used (e.g., optimal temperature may be determined prior to this protocol or published in literature).*

7. To initiate the reaction, add 2.5 µL of each JmjC KDM dilution to a specific microplate well containing the 2X Reaction Buffer/Peptide Mix (Table 2.2 shows an example plate layout).
    a. Set the 384-well microplate on a plate shaker for 2 min at 22-25°C.
    b. Centrifuge the plate at 240 x g for 1 min at 23°C.

***Note:*** *For the no enzyme control (NEC), add 2.5 µL of storage buffer containing an absence of JmjC KDM protein.*

**Table 2.2. Example of plate layout for JmjC 2x dilution series**. Initial concentration of purified KDM5A was 20 µM, yielding a final reaction concentration of 10 µM. Concentrations listed are in µM.

| Column | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row A-C (n=3) | NEC | 10 | 5.0 | 2.5 | 1.3 | 0.63 | 0.31 | 0.16 | 0.078 | 0.039 | 0.02 | 0.01 |

8. Allow the demethylation reaction to occur for 60 min at 23 °C.
9. Add 5 µL of Succinate Detection Reagent I to each assay well.
    a. Shake the 384-well microplate for 2 min at 22-25°C.
    b. Centrifuge the plate at 240 x g for 1 min at 23 °C and incubate for 60 min at 23 °C.

***Note:*** *The time of JmjC reaction may be adjusted, however, the reaction time used in this step must be kept consistent when performing all downstream JmjC KDM reactions. For example, reactions with KDM3A occurred for 3 hours.*

10. Add 10 µL of Succinate Detection Reagent II.
    a. Shake the 384-well microplate for 2 min at 22-25°C.
    b. Centrifuge the plate at 240 x g for 1 min at 23 °C and incubate for 10 min at 23°C.
11. Read the luminescence from each well using a microplate reader.
12. Calculate the average luminescence signal across technical replicates.
13. For each condition, subtract the average signal from that of the average of the NEC to account for background luminescence.
14. Using data analysis/graphing software, such as GraphPad Prism, visualize the data to determine non-saturating and linear dose responsive JmjC KDM concentrations (Fig. 2.1). Choose a concentration of protein within the mid-linear range of relative activity.
    a. See **Troubleshooting 1** if JmjC KDM activity is difficult to observe.

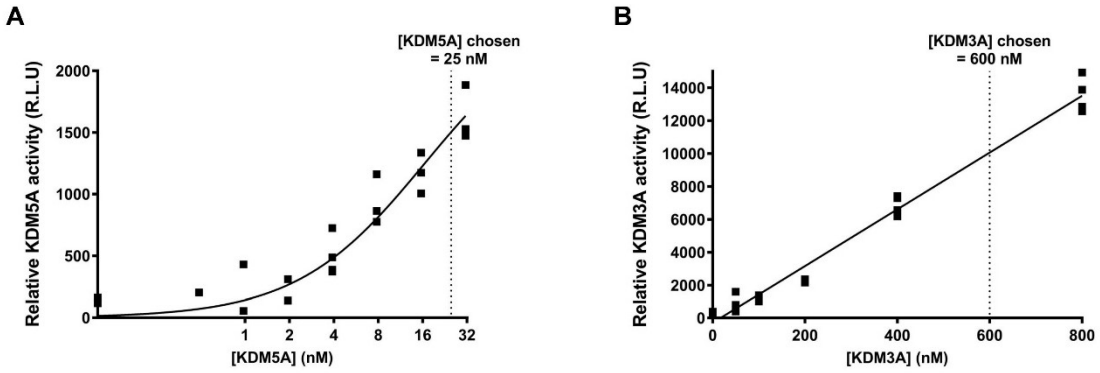



**Figure 2.1. Determining non-saturating JmjC KDM concentration.** Two-fold dilution series of recombinant (A) KDM5A and (B) KDM3A enzymes, observing activity towards H3-K4me3 and H3-K9me2 peptides, respectively. The data represents the average of three luminescent readings and error bars represent the SEM (n=3). Figure for KDM5A was adapted from [124]. Reprinted with permission of *Star Protocols – Cell Press.*

*Validate KDM methyl-state preference*

**Timing: 4 hours**

This step aims to validate the methyl-state preference (i.e., mono-, di-, tri-methylation) of the recombinant JmjC KDM for an established substrate. This is important as many JmjC KDMs have been shown to be capable of discriminating between different methyl-states of the same substrate. As the peptide libraries used for this method are synthesized as one methyl-state, this validation should be performed prior to commercially ordering or synthesizing the full peptide permutation library. Given the optimal concentration of the JmjC KDM determined in the previous step, the JmjC KDM activity can be monitored by performed reactions with peptide substrates of different methylation states. For assay validation, it is important to include an unmethylated substrate and a 'no peptide control' (NPC) alongside the standard NEC.

***Critical:*** *Some JmjC KDMs can convert 2-oxoglutarate to succinate in an appreciable extent in the absence of methylated peptide substrate. As a result, a 'no peptide control (NPC) is critical to use going forward to assess the true level of relative demethylation activity (Table 2.3).*

15. Thaw 40 µM wild-type substrate (including null-, mono-, di-, and tri-methyl state peptides) on ice.
16. Repeat step 1-3, except using 2X Reaction Buffer (similar to 2X Reaction Buffer/Peptide Mix minus peptide substrate).
    a. Keep 2X Reaction Buffer on ice or at 4°C for same day use.

17. Add 1.25 µL of 40 µM wild-type peptide substrates, individually, to each experimental well containing 2X Reaction Buffer.
    a. For the NPC, use 1.25 µL of 1X PBS.

***Note:*** *If peptides were dissolved in a buffer other than 1X PBS, use the peptide buffer for the NPC.*

18. Thaw JmjC KDM stock protein on ice and dilute to 4X the optimal concentration determined in Step 14.
    a. Considering the number of experimental conditions requiring enzyme (5 in total), the volume of JmjC KDM needed per reaction (1.25 µL), and the number of replicates (e.g., n=3), make at least 20.63 µL of 2X JmjC KDM (this is 10% excess of the minimum volume needed (18.75 µL)).
    b. Perform the protein dilution with the JmjC storage buffer.
19. Set Peltier device to 23°C (or appropriate constant temperature).
20. Add 1.25 µL of 4X JmjC KDM dilution to all experimental wells, except for the NEC. Add 1.25 µL JmjC KDM storage buffer to the NEC.
21. Repeat Steps 8-13.

**Table 2.3. Example of plate layout for testing JmjC KDM activity**. KDM activity is towards differentially methylated peptides (monomethylation [Kme1], demethylation [Kme2], and trimethylation [Kme3]).

| **Column** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|
| Row A-C (n=3) | NEC | NPC | Null | Kme1 | Kme2 | Kme3 |

22. For each condition, subtract the signal from that of the NPC to account for demethylation-uncoupled 2-oxoglutarate turnover.

***Optional:*** *Represent methyl-state preference as relative JmjC KDM activity; normalize subtracted luminescent signals to that of the peptide displaying the highest signal. Again, using data analysis/ graphing software visualize the data to represent methyl-state preference (Figure 2.2).*

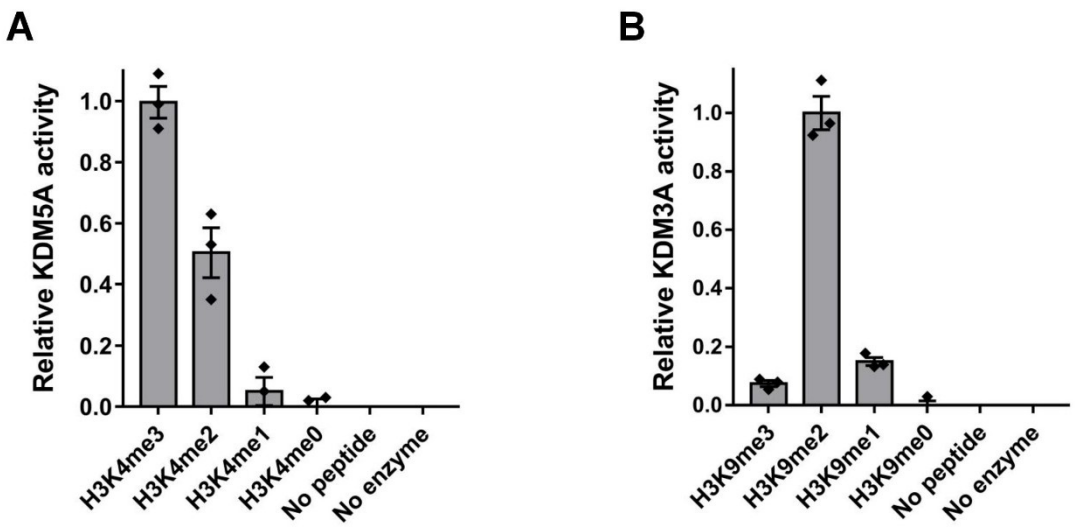



**Figure 2.2. Validation of JmjC KDM methyl-state preference.** Methyl-state specificity of (A) KDM5A and (B) KDM3A towards differentially methylated H3-K4 and H3-K9 peptides, respectively. KDM activity towards non-methylated, mono-, di-, and tri-methylated (i.e., me0/1/2/3) peptides at the indicated residues was assessed. Data represents the mean luminescence and standard deviation (n=3) normalized to the max signal. Figure for KDM5A was adapted from [124]. Reprinted with permission of *Star Protocols – Cell Press.*

*Determine JmjC KDM substrate preference by permutated peptide substrate (PPS) library*

**Timing: 4 hours (per mutation position)**

These steps aim to map JmjC KDM substrate preference by monitoring activity towards all peptides in the PPS library. At this point, all peptides within the library should be methylated according to the preferred methyl-state (e.g., H3-K4me3 and H3-K9me2 for KDM5A and KDM3A, respectively). The experimental set-up is performed in the same manner as the previous section (methyl-state validation), except using permutated peptides alongside the wild-type peptide and reaction controls (Table 2.4). In analyzing the data relative to the positive control (wild-type peptide) and negative controls (no peptide and no enzyme controls), one can determine the relative effect of individual amino acid mutations on JmjC KDM activity.

***Note:*** *To limit variability in the start of reaction time between experimental conditions; it is recommended to assess JmjC KDM activity towards peptides representing one residue position of the PPS library at a time.*

**Table 2.4. Example of plate layout for PPS library experiments.** The following is set up for mutations in one residue position. Each single letter amino acid designation is representative of the amino acid substitution occurring in the mutated peptide substrate.

| | **Experimental permutated peptides** | | | | | | | | | | | | | | | | | | | **Controls** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Column** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** |
| Row A-D (n=4) | R | H | D | E | N | Q | A | G | P | I | L | V | M | F | Y | W | S | T | C | K | NPC | NEC |

23. Thaw 40 µM peptide stocks for the PPS library position being tested on ice.
24. Make fresh 2X Reaction Buffer (e.g., for each residue position, 242 µL of 2X Reaction Buffer is required (22 conditions x 4 replicates x 2.5 µL = 220 µL + 10% extra).
25. Repeat Steps 1-3, using the freshly prepared 2X Reaction Buffer.
26. Add 1.25 µL of 4X (40 µM) peptide, individually, to experimental wells containing 2X Reaction Buffer. For the NPC, use 1.25 µL of 1X PBS.

***Note:*** *Again, if peptides were dissolved in a buffer other than 1X PBS, use that buffer for the NPC.*

***Critical****: Ensure each run of the experiment contains a reaction condition with a wild-type peptide. All mutation positions have a wild-type peptide. However, if mutation positions are not run in full, a wild-type peptide must be included.*

27. Thaw JmjC KDM stock on ice and dilute to 4X the optimal concentration determined in the previous step in JmjC storage buffer.

***Note:*** *Consider the number of experimental conditions requiring enzyme (e.g., the volume of JmjC KDM needed per reaction well (1.25 µL), and the number of replicates (e.g., n=4), make 115.5 µL of 2X JmjC KDM (this is 10% excess of what is needed; 21 conditions x 4 replicates x 1.25 µL = 105 µL)).*

28. Set Peltier device to 23°C (or appropriate constant temperature).

29. Add 1.25 µL of 4X JmjC KDM solution to all experimental peptide conditions (wild-type peptide and mutant peptides) and the NPC well. Add 1.25 µL JmjC KDM storage buffer to the NEC well.
30. Repeat Steps 8-13.
31. For each condition, subtract the average signal of the NPC to account for demethylation-uncoupled 2-oxoglutarate turnover.
32. To assess the relative effect of mutations on JmjC KDM activity; normalize substrate luminescent signals to that of the wild-type peptide.

***Note**: Normalization is achieved from dividing the luminescent signals of reactions with experimental peptides by that of the WT peptide. This is critical to gauge the relative effect of individual mutations on enzyme activity.*

33. Repeat Steps 23-32 for each mutation position.

***Optional:** Represent relative activity data as a heat-map to visualize JmjC KDM substrate preference (Figure 2.3) using basic data analysis software (e.g., Microsoft Excel).*

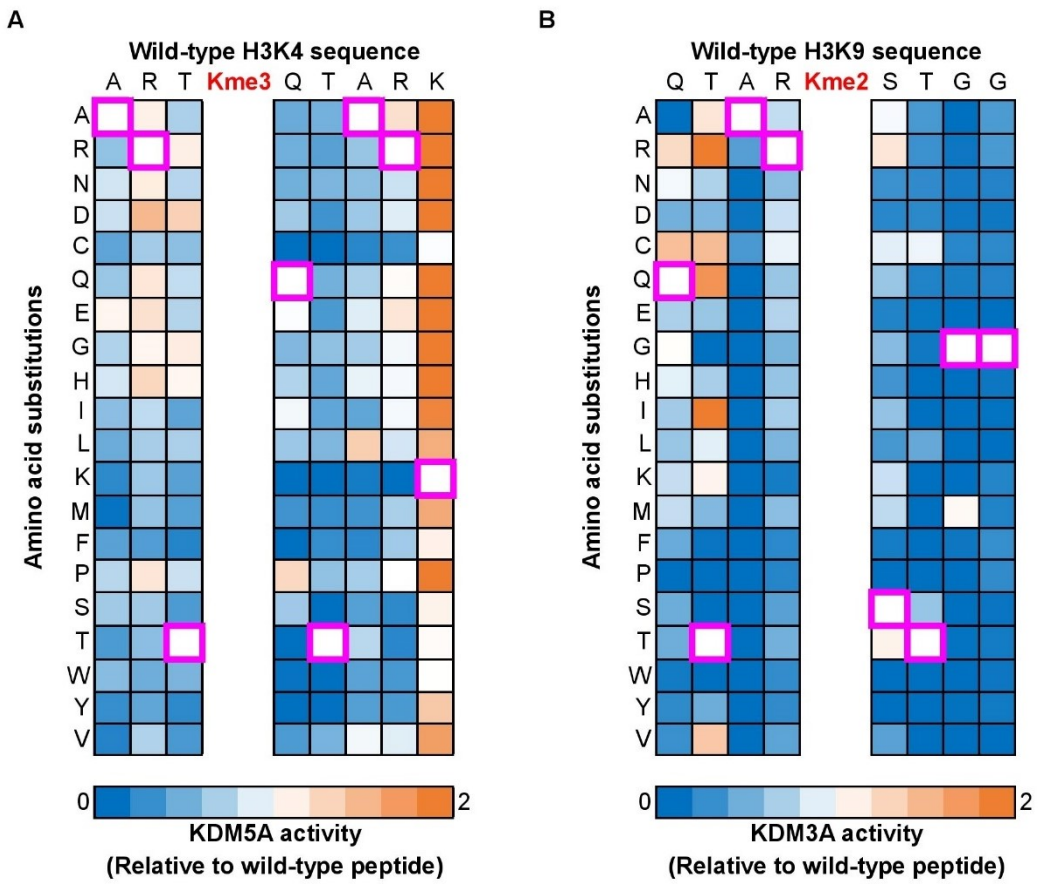



**Figure 2.3. Representation of KDM substrate preference as a heat-map.** Relative activity of recombinant (A) KDM5A and (B) KDM3A were assessed towards permutation libraries of their corresponding canonical substrates of preferred methyl state (i.e., H3-K4me3 and H3-K9me2, respectively). Data represents the mean luminescence normalized to the luminescence of their corresponding canonical substrates (n=3). X-axis and y-axis represent the wild-type peptide sequence and amino acid substitutions, respectively. Location of wild-type peptide (i.e., relative activity 1.0) spots are defined by the pink borders. Figure for KDM5A was adapted from [124]. Reprinted with permission of *Star Protocols – Cell Press.*

*JmjC KDM recognition motif generation, substrate prediction & in vitro validation*

**Timing: 1-2 hours**

The major aim of this step is to use JmjC KDM substrate preference to prioritize peptides to be tested for *in vitro* KDM activity. Peptide Specificity Analyst (PeSA) software can be used to easily produce a candidate KDM recognition motif [125]. In turn, these candidate recognition motifs can then be used to score and prioritize the methyllysine proteome for candidate substrates.

34. Download and open PeSA software (Topcu and Biggar, 2019; also available at https://github.com/EmineTopcu/PeSA).
35. Select “Permutation Array” tab at the menu.
36. Click “Load from File” and upload your JmjC KDM quantification matrix.

***Note:*** *The quantification matrix can be uploaded with either the relative activity values or the background subtracted luminescent values. If the latter, normalization to the wild-type luminescent values in each position can be performed within the software to obtain relative activity values. Under “Normalized”, select “Per Row/Column”*

***Note:*** *Experimentally, only the residue positions flanking the methylation site are assessed. However, to maintain the central target lysine within the visualization of the recognition motif; in the quantification matrix one can assign the central methylated K the relative value of “1” and all other mutations in this central position as “0”.*

37. Set the “Threshold” to the desired the value.

***Note**: In this context the "Threshold" defines the minimum level of relative JmjC KDM activity for an amino acid substitution to be included within the candidate recognition motif. Keep in mind that stringency increases with higher threshold values.*

***Alternative:** Another threshold value may be used and should be considered when troubleshooting Problem 2. Figure 2.4 represents KDM5A and KDM3A recognition motifs set at various threshold values (e.g., 0.25, 0.5, 0.75, and 1.0).*

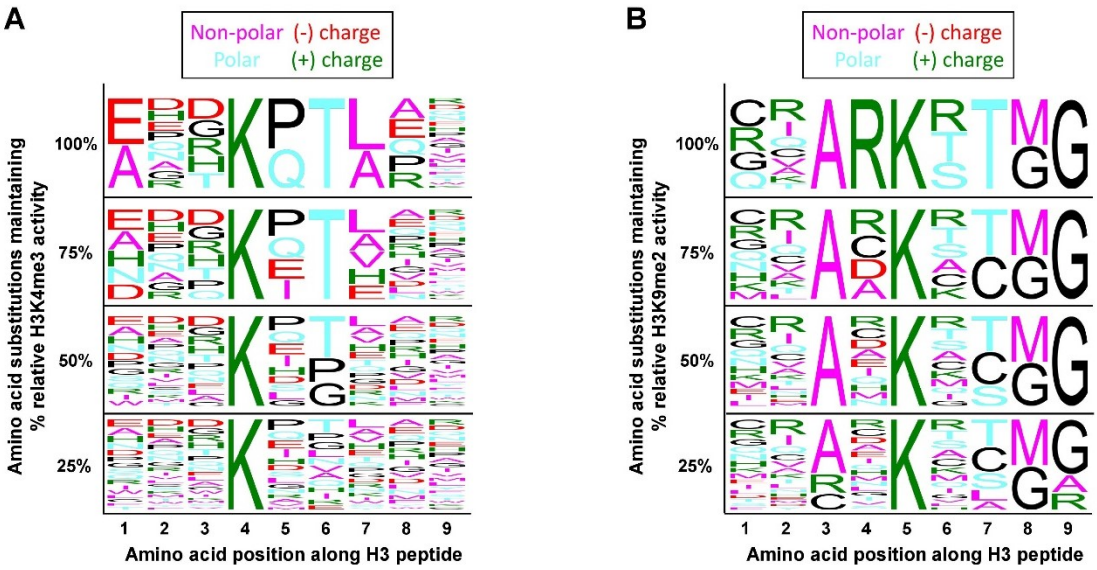



**Figure 2.4**. **Representation of the screening data as PeSA generated sequence motif.** The recognition motifs for (A) KDM5A and (B) KDM3A depict amino acid substitutions maintaining a minimum level relative activity defined by the threshold values on the y-axis. Motif images created using PeSA software[125]. Figure for KDM5A was adapted from [124]. Reprinted with permission of *Star Protocols – Cell Press.*

38. Obtain the recognition motif under the “motif” tab on the left-hand side of the window.
39. Score peptides that possess windows which reflect known methylation sites in the methylproteome, using the PeSA generated candidate recognition motif.
    a. Each position in the queried sequence which shares an amino acid with the recognition motif contributes a value of 1 to the total score. An example of scoring is provided in Table 2.5.

***Note:*** *Given the recognition motif produced and Kme position, ensure that the queried windows are of the same length and Kme residue in question is in the correct position. The PeSA score reflects the number of residues in the queried peptide matching those in the recognition motif at their corresponding positions. As a result, the predictions were made based on a 9-residue recognition motif, and therefore, the highest possible score obtainable is a total PeSA score of 9 (i.e., for windows exactly matching the recognition motif) (Table 2.5).*

**Table 2.5. Example of motif-based scoring.** Example of motif-based scoring. Example of scoring 9-mer sequences with KDM3A motif (defined by amino acids maintaining at least 100% relative activity). Tolerable amino acids at each position are shown in brackets. Matching and mismatching residues, compared to the motif, are shown in green and red, respectively. PeSA score is defined by the number of matching residues.

| Protein-site | Queried Sequence | Position and tolerable substitutions within motif | | | | | | | | | PeSA Score |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | -4 [CRGQ] | -3 [RIQCVAKT] | -2 [A] | -1 [R] | 0 [K] | +1 [RTS] | +2 [T] | +3 [MG] | +4 [G] | |
| H3-K9 (known) | QTARKSTGG | Q | T | A | R | K | S | T | G | G | 9 |
| RPA2-K693 (unknown) | CQMGKQTMG | C | Q | M | G | K | Q | T | M | G | 6 |

40. Determine mean (μ) and standard deviation (σ) of PeSA scores across all queried peptide sequences.
41. To define high-ranking peptide sequences, determine the PeSA score corresponding to 2 standard deviations above the mean.
    a. For KDM5A, using the 0.5 relative activity recognition motif, a PeSA score of 9 reflects high-ranking peptide sequences (μ = 6.6, σ = 1.2) (Fig. 2.5).
42. Repeat steps 23-32 in the previous section, except using high-ranking peptides instead of PPS library peptides to validate the *in vitro* activity of substrate predictions/rankings.

a. See **Troubleshooting 2** if JmjC KDM shows no activity to none, or few, of the predicted JmjC KDM substrates

***Note:*** *If the focus is purely on JmjC KDM substrate discovery, we recommend testing as many of the high-ranking peptides as possible, if not all. However, if the user is also interested in assessing whether the method accurately predicts or enriches in highly active in vitro substrates for the select JmjC KDM, one can assess a select number of random substrates with high-, medium-, and low-ranking PeSA scores.*

## Expected outcomes

In our experience, PeSA scores of large lists of peptide sequences (e.g., methylproteome; n=2155 sequences) follow a Gaussian distribution. For example, Figure 5 shows the distribution of PeSA scores of methylproteome sequences, scored with the 0.5 relative activity recognition motif of KDM5A. Thus, applying appropriate thresholds, based on standard deviation, to classify substrates as 'high-ranking' will result in a small number of candidate substrates to test *in vitro* JmjC KDM activity.

If testing relative JmjC KDM activity towards predicted substrates, we expect PeSA scores to correlate with *in vitro* JmjC KDM activity. For example, you could test JmjC KDM activity towards various high, medium- and low-ranking substrates. Ideally, one should observe no or minimal displayed JmjC KDM activity towards any of the medium and low-ranking substrates. Likewise, there should be observed JmjC activity towards a handful of the high-ranking

substrates. Ultimately, this added step of validation does prioritize several potential substrates to refine and focus further validation efforts on.

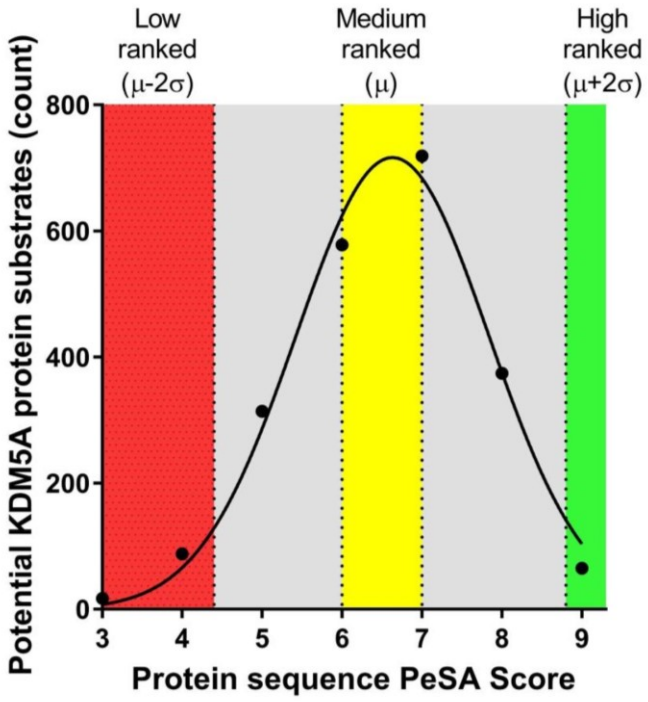



**Figure 2.5**. **Distribution of peptide scores for predicted KDM5A substrates.** Gaussian distribution of PeSA scores of methylproteome peptides scored with the 0.5 relative activity KDM5A recognition motif. High- and low-ranking substrates are defined by PeSA scores occurring 2 standard deviations (σ = 1.2) above (or equal to) or below (or equal to) the population mean (μ = 6.6), respectively. Reprinted with permission of *Star Protocols – Cell Press.*

## Limitations

*Limitation 1*

A limitation to consider is that the relative importance of the individual amino acids at each position in the recognition motif is determined in the context of a fixed sequence and thus sequence bias may influence the results. This is a consequence of using permutation-based exploration of enzyme specificity. It is certainly possible that the relative importance of a given amino acid is only observable when occurring in the presence of specific amino acids at other positions. Supporting this, and in the context of methyl-binding domains, JMJD2A-double Tudor domain binding to permutations of H3-K23me3, H3-K4me3, and H4-K20me3 show distinct specificities depending on wild-type sequence used [126].

*Limitation 2*

The protocol defined here assumes recognition of substrates by a given enzyme is specified by the residues directly proximal to the modification site (i.e., -4 to +5 positions). Although this has been shown to be the case for many KMT enzymes, and for Suv39H2 the +5 position also determines specificity [127], recognition of substrates may also be determined by more distant interactions. Additionally, on the H3 tail, more distal sequence elements were recently shown to be important for KDM5A-dependent demethylation of the H3-K4 site [74].

## Troubleshooting

## Problem 1:

Low level of observable enzyme activity (i.e., low signal-to-background ratio). This problem is referring to corresponding protocol step 14.

*Potential solution 1:*

A low observable enzyme activity could be due to several factors such as reaction buffer composition, pH, temperature, time, as well as substrate and cofactor concentration. Each factor mentioned may be optimized individually. Additionally, additives may affect enzyme activity. For example, KDM3A activity may be improved nearly two-fold by addition of TCEP at certain concentrations, whereas higher concentrations of sodium chloride hinder KDM3A activity [128]. Furthermore, peptide length may influence the level of enzyme activity. For example, the affinity of KDM3A for the H3-K9 substrate has been observed to be nearly 200-fold greater when using 21-mer peptides compared to 15-mer peptides [128,129].

Furthermore, if the level of uncoupled 2-oxoglutarate conversion (i.e., succinate production in absence of peptide) is relatively high compared to succinate formation in the presence of peptide; true demethylation activity may be difficult to observe using this assay or other assays detecting 2-oxoglutarate turnover. If this is the case, we suggest using assays directly assessing demethylation (e.g., formaldehyde turnover). However, if a significant difference is still able to be observed between the two conditions (i.e., JmjC KDM activity in the presence of peptide verses no peptide), succinate detection may still be a

feasible technique. In this case, the uncoupled 2-oxoglutarate conversion must be considered as a baseline.

*Problem 2:*

The JmjC KDM assessed shows no activity to none, or few, of the predicted JmjC KDM substrates. This problem is referring to corresponding protocol step 42.

*Potential solution 2:*

No activity towards any of the substrate predictions could be due to several factors, such as limitations inherent to permutation-based exploration of specificity. The latter cannot be addressed by adjusting variables within this protocol and would need to be resolved by assessing specificity in another manner or by using multiple substrate permutations. However, the adjustable variables include reducing the stringency of substrate predictions to be more permissive for substrate discovery (i.e., use recognition motif defined by a lower activity threshold). It may also be beneficial to test activity of the given JmjC KDM at multiple substrate concentrations. Finally, and related to Limitation 2, the critical positions defining specificity may exist outside of the permutated positions and thus it may be beneficial to expand this window and re-predicting substrates based on this expanded recognition motif.

# Resource availability

### *Lead contact*

Further information and requests should be directed to and will be fulfilled by the lead contact Dr. Kyle Biggar (kyle_biggar@carleton.ca)

### *Materials availability*

This protocol is not associated with any newly generated materials.

### *Data and code availability*

The protocol includes the methylproteome accessed from PhosphoSitePlus on 12-03-2020[130].

# Acknowledgments

This work was supported by a National Science and Engineering Research Council (NSERC) Canada Discovery grant to K.K.B (grant no. RGPIN-2016-06151) and W.G.W (grant no. RGPIN-2017-06414). A.C. held a Canada Graduate Scholarship – Doctoral (CGS-D) from the NSERC of Canada. The KDM5A Sf9 construct was a generous gift from the Structural Genomics Consortium Oxford group. The KDM3A construct was a generous gift from Nicola Burgess-Brown.

## Author contributions

**Matthew Hoekstra:** Conceptualization, Methodology, Data curation, Validation, Writing – Original draft.

**Anand Chopra:** Methodology, Data curation, writing- original draft, review & editing

**William G. Willmore:** Supervision, Writing - Review & Editing.

**Kyle K. Biggar:** Conceptualization, Resources, Funding acquisition, Supervision, Writing - Review & Editing.

# Chapter 3

# Identification of *in vitro* JMJD lysine demethylase candidate substrates via systematic determination of substrate preference

**Matthew Hoekstra[1,2], Kyle Biggar[1,2]**

[1]Institute of Biochemistry, Carleton University, Ottawa, ON, K1S 5B6, Canada

[2]Department of Biology, Carleton University, Ottawa, ON, K1S 5B6, Canada

## Published manuscript

Hoekstra, M., Biggar, K.K. (2021). Identification of in vitro JMJD lysine demethylase candidate substrates via systematic determination of substrate preference. ***Anal. Biochem.*** 633, 114429.

## Chapter Summary

As demonstrated in the Chapter 2, we were able to develop a robust assay to relatively quantify JmjC KDM substrate preference. Using this approach, we demonstrated an ability to predict and rank methyllysine-containing proteins in the methyllysine proteome. Once optimized and developed, we then sought out to apply our method and determine if we could validate the JmjC KDM-mediated demethylation of a high-ranking substrate in an *in vitro* context. The result was a successful prediction and assessment analysis of the ability of $KDM5A_{1-801}$ to demethylate a handful of high-ranking substrates *in vitro.* This Chapter is based on a paper published in *Analytical Biochemistry* that describes how we applied this method to $KDM5A_{1-801}$ to accurately predict high-ranking substrates, as well as validate select substrates *in vitro*. This paper represents our first attempt to apply our full method to a KDM5 enzyme in hopes to determine potential novel KDM5 substrates to focus further validation efforts.

Briefly, through the application of a library of permutated peptide substrates coupled with a succinate-based detection assay, we were able to systematically determine $KDM5A_{1-801}$ substrate preference. To reiterate, our method utilizes a library of tri-methylated peptides that are systematically mutated from an established substrate peptide. By using the well-characterized H3K4me3 peptide substrate, 180 mutated H3K4me3 peptide substrates containing a single amino acid mutation from the -3 position, to +5 position around the K4 tri-methylated lysine residue were produced. Using this highly specialized permutated peptide substrate (PPS) library, we were able to assess

position-specific amino acid preferences displayed by KDM5A$_{1-801}$ through observing its activity towards each of the systematically mutated substrates. The result of obtaining this relative activity is the production of a recognition motif which was used to predict potential KDM5A$_{1-801}$ substrates. Finally, we synthesized trimethylated peptides that represented high-ranking predicted KDM5A substrates and assessed KDM5A$_{1-801}$ activity towards these substrates *in vitro*. Our analysis showed that our approach was 90% accurate in predicting *in vitro* substrates. Ultimately, we demonstrated the first systematic analysis of any JmjC KDM substrate preference and used this data to accurately predict potential novel substrates for a KDM5 demethylase family member.

## Manuscript Summary

A major regulatory influence over gene expression is the dynamic post translational methylation of histone proteins, with major implications from both lysine methylation and demethylation. The KDM5/JARID1 sub-family of Fe(II)/2-oxoglutarate dependent lysine-specific demethylases is, in part, responsible for the removal of tri/dimethyl modifications from lysine 4 of histone H3 (i.e., H3K4me3/2), a mark associated with active gene expression. Although the relevance of KDM5 activity to disease progression has been primarily established through its ability to regulate gene expression via histone methylation, there is evidence that these enzymes may also target non-histone proteins. To aid in the identification of new non-histone substrates, we examined KDM5A *in vitro* activity towards a library of 180 permutated peptide substrates derived from the

H3K4me3 sequence. From this data, a recognition motif was identified and used to predict candidate KDM5A substrates from the methyllysine proteome. High-ranking candidate substrates were then validated for *in vitro* KDM5A activity using representative trimethylated peptides. Our *in vitro* approach correctly identified activity towards 90% of high-ranked substrates. Here, we have demonstrated the usefulness of our method in identifying candidate substrates that is applicable to any Fe(II)- and 2-oxoglutarate dependent demethylase.

## Introduction

Histone proteins and their diverse array of post-translational modifications (PTMs) have been subject to exquisite evolutionary conservation in eukaryotes. Accordingly, factors controlling the deposition, removal, and interpretation of histone modifications are themselves deeply conserved. Additionally, many of such factors are strongly involved in the development of cancer and other diseases in humans[80,93,131]. One of the most well-studied types of PTMs occurring on histones is the methylation of lysine residues. Specific lysine residues on histone proteins can be mono-, di-, or tri-methylated (Kme1, 2, 3) and these methylation states have distinctive mechanistic consequences reflected by the recruitment of binding proteins with differential affinities for each methylation state[30,58,115,132]. The regulation of histone lysine methylation requires enzymes that read, write, and erase methyl marks on histone proteins in a sequence-specific manner. Methyl-lysine binding domains (MBD) are commonly referred to as the “readers” of the modification[3,30]. In this way, lysine

methyltransferase (KMT) enzymes can be considered the “writers” and lysine demethylase (KDM) enzymes the “erasers” of lysine methylation[3,29,51]. To date, there are currently two known families of KDM enzymes; flavin adenine dinucleotide (FAD)-dependent amine oxidases and the larger Jumonji C domain (JMJD) containing demethylases[58]. This large JMJD-containing family of KDMs functions as $Fe^{2+}$ and 2-oxoglutarate-dependent dioxygenases, which is represented by a diverse family of over 30 demethylases enzymes classified into several subfamilies[29,48,51,69]. Overall the JMJD family of KDM enzymes demonstrate a wide range of histone-lysine-containing sequence specificity and demethylation of different methylated lysine states[51].

The number of non-histone methyl-lysine sites for which there is an identified KMT or KDM enzyme continues to grow, but remains outpaced by the discovery of new methylation sites via mass spectrometry (MS)-based methods[99,100,133,134]. To date, more than 6,000 unique methyl-lysine sites, across over 2,000 unique human proteins have been identified, but very few of these identified sites have been linked to a methyl-modifying enzyme[71]. A number of proteomic methods have been developed to help identify substrates of both KMTs and KDMs, yet the discovery of KMT substrates have vastly outpaced those of KDM enzymes due to the absence of a systematic screening platform to map substrate selectivity[90,95,96,99]. For KMT enzymes, the use of peptide arrays (either systematic mutations of a known substrate or a lysine-oriented peptide libraries (K-OPL)) have been successfully used to generate substrate selectivity profiles for any KMT enzyme[135,136]. *In vitro* sequence preferences can then be

used to rank all lysine methylation sites within the methyl-lysine proteome by the likelihood of its use as a substrate. Similar rapid screening platforms have not been developed for KDM substrate discovery, resulting in a comparatively stark lack of non-histone substrates identified for these enzymes.

To expand the enzyme-substrate information available for KDM enzymes, there has been a number of methodologies developed that examine and characterize catalytic activity towards their substrate protein(s). Traditionally, radioactive and enzyme-coupled assays have been well-documented methods to study histone lysine demethylation[85,87]. These assays function on the principle of measuring either the depletion or production of reaction products, such radioactively labelled $^{3}$H-SAM or 2-oxoglutarate and peptide substrates[85,87]. Recently, there has been an increase in the use of bead-based techniques, including the Amplified Luminescent Proximity Homogenous Assay (ALPHA) and Time-Resolved Fluorescence Resonance Energy Transfer (TR-FRET)[87,89]. These bead-based methods offer signal sensitivity and large library size which allows for high-throughput screening[88,89]. However, the high cost and labor-intensive initial screening and library preparation are key disadvantages to these techniques[87,89,90]. In contrast, MS-coupled assays have also been a commonly used method to derive quantitative data for KDM activity, although low throughput is a major disadvantage with this approach[87,98,100]. Here, we report on the development of an *in vitro* platform that relies on succinate-based detection to measure KDM activity to map KDM selectivity and predict new candidate substrates[123].

To demonstrate the utility of our approach to identify *in vitro* substrate preference, we used KDM5A as our prototype example of a JMJD-containing demethylase with well-documented activity[43,64]. The KDM5/JARID (Jumonji, AT-rich interactive domain) sub-family of JMJD-containing KDMs consists of 4 members: KDM5A/B/C/D[29,48,69]. The KDM5 sub-family catalyzes the demethylation of tri- and di-methylation groups from lysine 4 of histone H3 (i.e., H3K4me3/2), the canonical and only known substrate of this family[28,48,69]. As a result, majority of studies to date have revolved around the ability of KDM5 to regulate gene expression through regulation of H3K4 methylation status.[69]. More recently, KDM5 research has begun to focus further on characterising the KDM5 family. This research has demonstrated growing evidence that KDM5 enzymes may be capable of demethylating non-histone proteins[69,137]. Such findings are vital as the identification of substrates beyond H3K4me3/2 could help better define the role(s) of KDM5 in both normal and disease biology[6,69,93]. To support the growing need of a systematic method to predict KDM substrates, we developed a method to study substrate preference. This method is derived from peptide permutation array-based assays that are typically used for KMT substrate prediction, as well as KDM demethylation assays to relatively determine activity. Our method utilizes a panel of unique lysine methylated peptides derived by systematically mutating an established substrate peptide. Using this highly specialized permutated peptide substrate library, it is possible to determine position-specific amino acid preferences displayed by any JMJD KDM enzyme with at least one established substrate. Through observing relative

activity towards each of the mutated substrates it can be concluded which mutations result in a relative increase or decrease in KDM activity. Collectively, through using this information of *in vitro* substrate preference, a KDM recognition motif can be produced and used to determine a ranked list of candidate demethylation substrates.

## Materials and Methods

*Recombinant KDM5A expression and purification*

*Spodoptera frugiperda* (Sf9) cells (Novagen) were maintained at 27 $^{o}$C in a humidified chamber in Grace's Insect Media (Gibco) supplemented with 10% fetal bovine serum (Gibco) and Penicillin-Streptomycin (Gibco). For recombinant KDM5A$_{1\text{-}801}$ protein purification, Sf9 cells (500 mL at $10^6$ cells/mL) were infected with KDM5A-6XHis P3 baculovirus (Bacmid received from SGC) at a 1:100 ratio. Cells were grown propelling at 130 rpm for 60 hrs at 27$^{o}$C. Cells were collected by centrifugation, snap frozen and stored at -80 $^{o}$C. Cells were then resuspended in P5 buffer (50mM $NaHPO_4$ (pH 7), 500 mM NaCl, 10% glycerol, 0.05% TritonX-100, 1 mM DTT, 5 mM Imidazole) supplemented with protease and phosphatase inhibitors (1 mM PMSF, 0.5 µM Aprotinin, 10 µM E-64, 1 µM Pepstatin, 1 mM $Na_3VO_4$). Cells were lysed through 15 passes through a Dounce homogenizer at 4 $^{o}$C. Cells were subsequently sonicated 3x for 30 sec at 40% amplitude intensity followed by another 15 passes through a Dounce homogenizer at 4 $^{o}$C. Cells were then centrifuged at 16,000 RPM for 45 min at 4 $^{o}$C. The recombinant protein was purified via standard Ni-NTA

chromatography. Briefly, the soluble cell lysate was incubated with 250 µL HisPur™ Ni-NTA Resin (Thermo Fisher Scientific) for 2.5 hrs at 4 $^{0}$C and washed with P40 buffer (P5 buffer containing 40 mM imidazole). Recombinant protein was eluted with P500 buffer (P5 buffer containing 500 mM imidazole) then dialyzed overnight into storage buffer (25 mM HEPES (pH 7.5), 150 mM NaCl, 10% glycerol, 1 mM DTT) using 3.5 kDa molecular weight cut-off dialysis tubing (SnakeSkin® Dialysis tubing, Thermo Scientific). Protein concentration was quantified through a standard Bradford assay[138]. Proteins were snap-frozen in liquid nitrogen and stored at -80 $^{o}$C for later use.

*Peptide Synthesis*

Peptides were synthesized *de novo* in house using a ResPep SL (Intavis Bioanalytical Instruments) peptide synthesizer. Fmoc protected amino acids, including lysine methylated amino acids, were purchased from p3bio systems. Standard Fmoc chemistry was used to produce peptides on Rink amide resin (p3bio systems). After synthesis, peptides were cleaved from the resin and side chains were deprotected using an acidic cleavage solution (95% trifluoroacetic acid, 3% triisopropylsaline, and 2% water). Peptides were precipitated and washed with ice-cold diethyl ether. Peptides were resuspended in 1X PBS solution, quantified, aliquoted and frozen until needed. SDS-PAGE was used to visualize synthesized peptides and MS was used to validate mass (Appendix B, Fig. B1).

To design our library of mutated substrate peptides, we systematically mutated the H3K4 sequence (ARTKQTARKSTGGKA; K4 position underlined). In essence, each peptide, trimethylated at K4, contains a single amino acid mutation to one of the remaining 19 amino acids in the positions surrounding the substrate site. For our analysis of H3K4, we mutated the -3 position to +5 position around the trimethylated K4 position (i.e., ARTKme3QTARK; K4 position underlined).

*KDM demethylation assay*

All KDM5A$_{1\text{-}801}$ *in vitro* enzyme characterization assays and permutated peptide substrate library experiments were performed as described. All reactions took place in demethylase reaction buffer (25 mM HEPES (pH 7.5), 10 μM peptide substrate, 100 μM ascorbate (J.T Baker Chemical Co), 10 μM Fe(II) sulfate (BDH chemicals), 10 μM α-ketoglutarate (Sigma Aldrich), 1% v/v DMSO) for 1 hr at 23 $^{o}$C in a 384-well Corning flat-bottom microplate. Following incubation, succinate detection reagents were added (Succinate Glo™ JMJC Demethylase Assay; Promega Corp.) according to manufacturer's instruction. Briefly, 5 μL of Succinate detection reagent I was added and incubated for 60 min to quench the demethylation reaction and convert succinate product to ATP. After, 10 μL of Succinate detection reagent II was for 10 min added to convert ATP to light. After 10 min, luminescence was read using a Cytation 5 plate reader (Bio-Tek, Winooski, VT). A KDM5A$_{1\text{-}801}$ 2x dilution curve (Fig. 3.1) was used to determine the concentration of KDM5A$_{1\text{-}801}$ we observed maximum

activity with established reaction conditions. Once the maximum activity was observed, 80% of this enzyme concentration was used for all further activity assays.

All KDM5A$_{1\text{-}801}$ permutated peptide substrate library experiments were performed similarly as described above. Briefly, 2.5 µL of 2X reaction buffer (without peptide substrate) was aliquoted to a 384-well plate at 4 $^{o}$C. While at 4$^{o}$C, 1.25 µL of 40 µM permutated substrate peptide were added. Temperature was increased to 23 $^{o}$C and 1.25 µL of 4X (100 nM) KDM5A in 25 mM HEPES (pH 7.5) were added to the reaction. The reactions proceeded for 1 hr at 25 $^{o}$C, followed by the addition of the succinate detection reagents, quenching, and measuring luminescence as described above.

*Western blotting*

Proteins were separated on a standard 8% tris-glycine SDS-PAGE gel and transferred to 0.45 µm PVDF membrane (GE Healthcare Life Sciences). Immunoreactive bands specific to KDM5A were detected using anti-KDM5A primary antibody (1:800; Santa Cruz Biotechnology, sc-365993) and rabbit anti-mouse HRP-conjugated secondary antibody (1:10,000; ThermoFisher Scientific, 31450). Bands were visualized via chemiluminescence using Clarity™ Western ECL Substrate with a Gel Doc™ XRS+ Imaging system (Bio-Rad).

*Data analysis*

Demethylation activity data derived from the permutated H3K4 peptide substrate library was used to determine the KDM5A substrate recognition motif using Peptide Specificity Analyst (PeSA) software[125]. Weight-based KDM5A substrate recognition motifs were constructed using Peptide Specificity Analyst (PeSA) software. PeSA allows a user-defined threshold that is set relative to a defined positive control (i.e. H3K4me3 peptide) [125]. The threshold reflects the minimum activity required for any mutated peptide substrate to demonstrate $KDM5A_{1\text{-}801}$ activity. The recognition motif was used to identify candidate substrates and works by identifying specific combinations of amino acids in potential substrates that are predicted to display KDM5A activity[125].

## Results

*Succinate-based demethylation detection assay*

The commercially available *Promega Succinate-Glo™* assay monitors succinate production as a measure of KDM activity[92,93,123,139] (Fig. 3.1A). First, an optimal enzyme concentration was identified via an enzyme titration curve (Fig. 3.1B). A $KDM5A_{1\text{-}801}$ enzyme concentration range between 4 – 32 nM was found to produce a suitable dose-responsive signal (Fig. 3.1C). As a result, an optimal $KDM5A_{1\text{-}801}$ concentration of 25 nM was used for all subsequent demethylation reactions.

As KDM5A exhibits catalytic activity towards specific methylation states, (i.e., Kme1 vs. Kme3), we aimed to then confirm the methyl-state specificity of

KDM5A$_{1-801}$ for H3K4me0/1/2/3 peptides (Fig. 3.1D). As expected, KDM5A$_{1-801}$ demethylation of H3K4me3 was the most active, whereas the H3K4me2 substrate demonstrated 51% of the relative H3K4me3 activity ($p < 0.05$). The H3K4me1/0 substrates did not produce a signal significantly above background level. Similarly, no measurable activity was found for either the no peptide or no enzyme negative controls. From this data, we were confident the KDM5A$_{1-801}$ enzyme concentration, reaction parameters and succinate-based detection assay were optimized to study the KDM5A$_{1-801}$ substrate preference through the permutated peptide substrate library.

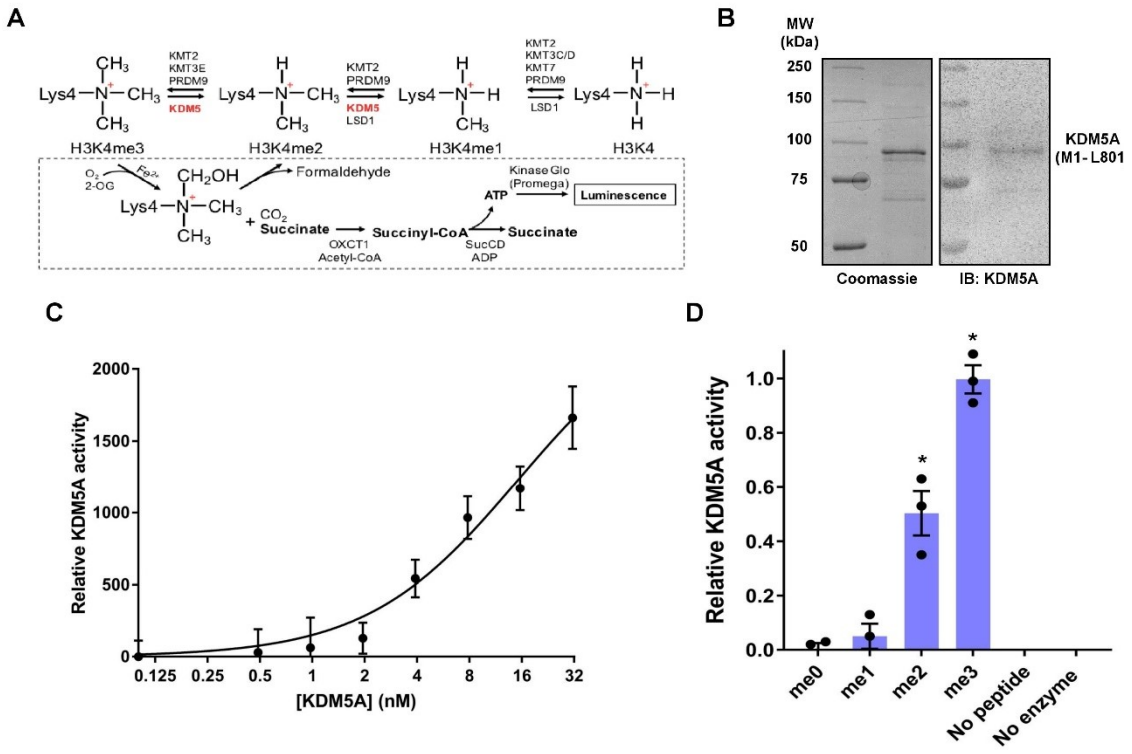



**Figure 3.1. KDM5A activity toward H3K4 peptides.** ***A.*** Reaction mechanism for histone H3K4 methylation/demethylation. *In vitro* assay performed in Demethylase assay buffer (10 μM tri-methylated histone H3 peptide substrate, 100 μM ascorbate, 10 μM Fe (II) and 10 μM 2-oxoglutarate). Succinate detection reagents added in a 1:1:2 ratio. ***B.*** SDS-PAGE and western blot images showing purified KDM5A$_{1\text{-}801}$ used in this study. ***C.*** Luminescence-based detection of KDM5A$_{1\text{-}801}$ demethylation of WT H3K4me3 peptide. Luminescence averages were normalized by background signal. Results are mean ± std. error (n=3). ***D.*** Luminescence detection of KDM5A$_{1\text{-}801}$ demethylation of H3K4me0, H3K4me1, H3K4me2, and H3K4me3 peptides. Luminescence averages were normalized by background signal, and relatively quantified to H3K4me3 peptide luminescence (set to 1). Results are mean ± std. error (n=3). * indicates $p < 0.05$ from me3 data. Reprinted with permission of *Analytical Biochemistry.*

*Application of a peptide library to determine substrate preference*

As H3K4me3/2 is the only known protein substrate for KDM5A, our novel permutated peptide substrate library was based on the systematic mutation of the H3K4 sequence surrounding the lysine 4 methylation site (ART<u>K</u>QTARKSTGGKA; methylated K4 position underlined). This resulted in a library of 180 unique peptide sequences, whereby each peptide contains a single amino acid mutation to one of the remaining 19 amino acids, from the -3 position to +5 position around the trimethylated K4 position (i.e., *ART*<u>Kme3</u>*QTARK*STGGKA; K4 position underlined, permutated region in italics). KDM5A$_{1\text{-}801}$ activity towards each permutated peptide was then assessed by demethylation assay (Fig. 3.2A). Figure 3.2A demonstrates a visual representation of the mean KDM5A$_{1\text{-}801}$ activity towards each permutated peptide in a virtual array-based format. Activity was monitored relative to the wild-type (WT) H3K4me3 (positive control, set to 1) and the H3K4me0 (negative control, set to 0) peptides. No peptide/enzyme controls were also included in all assays. This normalization strategy allows for the discovery of amino acid mutations that result in either an increase or decrease in *in vitro* KDM5A$_{1\text{-}801}$ activity. With the relative demethylation activity derived from our permutated H3K4 peptide library, we then used Peptide Specificity Analyst (PeSA) software to determine a candidate substrate recognition motif[125]. We compared recognition motifs between 25%, 50%, 75% and 100% relative H3K4me3 activity thresholds (Fig. 3.2B). As observed, the further the recognition motif is restricted (i.e., higher threshold), the less tolerable amino acid substitutions become. At the lowest

activity threshold (25%), no motif could be discerned making further analysis with this threshold non-consequential. We decided to choose a relative H3K4me3 activity threshold of 50% for our substrate predictions. This value was chosen to provide flexibility on amino acid substitutions and reduce stringency of downstream predictions.

*Identification of candidate demethylation substrates*

Using the candidate substrate recognition motifs, we then predicted candidate KDM5A substrates within the *Homo sapiens* methyllysine proteome (PhosphositePlus database; accessed on 12-03-2020)[130]. Using this database of over 2,000 lysine methylated protein sites, a list of candidate substrates was produced and scored against each of three recognition motifs (50%, 75%, or 100% relative H3K4me3 activity). Candidate substrate scoring was accomplished using the PeSA algorithm[125]. PeSA predictions are based on the number of amino acid residues within a predicted substrate that match the tolerable amino acid substitutions determined by the KDM5A recognition motif. As we collected relative $KDM5A_{1-801}$ activity from H3A1 (i.e., -3 position) to H3K9 (+5 position), surrounding the central trimethylated H3K4, predictions were made based on 9 residues. Thus, the highest possible score obtainable, for windows exactly matching the recognition motif, is a PeSA score of 9. As we were able to identify a set number of candidate proteins each with a respective PeSA score, we applied a Gaussian distribution to each motif threshold (Fig. 3.3A).

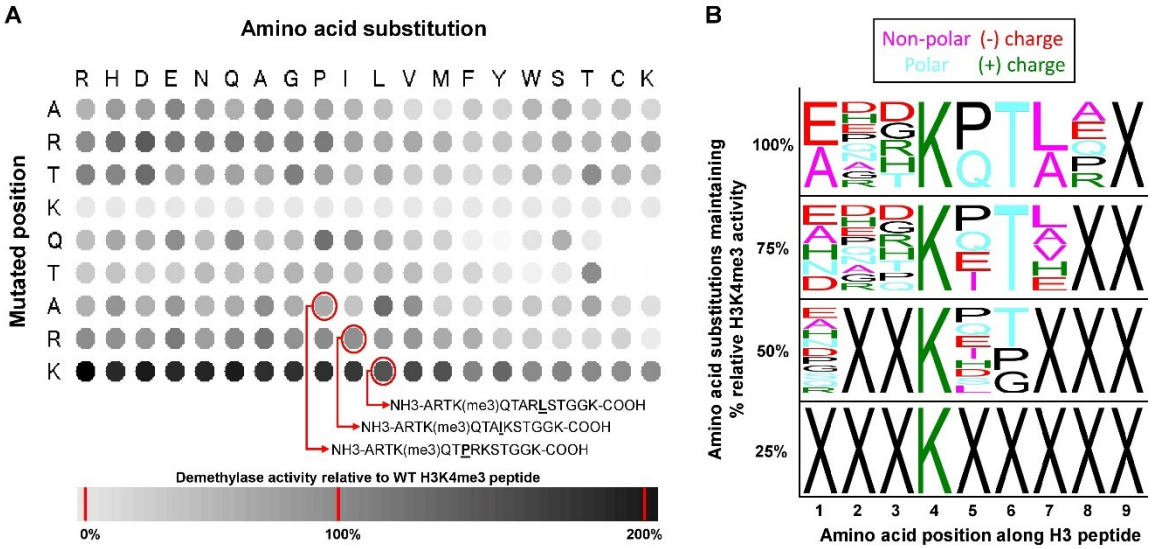



**Figure 3.2. Predicted KDM5A$_{1\text{-}801}$ specificity by permutated peptide substrate library.** ***A.*** Detection of KDM5A (M1-L801) demethylation activity with H3K4me3 peptide substrates that contain single amino acid mutation. Data was normalized to the H3K4me3 peptide at each mutation position (set to 1 for each relative position). Results are mean of (n=3-4). Digitized depiction of permutation array was created using PeSA software and shows mean demethylation activity [125] ***B.*** KDM5A recognition motifs derived from relative KDM5A demethylation activity towards permutated peptide substrate library. Each motif is reflective of amino acid substitutions that maintained 25%, 50%, 75%, and 100% of KDM5A activity relative to H3K4me3. Motif images created using PeSA software [125]. Reprinted with permission of *Analytical Biochemistry*.

This allowed us to rank candidate substrates and defined high, medium and low-ranking predictions. As a result, we observed how restricting, or relaxing, the relative H3K4me3 activity threshold can have adverse effects on the amount of candidate proteins considered to be high ranking. For example, when applying the 50% relative H3K4me3 activity threshold, as there are 54 protein sites that have a perfect PeSA score of 9. In contrast, only H3K4 has a perfect PeSA score of 9 at the 75% relative H3K4me3 activity threshold.

To identify candidate KDM5A substrates, we used a permissive threshold of 50% activity relative to the H3K4me3 peptide. At a threshold of 50%, the methyl-lysine proteome has a mean PeSA score of 6.6 ± 1.2 (Fig. 3.3A). Therefore, we determined any substrate with a score greater than 2 SD of the mean PeSA score as “high ranked” (i.e., PeSA score = 9). Likewise, “middle ranked” substrates were classified with a score near the mean, and “low ranked” substrate with a score of lower than 4.2. In total, 54 substrates were ranked as high-confidence, 1,297 as medium-confidence, and 105 as low-confidence predictions as KDM5A substrates based on this approach.

To validate the ability of our approach to rank *in vitro* substrates, we then assessed $KDM5A_{1-801}$ activity towards a random 10 substrates from high, and 5 substrates from both the low and medium-ranked classifications (Fig. 3.3B). This was accomplished by synthesizing trimethyl-lysine containing peptides with sequences matching the predicted demethylation site (+/- 5 residues: 11 total, including the trimethyl-lysine site). As seen in Fig. 3.3B, within the context of *in vitro* KDM5A activity, this method accurately differentiated between substrate

candidates ranked highly compared to those in the medium and low-ranked classifications (Fig. 3.3B). We observed that KDM5A$_{1-801}$ displayed activity towards 90% of high ranked peptides (activity defined as greater than 50% of H3K4me3 activity). The method also successfully ranked substrates predicted not to have significant activity; no significant activity was detected for any medium or low-ranked substrates. Therefore, our method has provided nine candidate substrates in which KDM5A may actively demethylate, 2 of which are known to be functionally di-methylated (Table 3.1). Although *in vitro* reactions with peptide substrates do not fully encompass the specificity that exists between enzyme-substrate interactions, this method does prioritize a number of potential substrates to refine and focus further validation efforts[95].

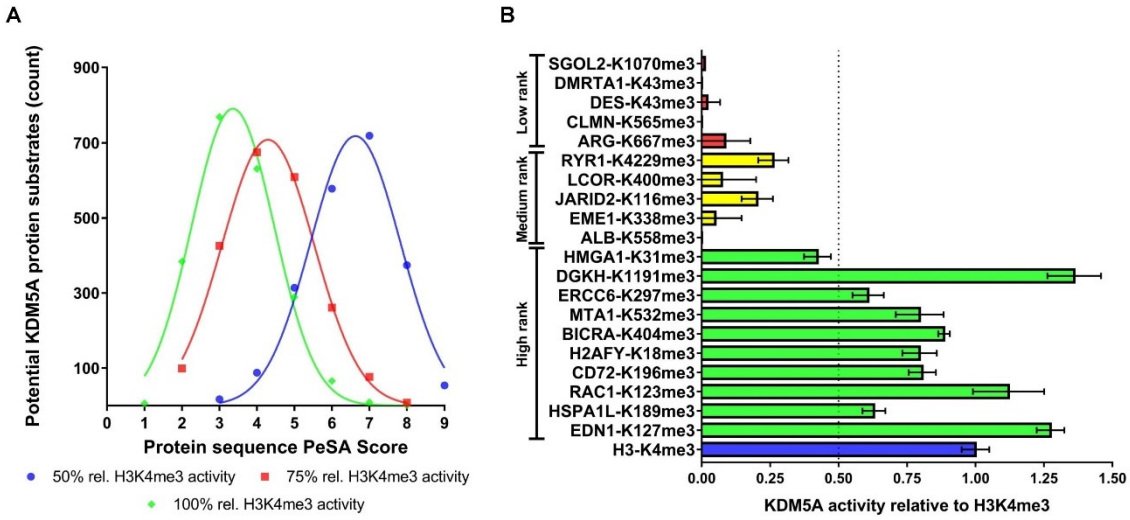



**Figure 3.3. Quantification of relative KDM5A$_{1\text{-}801}$ activity towards high, medium and low ranked substrates.** ***A.*** KDM5A predicted substrates based on KDM5A recognition motif. A total of 2,155 known methylation sites (from the PhosphoSitePlus database) were scored against each KDM5A recognition motif. ***B.*** Detection of KDM5A$_{1\text{-}801}$ demethylation activity towards high, medium and low ranked synthesized substrate peptides. Results are mean ± std. error (n=3). Reprinted with permission of *Analytical Biochemistry.*

**Table 3.1. Candidate protein substrates of KDM5A**

| Uniprot ID | Gene | Residue | Methylation state | Protein function |
|---|---|---|---|---|
| P05305 | EDN1 | K127 | Kme2[140] | Vasoconstrictor |
| P34931 | HSPA1L | K189 | Kme1[140] | Protein folding |
| P63000 | RAC1 | K123 | Kme1[140] | GTPase-mediated cell signaling |
| P21854 | CD72 | K196 | Kme1[140] | B-cell proliferation + differentiation |
| O75367 | H2AY | K18 | Kme1[141] | Chromatin regulator |
| Q9NZM4 | BICRA | K404 | Kme1[140] | Chromatin remodelling |
| Q13330 | MTA1 | K532 | Kme1, Kme2[142] | Transcription co-regulator |
| Q03468 | ERCC6 | K297 | Kme1, Kme2[111] | DNA damage repair |
| Q86XP1 | DGKH | K1191 | Kme3[140] | Lipid metabolism |

## Discussion

While there have been many advancements dedicated to expanding knowledge of the methyl-lysine proteome, limitations surrounding systematic identification of KDM substrates persist. Identification of sites of protein methylation by MS has assisted in this hurdle, although many sites identified lack of a methyl-modifying enzyme that facilitates the addition or removal of these methylation modifications[71]. Various proteomic methods such as K-OPL arrays have been used to aid in the identification of KMT substrates[135]. The result is a vastly greater identification of KMT substrates compared to KDM substrates due to a lack of standard proteomic methods developed for rapid, yet systematic identification of KDM substrates. Here, we demonstrate a systematic method to overcome barriers in KDM substrate discovery through a combination of techniques using peptide permutation array-based binding assays along with demethylation activity assays to relatively determine KDM activity. However, it should be highlighted that this method does require the synthesis of many methylated peptides that can be cost-prohibitive without access to a peptide synthesizer.

Through the analysis of KDM activity on systematic mutations of an established substrate (i.e., H3K4 for KDM5), we show that is possible to rank potential methylation sites based on their predicted potential to be demethylated *in vitro*. Our method ultimately led to an enrichment of candidate substrates that show potential as viable targets of KDM5A demethylation (Fig. 3.3). Based on an activity threshold of 50% of H3K4me3 demethylation activity, we determined the

candidate substrate recognition motif for KDM5A. This motif became the basis of our efforts to predict new substrates. Unlike other indirect methods of KDM substrate discovery that are based on protein interaction data, our permutation-based method is able to provide insight into the residue preferences and may help to predict new substrates as the methyl-lysine proteome continues to grow[95,96,136]. Given the high effort and time required to validate such enzyme-substrate relationships, such information is critical to prioritize the identification of new substrates. Ultimately, we successfully validated 90% of our high-confidence substrate predictions *in vitro*. Although we restricted our analysis to solely the testing of our method and not a complete exploration of all 54 high-confidence substrates, we identified nine new candidate substrates of KDM5A that should be further investigated within a biologically relevant system.

Almost all studies to date have focused on the role of KDM5A to regulate gene expression through removal of the Kme3 and Kme2 modifications at the K4 site in histone H3[69]. Although this is clearly pertinent to the growing implications of how KDM5A may contribute to disease progression, how KDM5A specifically promotes cancer cell proliferation and drug tolerant persistent cells is still not yet entirely clear[69,143]. Given how dynamic and complex the methyl-lysine proteome has been shown to be, and the growing number of non-histone KMT substrates, it is critical to begin exploring new substrates that might facilitate the roles of KDM5A in disease[69,93,103]. Indeed, previous studies have already identified how the KDM5 family can manipulate a number of signaling pathways involved in cell growth, differentiation, apoptosis, and general cellular homeostasis[78,80,103,104,144].

For example, it was observed that cytoplasm-localized KDM5C was shown to interact with Smad3, a mediator of the TGFβ signaling pathway[137]. Further, it was demonstrated that this cytoplasmic KDM5C-Smad3 interaction was dependent on catalytic activity[137]. Taken together, this provides the intriguing possibility that a KDM5 family member may target substrates outside of histones. In recent years, there has been a increase in evidence that histone methyl-modifying enzymes can also target non-histone substrates for methylation/demethylation[56]. Proteomic databases (such as PhosphositePlus) have collected thousands of lysine methylation sites in human proteins, majority of which are non-histone proteins[130]. With all these lysine methylation sites now identified, it is critical that methods be developed to help determine the methyl-modifying enzymes that regulate methylation at these sites and their functional implications[69,93].

Supporting our attempt to systematically identify non-histone targets of KDM demethylation, our analysis yielded the identification of metastasis-associated protein 1 (MTA1)-K532 methylation as a candidate substrate for KDM5A. This analysis included the formation of an interaction network assessed through the STRING protein interaction database, in order to determine if any of the new 9 candidate substrates have been shown to physically interact or be associated in some regard with KDM5A (Fig. 3.4)[145]. MTA1-K532 methylation has been previously characterized as a non-histone substrate of the KMT G9a with functional implication in the association of MTA1 with NuRD co-repressor and NuRF co-activator complexes as a result of regulating methylation status of the K532 residue [142,146]. This dynamic relationship between MTA1 methylation

state and formation of co-activator/-repressor complex plays an essential role in chromosomal remodelling and the subsequent fate of transcription[142,146]. KDM5A is also known to physically interact with the NuRD complex to cooperatively control the expression of developmentally regulated genes[147]. Furthermore, MTA1-K532 is a substrate of the H3K4me1 monomethyl-demethylase, LSD1 site[142]. Given that both LDS1 and KDM5A target the same H3K4 site, it is intriguing to hypothesize that KDM5A may also target MTA1-K532 for demethylation; insight that has been uniquely provided through our substrate preference analysis (Appendix C, Fig. C1).

## Conclusion

As demonstrated, this method permits the systematic characterization of substrate preference for JMJD-type KDM enzymes. This is significant as it has been recently estimated that the methyl-lysine proteome may contain significantly more modification sites than that are currently known[71]. Using KDM5A as our model demethylase, our method has shown potential for refining substrates to a testable number of high-confidence sites. Ultimately, this approach is applicable to all JMJD KDM enzymes with at least one known substrate. It also holds potential as a method that can shed light onto new KDM substrates and spark much needed growth in the discovery of KDM-substrate networks, beyond histones.

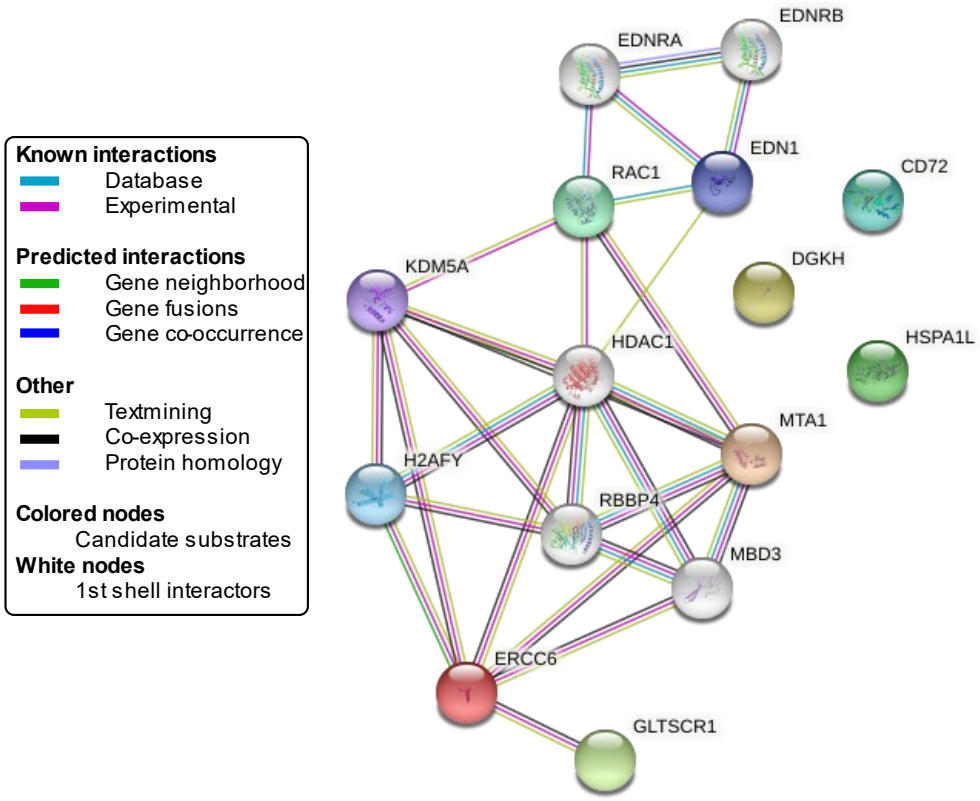



**Figure 3.4. Insight into biological relevance of predicted KDM5A substrates.** STRING interaction network of *in vitro* candidate substrates of KDM5A. Interactions derived from the String protein interaction database [145]. Reprinted with permission of *Analytical Biochemistry.*

# Chapter 4

# Characterization of KDM5 lysine demethylase family substrate preference and identification of novel substrates

**Matthew Hoekstra[1,2], Nashira H. Ridgeway[1,2], Kyle Biggar[1,2]**

[1]Institute of Biochemistry, Carleton University, Ottawa, ON, K1S 5B6, Canada

[2]Department of Biology, Carleton University, Ottawa, ON, K1S 5B6, Canada

## Submitted manuscript

Hoekstra, M., Ridgeway, N.H., Biggar, K.K. (2022). Characterization of KDM5 lysine demethylase family substrate preference and identification of novel substrates. ***J. Biochem.,*** *In review at time of thesis submission.*

## Chapter Summary

In Chapter 3, we demonstrated the ability to apply our method and pipeline to determine if our approach can accurately predict high-ranking substrates, as well as validate substrates *in vitro*. In this chapter, we will summarize how we have applied our method and pipeline to every KDM5 family member to identify 66 high ranking candidate substrates from within the KDM5 family. This includes defining the substrate preference of each KDM5, prediction of substrates, and *in vitro* validation of all high-ranking substrates for each KDM5 demethylase. Moreover, we have analyzed prediction success rate, as well as shown that we can monitor changes in cellular methylation in a handful of the 66 high ranking candidate substrates as a response to KDM5 inhibition.

Ultimately, this chapter defines specific cellular substrates that appear to be influenced by KDM5 inhibition. The following section are from my submitted manuscript that describes the full analysis of characterizing KDM5A/B/C/D substrate preference and identifying potential substrates. It also describes our *in vitro* validation of highly ranked substrates as well as investigating these substrates in a cellular context. Notably, this manuscript represents the first evidence that members of the KDM5 demethylase family may target non-histone proteins.

## Manuscript Summary

The KDM5/JARID1 sub-family are 2-oxoglutarate and Fe(II)-dependent lysine-specific histone demethylases that are characterized by their Jumonji catalytic domains. The KDM5 family is known to remove tri-/di-methyl modifications from lysine-4 of histone H3 (i.e., H3-K4me2/3), a mark associated with active gene expression. As a result, studies to date have revolved around KDM5's influence on disease through their ability to regulate H3-K4me2/3. Recently, evidence demonstrated that KDM5's may influence disease beyond H3-K4 demethylation, making it critical to further investigate KDM5 demethylation of non-histone proteins. To help identify potential non-histone substrates for the KDM5 family, we developed a library of 180 permutated peptide substrates (PPS), with sequences that are systematically altered from the wild-type H3-K4me3 sequence. From this library, we characterized recombinant KDM5A/B/C/D substrate preference and developed recognition motifs for each KDM5 demethylase. The recognition motifs developed were used to predict potential substrates for KDM5A/B/C/D and profiled to generate a list of high-ranking and medium/low-ranking substrates for further *in vitro* validation for each KDM5. Through this approach, we identified 66 high-ranking substrates in which KDM5 demethylases displayed significant *in vitro* activity towards. Herein, we looked to explore the potential of cellular validation of the KDM5 family-mediated demethylation of several of these candidate substrates.

## Introduction

Histone proteins and their diverse array of post-translational modifications (PTMs) have been subject to exquisite evolutionary conservation in eukaryotes[148]. Accordingly, the factors that control the deposition, removal, and interpretation of histone modifications are themselves deeply conserved, with many strongly impacting disease development[148,149]. As the dysregulation of these histone modifying enzymes become increasingly recognized as drivers of disease development, there has been a fundamental need to characterize the specific mechanisms that facilitate their function(s)[115,150]. One of the more recent well-studied PTMs on histones is the methylation of lysine residues[115,150]. Specific lysine residues on histone proteins can be mono-, di-, or tri-methylated (Kme1, 2, 3) and these methylation states have distinctive mechanistic consequences that can result from the recruitment of binding proteins specific to each methylated state[115,151]. The regulation of histone lysine methylation requires specific enzyme families that read, write, and erase methyl marks on histone proteins in a sequence-specific approach. The methyl readers of the histone code, or methyl-lysine binding domain (MBDs) proteins, bind to various lysine residues of varying degrees of methylation[3,30]. As a result, the methyl writers of the histone code are lysine methyltransferase (KMTs) enzymes and the methyl erasers are lysine demethylase (KDMs) enzymes[3,29,51]. There are two families of KDM enzymes. One of the KDM families are the flavin adenine dinucleotide (FAD)-dependent amine oxidase family, referred to as the KDM1/LSD family of demethylases[29,51]. The other KDM family is the Jumonji domain (JmjD) containing family of enzymes[29,51]. This family of JMJD-containing KDM functions

as $Fe^{2+}$ and 2-oxoglutarate-dependent oxygenases[29,48,69] and is the largest of the KDM families, containing 30 demethylases classified into several subfamilies[29,51]. In whole, the JMJD family of KDMs target a wide array of lysine methylation degrees and lysine-containing sequences for demethylation[51].

Recently there has been a large increase in experimental evidence through primary tumors and model systems that suggests a potential dynamic role the KDM5 family has as oncogenic drivers[28,48,69,73,102,131,143,152]. The KDM5 family catalyzes the removal of methyl groups from tri- and di-methylated lysine 4 of histone 3 (H3K4me3/2)[48,69] and consists of four members: KDM5A, KDM5B, KDM5C and KDM5D. Evidence demonstrating how the KDM5 family functions to drive chemoresistance and drug-tolerant persister (DTP) cancer cells has supported these claims[102,143]. These revelations have made it clear how critical investigation into the molecular functions of this KDM5 family of demethylases is. As this family was only discovered in the last 15 years or so, there has been minimal research completed outside the scope of KDM5's only known substrate, H3K4[48,70,79,153]. This includes substrate interaction and gene transactivation in response to KDM5 expression[48].

As KDM enzymes have been identified to influence disease progression by their ability to regulate gene expression through removing histone lysine methyl modifications, there is also growing evidence that KDM enzymes are capable of demethylating non-histone proteins[67,154]. In fact. the number of identified KMT and KDM enzymes with non-histone substrates is steadily growing but remains outpaced by the discovery of methylated proteins by mass

spectrometry (MS)[71,99,100]. Currently more then 5000 unique sites of lysine methylation on more then 2500 unique proteins have been identified[71]. Unfortunately, very few of these identified sites have been linked to any methyl-modifying enzyme[71]. As a result a number of proteomic methods have been developed to help identify substrates of both KMTs and KDMs, yet the discovery of KMT substrates have vastly outpaced those of KDM enzymes due to a lack of a systematic screening platform to map substrate selectivity[90,95,96,99].

Studies to date have revolved around the ability of KDM5 to regulate gene expression through removal of methyl-modifications at the histone H3K4 site[48,69]. However, the mechanism(s) through which the KDM5 family members differentially promote cancer cell proliferation and drug tolerant persistent cells is still not entirely clear[69,143]. Recently, KDM5 research has began to branch away from the ability of KDM5 to regulate specific gene expression and more towards characterising the contributions within the KDM5 family[48,69]. As a result, recent research has demonstrated growing evidence that KDM5 enzymes might too be capable of demethylating non-histone proteins[6,69,137]. Indeed the discovery of new substrates could help identify the new role(s) of KDM5 in diseased progression, beyond solely its function as solely a H3K4me3/2 demethylase[6,69,93]. To systematically identify potential novel protein substrates of the KDM5 family, we developed a method to study *in vitro* substrate specificity[124]. Our approach is derived from combining demethylation activity assays to determine relative KDM5 activity and peptide permutation array-based activity assays used for KMT substrate identification[84,124]. The method utilizes a library of trimethylated lysine-

containing peptides derived by systematically mutating the canonical H3K4me3 substrate peptide. Using this permutated peptide substrate (PPS) library, we can determine specific amino acid preferences displayed by the KDM5 family of demethylases. Using this information, recognition motifs can be proposed for each of KDM5A/B/C/D and used to predict a ranked list of candidate demethylation substrates to validate *in vitro*. Using this approach, we aim to prioritize the methyllysine proteome for KDM5 substrates and focus further validation research efforts on.

## Materials and Methods

*Recombinant KDM5A/B/C/D expression and purification*

*Spodoptera frugiperda* (Sf9) cells (Novagen) were maintained at 26 ºC in a humidified chamber in Grace's Insect Media (ThermoFisher Scientific, Cat# 11605094) supplemented with 10% fetal bovine serum (ThermoFisher Scientific, Cat# 12484028) and 1% Penicillin-Streptomycin (ThermoFisher Scientific, Cat# 10378016). For recombinant protein purification, Sf9 cells (500 mL at $10^6$ cells/mL) were infected with P3 baculovirus for either KDM5A$_{1\text{-}801}$-6XHis (JARID1AA-c180 pFB-LIC-Bse), KDM5B$_{25\text{-}770}$-6XHis (JARID1BA-c165 pFB-LIC-Bse), KDM5C$_{1\text{-}765}$-6XHis (JARID1CA-c022 pFB-CT10HF-LIC) or KDM5D$_{1\text{-}775}$-6XHis (JARID1DA-c066 pFB-LIC-Bse) (generously gifted from the Structural Genomics Consortium (SGC) Toronto) at a 1:100 ratio. Cells were grown spinning at 120 rpm for 65 hours at 26ºC. Cells were collected by centrifugation, snap frozen and stored at -80ºC. Cells were resuspended in P5 buffer (50mM $NaHPO_4$ (pH 7), 500 mM NaCl, 10% glycerol, 0.05% TritonX-

100, 1 mM DTT, 5 mM Imidazole) supplemented with protease and phosphatase inhibitors (1 mM PMSF, 0.5 µM Aprotinin, 10 µM E-64, 1 µM Pepstatin, 1 mM $Na_3VO_4$). Cells were lysed through 15 passes through a Dounce homogenizer at 4°C. Cells were subsequently sonicated 3x for 30 sec at 40% amplitude intensity followed by another 15 passes through a Dounce homogenizer at 4°C. Cells were then centrifuged at 18000 x g for 45 min at 4°C. The recombinant protein was purified via standard Ni-NTA chromatography. Briefly, the soluble cell lysate was incubated with 250 µL HisPur™ Ni-NTA Resin (ThermoFisher Scientific, Cat# 88221) for 2.5 hour at $4^0$C and washed with P40 buffer (P5 buffer containing 40 mM imidazole). Recombinant protein was eluted with P500 buffer (P5 buffer containing 500 mM imidazole) and dialyzed overnight into storage buffer (25 mM HEPES (pH 7.5), 150mM NaCl, 10% glycerol, 1mM DTT) using 12-16kDa molecular weight cut-off dialysis tubing (FisherScienceEducation™, Cat# S25645F). Protein concentration was quantified through a standard Bradford assay[138]. Proteins were snap-frozen in liquid nitrogen and stored at -80°C for later use.

*Peptide Synthesis*

Peptides were synthesized *de novo* in house using a ResPep SL (Intavis Bioanalytical Instruments) peptide synthesizer. Fmoc protected amino acids, including lysine methylated amino acids, were purchased from p3bio systems. Standard Fmoc chemistry was used to produce peptides on Rink amide resin (p3bio systems). After synthesis, peptides were cleaved from the resin and side

chains were deprotected using an acidic cleavage solution (95% trifluoroacetic acid, 3% triisopropylsaline, and 2% water). Peptides were precipitated overnight at -20$^{o}$C with ice-cold diethyl ether and washed the next day with ice-cold diethyl ether and allowed to dry overnight at 23$^{o}$C. Peptides were resuspended in 1X PBS solution, quantified, aliquoted and frozen at -20$^{o}$C until needed.

*Permutated peptide substrate (PPS) library and ranked peptide design*

To design our library of permutated peptide substrates, we systematically mutated the H3K4 sequence (ART<u>K</u>QTARKSTGGKA; K4 position underlined) in 180 peptides to cover a 9-residue peptide window. In essence, each peptide, trimethylated at K4, contains a single amino acid mutation to one of the remaining 19 amino acids in the positions surrounding the trimethylated lysine site. For our analysis of H3K4, we mutated the -3 position to +5 position surrounding the trimethylated K4 position (i.e., $A_{-3}R_{-2}T_{-1}$<u>Kme3</u>$Q_{+1}T_{+2}A_{+3}R_{+4}K_{+5}$; K4 position underlined).

To design our ranked peptides, we first identified how to classify our high, medium, and low-ranking peptides. We then designed our ranked substrate peptides by synthesizing 11-mer peptides whose sequences reflected a known methylated lysine site in a specific protein. Essentially, we took the specific lysine residue and synthesized the surrounding 5 residues on either side of a specific lysine site. It should be noted that for each plate synthesis, a WT H3K4me3 and WT H3K4me0 peptide was included to compare relative demethylase activity against.

*KDM demethylation assay*

All KDM5A$_{1-801}$, KDM5B$_{25-770}$, KDM5C$_{1-765}$ and KDM5D$_{1-775}$ *in vitro* enzyme characterization assays and permutated peptide substrate library experiments were performed as described. All reactions took place in demethylase reaction buffer (25 mM HEPES (pH 7.5), 10 µM peptide substrate, 100 µM ascorbate (J.T Baker Chemical Co), 10 µM Fe(II) sulfate (BDH chemicals), 10 µM α-ketoglutarate (Sigma Aldrich), 1% v/v DMSO) for 1 hour at 23$^{o}$C in a 384-well Corning flat-bottom microplate. Following incubation, succinate detection reagents were added (Succinate Glo™ JMJC Demethylase Assay; Promega Corp, Cat# V7991) according to manufacturer's instruction. Briefly, 5 µL of Succinate detection reagent I was added and incubated for 60 min followed by 10 µL of Succinate detection reagent II. After 10 minutes, luminescence was read using a Cytation 5 plate reader (Bio-Tek, Winooski, VT) with a plate height of 1.25 mm and a gain set to 170.

The permutated peptide substrate library experiments were carried essentially the same as described above. Briefly, 2.5 µL of 2X reaction buffer (without peptide substrate) was aliquoted to a 384-well plate at 4$^{o}$C. While at 4$^{o}$C, 1.25 µL of 40 µM permutated H3K4me3 peptide were added to separate wells. Temperature was increased to 23$^{o}$C and 1.25 µL of 4X KDM5 enzyme (100, 640, 1600, and 1400 nM for KDM5A/B/C/D, respectively) in 25 mM HEPES (pH 7.5) was added. The reactions proceeded for 1 hour at 25$^{o}$C, followed by the addition of the succinate detection reagents, quenching the reaction, and reading

luminescence as described above. KDM5A/B/C/D activity towards ranked protein peptide substrate experiments were carried out in the same manner.

*SKBR3 cell treatment with CPI-455*

SKBR3 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 4.5g/L glucose, L-glutamine and sodium pyruvate (Wisent Inc, Cat# 319-005-CL) plus 10% fetal bovine serum (ThermoFisher Scientific, Cat# 12484028) and 1% penicillin-streptomycin (ThermoFisher Scientific, Cat# 10378016) and cultured in a humidified incubator at 37$^{o}$C and 5% $CO_2$. CPI-455 (Caymen Chemical Company, Cat# 22127) were dissolved in 100% DMSO to 10mM and stored at -20$^{o}$C. SKBR3 cells were seeded in duplicated in 10cm$^2$ plates (4.2 million cells/plate) and allowed to adhere overnight. The next day, cells were treated with 12.5 µM CPI-455 or DMSO (vehicle) for 24 hours. After 24 hours, cells were trypsinzed, collected and washed once with ice cold sterile 1X PBS and centrifuged (1000 RPM for 5 min). Supernatant was removed and pellets were snap frozen and stored at -80$^{o}$C for subsequent cell lysis.

*Histone isolation*

Histone isolation generally followed the protocol described in Shechter et al., 2007[155]. Cell pellets were thawed on ice and resuspended in hypotonic lysis buffer containing 10 mM-Tris-HCl solution (pH 8.0), 1 mM KCl 1.5mM $MgCl_2$, 1mM DTT and protease inhibitors (ThermoFisher Scientific, Cat# A32953). Cells were lysed at 4$^{o}$C for 90 min while rotating and collected by centrifugation

(10,000 x g for 10 min at 4°C). The pellet was resuspended with 400µL 0.4N $H_2SO_4$ and allowed to rotate overnight 4°C. Cells were collected by centrifugation (16,000 x g for 10 min at 4°C) and supernatant was removed and placed in a new 1.5 mL tube. 132 µL of 6.1N trichloroacetic acid solution was added slowly dropwise into middle of solution and inverted 6 times. Histones were allowed to precipitate at 4°C overnight. Precipitates were pelted by centrifugation (16,000 x g for 10 min at 4°C) and supernatant was poured off. Pelleted histones were washed twice with 400µL ice-cold acetone. Histone pellets was air-dried, and histones were resuspended with 60 µL $H_2O$. Quality of histone prepped was assessed by SDS-PAGE and western blot analysis.

*Mass spectrometry analysis of SKBR3 cell protein lysate*

Cell pellets were thawed on ice and resuspended in lysis buffer containing 50 mM-Tris-base solution (pH 7.6), 150 mM NaCl 10% glycerol, 0.5% Triton X-100, 1mM DTT, and protease inhibitors (1 mM PMSF, 0.5 µM Aprotinin, 10 µM E-64, 1 µM Pepstatin). Cells were lysed at 4°C for 60 min while rotating and collected by centrifugation (18,000 x g for 30 min at 4°C). Protein supernatant was collected and quantified through Bradford assay[138]. A total of 1 µg of total protein lysate was subjected to trypsin proteolytic digestion. Briefly, 1 ug of total protein was diluted to a volume of 100µL with $(NH_4)HCO_3$ (pH 7.8) to a final concentration of 50 mM $(NH_4)HCO_3$. Trypsin was added to a final working concentration of 20 ng/µL (or 833 nM) and incubated overnight while rotating at 37°C. Samples were C18 cleaned using zip-tip C18 cleaning tips (EMD Millipore,

Cat# ZTC18S096). Samples were centrifuged in a roto-vac to evaporate solvent and dried peptides were resuspended in 20 µL of 0.1% formic-acid $H_2O$. For this analysis, digests from two biological replicates were combined to go forward with.

Digests were analyzed by positive ESI LC-MS/MS on a Thermo Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer. Briefly, peptide digests were loaded onto a Thermo Easy-Spray analytical column with an Easy-nLC 1000 chromatography pump coupled to a Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer. Peptides were separated on a 125-min (5–40% acetonitrile) gradient. The spectrometer was set in full MS/data-dependent-MS2 TopN mode: mass analyzer over a mass-to-charge ratio (m/z) range of 400–1,600 with a mass resolution of 70,000 (at m/z = 200), 35 NCE (normalized collision energy), 2.0 m/z isolation window, and 15-s dynamic exclusion. The isolation list with the mass [m/z] and the sequences of the peptides used to identify candidate substrates by PRM-MS were recorded. Each trace on the chromatograph represents the detection of each individual transition ion used to monitor site-specific methylation of a candidate substrate. The retention time provides an indication that the transition ions result from the same parent peptide (correlates with the time that the parental peptide had eluted from the C18 column). The *in silico* protease digest patterns (i.e., to generate precursor ions) and the corresponding PRM transitions were compiled using the Skyline software[156]. Transitions that are larger than the precursor ion was selected based on the Skyline predictions and the specific b/y ions that allow unambiguous identification of the methylated Lys site were included and used in the isolation list. Positive

identification of a new methylation site required the successful detection of at least three transition ions.

*Western blotting*

Histone samples isolated from SKBR3 cells were separated on a standard 15% tris-glycine SDS-PAGE gel and transferred to 0.2 µm PVDF membrane (Bio-Rad, Cat# 1620177) for western blotting analysis. Immunoreactive bands specific to H3 were detected using anti-H3 primary antibody (1:2000; ThermoFisher Scientific, Cat# 702023). Immunoreactive bands specific to H3K4me3 were detected using anti-H3K4me3 primary antibody (1:2000; ThermoFisher Scientific, Cat# 703954).

Purified KDM5A/B/C/D were separated on a standard 8% tris-glycine SDS-PAGE gel and transferred to 0.45 µm PVDF membrane (GE Healthcare Life Sciences, Cat# 10600023). Immunoreactive bands specific to KDM5A were detected using anti-KDM5A primary antibody (1:800; Santa Cruz Biotechnology, Cat# sc-365993). Immunoreactive bands specific to KDM5B were detected using anti-KDM5B primary antibody (1:1000; Santa Cruz Biotechnology, Cat# sc-517291). Immunoreactive bands specific to KDM5C were detected using anti-KDM5C primary antibody (1:1000; Santa Cruz Biotechnology, Cat# sc-376255). Immunoreactive bands specific to KDM5D were detected using anti-KDM5D primary antibody (1:1000; ThermoFisher Scientific, Cat# PA5-40120). All histone and KDM5 primary antibodies were detected with either of rabbit anti-mouse HRP-conjugated secondary antibody (1:10,000; ThermoFisher Scientific, Cat#

31450) or goat anti-rabbit HRP-conjugated secondary antibody (1:10,000; ThermoFisher Scientific, Cat# G21234). Bands were visualized via chemiluminescence using Clarity™ Western ECL Substrate (Bio-Rad, Cat# 170-5060) with a Gel Doc™ XRS+ Imaging system (Bio-Rad).

*Data analysis*

With the relative demethylation activity derived from our mutated H3K4 peptide library, we used Peptide Specificity Analyst (PeSA) software to determine a candidate substrate recognition motif for *in vitro* substrates (i.e., a specific sequence of amino acids that required to be a KDM5A/B/C/D substrate)[125]. These weight-based motifs are constructed using the activity data and a user defined threshold that is set relative to the positive control[125]. The threshold reflects the minimum activity required for any mutated peptide substrate to have KDM5A/B/C/D activity exhibited towards it. For example, if the user would like to include any amino acid mutations that retained 50% of the WT H3K4me3 demethylation activity, then they could produce a 50% activity threshold recognition motif.

The KDM5A/B/C/D substrate specificity matrices (Appendix D, Fig. D3) were utilized to apply hierarchical clustering, to compare substrate specificity for KDM5A/B/C/D. Using the Pandas[157,158] package for Python 3[159] the motifs were represented sequentially. Two randomized motifs were added to the analysis to serve as negative controls. The Scipy[160] package was then used to apply the Ward variance minimization algorithm[161] to hierarchically cluster each motif.

These results were visualized through Matplotlib[162] for Python 3[159]. The code is available via https://github.com/nashirag/Substrate_Specificity_Clustering.

## Results

*Quantification of KDM5 family relative activity towards WT H3K4 substrates*

We have developed a method that combines a luminescence-based demethylation assay (Appendix D, Fig. D1A) with 15-mer WT H3K4 tri-methylated peptide substrates (Fig 4.1A) to measure relative KDM5 demethylase activity. KDM5 proteins were expressed via *Spodoptera frugiperda* (Sf9) cells and purified through standard Ni-NTA chromatography (Appendix D, Fig. D2**)**. Once KDM5A$_{1\text{-}801}$, KDM5B$_{25\text{-}770}$, KDM5C$_{1\text{-}765}$ and KDM5D$_{1\text{-}775}$ were purified, an optimum enzyme concentration required for the selected assay conditions was identified via an 2x enzyme dilution curve (Appendix D, Fig. D1B). As we observed a dose-responsive signal for all of KDM5A/B/CD, we selected concentrations within the linear range to carry out our PPS library assays. Specifically, a concentration of 25, 160, 400, and 350 nM were chosen for KDM5A/B/C/D, respectively.

As all KDM5s have been shown to display catalytic activity towards different methylation states, (i.e., Kme3 vs. Kme1), we aimed to confirm the specificity of KDM5A/B/C/D for H3K4me0/1/2/3 peptides (Fig 4.1B). As expected, all the KDM5's displayed the highest demethylation activity towards the H3K4me3 substrate. Similarly, for the H3K4me2 substrate, KDM5A/B/C/D displayed 51%, 60%, 43%, and 45% of the luminescent signal relative to the H3K4me3 substrate reaction ($p < 0.05$). For KDM5A/C/D, the H3K4me1/0

substrates were shown to not produce a signal significantly above background level. KDM5B displayed a low level of *in vitro* activity for the H3K4me1 substrate. From these results we were confident that all the KDM5 enzyme concentrations and the reaction parameters for the succinate-based detection assay were optimized to determine which amino acid substitutions in the WT H3K4me3 substrate would alter KDM5 substrate recognition.

*Application of H3K4me3 PPS library to determine KDM5 family substrate specificity and preference*

As H3K4me3/2 to date is the only known protein substrate for the KDM5 family, our permutated peptide substrate (PPS) library was based on the systematic mutation of the H3K4 sequence. This resulted in a library of 180 unique peptide sequences, whereby each peptide contains a single amino acid mutation to one of the remaining 19 amino acids. The activity of KDM5A/B/C/D towards each permutated peptide was then assessed via demethylation assay. Figure 4.2A demonstrates the $KDM5A_{1\text{-}801}$ substrate specificity matrix that illustrates the changes in $KDM5A_{1\text{-}801}$ activity as a result from each permutated peptide substrate. All $KDM5A_{1\text{-}801}$, $KDM5B_{25\text{-}770}$, $KDM5C_{1\text{-}765}$ and $KDM5D_{1\text{-}775}$ substrate specificity matrices can be seen in Appendix D (Fig. D3). KDM5A/B/C/D activity was monitored relative to the wild-type (WT) H3K4me3 (i.e., positive control) and the WT H3K4me0 (i.e., negative control) peptides. This strategy of normalization permits us to determine which amino acid mutations

resulted in an either null, positive, or negative influence on *in vitro* KDM5A/B/C/D activity.

After we derived relative KDM5A/B/C/D demethylation activity from our H3K4me3 PPS library, we used PeSA software to determine candidate substrate recognition motifs for *in vitro* substrates for each KDM5 demethylase[125]. As our goal was to predict potential substrates, we chose to continue to substrate prediction using the 50% recognition motifs for each of KDM5A/B/C/D (Fig 4.2B; Appendix D, Fig. D4). Therefore, any amino acid substitutions maintaining 50% relative H3K4me3 activity were factored into our substrate predictions. Moreover, a threshold of 50% relative activity can be considered permissive when predicting substrates and reduces the stringency of downstream predictions.

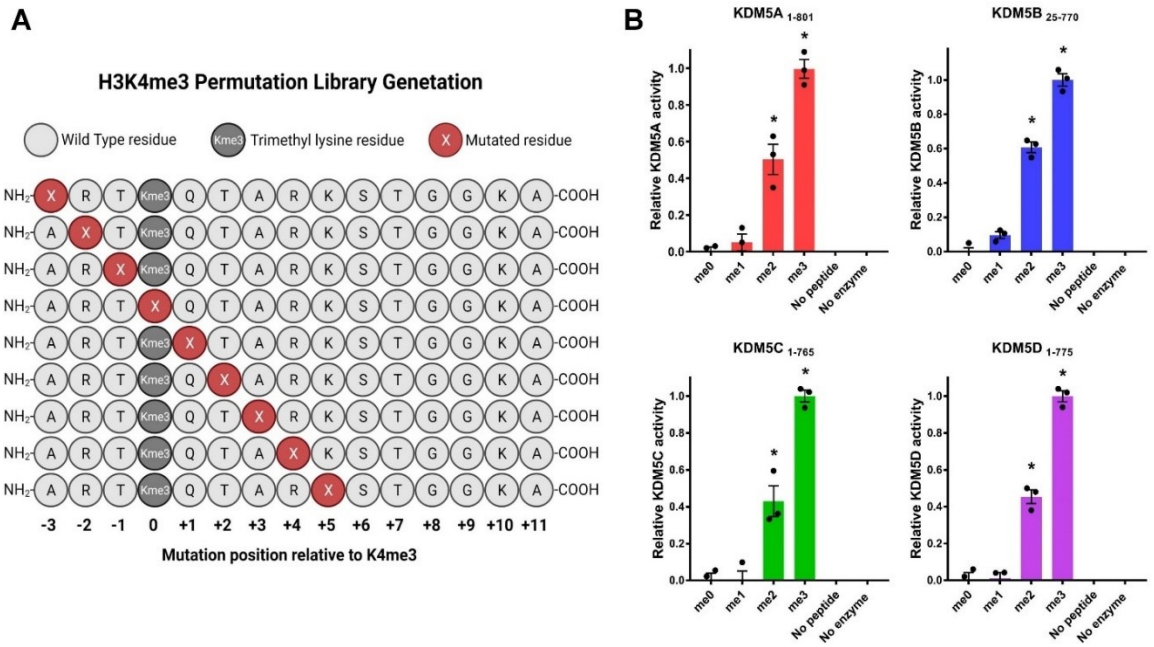



**Figure 4.1. Setup of PPS System.** ***A.*** Graphical representation of generation of H3K4me3 permutation library. ***B.*** Luminescence-based detection of KDM5A$_{1\text{-}801}$, KDM5B$_{25\text{-}770}$, KDM5C$_{1\text{-}765}$ and KDM5D$_{1\text{-}775}$ demethylation of WT H3K4me0 peptide, WT H3K4me1 peptide, WT H3K4me2 peptide, WT H3K4me3 peptide. Luminescence averages were normalized by background signal, followed by normalization to WT H3K4me3 peptide luminescence (considered as 1.00). Results are mean ± S.E.M of n=3. (* indicates $p < 0.05$ from me0 data).

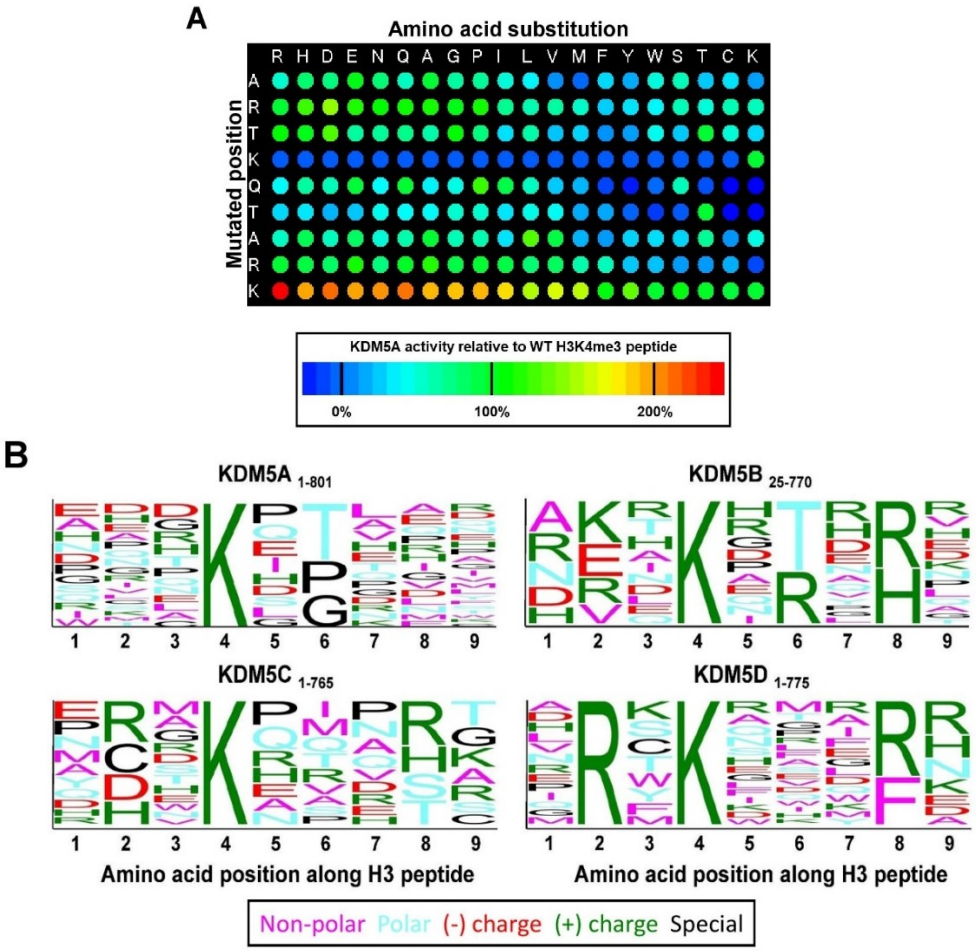



**Figure 4.2. Assessing KDM5A/B/C/D substrate specificity through a permutated peptide substrate (PPS) library.** ***A.*** Detection of relative KDM5A$_{1\text{-}801}$ demethylation activity with mutated Lys4 tri-methylated histone H3 peptide substrates. Data was normalized to the WT H3K4me3 peptide at each mutation position (considered as 1.00 for each position). Results are mean of n=3. Array image developed from PeSA[37] ***B.*** KDM5A$_{1\text{-}801}$, KDM5B$_{25\text{-}770}$, KDM5C$_{1\text{-}765}$ and KDM5D$_{1\text{-}775}$ recognition motifs derived from relative KDM5A/B/C/D demethylation activity towards PPS library. Each motif is reflective of amino acid substitutions that maintained 50% of KDM5A/B/C/D's activity relative to their H3K4me3 activity. Motif images developed from PeSA[37].

*Identifying potential high-ranking substrates for KDM5 family*

Using our KDM5A/B/C/D 50% recognition motifs, we then predicted candidate KDM5A/B/C/D substrates within the *Homo sapiens* methyllysine proteome (PhosphositePlus database; accessed on 12-03-2020)[130]. Using this catalog of over 2000 methylated lysine sites, a list of candidate substrates was produced and scored against the 50% recognition motif for each of KDM5A/B/C/D. Candidate substrate scoring was done using the PeSA algorithm[125]. PeSA predictions are based on the number of amino acid residues within a predicted substrate that match the tolerable amino acid substitutions in the KDM5A/B/C/D recognition motifs. As we collected relative KDM5A/B/C/D activity using a peptide substrate containing only 9 mutated positions, predictions were made based on whether a sequence matched our motif (X/9 positions). Therefore, the highest PeSA score that could be obtained for any substrate which contains a peptide window exactly matching one of the recognition motifs, is 9/9. When the threshold is set to 50%, the Gaussian distribution of the known methylation sites and their PeSA scores results in a mean score of 6.6 (±1.2), 4.4 (±1.2), 4.2 (±1.3), 5.2 (±1.2) for KDM5A/B/C/D respectively (Fig 4.3A). We decided to define any substrate with a score greater than 2 standard deviations above the mean score as a “high ranked” candidate substrate. Therefore, this denoted that any substrates with PeSA scores of 9 (KDM5A), 7-9 (KDM5B), 7-9 (KDM5C), and 8-9 (KDM5D) were classified as “high rank”. Likewise, “middle ranked” substrates were classified with a score near the mean, and “low ranked” substrate with a score of less than 2 standard deviations below the mean. As a

result of this analysis, the number of substrates ranked as high-confidence substrates for KDM5A was 54 substrates, KDM5B had 52 substrates, KDM5C had 77 substrates and KDM5D had 17 substrates (Fig 4.3B).

*Hierarchical Clustering Analysis of KDM5 substrate specificity matrices*

Once substrate homology between KDM5's was assessed, we then began to investigate whether the KDM5A/B/C/D substrate specificity matrices (Appendix D, Fig. D3) themselves possessed a degree of homology between them. This inquiry stems from the observed sequence homology between KDM5's. Specifically, as there appears to be a high degree of sequence conservation (>80%) that exists between the KDM5 catalytic *Jumonji* domains (Table 4.1), we wanted to determine whether a similar homology would be observed in substrate specificity. Clustal Omega[77,163,164] was used on human KDM5A/B/C/D full length and catalytic *Jumonji* domain sequences obtained from UniProt[12] to effectively compare sequence homology in Table 4.1. Moreover, given that KDM5C and KDM5D possess a >98% sequence homology between their *Jumonji* domains, we hypothesized they would display the greatest homology in substrate specificity. To compare sequence vs substrate specificity matrices, we applied hierarchical clustering to compare substrate preference matrices for KDM5A/B/C/D[161] (Fig 4.3C). Our analysis demonstrated our hypothesis was correct and that the KDM5C and KDM5D substrate specificity matrices were the most similar among the KDM5 family. Furthermore, the two random matrices applied serving as our control matrices appear to be the most dissimilar when

compared to the KDM5A/B/C/D matrices. This suggests that significant clusters were formed throughout the KDM5 matrices that allowed for the algorithm to determine they were most similar. Ultimately, this demonstrates that there is some degree of homology in substrate specificity exhibited between the KDM5 family.

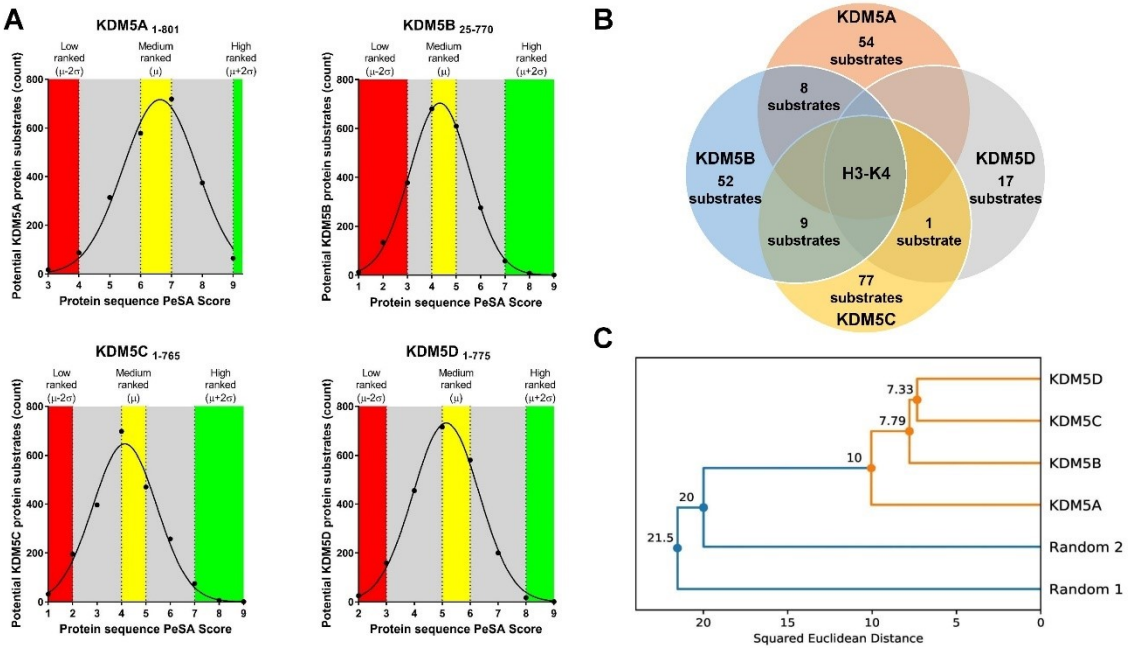



**Figure 4.3. Identifying high-ranking potential substrates for KDM5A/B/C/D.**

***A.*** Summary of KDM5A$_{1\text{-}801}$, KDM5B$_{25\text{-}770}$, KDM5C$_{1\text{-}765}$ and KDM5D$_{1\text{-}775}$ predicted substrates based on KDM5A/B/C/D substrate specificity matrices. Approximately 2100 predicted substrates were fit to a Gaussian distribution for further statistical analysis. ***B***. Number of high-ranking predicted substrates for the KDM5 demethylase family. ***C.*** Dendrogram representing of KDM5 substrate selection matrix homology. Analysis completed through Ward variance minimization algorithm[161].

**Table 4.1**. **KDM5 full sequence homology and Jumonji domain homology analysis**. *ClustalΩ multiple sequence alignment* indicates 564 identical positions for an overall 32% full sequence homology and 166 identical positions for an overall 80% Jumonji-domain homology between KDM5 demethylases.

| **Full sequence homology Jumonji domain homology** | **KDM5A** | **KDM5B** | **KDM5C** | **KDM5D** |
|---|---|---|---|---|
| **KDM5A** | | 48%<br>828 identical positions | 46%<br>813 identical positions | 45%<br>788 identical positions |
| **KDM5B** | 86%<br>179 identical positions | | 44%<br>726 identical positions | 43%<br>698 identical positions |
| **KDM5C** | 86%<br>180 identical positions | 84%<br>175 identical positions | | 84%<br>1321 identical positions |
| **KDM5D** | 86%<br>179 identical positions | 83%<br>173 identical positions | 99%<br>206 identical positions | |

*Assessing KDM5 family activity towards high ranked substrates to identify candidate substrates*

To further assess which of the potential high-ranking substrates are candidate substrates to further investigate in a cellular context, we then assessed KDM5 activity towards each of their respective high-ranking substrates, as well as compared the activity of 10 substrates from their middle and low ranked classifications to provide insight into the specificity of our predictions (Appendix D, Fig. D5). This was accomplished by synthesizing 11-mer tri-methyl-lysine containing peptides with sequences matching the predicted demethylation site (+/- 5 residues: 11 residue peptides). To visualize percent prediction success rate, we then plotted the amount of successful vs unsuccessful predictions for each of high, middle and low ranked substrate prediction sets for each of KDM5 enzyme (Fig 4.4A). We observed that KDM5A/B/C/D respectively displayed activity (defined as >50% H3K4me3 activity) towards 39%, 33%, 29% and 24%, respectively, of their high ranked substrates (Fig 4.4A). As seen in Fig 4.4A, our method accurately differentiated between substrate candidates ranked highly compared to those in the medium/ low scoring classifications for KDM5A/C/D. We observed that KDM5A/C/D displayed no activity towards all medium and low ranked peptides. However, there were 2 medium ranked substrates that demonstrated activity from our KDM5B predictions. Collectively, our method has provided 66 candidate peptide substrates for KDM5 demethylases *in vitro*. Individually, this corresponds to 21 candidate substrates for $KDM5A_{1-801}$, 17 for $KDM5B_{25-770}$, 24 for $KDM5C_{1-765}$, and 4 for $KDM5D_{1-775}$ (Fig 4.4B). Although there

is still further validation to be done in a cellular context to truly identify KDM5-driven demethylation of novel candidate substrates, this method does prioritize the number of potential substrates to refine and focus further validation efforts[95].

*Assessing KDM5 family activity towards candidate substrates in SKBR3 cells*

From our list of candidate substrates for the KDM5 family, we next sought to confirm cellular demethylation. Using our prioritized list of 66 candidate *in vitro* peptide substrates, we developed a targeted parallel reaction monitoring (PRM) mass spectrometry approach to confirm the activity of several KDM5 substrates in cells. The chemical inhibitor, CPI-455, was used to readily inhibit of the KDM5 family in a facile way[143]. Our aim was to determine whether we could observe changes in protein methylation as a response to CPI-455 treatment in SKBR3 cells, a cell line with a previously documented response to this drug (Fig 4.5A)[143]. Vinogradova et al., displayed an optimal CPI-455 concentration to use in SKBR3 cells was 12.5 μM and the $IC_{50}$ for KDM5A/B/C *in vitro* were respectfully ~10 ± 1 nM, ~3 ± 1 nM and ~15 ± 1 nM[143]. We were able to measure changes in protein methylation though PRM-MS analysis of total soluble protein lysate of SKBR3 cells subjected to proteolytic digestion with trypsin[71,100]. Due to several factors such as lysine-rich regions and distribution, limitations in trypsin coverage, peptide size/charge, etc., it was not possible to examine changes in all our candidate KDM5 substrates using this approach. This approach is further complicated as the inhibition of any KDM does not guarantee an opposing increase in the methylation on a specific substrate. Despite these limitations, we

were successfully able to identify a CPI-455 responsive increase in tri-, and di-methylation levels in 6 candidate KDM5 family substrates in the SKBR3 cells (Fig 4.5B and C; Appendix D, Fig. D6). Specifically, we observed an increase in tri-methylation levels in BICRA-K404me3 by 1.29-fold, FGD5-K1301me3 by 1.85-fold, and HECW2-K199me3 by 1.57-fold, as a response to KDM5 family inhibition. Further, we observed an increase in di-methylation levels in WNK4-K226me2 by 5.33-fold, H2AY-K18me2 by 1.41-fold and HSPA1L-K189me2 by 1.37-fold.

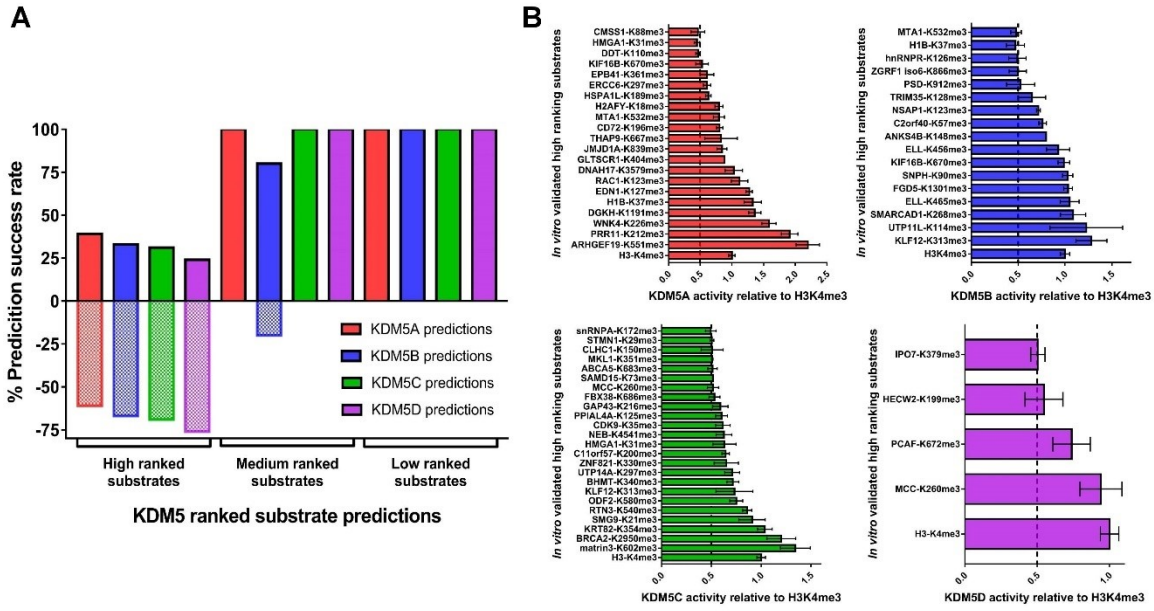



**Figure 4.4. Quantifying relative KDM5A/B/C/D activity towards high, medium and low ranked substrates.** ***A.*** Percent prediction success rates of $KDM5A_{1\text{-}801}$, $KDM5B_{25\text{-}770}$, $KDM5C_{1\text{-}765}$ and $KDM5D_{1\text{-}775}$ towards each of KDM5A/B/C/D high, medium and low ranked substrate peptides. ***B.*** Detection of $KDM5A_{1\text{-}801}$, $KDM5B_{25\text{-}770}$, $KDM5C_{1\text{-}765}$ and $KDM5D_{1\text{-}775}$ demethylation activity validated *in vitro* high ranked synthesized substrate peptides. Results are mean ± std. error (n=3).

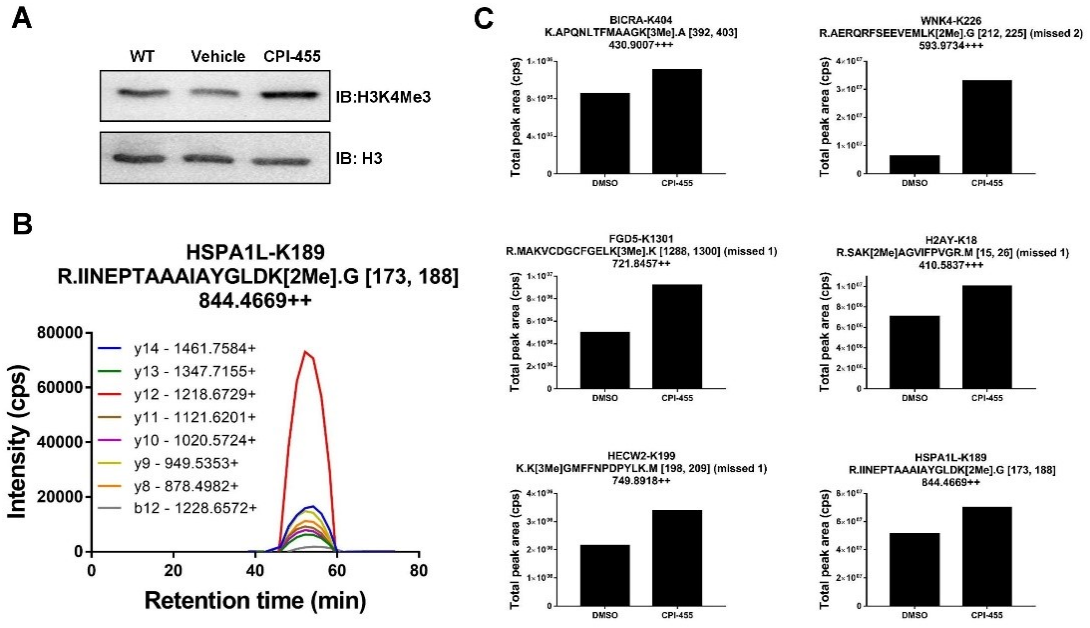



**Figure 4.5. Detecting changes in lysine methylation as a response to CPI-455 inhibition.** ***A.*** Validation of CPI-455 inhibition in SKBR3 cells through monitoring of H3K4me3 levels. ***B.*** Example of transitions observed to measure abundance of specific trimethyl lysine residue-containing peptide. ***C.*** Detection of changes in tri-methylation and di-methylation as a response to CPI-455 inhibitor treatment in BICRA-K404me3, FGD5-K1301me3, HECW2-K199me3, WNK4-K226me2, H2AY-K18me2 and HSPA1L-K189me2. Results are from combined biological replicates (n=2).

## Discussion

It has been reported that KDM5 substrate recognition is determined through the PHD1 and PHD3 (KDM5A/B) methyl binding domains, whereas the jumonji domain solely facilitates the catalytic demethylation of substrate(s)[43,48,64,65,74–76]. Although the presence of a methylated lysine residue is a key factor in PHD3 binding and subsequent JMJD domain-mediated demethylation, previous studies analyzing PHD1 binding have demonstrated that KDM5A is able to accommodate substrates with sequences that vary from histone H3K4[74]. This permissible flexibility in sequence surrounding the methylysyl residues justified the use of a systematic exploration of amino acid residues that either increase, maintain, or decrease binding and subsequent catalyzed demethylation reaction. Using this approach to studying substrate preference, we were able to generate recognition motifs for each KDM5 demethylase relative to the histone H3K4 sequence. Uniquely, our analysis of KDM5 substrate preference resulted in an enrichment of candidate KDM5 substrates from the thousands unique sites within the currently methyllysine proteome[130]; identifying a collective 66 *in vitro* substrates for the KDM5 family (Fig. 4.4B). Among these 66 high-ranking candidate substrates, we were able to observe changes in cellular methylation levels of 6 of KDM5 candidate substrates, because of chemical KDM5 inhibition (Fig. 4.5B-C). Ultimately, we've found that the primary sequence surrounding methylated lysyl residues can vary to maintain KDM5 activity *in vitro*. Moreover, similar analysis have demonstrated similar affects observed in other KDM families, such as KDM3A[165]. We have also

demonstrated the ability for KDMs to demethylate non-histone peptide substrates *in vitro*.

Recently there has been increasing evidence that KDM proteins might facilitate dynamic roles in the cell, existing well-beyond histone demethylation[67,70,166–171]. It is now apparent that both families of histone demethylase enzymes, flavin adenine dinucleotide (FAD)-dependent amine oxidases and the larger Jumonji C domain (JMJD) containing demethylases are capable of targeting non-histone proteins for demethylation[67,167–171]. As these KDM families are shown to not only demethylate non-histone proteins, but as well have biological functional consequences of these demethylation events. For example, the apparent functional redundancy of having 4 distinct KDM5 family members performing a catalytic reaction one shared substrate may be viewed as inefficient if all KDM5 substrates displayed the same level of conservation. As supported by our findings, we posit that it is likely that KDM5 family members have diverse cellular functions, beyond histone H3K4me3/2 demethylation[172]. Indeed, previous studies have already identified how KDM5 demethylases are implicated in a variety of biological functions. For example, although no direct substrates have been identified,  there is evidence that KDM5B associates with the NuRD complex which mediates transcriptional repression[65]. This is supported by our findings as we demonstrated KDM5B and KDM5A activity towards the MTA1-K532 peptide *in vitro*. Interestingly, dimethylation at K532 in MTA1 has been documented to encourage the association of MTA1 with NuRD co-repressor complex and demethylation of this site would have clear functional

implication[142,146]. Furthermore, it was shown that cytoplasm-localized KDM5C interacts with Smad3, a mediator of the TGFβ signaling pathway[137]. This cytoplasmic KDM5C-Smad3 interaction was dependent on catalytic activity, although no substrate had been identified[137]. These complementary studies demonstrate how critical the discovery of new substrates for the KDM5 family is to help further elucidate our understanding of this family.

To provide further insight into our prioritized list of high-ranking *in vitro* peptide substrates for the KDM5 family, we utilized a KDM5-family specific inhibitor (CPI-455) to help monitor cellular methylation events by targeted mass spectrometry. Previous studies that have looked to translate *in vitro* research to cellular validation efforts have made it apparent that the reaction parameters may be vastly different in a cellular environment[173]. Further, not all our prioritized substrates exist in protein regions amenable to trypsin digestion for mass spectrometry detection. Despite these logistical limitations, we successfully observed an increase in either the Kme2 or Kme3 status of 6 methyllysine sites following CPI-455 treatment (Fig. 4.5). Although we have identified several high-ranking candidate substrates through our pipeline, identification of demethylation events on these non-histone proteins via KDM5 demethylases remained to be confirmed through further cellular analysis. *In vivo* factors including spatial and/or temporal differences in cellular localization and expression might be driving factors in limiting these identified interactions in cells[173]. By monitoring specific protein lysyl-methylation in response to one specific KDM overexpression or knockdown, one would be able to identify whether a change in methylation was

seen as a result of KDM expression[67]. Ideally, changes in methylation could be observed through methyl specific antibodies, or mass spectrometry analysis[67,71]. The authors recognize that the cell line chosen to monitor changes in cellular methylation events is not optimal for KDM5D expression[174]. The change in methylation seen for a KDM5D predicted substrate might have been facilitated through inhibition of another KDM5 family member. We find this plausible as we demonstrated the possibility that some of the KDM5s might have shared substrates between them, as shown through our *in vitro* data.

Through measuring relative KDM5 activity towards a highly specialized PPS library, we have demonstrated the ability to determine KDM5 substrate specificity and preference. In our analysis of KDM5 substrate preference, we generated permissible recognition motifs that we used to predict and rank potential methylated lysine residues in which the KDM5 family may demethylate in the methylproteome. Moreover, we were able to demonstrate the overall observed homology between KDM5 substrate specificity attributed most likely to their high catalytic *Jumonji* domain sequence homology. In assessing KDM5 activity *in vitro* towards peptides reflective of high-ranked protein substrates, we were able to identify 66 candidate KDM5 substrates from the thousands unique sites within the currently methyllysine proteome[130]. We further validated the likelihood that 6 of these candidate KDM5 substrates might be targeted by KDM5 demethylases through examining changes in methylation via CPI-455-mediated inhibition of the KDM5 family. Although we have not definitively proven any of these high-ranking candidate KDM5 substrates are targeted by KDM5s in a

cellular environment, we have defined specific, potential substrates for which validation efforts should be focused on. Ultimately, we have provided a novel approach for systematic characterization of substrate preference for any JMJD-type KDM demethylase with at least one established substrate. Moreover, we've demonstrated the ability to use the substrate preference data to predict high-ranking, methyllysine-containing protein substrates in the methylproteome, for any JMJD-type KDM demethylase. Future efforts should be focused on demonstrating the ability for these KDM5 enzymes to actively target non-histone protein substrates for demethylation, in a cellular context. This type of investigation is critical as to have a firm understanding of how these enzymes are altered and dysregulated in disease provides scientists and researchers with the proper tools to develop therapeutics to target the KDM5 family and its specific interactions with substrates, apart from H3K4.

## Author Contributions

**Matthew Hoekstra**: Methodology, Data curation, Validation, Writing – original draft

**Nashira H. Ridgeway**: Data analysis, Writing – review and editing.

**Kyle Biggar**: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review and editing.

## Acknowledgements

We would like to acknowledge the SGC consortium for the generous gift of JARID1AA-c180 pFB-LIC-Bse, JARID1BA-c165 pFB-LIC-Bse, JARID1CA-c022 pFB-CT10HF-LIC and JARID1DA-c066 pFB-LIC-Bse.

## Funding source

This work was supported by a National Science and Engineering Research Council (NSERC) Canada Discovery grant to K.K.B (06151).

# Chapter 5
# Identification of novel KDM5A non-histone substrate: p53-K370me3

**Matthew Hoekstra[1,2], Nicolas Reynoird[3], Kyle Biggar[1,2]**

[1]Institute of Biochemistry, Carleton University, Ottawa, ON, K1S 5B6, Canada

[2]Department of Biology, Carleton University, Ottawa, ON, K1S 5B6, Canada

[3]Institute for Advanced Biosciences - Grenoble Alpes University, CNRS UMR5309, INSERM U1209 - Grenoble, FRANCE

## Chapter summary

In this Chapter, we demonstrated the applicability of our novel method and pipeline to identify and functionally validate a high-ranking KDM5A substrate. Once identified, we were able to validate the KDM5A-mediated demethylation of the high-ranking p53-K370me3 substrate in a cellular context using HCT 116 and HEK 293 cells. Using a combination of wild-type (WT) KDM5A and inactive, mutant (Mut) $KDM5A_{H483A}$ overexpression experiments, along with affinity purification tandem mass spectrometry (AP-MS/MS) analysis, we were able to analyze the KDM5A-mediated demethylation of p53-K370me3/2. Moreover, we were able to demonstrate how KDM5A expression has an influence on an established p53K370me2-53BP1 protein-protein interaction (PPI) and identify a novel TAF5-p53 PPI dominated through p53-K370me3. Ultimately, this chapter defines the novel KDM5A demethylation of p53-K370me3. This is the first evidence that any KDM5 family member can target a non-histone substrate for demethylation. Notably, it is also the first indication that KDM5 can functionally regulate non-histone proteins via direct lysine demethylation.

## Introduction

Previous studies of KDM-mediated demethylation of non-histone substrates have led us to investigate whether members of the KDM5 family of histone lysine demethylases are capable of this activity as well. As previously mentioned in Chapter 1, the first KDM identified to target a non-histone substrate for demethylation was LSD1/KDM1A[67]. In this study, it was demonstrated that KDM1A catalyzes the removal of both the mono- and di-methylation modifications on p53 at lysine 370 (i.e., p53-K370me1/2)[67]. In Chapter 4, we touch on the increasing evidence that not only are flavin adenine dinucleotide (FAD)-dependent amine oxidases capable of targeting non-histone substrates, but there is growing evidence that Jumonji C domain (JMJD) containing demethylases have activity that likely extends beyond histone demethylation[67,70,166–171]. For example, it was shown that JMJD-containing KDM3A targets the p53-K372me1 substrate for demethylation[68]. Insights such as these have made it critical to investigate the molecular targets of these KDM families. Identification of functional KDM targets can aid in uncovering the role(s) of these enzymes in cell biology and disease development.

As stated in Chapter 4, it's been shown recently that KDM5B associates with the NuRD complex to mediate transcriptional repression[65]. Moreover, KDM5C was shown to interact with cytoplasmic Smad3 and that this interaction is dependent on catalytic activity[137]. Although neither of these interactions demonstrate active KDM5 demethylation on a non-histone substrate, they do showcase that the KDM5 possess function(s) beyond solely H3K4me3/2

demethylation. Although KDM5 can influence specific gene transcription through changes in H3K4me3/2, there are still unknowns surrounding how KDM5 is able to regulate both cellular homeostasis and disease progression.

As previously mentioned, there have been several KDM enzymes that target p53 lysine residues, resulting in either the activation or inactivation of p53; this includes KDM1A and KDM3A[67,68]. In many cases, the dynamic change in lysine methylation status regulates the recruitment of methyl-binding domain (MBD) proteins that help to direct p53 function. There are also several KMT enzymes that target p53 for methylation; for example, SET8-mediated p53-K382 mono-methylation[175]. This modification results in the suppression of p53 transcriptional activity through the recruitment of the MBD protein, L3MBTL1[176]. Generally, methylation sites are clustered within the C-terminal regulatory domain (REG) of p53 and results in variations on p53 acitvity[177]. Collectively, these studies have demonstrated how a single change in a p53 lysine methylation site and state can dramatically alter activity through changes in the recruitment of protein interactions. This has been observed with the KDM1A-mediated demethylation of the p53-K370me2 site; as p53-K370me2 promotes a unique function which is not observed in the p53-K370me1 state[67]. To elaborate, when p53-K370me2 is demethylated to K370me1, p53 no longer interacts with the transcriptional co-activator 53BP1; an interaction mediated through the 53BP1 tandem Tudor MBDs[67]. When this occurs, 53BP1 no longer stabilizes p53 for stimulated genome-wide p53-dependent gene transactivation[67]. As well p53-

mediated repression events in response to genotoxic stress and ionization radiation are no longer triggered[67,178].

Herein, we look to examine the possibility of KDM5A-mediated demethylation of the p53-K370me3 substrate. Specifically, we look to validate the demethylation event *in vitro* and subsequently investigate in a cellular context in HCT 116 cells, given their WT p53 status , known p53-K370me1/2 methylation, and history in KDM5 research[104,179]. Through the use of methyl specific p53-K370me1 and p53-K370me2 antibodies, and targeted PRM-MS/MS analysis to study p53-K370me3 dynamics, we are confident that we can determine whether KDM5A activity influences the p53-K370 degree of methylation[67,180]. Moreover, given the documented methylation dependent 53BP1 interaction with p53-K370me2 (but not K370me3), we seek to determine whether KDM5A expression has an influence on this PPI. Finally, as 53BP1 specifically interacts with p53 through its K370me2 state, we wish to investigate the likelihood of a novel PPI that occurs with p53 through its K370me3 state.

## Materials and Methods

*Recombinant KDM5A$_{1\text{-}588\Delta AP}$ expression and purification*

pET28 plasmid containing KDM5A$_{1\text{-}588\Delta AP}$ construct (obtained from Xiaodong Cheng of Emory University, Atlanta, Georgia) was transformed in *Escherichia coli* strain BL21 (DE3)C+ for protein expression and grown in LB media. Cultures for protein expression were inoculated with 0.5-1% starter culture and grown at 37$^{o}$C to an $A_{600}$ of 0.5-0.6 upon which the temperature was

decreased to 16°C and expression was induced with 0.2 mM isopropyl ß-D-1-thiogalacttopyranoside (IPTG). Following IPTG induction, cultures were incubated overnight at 16°C. The next day, cells were pelleted by centrifugation (5000 x g for 15 min at 4°C), media decanted, pellets snap frozen in liquid nitrogen and then stored at -80°C.

For protein purification, pellets were resuspended in 25 mM HEPES, pH 8.0, 10% glycerol, 400 mM NaCl, 1 mM DTT, 10 µg/mL DNase and supplemented with protease and phosphatase inhibitors (1 mM PMSF, 0.5 µM Aprotinin, 10 µM E-64, 1 µM Pepstatin). Cells were lysed through 15 passes in a Dounce homogenizer on ice. Cells were subsequently sonicated 3x for 20 sec at 70% amplitude intensity on ice. The cell lysate was cleared via centrifugation (18000 x g for 45 min at 4°C) and the soluble lysate was loaded onto a 500 µL HisPur™ Ni-NTA Resin (ThermoFisher Scientific, Cat# 88221) affinity column. The affinity column was washed and eluted with an imidazole gradient (40-250 mM imidazole in 1X PBS) and protein fractions were collected. Protein fractions were dialyzed overnight in a protein storage buffer (25 mM HEPES, 150 mM NaCl, 10% glycerol and 1 mM DTT) using 3.5kDa molecular weight cut-off dialysis tubing (SnakeSkin® Dialysis tubing, Thermo Scientific). Protein purity was assessed through 10 % SDS-PAGE stained with Coomasie Brilliant Blue G-250 stain. Protein concentration was quantified through a standard Bradford assay[138]. Proteins were snap-frozen in liquid nitrogen and stored at -80°C for later use.

*Recombinant KDM5A$_{1-801}$ expression and purification*

Please refer to Chapter 4

*Peptide Synthesis (check KDM5 paper)*

Please refer to Chapter 4

*Permutated peptide substrate (PPS) library and ranked peptide design*

Please refer to Chapter 4

*KDM demethylation assay*

Each of KDM5A$_{1-588\Delta AP}$ and KDM5A$_{1-801}$ (protein used in Chapter 4) *in vitro* demethylation assays, as well as permutated peptide substrate (PPS) library experiments, were performed as described. All reactions took place in demethylase reaction buffer (25 mM HEPES pH 7.5, 10 μM peptide substrate, 100 μM ascorbate (J.T Baker Chemical Co), 10 μM Fe(II) sulfate (BDH chemicals), 10 μM α-ketoglutarate (Sigma Aldrich), 1% v/v DMSO) for 1 hour at 23$^{o}$C in a 384-well Corning flat-bottom microplate. Following incubation, succinate detection reagents were added (Succinate Glo™ JMJC Demethylase Assay; Promega Corp, Cat# V7991) according to manufacturer's instruction. Briefly, 5 μL of Succinate detection reagent I was added and incubated for 60 min followed by 10 μL of Succinate detection reagent II. After 10 minutes, luminescence was read using a Cytation 5 plate reader (Bio-Tek, Winooski, VT) with a plate height of 1.25 mm and a gain set to 170.

The PPS library experiments were carried essentially the same as described above. Briefly, 2.5 µL of 2X reaction buffer (without peptide substrate) was aliquoted to a 384-well plate at 4ºC. While at 4ºC, 1.25 µL of 40 µM permutated H3K4me3 peptide were added to separate wells. Temperature was increased to 23ºC and 1.25 µL of 4X KDM5 enzyme (1.8µM for $KDM5A_{1\text{-}588\Delta AP}$ and 100 nM for $KDM5A_{1\text{-}801}$) in 25 mM HEPES (pH 7.5) was added. The reactions proceeded for 1 hour at 25ºC, followed by the addition of the succinate detection reagents, quenching the reaction, and reading luminescence as described above. $KDM5A_{1\text{-}588\Delta AP}$ and $KDM5A_{1\text{-}801}$ activity towards ranked protein peptide substrate or p53-K370me3/2/1/0 peptide substrate experiments were carried out in the same manner.

*Cloning and site-directed mutagenesis*

KDM5A open reading frame (ORF) were amplified from cDNA derived from HEK 293 cells via Phusion DNA polymerase (ThermoFisher Scientific, Cat# F530S) with primers that generated PCR products with KpnI and XbaI restriction sites flanking the ORF. The PCR product was digested with the specified restriction enzymes (New England Biolabs; Cat# R3142S and Cat# R0145S, respectively) and cloned into a pcDNA3.1-HA vector (Addgene, Plasmid #128034) (Appendix F, Fig. F1A). Site directed mutagenesis was carried out on the pcDNA3.1/HA-KDM5A plasmid to form and pcDNA3.1/HA-$KDM5A_{H483A}$ plasmid using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs; Cat# E0554S) according to manufacturer's instructions. Confirmation of

identification of constructs was verified by DNA sequencing (Appendix F, Fig. F1B-C). The following primers were used in the Site directed mutagenesis amplification: KDM5A$_{H483A}$ Forward (TTTTTGCTGGgccATTGAGGATCAC), KDM5A$_{H483A}$ Reverse (GAAGAGAAGCACATTCCC). The following primers were used in KDM5A amplification from cDNA: KMD5A Forward with KpnI restriction site (ATTGGTACCCATGGCGGGCGTGGGGC), KDM5A Reverse with XbaI restriction site (ATTTCTAGACTAATATTTTTTATTCTTCTCTTCTTCCGTTACAAAT)

*Cell culture*

HEK 293 and HCT 116 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 4.5g/L glucose, L-glutamine and sodium pyruvate (Wisent Inc, Cat# 319-005-CL) plus 10% fetal bovine serum (ThermoFisher Scientific, Cat# 12484028) and 1% penicillin-streptomycin (ThermoFisher Scientific, Cat# 10378016) and cultured in a humidified incubator at 37$^{o}$C and 5% $CO_2$.

*WT KDM5A and Mut KDM5A$_{H483A}$ overexpression*

HCT 116 and HEK 293 cells were seeded in a 10 cm$^2$ plate (2.2 x 10$^6$ cells/ plate) and adhered overnight. Fresh DMEM+/+ was added day of and cells were transfected with pcDNA3.1/HA vector only, pcDNA3.1/HA-KDM5A, and pcDNA3.1/HA-KDM5A$_{H483A}$ (Addgene, Plasmid # 14800) using jetOPTIMUS® reagent (Polyplus, Cat# 101000006) in a 1:1.5 (v/v) ratio, under manufacturers

recommendation. Media was changed 4 hours post-transfection, and 24 hours post-transfection. Cells were left to grow for 48 hours at 37°C. 48 hours post-transfection, cells were trypsinized, collected and washed with ice cold sterile 1X PBS and briefly centrifuged (0.2 x g for 5 min). Cell pellets were resuspended in lysis buffer containing 50 mM-Tris-base solution (pH 7.5), 150 mM NaCl, 10% glycerol, 0.5% Triton X-100, 1mM DTT and protease inhibitors (1 mM PMSF, 0.5 µM Aprotinin, 10 µM E-64, 1 µM Pepstatin). Cells were lysed at 4°C for 60 while rotating min and collected by centrifugation (18000 x g for 30 min at 4°C). Protein supernatant was collected and quantified through Bradford assay[138], followed by further western blot analysis and/or immunoprecipitation.

*Histone isolation*

Please refer to Chapter 4

*Reverse Transcription (RT) and polymerase chain reaction (RT-qPCR)*

WT KDM5A and Mut KDM5A$_{H483A}$ overexpression in HCT 116 and HEK 293 cells was performed, and cells were harvested as described above. After ice cold sterile 1X PBS wash and cell collection, RNA was isolated from cell pellets using QIAGEN RNeasy Mini Kit (Qiagen, Cat# 74104). RNA was quantified using a BIO-TEK POWERWAVE XS spectrophotometer with a Take3 micro-volume plate (BIO-TEK). Equal 1 µg amounts of RNA from each treated sample replicate were converted to cDNA using the Bio-Rad iScript™ reverse transcription kit (Bio-Rad, Cat# 1708840). Following reverse transcription, cDNA was diluted 10X

with RNAse free $dH_20$ and qRT-PCR was completed using SsoAdvanced universal SYBR green supermix (Bio-Rad, Cat# 1725274) with the following conditions: initial denaturation (1 min @ 95 $^oC$), PCR reaction (5 sec @ 95$^oC$, 20 sec @ 60$^oC$, repeat for 40 cycles), followed by a melt curve (65$^oC$- 95$^oC$, with degree increments every 5 sec). For RT-qPCR, the following primers were used: p53 Forward (CAGCACATGACGGAGGTTGT)[181], p53 Reverse (TCATCCAAATACTCCACACGC)[181], p21 Forward (GGAGACTCTCAGGGTCGAAA), p21 Reverse (GCTTCCTCTTGGAGAAGATCAG), and GAPDH Forward (GGAGCGAGATCCCTCCAAAAT), GAPDH Reverse (GGCTGTTGTCATACTTCTCATGG). These primers were synthesized by Integrated DNA Technologies (Coralville, IA) and were a gracious gift from the McKay lab (Carleton University). P53 primers were designed from Guo et al., 2021[181].

*Immunoprecipitation of p53*

Total soluble protein from HCT 116 cells with (and without) overexpression of either WT KDM5A and Mut $KDM5A_{H483A}$ was isolated and quantified as described above. After protein quantification, 1 mg of protein lysate in 500 µL p53-IP buffer (50 mM-Tris-base solution (pH 7.5), 150 mM NaCl, 10% glycerol, 0.05% Triton X-100, 1mM PMSF) was incubated with 2 µg of mouse monoclonal anti-p53-$IgG_{2a}$ antibody (DO-1, Santa Cruz Biotechnology, Cat# sc-126) with rotation at 4$^oC$ overnight. 30 µL of protein A-agarose beads (BioBasic) were added to each

sample for 4 hours at 4°C under rotation. After, the beads were washed 3 times in 300 µL p53-IP wash buffer (50 mM-Tris-base solution (pH 7.5), 150 mM NaCl, 10% glycerol, 0.01% Triton X-100, 1mM PMSF). Bound proteins were eluted using 30 µL of 2X Laemmli sample buffer (100 mM Tris-HCl pH 6.8, 4% SDS, 20% Glycerol, 200 mM β-mercaptoethanol, 0.2% Bromophenol blue) and heated for 5 min at 90-100°C. Samples were then subjected to subsequent western blot analysis, described below.

*Western blot analysis*

Proteins were separated on a standard 8%, 10% or 15% tris-glycine SDS-PAGE gel and transferred to 0.2 µm PVDF membrane (Bio-Rad, Cat# 1620177) for western blotting. Membranes were blocked in 5% non-fat skim milk in TBST (1X TBS pH7.6 with 0.1% Tween-20) at room temperature for 1 hour. Membranes were then incubated overnight at 4°C with primary antibodies: anti-KDM5A antibody (1:800; Santa Cruz Biotechnology, Cat# sc-365993), anti-H3K4me3 primary antibody (1:2000; ThermoFisher Scientific, Cat# 703954), anti-H3 primary antibody (1:2000; ThermoFisher Scientific, Cat# 702023), anti-p53 antibody (1:2000; Santa Cruz Biotechnology, Cat# sc-126), anti-p53-K370me1 primary antibody (1:1000; St Johns Laboratory, Cat# STJ97495), anti-p53-K370me2 primary antibody (1:2000; St Johns Laboratory, Cat# STJ90115), anti-p21 primary antibody (1:1000; CalBiochem, Cat# OP64), anti-Vinculin primary antibody (1:2000; Santa Cruz Biotechnology, Cat# sc-25336), anti-53BP1 primary antibody (1:1000; Santa Cruz Biotechnology, Cat#: sc-515841) or anti-TAF5 primary

antibody (1:1000; ThermoFisher Scientific, Cat# A303-686A). The next day, membranes were washed 3 x 5 min in TBST at room temperature. All membranes containing only total cell lysates were incubated with either of rabbit anti-mouse HRP-conjugated secondary antibody (1:10,000; ThermoFisher Scientific, Cat# 31450) or goat anti-rabbit HRP-conjugated secondary antibody (1:10,000; ThermoFisher Scientific, Cat# G21234) for 1 hour at room temperature. All membranes containing p53 immunoprecipitated samples, 53BP1-p53 co-immunoprecipitated samples (methods below), or TAF5-p53 co-immunoprecipitated samples (methods below) were incubated with either or of goat anti-mouse IgG light chain specific HRP-conjugated secondary antibody (1:8,000; Jackson ImmunoResearch, Cat# 115-035-174) or mouse anti-rabbit IgG light chain specific HRP-conjugated secondary antibody (1:8,000; Jackson ImmunoResearch, Cat# 211-032-171) for 1 hour at room temperature. Bands were visualized via chemiluminescence using Clarity™ Western ECL Substrate (Bio-Rad, Cat# 170-5060) with a Gel Doc™ XRS+ Imaging system (Bio-Rad).

*Mass spectrometry analysis.*

Overexpression of WT KDM5A and Mut KDM5A$_{H483A}$ in HCT 116 and HEK 293 cells was performed, and cells were harvested, lysed and total protein lysate was quantified as described above. After protein quantification, 1 mg of protein lysate in 500 µL p53-IP buffer (50 mM-Tris-base solution (pH 7.5), 150 mM NaCl, 10% glycerol, 0.05% Triton X-100, 1mM PMSF) was incubated with 2 µg of mouse monoclonal anti-p53-IgG$_{2a}$ antibody (DO-1, Santa Cruz Biotechnology, Cat# sc-

126) with rotation at 4°C overnight. 30 µL of protein A-agarose beads (BioBasic) were added into each reaction for 4 hours at 4°C. Beads were then washed in 200 µL trypsin digestion buffer (50mM $NH_4HCO_3$ pH 7.8). The protein-antibody-bead complex was then resuspended with 100µL 50 mM $(NH_4)HCO_3$ and 20 ng/µL trypsin (Promega, Cat# V5111) and incubated overnight while rotating at 37°C. Following digestion, beads were pelleted and soluble peptides were C18 cleaned using zip-tip C18 cleaning tips (EMD Millipore, Cat# ZTC18S096). Samples were centrifuged in a roto-vac to evaporate solvent and peptides were resuspended in 20 µL of 0.1% formic-acid $H_2O$.

Digests were analyzed by positive ESI LC-MS/MS on a Thermo Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer at the John Holmes Mass Spectrometry Center (University of Ottawa, ON). Briefly, peptide digests were loaded onto a Thermo Easy-Spray analytical column with an Easy-nLC 1000 chromatography pump coupled to a Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer. Peptides were separated on a 125-min (5–40% acetonitrile) gradient. The spectrometer was set in full MS/data-dependent-MS2 TopN mode: mass analyzer over a mass-to-charge ratio ($m/z$) range of 400–1,600 with a mass resolution of 70,000 (at $m/z = 200$), 35 NCE (normalized collision energy), 2.0 $m/z$ isolation window, and 15-s dynamic exclusion. The isolation list with the mass [$m/z$] and the sequence p53 peptides was generated using Skyline software. Each trace on the chromatograph represents the detection of each individual transition ion used to monitor p53-K370me3 site-specific methylation. Relative changes in p53-K370me3 methylation (EPGGSRAHSSHLK(me3)) were

determined by the total transition peak area. Two internal p53 peptides were used to normalize p53-K370me3 data for each sample. Positive identification required the successful detection of at least three transition ions.

*Doxorubicin treatment*

HCT 116 cells were seeded in a 10 $cm^2$ plate (2.2 x $10^6$ cells/ plate) and adhered overnight. Cells were first treated with 0.5-10 μM doxorubicin and 0 μM (DMSO vehicle control) for 24 hours. After 24 hours, cells were trypsinized, collected and washed with ice cold sterile 1X PBS and centrifuged (0.2 x g for 5 min at 4$^o$C). Cell pellets were resuspended in lysis buffer containing 50 mM-Tris-base solution (pH 8.0), 150 mM NaCl, 10% glycerol, 0.5% Triton X-100, 1 mM DTT and protease inhibitors (1 mM PMSF, 0.5 μM Aprotinin, 10 μM E-64, 1 μM Pepstatin). Cells were lysed at 4$^o$C for 60 min while rotating and collected by centrifugation (18000 x g for 30 min at 4$^o$C). Protein supernatant was collected and quantified through Bradford assay[138]. Total protein lysates were diluted to the same concentration and had equal volume mixed with 2X Laemmli sample buffer (100 mM Tris-HCl pH 6.8, 4% SDS, 20% Glycerol, 200 mM β-mercaptoethanol, 0.2% Bromophenol blue) and heated for 5 min at 90-100$^o$C. Samples were then subjected to subsequent western blot analysis.

To determine the response of p53-K370me following doxorubicin treatment in KDM5A OE cells, HCT 116 cells were transfected with and without pcDNA3.1/HA vector only, pcDNA3.1/HA-KDM5A, and pcDNA3.1/HA-KDM5A$_{H483A}$ (Addgene, Plasmid # 14800) using jetOPTIMUS® reagent (Polyplus,

Cat# 101000006) in a 1:1.5 (v/v) ratio. Media was changed 4 hours post-transfection. After 24 hours post-transfection, the media was changed, and cells were treated with 1 µM doxorubicin (concentration determined above). Cells were left to grow for another 24 hours at 37°C. 48 hours post-transfection, cells were trypsinized, collected and washed with ice cold sterile 1X PBS and centrifuged (0.2 x g for 5 min at 4°C). Cell pellets were resuspended in lysis buffer containing 50 mM-Tris-base solution (pH 7.5), 150 mM NaCl, 10% glycerol, 0.5% Triton X-100, 1mM DTT, protease inhibitors (1 mM PMSF, 0.5 µM Aprotinin, 10 µM E-64, 1 µM Pepstatin). Cells were then lysed at 4°C for 60 while rotating min and collected by centrifugation (18000 x g for 30 min at 4°C). Soluble protein was collected from the supernatant and quantified through Bradford assay[138].

For p53-53BP1 co-immunopreciptation, 1 mg of protein lysate in 500 µL p53-coIP buffer (50 mM-Tris-base solution (pH 7.5), 150 mM NaCl, 10% glycerol, 0.05% Triton X-100, 1mM PMSF) was incubated with 2 µg of mouse monoclonal anti-p53-$IgG_{2a}$ antibody (DO-1, Santa Cruz Biotechnology, Cat# sc-126) with rotation at 4°C overnight. 30 µL of protein A-agarose beads (BioBasic, Cat# msa015005) were added to each sample for 4 hours at 4°C. After, the beads were washed 2 times in 300 µL p53-coIP wash buffer (50 mM-Tris-base solution (pH 7.5), 150 mM NaCl, 10% glycerol, 0.005% Triton X-100, 1mM PMSF). Bound proteins were eluted using 30 µL of 2X Laemmli sample buffer (100 mM Tris-HCl pH 6.8, 4% SDS, 20% Glycerol, 200 mM β-mercaptoethanol, 0.2% Bromophenol blue) and heated for 5 min at 90-100°C.

*Nutlin 3A treatment*

HCT 116 cells were seeded in a 10 $cm^2$ plate (2.2 x $10^6$ cells/ plate) and adhered overnight. Cells were treated without and with 1-20 μM Nutlin 3A and 0 μM (DMSO vehicle control) for 24 hours. After 24 hours, cells were trypsinized, collected and washed with ice cold sterile 1X PBS and centrifuged (0.2 x g for 5 min at 4$^o$C). Cell pellets were resuspended in lysis buffer containing 50 mM-Tris-base solution (pH 7.6), 150 mM NaCl, 10% glycerol, 0.5% NP-40, 1mM DTT, and protease inhibitors (1 mM PMSF, 0.5 μM Aprotinin, 10 μM E-64, 1 μM Pepstatin). Cells were lysed at 4$^o$C for 60 min while rotating and collected by centrifugation (18000 x g for 30 min at 4$^o$C). Protein supernatant was collected and quantified through Bradford assay[138]. Total protein lysates were diluted to the same concentration and had equal volume mixed with 2X Laemmli sample buffer (100 mM Tris-HCl pH 6.8, 4% SDS, 20% Glycerol, 200 mM β-mercaptoethanol, 0.2% Bromophenol blue) and heated for 5 min at 90-100$^o$C. Samples were then subjected to subsequent western blot analysis, described below (Appendix G, Fig. G1).

For treatment of HCT 116 cells, an optimized treatment condition of 10μM Nutlin 3A and 1% DMSO (vehicle control) for 24 hours was used. Briefly, fresh DMEM+/+ was added day of and cells were transfected with and without pcDNA3.1/HA vector only, pcDNA3.1/HA-KDM5A, and pcDNA3.1/HA-$KDM5A_{H483A}$ (Addgene, Plasmid # 14800) using jetOPTIMUS® reagent (Polyplus, Cat# 101000006) in a 1:1.5 (v/v) ratio. Media was changed 4 hours post-transfection. After 24 hours post-transfection, the media was changed

again, and cells were treated with 10 μM Nutlin 3A. Cells were left to grow for another 24 hours at 37ºC. 48 hours post-transfection, cells were trypsinized, collected and washed with ice cold sterile 1X PBS and centrifuged (0.2 x g for 5 min at 4ºC). Cell pellets were resuspended in lysis buffer containing 50 mM-Tris-base solution (pH 7.6), 150 mM NaCl, 10% glycerol, 0.1% NP-40, 1mM DTT, and protease inhibitors (1 mM PMSF, 0.5 μM Aprotinin, 10 μM E-64, 1 μM Pepstatin). Cells were lysed at 4ºC for 60 while rotating min and collected by centrifugation (18000 x g for 30 min at 4ºC). Protein supernatant was collected and quantified through Bradford assay[138]

For p53-TAF5 co-immunopreciptation, after protein quantification, 1 mg of protein lysate in 500 μL p53-coIP buffer (50 mM-Tris-base solution (pH 7.6), 150 mM NaCl, 10% glycerol, 0.05% NP-40, 1mM PMSF) was incubated with 1 μg of mouse monoclonal anti-p53-$IgG_{2a}$ antibody (DO-1, Santa Cruz Biotechnology, Cat# sc-126) with rotation at 4ºC overnight. 30 μL of protein A-agarose beads (BioBasic, Cat# msa015005) were added to each sample for 4 hours at 4ºC. After, the beads were washed 2 times in 300 μL p53-coIP wash buffer (50 mM-Tris-base solution (pH 7.6), 150 mM NaCl, 10% glycerol, 0.005% NP-40, 1mM PMSF). Bound proteins were eluted using 40 μL of 2X Laemmli sample buffer (100 mM Tris-HCl pH 6.8, 4% SDS, 20% Glycerol, 200 mM β-mercaptoethanol, 0.2% Bromophenol blue) and heated for 5 min at 90-100ºC.

*P53-K370me AP-MS/MS*

HeLa cells were harvested with ice cold sterile 1X PBS and pelleted by centrifugation. Cell pellets were resuspended in 5x of pellet cell volume in lysis buffer containing 50 mM-Tris-base solution (pH 8.0), 150 mM NaCl, 1% NP-40, PMSF and antiproteases (Roche). Cells were then lysed at 4ºC for 30 under rotation and subjected to sonication at high intensity for 5 min, followed by resting on ice for 10 min. Cells were then collected by centrifugation (21,000 x g for 30 min at 4ºC) and the supernatant was collected and filtered using 0.2 µM filters.

To perform p53-K370me3/2/1/0 peptide pull-down, 10 µl of streptavidin sepharose beads (GE Healthcare) were saturated with 7.5 µg of either p53-K370me3/2/1 or Kme0 biotinylated peptides in peptide binding buffer (50 mM Tris pH 8, 150 mM NaCl, 0.5 % NP40, 0.5 mM DTT, 10 % glycerol, complete protease inhibitors (Roche)) for 2 h at 4 °C with rotation. Next, beads were washed in the peptide binding buffer and incubated with 1 mg of whole cell soluble protein extract for 4 h at 4ºC with rotation. Beads were then washed 3 times in peptide binding buffer. For mass spectrometry analyses, proteins still bound to beads were denatured and disulfide bonds reduced in digestion buffer (100 mM Tris (pH 8.0), 2 M urea and 10 mM DTT), after which cysteines were alkylated using 50 mM iodoacetamide. Then proteins were digested from the beads overnight using trypsin. Each digested sample (i.e., p53-K370me3/2/1 or Kme0 peptides interactors) was loaded on StageTip for purification, followed by differential isotope labeling with either light ($CH_2O$) or heavy ($CD_2O$) formaldehyde in the presence of sodium cyanoborohydride. Each pair of

corresponding peptide pull-downs (Forward= p53-K370me0/$CH_2O$ vs p53-K370me3/$CD_2O$; Reverse = p53-K370me3/$CD_2O$ vs p53-K370me0/$CH_2O$) were then pooled and analyzed using reverse phase Easy-nLC 1000 coupled online to a Thermo Fisher Orbitrap Fusion Tribrid mass spectrometer using a 140 minute gradient of buffer B (80% Acetonitrile, 0.1% TFA). Raw data were analyzed using MaxQuant (RRID:SCR_014485) software to quantify the ratio of each potential binder to the p53-K370me0 and p53-K370me3 or Kme2 or Kme1 peptides and further filtered for contaminants and reverse hits using Perseus software. Proteins identified as outlier in both experiments are assigned as significant interactors and written either in blue (methyl interactors) or red (methyl-repelled interactors).

*Data analysis*

With the relative demethylation activity derived from our mutated H3K4 PPS library, we used Peptide Specificity Analyst (PeSA) software to determine a candidate substrate recognition motif for *in vitro* substrates (i.e., a specific sequence of amino acids that required to be a KDM5A substrate)[125]. These weight-based motifs are constructed using the activity data and a user defined threshold that is set relative to the positive control[125]. The threshold reflects the minimum activity required for any mutated peptide substrate to have KDM5A activity exhibited towards it. For example, if the user would like to include any amino acid mutations that retained 50% of the WT H3K4me3 demethylation activity, then they could produce a 50% activity threshold recognition motif.

## Results

*Application of a PPS library to determine substrate preference of KDM5A$_{1\text{-}588\Delta AP}$*

As addressed in previous chapters, we have demonstrated the applicability of the commercially available *Promega Succinate-Glo*™ assay to monitor succinate production as a measure of KDM activity[92,93,123,139]. Here, we applied our PPS library to study KDM5A$_{1\text{-}588\Delta AP}$ substrate specificity, as well as predict potential substrates in which KDM5A$_{1\text{-}588\Delta AP}$ may demethylate *in vitro*. To begin, we successfully purified KDM5A$_{1\text{-}588\Delta AP}$ from *Escherichia coli* (Fig. 5.1A). We then validated its activity towards the WT H3K4me3 peptide substrate and identified an optimal concentration (1.8 μM) of KDM5A$_{1\text{-}588\Delta AP}$ to be used in our demethylase reaction conditions (Fig. 5.1B). We then validated the methyl-state specificity of KDM5A$_{1\text{-}588\Delta AP}$ for H3K4me0/1/2/3 peptides (Fig. 5.1C). As expected, KDM5A$_{1\text{-}588\Delta AP}$ demethylation of H3K4me3 was the most active, whereas the H3K4me2 substrate demonstrated 40% of the relative H3K4me3 activity ($p < 0.05$). The H3K4me1/0 substrates did not produce a signal significantly above background level. Furthermore, no measurable activity was found for either the no peptide or no enzyme negative controls. From this data, we were confident the KDM5A$_{1\text{-}588\Delta AP}$ enzyme concentration, reaction parameters and succinate-based detection assay were optimized to study the KDM5A$_{1\text{-}588\Delta AP}$ substrate preference through the PPS library.

As done previously, given the established H3K4me3/2 protein substrate for the KDM5 family, we applied our novel permutated peptide substrate library to assess KDM5A$_{1\text{-}588\Delta AP}$ substrate preference(s). Briefly, our novel PPS library was

based on the systematic mutation of the H3K4me3 substrate sequence surrounding the lysine 4 methylation site. This resulted in a library of 180 unique peptides, whereby each peptide contains a single amino acid mutation to one of the remaining 19 amino acids, from the -3 position to +5 position around the tri-methylated K4 position (i.e., ***ART***Kme3***QTARK***STGGKA; K4 position underlined, permutated region in italics/bold). KDM5A$_{1\text{-}588\Delta AP}$ activity towards each permutated peptide was then assessed by demethylation assay to determine substrate preference (Fig. 5.1D). KDM5A$_{1\text{-}588\Delta AP}$ activity was monitored relative to the wild-type (WT) H3K4me3 (positive control, set to 1) and the H3K4me0 (negative control, set to 0) peptides and no enzyme/peptide controls were included in all assays. We then determined candidate substrate recognition motifs reflective of 25%, 50%, 75% and 100% relative H3K4me3 activity thresholds using Peptide Specificity Analyst (PeSA) software[125] (Appendix E, Fig. E1A). With the goal to accurately predict substrates while maintaining some flexibility on amino acid substitutions, we decided to choose a relative H3K4me3 activity threshold of 75% for our substrate predictions (Fig 5.1E).

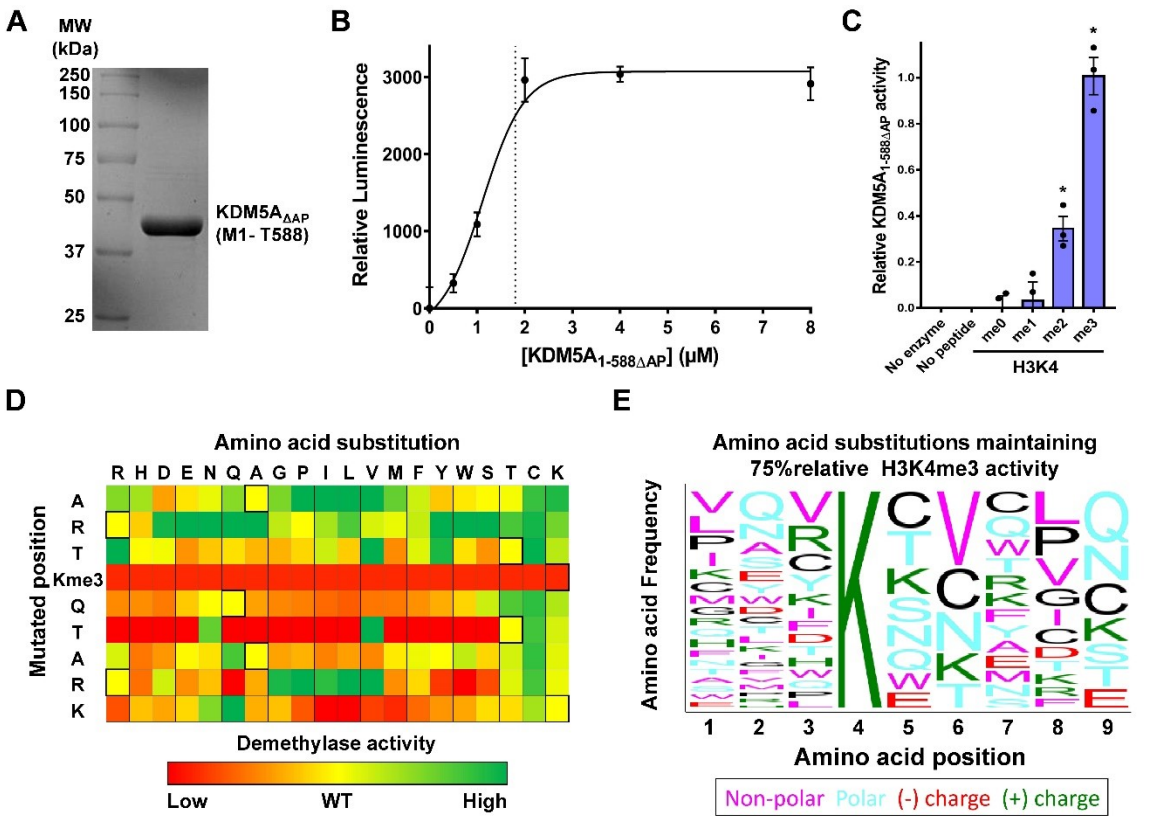



**Figure 5.1. Determining KDM5A$_{1\text{-}588\Delta AP}$ specificity by permutated peptide substrate library.** ***A.*** SDS-PAGE image showing purified KDM5A$_{1\text{-}588\Delta AP}$ used in this study. ***B.*** Luminescence-based detection of KDM5A$_{1\text{-}588\Delta AP}$ demethylation of WT H3K4me3 peptide. The dashed line indicates the optimal [KDM5A$_{1\text{-}588\Delta AP}$] concentration to use in subsequent *in vitro* experiments which was (1.8 μM). Luminescence averages were normalized by background signal. Results are mean ± std. error (n=3). ***C.*** Luminescence detection of KDM5A$_{1\text{-}588\Delta AP}$ demethylation of H3K4me0, H3K4me1, H3K4me2, and H3K4me3 peptides. Luminescence averages were normalized by background signal, and relatively quantified to H3K4me3 peptide luminescence (set to 1). Results are mean ± std. error (n=3). * indicates $p < 0.05$ from me0 data. ***D.*** Detection of KDM5A$_{1\text{-}588\Delta AP}$

demethylation activity with H3K4me3 peptide substrates that contain single amino acid mutation. Data was normalized to the H3K4me3 peptide at each mutation position (set to 1 for each relative position). Results are mean of (n=3-4). Digitized depiction of permutation array shows mean demethylation activity. ***E.*** $KDM5A_{1\text{-}588\Delta AP}$ recognition motif derived from relative $KDM5A_{1\text{-}588\Delta AP}$ demethylation activity towards permutated peptide substrate library. Motif is reflective of amino acid substitutions that maintained 75% of $KDM5A_{1\text{-}588\Delta AP}$ activity relative to H3K4me3. Motif image created using PeSA software[125]. Figure caption continued from Figure 5.1 (page 159).

*Identification of potential novel p53-K370me3 substrate for KDM5A*

Once we were able to construct our permissive KDM5A$_{1\text{-}588\Delta AP}$ recognition motif reflective of a 75% relative H3K4me3 activity threshold (Fig. 5.1E), we used this motif to identify high-ranking substrates in which we synthesize peptides of to monitor KDM5A$_{1\text{-}588\Delta AP}$ activity on (Fig. 5.2A). Applying this 75% motif yielded 20 high-ranking substrates to further assay *in vitro*. This was advantageous as it allowed us to reduce stringency of downstream analysis opposed to the application of other motifs (i.e 25%, 50%) which would have yielded high-ranking substrate numbers that were >100 (Appendix E, Fig. E1B). Once we synthesized peptides reflective of the high-ranking substrates, we assessed KDM5A$_{1\text{-}588\Delta AP}$ activity on these 20 high-ranking substrates *in vitro*. One of the peptides was reflective of the p53-K370me3 predicted substrate and we observed significant KDM5A$_{1\text{-}588\Delta AP}$ activity towards it (Fig 5.2B). As we were able to monitor KDM5A$_{1\text{-}588\Delta AP}$ activity towards the p53-K370me3 substrate peptide, we wanted to determine if KDM5A$_{1\text{-}588\Delta AP}$ activity was observed towards the p53-K370 peptide regardless of its methyl state. We observed significant KDM5A$_{1\text{-}588\Delta AP}$ activity ($p < 0.05$) towards only the p53-K370me3 substrate, and non-significant activity to the other p53-K370 peptides of varying methylation degrees, in relation to the H3K4me3 control peptide (Fig. 5.2C). As we had already purified KDM5A$_{1\text{-}801}$ and validated its activity and methyl-state specificity (shown in previous Chapter 3; Appendix E, Fig. E2A-C), we then wanted to further validate this potential p53-K370me3 substrate using sf9 insect-cell expressed KDM5A$_{1\text{-}801}$. We confirmed *in vitro* that *sf9* expressed KDM5A$_{1\text{-}801}$ displayed comparable activity ($p < 0.05$)

towards the p53-K370me3 substrate peptide that it did with the WT H3K4me3 substrate peptide activity (Fig. 5.2D). We also verified that KDM5A$_{1-801}$ activity towards the p53-K370me3 peptide was not only methyl specific, but site specific as well (Appendix E, Fig. E3). Given there are other methylated lysine residues in the C-terminal regulatory domain of p53[182], we wanted to ensure we were assessing the correct position with the potential KDM5A$_{1-801}$ demethylation of this substrate. Furthermore, we also confirmed that *in vitro,* KDM5A is the only member of the KDM5 family the appears to target the p53-K370me3 substrate (Fig. 5.2E). From these results we decided to go forth and begin examining if KDM5A targets the p53-K370me3 substrate in a cellular context.

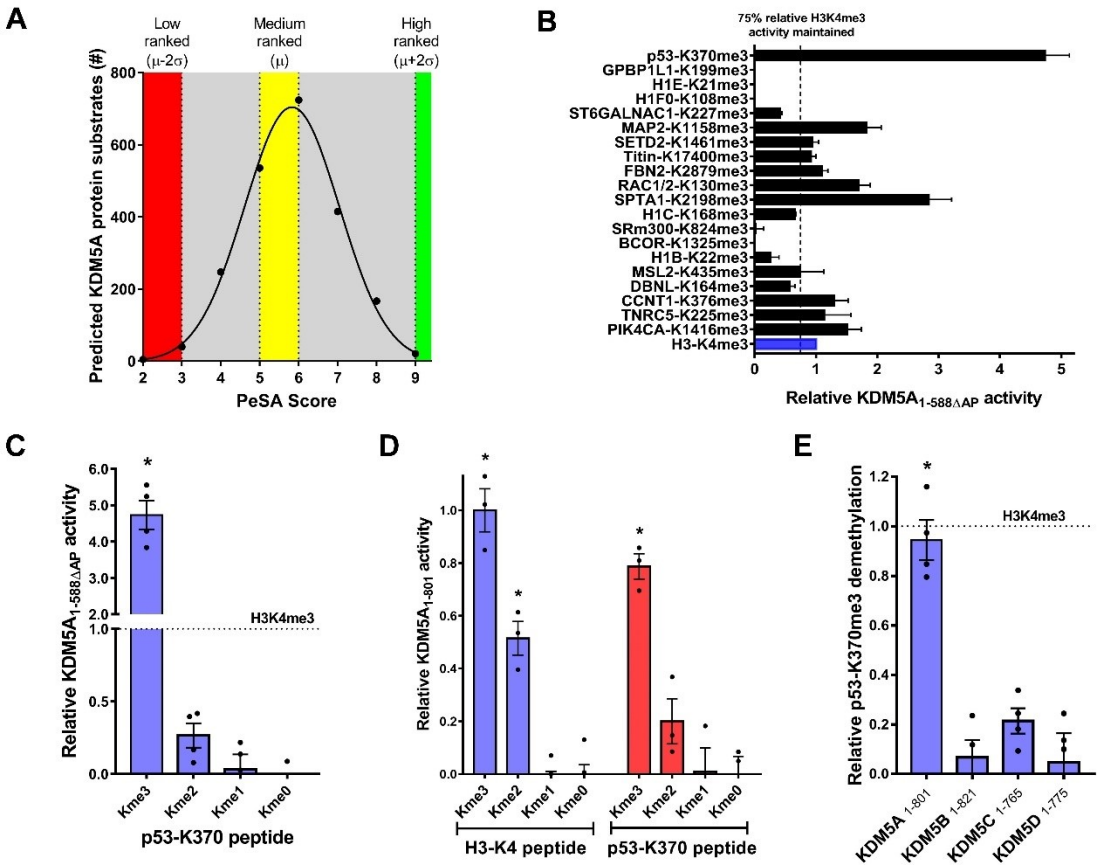



**Figure 5.2**. **Monitoring KDM5A demethylation of p53-K370me3 substrate *A.*** KDM5A predicted substrates based on ΔAP KDM5A recognition motif. Approximately 2100 predicted substrates were fit to a Gaussian distribution for further statistical analysis. ***B.*** Detection of KDM5A$_{1\text{-}588\Delta AP}$ demethylation activity towards high ranked synthesized substrate peptides. Results are mean ± std. error (n=3). ***C.*** Detection of KDM5A$_{1\text{-}588\Delta AP}$ demethylation of p53-K370me0, p53-K370me1, p53-K370me2, and p53-K370me3 peptides. Luminescence averages were normalized by background signal, and relatively quantified to H3K4me3 peptide luminescence (set to 1). Results are mean ± std. error (n=3). ***D.*** Detection of KDM5A$_{1\text{-}801}$ demethylation of H3K4me0, H3K4me1, H3K4me2, H3K4me3 and p53K370me0, p53K370me1, p53K370me2, and p53K370me3

peptides. Luminescence averages were normalized by background signal, and relatively quantified to H3K4me3 peptide luminescence (set to 1). Results are mean ± std. error (n=3). ***E.*** Detection of KDM5A$_{1-801}$, KDM5B$_{1-821}$, KDM5C$_{1-765}$, and KDM5D$_{1-775}$ demethylation of p53K370me3 peptide. Luminescence averages were normalized by background signal, and relatively quantified to H3K4me3 peptide luminescence (set to 1). Results are mean ± std. error (n=3). Figure caption continued from Figure 5.2 (page 163).

*Examining novel KDM5A mediated demethylation of p53-K370me3 substrate in HCT116 cells and subsequent biological effects*

We decided to examine the potential KDM5A mediated demethylation of p53-K370me3 in HCT 116 cells as it is a cell line with established WT p53-K370me function[179]. A full-length WT KDM5A overexpression construct, as well as a catalytically inactive KDM5A H483A mutant construct, was used to determine whether we could observe changes in the methylation levels of p53-K370 as a result of KDM5A activity. To begin, we first sought out to examine H3K4me3 levels in HCT 116 cells upon treatment of the cells with our overexpression plasmids. We overexpressed our pcDNA3.1 vector control, WT KDM5A and Mut KDM5A$_{H483A}$ overexpression plasmids and immunoblotted for H3K4me3 using isolated histone protein (Fig. 5.3A). As expected, when compared to the vector control, we observed a decrease in H3K4me3 levels upon WT KDM5A overexpression and an increase in H3K4me3 levels upon Mut KDM5A$_{H483A}$ overexpression. From these results we decided to proceed with examination of the p53-K370 methylation.

To examine p53-K370 methylation levels in HCT 116 cells, we overexpressed pcDNA3.1 vector control, WT KDM5A and Mut KDM5A$_{H483A}$ and examined p53-K370 methylation levels (Fig. 5.3B). Figure 5.3B shows a distinctive increase of p53-K370me2/1 levels in response to WT KDM5A overexpression, when compared to the vector only control and Mut KDM5A$_{H483A}$ overexpression. Specifically, we observed an increase of over 45% ($p < 0.05$; n=3) in p53-K370me1 levels (Fig. 5.3C) as a response to WT KDM5A

overexpression, when compared to the vector control. Similarly, we observe an increase of over 45% ($p < 0.05$; n=3) in p53-K370me2 levels (Fig. 5.3D) as a response to WT KDM5A overexpression when compared to the vector control. As there are no commercially available p53K370me3 antibody that exists, and we were unsuccessful in generating a specific custom antibody, we aimed to monitor p53-K370me3 via PRM-MS/MS. This was achieved through tryptic cleavage of the protein and C18-cleanup to monitor p53-K370me3 methylation as a response to overexpression of pcDNA3.1 vector control, WT KDM5A and Mut KDM5A$_{H483A}$ in HCT116 cells. We were able to see a decrease in the detection of *EPGGSRAHSSHLK(me3)* peptide in response to WT KDM5A overexpression in by over 60% and an increase by 20% upon Mut KDM5A$_{H483A}$ overexpression (Fig. 5.3E-F). This result is supported as we observed a decrease in p21 levels as a response to WT KDM5A overexpression when compared to the vector control by western blotting (Fig. 5.3B). This same result was observed when performing qPCR on *p21* and *p53* from HCT 116 cells (Fig. 5.3G). We observe a significant reduction ($p < 0.05$; n=4) in *p21* expression when WT KDM5A was overexpressed, whereas *p53* expression maintained relatively consistent suggesting that changes in p53 activity in response to KDM5 expression are not likely driven through changes in H3K4me3 at p53 promoter sites.

We then sought out to identify if KDM5A expression has subsequent effects on a known interactor of p53. Specifically, given the effect that p53-K370me has on directing p53 interaction with 53BP1, we wanted to see if

KDM5A mediated demethylation of p53K370me3 would encourage or antagonize this interaction[67]. To begin, as 53BP1 is recruited to interact with p53 in a DNA damage-responsive manner, we first had to encourage this interaction by inducing DNA damage[183]. Using doxorubicin, a well-established inducer of DNA damage through both intercalation into DNA and inhibition of topoisomerase-II-mediated DNA repair, we aimed to determine an optimal concentration to both activate p53 transcriptional activity, as well as increase 53BP1 expression[184]. We were able to demonstrate an optimal [doxorubicin] of 1 μM was sufficient in both increasing 53BP1 expression and activating p53 transcriptional activity (Fig 5.3H). We then overexpressed pcDNA3.1 vector control, WT KDM5A and Mut $KDM5A_{H483Ain}$ in tandem with treating 1 μM of doxorubicin in HCT 116 cells to determine if we could see if KDM5A expression had effects on the p53-53BP1 interaction. We were able to observe an increase in the amount of 53BP1 that was co-immunoprecipitated with p53, as a result of WT KDM5A expression (Fig. 5.3I). This data suggests that KDM5A overexpression drives the demethylation of the p53-K370me3 down from the tri-methylated state, allowing for 53BP1 to interact with p53. This is significant as it is the first documentation of how KDM5A can not only target a non-histone protein for demethylation, but as a result influences the subsequent biological interaction with methyl binding domain-containing proteins.

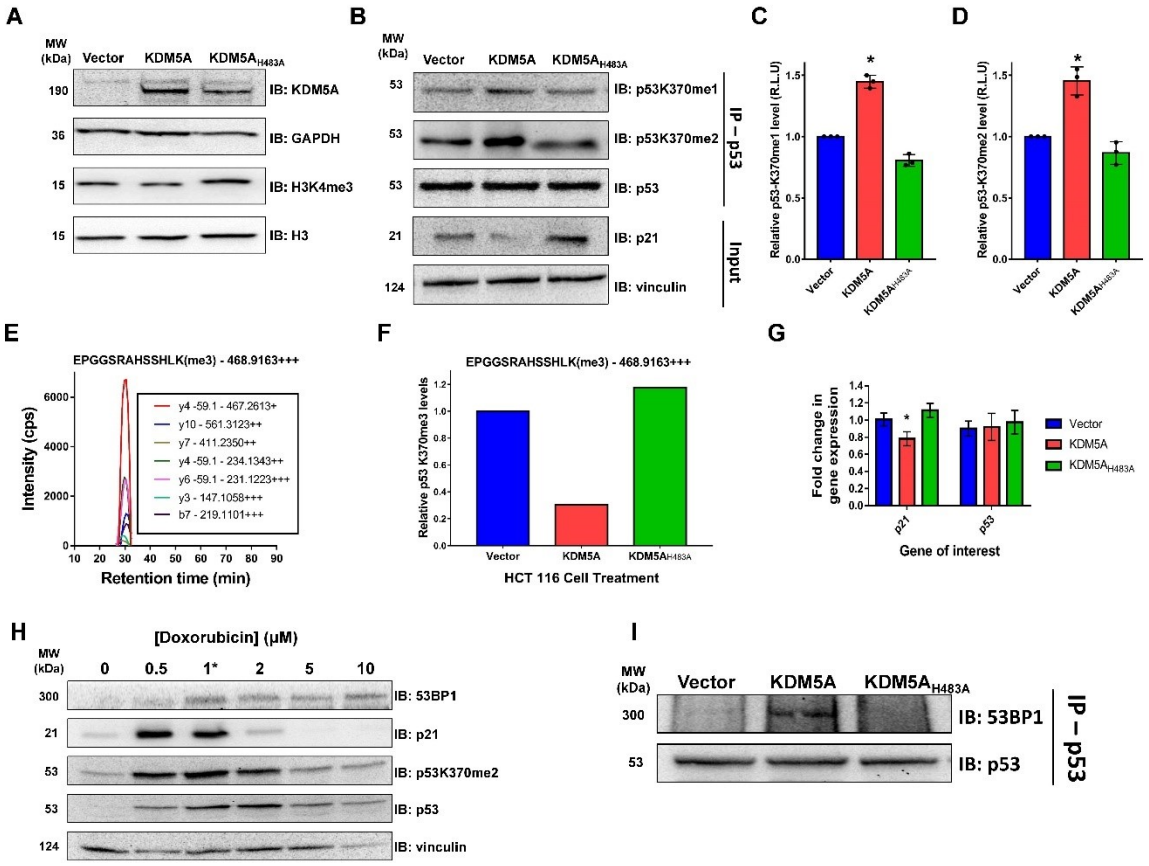



**Figure 5.3. Analysis of KDM5A mediated demethylation of p53-K370me3 substrate in HCT116 cells *A.*** Western blot analysis to confirm overexpression of WT KDM5A and Mut H483A KDM5A plasmids in HCT 116 cells. ***B.*** Western blot analysis of p53 immunoprecipitation (IP) samples to test effects of KDM5A over-expression on p53-K370me1/2, total p53 and p21 expression in HCT 116 cells. ***C.*** Densitometry of p53-K370me1 levels in response of WT KDM5A and Mut H483A KDM5A overexpression. Results are mean ± std. error (n=3). * Indicates $p < 0.05$ from vector control. ***D.*** Densitometry of p53-K370me2 levels in response of WT KDM5A and Mut H483A KDM5A overexpression. Results are mean ± std. error (n=3). * Indicates $p < 0.05$ from vector control. ***E-F.*** PRM-MS/MS analysis of p53 immunoprecipitation (IP) from HCT116 cells to determine

levels of p53 K370me3 expression upon KDM5A over-expression, analyzed through mass spectrometry. ***G.*** qPCR analysis of *p53* and *p21* gene expression in HCT116 cells to test p53 transcriptional activity because of KDM5A over-expression. Results are mean ± std. error (n=4). * Indicates $p < 0.05$ from vector control. ***H***. Doxorubicin does response in HCT 116 cells. Western blot analysis to determine optimal concentration of Doxorubicin to use in HCT 116 cells for p53 activation and to induce 53BP1 expression. * Indicates the concentration of doxorubicin used going forward. ***I.*** Western blot analysis of 53BP1-p53 co-immunoprecipitation in HCT116 cells in response to WT KDM5A and Mut $KDM5A_{H483A}$. Figure caption continued from Figure 5.3 (page 168).

*Examining novel KDM5A mediated demethylation of p53-K370me3 in HEK 293*

Upon examining these results in HCT 116 cells, we then looked to investigate whether these results could be duplicated in another cell line. We chose HEK293 cells as they are also a cell line expressing WT p53[185] and we followed the same experiments as with HCT 116 cells. Briefly, we first validated the expected changes in H3K4me3 methylation is response to WT/Mut KDM5A overexpression (Appendix F, Fig. F2A). We then examined p53-K370 methylation levels in HEK 293 cells in response to WT/Mut KDM5A overexpression (Appendix F, Fig. F2B). In HEK 293 cells, we observed an increase of 80% ($p < 0.05$; n=3) in p53-K370me1 levels in response to KDM5A overexpression, when compared to the vector control (Appendix F, Fig. F2C). Likewise, we observe an increase of over 100% ($p < 0.05$; n=3) in p53-K370me2 levels in response to KDM5A overexpression when compared to the vector control (Appendix F, Fig. F2D). Upon observing these increases in p53-K370me2/1 level, we then aimed to determine if p53-370me3 levels were responsive to WT and Mut KDM5A overexpression via PRM-MS/MS. Again, through tryptic cleavage of p53 IP and C18-cleanup, we were able to observe p53-K370me3 methylation as a response to overexpression of pcDNA3.1 vector control, WT KDM5A and Mut $KDM5A_{H483A}$ in HEK293 cells. We were able to see a decrease in the amount of detectable *EPGGSRAHSSHLK(me3)* peptide in response to WT KDM5A overexpression in by over 40% and an increase by over 100% upon Mut $KDM5A_{H483A}$ overexpression (Appendix F, Fig. F2F-G). Similarly, p21 protein expression was observed in HEK 293 treated cells and found to

decrease following WT KDM5A overexpression by western blotting (Appendix F, Fig. F2B). We also show that *p21* expression decreased when WT KDM5A was overexpressed, whereas *p53* expression maintained relatively consistent (Appendix F, Fig. F2E).

*Determining novel TAF5 interaction with p53 through K370me3 methylation status*

As it is unlikely that p53-K370me3 serves solely to generate lower methylation states, the last avenue we wanted to investigate was if we could identify any PPIs associated with the p53-K370me3 modification. To identify a potential PPI, collaborators at the Institute for Advanced Biosciences at Grenoble Alpes University performed a peptide pull down experiment coupled with reductive dimethylation for stable isotope labelling and mass spectrometry (LC-MS) analysis in order to identify any novel PPIs associated with the p53-K370me3 (Fig. 5.4A). Briefly, p53-K370me interacting proteins were enriched by incubating whole cell lysate with immobilized p53-K370me3 or p53-K370me0 peptide. While interacting proteins were still bound to the p53-K370me3/0-bead complex, they were denaturated, trypsin digested and C18 cleaned. The digested peptides were subjected to differential dimethyl labeling through either light ($CH_2O$) or heavy ($CD_2O$) labeled formaldehyde, which naturally reacts with primary amines which are present on the N-terminus of the digested peptides[186]. As these isotopes have an associated mass shift (+6.0377 Da / free amine), through pooling *heavy* and *light* treated formaldehyde peptides together one can

quantify relative interacting protein abundance between p53-K370me3/0 protein binders[186].

We investigated whether we could pull out interacting proteins from cell lysate using p53K370me3/2/1 peptide(s) as bait, and a p53K370me0 peptide as a control (Table 5.1; Appendix H, Table H1, H2). These experiments successfully identified TAF5, TAF9 and CDYL (Fig. 5.4B) as p53-K370me3 interactors; however, TAF5 was the only one quantifiable interaction to pulled out from both conditions (Table 5.1). Moreover, TAF5 had the largest number of unique peptides pulled out and identified with the p53-K370me3 peptide and therefore, was the most confident result (Table 5.1). TAF5 (Transcription initiation factor TFIID subunit 5) is a TATA-box binding protein (TBP) and part of the transcription factor IID (TFIID) complex and associates strongly with other TAF complexes involved in H3-related TAF complexes involving TAF9/TAFII31[187]. Fundamentally, TATA-box binding proteins binds with the negative charge in the phosphate groups in the backbone of DNA through positively charged arginine and lysine residues in the protein[187]. Recently, researchers have shown the ability to pull out a number of TAF proteins, including TAF5 with p53 in U2OS cells[188]. They identified that TAF5 appears to interact in the C-terminus of p53 but were unable to identify the mechanism of the interaction[188]. As a result of this insight and the data obtained from our collaborators, we look to determine if this interaction is dependent on the p53-K370me3 site.

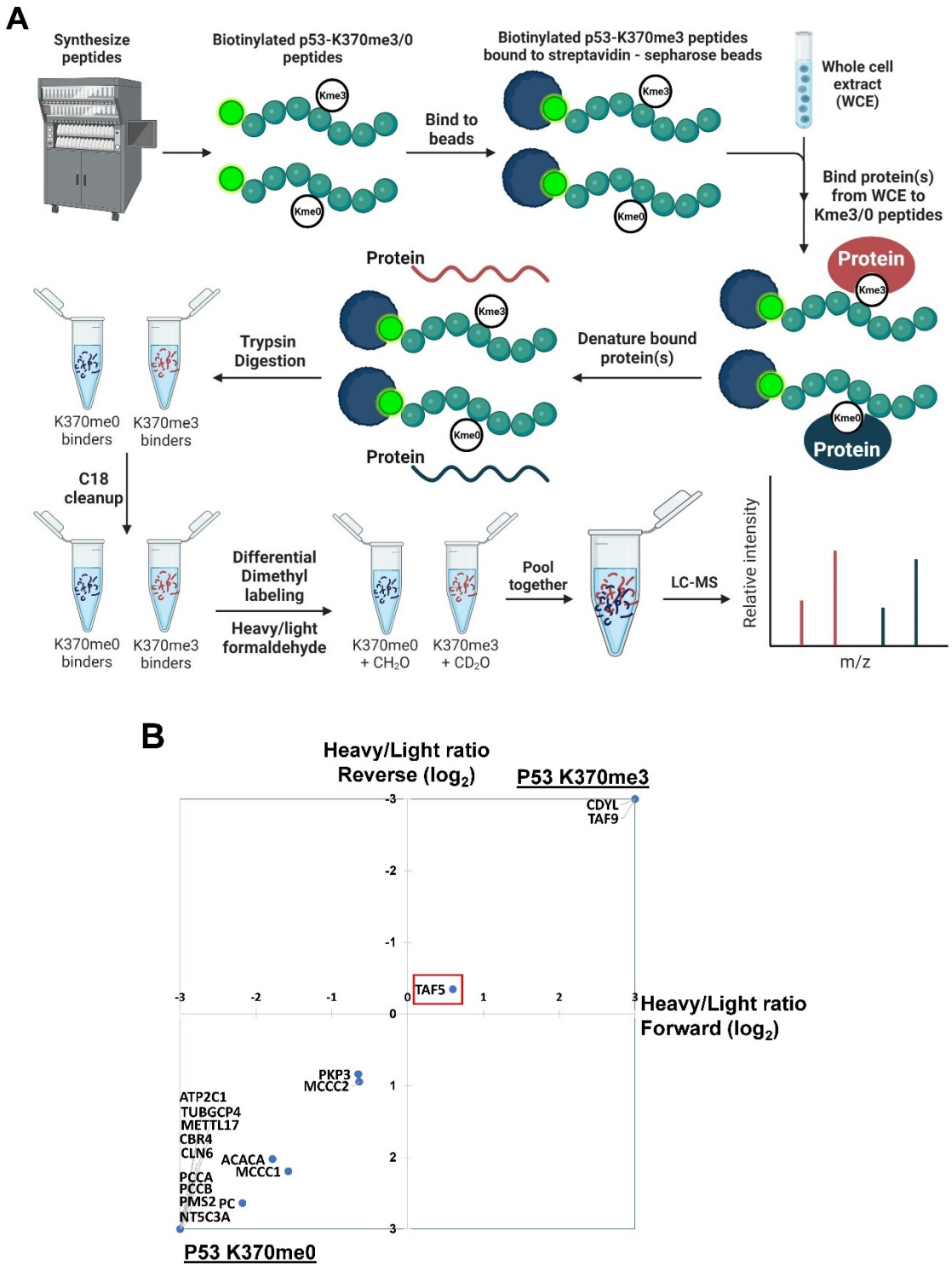



**Figure 5.4. Identification of new TAF5 interaction with p53 K370me3 site through p53-K370me3 peptide pull down coupled with reductive dimethylation.** ***A.*** Experimental overview of peptide pull down experiment

coupled with reductive dimethylation for stable isotope labelling and mass spectrometry (LC-MS) analysis. Biotinylated p53K370me3/2/1 peptides were used as bait (Biotin(N-terminal)-ahx-*RAHSSHLK(me3/2/1)SKKGQST)* and a biotinylated p53K370me0 (Biotin(N-terminal)-ahx-*RAHSSHLK(me0)SKKGQST)* peptide was used as a control. Created with BioRender.com. ***B***. Results of p53-K370me3 peptide pull down coupled with reductive dimethylation. MS/MS analysis identified 3 unique proteins that interacted strongly with p53-K370me3 peptide. TAF5 was only unique repeating peptide shown to display a quantifiable change in response to both heavy isotopic and light lysine labeling. Figure caption continued from Figure 5.4 (page 174).

**Table 5.1. Results of p53-K370me3 peptide interactomics through p53-K370me3 peptide pull down coupled with reductive dimethylation**. MSMS analysis identified 3 unique protein peptides that interacted strongly with p53-K370me3 peptide.

| | Gene name | Protein names | H/L ratio Forward Log2 | H/L ratio Reverse Log2 | Unique peptides Forward | Unique peptides Reverse |
|---|---|---|---|---|---|---|
| K370me3 | CDYL | Chromodomain Y-like protein | H only | L only | 1 | 1 |
| | TAF9 | Transcription initiation factor TFIID subunit 9;Transcription initiation factor TFIID subunit 9B | H only | L only | 2 | 1 |
| | TAF5 | Transcription initiation factor TFIID subunit 5 | 0,597507338 | -0,34587613 | 6 | 6 |
| K370me0 | MCCC2 | Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial | -0,638119373 | 0,945907297 | 18 | 15 |
| | PKP3 | Plakophilin-3 | -0,649525384 | 0,839072743 | 19 | 18 |
| | MCCC1 | Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial | -1,572350922 | 2,195221593 | 10 | 13 |
| | ACACA | Acetyl-CoA carboxylase 1 | -1,780710647 | 2,024071406 | 71 | 69 |
| | PC | Pyruvate carboxylase, mitochondrial | -2,177620628 | 2,638050661 | 42 | 45 |
| | PCCA | Propionyl-CoA carboxylase alpha chain, mitochondrial | L only | H only | 13 | 15 |
| | PCCB | Propionyl-CoA carboxylase beta chain, mitochondrial | L only | H only | 12 | 10 |
| | NT5C3A | Cytosolic 5-nucleotidase 3A;5-nucleotidase | L only | H only | 3 | 2 |
| | PMS2 | Mismatch repair endonuclease PMS2 | L only | H only | 3 | 3 |
| | CLN6 | Ceroid-lipofuscinosis neuronal protein 6 | L only | H only | 2 | 1 |
| | CBR4 | Carbonyl reductase family member 4 | L only | H only | 1 | 1 |
| | METTL17 | Methyltransferase-like protein 17, mitochondrial | L only | H only | 1 | 1 |
| | ATP2C1 | Calcium-transporting ATPase type 2C member 1;Calcium-transporting ATPase | L only | H only | 1 | 3 |
| | TUBGCP4 | Gamma-tubulin complex component 4 | L only | H only | 1 | 2 |

**P53 K370 me0 vs me3 interactors (>1600 identifications; 17 hits)**

*Characterization of the novel TAF5-p53-K370me3 interaction*

To investigate whether this p53-TAF5 interaction occurs in HCT 116 cells, we first began with attempting to co-IP TAF5 with p53. We were successful in co-immunoprecipitating TAF5 and p53 using an anti-p53 antibody (Fig. 5.5A). In order to determine whether we could further encourage/improve this interaction, we attempted to use Nutlin 3A to stabilize p53[189]. As typically MDM2 binds to p53 and negatively regulates its transcriptional activity and stability, Nutlin 3A functions by inhibiting MDM2 binding of p53[189]. Through chemical MDM2 inhibition, we wanted to observe whether we could further encourage the TAF5-p53 interaction. After performing a Nutlin 3A dose response in HCT 116 cells, we determined an optimal [Nutlin 3A] to use was 10 μM (Appendix G, Fig. G1). Upon treating cells with Nutlin 3A, we were able to see a slight (~10%) increase in the amount of TAF5 that was co-immunoprecipitated with p53 (Fig. 5.5B). From these results, we looked to determine what effect WT KDM5A and Mut $KDM5A_{H483A}$ expression would have on this TAF5-p53 interaction. We observed that both WT KDM5A and Mut $KDM5A_{H483A}$ overexpression increased the amount of TAF5 that was immunoprecipitated with p53 (Fig. 5.5C). When factoring in the amount of p53 used to pull out TAF5, we observed a >50% increase in the amount of TAF5 that was co-immunoprecipitated with p53, as a result of Mut $KDM5A_{H483A}$ overexpression (Fig. 5.5D). Here, we have demonstrated how by increasing the amount of mutant, inactive $KDM5A_{H483A}$ expression in HCT116 cells suggests the encouragement of the novel TAF5-p53K370me3 interaction. Ultimately, this is significant as not only have we provided proof of a novel PPI

between TAF5 p53 through the K370me3 site, but we have shown how KDM5A expression influences the amount of TAF5 that interacts with p53. Moreover, we have demonstrated how KDM5A can target this first identified non-histone substrate for the KDM5 family in p53-K370me3, but also how KDM5A mediated demethylation of this substrate can influence subsequent biological PPIs associated with p53.

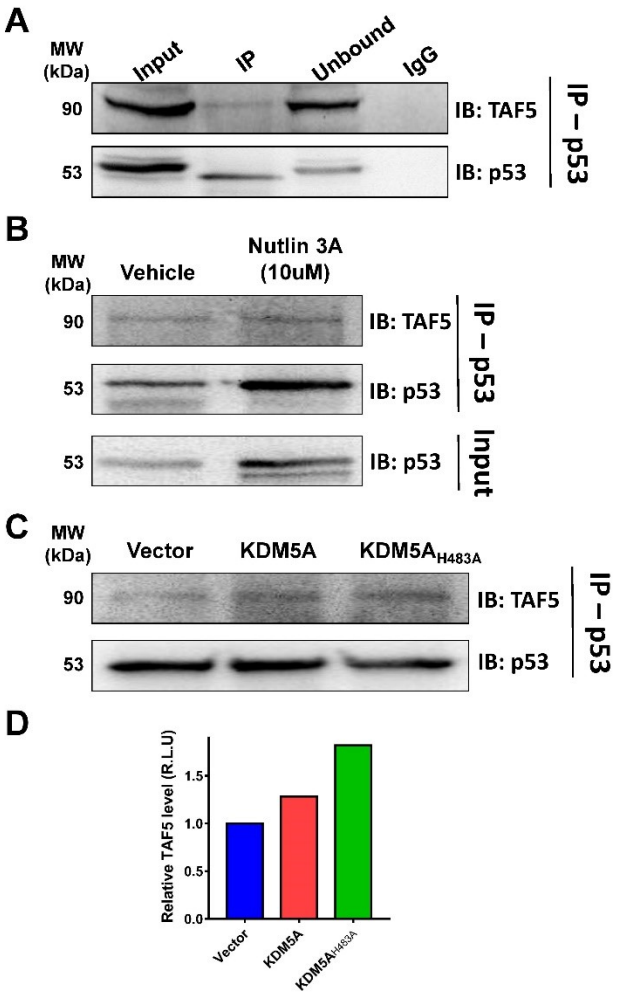



**Figure 5.5. Investigation of KDM5A influence on novel TAF5-p53 interaction.** ***A.*** Western blot examination of TAF5-p53 co-immunoprecipitation in HCT116. ***B.*** Western blot analysis of Nutlin 3A-mediated stability of TAF5-p53 interaction. TAF5-p53 were co-immunoprecipitation in HCT116. ***C.*** Western blot analysis of TAF5-p53 co-immunoprecipitation in HCT116 cells in response to WT KDM5A and Mut KDM5A$_{H483A}$ expression in tandem with 10 μM Nutlin 3A treatment. ***D.*** Densitometry of TAF5 levels relative to p53 levels in response of WT KDM5A and Mut H483A KDM5A overexpression in tandem with 10 μM Nutlin 3A treatment.

## Discussion

Using our approach to studying KDM5A$_{1\text{-}588\Delta AP}$ substrate preference using our highly specialized PPS library, we were able to determine KDM5A$_{1\text{-}588\Delta AP}$ substrate preference. Our analysis resulted in an enrichment of candidate KDM5A substrates from the thousands unique sites within the currently methyllysine proteome[130]; identifying 20 *in vitro* substrates for KDM5A that we assessed *in vitro* (Fig. 5.2A-B). From these, KDM5A$_{1\text{-}588\Delta AP}$ demonstrated significant *in vitro* activity towards the p53-K370me3 substrate (Fig. 5.2B). We subsequently demonstrated KDM5A$_{1\text{-}801}$ activity towards the p53-K370me3 substrate and consequently, began investigating this enzyme-substrate relationship in a cellular context. Our results demonstrate that KDM5A demethylates p53 at the lysine 370 residue in HCT 116 cells (Fig. 5.3B-F), as well as HEK 293 cells (Appendix F, Fig. F2B-D, F-G) driving the methylation state down to the lower K370me2 and K370me1 states. We also demonstrated how KDM5A expression influences p53's interaction with 53BP1 (Fig. 5.3I) and how KDM5A expression influences a novel p53 PPI with TAF5 (Fig. 5.5C); an interaction not previously reported in literature.

It is well established that KDM5 demethylases target the histone H3K4me3/2 substrate for demethylation, leading to transcriptional regulation[48]. As a result, this is where majority of the research surrounding the cellular function of KDM5 have been devoted towards[69]. Here, we have provided the first evidence that a member of the KDM5 family, KDM5A, can target a non-histone substrate for demethylation. This study effectively demonstrates another

mechanism of how KDM5A might influence cellular regulation. In this study, we have found that the p53-K370me2/1 level increase in response to WT KDM5A overexpression, suggesting that KDM5A targets the p53-K370me3/2 substrate (Fig. 5.3B). These results compliment the study done by Huang et al., who demonstrated that KDM1A/LSD1 targets the p53-K370me2/1 substrate for demethylation[67]. It should be noted that KDM5A and KDM1A both target the H3K4me3/2 and H3K4me2/1 substrates, respectfully[48,61]. This is significant as this is the first evidence of KDMs from different families that target two of the same substrates, one histone and one non-histone.

There has been extensive research done on how different methylation states (e.g., Kme1 vs. Kme2) can dictate the binding of specific effector proteins containing MBDs[30,35,39,50,52]. In this study, we found that KDM5A-mediated changes to p53-K370 methylation levels effected the interaction between p53 and the Tudor MBDs of 53BP1 (Fig. 5.3I). During stress such as DNA damage, 53BP1 binds to p53-K370me2 through the 53BP1 tandem Tudor MBDs, an interaction specific to that methylation site and state[67]. The demethylation activity of KDM5A therefore appears to encourage the p53-53BP1 interaction upon DNA damage. This can be seen in the decrease in p21 levels in Figure 5.3B. The decrease in p21 levels is significant as the KDM5A demethylation of p53 to the K370me1 state has been demonstrated to repress the association of p53 with coactivator 53BP1[67].This collectively has been shown to result in a decrease in p53 activity and a decrease in p53-responsive genes, such as *p21*[67,190]. As we did not induce DNA damage in Figure 5.3B, this would explain the low levels

seen in p21 expression. As a result of activating p53 transcriptional activity through doxorubicin treatments, we see an increase in p21 levels (Fig. 5.3H). This suggests that KDM5A overexpression is driving p53-K370 to the mono-methyl state (p53-K370me1) when DNA damage is not induced, and to the p53-K370me2 state in the presence of DNA damage (Fig. 5.3I). However, the fine-tune control of methylation dynamics is complex and is likely influenced through the opposing actions of KDM5A and an unknown KMT.

The p53 protein has been well investigated to be functionally regulated through its highly basic C-terminal regulatory domain (CTD), effectively altering its transcriptional activity[182,191]. This dynamic regulation can be facilitated through the variety of effector proteins that bind to p53 in response to one, or many, specific PTMs[182,191]. In this Chapter, we have provided the possible mechanism of how the TAF5 transcriptional activator can interact with p53, an interaction facilitated by p53-K370me3, and that KDM5A can remove this modification to encourage p53 interaction with 53BP1. Using AP-MS/MS with p53-K370me0/3 peptide as bait, the ability to enrich TAF5 from a total cell lysate allowed us to dive deeper into this potential novel PPI. Interestingly, TAF5 has already been shown to interact with p53 c-terminus and contains six WD40 MBDs that form a closed β-propeller structure to mediate methyllysine binding[36,187,188]. In fact, complementary research has shown the ability of the WD40 MBD to be able to bind the Kme3 mark, a trait not observed in every MBD family[192]. In this study, we found that KDM5A$_{H483A}$ overexpression increased the amount of TAF5 pulled out with p53 upon p53 stabilization with Nutlin 3A (Fig. 5.5C-D). This is significant

as not only have we identified a MBD-containing transcription factor that interacts with this p53-K370me3 site, but have shown that this interaction can be influenced by KDM5A expression. Regardless, this interaction and KDM5A-mediated demethylation of the p53-K370me3 substrate adds an extra layer into the dynamic and complex methyl-mediated regulation of the p53 CTD.

As a result of our findings, we propose a dynamic p53 regulation through KDM5A-mediated demethylation (Fig. 5.6). We show that KDM5A can demethylate p53-K370me3 to lower K370me1/2 states. In the presence of DNA damage, this KDM-substrate relationship is likely further influenced by an unknown K370me1 KMT to stabilize the p53-K370me2 state to encourage the p53-53BP1 interaction[177]. Our findings further propose a novel interaction between TAF5 and p53-K370me3 whose interaction does not depend on DNA damage and may possibly influence p53 transcriptional activity. Ultimately, our data suggests that the p53-K370me3/2 is the first non-histone substrate identified for any KDM5 enzyme and given the status of p53 as an oncoprotein, represents a modification of biological interest.

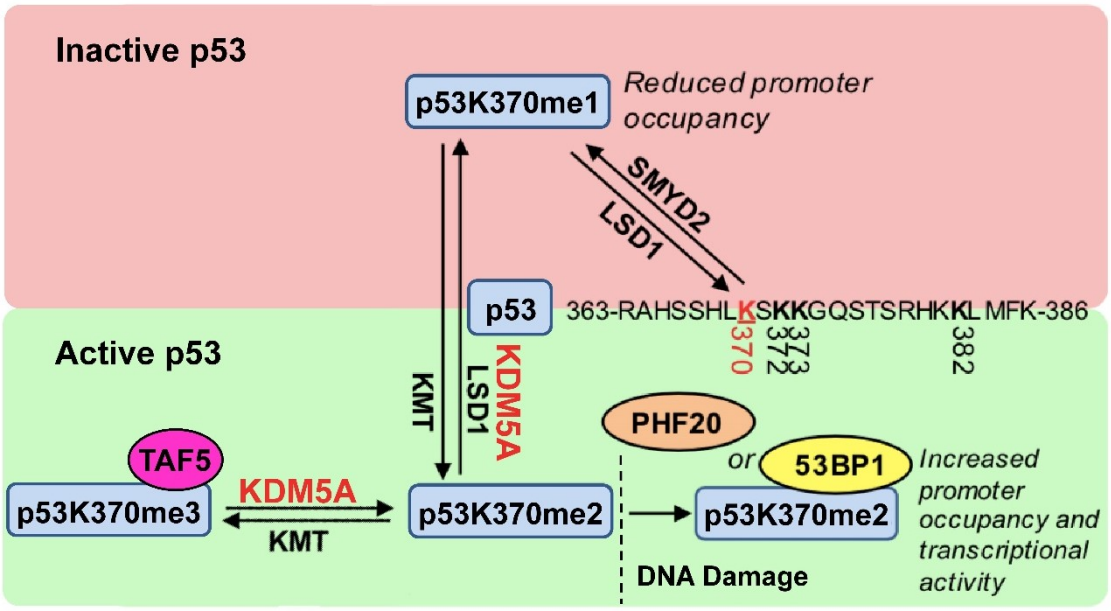



**Figure 5.6. Schematic model for the dynamic regulation of p53-K370 methylation and the role in regulating p53 activity and p53 protein interactors.** When p53 is in the K370me1 state it is not transcriptionally active (red zone). When p53 is in the K370me2 state, it remains transcriptionally active (green zone). We propose when p53 is in the K370me3 state, it remains transcriptionally active as well as p53-K370me3 interacts with TAF5.

# Chapter 6

# Conclusion

**Matthew Hoekstra[1,2], Kyle Biggar[1,2]**

[1]Institute of Biochemistry, Carleton University, Ottawa, ON, K1S 5B6, Canada

[2]Department of Biology, Carleton University, Ottawa, ON, K1S 5B6, Canada

*Novel systematic analysis of KDM5 family substrate preference*

Just as other groups have done for the prediction of KMT substrates using permutated peptide arrays, we were able to apply the same underlying theoretical concepts to KDM substrate identification[84,90,95,96,135]. We successfully altered this permutated peptide array-based approach by transitioning this system to in-solution peptides opposed to peptides immobilized to a membrane to analyze KDM substrate recognition and activity. As such, by assessing KDM *in vitro* demethylation activity towards in-solutions peptides using an established substrate as a "positive control", we can ensure any positive signal around or equal too that of the "positive control" is in fact a sign of active *in vitro* demethylation. Moreover, by using systemically permutated peptides that deviate by one residue of an established substrate(s), we can determine the influence neighbouring amino acid mutations have on promoting or inhibiting activity[84,90,95,96,135].

**Using our novel PPS library, we have performed the first systematic analysis of the substrate preference of any KDM enzyme** (Appendix D, Fig. D3). While other groups have performed alanine mutagenesis scans or mutated specific residues in order to explore KDM5 PHD domain binding specificity, no one has performed this extensive of an analysis of KDM5 (or any KDM) substrate recognition[74]. Moreover, we demonstrated how our PPS library method can be used for any JMJD-type KDM demethylase, granted it has 1+ established *in vitro* substrate(s). The outcome of applying our PPS library method is a detailed, relative quantification of specific amino acid mutations surrounding a central

methyllysine residue that effect substrate demethylation. As matrix and recognition motif analysis programs has been previously developed by the Biggar lab, we were able to input our data and provide visual representations of the influence each individual mutation has on KDM5 activity (Appendix D, Fig. D3, Fig. D4)[125]. This allowed for an analysis of key residues or physio-chemical properties (i.e hydrophobic, charged, etc) that were critical in substrate recognition of a KDM5 demethylase. This has further become beneficial for KDM families such as the KDM5s as they have been elusive to having full crystal structure models obtained.

*Identification of KDM5 activity towards non-histone substrates*

Using data derived from our PPS library analysis, we were able to predict which methyllysine-containing proteins in the methyl proteome based are likely to be KDM5 substrates, based on sequence similarity to our KDM5 recognition motifs (Fig. 4.3A-B). Databases such as PhosphoSite have large datasets of known protein lysine methylation sites[130]. By scoring these methyllysine-modification sites, and their surrounding residues, against our substrate recognition motifs, we were able to predict substrates that are most likely to be demethylated by KDM5 *in vitro* (i.e. high-ranking substrates). Our analysis shows that our method is very effective (80-100%) in predicting which substrates will most likely not be demethylated *in vitro* (i.e medium/low-ranking substrates). Opposingly, it was fairly accurate (25-40%) in predicting which substrates would be demethylated by KDM5 *in vitro* (Fig. 4.4A). In other words, our method is able

to accurately identify true negatives to help refine and guide downstream identification efforts. Through our analysis, we have identified 66 unique candidate KDM5 substrates in which KDM5s displayed significant activity towards *in vitro* (Fig. 4.4B). Of these 66, we further validated the likelihood that, at least, 6 of these candidate KDM5 substrates might be targeted by KDM5 demethylases in cells (Fig. 4.5B-C). This was accomplished through examining changes in methylation status via chemical inhibition of KDM5A/B/C/D.

*Investigation of novel KDM5A non-histone substrate: p53-K370me3*

To truly show the potential of our PPS library-based method and substrate identification pipeline, we decided to focus much of our validation efforts on a single high-ranking KDM5A novel substrate; namely, p53-K370me3. After validating the ability for both recombinant KDM5A$_{1\text{-}588\Delta AP}$ and KDM5A$_{1\text{-}801}$ to demethylate the p53-K370me3 substrate *in vitro* (Fig. 5.2B-D), we further investigated in HCT 116 and HEK 293 cells. We revealed that KDM5A activity can dramatically influence the p53-K370 methylation state. Specifically, we showed that both p53-K370me2/1 levels increase, and p53-K370me3 levels decrease when WT KDM5A is overexpressed; these changes fail to be seen when overexpressing catalytically inactive KDM5A$_{H483A}$ (Fig. 5.3B-F; Appendix F, Fig. F2B-G). Moreover, we showed that upon DNA damage, KDM5A expression encourages the p53-53BP1 interaction (Fig. 5.3I); likely as a result from demethylating p53-370me3 to the lower Kme2 state, permitting 53BP1 interaction with p53 via its tandem Tudor MBD[67]. Beyond the scope of this single

substrate, our data indicates the first evidence of any KDM5-mediated demethylation of a non-histone substrate. At this time, the KDM5A-mediated demethylation of p53-K370me3 is the only evidence of a KDM5 non-histone substrate. Such a significant discovery supports our hypothesis that the KDM5 family has a more diverse function(s) in the cell then just solely histone H3K4me3/2 demethylation and subsequent transcriptional regulation.

*Identification of novel TAF5-p53K370me3 interaction*

As we had identified the p53-K370me3 site to be a KDM5A substrate, the biological significance of this Kme3 modification was still unknown. To gain insight into biological function, we set out to determine if there were any novel protein-protein interactions (MBD or otherwise) that could specifically exist via the p53-K370me3 site. Through collaboration with the Reynoird lab, a peptide pull down experiment coupled with reductive dimethylation for stable isotope labelling and mass spectrometry (LC-MS) analysis was applied to identify any protein that interacted with peptide corresponding to the p53-K370me3 sequence (and not the p53-K370me0/1/2 versions). From this analysis, we were able to identify a novel p53-TAF5 interaction existing only with the p53-K370me3 peptide (Fig. 5.4B). Although earlier groups have shown the interaction between TAF5 and p53, the specific residue and mechanism through which this PPI was governed was unknown[188]. Once we were able to co-IP TAF5 with p53 in HCT 116 cells, we then determined how WT KDM5A and Mut KDM5A$_{H483A}$ expression might influence this interaction. Our data shows that maintaining p53-K370 in the

Kme3 state permits more of the TAF5-p53 interaction to occur, suggesting that this interaction is more favoured by the Kme3 state (Fig. 5.5C-D). Other data showing how the WD40 repeat domains (present in TAF5) are able to bind to Kme3-containing peptides supports our findings[192]. Therefore, our data demonstrates the first evidence of a novel p53 PPI with TAF5 which is conveyed through the p53-K370me3 state.

*Conclusion(s) of thesis*

As a result of my thesis research, we have demonstrated that:

**1)** We have provided a novel approach for systematic characterization of substrate specificity and substrate prediction for any JMJD-type KDM demethylase with at least one established substrate.

**2)** Our method and pipeline were efficient at relatively quantifying KDM5 substrate selection and predicting KDM5 substrates *in vitro.*

**3)** KDM5A targets the non-histone p53-K370me3 substrate for demethylation.

**4)** KDM5A may have an influence on DNA damage response through p53-K370me3 demethylation to mediate p53 interaction with 53BP1.

**5)** TAF5 interacts with p53 through, at least, the K370me3 state

*Future Experiments*

As previously stated, although we have identified TAF5 to interact with p53 through its K370me3 residue, the effect TAF5 has on p53 transcriptional activity and/or other PPIs is unknown. As a result, one avenue to investigate is

whether TAF5 influence p53 transcriptional activity, such as other TAF proteins (e.g., TAF1 and TAF9) that are also known to interact with p53[193]. TAF9 was another protein identified in our analysis to potentially interact with p53-K370me3. TAF9 was demonstrated to be a transcriptional co-activator of p53, although this was shown to be mediated through p53's N-terminal domain[193]. TAF1 is another TAF protein shown to interact with p53[188]. TAF1 was shown to induce cell cycle G1 progression in a p53-depedent manner[188]. Moreover, this TAF1 kinase-mediated interaction contributes to p53 degradation, except upon DNA damage where p53 phosphorylation by TAF1 leads to p53 stabilization[188]. Both studies demonstrate how TAF9/1 have dynamic regulation of p53 function and stability[188,193]. These studies consequently open the door to investigate whether other members of the TFIID transcription factor complex, such as TAF5, have the potential to have dynamic regulation of p53 function and stability. As we have already identified the potential site of interaction, next steps would be to 1) confirm this interaction through means of crystal structure or p53-K370 mutagenesis and 2) determine if TAF5 influences p53 transcriptional activity or p53 stability. This could be accomplished by examining p53 target genes involved in pathways such as cell cycle arrest/regulation (coupled with flow cytometry), apoptosis and DNA damage in a TAF5 inducible cell line.

Although the aims of my thesis have been completed and we have conclusions to our original hypotheses, this research has opened the potential of other various KDM5 demethylation targets of other non-histone proteins. As we have only validated the KDM5A demethylation of the p53-K370me3 substrate,

there remains, at least, another 6 non-histone proteins with lysine residues where we observed a change in methylation upon in response to CPI-455-induced KDM5 inhibition. Therefore, future efforts should be focused on demonstrating the ability for these KDM5 enzymes to actively target non-histone protein substrates for demethylation, in a cellular context. As cross talk between PTMs become more apparent in recent years, one can look at these substrates and wonder if methylation might factor in the regulation of other proteins[194].

A variety of complex members had arisen in our analysis as high-ranking substrates, and some even had a response to KDM5 inhibition. For starters, it is known that KDM5B associates with the NuRD complex and one of our KDM5A/B-responsive peptide substrates, MTA1-K532, is also a known member of this complex[65]. Critically, MTA1-K532 was a high-ranking substrate and the MTA1-K532me2 has a functional significance in the formation of the NuRD complex[142]. Although the MTA1-K532me was not successfully detected in a PRM-MS/MS screen in any cellular condition, it should still be investigated whether this protein is a true substrate of KDM5B and whether KDM5B expression influences NuRD complex formation and regulation in an MTA-demethylation dependent context[65].

# References


1. Ponomarenko, E. A. *et al.* The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* **2016**, 1–6 (2016).
2. Legrain, P. *et al.* The Human Proteome Project: Current State and Future Direction. *Mol. Cell. Proteomics.* **10**, M111.009993 (2011).
3. Zoghbi, H. Y. & Beaudet, A. L. Epigenetics and Human Disease. *Cold Spring Harb. Perspect. Biol.* **8**, a019497 (2016).
4. Valencia, A. M. & Kadoch, C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. *Nat. Cell Biol. 2019 212* **21**, 152–161 (2019).
5. Moosavi, A. & Motevalizadeh Ardekani, A. Role of Epigenetics in Biology and Human Diseases. *Iran. Biomed. J.* **20**, 246–58 (2016).
6. Mosammaparast, N. & Shi, Y. Reversal of Histone Methylation: Biochemical and Molecular Mechanisms of Histone Demethylases. *Annu. Rev. Biochem.* **79**, 155–179 (2010).
7. Strahl, B. D. & David Allis, C. The language of covalent histone modifications. *Nat.* **403**, 41-45 (2000).
8. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications 564 Readers of histone modifications. *Nat. Publ. Gr.* **21**, 564–578 (2011).
9. Taverna, S. D. *et al.* How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers HHS Public Access Author manuscript. *Nat Struct Mol Biol* **14**, 1025–1040 (2007).

10. Xu, YM., Yao, Y., Lau, A.T.Y. (2018). Histone H3. In: Choi, S. (eds) *Encyclopedia of Signaling Molecules*. Springer, Cham. https://doi.org/10.1007/978-3-319-67199-4_101644

11. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).

12. Bateman, A. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

13. Gujral, P., Mahajan, V., Lissaman, A. C. & Ponnampalam, A. P. Histone acetylation and the role of histone deacetylases in normal cyclic endometrium. *Reprod. Biol. Endocrinol.* **18**, 1–11 (2020).

14. Gräff, J. & Tsai, L. H. Histone acetylation: molecular mnemonics on the chromatin. *Nat. Rev. Neurosci. 2013 142* **14**, 97–111 (2013).

15. Eberharter, A. & Becker, P. B. Histone acetylation: a switch between repressive and permissive chromatin. *EMBO Rep.* **3**, 224-229 (2002).

16. Seto, E. & Yoshida, M. Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb. Perspect. Biol.* **6**, 1-26 (2014).

17. Timmermann, S., Lehrmann, H., Polesskaya, A. & Harel-Bellan, A. Histone acetylation and disease. *Cell. Mol. Life Sci.* **58**, 728–736 (2001).

18. Meas, R. & Mao, P. Histone ubiquitylation and its roles in transcription and DNA damage response. *DNA Repair (Amst).* **36**, 36-42 (2015).

19. Cao, J. & Yan, Q. Histone Ubiquitination and Deubiquitination in Transcription, DNA Damage Response, and Cancer. *Front. Oncol.* **2**, 1-9 (2012).

20. Zhang, Y. Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Dev.* **17**, 2733–2740 (2003).

21. Zhou, W. *et al.* Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol. Cell.* **29**, 69–80 (2008).

22. Mattiroli, F. *et al.* RNF168 ubiquitinates K13-15 on H2A/H2AX to drive DNA damage signaling. *Cell.* **150**, 1182–1195 (2012).

23. Rossetto, D., Avvakumov, N. & Côté, J. Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics.* **7**, 1098-1108 (2012).

24. Kschonsak, M. & Haering, C. H. Shaping mitotic chromosomes: From classical concepts to molecular mechanisms. *Bioessays.* **37**, 755-766 (2015).

25. Lowndes, N. F. & Toh, G. W. L. DNA Repair: The Importance of Phosphorylating Histone H2AX. *Curr. Biol.* **15**, R99–R102 (2005).

26. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Nat. Publ. Gr.* **21**, 381–395 (2011).

27. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49**, e324-e346 (2017).

28. Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.* **11**, 384–400 (2012).

29. Blair, L. P., Cao, J., Zou, M. R., Sayegh, J. & Yan, Q. Epigenetic

Regulation by Lysine Demethylase 5 (KDM5) Enzymes in Cancer. *Cancers (Basel).* **3**, 1383–1404 (2011).

30. Teske, K. A. & Hadden, M. K. Methyllysine binding domains: Structural insight and small molecule probe development. *Eur. J. Med. Chem.* **136**, 14–35 (2017).
31. Herold, J. M. *et al.* Small Molecule Ligands of Methyl-Lysine Binding Proteins. *J. Med. Chem.* **54**, 2504-2511 (2011).
32. Nady, N. *et al.* Histone Recognition by Human Malignant Brain Tumor Domains. *J. Mol. Biol.* **423**, 702–718 (2012).
33. Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. & Peng, Z. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13**, 1435-1448 (2004).
34. Lux, S. E., John, K. M. & Bennett, V. Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nat.* **344**, 36–42 (1990).
35. Collins, R. E. *et al.* The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat. Struct. Mol. Biol.* **15**, 245–250 (2008).
36. Suganuma, T., Pattenden, S. G. & Workman, J. L. Diverse functions of WD40 repeat proteins in histone recognition. *Genes Dev.* **22**, 1265-1268 (2008).
37. Zou, X. D. *et al.* Genome-wide Analysis of WD40 Protein Family in Human. *Sci. Reports.* **6**, 1–11 (2016).

38. Song, J. J., Garlick, J. D. & Kingston, R. E. Structural basis of histone H4 recognition by p55. *Genes Dev.* **22**, 1313–1318 (2008).

39. Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell.* **121**, 859–872 (2005).

40. Cho, Y. W. *et al.* PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J. Biol. Chem.* **282**, 20395–20406 (2007).

41. Jain, K. *et al.* Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. *Epigenetics and Chromatin.* **13**, 1–11 (2020).

42. Baker, L. A., Allis, C. D. & Wang, G. G. PHD fingers in human diseases: Disorders arising from misinterpreting epigenetic marks. *Mutat. Res. Mol. Mech. Mutagen.* **647**, 3–12 (2008).

43. Torres, I. O. *et al.* Histone demethylase KDM5A is regulated by its reader domain through a positive-feedback mechanism. *Nat. Commun.* **6**, 6204-6228 (2015).

44. Shi, X. *et al.* Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J. Biol. Chem.* **282**, 2450–2455 (2007).

45. Karagianni, P., Amazit, L., Qin, J. & Wong, J. ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin

formation. *Mol. Cell. Biol.* **28**, 705–717 (2008).

46. Musselman, C. A. & Kutateladze, T. G. PHD Fingers: Epigenetic Effectors and Potential Drug Targets. *Mol. Interv.* **9**, 314-323 (2009).

47. Tahiliani, M. *et al.* The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nat.* **447**, 601–605 (2007).

48. Horton, J. R. *et al.* Characterization of a Linked Jumonji Domain of the KDM5/JARID1 Family of Histone H3 Lysine 4 Demethylases. *J. Biol. Chem.* **291**, 2631–2646 (2016).

49. Husmann, D. & Gozani, O. Histone lysine methyltransferases in biology and disease. *Nat. Struct. Mol. Biol.* **26**, 880–889 (2019).

50. Morera, L., Lübbert, M. & Jung, M. Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clin. Epigenetics.* **8**, 1–16 (2016).

51. Li, K. K., Luo, C., Wang, D., Jiang, H. & Zheng, Y. G. Chemical and biochemical approaches in the study of histone methylation and demethylation. *Med. Res. Rev.* **32**, 815–867 (2012).

52. Dillon, S. C., Zhang, X., Trievel, R. C. & Cheng, X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* **6**, 227-237 (2005).

53. Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12**, 1052–1058 (2002).

54. Wang, X. *et al.* Depletion of H3K79 methyltransferase Dot1L promotes cell invasion and cancer stem-like cell property in ovarian cancer. *Am. J.*

*Transl. Res.* **11**, 1145-1153 (2019).

55. Hamamoto, R., Saloura, V. & Nakamura, Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat. Rev. Cancer* **15**, 110–124 (2015).

56. Biggar, K. K. & S-C Li, S. Non-histone protein methylation as a regulator of cellular signalling and function. *Nat. Rev. Mol. Cell. Biol.* **16**, 5-17 (2015).

57. Huang, J. & Berger, S. L. The emerging field of dynamic lysine methylation of non-histone proteins. *Curr. Opin. Genet. Dev.* **18**, 152–158 (2008).

58. Hou, H. & Yu, H. Structural insights into histone lysine demethylation. *Curr. Opin. Struct. Biol.* **20**, 739-748 (2010).

59. Shi, Y. G. & Tsukada, Y. The Discovery of Histone Demethylases. *Cold Spring Harb. Perspect. Biol.* **5**, 1-3 (2013).

60. Trojer, P. & Reinberg, D. Histone Lysine Demethylases and Their Impact on Epigenetics. *Cell* **125**, 213–217 (2006).

61. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).

62. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7**, 715–727 (2006).

63. Tsukada, Y. I. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nat.* **439**, 811–816 (2006).

64. Longbotham, J. E. *et al.* Histone H3 binding to the PHD1 domain of histone demethylase KDM5A enables active site remodeling. *Nat. Commun.* **10**, 1-12 (2019).

65. Klein, B. J. *et al.* The histone-H3K4-specific demethylase KDM5B binds to its substrate and product through distinct PHD fingers. *Cell Rep.* **6**, 325–335 (2014).

66. Kumar, A., Kumari, N., Nallabelli, N. & Prasad, R. Pathogenic and Therapeutic Role of H3K4 Family of Methylases and Demethylases in Cancers. *Indian J. Clin. Biochem.* **34**, 123-132 (2019).

67. Huang, J. *et al.* p53 is regulated by the lysine demethylase LSD1. *Nat.* **449**, 105–108 (2007).

68. Ramadoss, S., Guo, G. & Wang, C.-Y. Lysine demethylase KDM3A regulates breast cancer cell invasion and apoptosis by targeting histone and the non-histone protein p53. *Oncogene.* **36**, 47–59 (2017).

69. Rasmussen, P. B. & Staller, P. The KDM5 family of histone demethylases as targets in oncology drug discovery. *Epigenomics.* **6**, 277–286 (2014).

70. Dimitrova, E., Turberfield, A. H. & Klose, R. J. Histone demethylases in chromatin biology and beyond. *EMBO Rep.* **16**, 1620–1639 (2015).

71. Biggar, K.K. *et al.* Proteome-wide Prediction of Lysine Methylation Leads to Identification of H2BK43 Methylation and Outlines the Potential Methyllysine Proteome. *Cell Rep.* **32**, 107896 (2020).

72. Kaniskan, H. Ü., Martini, M. L. & Jin, J. Inhibitors of Protein Methyltransferases and Demethylases. *Chem. Rev.* **118**, 989–1068 (2018).

73. Tumber, A. *et al.* Potent and Selective KDM5 Inhibitor Stops Cellular Demethylation of H3K4me3 at Transcription Start Sites and Proliferation of MM1S Myeloma Cells. *Cell Chem. Biol.* **24**, 371–380 (2017).

74. Petronikolou, N., Longbotham, J. E. & Fujimori, D. G. Extended Recognition of the Histone H3 Tail by Histone Demethylase KDM5A. *Biochemistry.* **59**, 647–651 (2020).

75. Jain, K. *et al.* Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. *Epigenetics and Chromatin* **13**, 1-11 (2020).

76. Liu, X. & Secombe, J. The histone demethylase KDM5 activates gene expression by recognizing chromatin context through its PHD reader motif. *Cell Rep.* **13**, 2219-2231 (2015).

77. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 1-6 (2011).

78. Shokri, G., Doudi, S., Fathi-Roudsari, M., Kouhkan, F. & Sanati, M.-H. Targeting histone demethylases KDM5A and KDM5B in AML cancer cells: A comparative view. *Leuk. Res.* **68**, 105–111 (2018).

79. Klose, R. J. *et al.* The Retinoblastoma Binding Protein RBP2 Is an H3K4 Demethylase. *Cell.* **128**, 889–900 (2007).

80. Cao, J. *et al.* Histone Demethylase RBP2 Is Critical for Breast Cancer Progression and Metastasis. *Cell Rep.* **6**, 868–877 (2014).

81. Zeng, J. *et al.* The Histone Demethylase RBP2 Is Overexpressed in Gastric Cancer and Its Inhibition Triggers Senescence of Cancer Cells. *Gastroenterology* **138**, 981–992 (2010).

82. van Zutven, L. J. C. M. *et al.* Identification ofNUP98 abnormalities in acute leukemia:JARID1A (12p13) as a new partner gene. *Genes, Chromosom.*

*Cancer* **45**, 437–446 (2006).

83. Secombe, J., Li, L., Carlos, L. & Eisenman, R. N. The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. *Genes Dev.* **21**, 537–551 (2007).

84. Cornett, E. M. *et al.* Substrate Specificity Profiling of Histone-Modifying Enzymes by Peptide Microarray. *Methods in Enzymology.* **574**, 31–52 ( 2016).

85. Tsukada, Y. ichi & Nakayama, K. I. In vitro histone demethylase assay. *Cold Spring Harb. Protoc.* **5**, 1-7 (2010).

86. Ullman, E. F. *et al.* Luminescent oxygen channeling immunoassay: Measurement of particle binding kinetics by chemiluminescence. *Proc. Natl. Acad. Sci.* **91**, 5426–5430 (1994).

87. Gale, M. & Yan, Q. High-throughput screening to identify inhibitors of lysine demethylases. *Epigenomics.* **7**, 57–65 (2015).

88. Garske, A. L. *et al.* Combinatorial profiling of chromatin-binding modules reveals multi-site discrimination HHS Public Access Author manuscript. *Nat. Chem. Biol.* **6**, 283–290 (2010).

89. Kawamura, A. *et al.* Development of homogeneous luminescence assays for histone demethylase catalysis and binding. *Anal. Biochem.* **404**, 86–93 (2010).

90. Rothbart, S. B., Krajewski, K., Strahl, B. D. & Fuchs, S. M. Peptide microarrays to interrogate the ‘histone code’. *Methods in Enzymology.* **512**, 107–135 (2012).

91. Glickman, J. F. Assay Development for Protein Kinase Enzymes. *Assay Guidance Manual (Eli Lilly & Company and the National Center for Advancing Translational Sciences)* (2004).

92. Krishnan, S. & Trievel, R. C. Purification, Biochemical Analysis, and Structure Determination of JmjC Lysine Demethylases. *Methods in Enzymology.* **573**, 279–301 (2016).

93. Horton, J. R. *et al.* Structural Basis for KDM5A Histone Lysine Demethylase Inhibition by Diverse Compounds. *Cell Chem. Biol.* **23**, 769–781 (2016).

94. Hauser, A. T. *et al.* Discovery of Histone Demethylase Inhibitors. *Epigenetic Technological Applications* 397–424 (Elsevier Inc., 2015).

95. Szymczak, L. C., Kuo, H. Y. & Mrksich, M. Peptide Arrays: Development and Application. *Analytical Chemistry.* **90**, 266–282 (2018).

96. Katz, C. *et al.* Studying protein–protein interactions using peptide arrays. *Chem. Soc. Rev.* **40**, 2131–2145 (2011).

97. Bagowski, C. P., Bruins, W. & Te Velthuis, A. J. The Nature of Protein Domain Evolution: Shaping the Interaction Network. *Curr. Genomics* **11**, 368–376 (2010).

98. Plant, M. *et al.* Screening for lysine-specific demethylase-1 inhibitors using a label-free high-throughput mass spectrometry assay. *Anal. Biochem.* **419**, 217–227 (2011).

99. Hutchinson, S. E. *et al.* Enabling Lead Discovery for Histone Lysine Demethylases by High-Throughput RapidFire Mass Spectrometry. *J.*

*Biomol. Screen.* **17**, 39–48 (2012).

100. Leveridge, M. *et al.* Demonstrating enhanced throughput of RapidFire mass spectrometry through multiplexing using the JmjD2d demethylase as a Model System. *J. Biomol. Screen.* **19**, 278–286 (2014).
101. Ohguchi, H. *et al.* Lysine Demethylase 5A is Required for MYC Driven Transcription in Multiple Myeloma. *Blood cancer Discov.* **2**, 370–387 (2021).
102. Plch, J., Hrabeta, J. & Eckschlager, T. KDM5 demethylases and their role in cancer cell chemoresistance. *Int. J. Cancer* **144**, 221–231 (2019).
103. Oser, M. G. *et al.* The KDM5A/RBP2 histone demethylase represses NOTCH signaling to sustain neuroendocrine differentiation and promote small cell lung cancer tumorigenesis. *Genes Dev.* **33**, 1718-1738 (2019).
104. Hu, D. *et al.* KDM5A Regulates a Translational Program that Controls p53 Protein Expression. *IScience* **9**, 84–100 (2018).
105. Compton, B. J. & Purdy, W. C. Fluoral-P, a member of a selective family of reagents for aldehydes. *Anal. Chim. Acta.* **119**, 349–357 (1980).
106. Li, Q., Sritharathikhum, P., Oshima, M. & Motomizu, S. Development of novel detection reagent for simple and sensitive determination of trace amounts of formaldehyde and its application to flow injection spectrophotometric analysis. *Anal. Chim. Acta* **612**, 165–172 (2008).
107. Yoshikawa, K., Oshima, Y., Inagaki, A. & Sakuragawa, A. Determination of Formaldehyde in Water Samples by High-Performance Liquid Chromatography with Methyl Acetoacetate Derivatization. *Environ.*

*Contam. Toxicol.* **101**, 672–677 (2018).

108. Gholami, A., Mohsenikia, A. & Masoum, S. Determination of Very Low Level of Free Formaldehyde in Liquid Detergents and Cosmetic Products Using Photoluminescence Method. *J. Anal. Methods Chem.* **2016**, 1-8 (2016).
109. Luo, L. *et al.* An assay for Fe(II)/2-oxoglutarate-dependent dioxygenases by enzyme-coupled detection of succinate formation. *Anal. Biochem.* **353**, 69–74 (2006).
110. Rathert, P., Zhang, X., Freund, C., Cheng, X. & Jeltsch, A. Analysis of the substrate specificity of the Dim-5 histone lysine methyltransferase using peptide arrays. *Chem. Biol.* **15**, 5–11 (2008).
111. Rathert, P. *et al.* Protein lysine methyltransferase G9a acts on non-histone targets. *Nat. Chem. Biol.* **4**, 344–346 (2008).
112. Dhayalan, A., Kudithipudi, S., Rathert, P. & Jeltsch, A. Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chem. Biol.* **18**, 111–120 (2011).
113. Kudithipudi, S., Lungu, C., Rathert, P., Happel, N. & Jeltsch, A. Substrate Specificity Analysis and Novel Substrates of the Protein Lysine Methyltransferase NSD1. *Chem Biol.* **21**, 226-237 (2014).
114. Lanouette, S. *et al.* Discovery of substrates for a SET domain lysine methyltransferase predicted by multistate computational protein design. *Structure* **23**, 206–215 (2015).
115. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone

lysine methylations. *Exp. Mol. Med.* **49**, e324–e324 (2017).

116. Rose, N. R. *et al.* The Plant Growth Regulator Daminozide is a Selective Inhibitor of the Human KDM2/7 Histone Demethylases. *J. Med. Chem.* **55**, 6639-6643 (2012).
117. Krishnan, S., Collazo, E., Ortiz-Tello, P. A. & Trievel, R. C. Purification and assay protocols for obtaining highly active Jumonji C demethylases. *Anal. Biochem.* **420**, 48–53 (2012).
118. Hirsilä, M., Koivunen, P., Günzler, V., Kivirikko, K. I. & Myllyharju, J. Characterization of the Human Prolyl 4-Hydroxylases That Modify the Hypoxia-inducible Factor. *J. Biol. Chem.* **278**, 30772–30780 (2003).
119. Kivirikko, K. I. & Myllylä, R. Posttranslational enzymes in the biosynthesis of collagen: Intracellular enzymes. *Methods Enzymol.* **82**, 245–304 (1982).
120. Yang, Z. *et al.* Structure of the arabidopsis JMJ14-H3K4me3 complex provides insight into the substrate specificity of KDM5 subfamily histone demethylases. *Plant Cell* **30**, 167–177 (2018).
121. McKenna, M., Balasuriya, N., Zhong, S., Li, S. S. C. & O'Donoghue, P. Phospho-Form Specific Substrates of Protein Kinase B (AKT1). *Front. Bioeng. Biotechnol.* **8**, 619252 (2021).
122. Wei, R. *et al.* Interactome Mapping Uncovers a General Role for Numb in Protein Kinase Regulation. *Mol. Cell. Proteomics* **17**, 2216–2228 (2018).
123. Alves, J., Vidugiris, G., Goueli, S. A. & Zegzouti, H. Bioluminescent High-Throughput Succinate Detection Method for Monitoring the Activity of JMJC Histone Demethylases and Fe(II)/2-Oxoglutarate-Dependent

Dioxygenases. *SLAS Discov.* **23**, 242–254 (2018).

124. Hoekstra, M. & Biggar, K. K. Identification of in vitro JMJD lysine demethylase candidate substrates via systematic determination of substrate preference. *Anal. Biochem.* **633**, 114429 (2021).
125. Topcu, E. & Biggar, K. K. PeSA: A software tool for peptide specificity analysis. *Comput. Biol. Chem.* **83**, 107145 (2019).
126. Liu, H. *et al.* A method for systematic mapping of protein lysine methylation identifies new functions for HP1β in DNA damage repair. *Mol. Cell* **50**, 723-735 (2013).
127. Schuhmacher, M. K., Kudithipudi, S., Kusevic, D., Weirich, S. & Jeltsch, A. Activity and specificity of the human SUV39H2 protein lysine methyltransferase. *Biochim. Biophys. Acta* **1849**, 55–63 (2015).
128. Yu, W. *et al.* A scintillation proximity assay for histone demethylases. *Anal. Biochem.* **463**, 54–60 (2014).
129. Goda, S., Isagawa, T., Chikaoka, Y., Kawamura, T. & Aburatani, H. Control of Histone H3 Lysine 9 (H3K9) Methylation State via Cooperative Two-step Demethylation by Jumonji Domain Containing 1A (JMJD1A) Homodimer. *J. Biol. Chem.* **288**, 36948–36956 (2013).
130. Hornbeck, P. V. *et al.* PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **40**, D261-D270 (2012).
131. Shokri, G., Doudi, S., Fathi-Roudsari, M., Kouhkan, F. & Sanati, M-H.

Targeting histone demethylases KDM5A and KDM5B in AML cancer cells: A comparative view. *Leuk Res.* **68**, 105-111 (2018).

132. Völkel, P. & Angrand, P. O. The control of histone lysine methylation in epigenetic regulation. *Biochimie.* **89**, 1–20 (2007).
133. Carlson, S.M., Moore, K. E., Green, E.M., Martin, G.M., & Gozani, O. Proteome-wide enrichment of proteins modified by lysine methylation. *Nat. Protoc.* **9**, 37–50 (2014).
134. Ning, Z. *et al.* A charge-suppressing strategy for probing protein methylation. *Chem. Commun.* **52**, 5474–5477 (2016).
135. Cornett, E. M. *et al.* A functional proteomics platform to reveal the sequence determinants of lysine methyltransferase substrate selectivity. *Sci. Adv.* **4**, eaav2623 (2018).
136. Trojer, P. & Reinberg, D. A gateway to study protein lysine methylation. *Nat. Chem. Biol.* **4**, 332–334 (2008).
137. Kim, T.-D., Shin, S. & Janknecht, R. Repression of Smad3 activity by histone demethylase SMCX/JARID1C. *Biochem. Biophys. Res. Commun.* **366**, 563–567 (2008).
138. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
139. Couture, J. F., Collazo, E., Ortiz-Tello, P. A., Brunzelle, J. S. & Trievel, R. C. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **14**, 689–695 (2007).

140. Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V. & Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512-D520 (2015).

141. Chu, F. *et al.* Mapping Post-translational Modifications of the Histone Variant MacroH2A1 Using Tandem Mass Spectrometry. *Mol. Cell Proteomics.* **5**, 194-203 (2006).

142. Nair, S. S., Li, D. Q. & Kumar, R. A. Core Chromatin Remodeling Factor Instructs Global Chromatin Signaling through Multivalent Reading of Nucleosome Codes. *Mol. Cell.* **49**, 704–718 (2013).

143. Vinogradova, M. *et al.* An inhibitor of KDM5 demethylases reduces survival of drug-tolerant cancer cells. *Nat. Chem. Biol.* **12**, 531–538 (2016).

144. Chang, S., Yim, S. & Park, H. The cancer driver genes IDH1/2, JARID1C/ KDM5C, and UTX/ KDM6A: crosstalk between histone demethylation and hypoxic reprogramming in cancer metabolism. *Exp. Mol. Med.* **51**, 1-17 (2019).

145. D, S. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

146. Sen, N., Gui, B., Kumar, R. & Kumar, R. Role of MTA1 in Cancer Progression and Metastasis. *Cancer Metastasis Rev.* **33**, 879–889 (2014).

147. Nishibuchi, G. *et al.* Physical and Functional Interactions between the Histone H3K4 Demethylase KDM5A and the Nucleosome Remodeling and Deacetylase (NuRD) Complex. *J. Biol. Chem.* **289**, 28956–28970 (2014).

148. Woo, Y. H. & Li, W. H. Evolutionary Conservation of Histone Modifications in Mammals. *Mol. Biol. Evol.* **29**, 1757–1767 (2012).

149. Hendrich, B. & Bickmore, W. Human diseases with underlying defects in chromatin structure and modification. *Hum. Mol. Genet.* **10**, 2233–2242 (2001).

150. Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–357 (2012).

151. Sims Iii, R. J., Nishioka, K. & Reinberg, D. Histone lysine methylation: a signature for chromatin function. *Trends Genet.* **19**, 629–639 (2003).

152. Wu, L. *et al.* KDM5 histone demethylases repress immune response via suppression of STING. *PLOS Biol.* **16**, 1-29 (2018).

153. Secombe, J. & Eisenman, R. N. The Function and Regulation of the JARID1 Family of Histone H3 Lysine 4 Demethylases: the Myc Connection. *Cell Cycle.* **6**, 1324-1328 (2007).

154. Zhang, K. & Dent, S. Y. R. Histone modifying enzymes and cancer: going beyond histones. *J. Cell. Biochem.* **96**, 1137–1148 (2005).

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

156. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).

157. The Pandas development team. Pandas. (2020).

158. Mckinney, W. Data Structures for Statistical Computing in Python. (2010).

159. Van Rossum, G. *et al.* Python 3 Reference Manual. *Nat.* **585**, 357–362 (2009).

160. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods.* **17**, 261–272 (2020).

161. Murtagh, F. & Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion. *J. Classif.* **31**, 274–295 (2014).

162. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

163. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597-W600 (2013).

164. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**, W695-W696 (2010).

165. Chopra, A., Willmore, W. G. & Biggar, K. K. Insights into a Cancer-Target Demethylase: Substrate Prediction through Systematic Specificity Analysis for KDM3A. *Biomol.* **12**, 641-654 (2022).

166. Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13**, 297–311 (2012).

167. Ponnaluri, V. K. C., Vavilala, D. T., Putty, S., Gutheil, W. G. & Mukherji, M. Identification of non-histone substrates for JMJD2A-C histone demethylases. *Biochem. Biophys. Res. Commun.* **390**, 280–284 (2009).

168. Lu, T. *et al.* Regulation of NF-κB by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proc. Natl. Acad. Sci.* **107**, 46–51 (2010).

169. Zhao, L. *et al.* JMJD3 promotes SAHF formation in senescent WI38 cells by triggering an interplay between demethylation and phosphorylation of RB protein. *Cell Death Differ.* **22**, 1630–1640 (2015).

170. Lu, L. *et al.* Kdm2a/b Lysine Demethylases Regulate Canonical Wnt Signaling by Modulating the Stability of Nuclear β-Catenin. *Dev. Cell.* **33**, 660–674 (2015).

171. Jung, E. S. *et al.* Jmjd2C increases MyoD transcriptional activity through inhibiting G9a-dependent MyoD degradation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1849**, 1081–1094 (2015).

172. Brier, A. S. B. *et al.* The KDM5 family is required for activation of pro-proliferative cell cycle genes during adipocyte differentiation. *Nucleic Acids Res.* **45**, 1743–1759 (2017).

173. Takeda, H. *et al.* Comparative analysis of human SRC-family kinase substrate specificity in vitro. *J. Proteome Res.* **9**, 5982–5993 (2010).

174. Agulnik, A. I., Longepied, G., Ty, M. T., Bishop, C. E. & Mitchell, M. Mouse H-Y encoding Smcy gene and its X chromosomal homolog Smcx. *Mamm. Genome.* **10**, 926–929 (1999).

175. Shi, X. *et al.* Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol. Cell.* **27**, 636–46 (2007).

176. West, L. E. *et al.* The MBT Repeats of L3MBTL1 Link SET8-mediated p53 Methylation at Lysine 382 to Target Gene Repression. *J. Biol. Chem.* **285**, 37725–37732 (2010).

177. West, L. E. & Gozani, O. Regulation of p53 function by lysine methylation.

*Epigenomics.* **3**, 361–369 (2011).

178. Cuella-Martin, R. *et al.* 53BP1 Integrates DNA Repair and p53-Dependent Cell Fate Decisions via Distinct Mechanisms. *Mol. Cell.* **64**, 51–64 (2016).
179. O'Connor, P. *et al.* Characterization of the p53 Tumor Suppressor Pathway in Cell Lines of the National Cancer Institute Anticancer Drug Screen and Correlations with the Growth-Inhibitory Potency of 123 Anticancer Agents. *Cancer Res.* **57**, 4285–4300 (1997).
180. Huang, J. *et al.* Repression of p53 activity by Smyd2-mediated methylation. *Nat.* **444**, 629–632 (2006).
181. Guo, Y., Wu, Y., Li, N. & Wang, Z. Up-regulation of miRNA-151-3p enhanced the neuroprotective effect of dexmedetomidine against β-amyloid by targeting DAPK-1 and TP53. *Exp. Mol. Pathol.* **118**, 104587 (2021).
182. Hamard, P. J., Lukin, D. J. & Manfredi, J. J. p53 Basic C Terminus Regulates p53 Functions through DNA Binding Modulation of Subset of Target Genes. *J. Biol. Chem.* **287**, 22397–22407 (2012).
183. Schultz, L. B., Chehab, N. H., Malikzay, A. & Halazonetis, T. D. P53 Binding Protein 1 (53bp1) Is an Early Participant in the Cellular Response to DNA Double-Strand Breaks. *J. Cell Biol.* **151**, 1381–1390 (2000).
184. Thorn, C. F. *et al.* Doxorubicin pathways: pharmacodynamics and adverse effects. *Pharmacogenet. Genomics.* **21**, 440–446 (2011).
185. Sasaki, Y. *et al.* P53 negatively regulates the hepatoma growth factor HDGF. *Cancer Res.* **71**, 7038–7047 (2011).

186. Tolondn, A. C. & Haas, W. Quantitative Proteomics Using Reductive Dimethylation for Stable Isotope Labeling. *J. Vis. Exp.* **89**, 1–7 (2014).

187. Bhattacharya, S., Takada, S. & Jacobson, R. H. Structural analysis and dimerization potential of the human TAF5 subunit of TFIID. *Proc. Natl. Acad. Sci.* **104**, 1189–1194 (2007).

188. Li, H. H., Li, A. G., Sheppard, H. M. & Liu, X. Phosphorylation on Thr-55 by TAF1 Mediates Degradation of p53: A Role for TAF1 in Cell G1 Progression. *Mol. Cell.* **13**, 867–878 (2004).

189. Vassilev, L. T. *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science.* **303**, 844–848 (2004).

190. Zilfou, J. T. & Lowe, S. W. Tumor suppressive functions of p53. *Cold Spring Harb. Perspect. Biol.* **1**, a001883 (2009).

191. Kim, H. *et al.* p53 Requires an Intact C-Terminal Domain for DNA Binding and Transactivation. *J. Mol. Biol.* **415**, 843–854 (2012).

192. Musselman, C. A., Lalonde, M. E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* **19**, 1218–1227 (2012).

193. Lu, H. & Levine, A. J. Human TAFII31 protein is a transcriptional coactivator of the p53 protein. *Proc. Natl. Acad. Sci.* **92**, 5154–5158 (1995).

194. Wu, Z., Huang, R. & Yuan, L. Crosstalk of intracellular post-translational modifications in cancer. *Arch. Biochem. Biophys.* **676**, 108138 (2019).

# Appendices

## Appendix A

### Detection of JMJD-type KDM demethylation *in vitro*

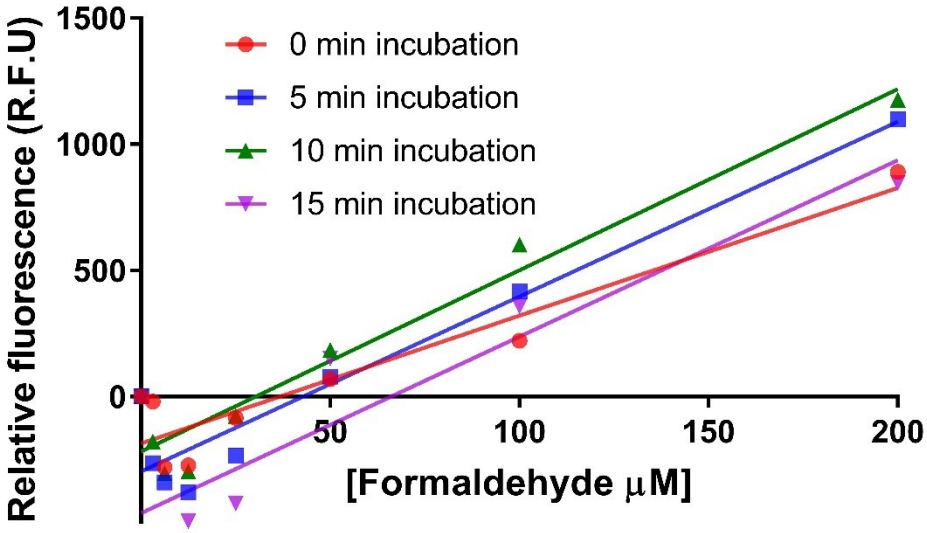



**Figure A1. Application of Fluoral-P based assay.** Detection of formaldehyde standard curve using Fluoral-P. Formaldehyde standards were prepared in a 1X KDM5A reaction buffer. Fluorescent averages were normalized by background signal. Results are mean (n=3).

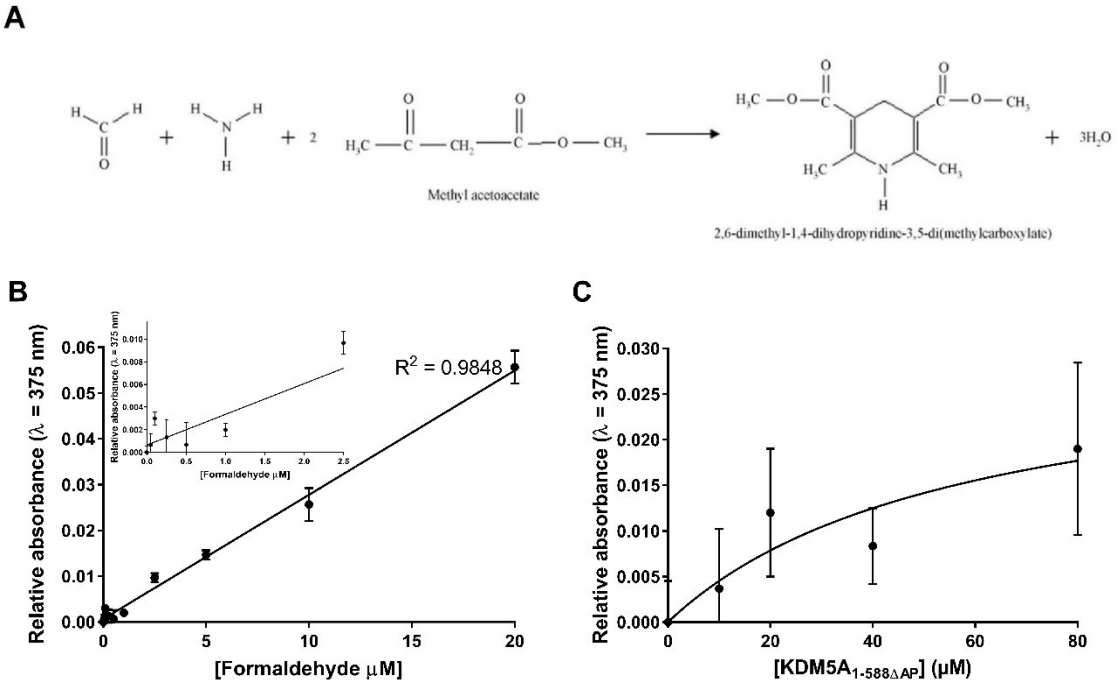



**Figure A2. Application of methyl acetoacetate-based assay.** ***A.*** Reaction mechanism formaldehyde detection. Formaldehyde production is coupled with 2-methyl acetoacetate in the presence of ammonia to form compound which absorbs light at λ = 375 nm. Reaction mechanism derived from Li et al., 2008[106]. ***B.*** Detection of formaldehyde standard curve. Formaldehyde standards were prepared in a 1X KDM5A reaction buffer. Absorbance averages were normalized by background signal. Results are mean ± S.E.M (n=3). ***C.*** Absorbance-based detection of $KDM5A_{1\text{-}588\Delta AP}$ demethylation of WT H3K4me3 peptide. Absorbance averages were normalized by background signal. Results are mean ± S.E.M (n=3).

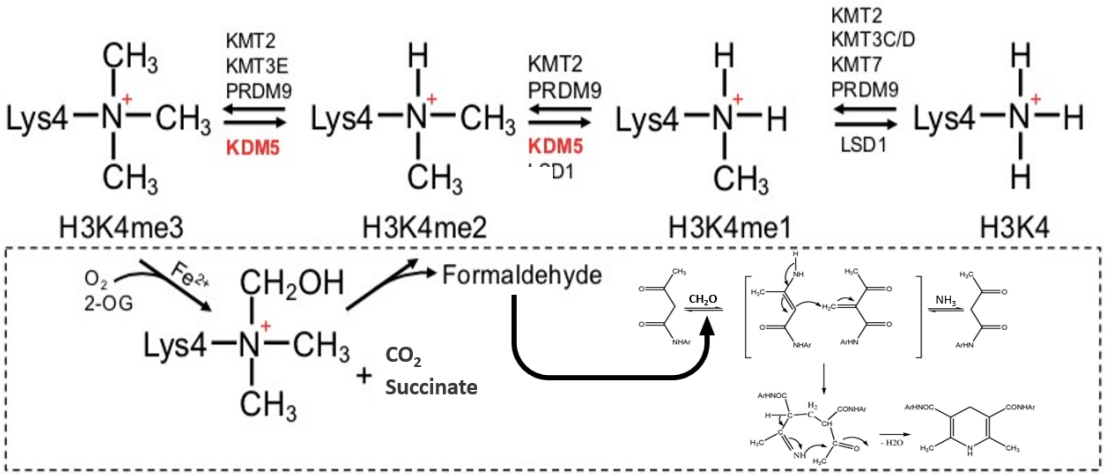



**Figure A3. 2-methyl acetoacetanilide assay mechanism**. Reaction mechanism for histone H3K4 demethylation. *In vitro* assay performed in Demethylase assay buffer (10 μM tri-methylated histone H3 peptide substrate, 100 μM ascorbate, 10 μM Fe (II) and 10 μM 2-oxoglutarate). Formaldehyde detection reagents added in a 1:2 ratio. Formaldehyde production is coupled with 2-methyl acetoacetanilide and ammonium acetate to form fluorescent compound which can be detected using excitation and emission wavelengths 260nm and 460nm, respectfully. Figure adapted from Gholami et al., 2016[108].

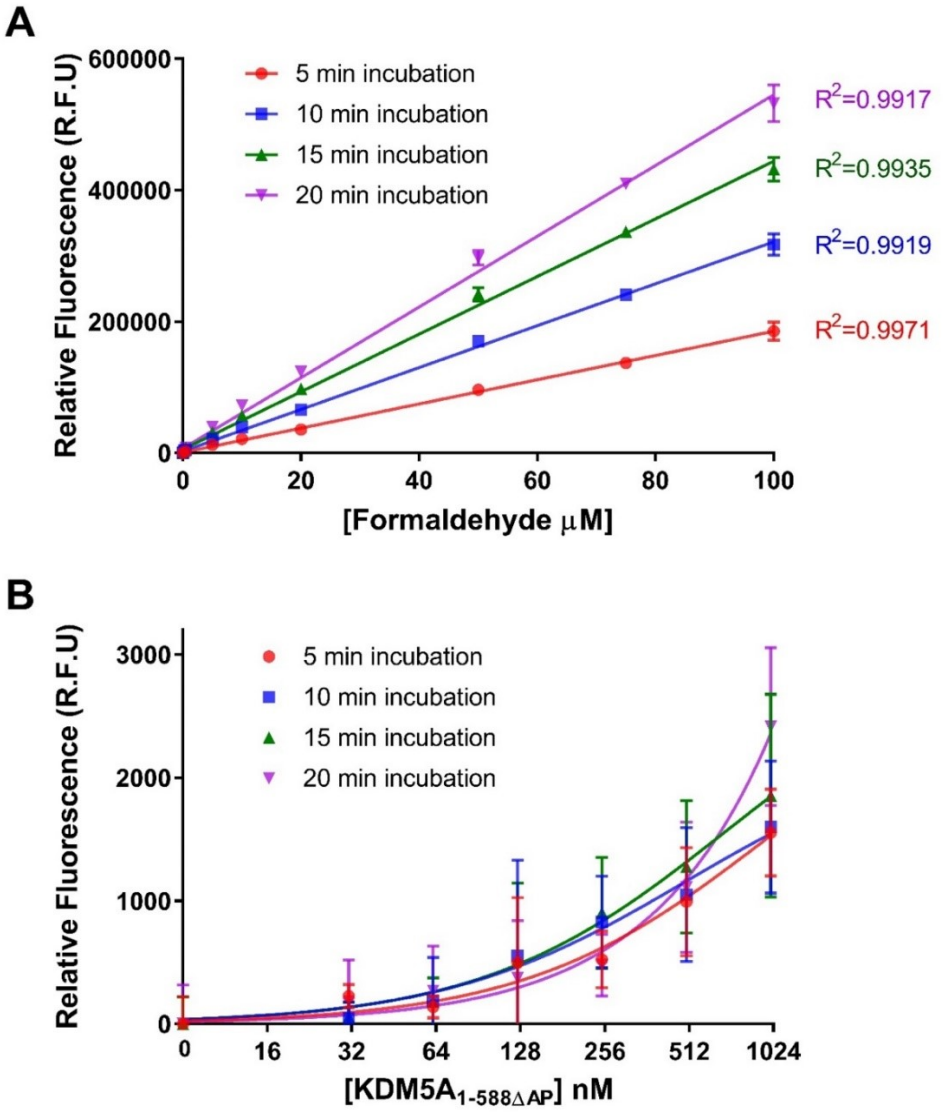



**Figure A4. Application of 2-methyl acetoacetanilide-based assay. *A.*** Detection of formaldehyde standard curve. Formaldehyde standards were prepared in a 1X KDM5A reaction buffer. Fluorescent averages were normalized by background signal. Results are mean ± S.E.M (n=3). ***B.*** Fluorescent-based detection of $KDM5A_{1\text{-}588\Delta AP}$ demethylation of WT H3K4me3 peptide. Fluorescent averages were normalized by background signal. Results are mean ± S.E.M (n=3).

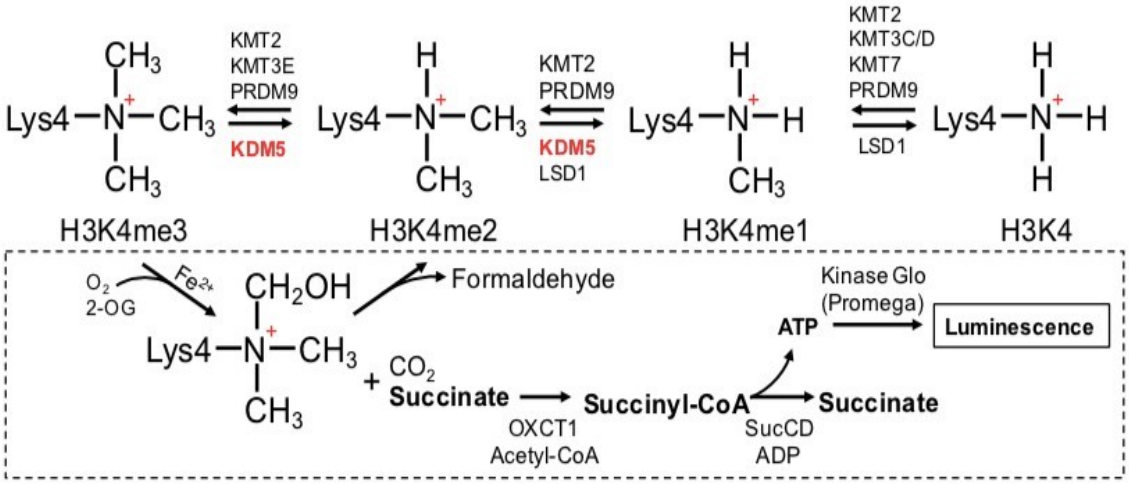



**Figure A5. Succinate-based assay mechanism.** Reaction mechanism for histone H3K4 demethylation. *In vitro* assay performed in Demethylase assay buffer (10 μM tri-methylated histone H3 peptide substrate, 100 μM ascorbate, 10 μM Fe (II) and 10 μM 2-oxoglutarate). Succinate detection reagents added in a 1:1:2 ratio. Reprinted with permission of *Analytical Biochemistry.*

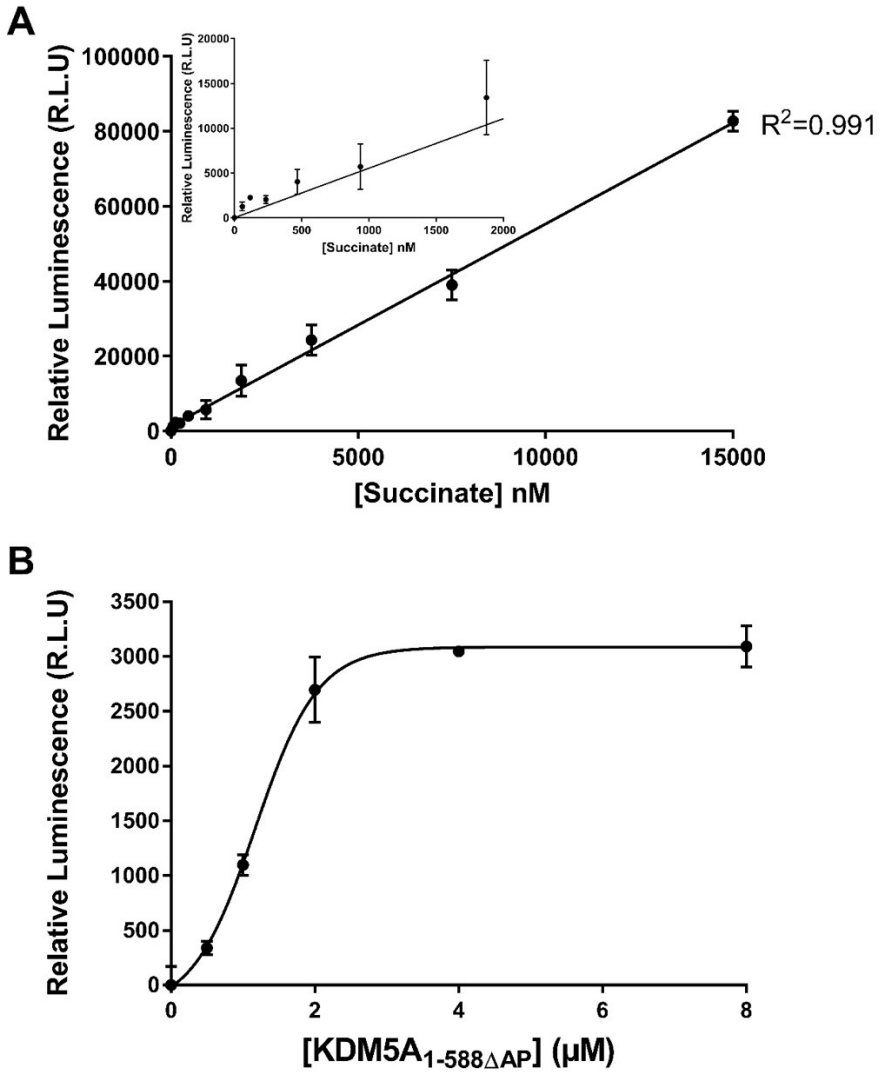



**Figure A6. Application of succinate-based assay.** ***A.*** Detection of succinate standard curve. Succinate standards were prepared in a 1X KDM5A reaction buffer. Luminescence averages were normalized by background signal. Results are mean ± S.E.M (n=3). ***B.*** Luminescence-based detection of $KDM5A_{1\text{-}588\Delta AP}$ demethylation of WT H3K4me3 peptide. Luminescence averages were normalized by background signal. Results are mean ± S.E.M (n=3).

**Table A1.** Reaction parameters for 2-methyl acetoacetanilide-based KDM demethylation assay.

| Reaction Variable | Optimized Value |
|---|---|
| KDM5A Reaction Buffer | 50 mM HEPES (pH 7.5)<br>10 μM H3K4me3 peptide substrate<br>10 μM alpha-ketoglutarate<br>10 μM Fe (II)<br>100 μM Ascorbic acid<br>1% DMSO |
| KDM5A reaction volume | 50 μL |
| pH | 7.5 |
| Temperature | 4°C during reaction set-up<br>Room temperature for reaction duration |
| Reaction time | 60 min |
| Detection mixture | 35mM 2-methyl acetoacetanilide<br>2M ammonium acetate |
| Enzyme Reaction : Detection Mixture Ratio | 1 : 2 (v/v) |

# Appendix B
## Quantifying peptide concentration

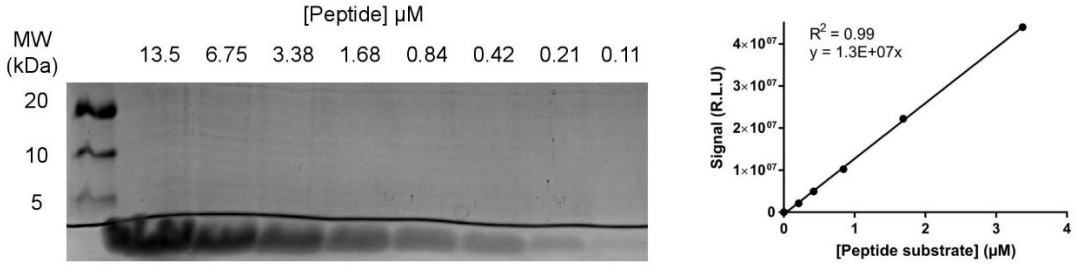



**Figure B1**. **Peptide substrate dilution curve**. Each mutated H3K4me3 peptide substrate was assessed by 20% SDS-PAGE to visualize synthesis product and concentration. Reprinted with permission of *Analytical Biochemistry.*

# Appendix C
## Proposed KDM5A-mediated regulation of MTA1-NuRD co-repressor complex

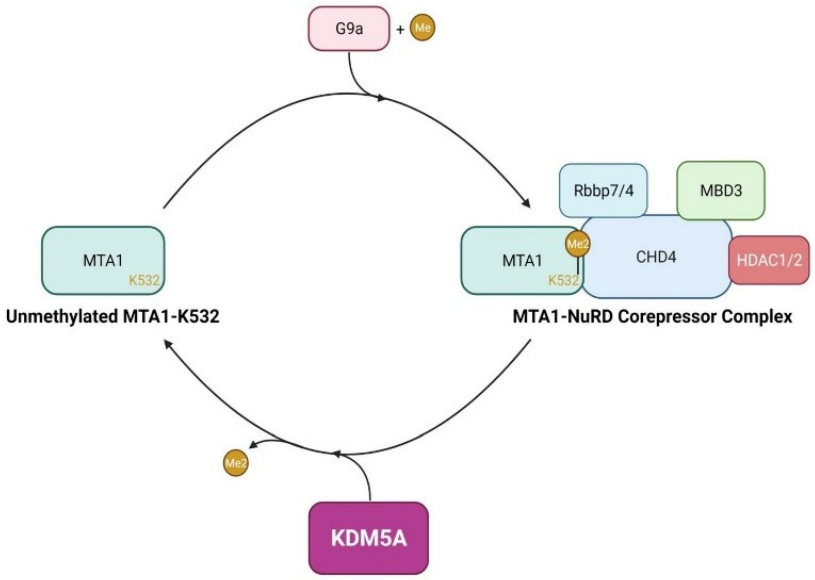



**Figure C1**. Proposed KDM5A-mediated regulation of MTA1-NuRD co-repressor complex dissociation and formation. Created with BioRender.com. Reprinted with permission of *Analytical Biochemistry.*

# Appendix D

## KDM5 *in vitro* activity towards PPS library assessment of activity to predicted ranked substrates

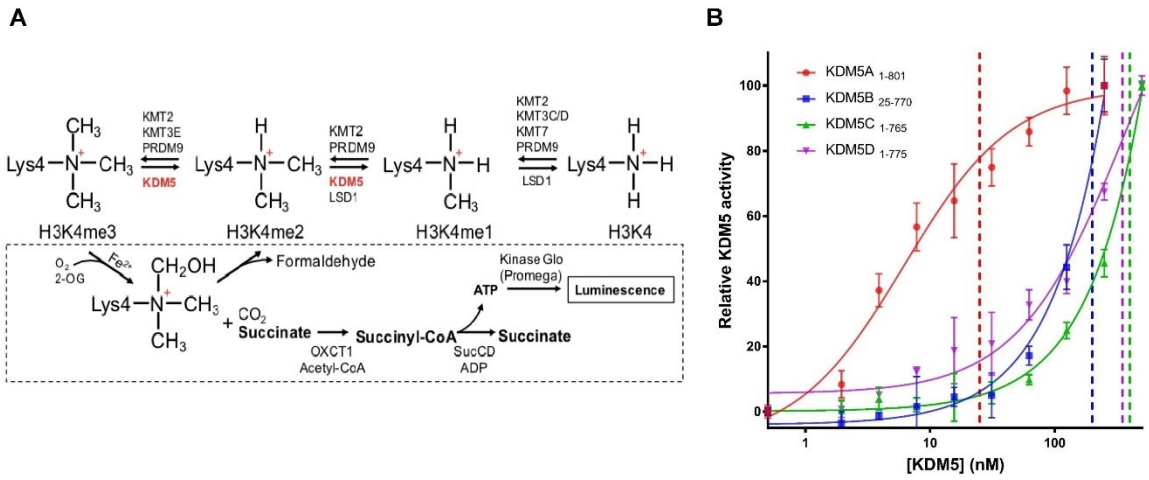



**Figure D1.** ***A***. *In vitro* assay performed in Demethylase assay buffer (10 μM tri-methylated histone H3 peptide substrate, 100 μM ascorbate, 10 μM Fe (II) and 10 μM 2-oxoglutarate). Succinate detection reagents added in a 1:1:2 ratio. ***B.*** Luminescence-based detection of KDM5A$_{1-801}$, KDM5B$_{25-770}$, KDM5C$_{1-765}$ and KDM5D$_{1-775}$ demethylation of WT H3K4me3 peptide. Luminescence averages were normalized by background signal. Results are mean ± S.E.M of n=3.

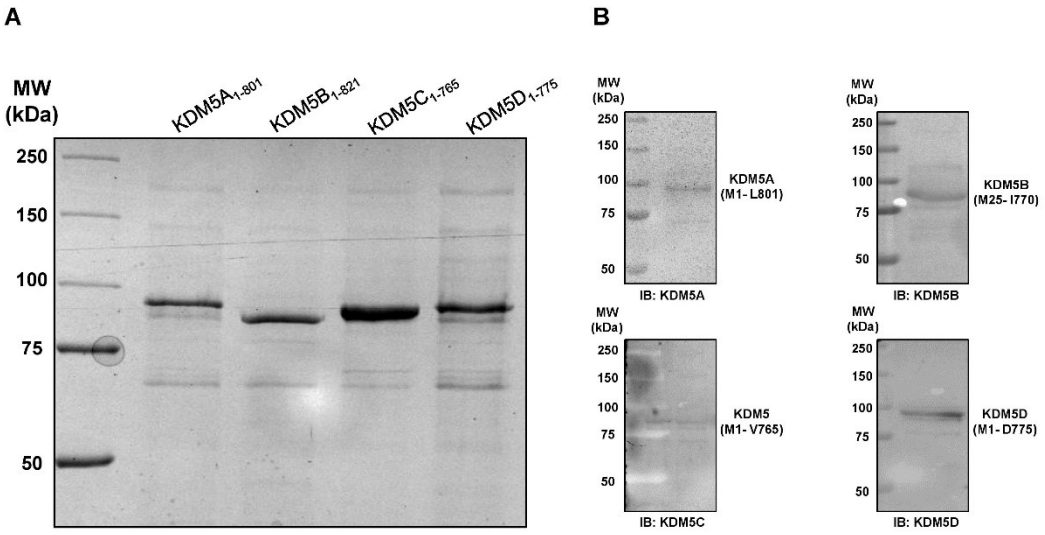



**Figure D2. Assessment of KDM5A/B/C/D purification *A.*** SDS-PAGE gel of recombinant KDM5 demethylases. ***B.*** Immunoblotting of recombinant KDM5 demethylases.

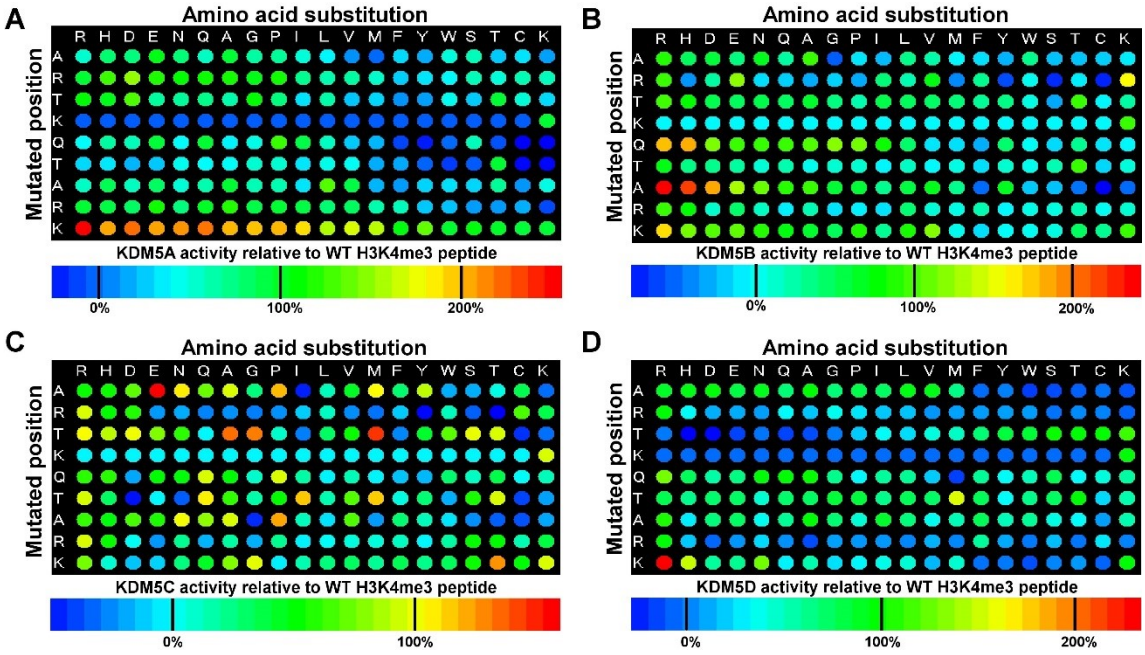



**Figure D3. KDM5A/B/C/D substrate specificity matrices. A.** Detection of relative KDM5A$_{1\text{-}801}$ demethylation activity with mutated Lys4 tri-methylated histone H3 peptide substrates. Data was normalized to the WT H3K4me3 peptide at each mutation position (considered as 1.00 for each position). Results are mean of n=3. Array image developed from PeSA[125]. ***B.*** Detection of relative KDM5B$_{25\text{-}770}$ demethylation activity with mutated Lys4 tri-methylated histone H3 peptide substrates. Data was normalized to the WT H3K4me3 peptide at each mutation position (considered as 1.00 for each position). Results are mean of n=3. Array image developed from PeSA[125]. ***C.*** Detection of relative KDM5C$_{1\text{-}765}$ demethylation activity with mutated Lys4 tri-methylated histone H3 peptide substrates. Data was normalized to the WT H3K4me3 peptide at each mutation position (considered as 1.00 for each position). Results are mean of n=3. Array image developed from PeSA[125]. ***D.*** Detection of relative KDM5D$_{1\text{-}775}$ demethylation activity with mutated Lys4 tri-methylated histone H3 peptide

substrates. Data was normalized to the WT H3K4me3 peptide at each mutation position (considered as 1.00 for each position). Results are mean of n=3. Array image developed from PeSA[125]. Figure caption continued from Figure D3 (page 225).

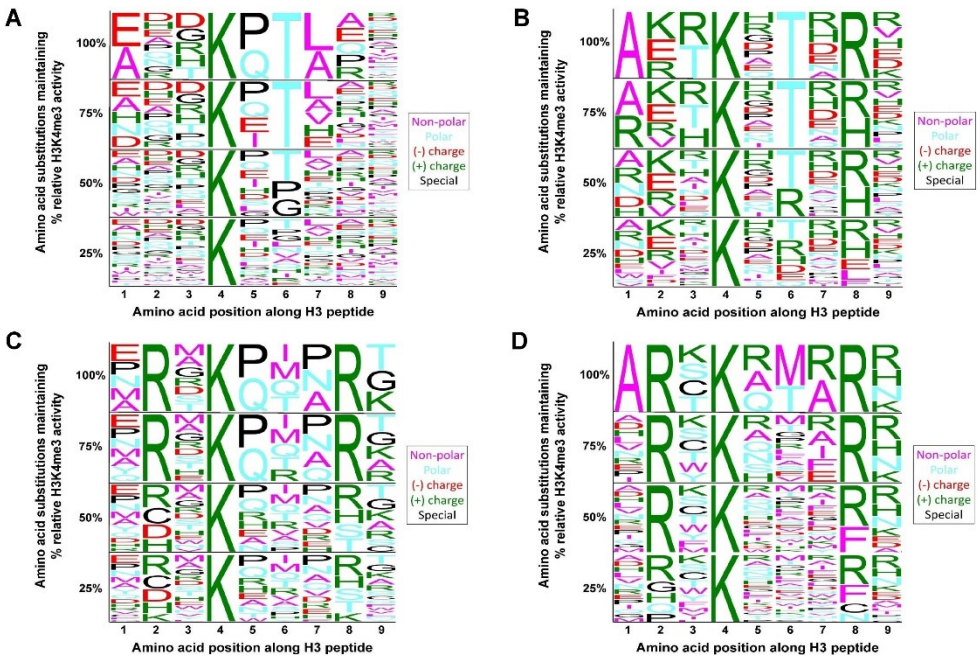



**Figure D4. KDM5A/B/C/D recognition motifs. *A*.** KDM5A$_{1\text{-}801}$ recognition motifs derived from relative KDM5A$_{1\text{-}801}$ demethylation activity towards PPS library. Each motif is reflective of amino acid substitutions that maintained 25%, 50%, 75% and 100% of KDM5A$_{1\text{-}801}$ activity relative to the H3K4me3 activity. Motif images developed from PeSA[125]. ***B.*** KDM5B$_{25\text{-}770}$ recognition motifs derived from relative KDM5B$_{25\text{-}770}$ demethylation activity towards PPS library. Each motif is reflective of amino acid substitutions that maintained 25%, 50%, 75% and 100% of KDM5B$_{25\text{-}770}$ activity relative to the H3K4me3 activity. Motif images developed from PeSA[125]. ***C.*** KDM5C$_{1\text{-}765}$ recognition motifs derived from relative KDM5C$_{1\text{-}765}$ demethylation activity towards PPS library. Each motif is reflective of amino acid substitutions that maintained 25%, 50%, 75% and 100% of KDM5C$_{1\text{-}765}$ activity relative to the H3K4me3 activity. Motif images developed from PeSA[125]. ***D.*** KDM5D$_{1\text{-}775}$ recognition motifs derived from relative KDM5D$_{1\text{-}775}$ demethylation

activity towards PPS library. Each motif is reflective of amino acid substitutions that maintained 25%, 50%, 75% and 100% of $KDM5D_{1-775}$ activity relative to the H3K4me3 activity. Motif images developed from PeSA[125]. Figure caption continued from Figure D4 (page 227).

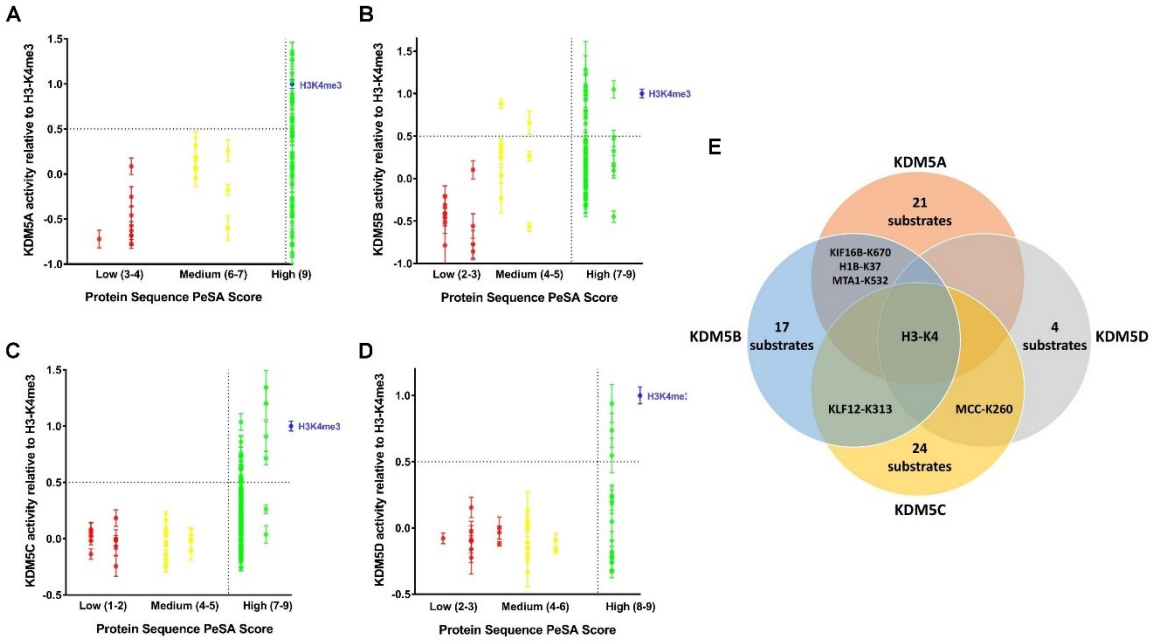



**Figure D5. KDM5A/B/C/D activity towards high, medium and low ranked substrates.** ***A.*** Detection of KDM5A$_{1\text{-}801}$ demethylation activity towards high, medium and low ranked synthesized substrate peptides. Results are mean ± std. error (n=3). ***B.*** Detection of KDM5B$_{25\text{-}770}$ demethylation activity towards high, medium and low ranked synthesized substrate peptides. Results are mean ± std. error (n=3). ***C.*** Detection of KDM5C$_{1\text{-}765}$ demethylation activity towards high, medium and low ranked synthesized substrate peptides. Results are mean ± std. error (n=3). ***D.*** Detection of KDM5D$_{1\text{-}775}$ demethylation activity towards high, medium and low ranked synthesized substrate peptides. Results are mean ± std. error (n=3).

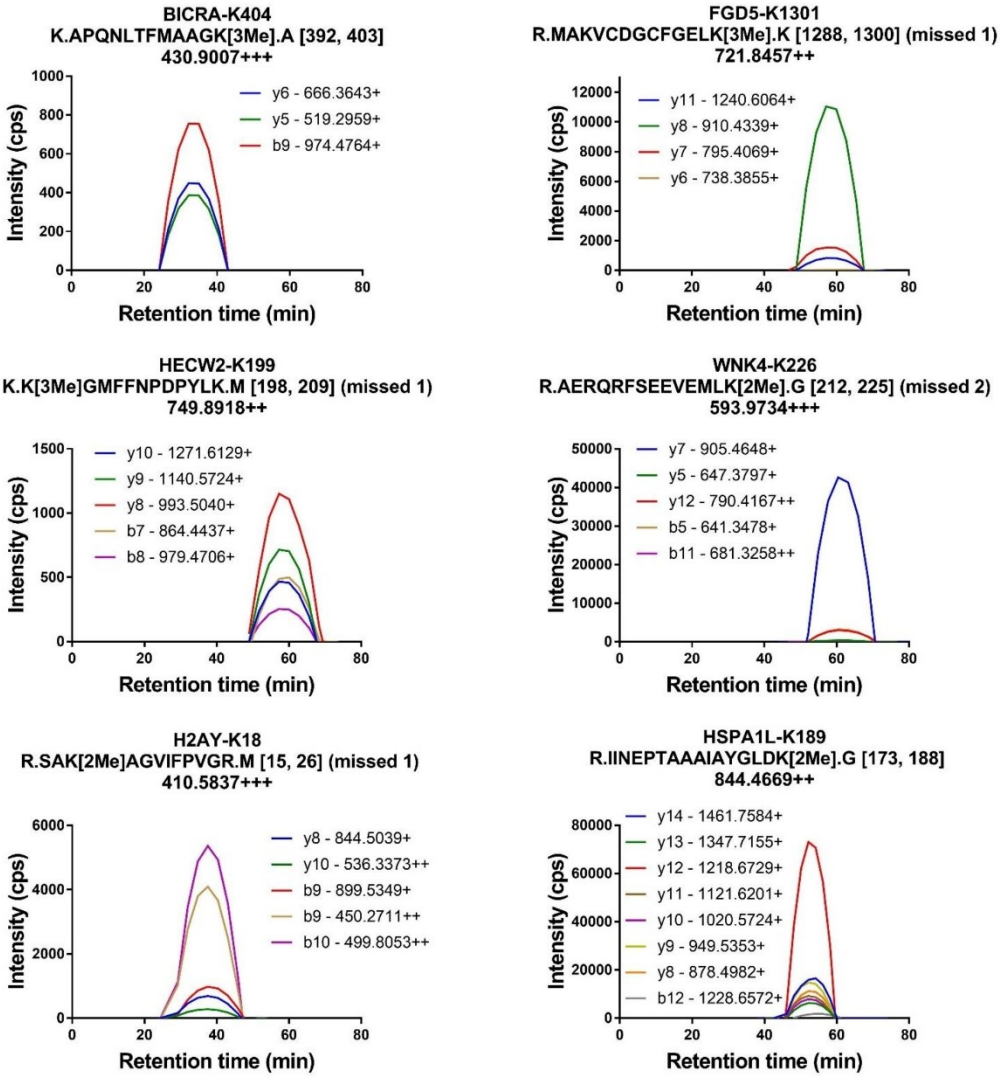



**Figure D6. Chromatographs of PRM-MS/MS analysis of CPI treated SKBR3 cells.**

# Appendix E

## Development of KDM5A recognition motifs and assessment of *in vitro* activity to methylated lysine sites in p53 C-terminal regulatory domain

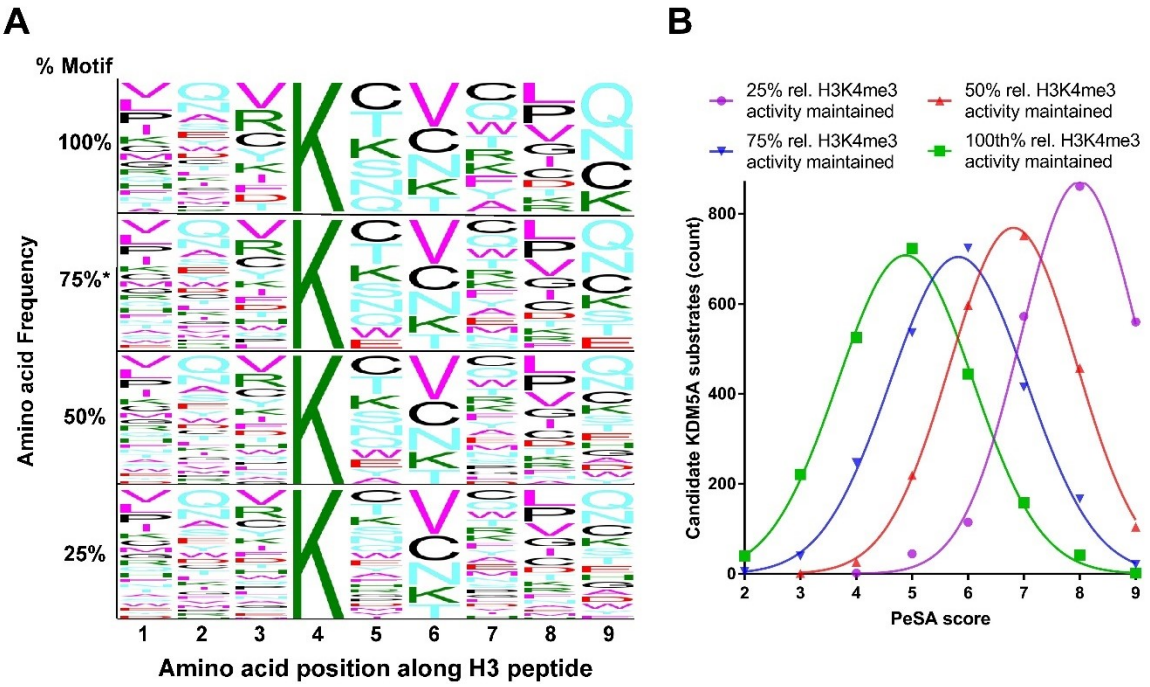



**Figure E1. Examining KDM5A$_{1\text{-}588\Delta AP}$ permissive and restrictive recognition motifs to predict substrates.** ***A.*** KDM5A recognition motifs derived from relative KDM5A$_{1\text{-}588\Delta AP}$ demethylation activity towards permutated peptide substrate library. Each motif is reflective of amino acid substitutions that maintained 25%, 50%, 75%, and 100% of KDM5A activity relative to H3K4me3. Motif images created using PeSA software[125]. ***B.*** KDM5A predicted substrates based on KDM5A recognition motif. A total of 2,155 known methylation sites (from the PhosphoSitePlus database[130]) were scored against each KDM5A recognition motif.

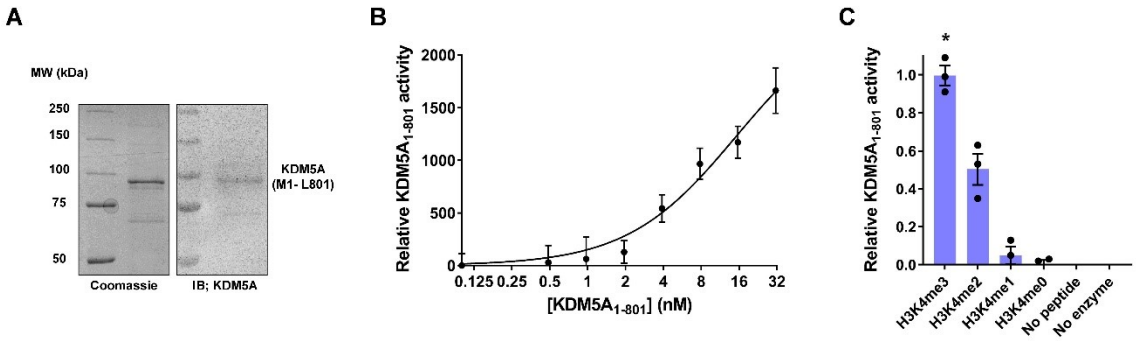



**Figure E2. KDM5A activity toward H3K4 peptides**. ***A.*** SDS-PAGE and western blot images showing purified KDM5A$_{1-801}$ used in this study. ***B.*** Luminescence-based detection of KDM5A$_{1-801}$ demethylation of WT H3K4me3 peptide. Luminescence averages were normalized by background signal. Results are mean ± std. error (n=3). ***C.*** Luminescence detection of KDM5A$_{1-801}$ demethylation of H3K4me0, H3K4me1, H3K4me2, and H3K4me3 peptides. Luminescence averages were normalized by background signal, and relatively quantified to H3K4me3 peptide luminescence (set to 1). Results are mean ± std. error (n=3). * indicates $p < 0.05$ from me0 data. Figures derived from Hoekstra et al., 2021[124].

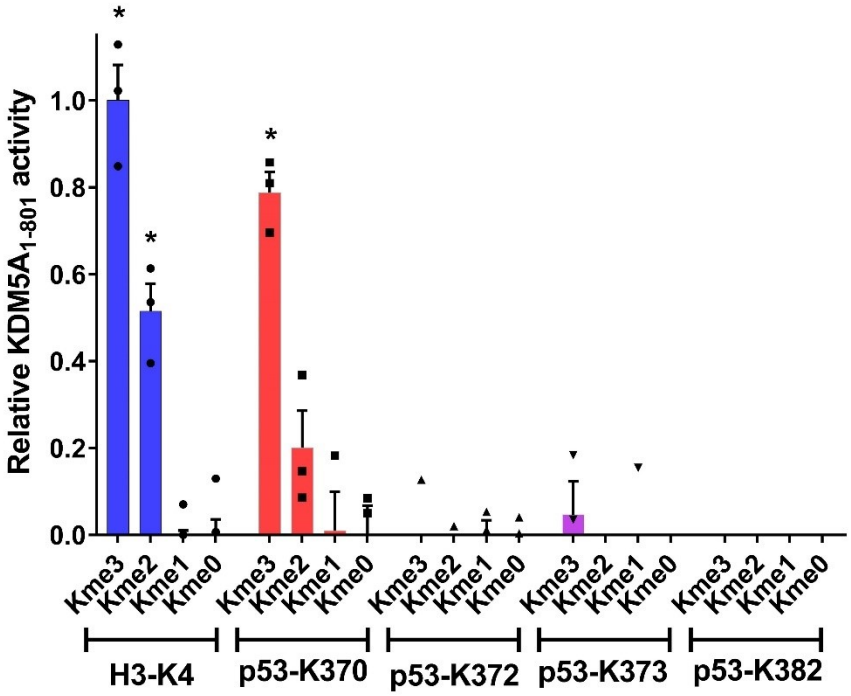



**Figure E3. Detection of KDM5A$_{1\text{-}801}$ demethylation activity to various methylated lysine sites in p53 C-terminal regulatory domain.** Detection of KDM5A$_{1\text{-}801}$ demethylation of H3-K4, p53-K370, p53-K372, p53-K373, p53-K382 peptides with Kme3, Kme2, Kme1, and Kme0 degrees of methylation. Luminescence averages were normalized by background signal, and relatively quantified to H3K4me3 peptide luminescence (set to 1). Results are mean ± std. error (n=3).

# Appendix F

## Development of WT and Mutant KDM5A overexpression plasmids and assessment of KDM5A-mediated demethylation in HEK 293 cells

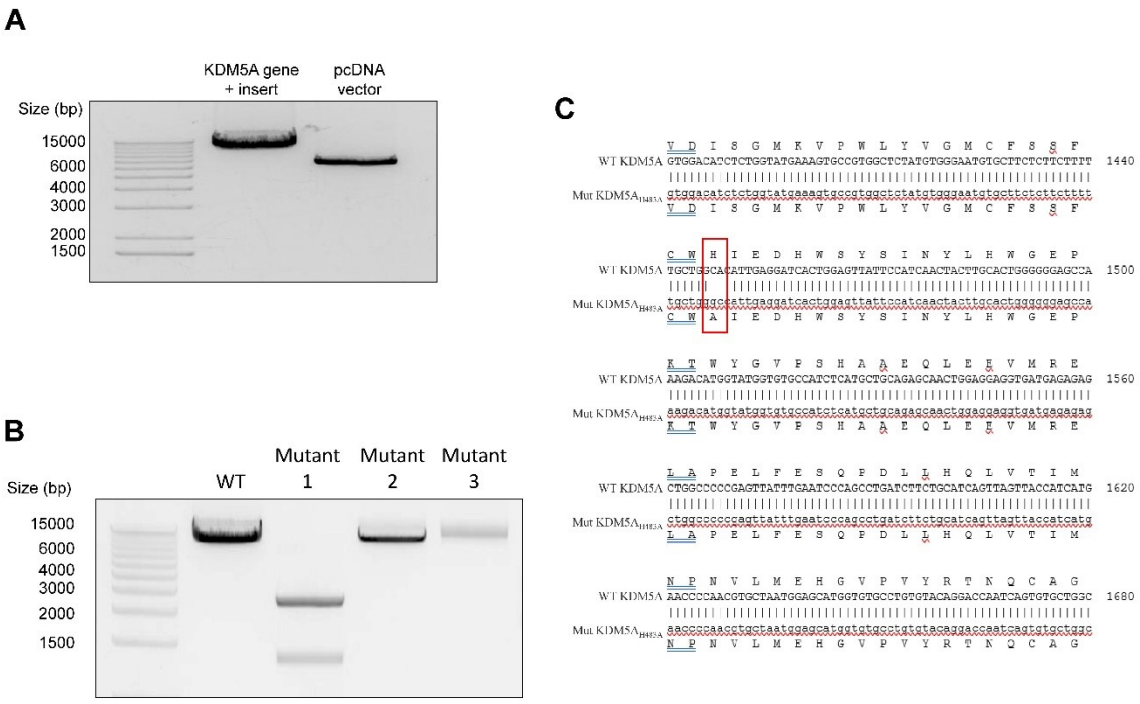



**Figure F1. Sequencing confirmation of mutant inactive pcDNA3.1/HA-KDM5A$_{H483A}$ overexpression plasmid.** ***A.*** Restriction digest to confirm successful cloning of pcDNA3.1/HA-KDM5A overexpression plasmid. ***B.*** Restriction digest to confirm successful site directed mutagenesis of pcDNA3.1/HA-KDM5A$_{H483A}$ overexpression plasmid. ***C.*** Sequencing results to confirm successful site directed mutagenesis of pcDNA3.1/HA-KDM5A$_{H483A}$ overexpression plasmid. Mutation location highlighted by red box.

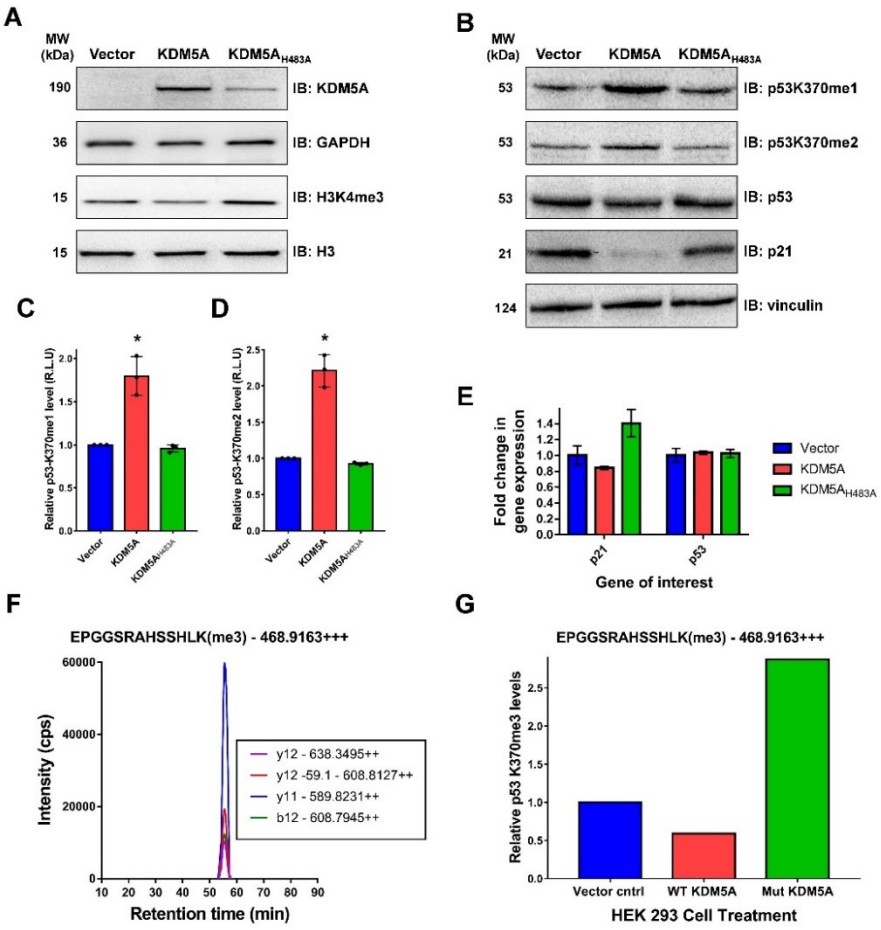



**Figure F2. Analysis of KDM5A mediated demethylation of p53-K370me3 substrate in HEK293.** ***A.*** Western blot analysis to confirm overexpression of WT KDM5A and Mut H483A KDM5A plasmids in HEK 293 cells. ***B.*** Western blot analysis to test effects of KDM5A over-expression on p53-K370me1/2, total p53 and p21 expression in HEK 293 cells. ***C.*** Densitometry of p53-K370me1 levels in response of WT KDM5A and Mut H483A KDM5A overexpression. Results are mean ± std. error (n=3). * indicates $p < 0.05$ from vector control. ***D.*** Densitometry of p53-K370me2 levels in response of WT KDM5A and Mut H483A KDM5A overexpression. Results are mean ± std. error (n=3). * indicates $p < 0.05$ from vector control. ***E***. qPCR analysis of p53 and p21 gene expression in HEK293

cells to test p53 transcriptional activity as a result of KDM5A over-expression. Results are mean ± std. error (n=3). ***F-G.*** PRM-MS/MS analysis of p53 immunoprecipitation (IP) from HEK293 cells to determine levels of p53 K370me3 expression upon KDM5A over-expression, analyzed through mass spectrometry. Figure caption continued from Figure F2 (page 235).

# Appendix G

## Nutlin 3A optimization in HCT 116 cells.

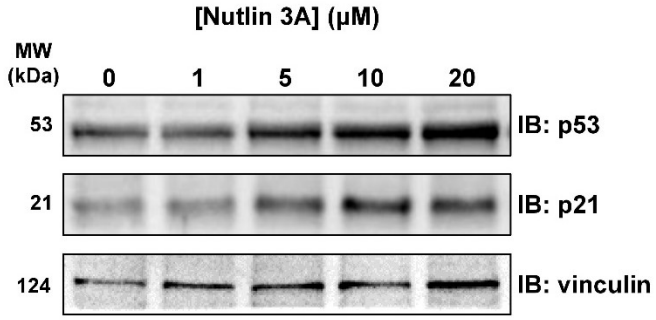



**Figure G1. Nutlin 3A dose response in HCT 116 cells**. Western blot analysis to determine optimal concentration of Nutlin 3A to use in HCT 116 cells for p53 stabilization and activation.

# Appendix H

## Analysis of p53-K370me2/1 peptide interactomics

**Table H1. Results of p53-K370me2 peptide interactomics through p53-K370me2 peptide pull down coupled with reductive dimethylation**. MSMS analysis identified 3 unique protein peptides that interacted strongly with p53-K370me2 peptide.

| | Gene name | Protein names | H/L ratio Forward Log2 | H/L ratio Reverse Log2 | Unique peptides Forward | Unique peptides Reverse |
|---|---|---|---|---|---|---|
| K370me2 | RGPD5 | RANBP2-like and GRIP domain-containing protein 5/6 | H only | L only | 1 | 1 |
| | PSMA5 | Proteasome subunit alpha type-5;Proteasome subunit alpha type | H only | L only | 1 | 1 |
| | SFPQ | Splicing factor, proline- and glutamine-rich | 0,706994616 | -0,429381635 | 5 | 4 |
| K370me0 | POM121C | Nuclear envelope pore membrane protein POM 121C | -0,360360327 | 0,484086559 | 7 | 6 |
| | PKP3 | Plakophilin-3 | -0,711842277 | 0,755913054 | 16 | 12 |
| | PML | Protein PML | -0,812029606 | 0,706198987 | 3 | 3 |
| | MCCC2 | Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial | -0,671002735 | 1,09173368 | 22 | 14 |
| | MCCC1 | Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial | -1,592092577 | 2,029664901 | 14 | 6 |
| | ACACA | Acetyl-CoA carboxylase 1 | -2,226003675 | 2,491570942 | 65 | 58 |
| | PC | Pyruvate carboxylase, mitochondrial | -2,427246121 | 2,858219766 | 44 | 41 |
| | PCCB | Propionyl-CoA carboxylase beta chain, mitochondrial | L only | H only | 10 | 9 |
| | PCCA | Propionyl-CoA carboxylase alpha chain, mitochondrial | L only | H only | 9 | 13 |
| | SRSF9 | Serine/arginine-rich splicing factor 9 | L only | H only | 1 | 1 |
| | MRPS21 | 28S ribosomal protein S21, mitochondrial | L only | H only | 1 | 1 |
| | SLC35B2 | Adenosine 3-phospho 5-phosphosulfate transporter 1 | -0,164932907 | H only | 4 | 1 |
| | COX4I1 | Cytochrome c oxidase subunit 4 isoform 1, mitochondrial | -0,33562639 | H only | 2 | 1 |

**P53 K370 me0 vs me2 interactors (>1300 identifications; 16 hits)**

**Table H2. Results of p53-K370me1 peptide interactomics through p53-K370me1 peptide pull down coupled with reductive dimethylation**. MSMS analysis identified 4 unique protein peptides that interacted strongly with p53-K370me1 peptide.

| | Gene name | Protein names | H/L ratio Forward Log2 | H/L ratio Reverse Log2 | Unique peptides Forward | Unique peptides Reverse |
|---|---|---|---|---|---|---|
| K370me1 | PYCR1 | Pyrroline-5-carboxylate reductase 1, mitochondrial;Pyrroline-5-carboxylate reductase | H only | L only | 1 | 3 |
| K370me1 | SAMD1 | Atherin | H only | L only | 1 | 2 |
| K370me1 | UFL1 | E3 UFM1-protein ligase 1 | H only | L only | 1 | 1 |
| K370me1 | LOXL2 | Lysyl oxidase homolog 2 | H only | L only | 1 | 1 |
| K370me0 | PKP2 | Plakophilin-2 | -0,407344438 | H only | 5 | 2 |
| K370me0 | MCCC2 | Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial | -0,598129286 | 1,008271738 | 16 | 19 |
| K370me0 | PKP3 | Plakophilin-3 | -0,83040846 | 0,831715132 | 16 | 12 |
| K370me0 | MCCC1 | Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial | -1,619758181 | 2,062190805 | 12 | 12 |
| K370me0 | ACACA | Acetyl-CoA carboxylase 1 | -2,269512201 | 2,277776508 | 71 | 69 |
| K370me0 | PC | Pyruvate carboxylase, mitochondrial | -2,518039803 | 2,782366626 | 44 | 46 |
| K370me0 | PCCB | Propionyl-CoA carboxylase beta chain, mitochondrial | L only | H only | 9 | 15 |
| K370me0 | PCCA | Propionyl-CoA carboxylase alpha chain, mitochondrial | L only | H only | 14 | 11 |
| K370me0 | PSMD13 | 26S proteasome non-ATPase regulatory subunit 13 | L only | H only | 1 | 1 |
| K370me0 | PML | Protein PML | L only | 0,788936229 | 3 | 5 |

**P53 K370 me0 vs me1 interactors (>1400 identifications; 14 hits)**

# Appendix I.

## List of publications and patents

### Publications

- **Hoekstra, M**., Biggar, K.K. (2022). Characterisation of KDM5 family substrate specificity and identification of potential novel substrates. *J. Biochem*. In submission.
- Chopra, A., **Hoekstra, M**, Willmore, W., Biggar, K.K. (2022). Mapping Protein Lysine Methylation in Cellular Reprogramming using Targeted-Mass Spectrometry. *Methods in Molecular Biology*. In submission.
- **Hoekstra, M**., Chopra, A., Willmore, W., Biggar, K.K. (2022). Evaluation of Jumonji C lysine demethylase substrate preference to guide identification of *in vitro* substrates. *Star Protocols: Cell Press.* 3: 101271.
- **Hoekstra, M**., Biggar, K.K. (2021). Identification of *in vitro* JMJD lysine demethylase candidate substrates via systematic determination of substrate preference. *Analytical Biochem*. 633: 114429.
- Biggar, K.K., Charih, F., Liu, H., Ruiz-Blanco, Y.B., Stalker, L., Chopra, A., Connolly, J., Adhikary, H., Frensemier, K., **Hoekstra, M**., Galka, M., Fang, Q., Wynder, C., Stanford, W.L., Green, J.R., Li, S.S. (2020). Proteome-wide prediction of lysine methylation leads to identification of H2BK43 methylation and outlines the potential methyllysine proteome. *Cell Reports.* 32: 107896.

**Patents**

- Peptide-derived therapeutics targeting KDM5C for the treatment of cancer. United States. WO2020186345A1, PCT/CA2020/050332, Patent Status: Published, 2020-09-24

# Appendix J.

## Communication at scientific meetings

- **Hoekstra, M*.**, Biggar, K.B. KDM5 lysine demethylase family substrate preference and identification of potential methyllysine-containing protein substrates through PRM-MS. *19th annual OCIB symposium,* University of Ottawa (virtual). May 2022. (PhD work)
- **Hoekstra, M*.**, Biggar, K.B. KDM5 lysine demethylase family substrate preference and identification of potential methyllysine-containing protein substrates. *7th annual Canadian Conference on Epigenetics*, CEEHRC Network. November 2021. (PhD work)
- Adhikary, H*., **Hoekstra M**., Jagadeesan, S., Chopra, A., Mckay, B., Biggar, K.K. The systematic development of peptide-based inhibitors for cancer treatment: A cell-active KDM5C/JARID1C-specific peptide inhibitor for *VHL-mutant* ccRCC. *5th annual Canadian Cancer Research Conference.* Ottawa, Canada. November 2019.
- **Hoekstra, M*.**, Biggar, K.B. Development of an Enzyme Activity Assay to Study Substrate Selection of KDM5/JARID1 Family of Lysine-Specific Histone Demethylases. *21st annual Chemistry and Biochemistry Graduate Research Conference,* Concordia University. December 2018. (PhD work)
- **Hoekstra, M*.**, Biggar, K.B. Development of Enzyme Activity and Inhibitor Assay for KDM5A/JARID1A Lysine-specific histone demethylase*. 2nd annual Life Science Research Day,* Carleton University. May 2018. (PhD work)
- **Hoekstra, M*.**, Biggar, K.B. Development of enzyme activity assay for KDM5A/JARID1A lysine-Specific Demethylase. *15th annual OCIB symposium,* University of Ottawa. April 2018. (PhD work)

- **Hoekstra, M*.**, Willmore, W.G. (2017). Analysis of NFE2L1: Structure, Function, PTMs and homology modelling. *Carleton Research Day*, Carleton University. April 2017. (B.Sc. work)