Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

10-3-2019 2:00 PM

Systematic identification of the lysine methylome using methyllysine binding domains

Wen Qin
The University of Western Ontario

Supervisor
Shawn Li
The University of Western Ontario

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science
© Wen Qin 2019

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

Part of the Biochemistry Commons

Recommended Citation

Qin, Wen, "Systematic identification of the lysine methylome using methyllysine binding domains" (2019). *Electronic Thesis and Dissertation Repository*. 6625.
https://ir.lib.uwo.ca/etd/6625

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

Abstract

Post-translational modifications (PTM) are vital regulators of protein function and homeostasis. The role of dynamic regulations of non-histone lysine methylated proteins (NHKMP) recently began to be recognized in DNA damage repair, apoptosis and transcriptional pathways. My goal was to identify components of the NHKMP network to understand its importance in a healthy versus diseased cellular state. I used membrane peptide arrays to systematically characterize nine naturally occurring lysine methyl binding domains (KMBD). Five KMBDs were chosen based on their overlapping specificities to achieve maximum coverage of lysine methylated peptides. These five KMBDs was used to enrich for methylated lysine peptides from a trypsinized HEK293 cell lysate and followed by mass spectrometry identification. We identified 229 NHKMP and 301 novel sites from HEK293. The amount of NHKMPs and sites that I have identified in total was unprecedented: this allows us to gain valuable insights into components of the lysine methylome network.

Keywords

Protein methylation, lysine methylation, non-histone protein methylation, peptide array, mass spectrometry

Summary for lay audience

Proteins can have many modifications post-synthesis, called post-translational modification. These PTMs are responsible for regulating vital pathways within the cells and ensure the survival of the cells. In this thesis, I studied one type of PTM called protein lysine methylation. Protein lysine methylations are found in DNA regulation, cell regulation and other pathways. It is also very common to have dysregulation of protein methylation within cancer cells. To understand how protein lysine methylation works within the cell, we first need to identify the proteins that are methylated. Current methods cannot identify a huge range of methylated lysine proteins within one experiment. Therefore, for my thesis, I developed a method to enrich and identify methylated lysine proteins. In this thesis, I was able to identify several novel proteins that were methylated within the cell.

Co-Authorship Statement

Dr. Suzanne Ackloo from SGC Toronto provided the plasmids for KMBDs used in this projected. Dr. Kyle Biggar generated Fig 1.5 and the peptide list for the first 70-protein peptide array used in Fig. 3.5. Dr. Lei Li and Eric Liu designed and generated the phylogenetic clustering of methylated peptides. Dr. Xuguang Liu helped with the mass spectrometry experiments and the identification of lysine methylated proteins.

Acknowledgments

I want to thank my supervisor Dr. Shawn Li, for the mentorship and support provided me in the past three years. It has been a privilege to be able to a member of this lab to learn and grow.

I want to thank my committee members Dr. Patrick O'Donoghue and Dr. Madhulika Gupta, for their help, insight and guidance throughout my project.

I want to give special thanks to Dr. Suzanne Ackloo from the Structural Genomics Consortium at the University of Toronto for the biotinylated binding domains. Without you, this project would not be possible.

I want the thank the current and former members of the Li lab for their ongoing support, advice and guidance throughout my graduate career. Thank you, Dr. Kyle Biggar, for your valuable guidance and support to help me get the project started. Thank you, Dr. Xuguang Liu, for your help and your expertise in mass spectrometry, helping me with troubleshooting and data collection and analysis. Thank you, Dr. Lin Zhao, for your guidance in cell culture, experiment setup and fun and tricks in and out of the lab. Thank you, Dr. Huadong Liu, for your guidance in peptide array and peptide chemistry. Thank you, Dr. Lei Li, for your computational work and for designing the phylogenetic clustering of methylated peptides. Thank you, Courtney Voss, for lab maintenances and planning fun activities for the lab. Thank you, Dr. Tomonori Kaneko, Dr. Xiaoling Liu, Dr. Lyugao Qin, Fang Qi, Owen Hovey, Shanshan Shen, for your support and the brainstorm sessions to figure out a problem.

I want to give a special thank you to my girlfriend, Kyla Yang, for your love, encouragement and reminders.

I want to thank my parents for their unwavering love and support throughout my graduate career.

Table of Content

Abstract i

Summary for lay audience ii

Co-Authorship Statement iii

Acknowledgments iv

List of Tables viii

List of Figures ix

List of Abbreviation xi

Chapter 1 1

1 General Introduction 1

1.1 Protein methylation 1

1.2 Protein lysine methylation 4

1.2.1 Lysine methyltransferases and demethylases 7

1.2.2 Lysine methyl binding domains 13

1.3 Histone methylation – methyllysine 16

1.4 Non-histone methyllysine proteins 17

1.5 Previous studies 21

1.6 Rationale and objectives 22

Chapter 2 24

2 Methods 24

2.1 Protein expression and purification 24

2.2 Bradford Assay 25

2.3 Bis-Tris SDS-PAGE and Coomassie staining 25

2.4 Western blot 26

2.5 Peptide membrane array synthesis 26

2.6 Peptide membrane binding and analysis 27

2.7 In-solution peptide synthesis 28
2.8 Mammalian cell culture 28
2.9 Streptavidin pull-down 30
2.10 Silver staining 30
2.11 Mass spectrometry 31
2.12 Fluorescence polarization assay 31
2.13 Systematic clustering of lysine methylated peptides 32
Chapter 3 33
3 Results 33
3.1 Characterizing lysine methyl binding domain specificity 33
3.1.1 Expression and purification of lysine methyl binding domains 33
3.1.2 Biotinylation of KMBD 36
3.1.3 53BP1-tudor binding to standard peptides 39
3.1.4 Peptide membrane array to probe KMBD specificity 41
3.2 Lysine methyl binding domain substrate prediction 45
3.2.1 Construction of lysine methylated peptide based on a phylogenetic tree approach 45
3.2.2 Results from the 140-cluster peptide array 45
3.2.3 KMBD biding specificity from 140-cluster array versus previous 70-protein array 51
3.2.4 53BP1-tudor domain binding substrate prediction 52
3.2.5 53BP1-tudor domain predicted substrates 56
3.3 Lysine methylome identification 60
3.3.1 Elution conditions for KMBD using standard peptides 60
3.3.2 Methyl peptide enrichment 69
4 Discussion and future directions 83
4.1 General discussion and conclusions 83

4.1.1 Systematic characterization of KMBD binding profile 83
4.1.2 53BP1-tudor substrate prediction 87
4.1.3 Methyllysine peptide enrichment by KMBDs 88
4.1.4 Strong cation exchange column in methyllysine enrichment 91
4.1.5 Mass spectrometry identification and database search 92
4.2 Future Directions 94
5 Reference 96
6 Supplemental Data 108
Curriculum Vitae 140

List of Tables

Table 1.1 A list of human lysine methyltransferase and demethylases with their substrates, inhibitors and common cancers. 8

Table 3.1. List of lysine methyl binding domains used in this study 35

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs. 74

Table S.1. List of peptides used in the 70-protein array 108

Table S.2. List of peptides used in the 140-cluster array 118

List of Figures

Figure 1.1. A general overview of protein methylation..... 3

Figure 1.2. Comparison between protein lysine and arginine methylation 6

Figure 1.3. Overview of lysine methyl binding domain 15

Figure 1.4. Published papers on histone methylation versus nonhistone methylation in the past 50 years 19

Figure 1.5. BiNGO analysis of major biological processes where lysine methylation was enriched..... 20

Figure 3.1. Coomassie blue-stained gel of his-tagged lysine methyl binding domain purification: CHD1-CD, MPP8-CD, PHF1-tudor..... 34

Figure 3.2. MPP8-CD purification and western blot with streptavidin-HRP 37

Figure 3.3. Coomassie blue staining of L3MBTL2-3xMBT binding to streptavidin beads... 38

Figure 3.4. The kinetic dissociation constant of 53BP1-tudor with H4K20 peptides was determined by fluorescence polarization 40

Figure 3.5. Lysine methyl binding domain binding heat map to 70 proteins peptide array ... 44

Figure 3.6. 140-cluster peptide array binding with the 53BP1-tudor domain 47

Figure 3.7. Logos of top and bottom 5% of Kme1 peptides to the 53BP1-tudor domain from 140-cluster peptide array 48

Figure 3.8. Lysine methyl binding domain heatmap of 140-cluster peptide array 50

Figure 3.9. Using 140-cluster data to construct binding prediction logos for the 53BP1-tudor domain 54

Figure 3.10. Permutation array of NHEJ1, the number one predicted substrate for the 53BP1-tudor domain 55

Figure 3.11. Prediction ranking of all known methyllysine sites to a 53BP1-tudor domain based on peptide array data 58

Figure 3.12. Biological pathway analysis of the top 100 predicted 53BP1-tudor binding substrate 59

Figure 3.13. Coomassie blue staining of MPP8-CD binding to streptavidin beads at varying urea concentration and pH. Purified MPP8-CD were conjugated to streptavidin beads 61

Figure 3.14. Silver staining of p100 binding to streptavidin beads at varying pH and urea concentrations 62

Figure 3.15. Using fluorescein tagged H4K20me0/2 peptide testing elution conditions with varying pH 64

Figure 3.16. Using fluorescein tagged H4K20me0/2 peptide testing SCX conditions with varying pH 66

Figure 3.17. Western blot and a far-western blot of intact protein pulldown from HEK293 lysate via 53BP1-tudor domain 68

Figure 3.18. Methyllysine sites and proteins identified via 53BP1-tudor affinity purification of HEK293 cell lysate followed by MS 71

Figure 3.19. Methyllysine sites and proteins identified from HEK293 cell lysate via MPP8-CD affinity purification and MS 73

Figure 3.20. Methyllysine sites and proteins identified from HEK293 cells via five KMBD super-column followed by MS 76

Figure S.1. Stripped 140-cluster array 139

List of Abbreviation

53BP1	p53 binding protein 1
6-FAM	6-Carboxyfluorescein
ACN	Acetonitrile
AP-MS	Affinity-purification mass spectrometry
APS	Ammonium persulfate
APS	Ammonium persulfate
BCA	Bicinchoninic acid assay
BSA	Bovine serum albumin
CBX7	Chromobox protein homolog 7
CBX8	Chromobox protein homolog 8
CD	Chromatin-binding domain
CHD1	Chromodomain-helicase-DNA-binding protein 1
DCM	Dichloromethane
DDR	DNA damage repair
DMEM	Dulbecco's modified eagle medium
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol

ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
FA	Formic acid
FAD	Flavin adenine dinucleotide
FBS	Fetal bovine serum
FP	Fluorescence polarization
GSH	Glutathione
GST	Glutathione S-transferase
H4K20	Histone H4 lysine 20
His-Tag	6x histidine tag
HRP	Horseradish peroxidase
IAA	Iodoacetamide
IPTG	Isopropyl β-D-a-thiogalactopyranoside
JAK	Janus kinase
JMJD	JmjC domain
K_D	Equilibrium dissociation constant
KDM	Lysine demethylase
KMBD	Lysine methyl-binding domain
Kme1	Monomethylated lysine
Kme2	Dimethylated lysine

Kme3	Trimethylated lysine
KMT	Lysine methyltransferase
L3MBTL2	Lethal (3) malignant brain tumor-like protein 2
LB	Lysogeny broth medium
LSD1	Lysine-specific demethylase 1
MALDI-MS	Matrix-assisted laser desorption/ionization – mass spectrometry
MBT	Malignant brain tumor domain
MLL	Mixed-lineage leukemia
MOPS	3-(N-morpholino)propanesulfonic acid
Mpp8	M-phase phosphoprotein 8
NHKMP	non-histone lysine methylated protein
NSD	nuclear-receptor-binding SET domain
P100	Staphylococcal nuclease domain-containing protein 1
PAGE	polyacrylamide gel electrophoresis
PBS	Phosphate buffered salt
PBST	Phosphate buffered salt with Tween 20
PHD	plant homeodomain
PHF1	PHD finger protein 1
PHF19	PHD finger protein 19
PMSF	Phenylmethylsulfonyl fluoride

PPI	protein protein interaction
PTK	Protein tyrosine kinase
PTM	Post-translational modification
PTP	Protein tyrosine phosphatase
pTyr	Phosphotyrosine
PVDF	Polyvinylidene fluoride
PWWP	Proline-tryptophan-tryptophan-proline motif
SAM	S-adenosyl-L-methionine
SCX	Strong cation exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SET	[Su(var)3-9, Enhancer-of-zeste and Trithorax]
Set8	N-lysine methyltransferase SETD8
SMYD	SET and MYND domain-containing proteins
STAT	Signal transducer and activator of transcription
TBS	Tris buffered saline
TBST	Tris buffered saline with Tween 20
TD	Tudor domain
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine

TFA	Trifluoroacetic acid
TIPS	Tri-isopropylsilane
TOF	Time-of-flight
UTX	Ubiquitously transcribed tetratricopeptide repeat on chromosome X
WT	Wild type

Chapter 1

1 General Introduction

Protein post-translational modifications (PTM) are vital regulators of protein function. These PTMs can give a protein a specific function and can alter the protein-protein interaction (PPI) network. PTMs are a very energy-efficient way to alter protein function [1]. Thus, PTMs can give a protein the potential of interacting with a variety of partners depending on the needs of the cell. Such as turning an enzyme on or off with one or more PTMs. PTMs are tightly regulated in cellular processes to maintain homeostasis. Aberrations in PTMs can result in a diseased cell state through dysregulation of DNA damage repair (DDR), apoptosis, metabolic and other essential pathways [2,3].

There are many known PTMs such as phosphorylation, acetylation, ubiquitination, glycosylation, methylation and others. One of the most studied PTM is protein phosphorylation [4]. Protein phosphorylation is highly regulated, highly dynamic and commonly found perturbed in diseased cells, such as cancer. Protein phosphorylation is known for signalling cascades and signalling transduction. Phosphorylation is a relatively fast PTM that can occur seconds to hours after stimulus [5]. Protein phosphorylation is regulated by a protein kinase, a "writer," where it puts on the signal. Then the signal gets read by a "reader" protein. Finally, the signal gets removed by a phosphatase, an "eraser." Some of the other PTMs and other pathways are regulated via the writer-reader-eraser mechanism.

1.1 Protein methylation

Protein methylation is a very small, hydrophobic modification onto the side chains of many residues. There are at least nine different residues that could be methylated in the cell: Met, Cys, Lys, Arg, His, Gln, Asn, Glu and Asp [6]. However, the most common methylated residues are found on lysine and arginine. The addition of one or more methyl-groups onto Arg or Lys residues will not affect their overall charge (Fig. 1.1) [7].

The first protein methylation was reported in 1959 [8], and it was fascinating at the time. However, not much interest soon followed. It was not until the late 1990s and early 2000s when protein methylation gained more popularity [9]. The stagnation in the study of protein methylation was partly due to the dogma that protein methylation was irreversible, and it does not have much biological relevance [10]. With technological advances of the late 1990s, gene regulation by histones was a hot topic [11]. Histone methylation on lysine and arginine are important in gene regulation [12]. A hypothesis was conceived, the "histone code," where histone PTMs, like methylation, acetylation and phosphorylation, contributing to gene regulation [13]. It catalyzed an interest to investigate further about protein methylation. Several histone methyltransferases, both arginine and lysine, were discovered at the time [9]. However, it was not known whether methylation is dynamic or static modification until 2004, where the first lysine demethylase was discovered [14]. The discovery of a demethylase, an eraser, abolished the 45 years of dogma that protein methylation was a static and irreversible modification.

Protein methylation gained a lot of interest in the past few years because of its biological relevance, finally being unravelled, starting with the discovery of the first lysine methyltransferase (KMT) in 1996 [15]. Protein methylation is involved in many key essential pathways [9], for example, DNA damage repair, RNA processing, metabolism, cell cycle regulation, apoptosis and others [3,16]. Aberrations in protein methylation often lead to diseased cell states such as inflammation and cancer [3,17].

Unlike other PTMs, protein methylation is seen as a slow form of regulation that lasts for a longer time than some other PTMs [18]. Some methylation events can take hours to even days after a stimulus is introduced, and the turnover rate may also not be as fast as other PTMs, such as phosphorylation [19]. Protein methylation is regulating genes on an epigenetic level through histone modifications [20].

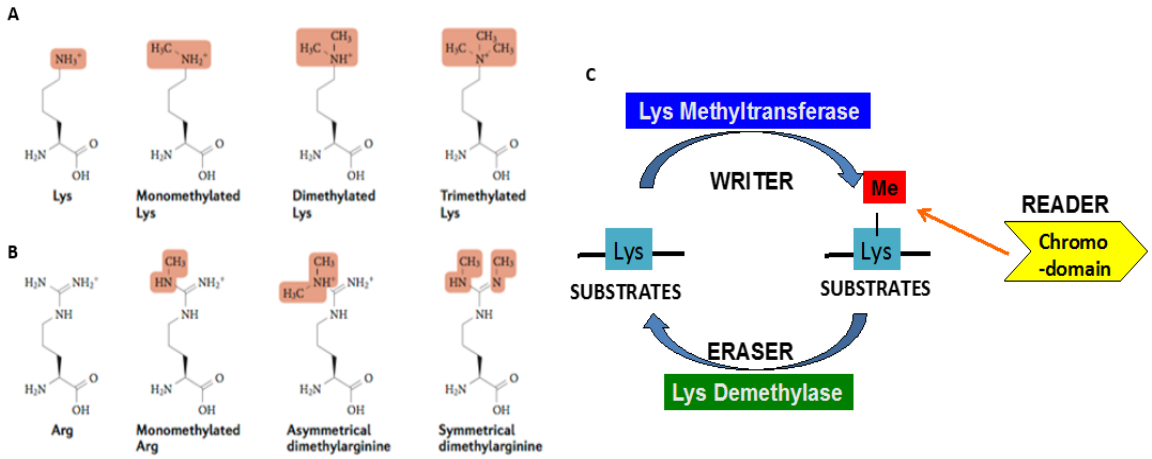

Figure 1.1. A general overview of protein methylation. A) Lysine methylation can occur in four modification states, none, mono-, di- or tri-methylation on the ε-N-methyl lysine. B) Arginine methylation can also have four different states, none, mono-, di-methylation. However, dimethylarginine can be asymmetrical or symmetrical. C) The writer, reader, eraser of a dynamic post-translational modification. The writer puts on a methyl-group, it is read by a reader to transduce the signal, and finally, the methyl-group gets removed by an eraser to stop the signal. It is similar to other dynamic PTMs, such as phosphorylation and acetylation. Figure adapted from Dr. Kyle Biggar [21].

1.2 Protein lysine methylation

Sanger first observed methylated lysine residue in 1943 [22]. He observed that the methyl-group is attached to the side chain amino group of the lysine residue (ε-N-methyl-lysine) and not on the terminal amino group (Fig. 1.1A). Sanger observed that ε-methyl-lysine residues could substitute lysine residues and were demethylated in the kidney slices [23]. However, his discovery was very advanced in his times. Now, more about lysine methylated proteins are being discovered and understood. Dedicated lysine methyltransferases and demethylases (KDM) were discovered[3].

Lysine methylation follows many PTMs with the writer-reader-eraser model. For the lysine residue, one, two or three methyl-groups could be attached to the ε-amino group, named mono, di, tri-methylated lysine, respectively. (also known as Kme1, Kme2, Kme3). The transfer of a methyl group to a lysine residue is facilitated by a lysine methyltransferase (KMT), the writer. It generates three different modifications with increasing hydrophobicity as more methyl-groups are added [21]. The addition of a methyl-group does not affect the charge of the residue. After the modification is added, there are about 148 different lysine methyl-binding domains (KMBD) that can read the modification[24]. These reader domains on effector proteins, transducing the signal or as a recruiting site to recruit other proteins [25–27]. Furthermore, these reader domains are also found on the writers and erasers to facilitate binding to their substrates. The removal of the methyl-group requires a lysine demethylase, the eraser. There are several mechanisms involved in removing a methyl-group [28]. Some demethylases can remove mono- or di-methylation; some can remove tri-methylation only, and some can remove all three [3].

There are a lot more arginine methylation sites that have been uncovered, and a lot of these correspond to important regulatory pathways within the cell [29]. However, protein lysine methylation sites are far fewer compared to arginine methylation, 5,300 to 14,100 sites, respectively [24]. There are roughly three times more methylated arginine sites than lysine sites currently found (Fig. 1.2A), and the majority of both are mono-methylated sites. However, by comparing the number of unique proteins, the discrepancy is not as drastic (Fig. 1.2B). An interesting comparison is that there are a lot of proteins containing

both arginine and lysine methylation (Fig. 1.2C). It adds to the complexity of methylation and crosstalk between lysine and arginine methylation as well as other PTMs.

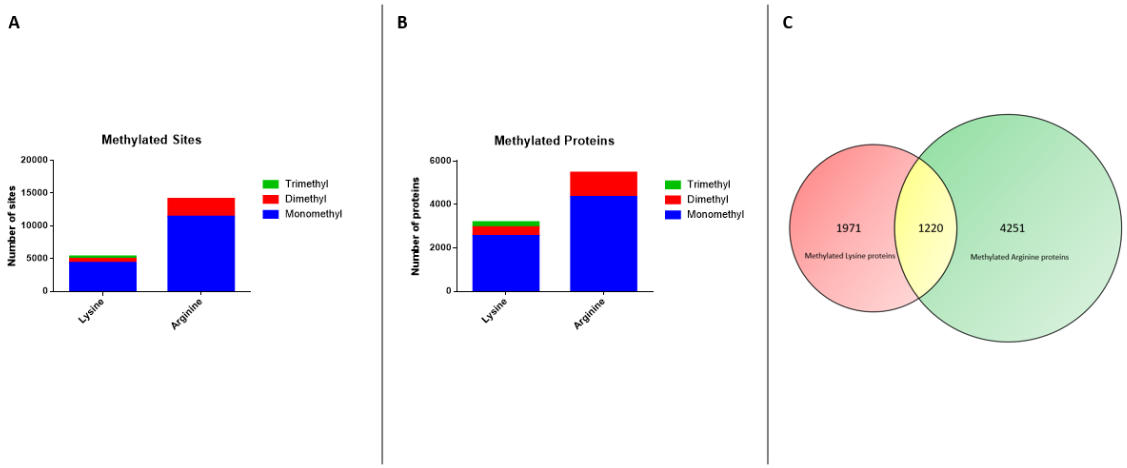

Figure 1.2. Comparison between protein lysine and arginine methylation. All the data were obtained from PhosphoSitePlus [24]. A) Current protein lysine and arginine methylated sites, lysine methylation, only has about 5,300 sites discovered so far. Whereas about 14,100 sites were found for arginine methylation. B) Comparing lysine and arginine methylation when those sites are corresponding to proteins. Several arginine methylated protein has many sites. C) Looking at the crosstalk between methylated lysine and arginine proteins. There are about 1,220 proteins that have both lysine and arginine methylation sites.

1.2.1 Lysine methyltransferases and demethylases

Lysine methyltransferase adds one or more methyl-groups onto a lysine residue. The cofactor used by KMTs is S-adenosyl-L-methionine (SAM), the sulphur on the methionine serves as the electrophile that transfers the methyl-group in an S_N2 reaction [30]. There are several families of KMTs, the [Su(var)3-9, Enhancer-of-zeste and Trithorax] (SET), nuclear-receptor-binding SET domain (NSD), mixed-lineage leukemia (MLL), Suv, and the SET and MYND domain-containing proteins (SMYD) family transferases [31]. Some of the KMTs have different specificities where it can be only able to add one out of the three modification states. Whereas other KMTs may add more than one modification states [31]. Moreover, there is an overlap between several KMTs, where they transfer to the same substrates, but they are expressed in different tissues or at different cell cycle stages [32]. For instance, SETD7 is known only to monomethylate lysine residues [33]. It monomethylates H3K4, and a starts a building block for other KMTs, such as MLL1/2 to di- or tri-methylate the substrate [34]. There are about 49 known KMTs, some are reviewed, and some are predicted and unvalidated [24]. Table 1.1 lists KMTs, as well as KDMs with their histone substrates and their modification preferences, common cancers that are linked to KMTs and KDMs.

Lysine demethylase removes the methyl-group from Kmes. There are fewer erasers known than writers. There are about 28 KDMs, and three of them are unvalidated yet [24]. There are several families of KDMs as well, and they use a different mechanism to remove the methyl group. There are four families of KDMs: lysine-specific histone demethylase 1 (LSD1), JmjC domain (JMJD), ubiquitously transcribed tetratricopeptide repeat on chromosome X (UTX) and JARID family [35]. The first-ever discovered KDM is lysine-specific histone demethylase 1 (LSD1) [14]. LSD1 demethylates mono- and di-methylated lysine of H3K4 and H3K9. LSD1 uses a cofactor flavin adenine dinucleotide (FAD) to oxidize the methylated lysine to remove the methyl-group [14]. Since LSD1 cannot remove Kme3, there must be other enzymes with another mechanism to remove Kme3s. Another family, the JMJD family, uses iron (II) and 2-oxoglutarate-dependent deoxygenation [35]. The JMJD family can remove methyl-groups from all three

modification states. The UTX and JARID core catalytic domain is JmjC, using a similar mechanism as JMJD family [31].

Table 1.1 A list of human lysine methyltransferase and demethylases with their substrates, inhibitors and common cancers.

LYSINE METHYLTRANSFERASES

FAMILY	Gene	Alias	Substrate	Non-histone	Kme Mods	Inhibitors	Cancers	Ref
SET1	KMT2E	MLL5	H3K4	-	kme1/2	-	Colorectal, breast, prostate, lung, liver	[24,36,37]
SET1	KMT2C	HALR	H3K4	-	kme1/2/3	-	Breast, glioblastoma, melanoma, MLL, oesophageal, pancreatic, stomach	[24,36–38]
SET1	SETD1A	Set1	H3K4	HSP70	kme3	-	Bladder, CRC, HCC, lung and RCC	[24,36,37,39]
SET1	EZH2	WVS	H1.4K25; H3K4; H3K9; H3K27	RORα, STAT3	kme1/2/3	E7438, GSK2816126, CPI-1205, GSK343, GSK126, UNC1999	AML, bladder, breast, CCC, CML, CRC, glioblastoma, lymphoma, NSCLC, oesophageal, osteosarcoma, SCLC, RCC	[3,24,36,37,40–45]
SET1	EZH1	KMT6B	H3K27	-	kme1/2/3	-	Colorectal, breast, prostate, lung, liver	[24,36,37]
SET1	KMT2A	HRX	H3K4	-	kme1	-	AML, ALL, MLL	[24,36,37]
SET1	SETD1B	KMT2G	H3K4	-	kme3	-	Colorectal, breast, prostate, lung, liver	[24,36,37]
SET1	KMT2D	ALR	H3K4	ERα	kme1/kme3	-	Hepatocellular carcinoma, HCC,	[24,36,37]
SET1	KMT2B	HRX2	H3K4	-	kme3	-	Stomach, endometrial, bladder, lung	[24,36,37]
SET2	NSD2	WHSC1	H3K27, H3K26, H4K20	-	kme1	-	Bladder, breast, CCC, CML, HCC, multiple myeloma, NSCLC, oesophageal, osteosarcoma, prostate, RCC and SCLC	[24,36,37,46]
SET2	ASH1L	KMT2H	H3K4, H3K36	-	kme1	-	Colorectal, breast, prostate, lung, liver	[24,36,37]
SET2	NSD3	WHSC1L1	H3K36, H3K4, H3K27	-	kme2	-	Bladder, breast, lymphoma, SCLC	[24,36,37,46]
SET2	NSD1	STO	H3K36, H4K20	-	kme1/2/3	-	AML, glioblastoma, lung, multiple myeloma	[24,36–38,46]
SET2	SETD2	HYPB	H3K36	-	kme3	-	Colorectal, breast, prostate, lung, liver	[24,36,37]

Table 1.1. List of lysine methyltransferases and demethylases (continued)

RIZ	PRDM1	BLIMP 1	H3K9	-			RENAL, LUNG, ENDOMETRIAL, STOMACH	[24,36,37,47]
RIZ	PRDM2	RIZ	H3K9	-	kme1/2	-	Testicular, breast, liver, bone, lung, colon	[24,36,37,48–50]
RIZ	PRDM12	PFM9	H3K9	-	kme2	-	Lung, colorectal, endometrial	[24,36,37]
RIZ	PRDM5	BCS2	H3K9	-	kme2	-	Endometrial, stomach, bladder, lung	[24,36,37]
RIZ	PRDM16	MEL1	K3K9	-	kme1	-	Bladder, endometrial, stomach, lung	[24,36,37]
RIZ	PRDM8	PFM5	H3K9	-			Stomach, kidney, lung, endometrial	[24,36,37,51]
RIZ	PRDM9	PFM6	H3K4	-	kme2/3	-	Lung, colorectal, endometrial, stomach	[24,36,37]
SUV39	SUV39H1	MG44	H3K9	-	kme3	-	Endometrial, stomach, lung	[24,36,37]
SUV39	SETDB2	CLLD8	H3K9	-	kme3	-	Colorectal, bladder, stomach, lung	[24,36,37]
SUV39	SETDB1	ESET	H3K9	-	kme3	-	Endometrial, stomach, lung, colorectal	[24,36,37]
SUV39	SUV39H2	KMT1B	H3K9	-	kme3	-	Bladder, cervical, NSCLC, oesophageal, osteosarcoma, prostate and STT	[24,36,37]
SUV39	SETMAR	Mar1	H3K36	-	kme2	-	Endometrial, stomach, colorectal	[24,36,37]
SUV39	EHMT1	GLP	H3K9, H1.2K186	p53	kme1/2	BIX-01294, E72, UNC0321, UNC0638	Endometrial, Stomach, lung, head/neck, lung	[24,36,37,52–56]
SUV39	EHMT2	G9A	H3K9, H3K27, H3K56, H1.2K186, H1.4K25	p53, CDYL, WIZ, ACIN1, DNMT1, HDAC1, ERCC6, KLF12	kme1/2	BIX-01294, UNC0642, A-366, BRD9539	AML, bladder, breast, CCC, CML, NSCLC, oesophageal, prostate, SCLC	[3,24,36,37,52,53,55–57]
SUV4-20	SUV420H1	KMT5B	H4K20	-	kme3	A-196	Breast, esophagus, prostate, bladder	[24,36,37,58]
SUV4-20	SUV420H2	KMT5C	H4K20	-	kme3	A-196	Stomach, head/neck, breast	[24,36,37,58]
								[24,36,37]
SMYD	SMYD1	BOP	H3K4	-	kme1/2/3	-		[24,36,37]
SMYD	SMYD2	KMT3C	H3K4, H3K36	p53, RB1, PARP1, HSP90AB1, Erα	kme1/2	AZ505, BAY-598, LLY-507	Bladder, breast, cervical, CRC, HCC, head and neck, lymphoma, oesophageal, ovarian, pancreatic, prostate, seminoma and skin	[3,17,24,36,37,59–61]
SMYD	SMYD3	KMT3E	H3K4	VEGFR1, MAP3K2	kme2/3	-	AC, breast, CCC, cervical, CRC, HCC, lung, MTC, oesophageal, pancreatic and prostate	[3,24,36,37]

Table 1.1. List of lysine methyltransferases and demethylases (continued)

OTHER	SETD6		H2A.ZK4me 1;	NF-Kβ	kme1	-		[24,36,37,62]
OTHER	SETD7	KMT7	H3K4; H4K20	TAF10, p53, PPP1R12A , NF-Kβ, E2F1, DNMT1, STAT4	kme1	PFI-2	breast, multiple myeloma	[3,24,33,36,37,63]
OTHER	SETD8	SET8	H4K20	p53, PCNA	kme1	-	breast, CML, HCC, NSCLC and SCLC	[3,24,36,37]
OTHER	DOT1L	KMT4	H4K79	-	kme1/2/ 3	EPZ004777, EPZ-5676, SGC0946	MLL, bladder, stomach, lung	[24,36,37,64–66]

Table 1.1. List of lysine methyltransferases and demethylases (continued)

LYSINE DEMETHYLASES

FAMILY	Gene	Alias	Substrate	Non-histone	Kme	Inhibitors	Cancers	Ref
LSD_KDM1	KDM1A	LSD1	H3K4me1/2, H3K9	PPP1R12A, p53, ERα and STAT3	kme1/2	trans-2-phenylcyclopropylamine, polyamine analogues, CBB1007	Bladder, lung, endometrial, stomach	3,24,36,37,67–69
JARID	KDM5D	SMCY	H3K4me2/3	-	kme2/3	-	Lung, bladder, colorectal, lung	24,36,37
JARID	KDM5A	RBP2	H3K4me1/2/3	-	kme1/2/3	-	AML, lung, endometrial, stomach, bladder	24,36,37
JARID	KDM5B	PLU-1	H3K4	-	kme1/2/3	PDCA	Endometrial, stomach, colorectal, lung	24,36,37,70,71
JARID	KDM5C	SMCX	H3K4me2/3	-	kme2/3	-	Kidney, endometrial, stomach, colorectal	24,36,37
								24,36,37
JHDM1	KDM2B	FBXL10	H3K4me3, H3K36me1/2	-	kme2	-	Stomach, endometrial, lung	24,36,37
JHDM2	KDM2A	KDM2A	H3K4me3, H3K36me1/2	-	kme2	-	Endometrial, lung, stomach, bladder	24,36,37
JHDM2	KDM3A	JHDM2A	H3K9, H3K36	-	kme2	-	Stomach, bladder, colorectal, endometrial	24,36,37
JHDM2	KDM3B	JHDM2B	H3K9	-	kme3	-	Endometrial, stomach, bladder, colorectal	24,36,37
JHDM2	JMJD1C		H3K9	-	kme2	-	Stomach, endometrial, bladder, head/neck	24,36,37
JHDM3	KDM4A	JMJD2A	H3K9, H3K36	-	kme3	N-oxalylglycine, 8-hydroxyquinoline	Lung, endometrial, bladder, stomach	24,36–38,72,73
JHDM3	KDM4C	JMJD2C	H3K9me3, H3K36me3	-	kme2/3	N-oxalylglycine, 8-hydroxyquinoline, methylstat	Endometrial, stomach, colorectal, lung	24,36–38,72–74
JHDM3	KDM4B	JHDM3B	H3K9me3	-	kme3	-	Stomach, lung, endometrial	24,36,37
JHDM3	KDM4D	JMJD2D	H3K9	-	kme2/3	8-hydroxyquinoline	Endometrial, stomach, lung, bladder	24,36–38,73
JHDM3	KDM4E		H3K9	-	kme2	-	Endometrial, lung	24,36,37
PHF2_PHF8	KDM7A	JHDM1D	H3K9me2, H3K27me2, H4K20	-	kme1/2/3	-	Stomach, endometrial, lung, head/neck	24,36,37
PHF2_PHF8	PHF8	ZNF422	H3K9me1/2, H3K27me2, H4K20me1, H3K4me3	-	kme1/2/3	-	Endometrial, stomach, lung, colorectal	24,36,37
								24,36,37
UTX_UTY	KDM6B	JMJD3	H3K4, H3K27me2/3	-	kme2/3	Methylstat, GSK-J1	Stomach, lung, endometrial, colorectal	24,36,37,74,75
UTX_UTY	KDM6A	UTX	H3K27me2/3, H3K4	-	kme2/3	GSK-J1	Bladder	24,36,37,75

Table 1.1. List of lysine methyltransferases and demethylases (continued)

UTX_UTY	KDM6C	UTY	H3K27me3	-	kme3	-	Bladder, colorectal	[24,36,37]
JMJC_ONLY	KDM8	JMJD5	H3K36me2	-	kme2	-	Endometrial, lung, stomach	[24,36,37]
JMJC_ONLY	NO66	ROX, MAPJD	H3K4me1/2/3, H3K36me2/3	-	kme1/2/3	-	CRC	[24,36,37]
JMJC_ONLY	MINA		H3K9me3	-	kme3	-	Endometrial, stomach, bladder	[24,36,37]

Table 1.1. A list of lysine methyltransferases and demethylases. This list contains some of the lysine methyltransferases and demethylases, with their histone substrates, their catalytic modifications, any known non-histone proteins, the cancer types that it occurs most in, and if there are any inhibitors for this enzyme. Abbreviations: CRC, colorectal cancer; HCC, hepatocellular carcinoma; RCC, renal cell carcinoma; MLL, mixed-lineage leukaemia; AML, acute myeloid leukaemia; CCC, cholangiocarcinoma; CML, chronic myelogenous leukaemia; NSCLC, non-small cell lung carcinoma; SCLC, small cell lung carcinoma; ALL, acute lymphocytic leukemia; STT, soft tissue tumour; MTC, medullary thyroid cancer.

1.2.2 Lysine methyl binding domains

The reader domains are the modular domains that recognize methylated lysine residues (Fig. 1.3). Since there are four different lysine-methylated states, these reader domains need to be able to recognize Kme0, Kme1, Kme2 or Kme3. There are several families of lysine methyl-binding domains: tudor domain, chromodomain, plant homeodomain (PHD) finger domain, PWWP (Pro-Trp-Trp-Pro), WD-40 and others [24]. There are about 148 reader domains. However, about 40% of them are unvalidated [24]. Reader domains are found on KMTs, KDMs or other effector proteins. They help the writer and erasers to find their substrate and others to conduct their signalling cascade.

One of the big family of KMBD is the Tudor domain family. There are about thirty mammalian Tudor containing proteins, for example, p53 binding protein 1 (53BP1), KDMs, PHD finger protein family and others [24]. Tudor domain was first found in *Drosophila* on the *tud* gene [76]. Tudor domain has about sixty residues, and it is folded in 4-5 antiparallel beta-strand sheets [25]. It can read methylated lysine, as well as some methylated arginine residues [77]. Tudor domains attached to different proteins have different specificities. For example, 53BP1-tudor are known to read Kme1/2 [78]. Whereas, PHF1 and PHF19 tudor domains can read H3K36me2/3 modifications [79].

Another big family reader domain is chromodomain (CD). CDs bind to Kmes using an "aromatic cage" [27]. Just like tudor family protein, where structurally conserved CD has diverse binding specificity to Kmes [80]. CDs can bind to all three Kmes, but varying affinity depending on the protein [81]. One of the most studied CD is HP1β, where it has a high affinity to H3K9 [82]. The CD of HP1β can bind to all three Kmes, but highest affinity on Kme3, then Kme2 with the worst affinity for Kme1 residues [81].

Different readers have different specificity and affinity towards their substrates. However, for most reader domains, their binding specificities were only tested against methylated histones and histone peptides [20]. Only a few were tested for non-histone

methylated proteins and peptides [4]. There has not been a study to investigate non-histone methylated substrates for these reader domains.

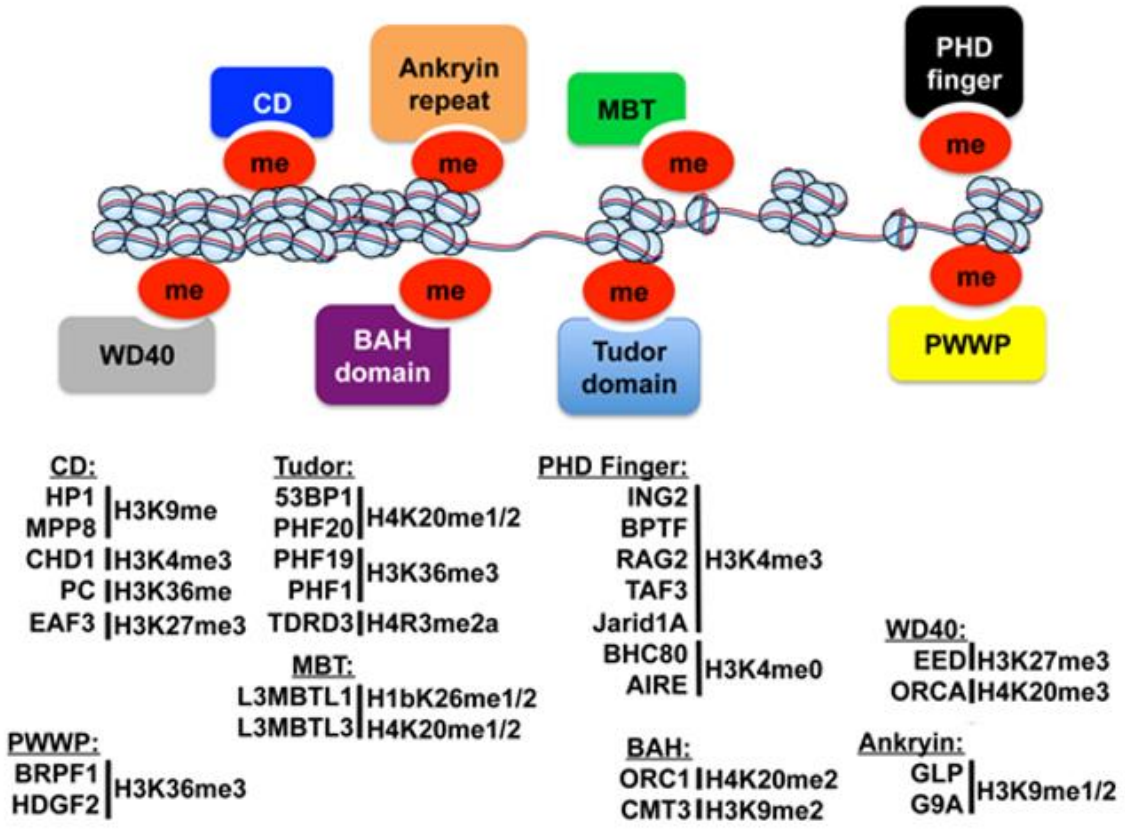

Figure 1.3. Overview of lysine methyl binding domain. Lysine methyl binding domains (KMBD) are mostly known to bind to histone lysine methylated proteins. There are several big family groups of reader groups listed above. They are part of bigger enzyme complexes to read the information and pass it onwards. These KMBDs all have different specificities and different expression rates. Chromodomain uses an aromatic cage to bind to its substrates. 53BP1-tudor domain forms a similar cage structure [81,83,84].

1.3 Histone methylation – methyllysine

With the advancement of genomic techniques, a lot of protein methylation found in the 1990s and 2000s was on histone tails [9]. These histone methylations, along with other PTMs, contribute to the global chromatin organization [13]. Histone lysine methylation contributes to gene repression and gene activation. For example, H3K9 methylation contributes to gene silencing and forming heterochromatin [85]. There are several KMTs which target H3K9, such as SETDB1 facilities mono-methylation, providing a substrate for SUV39H1/2 to di and tri-methylated lysine residues on H3K9 in heterochromatin region [86–88]. Several KDMs can remove the attached methyl-groups to H3K9 [31]. Another well-studied histone methylation is H3K4, where it is found to activate a gene [89]. The degrees of methylation is found at distinct regions of an active gene; Kme1s at the enhancer region, Kme2s at 5' end of the transcribing gene and Kme3s are mostly found at the promoter region of the active gene [90]. Several methyltransferases and demethylases are responsible for adding and removing the methyl-groups. Table 1.1 has a list of KMTs and KDMs for the most commonly known lysine histone modifications. H3K27, H3K36 methylation contributes to gene repression [31]. Whereas H3K79 methylation is associated with active transcription [91]. Moreover, H4K20 methylation is associated with genome stability [92].

Many cancer cells are found to have dysregulated histone methylations[31]. For example, the dysregulation of KMTs at the H3K4 site increases the chance of various cancers, such as infant leukemia [93]. The majority of the infant leukemia resulted from a translocation of the MLL1 gene onto different genes [94]. However, uncontrolled H3K4 methylation by MLL1-AF9 requires the entire MLL1 gene [95]. Efforts are pouring into finding MLL1 inhibitors to combat this subtype of leukemia cancer cells [96]. There are also diseases associated with dysregulated KDMs on the H3K4 site, where overexpression of JARID1 contributes to breast and testicular cancer [97]. Other dysregulated KMT and KDM on other histone sites are summarized in table 1.1.

Since the early 2000s and for the next decade, a lot of focus in the field of protein methylation was on histone methylation [9]. Indeed, histone methylation incorporates a lot of information, and its regulation is still far from being understood.

1.4 Non-histone methyllysine proteins

Early on, most of the methylation sites were found on histones and thought they only occur on histones [31]. However, with the gaining of interest in the protein methylation and their dynamic regulation, other non-histone methylated proteins were identified [9]. One of the most studied oncogenic protein p53 was found to have four methylation sites: K370, K372, K373 and K382 [98]. P53K372me1 increases the stability of p53, whereas p53K373me2 is an inhibitory mark [99]. Different degree of methylation of K370 and K382 determines the stability and activity of p53 [100]. K370 and K382 di-methylation contribute to p53 stabilization via p53BP1 binding protein. Whereas their mono-methylation reduces p53 activity [98]. One fascinating crosstalk between p53 PTM is by methylating K370/382; it blocks the lysine sites from being ubiquitinated and subsequent degradation [101]. The dynamic regulation of lysine methylation contributes to another level of protein regulation. Thus, it begs the question of what the other non-histone methylated proteins are and what are their functions.

The Janus kinase (JAK) – signal transducer and activator of transcription (STAT) signalling pathway play an important role in development and homeostasis in humans [102]. Lysine methylation on STAT3 plays an important regulatory role [21]. There are two key residues on STAT3 for lysine methylation: K180 and K140. STAT3 can be activated via the addition of a methyl-group by EZH2 methyltransferase to K180 [103]. The phosphorylation facilitates the activation of EZH2 at S21 by protein kinase B. STAT3 activity could also be shut down by di-methylation at K140 by SETD7 methyltransferase [104]. This pathway demonstrates the importance of methylation, which can modulate the activity of STAT3. It also highlights the crosstalk between methylation and other PTMs.

However, the disparity in the amount of research between histone and nonhistone methylation is drastic (Fig. 1.4). There are thousands of papers being published on histone methylation where only a few dozen in the recent years on nonhistone methylated proteins and even less for lysine methylated proteins.

There are more and more lysine methylated sites and proteins being discovered. Figure 1.5 shows the biological processes with enriched lysine methylated proteins [105]. These proteins are involved in many vital pathways such as DNA damage repair, apoptosis, metabolism, cell cycle regulation etc. It demonstrates the importance of lysine methylation.

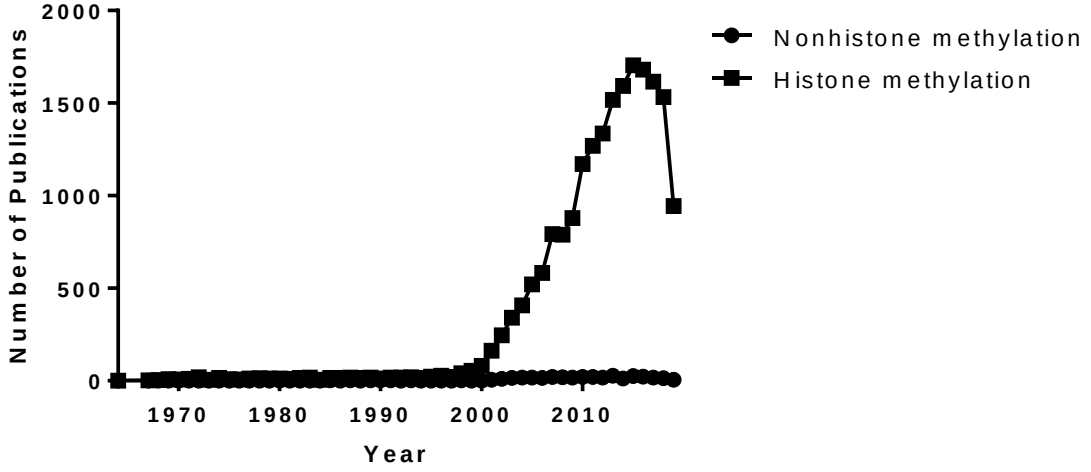

Figure 1.4. Published papers on histone methylation versus nonhistone methylation in the past 50 years. The number of publications in histone methylation versus nonhistone methylation in the past 50 years. The data was obtained through Pubmed. There are about a dozen or so papers published on nonhistone methylation while hundreds and in the thousands of histone methylation papers being published in the past ten years.

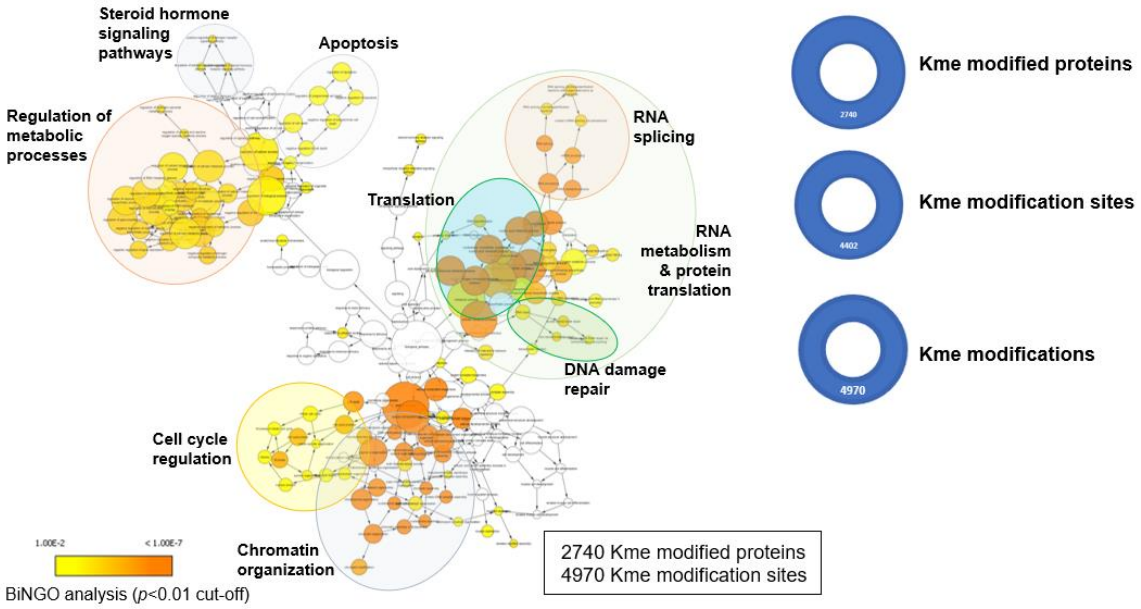

Figure 1.5. BiNGO analysis of major biological processes where lysine methylation was enriched. This BiNGO analysis shows overrepresented Gene Ontology (GO) categories [105]. There are many NHKMPs involved in all the circled pathways: DNA repair, apoptosis, cell cycle regulation, chromatin organization, translation, RNA processing steroid pathways and metabolic processes. Image adapted from Dr. Kyle Biggar.

1.5 Previous studies

Due to the limited amount of methyllysine antibodies, pan-specific antibodies and the huge cost associated with developing these antibodies, another way of enriching for methyllysine proteins must be developed [9]. Previously our lab and other labs have attempted to use naturally occurring lysine methyl-binding domains to enrich for methyllysine peptides [4,106].

In our lab previously, Dr. Liu used a chromodomain (CD), HP1β, to enrich for Kme peptides [4]. HP1β CD binds to H3K9 and H3K23, and it is a good candidate to look at DNA damage repair where protein methylation is involved [82]. Dr. Liu covalently linked his HP1β CD onto magnetic beads to enrich for Kme peptides from digested HEK293 lysate. They used a variety of proteases: trypsin, GluC, ArgC, elastase and chymotrypsin. This enrichment, followed by mass spectrometry analysis, yielded a total of 109 methyllysine proteins [4]. Furthermore, Dr. Liu used a peptide array to characterize HP1β CD specificities. It gave us insights into HP1β CD binding profile and was able to use this to predict novel substrates.

Another group, about the same time as our lab, used 3x MBT domain [107]. They attempted to enrich for as many Kme peptides as possible and to monitor the changes between different treatment groups. Their technique showed better enrichment for Kme peptides from current Kme antibodies [107]. They conjugate their GST-3x MBT domain and GST-3x MBT_{D355N} onto GSH-sepharose beads. They used stable isotope labelling by amino acids (SILAC) of heavy and light conditions to compare different treatments. However, they had only enriched 79 proteins. Within the 79 proteins, 30% were ribosomal proteins and some histone proteins; only about 40 nonhistone methylated proteins.

Another popular method of enriching methylated peptide is the strong cation exchange (SCX) [108]. However, this method works best for methyl-arginine peptides and does not work very well for methyl-lysine peptides. The methylated residue still carries a positive charge and usually not at the C-terminal end of the peptide due to less efficient digestion

by trypsin [109]. One group enrichment using SCX, the majority of the peptides enriched were methyl-arginine peptides [108]. They enriched 250 methylated arginine proteins and only about 11 lysine methylated proteins.

1.6 Rationale and objectives

Currently, there are about five thousand methylated lysine sites recorded, and less are validated. By comparing lysine methylation to tyrosine phosphorylation, we believe there is a lot of non-histone lysine methylated protein that has not been discovered. For tyrosine phosphorylation, there are around 90 tyrosine kinases, 31 protein tyrosine phosphatases and over 100 tyrosine binding domains [110]. More than 47,000 tyrosine phosphorylation sites have been discovered [111]. Compared to around 50 lysine methyltransferases, 30 lysine demethylases and over 100 KMBDs. But, only 4,500 sites recorded, corresponding to 2,740 proteins [24]. Drawing parallel from tyrosine phosphorylation, with a half the writers and similar erasers and readers, we believe there should be at least 25,000 sites.

Many of the currently known lysine methylated proteins are involved in essential pathways in the cell, and when they are perturbed, cells are in stress and can result in diseases [3]. Thus, by expanding on the current lysine methylome, it can give us insight into other components of the lysine methylome and potential hubs of regulation.

One of the impediments in discovering the lysine methylome is missing a high-throughput, high-coverage enrichment method [9]. The low abundance of methyl-lysine within the cell means we must enrich them first before mass spectrometry. Previous attempts using antibody enrichment methods have its downfalls. First, the cost of making a specific antibody to each type of methylation state (Kme1, Kme2 and Kme3) is high [112]. Second, the methyl-groups are small, hydrophobic modifications and a pan-specific antibody for all three will have a weak affinity. Third, even for single methylation state antibodies, their affinity is not great, due to the size and non-charge nature of the

modification [7]. Other approaches using a single natural KMBD were not able to capture all the targets [4,107].

The goals of this thesis were to develop a high-throughput, high-coverage enrichment and identification method to monitor the lysine methylome between different treatment groups. First, I aimed to establish a method to enrich as many Kme peptides as possible. Since using a single domain gave a limited set of data, my objective was to conjugate several different KMBDs onto beads and then combine them into a super-column. Subsequently, I used this super-column to enrich for Kme peptide from HEK293 cell lysate to see how many peptides I can identify in a single experiment.

My Specific aims were to:

1. Systematically characterize the binding specificities of KMBDs to determine the combination to be used for super-column.
2. Test different binding and eluting conditions for Kme peptides onto KMBDs for mass spectrometry
3. Analyze the identified peptides from mass spectrometry data of the enriched Kme peptides

Chapter 2

2 Methods

2.1 Protein expression and purification

Plasmid construction was done by our collaborator from Structure Genomics Consortium at the University of Toronto. Nine lysine methyl-binding domains were received (Table. 3.1). In brief, the p28BIOH-LIC plasmids were constructed from pET28a (Novagen), with the protein of interest, N-terminal AviTag for *in vivo* biotinylation and C-terminal 6X HIS tag. Plasmids were amplified using *E. coli* DH5α via standard transformation protocol. The plasmids were purified using the EZ-10 Spin Column Plasmid DNA Miniprep kit (Bio Basic) according to the manufacturer's instructions. The KMBDs were biotinylated *in vivo* via the bacteria. Another plasmid pBirAcm (Avidity), encodes a biotin ligase, chloramphenicol resistant (10 µg/mL). Both p28BIOH-LIC and pBirAcm were transformed into *E. coli* BL21, with chloramphenicol (10 µg/mL), kanamycin (50 µg/mL) and D-biotin (50 µg/L) in Luria Broth (Bio Basic). Bacteria culture was grown at 37 °C with shaking to an OD600 of 0.6-0.8 absorbance before induction. 1 mM IPTG was used to induce the expression of BirA and KMBD and temperature was lowered to 18 °C to grow overnight. Bacteria were pelleted by 8000g centrifugation at 4 °C for 15 minutes and washed with 4 °C PBS. These bacteria were lysed with 1 mg/mL lysozyme (Bio Basic), with 2% (v/v) Triton-X100 (Sigma), 25 units/mL benzonase (Sigma) and 100 mM PMSF in PBS at 37 °C on a shaker for 40 minutes. The lysate was centrifuged at 25,000g for 30 minutes at 4 °C. His-tagged KMDBs were purified by His-Select Nickel Affinity Gel (Sigma). The column was equilibrated with 20 column volumes of PBS. The supernatants were loaded onto the equilibrated column, washed with 20 column volumes of PBS, 10 column volumes of 10 mM imidazole (Sigma) and 10 column volumes of 20 mM imidazole. The desired proteins were eluted with 250 mM imidazole in 1 mL fractions until no proteins were detected via simple Bradford assay (5 µL protein into 200 µL Bradford assay). The protein quality was analyzed via Coomassie-stained bis-tris

SDS-PAGE. The purified proteins were dialyzed in PBS overnight at 4 °C. The dialyzed proteins were also analyzed with a western blot to see if they are biotinylated.

2.2 Bradford Assay

Concentrated Bradford solution was purchased from (Bio-Rad), the working solution was 4-part water, 1-part Bradford concentrate. Protein standard were diluted from 2 mg/mL albumin standard (Thermo) to 1000, 750, 500, 250, 125, 62.5, 31.25, 0 µg/mL albumin. The assay was done in 96-well plates, 200 µL of Bradford reagent was added to each well. 5 µL of sample and standards were loaded into the wells, with triplicate repeats and taking the average of the three for best accuracy. The plate was shaken on a microplate shaker to ensure adequate mixing. The readings were read on EnVision 2103 Multilabel Reader (Perkin Elmer) at 595 nm.

2.3 Bis-Tris SDS-PAGE and Coomassie staining

Bis-Tris SDS-PAGE was used with MOPS running buffer. For the best protein separation, 10% gel was chosen. Brief recipe: 0.36 M bis-tris, 10% (w/v) acrylamide adjust total volume to 5 mL with water, 25 µL 10% (w/v) APS, 10 µL TEMED. A 20x MOPS running buffer was prepared: 1 M MOPS (Bio Basic), 1 M tris-base (Bio Basic), 69.3 mM SDS (Sigma), 20.5 mM EDTA (Bio Basic). 1x running buffer was diluted on the day of gel and used for only one run. The Coomassie blue staining recipe was composed of 45% (v/v) methanol (Sigma) with 0.05% (w/v) Coomassie Brilliant Blue R-250 (Bio Basic) and 10% (v/v) acetic acid (Sigma) and 45% (v/v) dH_2O. The destain composed of 10% (v/v) acetic acid, 20% (v/v) methanol and 70% (v/v) dH_2O. The gels were running at 120V for 50 minutes (Bio-Rad Mini-Protean) in 1x MOPS. Gels were stained with Coomassie blue for 30 minutes on a shaker, and destained overnight with Kimwipes (Kimtech) to soak the excess dye. Images were taken via ChemiDoc™ XRS+ system (Bio-Rad).

2.4 Western blot

Samples were prepared, and an equal amount of protein was loaded into each lane following the bis-tris SDS-PAGE method. The proteins were transferred onto polyvinylidene fluoride (PVDF) membrane (Sigma) pre-activated with methanol. The transfer cassette was placed in a cold room with an ice bath on the outside to cool down the transfer buffer. Transfer performed at 100V for 1 hour in 1x transfer buffer: 25 mM Tris, 190 mM glycine and 20% (v/v) methanol. After transfer, the PVDF membrane was washed in TBS-T (20 mM Tris, pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20, all chemicals from Sigma) for three times for 5 minutes each time. The membrane was blocked in 15 mL of 5% (w/v) skim milk powder (Bio Basic) or 3% (w/v) BSA (Sigma) in TBS-T for 1-2 hours at room temperature with shaking. The membrane was incubated with primary antibody diluted in the blocking buffer at 4 ℃ overnight. The next day, the membrane was washed with 20 mL TBS-T three times for 5 minutes each time. The membrane was incubated with 10 mL of secondary antibody crosslinked to horseradish peroxidase (HRP) diluted in blocking buffer for 30 minutes at room temperature with shaking. The membrane was washed three times again for 5 minutes, each with 20 mL of TBS-T. For detection, using 5 mL of our in-house HRP substrate at room temperature for 2 minutes. Solution A: 2.5 mM Luminol, 400 µM p-Coumaric acid and 100 mM Tris-Cl, pH 8.8; Solution B: 0.02% (v/v) hydrogen peroxide in 100 mM Tris-Cl, pH 8.8 (all chemicals purchased from Sigma). Images were taken via ChemiDoc™ XRS+ system.

2.5 Peptide membrane array synthesis

The methods used here were previously described methods[4]. In brief, the peptide membrane was synthesized in-house from 24 cm Whatman 1 filter paper (Whatman) to make them amino-functional; they were cut into 12 cm by 8 cm rectangular shapes to fit the Intavis MultiPep SRi peptide synthesizer. The peptides were chosen from known and predicted methylated peptides from the human proteome. The membranes were synthesized according to the manufacturer's instructions using Fmoc (N-(90fluorenyl) methoxycarbonyl) chemistry[113]. In brief, the peptides were synthesized from C-terminal

to N-terminal due to the amino-functional processed filter membrane. A total of 600 spots were available to be synthesized into any peptide; each amino acid was spotted twice to ensure no miss dotting. After membrane synthesis was completed, the membrane needed to go through post-synthesis treatment. First, this involves a TFA cocktail: 95% (v/v) trifluoracetic acid (TFA), 3% (v/v) tri-isoproplsilane (TIPS) and 2% (v/v) dH_2O (chemicals bought from Fisher). Incubate the newly synthesized peptide membrane with 100 mL of 50% (v/v) TFA cocktail and 50% (v/v) dH_2O for 1 hour with no shaking, and this cleaves the protected groups on the peptides. This step is important, and the TFA cocktail will make membrane fragile, too much shaking can break the membrane. Pour off the TFA cocktail carefully and incubated with 100 mL of 100% TFA cocktail for 30 minutes, no shaking. After, the membrane was washed with 100 mL of dichloromethane (DCM), three times for two minutes each time. Followed by washing the membrane with 100 mL of DMF, three times for two minutes each time. Finally washed with 100 mL of ethanol, three times for two minutes each. Let the membrane air dry and stored in darkness in -20 °C.

2.6 Peptide membrane binding and analysis

The processed membrane was taken out of -20 °C and incubated with 40 mL of 100% ethanol for 1 second, and it was diluted with 40 mL of dH_2O for 15 minutes at room temperature with shaking. After the membrane was washed 60 mL with 100% dH_2O, three times, each for 15 minutes with shaking at room temperature. After the membrane was incubated with 60 mL of PBS-T, washed three times, each for 5 minutes with shaking. The membrane was blocked with 5% (w/v) BSA in PBS-T overnight with shaking at 4 °C. The membrane was blotted with the protein of interest at a pre-determined concentration in 40 mL of 5% (w/v) BSA in PBS-T overnight at 4 °C with shaking. The next day, the primary antibody was diluted into the blocking buffer and incubated for two hours with shaking. The membrane was washed three times, with 40 mL of PBS-T for 5 minutes each. The secondary antibody was diluted in blocking buffer for one hour with shaking at room temperature. If no primary antibody required, one can

proceed directly to the secondary antibody. For HRP linked secondary antibody, the method was covered in 2.4. The image was saved in a TIFF file and transferred to the analysis computer. The intensity of each spot was read by Array-Pro Analyzer 4.5 (Meyer Instruments). It analyzed the intensity of each spot, and the program detected the location of the spots and numbered them following the synthesis file. The darker the spot, the more protein it bound, but we could not determine the K_D from the peptide array. The intensities were compared with known substrates to the binding domain, for example, H4K20me2 for the 53BP1-tudor domain.

2.7 In-solution peptide synthesis

The in-solution peptides were synthesized either on Tentagel resin (Intavis) or Wang resin (Sigma) on the MultiPep SRi peptide synthesizer (Intavis). It also used the Fmoc chemistry[113]. These peptides were also synthesized from the C-terminal to the N-terminal. The yield for each peptide was about 2 µmol. However, the uneven distribution of resin could result in varying yield of the final peptide amount. These peptides were mostly used to track binding to protein and fluorescence polarization; thus, at the end of their N-terminal, they were labelled with either biotin (Sigma) or 6-carboxyfluorescein (Sigma). After the synthesis, the peptides were cleaved with the TFA cocktail mentioned in 2.5. The peptides were precipitated with cold-ether and centrifuged at 1000g for 5 minutes. They were washed three more times with cold-ether and dried overnight in the fume hood. The peptides were stored in darkness at 4 °C. The peptides can be resuspended in 100 µL of water or DMSO depending on their solubility.

2.8 Mammalian cell culture

HEK293 cells were cultured in DMEM – high glucose (Sigma) supplemented with 10% (v/v) FBS (Wisent), 100 U/mL penicillin-streptomycin (Gibco), 2 mM L-glutamine (Wisent), 1 mM sodium pyruvate (Gibco) at 37 °C humidified incubator (Napco) with 5% CO_2. Cells were grown in T-75 flasks (Thermo) to about 80% confluent and passaged to avoid senescence. The cells were washed carefully with pre-warmed 37 °C PBS

(Wisent) once and quickly aspirated. Cells were incubated with 10 mL of 37 °C PBS for 3 minutes and shaken to dislodge cells from the flasks. Cells were collected into a 15 mL falcon tube and centrifuged at 1000 RPM for 3 minutes. The supernatant was aspirated, and the cell pellet was resuspended in 1 mL of media. The cell pellet was pipetted several times to break cell clumps to allow the cells to separate into singlets. A hemocytometer checked the quality of separation under the microscope. Cells were passaged 1:5 to remain at a higher rate of cell growth.

Cell lysis for mass spectrometry analysis was done by washing 80-90% confluent HEK293 cells with warm PBS, followed by a three-minute incubation of PBS. The flasks were shaken to release the cells from the bottom. Cells were collected into a 15 mL falcon tube and centrifuged at 1000 RPM for 3 minutes. The supernatant was aspirated and resuspended in 1 mL of PBS and transferred into a 1.5 mL Eppendorf tube. The cells were pelleted at 1000 RPM for 3 minutes, and the supernatant aspirated. The cell pellet was lysed using 1 mL of freshly made 8 M urea with 50 mM Tris-HCl at pH 7.4. The cell lysate was put on an end-over-end rotator for 30 minutes in 4°C cold room to allow lysis to complete. The lysate was sonicated at strength 4-5 for 10 seconds on ice and rest for 30 seconds on ice three times. The lysates were centrifuged at 15,000 RPM on a tabletop centrifuge (Eppendorf) at 4 °C for 1 hour. After the supernatants were precipitated with five times the volume of supernatant with ice-cold acetone/ethanol/acetic acid (50%/50%/0.1% v/v), the solution was stored in -20 °C. Proteins were left to precipitate overnight at 4 °C. The precipitated proteins were pelleted with 5000 RPM at 4 °C for 15 minutes. Pellet was resuspended in freshly made 8 M urea and 50 mM Tris. The final protein concentration was determined via Bradford assay. The proteins were reduced by 5 mM DTT (Sigma) with head-to-toe rotation for 1 hour at 37 °C and followed by alkylation by 14 mM IAA (Sigma) in darkness, where tubes were wrapped with tin foil with rotation for 1 hour. The addition of 5 mM DTT neutralized the unreacted IAA and rotated for 1 hour in darkness. These proteins were followed by trypsin digestion using sequencing grade modified trypsin (Promega), following manufacturer's instructions. The

enzyme to protein ratio used was 1:50 (w/w) and digested at 37 ℃ overnight in a shaker. The peptide mixture was stored at -80 ℃ for future experiments.

2.9 Streptavidin pull-down

For the pull-down experiment, streptavidin sepharose (GE Healthcare) was used. 30 µL of beads were used for one plate of HEK293 cells, about 1-1.5 mg of total protein. All the buffers used here were mass spectrometry grade to minimize mass spec contamination. The streptavidin beads were washed three times with 1 mL of water (Fisher), followed by three washes of 1 mL 50 mM ammonium bicarbonate pH 8. The beads were washed with 1 mL binding buffer 1x PBS pH 7.4 for three times. For each wash, the Eppendorf tubes were put on an end-over-end rotator at 4 ℃ for 3 minutes and centrifuged at 2000 RPM for 2 minutes at 4 ℃. After the beads were washed, 3 mg of the desired biotinylated KMDB was conjugated to the streptavidin beads at 4 ℃ on an end-over-end rotator for 2 hours. The protein-bead conjugate was washed stringently six times to remove any non-biotinylated proteins. After the wash, 10 µL of bead-protein was taken and incubated with a 4x protein loading buffer and boiled at 100 ℃ for 5 minutes. The sample was loaded onto bis-tris SDS-PAGE stained with silver staining to detect if there are any other contaminants bound to the beads.

2.10 Silver staining

The silver staining was used to detect if any trace amounts of non-specific proteins binding to the streptavidin beads. After the bis-tris SDS-PAGE of KMDB on streptavidin beads, the gel was stained for silver staining. This protocol was adapted from Rockefeller University. First, the gel was fixed in 150 mL of 50% (v/v) methanol with 10% (v/v) acetic acid for 30 minutes. Then washed with 150 mL of 50% methanol for 20 minutes. Followed by washing with dH_2O 3 times 10 minutes each. 150 mL of 0.02% (w/v) sodium thiosulfate was prepared and incubated the gel for 1 minute and washed three times with dH_2O for 1 minute each. For the silver reaction, the gel was incubated in 150 mL of 0.1% (w/v) silver nitrate with 0.08% (v/v) formaldehyde (37%) for 30 minutes.

The gel was washed three times in dH_2O for 2 minutes each. For developing, the gel was incubated in 150 mL of 2% (w/v) sodium carbonate with 0.04% (v/v) formaldehyde (37%) until the desired staining intensity. The reaction was stopped by 150 mL 5% (v/v) acetic acid for 10 minutes.

2.11 Mass spectrometry

The samples after KMBD enrichment were desalted via Strat-X columns (Phenomenex) following the manufacturer's instructions. In brief, the column was washed with 1 mL of 100% ACN, followed by 1 mL of 50% (v/v) ACN, after washed 3x with 1 mL MS grade H_2O. The samples were loaded and washed 3x with 1 mL MS grade H_2O. The enriched methylated peptides were eluted with 2x 200 µL of 75% (v/v) ACN in 0.1% (v/v) FA. The eluted peptides were tried via the speed vac and resuspend in 20 µL of 0.1% (v/v) FA. The mass spectrometry's parent mass tolerance was set to 10.0 ppm. Maxed missed cleavage was 3, with non-specific cleavages. The masses of Kme1, Kme2 and Kme3 are 14.02 Da, 28.03 Da, 42.05 Da, respectively.

2.12 Fluorescence polarization assay

Fluorescein labelled peptides were taken out of the 4 °C and carefully diluted to the working concentration. The working concentration can vary, but the fluorescence intensity should be between 20,000 and 60,000 units when measuring emission at 520 nm on Envision 2103 Multilabel Reader (Perkin Elmer). Due to the difficulty of measuring peptide concentration, we decided to dilute the peptides within a range. Using a 384-well plate, the total volume in each well was 35 µL, 30 µL of protein and 5 µL of the peptide. The peptide concentration was the same in all wells, and the protein concentration varied. Dialyzed KMBDs were concentrated to a range from 100 to 200 µM; they were subsequently diluted 14 times, half each time with PBS. The experiment was done in duplicate to minimize technical error. A multichannel peptide (Eppendorf) was used to minimize human error. The proteins were incubated in darkness with the peptide at room temperature for 10 minutes. The fluorescence polarization was measured at the emission

of 520 nm, and it was excited at 480 nm. The binding curve and K_D were determined using Prism 6.0 software (GraphPad Software Inc.) with formulas from Rossi [114].

2.13 Systematic clustering of lysine methylated peptides

All the known reported methylated peptides recorded in PhosphoSitePlus were collected. The known peptides were separated into unique peptide sequences with modified lysine at the center and ± 5 residues. Clusters were formed using an algorithm based on local amino acid similarities. Taking consideration of the charge, polarity and size of the amino acid residue. Two peptides with similar sequences will be grouped into the same cluster. At the time when the clustering was done, there was a total of 2018 unique methylysine peptides. The maximum number of clusters is 2018, where all the clusters have exactly one peptide. Due to the testing limit, we were only able to fit 140 clusters onto one membrane of 600 spots with adequate control peptides. Each of the peptides chosen will have four variants: none, mono-, di- or tri-methylated central lysine.

Chapter 3

3 Results

3.1 Characterizing lysine methyl binding domain specificity

3.1.1 Expression and purification of lysine methyl binding domains

To determine the best enrichment method and combination of domains to use, we first had to determine the specificities and binding patterns of each domain that we are going to use. With the help of Dr. Brown and his team from SGC Toronto, we were able to obtain nine different KMBD plasmids (Table. 3.1). These domains have an N-terminal biotinylation tag and will be biotinylated *in vivo* during the induction phase. All the domains contain a purification 6xHis-Tag at the C-terminal. Using a nickel-NTA column to purify the protein from the whole bacterial lysate. The average yield was about 10 mg of protein from 1 L of culture. From one single nickel-NTA purification, the proteins were above 90% pure. A sample purification of CHD1-CD, MPP8-CD, PHF1-tudor is shown in Figure 3.1. Proteins were dialyzed to remove imidazole and ready for the other experiments.

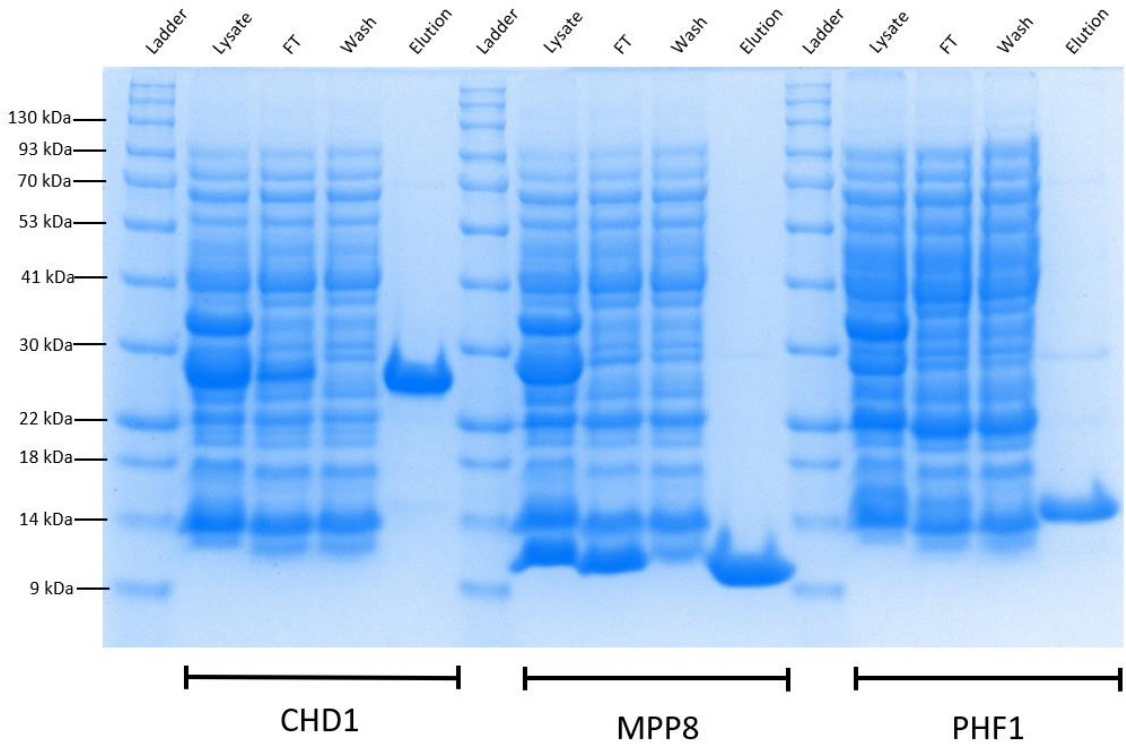

Figure 3.1. Coomassie blue-stained gel of his-tagged lysine methyl binding domain purification: CHD1-CD, MPP8-CD, PHF1-tudor. The proteins were grown in BL21 (DE3) *E. coli*, induced with 1 mM IPTG, grown for 18 hours after induction at 18 °C. Bacterial cells were lysed with 1 mg/mL lysozyme, 1 mM PMSF, 2% (v/v) triton-X. 5 µL of lysate, flow-through (FT) and 10 µL of wash and elution were mixed with 4x SDS loading buffer. Protein ladder is labelled on the left side, and the expected protein molecular weights are CHD1 25.1 kDa, MPP8 10.5 kDa, PHF1 16.5 kDa.

Name	*Domain*	*Residues*	*MW (kDa)*	*MW + His Tag (kDa)*	*Extinction Coefficient (M^{-1} cm^{-1})*	*Site*
53BP1	Tudor like	1480-1606	14.7	18.0	31985	H4K20me2
PHF1	Tudor	30-147	13.2	16.5	31440	H3K36me3
PHF19	Tudor	1-155	17.3	20.6	37400	H3K36me3
L3MBTL2	4x MBT	170-625	52.1	55.4	116155	H4K20me2
CBX7	Chromo	1-70	8.4	11.7	28420	H3K9me3
CBX8	Chromo	1-70	8.4	11.7	26930	H3K9me3
CHD1	2x Chromo	265-450	21.8	25.1	55600	H3K4me2
P100 (SND1)	Tudor	650-910	29.8	33.1	38975	H4K20me3
MPP8	Chromo	58-118	7.2	10.5	19940	H3K9me3

Table 3.1. List of lysine methyl binding domains used in this study. Table showing where the domains were from, the residues sequences, their molecular weight (MW), total MW with his-tag as well as biotinylation tag, their extinction coefficient and also the known histone binding sites for these binding domains [4,84,115–118]. All proteins used are all human proteins. Structural Genomics Consortium Toronto provided plasmids of these proteins.

3.1.2 Biotinylation of KMBD

The domains were used to link to streptavidin beads. Thus, I checked whether these proteins were properly biotinylated. A western blot was prepared for each KMBD using streptavidin-HRP (Thermo 21124) secondary antibody (Fig. 3.2). This western blot confirmed that the *in vivo* biotinylation process was conserved and little to no unintended biotinylated proteins. The exact amount of purified protein with biotin was not calculated; however, from Fig. 3.2, it demonstrated that the majority of the protein of interest was biotinylated.

The protein biotinylation was also tested by incubating L3MBTL2-3xMBT biotinylated proteins with streptavidin beads for them to conjugate for 30 minutes at room temperature. Due to the nature of biotin and streptavidin binding affinity, the conjugate was washed ten times to ensure all the proteins left were biotinylated. The beads conjugate was boiled in a 4x loading buffer for 15 minutes at 100 ºC to denature the streptavidin beads and the KMBD. The denatured sample was loaded onto a gel with original stock protein and the wash fractions (Fig. 3.2). From this, I observed the major band at the desired protein molecular weight. Very little proteins in the washes suggested that most of the bound proteins were biotinylated. The bottom band at around 14 kDa were streptavidin monomers from the streptavidin beads. The streptavidin proteins were released from the beads after they were boiled for 15 minutes in a 4x SDS loading buffer.

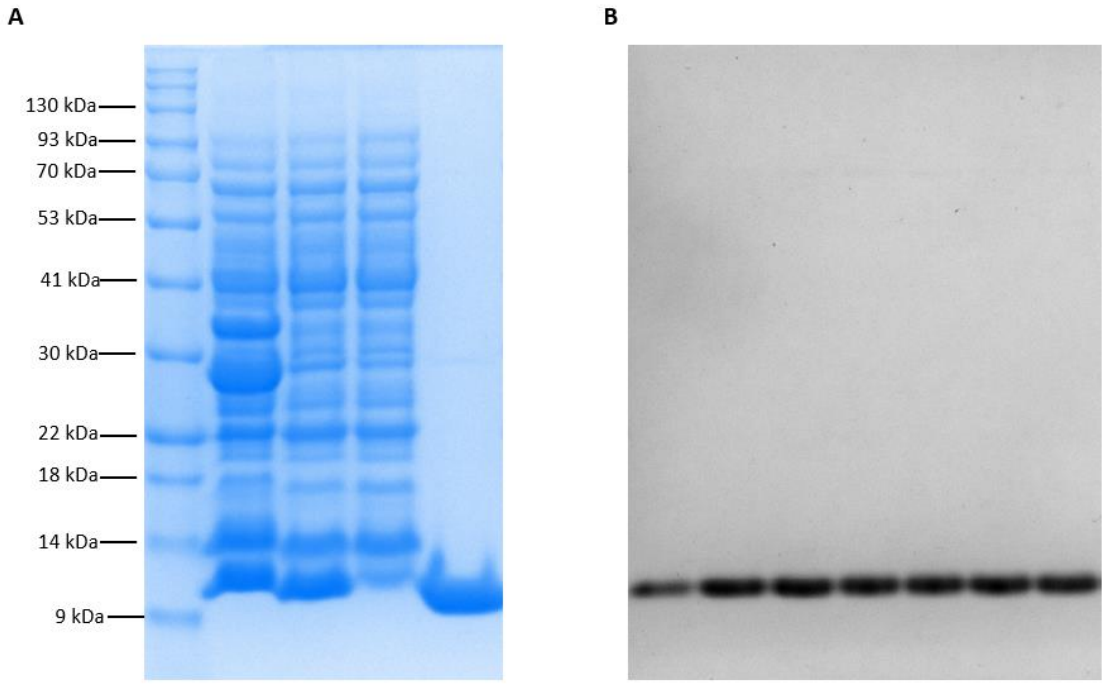

Figure 3.2. MPP8-CD purification and western blot with streptavidin-HRP. MPP8-CD is his-tagged and expressed in BL21 (DE3) *E. coli*, induced with 1 mM IPTG, grown for 18 hours after induction at 18 °C. Bacterial cells were lysed with 1 mg/mL lysozyme, 1 mM PMSF, 2% triton-X. A) The protein purification, as seen in Fig. 3.1. B) western blot with streptavidin-HRP, 2 µL of elution was taken and transferred onto the PVDF membrane. The membrane was blocked with TBS-T for 2 hours and incubated with streptavidin-HRP for 30 mins at room temperature with shaking.

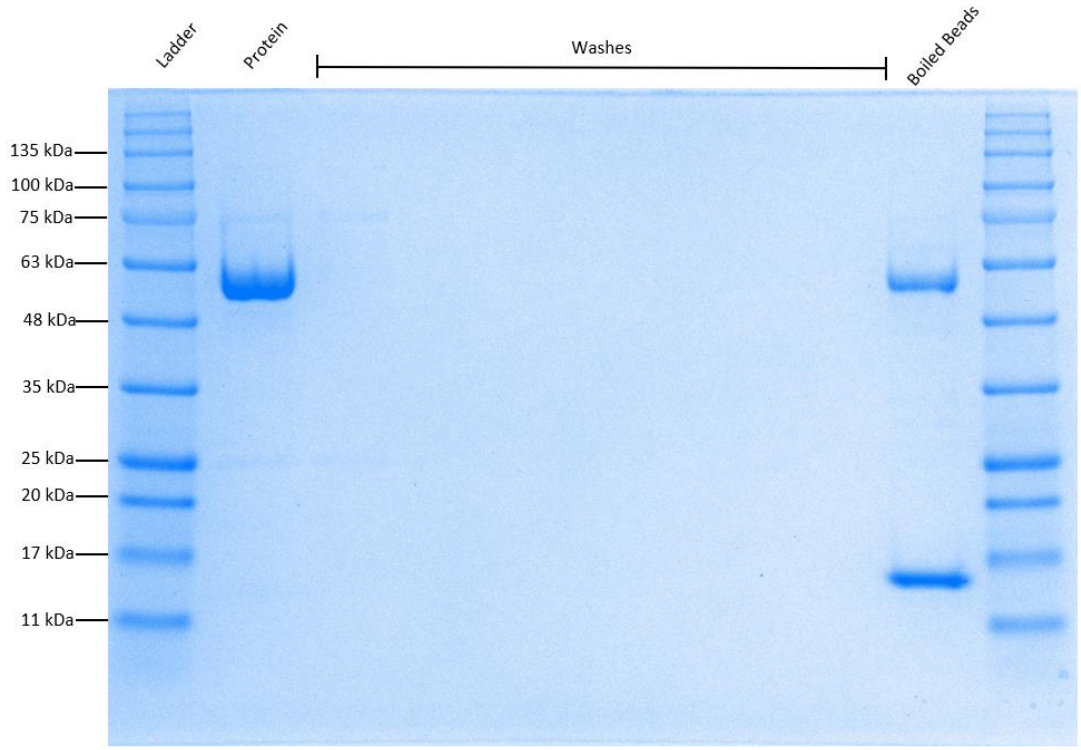

Figure 3.3. Coomassie blue staining of L3MBTL2-3xMBT binding to streptavidin beads. The nickel-IMAC column purified His-tagged L3MBTL2-3xMBT domain. 30 µL of purified proteins were incubated with 10 µL of streptavidin beads for 30 minutes at room temperature. The conjugated protein-beads were washed ten times. After, the conjugated beads were boiled for 15 minutes in a 4x SDS loading buffer. L3MBTL2-3xMBT had an expected molecular weight of 55.4 kDa. The bottom band around 14 kDa were streptavidin monomers from the streptavidin beads.

3.1.3 53BP1-tudor binding to standard peptides

To proceed with further experiments, I had to make sure the protein domains were functionally active. I first used the 53BP1-tudor domain to test its binding to H4K20 peptides. 53BP1-tudor domain is known to bind to H4K20me1/2 and with a 42.5 ± 24.6 µM affinity for H4K20me2 [106]. I synthesized four variants of H4K20, with K20 at the center and seven flanking residues on each side of the central K20, with fluorescein at the N-terminus of the peptides. The peptide sequence is Fluro-GGAKRHR**K**VLRDNIQ.

Fluorescence polarization was performed on the 53BP1-tudor domain with fluorescently labelled H4K20me0/1/2/3 peptides (Fig. 3.4). The purified 53BP1-tudor domains were concentrated to 100 µM and diluted 14 times to 0.006 µM. The experiment was performed on a 384-well dark plate, each well with 30 µL of protein with 5 µL of the peptide. The proteins and peptides were mixed via pipetting and left to incubate for 10 minutes at room temperature in the dark. This experiment was done in two replicates. The fluorescence polarization was measured at the emission of 520 nm, and it was excited at 480 nm. The binding curve and K_D was determined using the formula from [114]. From this result, the biotinylated 53BP1-tudor was active and can bind to H4K20me2 with 40.0 ± 12.1 µM affinity, which was similar to previously reported data. The binding to H4K20me1 was also relatively strong at 41.6 ± 18.1 µM. However, H4K20me0/3 did not have a determinable K_D. The undetermined K_D could be the result of their low-affinity nature, or they did not bind. Previously reported H4K20me0/3 also had no binding to the 53BP1-tudor domain [4].

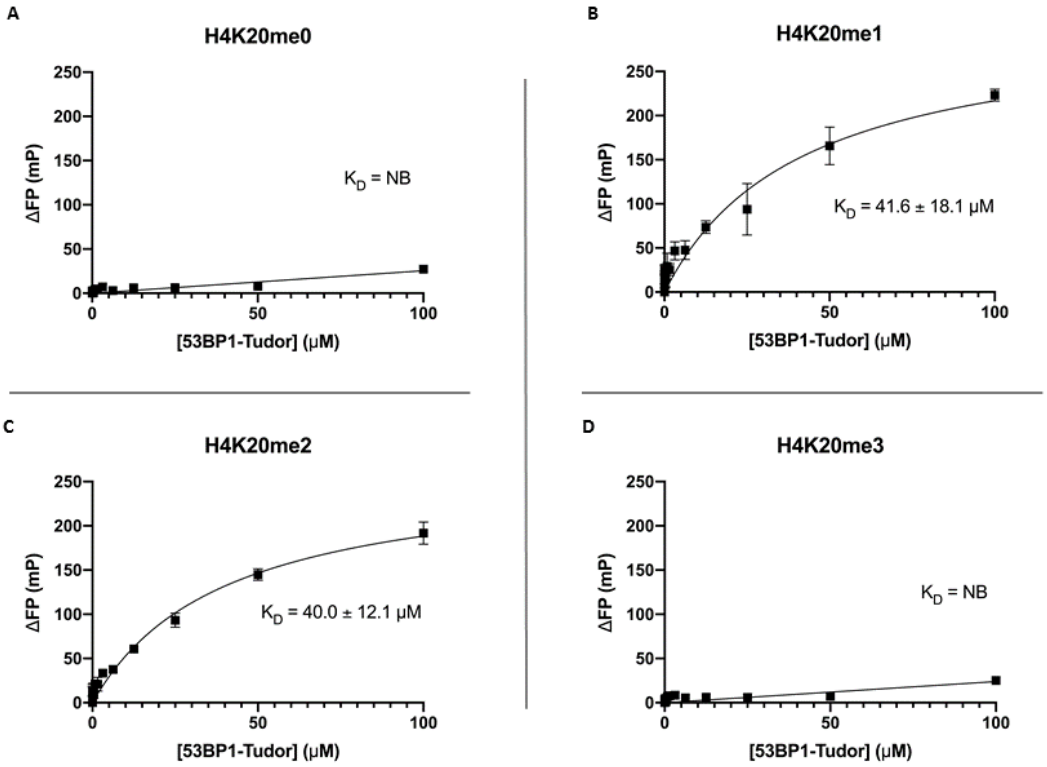

Figure 3.4. The kinetic dissociation constant of 53BP1-tudor with H4K20 peptides was determined by fluorescence polarization. H4K20 peptides were synthesized with fluorescein at the N-terminus: Fluro-GGAKRHR**K**VLRDNIQ. The bolded lysine has four different variants, Kme0/1/2/3. The kinetic dissociation constant (K_D) of the 53BP1-tudor domain to these four H4K20 peptides was determined. The calculations were based on two replicates. The concentrations of the 53BP1-tudor domain used starting from 100 μM, diluted 14 times to 0.006 μM and one control at 0 μM.

3.1.4 Peptide membrane array to probe KMBD specificity

To investigate the specificities and binding pattern of each KMBD in a high throughput manner, I used peptide membrane arrays. The initial array consisted of 70 proteins where their peptides were methylated or predicted to be methylated in cells. Each peptide was 15 amino acid residues in length with lysine at the center and seven residues flanking the lysine from both sides (Table S.1). Each peptide had four different variants: none methylated (Kme0), mono (Kme1), di (Kme2), or trimethylated (Kme3) center lysine residue. The protein peptides were picked based on a previous prediction experiment done in our lab, where they were potential lysine methyltransferase targets. Thus, these peptides were good candidates for the KMBD binding experiments. As well, histone peptides like H4K20 and H3K9 were included as controls.

All the KMBDs were expressed and purified from BL21 (DE3) *E. coli*. The peptide membrane was processed and blocked by 5% BSA overnight at 4 ℃. 1 µM of purified KMBD was incubated with the membrane overnight at 4 ℃. The next day, the binding was detected by streptavidin HRP and chemiluminescence via ECL solutions.

All the KMBDs were probed individually onto the membrane as well as some combination KMBDs to see whether it will impact the binding patterns (Fig. 3.5). To observe the preferred binding patterns, they were all compared to the Km0 variant of each peptide. Here, I assumed that the binding domains have a higher affinity to the methylated peptide than the non-methylated copy. Red means higher affinity compared to Kme0 and green means weaker affinity than Kme0, whereas yellow means neutral. Since these are methyl-binding domains, I assessed more carefully for the interactions that were red and orange.

From figure 3.5, there were some distinct binding patterns for each KMBD. For instance, the 53BP1-tudor domain and CBX7 chromodomain have a strong Kme1 binding preference, whereas PHF1 and CHD1 have more affinity towards Kme3. There were

others with more Kme1/Kme3 preferences like L3MBTL2 and MPP8. SET8 is a methyltransferase, and it was included as a control. I expected SET8 to bind to Kme0 with the strongest affinity since it adds a methyl group to its substrates. Thus, for SET8, the Kme1/2/3 was mostly green. 53BP1-tudor domain was known to bind to H4K20me1 and H4K20me2 [78]. In my experiment, H4K20me1/2 was strongly bound to by 53BP1-tudor domain compared to Kme0. Suggesting that the quality of my peptide membrane array was good, and the results were trustworthy.

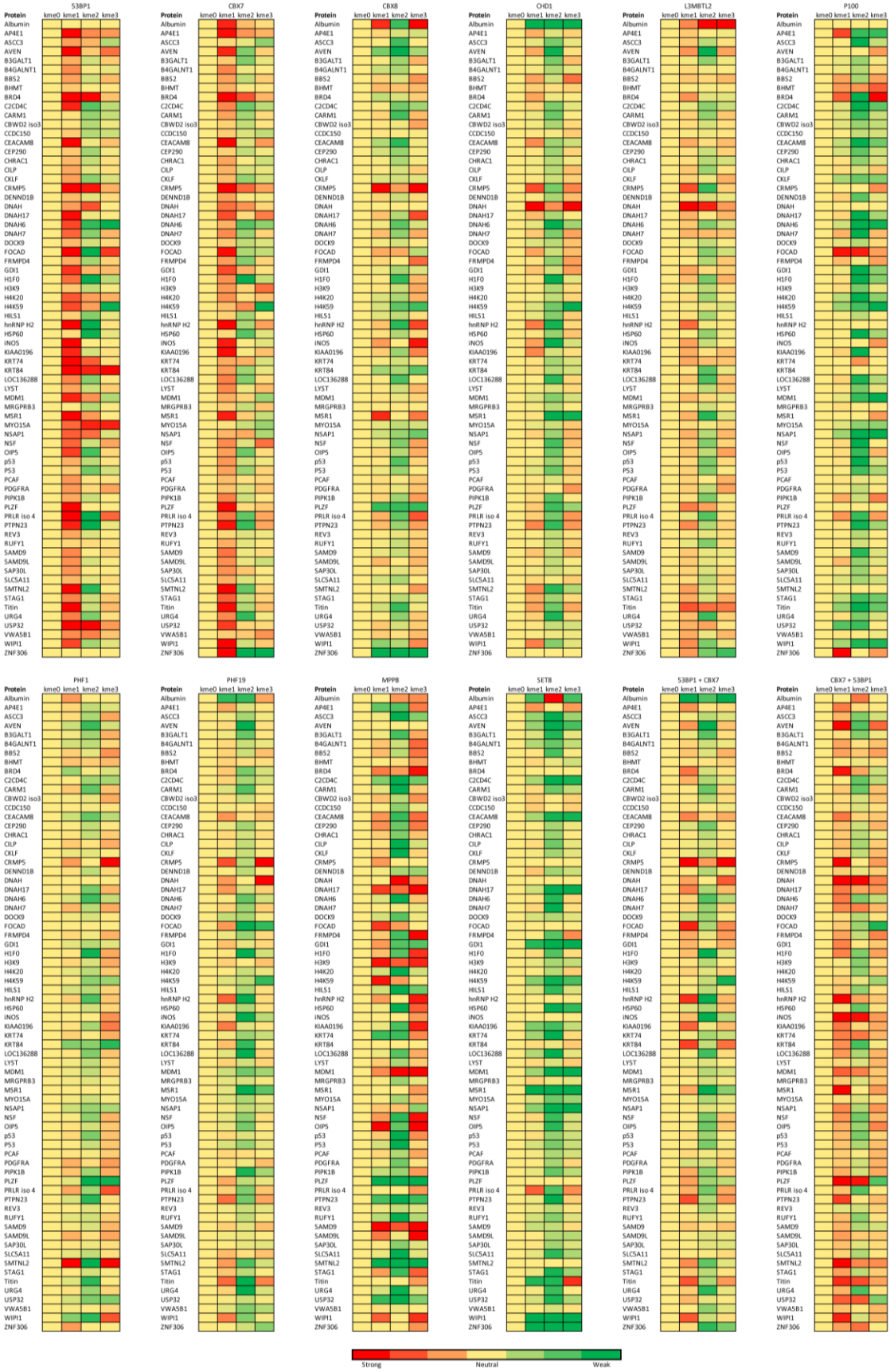
53BP1
CBX7
CBX8
CHD1
L3MBTL2
P100
PHF1
PHF19
MPP8
SET8
53BP1 + CBX7
CBX7 + 53BP1
Protein
kme0 kme1 kme2 kme3
Albumin
AP4E1
ASCC3
AVEN
B3GALT1
B4GALNT1
BBS2
BHMT
BRD4
C2CD4C
CARM1
CBWD2 iso3
CCDC150
CEACAM8
CEP290
CHRAC1
CILP
CKLF
CRMP5
DENND1B
DNAH
DNAH17
DNAH6
DNAH7
DOCK9
FOCAD
FRMPD4
GDI1
H1F0
H3K9
H4K20
H4K59
HIL51
hnRNP H2
HSP60
iNOS
KIAA0196
KRT74
KRT84
LOC136288
LYST
MDM1
MRGPRB3
M5R1
MYO15A
NSAP1
NSF
OIP5
p53
P53
PCAF
PDGFRA
PIPK1B
PLZF
PRLR iso 4
PTPN23
REV3
RUFY1
SAMD9
SAMD9L
SAP30L
SLC5A11
SMTNL2
STAG1
Titin
URG4
USP32
VWA5B1
WIPI1
ZNF306
Strong
Neutral
Weak

Figure 3.5. Lysine methyl binding domain binding heat map to 70 proteins peptide array. This 70 proteins peptide array was derived from a previous prediction experiment, where these proteins were predicted to be methylated within the cells. All KMBDs were expressed and purified from BL21 (DE3) *E. coli*. 1 µM of KMBD was incubated with the blocked peptide membrane at 4 ℃ overnight with shaking. The results of Kme1/2/3 were compared to Kme0. If the bindings were stronger than Kme0, it would be red; if they were weaker than Kme0, then it would be green. The Kme0 peptides were used as the negative control. 53BP1-tudor and CBX7 showed strong Kme1/2 preference. Whereas MPP8 showed Kme3 preference. SET8 was used as a control since SET8 is a Kme1 methyltransferase; it should have the greatest affinity for Kme0 substrates. The majority of the SET8 binding patterns were green.

3.2 Lysine methyl binding domain substrate prediction

3.2.1 Construction of lysine methylated peptide based on a phylogenetic tree approach

Mapping out and predicting KMBD domain substrate specificity requires a wild variety of substrates to be tested. I wanted to make a high-throughput screening method to map out the KMBD binding specificities. I used peptide membrane array for the assay. To pick the most representative group of peptides to test, I decided to use a phylogenetic-like approach. From the PhosphoSitePlus database, we downloaded all the currently validated lysine methylated peptides. With the help of Dr. Li Lei and Eric Liu from our lab, we grouped the peptides into different clusters based on their similarity to each other. In the beginning, each of the 2018 peptides was individually clustered, as we lessen the total amount of total clusters, peptides with similar residues were grouped into the same cluster. Due to the testing limit, we can only spot 600 peptides onto a membrane, and each peptide must have four modification states: none methylated, mono-, di- and tri-methylated. We picked 140 clusters of lysine methylated peptides to be spotted on a peptide array (560 spots), and 40 spots of controlled peptides (Table S.2). This peptide array later will be referred to as the 140-cluster array.

By grouping similar clusters together, I was able to test the widest array of different potential substrates in a systematic fashion.

3.2.2 Results from the 140-cluster peptide array

The peptide arrays were synthesized and probed with all nine of the KMBDs.

In figure 3.6, it showed the raw binding using the 53BP1-tudor domain on the 140-cluster array. Note the positive control H4K20me1/2 was shown in the red box (Fig. 3.6). From this experiment, by looking at the top 5% of highly bound substrates, I observed that the 53BP1-tudor domain bound strongly to Kme1 and Kme2 peptides. It was expected from my previous experiment. However, I did not see much binding of Kme3 peptides. Even though 53BP1-tudor domain bound well to Kme1/2 peptides, some Kme0 peptides were

bound. The binding to Kme0 peptides could suggest that local amino acid composition may play a more significant role in the 53BP1-tudor binding to its substrates (Fig. 3.6).

From the 53BP1-tudor 140-cluster result, I also wanted to assess what the amino acid compositions of the strong binding peptides versus the weak binding peptides. I looked at the residue logos of the top 5% binding Kme1 compared to the bottom 5% of Kme1 peptides (Fig. 3.7). The positions -2 to +3 has the most differences in terms of how the 53BP1-tudor domain bound to its substrates. In the top 5%, I saw several positively charged amino acids and no acidic residues within this region. Whereas in the bottom 5%, there were a lot of negatively charged amino acids within this region. To further validate these findings, I decided to take a look at the structure of the 53BP1-tudor domain [78]. From the 53BP1-tudor domain structure, I saw that the important residues in the binding interface were Trp1495, Tyr1502, Phe1519 and Asp1521. The negatively charged D1521 which complements well to substrate peptides with positively charged residue around the methyl-lysine. The data corresponded very well to the structural data. Thus, this added another level of validation for our peptide membrane array to tease out binding patterns of KMBDs.

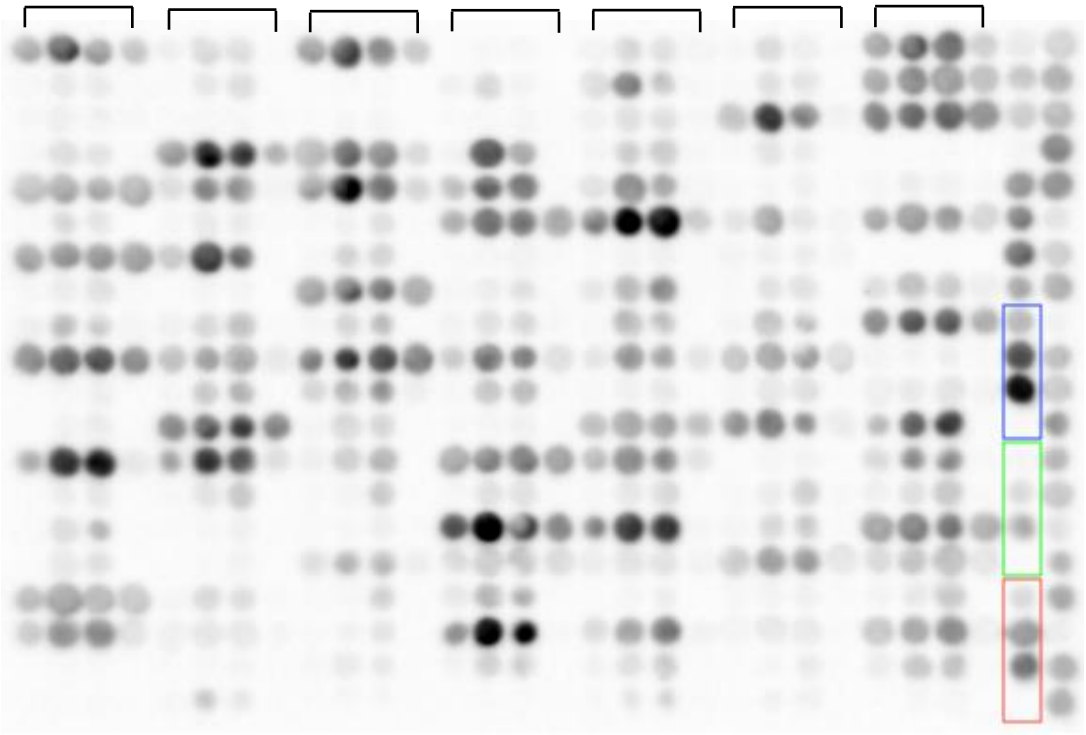

Figure 3.6. 140-cluster peptide array binding with the 53BP1-tudor domain. The 140-cluster peptide array membrane was blotted with 1 µM 53BP1-tudor domain at 4 °C overnight with shaking. Subsequently was bound by streptavidin-HRP and developed with an ECL solution. The darker the spots, the more binding it has occurred. The blue box contains the control peptides H3K9me0/1/2/3 vertically. The green box contains H3K27me0/1/2/3, and the red box contains the control peptide H4K20me0/1/2/3. From left to right, every four dots in a row are Kme0/1/2/3. The last two columns on the right control peptides.

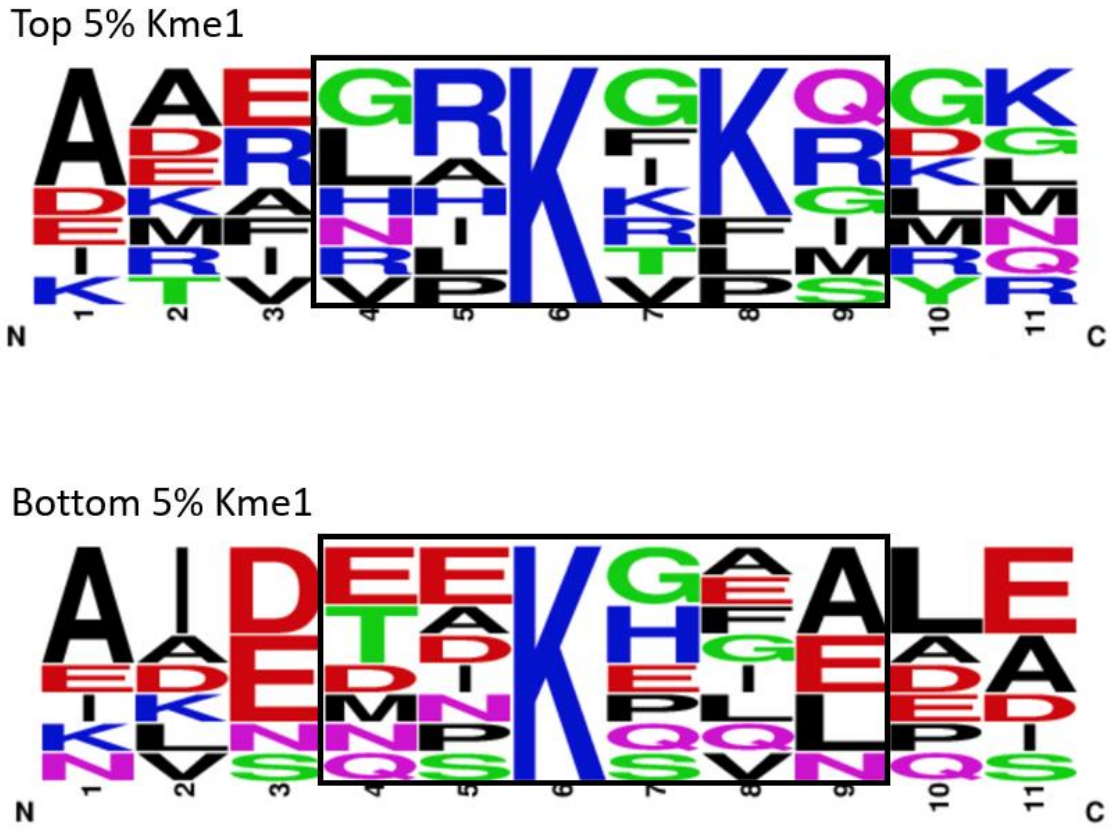

Figure 3.7. Logos of top and bottom 5% of Kme1 peptides to the 53BP1-tudor domain from 140-cluster peptide array. All the Kme1 peptides were ranked based on the binding intensity from the 140-cluster array. The top and bottom 5% were separated, and their sequences were compared using a sequence logo [119]. In the top 5%, Kme1s with strong binding to 53BP1-tudor in the -2 to +3 (4,5,7,8,9) positions contained mainly basic residue and no acidic residue. On the other hand, the bottom 5% Kme1s with little to no binding contained various acidic residues in the -2 to +3 region.

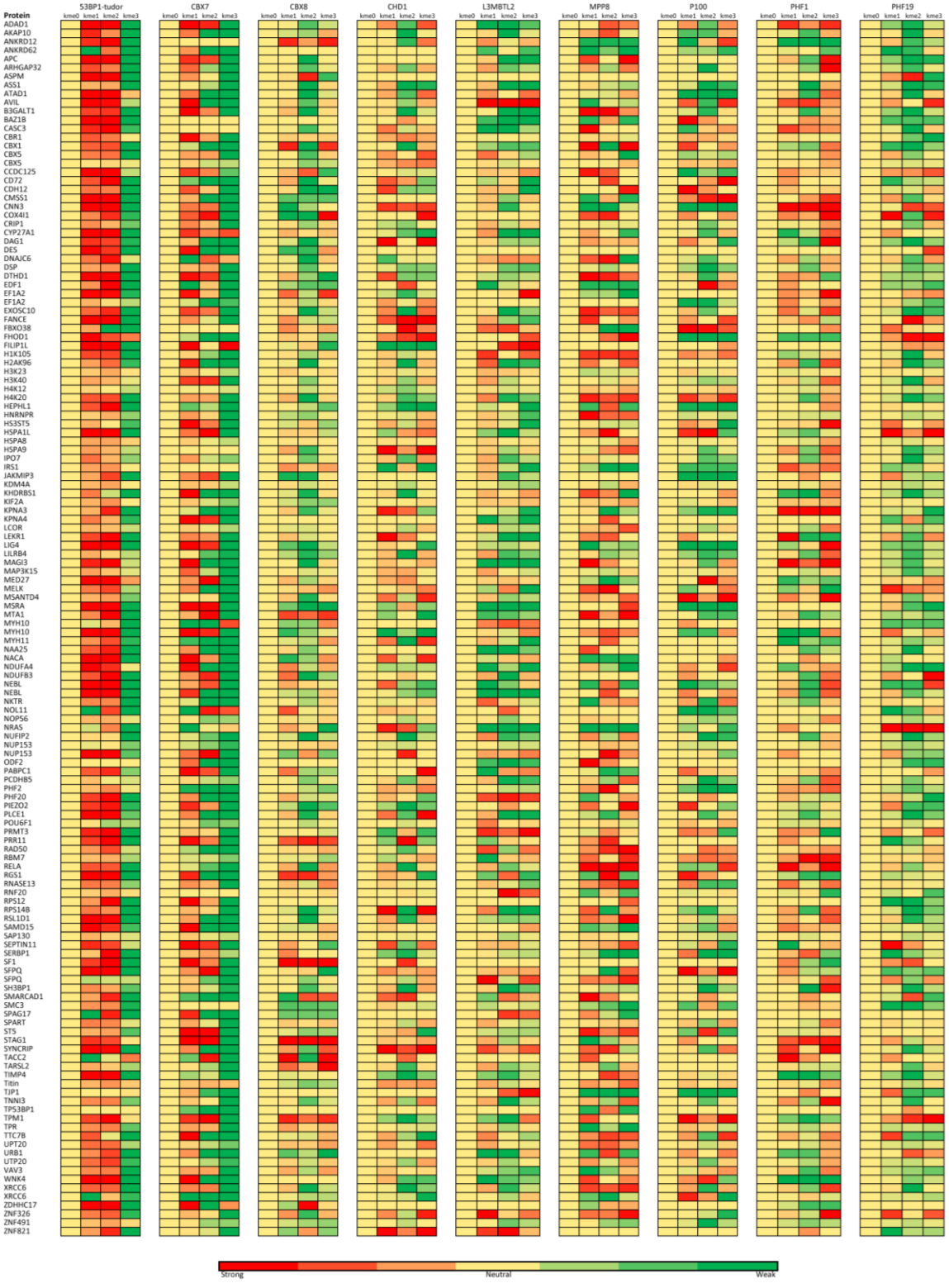
53BP1-tudor
CBX7
CBX8
CHD1
L3MBTL2
MPP8
P100
PHF1
PHF19
kme0 kme1 kme2 kme3
Protein
ADAD1
AKAP10
ANKRD12
ANKRD62
APC
ARHGAP32
ASPM
ASS1
ATAD1
AVIL
B3GALT1
BAZ1B
CASC3
CBR1
CBX1
CBX5
CBX5
CCDC125
CD72
CDH12
CMSS1
CNN3
COX4I1
CRIP1
CYP27A1
DAG1
DES
DNAJC6
DSP
DTHD1
EDF1
EF1A2
EF1A2
EXOSC10
FANCE
FBXO38
FHOD1
FILIP1L
H1K105
H2AK96
H3K23
H3K40
H4K12
H4K20
HEPHL1
HNRNPR
HS3ST5
HSPA1L
HSPA8
HSPA9
IPO7
IRS1
JAKMIP3
KDM4A
KHDRBS1
KIF2A
KPNA3
KPNA4
LCOR
LEKR1
LIG4
LILRB4
MAGI3
MAP3K15
MED27
MELK
MSANTD4
MSRA
MTA1
MYH10
MYH10
MYH11
NAA25
NACA
NDUFA4
NDUFB3
NEBL
NEBL
NKTR
NOL11
NOP56
NRAS
NUFIP2
NUP153
NUP153
ODF2
PABPC1
PCDHB5
PHF2
PHF20
PIEZO2
PLCE1
POU6F1
PRMT3
PRR11
RAD50
RBM7
RELA
RGS1
RNASE13
RNF20
RPS12
RPS14B
RSL1D1
SAMD15
SAP130
SEPTIN11
SERBP1
SF1
SFPQ
SFPQ
SH3BP1
SMARCAD1
SMC3
SPAG17
SPART
ST5
STAG1
SYNCRIP
TACC2
TARSL2
TIMP4
Titin
TJP1
TNNI3
TP53BP1
TPM1
TPR
TTC7B
UPT20
URB1
UTP20
VAV3
WNK4
XRCC6
XRCC6
ZDHHC17
ZNF326
ZNF491
ZNF821
Strong
Neutral
Weak

Figure 3.8. Lysine methyl binding domain heatmap of 140-cluster peptide array. The Kme1/2/3 peptides were compared to the Kme0 peptides. A bigger and more systematic expansion from the 70 proteins array. Here, 53BP1-tudor showed a strong Kme1/2 binding preference. While MPP8-CD appeared to be a balanced binder of all Kme modifications. PHF1-tudor preferred Kme3 peptides. As well as CBX7 preferred Kme1 substrates.

3.2.3 KMBD biding specificity from 140-cluster array versus previous 70-protein array

The 140-cluster array (Fig. 3.8) was a systemic expansion of the previous 70 proteins array (Fig. 3.5). The 140-cluster array peptides were chosen from a database of methylated peptides discovered from human cell lines. From the 140-cluster array, at least one variant of the methylation was observed in a cell. Whereas the 70 proteins array some were real, some were predictions and may not exist in a cell.

The 140-cluster array contained the patterns found in the 70-protein array and magnified it and became more systematic. Both data support each other, and the 140-cluster array gave us a big picture of the protein specificity and preferences.

Note from the figure. 3.6, the binding affinity was not raw intensity, but it was the Kme1/2/3 peptides compared to its Kme0 peptide. Thus, some of the weak binding substrates even though it showed an enriched preference for Kme1/2, but their actual affinity might not be very high. For example, in Fig. 3.6 in blue, it showed three dots, and in green, it showed two faint dots. Even though the blue and green boxes, their Kme1/2 were both red in Fig 3.8, but their actual binding affinity was not the same. 53BP1-tudor bound to the blue box was much tighter than the green box substrates. Another issue is some Kme0s have more binding than other peptide kme1/2s. It was observed again in Fig. 3.6, where the Kme0 of the blue box was darker than the kme1/2 of the green box. Fig. 3.8 gave us an idea of what kind of substrates the protein may prefer; however, it does not tell us about its affinity to the peptide.

The raw binding intensity can infer binding affinity. I could not determine the precise K_D, but I could determine the relative binding strength between the two peptides. The darker the spot, the stronger the protein binds to the substrate. The pattern was important, but we also need to keep in mind about the raw intensity (Fig. 3.9A). The majority of the Kme0/3s were not super high intensity; there were a few that bound well to 53BP1-tudor.

3.2.4 53BP1-tudor domain binding substrate prediction

53BP1-tudor domain shown to have the most diverse binding amongst all nine KMBDs that were tested. Thus, we wanted to use it to make a specificity pattern to predict potential substrates. We first organized the binding data, which allowed us to see the distribution of peptide binding separated into their individual modifications (Fig. 3.9A). Based on the positive control data for H4K20me2 having a net intensity of 107.8, we decided to use 100 as a cut-off to find the enriched amino acids amongst the peptides with at least 1 out of 4 modifications over 100 net intensity (Fig. 3.9C). From this enriched pattern, we saw a lot of basic residues that were enriched compared to the rest of the peptides having less than 100 net intensity. This enriched pattern will be the primary prediction feature. Subsequently, when there is a tie between a couple of peptides, we will need to use the secondary prediction features (Fig. 3.9D). The secondary prediction features were formed by taking a lower cut-off than the primary prediction features. For the secondary features, we took a cut-off of 50 and compared to the peptides with less than 50 net intensity.

Out of the first 140 clusters, we found some interesting binding substrates, and one that caught our eyes was NHEJ1, involved in the DNA damage repair pathway. From WT peptide, it had 4/6 basic residues enriched pattern, and it was a very strong binding substrate (Fig. 3.9). Thus, we wanted to perform a permutation array on this already strong binding peptide. Permutation array replaces each amino acid with 19 different amino acids at that spot to see if there were any critical amino acids involved and also to test our enriched features. In this permutation array (Fig. 3.10), the total protein used was significantly reduced due to the strong binding nature. This permutation array showed that our enriched features were accurate, with only a few places of differences. The interesting discovery was the residue at the -5 positions. By changing the arginine into a lysine, it showed a markedly increased intensity compared to the WT. Also, by changing the serine at the -1 position to arginine, the intensity went up compared to the WT. Out of the rest of the 18 amino acids, only the histidine residue increased binding a bit, as predicted, and everything else was either neutral or worse than WT. More interestingly,

when -2 to +3 region is permutated with an acidic residue, the peptide binding went down. Surprisingly by having proline, leucine and isoleucine, it could also increase the binding affinity towards the peptide.

Thus, I believe a consensus sequence exists for the identification of putative substrates of the 53BP1-tudor domain.

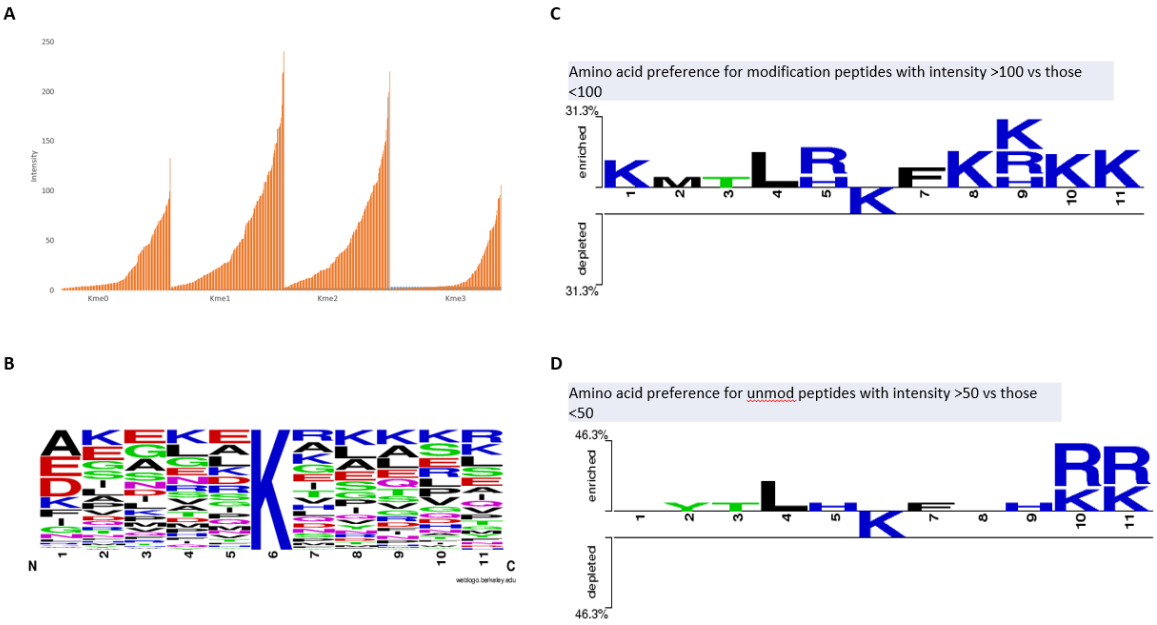

Figure 3.9. Using 140-cluster data to construct binding prediction logos for the 53BP1-tudor domain. A) The binding intensity was plotted based on their modification. For Kme1/2, there were a lot of high scoring peptides compared to Kme0/3. B) General peptide logo of all 600 peptides to see if there were any biases from our 140-cluster array. C) The primary prediction features, residue enrichment from peptides scoring above an intensity of 100 compared to those that were less. Many basic residues were enriched for 53BP1-tudor binding. D) Secondary prediction features were enriched using all peptides, with an intensity score of 50.

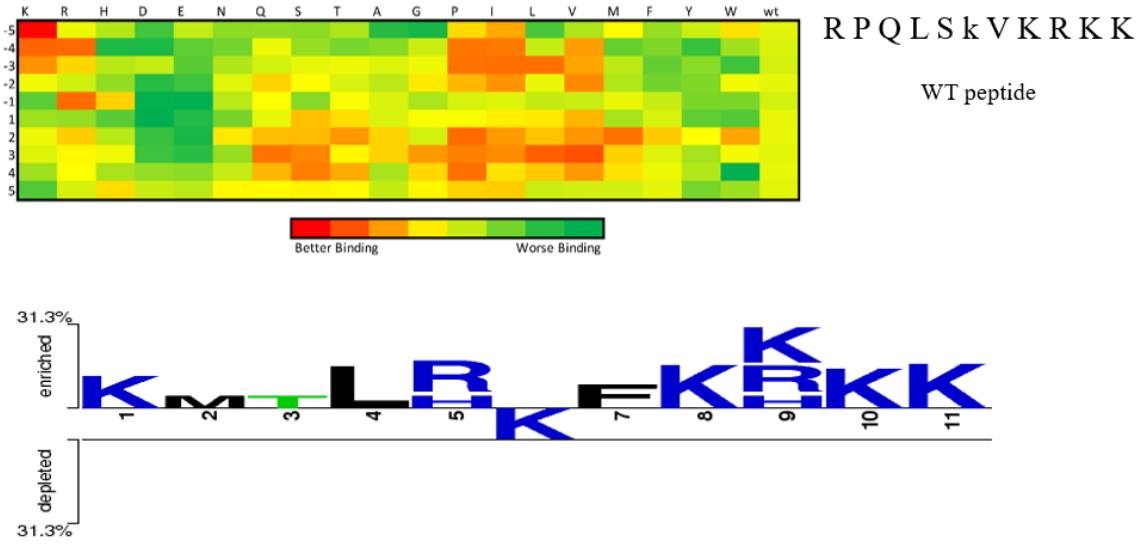

Figure 3.10. Permutation array of NHEJ1, the number one predicted substrate for the 53BP1-tudor domain. NHEJ1 was predicted to be the best substrate for the 53BP1-tudor domain. A permutation array was done on its WT sequence. By changing one amino acid at one time per position, using all amino acids except cysteine. The permutation array results agreed with our primary prediction features. By changing -5 to lysine, the binding affinity went up. As well as changing -1 position to an arginine. Acidic residues within -2 to +3 region decreased the binding affinity to 53BP1-tudor.

3.2.5 53BP1-tudor domain predicted substrates

By using the 53BP1-tudor binding features from above, I wanted to try to predict peptides substrates from the known methylated peptide list from PhosphoSitePlus [24]. There are currently 5,355 recorded Kme sites, and some are not validated. Out of the 5,355 sites, we filtered out the peptide sequences that were the same, to a list of 4,642 unique sequences. The amino acid residues from the peptide were separated from -5 to +5 with the modified lysine at the center, 0. Each position was assigned a score based on the enrichment logo features. First, the peptides with acidic residues from -2 to +3 were separated and moved to the bottom of the list. From our previous experiments, these peptides did not have a good binding with the 53BP1-tudor domain. The primary prediction score came from the 100 or higher net intensity enrichment logos. After the secondary scores came from the 50 or higher net intensity enrichment logo to break the ties in certain situations to rank the peptides. The sum of the score of each amino acid was added and ranked (Fig. 3.11). From this list, I saw there was a group of high scoring candidates and flanked by candidates scoring 0. Several of the known 53BP1-tudor binding substrates were all ranked in the top 15% of the prediction. For example, H3K9 ranked in the top 1%, p53K370 ranked in the top 3%, p53K382 ranked in the top 9%, and H4K20 ranked in the top 14% of all the peptides. With known binding substrates scoring in the top 15%, we have strong confidence in the predicted proteins.

We dove further into the top 100 predicted non-histone binding substrates of the 53BP1-tudor domain (Fig. 3.12). Here, we observed that the top 100 substrates belonged to five biological processes: Transcription regulation, RNA and mRNA processing, cell division, apoptosis regulation and DNA damage repair pathways. The predicted proteins fit into the current knowledge of the 53BP1-tudor domain, which binds to p53 and is involved in DNA damage repair [120].

Several known 53BP1-tudor binding domains scored high on our prediction ranking, and the predicted biological processes fit into the role of the 53BP1-tudor domain. Thus, I am

confident that there will be some valid and novel substrates for the 53BP1-tudor domain identified in this prediction.

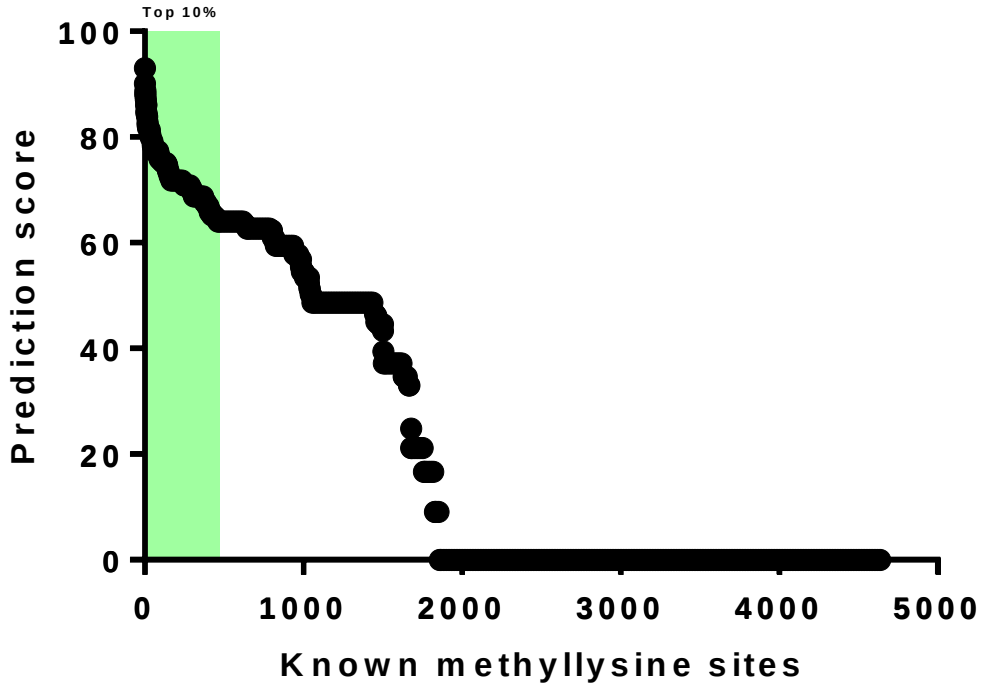

Figure 3.11. Prediction ranking of all known methyllysine sites to a 53BP1-tudor domain based on peptide array data. All of the currently known methyllysine peptides from PhosphoSitePlus were obtained [24]. All peptides were ranked using primary and secondary features developed in Figure 3.9. A peptide with an acidic residue in -2 to +3 region will receive a prediction score of 0, due to the less likely chance of binding. Each residue within the primary and secondary feature receives a score and a weight. In the end, the total score is added based on all residues within the peptide. Many of the known 53BP1-tudor domain substrates were within the top 10%.

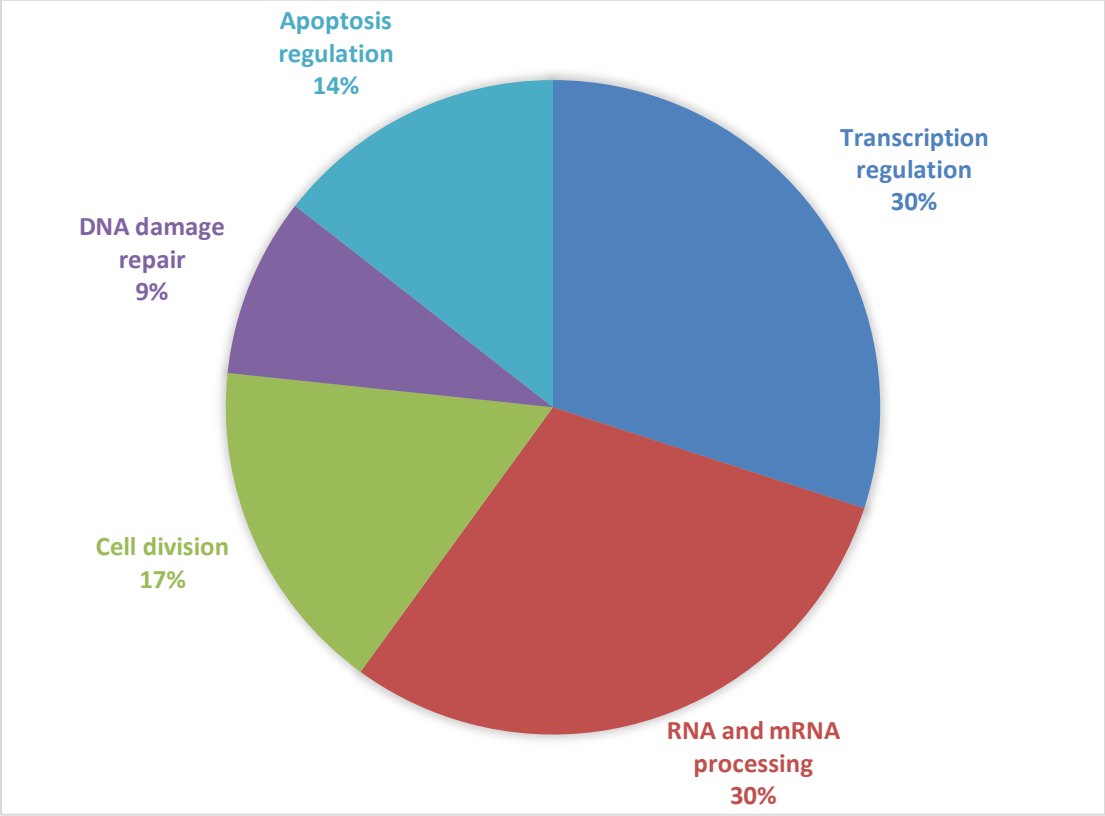

Figure 3.12. Biological pathway analysis of the top 100 predicted 53BP1-tudor binding substrate. Looking at the top 100 predicted substrates of the 53BP1-tudor domain and their biological processes. These five biological processes are vital to the cell but also corresponds to the function of 53BP1 within the cell.

3.3 Lysine methylome identification

3.3.1 Elution conditions for KMBD using standard peptides

Several parameters must be considered for peptide binding and eluting to ensure KMBD-biotin is linked to streptavidin beads during the elution condition but strong enough to release the bound peptides. Additionally, the binding and washing processes must be taken into consideration since the binding affinity for methylated substrates is quite low due to their small, non-polar modification. The final consideration was whether to pull-down whole proteins or digested peptides from the whole cell lysate.

3.3.1.1 Testing KMBD-biotin and streptavidin binding strength

During the enrichment process, we did not want to elute our binding protein off the streptavidin beads. It was critical to identify the conditions where their binding may be disrupted. Here, I used a varying concentration of urea (0 – 8 M) and varying pH (2 – 6 pH) (Fig. 3.13). In this experiment, MPP8-CD-biotin was purified and conjugated to the streptavidin beads. We saw that neither urea nor low pH was able to elute MPP8-CD-biotin off the streptavidin beads. To make sure that our protein was accurately bound to the beads, I boiled the bead conjugate with a loading buffer at 100 ºC for 15 minutes. Only when the beads were boiled, MPP8-CD-biotin was in the solution and at the same molecular weight as the stock protein.

To ensure that no protein was coming off during the harsh conditions, I performed silver staining on p100-tudor binding to streptavidin beads (Fig. 3.14). The proteins were purified and conjugated to streptavidin beads. Excess protein was washed off, followed by several washes to ensure all the nonbinding proteins were washed off. After increasing various concentrations of urea and decreasing the pH, no protein was eluted off the streptavidin beads. The final protein-beads were boiled, and p100-tudor was on the beads. Thus, confirming that pH and urea conditions tested were safe to use and will not elute off any of our KMBDs.

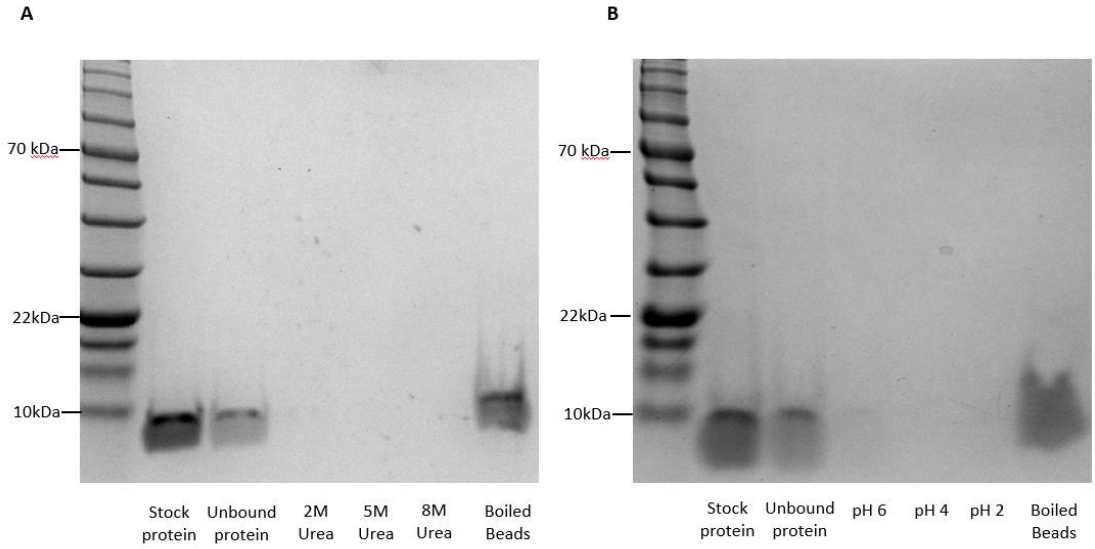

Figure 3.13. Coomassie blue staining of MPP8-CD binding to streptavidin beads at varying urea concentration and pH. Purified MPP8-CD was conjugated to streptavidin beads. 10 µL of streptavidin beads were used and were conjugated to saturation. A) Testing streptavidin and MPP8-CD-biotin binding to the beads with varying concentrations of urea. After MPP8-beads were conjugated, they were rigorously washed before starting the urea test. We saw that at concentrations 2, 5 and 8 M urea, no proteins came off. After the beads were boiled at 100 °C for 15 minutes, the proteins came off the beads. B) Testing streptavidin and MPP8 binding with varying pH. We saw at pH 2, the proteins were still intact and could only be found in the boiled beads. The wavy band was due to an unbalanced pH before running the gel.

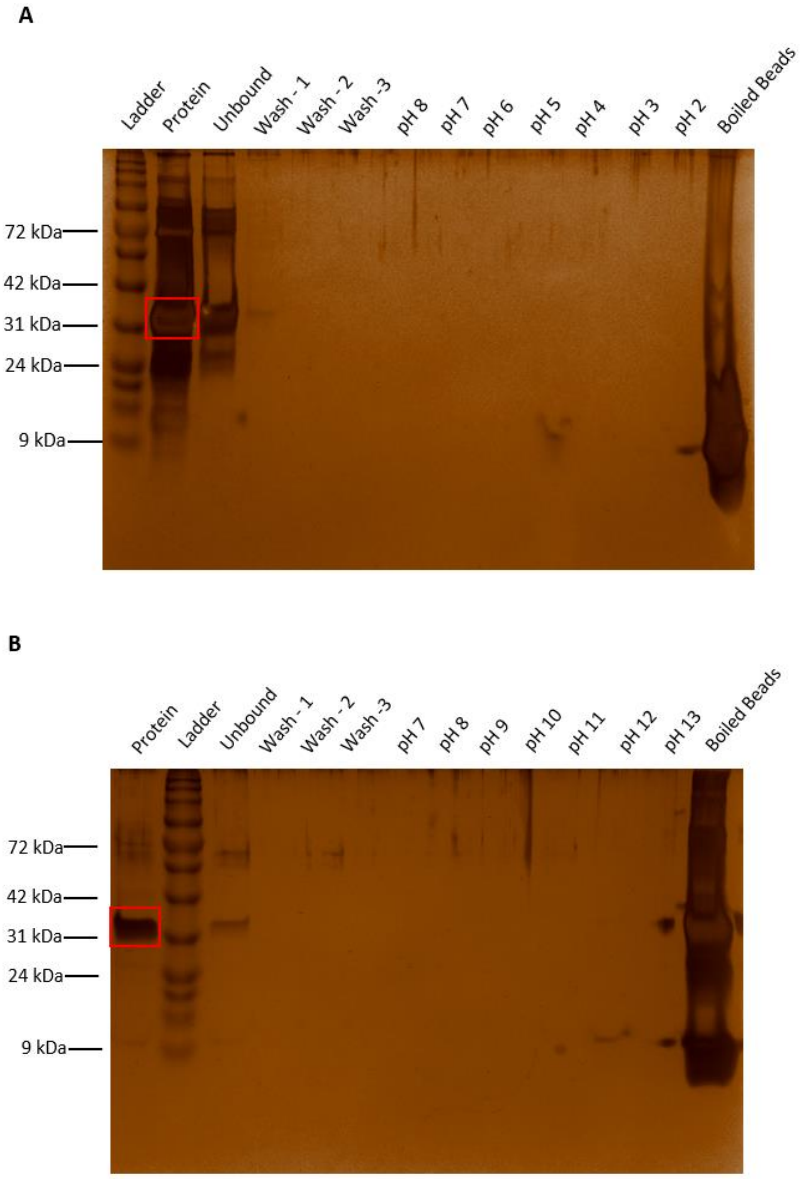

Figure 3.14. Silver staining of p100 binding to streptavidin beads at varying pH and urea concentrations. Testing streptavidin beads binding to the biotinylated protein, we used p100-tudor to test binding. The sensitivity was higher when using silver staining than Coomassie blue-staining. It allowed us to monitor small changes in p100-tudor and if it fell off the beads. The red box represents p100-tudor. A) Varying pH did not remove any conjugated proteins from the beads. B) Varying concentration of urea did not remove p100-tudor from streptavidin beads.

3.3.1.2 pH elution of standard peptides with 53BP1-tudor

After knowing that pH 2 will not elute our binding proteins from the streptavidin beads, I wanted to determine the conditions where I could elute the bound peptide from the KMBDs. I used the 53BP1-tudor domain for testing the conditions needed. Standard peptides of H4K20me0/2 were made with a fluorescein tag on the N-terminus. Peptides were added to the conjugated 53BP1-tudor-streptavidin beads in the dark. They were on an end-to-end rotator for 30 minutes at room temperature for binding. After washing several times, the peptides were eluted with varying pH comparing the two peptides (Fig. 3.15). In this experiment, it showed that the proportion of Kme0 peptides bound to the conjugated beads were much lower than Kme2 peptides. It showed us that the majority of peptides were eluted off 53BP1-tudor at pH 4 and most if not all were eluted at pH 3. Whereas on the basic spectrum, at pH 11, virtually eluted all the peptides off.

Taking consideration that low pH was better for the next step processing for mass spectrometry, I decided to use low pH elution for the rest of the experiments. Interestingly, 0.1% FA, used in MS processing, is about pH 2.7, which was great for eluting these peptides because most of the peptides will be eluted at that pH.

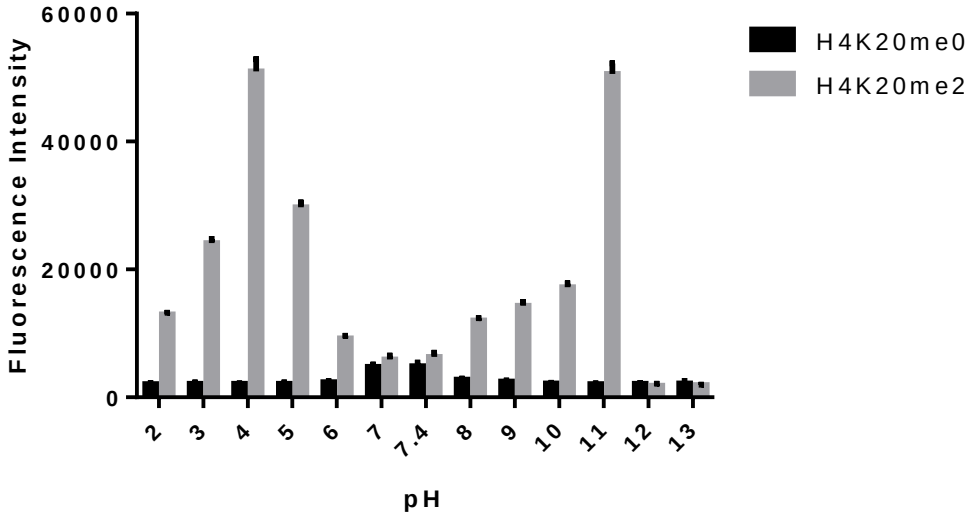

Figure 3.15. Using fluorescein tagged H4K20me0/2 peptide testing elution conditions with varying pH. Purified 53BP1-tudor domain was conjugated to streptavidin beads. Fluorescent H4K20me0/2 were incubated with the protein-beads, with the experiment done in duplicates. The starting point was pH 7.4 and serial addition of lower or high pH until pH 2 and 13, respectively. The pH of the elution was adjusted to pH 7.4 due to fluorescein sensitivity to change in pH. At pH 4 and pH 11 majority of the peptides will be eluted from 53BP1-tudor.

3.3.1.3 Strong cation exchange column

Several groups used strong cation exchange columns (SCX) to enrich their digested peptides for MS processing [108]. However, the majority of the paper where SCX applied was enriching for arginine methylated peptides. Although only a few have used SCX to enrich for lysine methylated peptides, I decided to see if I could utilize SCX to increase the amount of lysine methylated peptides in my enrichment process. Using the fluorescein tagged H4K20me0/2 peptides (Fig. 3.16), I found that no significant differences between the methylated peptide versus the non-methylated peptide. It could be because methyl-group does not affect the charge of the lysine residue being modified [29].

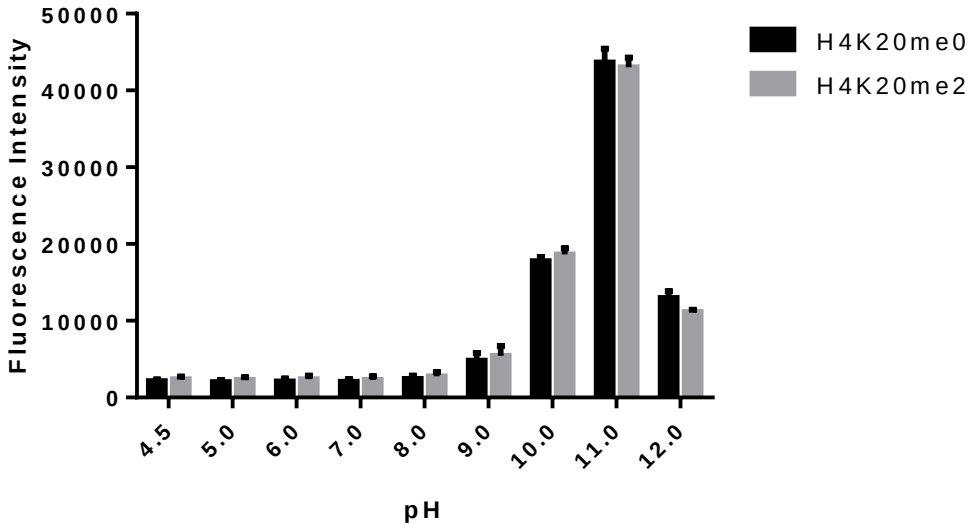

Figure 3.16. Using fluorescein tagged H4K20me0/2 peptide testing SCX conditions with varying pH. Testing if SCX could separate methylated or non-methylated peptides by using fluorescent H4K20me0 versus H4K20me2. The binding condition was with 50 mM sodium acetate at pH 4.5. By increasing pH from 4.5 to 12, we saw that most of the peptides were eluted around pH 11. However, SCX did not distinguish between methylated or non-methylated peptide.

3.3.1.4 Whole protein pull-down

The canonical method of affinity purification of cell lysate was trypsinized whole cell lysate. However, I wanted to try to do an undigested protein pull-down using one of our KMBDs, 53BP1-tudor domain. I wanted to see if and how many intact proteins I could enrich using our domains. I used a protocol by Cold Spring Harbor, a non-denaturing cell lysis protocol with lots of protease inhibitors [121]. Following their protocol, I was able to obtain non-denatured protein lysate from HEK293 cells. I incubated the lysate with conjugated 53BP1-tudor-streptavidin beads for 1 hour at room temperature in PBS pH 7.4. After, it was washed 3x with PBS and eluted with 0.1% FA. The elution was put onto a western blot with a pan-lysine-methyl antibody (Fig. 3.17A). The only detectable band was p53. The elution was trypsinized 1:50 concentration of trypsin to protein at 37 °C overnight with shaking. The sample was prepared and sent to MS. The results were not great, and we were only able to pull-down four non-histone methylated proteins.

Thus, I believe the majority of methyllysine proteins bind with μM to mM affinity [122]. The interactions were weak and transient. Pulling down the intact proteins would be difficult, and the proteins could be lost due to the washing step. Thus, I next decided to attempt enriching methyllysine proteins from trypsinized whole cell lysate.

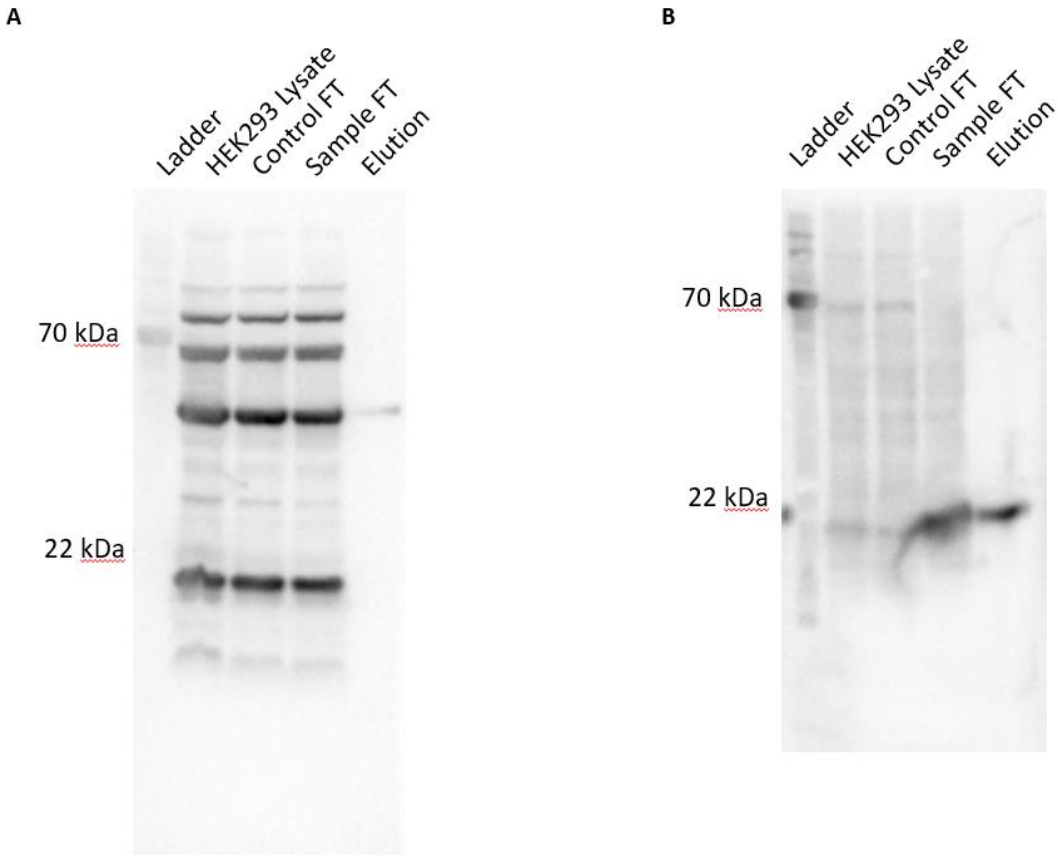

Figure 3.17. Western blot and a far-western blot of intact protein pulldown from HEK293 lysate via the 53BP1-tudor domain. HEK293 cells were lysed in a non-denaturing lysis buffer with protease inhibitors. The purified 53BP1-tudor domains were conjugated onto 30 µL of packed streptavidin beads. The intact proteins from HEK293 lysate were mixed with 53BP1-tudor-beads for 2 hours at room temperature on an end-over-end rotator. Control was streptavidin beads with biotin with HEK293 lysate. A) Using a pan-methyllysine antibody to monitor the elution results. We saw there were lots of methylated proteins in the cell. However, there was only one band around the size of p53 was observed. B) Using 53BP1-tudor-biotin to do a far-western blot to observe where 53BP1-tudor were binding to on the lysates and flow through. From this, we saw there were lots of proteins that 53BP1-tudor could have bound onto from the lysate. However, they were washed away or lost during elution.

3.3.2 Methyl peptide enrichment

Based on the peptide arrays data, we picked five KMBDs: 53BP1-tudor, CBX7-CD, L3MBTL2-3xMBT, MPP8-CD, PHF1-tudor. These domains complement each other in their binding specificity to have the highest overall coverage. The domains 53BP1-tudor and CBX7-CD preferred Kme1/2 peptides, especially 53BP1-tudor (Fig. 3.8). L3MBTL2-3xMBT domain was one of the balanced domains, where it had the potential to bind to all three Kme modifications. Lastly, MPP8-CD and PHF1-tudor domains preferred Kme3 peptides, which will complement the 53BP1-tudor and CBX7-CD. Thus, these five KMBDs had their preferences, and with these preferences, we can have the highest coverage of all three modifications of Kme.

I have done experiments of enriching HEK293 cell lysate with the single domain as well as a combination of five "in a super-column." In these experiments, the HEK293 cells were not stimulated with any stressors. HEK293 cells were grown to a confluence of 80% and lysed via 8 M urea. One replicate was done with two 10 cm culture dishes, to a final protein concentration of 2 mg of total protein. The sample was prepared and processed by mass spectrometry for peptide identification.

3.3.2.1 Methyl peptide enrichment by 53BP1-tudor

I wanted to test a few single domains to monitor the enrichment amount and compare it to the five KMBDs super-column. I used freshly purified and dialyzed 53BP1-tudor domain for this enrichment. The starting protein concentration was 2 mg/mL (111 µM) 53BP1-tudor. I used 120 µL of streptavidin beads mixture, 50:50 solution to beads, with a packed beads volume of 60 µL. The beads were washed vigorously to avoid any contaminants and polymers. After, 1 mg of protein was added to the beads and mixed for 10 minutes at room temperature. Excess proteins were washed off with PBS. To ensure the streptavidin beads were saturated, I added another 500 µg of protein onto the beads. The unbound protein was determined via Bradford assay to be around 500 µg; thus, I concluded that the streptavidin beads were saturated. The experiments were done with two biological repeats. The conjugated protein-beads were divided equally into two

Eppendorfs, each with 30 µL of packed protein-beads. Then, two biological repeats of 2 mg of digested HEK293 cell lysates were incubated with each correctly labelled Eppendorf tube. The tubes were put on an end-to-end rotator for 2 hours at room temperature for peptide binding. They were washed twice with 1 mL of PBS. The peptides were eluted twice with 100 µL of 0.1% FA. Samples were desalted via C18 column following the manufacturer's instructions. The eluted peptides were vacuum dried and resuspended in 20 µL of 0.1% FA for MS identification.

The MS results were analyzed by PEAKS 5.3 program [123]. By combining the identification from both replicates, I was able to identify 84 Kme sites and corresponding to 56 proteins. Out of the 84 sites, 29 were Kme1, 54 were Kme2 and 1 Kme3 peptide (Fig. 3.18A). These Kme sites further correspond to 20 Kme1 proteins, 37 Kme2 proteins and 1 Kme3 protein (Fig. 3.18B). From this experiment, I observed that not a lot of proteins overlap in terms of methyl modifications. There were 30 proteins overlap in both biological replicates, which consists of 54% of the total discovered proteins (Fig. 3.18C). Fig. 3.18D showed the percentages of replicate proteins and sites compared to the total number of proteins and sites found in this experiment. Out of the total 84 Kme sites identified, 77 sites were found to be novel by manually inserting into PhosphoSitePlus database [24]. Furthermore, out of the 56 proteins found, 23 proteins were unreviewed on Uniprot. There were 33 peptides associated with the 23 unreviewed proteins and were classified as a novel since there was no data on these proteins.

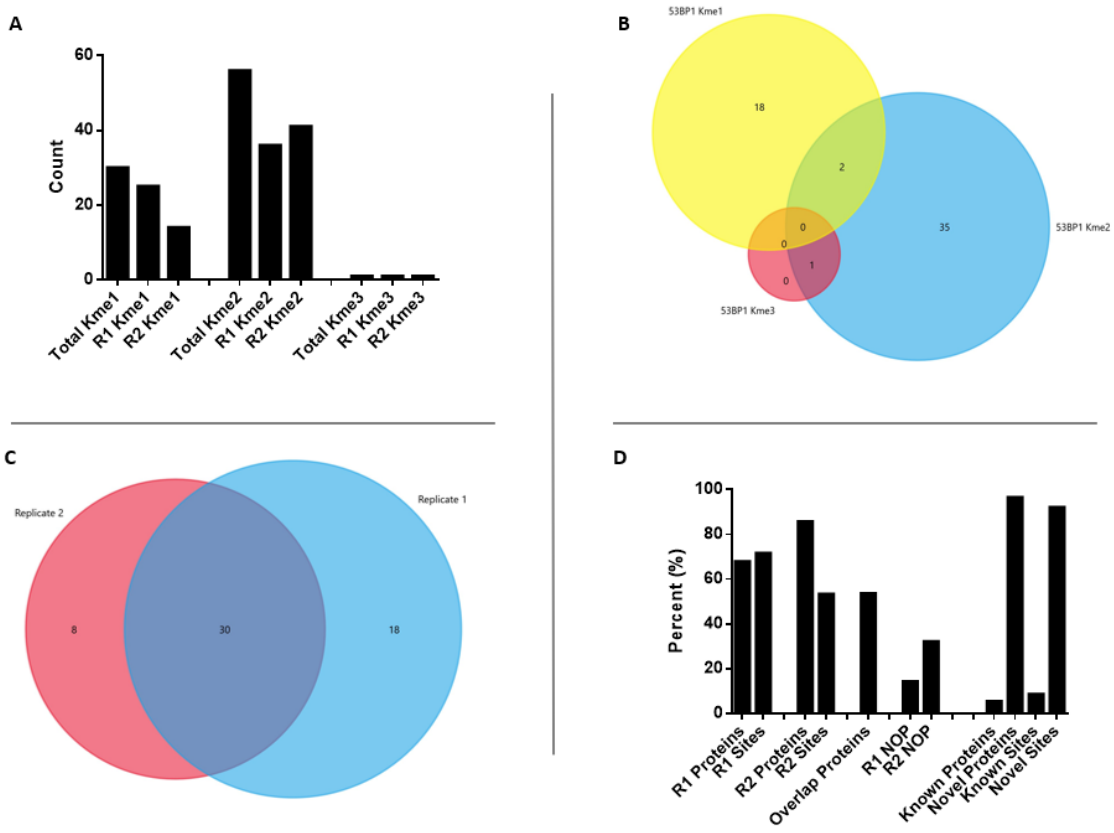

Figure 3.18. Methyllysine sites and proteins identified via 53BP1-tudor affinity purification of HEK293 cell lysate followed by MS. Two biological replicates of HEK293 cells were lysed using 8 M urea and processed for trypsinization. Purified 53BP1-tudor-biotin were conjugated to 30 µL of packed streptavidin beads. Trypsinized HEK293 peptides were mixed with protein-beads for 2 hours at room temperature on an end-to-end rotator. A) Total enriched Kme sites, as well as how many per replicate. B) The Kme sites were corresponding to proteins, a Venn diagram of Kme1/2/3 proteins. Not many proteins pulled out have more than one type of modification. C) The overlap of proteins between the two biological replicates. D) Percentage of replicate enriched proteins and sites compared to the combined proteins and sites across both replicates. As well as the known and novel sites.

3.3.2.2 Methyl peptide enrichment by MPP8-CD

Furthermore, we have tested another single domain enrichment of HEK293 trypsinized lysate with MPP8-CD. MPP8-CD purified well and had great results from the 140-cluster peptide array (Fig. 3.8). Moreover, it had a balanced preference for all Kme substrates instead of Kme1/2 focused like 53BP1-tudor. The same procedure was followed as the 53BP1-tudor domain enrichment. One process that was different from a 53BP1-tudor domain was the MS identification software. We switched from PEAKS to Maxquant software for identification [124]. The results are shown in Fig. 3.17. I was able to identify 48 Kme peptides, corresponding to 33 proteins. Interestingly, for MPP8-CD enrichment, the Kme peptides were balanced, with 17 Kme1, 16 Kme2 and 15 Kme3 peptides. This result also affirms our 140-cluster peptide array experiment. Moreover, all the peptides identified in this experiment were novel. It was particularly strange to not enrich any known sites for MPP8 in two biological replicates. Even though the overlap of the two experiments was low, neither of the two was able to enrich any known sites.

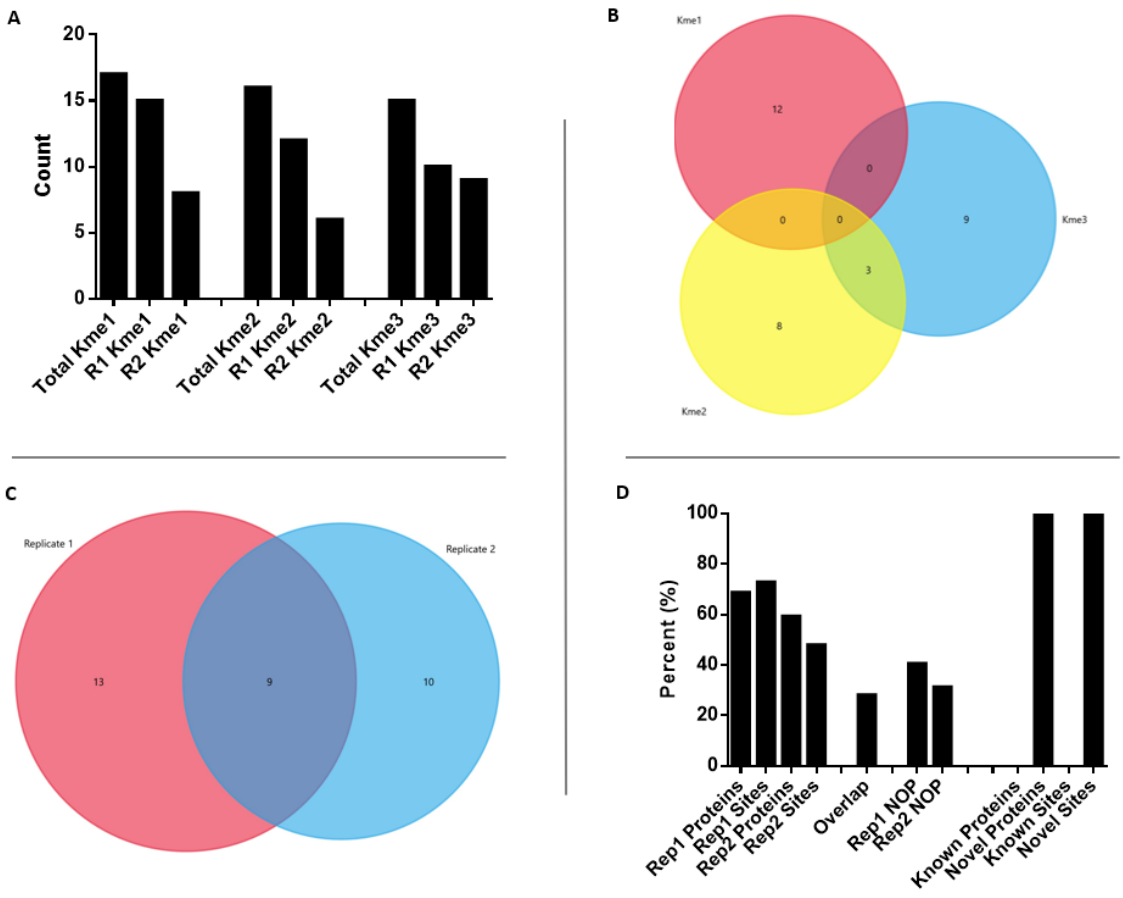

Figure 3.19. Methyllysine sites and proteins identified from HEK293 cell lysate via MPP8-CD affinity purification and MS. Two biological replicates of HEK293 cells were used. Purified MPP8-CD was conjugated to the streptavidin beads for affinity purification. A) The total number of Kme1/2/3 and the number of Kmes for each replicate. B) Venn diagram of Kme1/2/3 proteins and only three proteins overlap Kme2/3. C) The number of proteins being identified in each replicate, with an overlap of about 30%. D) Percentage of proteins and sites from individual replicate comparing to combined proteins and data from both replicates.

3.3.2.3 Methyllysine peptide enrichment by combined super column

After the single binding enrichments, the next step was to test the combination of five KMBDs to see if I could enrich more peptides. To keep everything consistent, I used 30 µL of packed beads. All KMBDs were 1:1:1:1:1 ratio, thus each KMBD was 6 µL of the total packed beads. The KMBDs were conjugated to streptavidin beads individually, and all beads were loaded to saturation. After the protein-beads were diluted and equivalent to 6 µL of packed beads were pipetted into an Eppendorf tube. Two biological replicates were used. The KMBD super-column experiment was performed twice, both were on HEK293 cells, but spanned multiple months apart. These experiments will be referred to as KMBD-1 (Exp 1) and KMBD-2 (Exp 2).

Overall, by combining both experiments, with four biological replicates, 8 mg of total cell lysate and 120 µL of conjugated protein-beads, I was able to enrich 214 Kme sites, corresponding to 148 proteins. Out of the 148 proteins, only three histone proteins were identified, and 145 NHKMPs were identified. From the 214 Kme sites identified, there were 82 Kme1, 103 Kme2 and 29 Kme3 sites (Fig. 3.20D). Those sites corresponded to 67 Kme1 proteins, 84 Kme2 proteins and 27 Kme3 proteins (Fig. 3.20C). From Fig. 3.20C, I observed that several proteins had multiple Kme states, and seven proteins had all three methylation states found on the same residue. These sites were very interesting because it was regulated heavily by various KMT and KDMs.

By looking at the two experiments separately, they did not have a high overlap in terms of proteins enriched in both experiments (Fig. 3.20B). Only 16 proteins were enriched from both experiments, and the majority of the proteins and sites were enriched from the individual experiments. In KMBD-1, it enriched 93 proteins and 134 sites. However, the overlap between replicate 1 and 2 was only about 25% (Fig. 3.20A). On the other hand, KMBD-2 had about 51% overlap between replicates, but the amount of protein and sites were a bit lower as well, 71 and 98, respectively. However, I ran into a problem during the MS injection. The signal from one injection was too strong that not all the sample was

processed. Thus, this injection problem could contribute to the variation between replicates and, as well, between experiments.

Nonetheless, the five KMBD super-column was far superior to the single domain enrichment. With similar amounts of proteins as a single domain enrichment, I was able to get twice as many sites and proteins from the KMBD super-column.

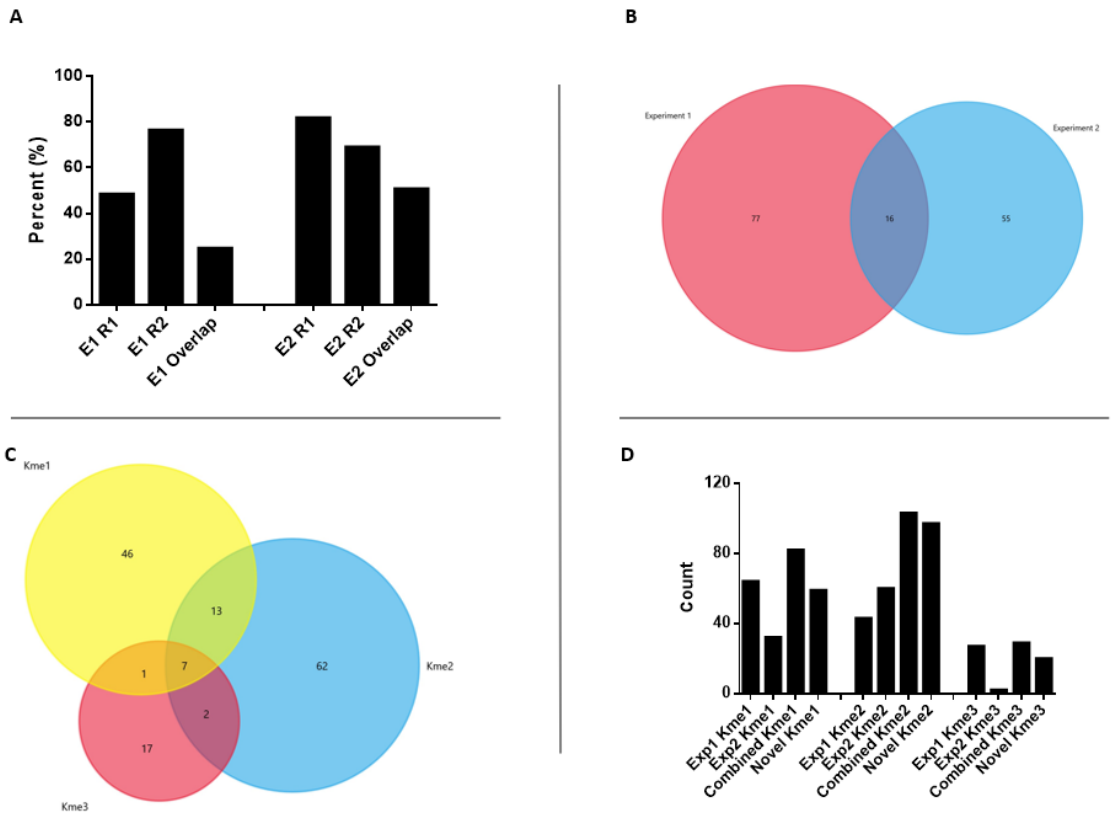

Figure 3.20. Methyllysine sites and proteins identified from HEK293 cells via five KMBD super-column followed by MS. 53BP1-tudor, MPP8-CD, CBX7-CD, PHF1-tudor and L3MBTL2-3xMBT domains were used to form the super-column. Each domain is 6 µL of packed streptavidin beads. Two separate experiments were done E1 (Experiment 1) and E2 (Experiment 2), each with two biological repeats. A) Percentage of proteins per replicate compared to the combined proteins identified from both replicates. E1 had about half the overlap between replicates as E2. B) The number of Kme proteins identified from each experiment with 16 proteins overlap. C) The Kme protein distribution. Interesting, seven proteins contained all three modifications, and some were on the same residue. D) The number of sites identified from both experiments as well as novel sites being discovered.

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs.

Exp	Mod	Uniprot ID	Gene	Site	Sequence	Found (Y/N)
53BP1	Kme2	A0A075		207	GGDRGGFKNFGGHRD	N
53BP1	Kme2	A0A087	recA	601	LQQSGTNKEFYYDIK	N
53BP1	Kme2	A0A0C4		679	QAWTDKQKGLEEKHR	N
53BP1	Kme2	A0A0C4		380	RLQAWTDKQKGLEEK	N
KMBD-2	Kme2	A1L390	PKHG3	953	RPPLQWEKVAPERDG	N
53BP1	Kme2	A6NC78	GOG8I	327	GALQAQVKNNQRISL	N
MPP8	Kme3	A8MV72	YH009	277	LQQGNMRKNMRVLSR	N
53BP1	Kme2	B4DUK8	PRKAG3	263	SPLRSTCKAASSLWS	N
53BP1	Kme2	B7ZBI5	COL20A1	124	RVGLQGPKGMRGLEG	N
53BP1	Kme2	C9JIG9	OXSR1	297	FFQKAKNKEFLQEKT	N
53BP1	Kme2	C9JIG9	OXSR1	295	HKFFQKAKNKEFLQE	N
53BP1	Kme2	D6REQ7	ZGRF1	34	KITHLGNKIQCL___	N
53BP1	Kme1	E9PFJ4	MLIP	181	KSRILLKKEEEVYEP	N
53BP1	Kme1	F8VU51	YLPM1	490	KTWQGHMKATQSYLQ	N
53BP1	Kme1	F8VU51	YLPM1	483	QEYEKQWKTWQGHMK	N
53BP1	Kme2	H0YAN8	ARHGEF10	815	PRLQGIPKVTGRGMV	N
53BP1	Kme2	H0YAW0	TRPA1	562	CKEYLLMKWLAYGFR	N
53BP1	Kme1	H0YES2	DEPDC1	292	FQERCAKKMQLVNLR	N
53BP1	Kme1	H0YFD6	HADHA	111	INMLAACKTLQEVTQ	N
53BP1	Kme2	H0YFD6	HADHA	625	LLTQMVSKGFLGRKS	N
53BP1	Kme2	H0YNE0	MORF4L1	80	VPESRVLKYVDTNLQ	N
53BP1	Kme2	H0YNE0	MORF4L1	88	YVDTNLQKQRELQKA	N
53BP1	Kme1	H3BNL9	TUBGCP4	506	QMQRKHLKSNQTDAI	N
53BP1	Kme1	H3BNL9	TUBGCP4	57	SNQTDAIKWRLRNHM	N
53BP1	Kme2	H7C5H8	CLIP2	23	AMRSCPDKAQDKH__	N
53BP1	Kme2	I3L2Y3	RNF167	112	IPTHDYQKGDQYDVC	N
53BP1	Kme2	I3L520	DNAH2	1125	IDYEGTQKLLALDPS	N
53BP1	Kme1	J3QSZ5	SON	356	GEGLGKNKEGNKEPI	N
53BP1	Kme2	K7EJJ0	ABCA9	23	LCKNCLKKWRMKRQT	N
53BP1	Kme2	K7EJJ0	ABCA9	22	LLCKNCLKKWRMKRQ	N
53BP1	Kme2	K7EJJ0	ABCA9	18	QTWALLCKNCLKKWR	N
53BP1	Kme1	K7EK07	H3F3B	79	REIAQDFKTDLRFQS	Y
53BP1	Kme2	K7EK07	H3F3B	79	REIAQDFKTDLRFQS	Y
53BP1	Kme2	M0QYA0	ZNF649	26	QFLSPAQKDLYRDVM	N
KMBD-1	Kme2	O00148	DDX39A	333	LSRYQQFKDFQRRIL	N
KMBD-1	Kme1	O00571	DDX3X	81	SRSDSRGKSSFFSDR	Y
53BP1	Kme2	O14578	CTRO	653	KAESLSDKLNDLEKK	N

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs (Continued)

MPP8	Kme1	O14734	ACOT8	204	QLQRMEPKQMFWVRA	N
KMBD-2	Kme2	O14979	HNRDL	405	GQQSTYGKASRGGGN	N
MPP8	Kme3	O43151	TET3	1446	QLVFYQHKNLNQPNH	N
KMBD-2	Kme2	O43704	ST1B1	259	SKSPFMRKGTAGDWK	N
MPP8	Kme1	O43819	SCO2	89	KERLQQQKRTEALRQ	N
KMBD-1	Kme3	O60476	MAN1A2	143	EIQTEKNKVVQEMKI	N
MPP8	Kme2	O60522	TDRD6	1930	LSLGVSQKAQESMCT	N
KMBD-2	Kme2	O60673	REV3L	1784	LNRSSVSKEVFLSLP	N
KMBD-2	Kme2	O60673	REV3L	1730	LSPEIFEKSTIDSNE	N
KMBD-1	Kme2	O60870	KIN	135	CKVDETPKGWYIQYI	Y
KMBD-1	Kme3	O60870	KIN	135	CKVDETPKGWYIQYI	N
KMBD-1	Kme1	O60882	MMP20	140	MSSVEVDKAVEMALQ	N
KMBD-1	Kme2	O60882	MMP20	128	TLTYRISKYTPSMSS	N
KMBD-2	Kme1	O60885	BRD4	177	EIMIVQAKGRGRGRK	Y
MPP8	Kme1	O60885	BRD4	349	PAPEKSSKVSEQLKC	N
MPP8	Kme1	O75179	ANR17	1394	EEIEAKNKENFELQA	N
MPP8	Kme1	O75179	ANR17	1254	KLEEIEAKNKENFEL	N
53BP1	Kme1	O75373	ZN737	435	EECGKAFKCFSILTT	N
53BP1	Kme1	O75373	ZN737	432	FKCEECGKAFKCFSI	N
53BP1	Kme1	O75373	ZN737	426	HTGQQPFKCEECGKA	N
KMBD-2	Kme2	O75586	MED6	51	TCNNEVVKMQRLTLE	N
53BP1	kme2	O75592	MYCB2	2885	KKMPLTEPLRGR	N
KMBD-1	Kme1	O75616	ERAL1	126	LGAPNAGKSTLSNQL	N
KMBD-2	Kme2	O75747	P3C2G	978	PDAVTLAKIHRHSGL	N
KMBD-1	Kme3	O76021	RSL1D1	440	KIKEEAVKEKSPSLG	N
KMBD-1	Kme1	O94776	MTA2	173	SDNRNQQKMEMKVWD	Y
53BP1	Kme1	O95628	CNOT4	207	LKNMQCPKPDCMYLH	N
KMBD-1	Kme1	O95801	TTC4	97	FKEKDYKKAVISYTE	N
KMBD-1	Kme3	P00451	F8	1091	SNKTTSSKNMEMVQQ	N
KMBD-2	Kme2	P01270	PTHY	57	ERVEWLRKKLQDVHN	N
KMBD-1	Kme1	P05060	CHGB	455	VQENQMDKARRHPQG	N
KMBD-2	Kme3	P05141	ADT2	52	ITADKQYKGIIDCVV	Y
KMBD-1	Kme2	P05423	POLR3D	231	RDEEEEAKMKAPPKA	N
KMBD-2	Kme1	P09651	ROA1	183	EVRKALSKQEMASAS	N
KMBD-2	Kme2	P09651	ROA1	183	EVRKALSKQEMASAS	N
KMBD-1	Kme1	P0CG12	CHTF8	2	_____MKEPRIFPR	Y
KMBD-2	Kme1	P0DN76	U2AF5	39	RCSRLHNKPTFSQTI	N
KMBD-2	Kme3	P0DP25	CALM3	116	VMTNLGEKLTDEEVD	Y
53BP1	Kme1	P11021	HSPA5	585	DKEKLGGKLSSEDKE	Y
53BP1	Kme1	P11021	HSPA5	591	GKLSSEDKETMEKAV	Y
KMBD-1	Kme3	P11021	HSPA5	585	DKEKLGGKLSSEDKE	Y
53BP1	Kme1	P11142	HSPA8	561	EDEKLQGKINDEDKQ	Y

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs (Continued)

53BP1	Kme1	P11142	HSPA8	159	DSQRQATKDAGTIAG	N
53BP1	Kme1	P11142	HSPA8	567	GKINDEDKQKILDKC	Y
53BP1	Kme1	P11142	HSPA8	557	KATVEDEKLQGKIND	Y
KMBD-1	Kme3	P11142	HSPA8	561	EDEKLQGKINDEDKQ	Y
53BP1	Kme2	P12036	NFH	886	AKAKEPSKPAEKKEA	N
53BP1	Kme2	P12036	NFH	953	EPSKPAEKKEAAPEK	N
KMBD-1	Kme1	P12883	MYH7	1895	QANTNLSKFRKVQHE	N
KMBD-1	Kme2	P13611	VCAN	2698	TSLPIPRKSATVIPE	N
KMBD-1	Kme1	P13639	EEF2	519	KNPADLPKLVEGLKR	N
KMBD-1	Kme1	P13639	EEF2	525	PKLVEGLKRLAKSDP	Y
KMBD-1	Kme2	P13639	EEF2	519	KNPADLPKLVEGLKR	N
KMBD-1	Kme3	P13639	EEF2	525	PKLVEGLKRLAKSDP	N
53BP1	Kme2	P15924	DESP	1054	ANSENCNKNKFLDQN	N
53BP1	Kme2	P15924	DESP	1056	SENCNKNKFLDQNLQ	N
MPP8	Kme3	P15924	DESP	1965	SLFQAMNKELIEKGH	N
KMBD-2	Kme2	P16104	H2AX	76	NAARDNKKTRIIPRH	N
KMBD-1	Kme2	P19622	EN2	242	PKKKNPNKEDKRPRT	N
KMBD-1	Kme2	P21399	ACO1	639	KLFFWNSKSTYIKSP	N
KMBD-2	Kme2	P21549	SPYA	61	YQIMDEIKEGIQYVF	N
MPP8	Kme1	P21817	RYR1	3942	DVIEEQGKRNFSKAM	N
KMBD-1	Kme2	P22059	OSBP	97	AREGWLFKWTNYIKG	N
KMBD-2	Kme1	P22061	PIMT	113	VIGIDHIKELVDDSV	N
KMBD-2	Kme2	P23246	SFPQ	703	EEYEGPNKKPRF___	N
53BP1	Kme2	P23416	GLRA2	271	RRQKRQNKEEDVTRE	N
53BP1	Kme2	P25101	EDNRA	30	MEFYQDVKDWWLFGF	N
MPP8	Kme3	P26583	HMGB2	141	SEQSAKDKQPYEQKA	N
KMBD-1	Kme2	P31150	GDI1	293	YIPDRVRKAGQVIRI	N
MPP8	Kme1	P32297	ACHA3	474	AQEIQQLKRKEKSTE	N
MPP8	Kme1	P32297	ACHA3	476	EIQQLKRKEKSTETS	N
MPP8	Kme1	P32297	ACHA3	466	NEAKEEQKAQEIQQL	N
KMBD-2	Kme1	P32302	CXCR5	339	DLSRLLTKLGCTGPA	N
KMBD-2	Kme1	P32302	CXCR5	50	GPLMASFKAVFVPVA	N
53BP1	Kme1	P33176	KINH	739	DLQDQNQKMMLEQER	N
KMBD-2	Kme2	P35637	FUS	306	ESVADYFKQIGIIKT	N
KMBD-2	Kme2	P35637	FUS	348	FDDPPSAKAAIDWFD	N
KMBD-1	Kme2	P38432	COIL	287	SKEEPSTKNTTADKL	N
KMBD-1	Kme1	P42681	TXK	68	TGRVQPSKRKPLPPL	N
KMBD-2	Kme2	P46783	RS10	59	LKSRGYVKEQFAWRH	N
KMBD-2	Kme2	P46783	RS10	139	VPPGADKKAEAGAGS	N
KMBD-2	Kme2	P50402	EMD	79	DMYDLPKKEDALLYQ	N
KMBD-2	Kme1	P50991	TCPD	65	DKMIQDGKGDVTITN	N
53BP1	Kme2	P61129	ZC3H6	770	IPRQDIRKPSESAPL	N

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs (Continued)

KMBD-1	Kme1	P61978	HNRNPK	198	RVVLIGGKPDRVVEC	N
KMBD-2	Kme2	P67936	TPM4	13	SLEAVKRKIQALQQQ	N
KMBD-2	Kme1	P68104	EF1A1	172	KRYEEIVKEVSTYIK	Y
KMBD-2	Kme1	P68104	EF1A1	165	TEPPYSQKRYEEIVK	Y
KMBD-2	Kme2	P68104	EF1A1	55	EMGKGSFKYAWVLDK	Y
KMBD-2	Kme2	P68104	EF1A1	179	KEVSTYIKKIGYNPD	N
KMBD-2	Kme1	P68431	H31	79	REIAQDFKTDLRFQS	Y
KMBD-1	Kme2	P68431	H31	27	LATKAARKSAPATGG	Y
KMBD-2	Kme2	P68431	H31	79	REIAQDFKTDLRFQS	Y
KMBD-1	Kme3	P68431	H31	27	LATKAARKSAPATGG	Y
MPP8	Kme3	P82987	ATL3	948	KSLIQWEKDGRCLQN	N
KMBD-2	Kme2	P98179	RBM3	7	_MSSEEGKLFVGGLN	N
KMBD-2	Kme1	Q00839	HNRPU	516	GKTTWVTKHAAENPG	Y
KMBD-1	Kme2	Q00839	HNRNPU	544	MMVAGFKKQMADTGK	N
KMBD-1	Kme3	Q00839	HNRNPU	543	KMMVAGFKKQMADTG	N
KMBD-1	Kme1	Q01082	SPTBN1	927	HPSEKEIKAQQDKLN	N
KMBD-2	Kme2	Q01804	OTUD4	1074	VRSEESWKGQPSRSR	N
KMBD-1	Kme1	Q01844	EWSR1	641	GGRGGPGKMDKGEHR	N
KMBD-2	Kme2	Q01844	EWS	441	DFQGSKLKVSLARKK	N
KMBD-1	Kme1	Q06033	ITIH3	352	QEARTFVKSMEDKGM	N
KMBD-1	Kme2	Q07065	CKAP4	283	LEESEGNKQDLKALK	N
KMBD-2	Kme2	Q07666	KHDR1	432	APPARPVKGAYREHP	N
KMBD-2	Kme2	Q07666	KHDR1	185	GPQGNTIKRLQEETG	N
KMBD-1	Kme2	Q0VDD8	DNAH14	216	EKKYEDVKPLETQPA	N
MPP8	Kme2	Q12789	TF3C1	1142	SYIINQAKKENTAAE	N
MPP8	Kme2	Q12789	TF3C1	1143	YIINQAKKENTAAEN	N
MPP8	Kme3	Q12789	TF3C1	1142	SYIINQAKKENTAAE	N
MPP8	Kme3	Q12789	TF3C1	1143	YIINQAKKENTAAEN	N
KMBD-1	Kme1	Q12830	BPTF	10	GRRGRPPKQPAAPAA	N
KMBD-1	Kme2	Q12830	BPTF	10	GRRGRPPKQPAAPAA	N
KMBD-2	Kme2	Q12834	CDC20	163	ATPGSSRKTCRYIPS	N
KMBD-2	Kme2	Q12834	CDC20	236	ISSVAWIKEGNYLAV	N
KMBD-1	Kme3	Q12873	CHD3	1582	EAENQEEKPEKNSRI	N
KMBD-1	Kme1	Q12906	ILF3	617	GGPKFAAKPHNPGFG	Y
KMBD-1	Kme1	Q12906	ILF3	613	VPVRGGPKFAAKPHN	Y
53BP1	Kme2	Q12933	TRAF2	295	KIEALSSKVQQLERS	N
KMBD-2	Kme2	Q13042	CDC16	67	LRSRKLDKLYEACRY	N
KMBD-2	Kme2	Q13151	ROA0	211	RDQNGLSKGGGGGYN	N
KMBD-1	Kme1	Q13185	CBX3	50	GKVEYFLKWKGFTDA	Y
KMBD-1	Kme1	Q13185	CBX3	34	PEEFVVEKVLDRRVV	Y
KMBD-2	Kme1	Q13185	CBX3	44	DRRVVNGKVEYFLKW	N
KMBD-2	Kme1	Q13185	CBX3	21	KQNGKSKKVEEAEPE	N

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs (Continued)

KMBD-1	Kme2	Q13185	CBX3	44	DRRVVNGKVEYFLKW	N
KMBD-1	Kme2	Q13185	CBX3	21	KQNGKSKKVEEAEPE	N
KMBD-1	Kme2	Q13185	CBX3	34	PEEFVVEKVLDRRVV	N
KMBD-1	Kme3	Q13185	CBX3	21	KQNGKSKKVEEAEPE	N
KMBD-1	Kme1	Q13895	BYSL	5	___MPKFKAARGVGG	Y
KMBD-2	Kme2	Q14103	HNRPD	341	NQQSGYGKVSRRGGH	N
KMBD-2	Kme2	Q14140	SRTD2	47	IFNISLMKLYNHRPL	N
KMBD-2	Kme1	Q14157	UBP2L	161	ENGLDGTKSGGPSGR	N
KMBD-2	Kme2	Q14157	UBP2L	161	ENGLDGTKSGGPSGR	N
MPP8	Kme1	Q14964	RB39A	60	LLEIEPGKRIKLQLW	N
KMBD-1	Kme1	Q14999	CUL7	211	SQQEAIEKHLDFDSR	N
KMBD-2	Kme2	Q15047	SETB1	1170	STRGFALKSTHGIAI	Y
KMBD-1	Kme1	Q15370	TCEB2	36	RIVEGILKRPPDEQR	Y
KMBD-2	Kme1	Q15758	AAAT	522	EEGNPLLKHYRGPAG	N
KMBD-1	Kme3	Q15858	SCN9A	1910	YRLRQNVKNISSIYI	N
KMBD-2	Kme2	Q1KMD3	HNRL2	342	GFDGRGLKAENGQFE	N
KMBD-2	Kme2	Q1KMD3	HNRL2	479	YAKENPEKRYNVLGA	N
KMBD-1	Kme3	Q2NL82	TSR1	476	AKMLEKYKQERLEEM	N
KMBD-1	Kme1	Q3BBV2	NBPF8	32	RPQLAENKQQFVNLK	N
KMBD-1	Kme2	Q4VCS5	AMOT	775	RSRKEPSKTEQLSCM	N
KMBD-1	Kme1	Q53EZ4	CEP55	29	KSETTLEKLKGEIAH	N
KMBD-1	Kme2	Q53GS7	GLE1	566	EQQDNFLKRMSGMIR	N
KMBD-1	Kme3	Q53RT3	ASPRV1	194	VFANSMGKGYYLKGK	Y
KMBD-1	Kme2	Q5BJF6	ODF2	512	KLENERLKASFAPME	N
53BP1	Kme2	Q5T2B5	CUL2	621	IKSLLDVKMINHDSE	N
53BP1	Kme2	Q5T2B5	CUL2	615	KELTKTIKSLLDVKM	N
KMBD-1	Kme2	Q5T8P6	RBM26	766	LEKNKTMKSEDKAEI	N
53BP1	Kme1	Q5THK1	PR14L	1186	SHYGQQDKGTSLRET	N
MPP8	Kme2	Q5UIP0	RIF1	1454	EEEKPLQKSPLHIKD	N
MPP8	Kme2	Q5UIP0	RIF1	1449	ERRKEEEKPLQKSPL	N
KMBD-1	Kme2	Q5VSY0	GKAP1	295	ARNAQLLKMLQEGEM	N
KMBD-1	Kme3	Q5VSY0	GKAP1	295	ARNAQLLKMLQEGEM	N
KMBD-1	Kme1	Q5VTE0	EF1A3	179	KEVSTYIKKIGYNPD	Y
KMBD-1	Kme1	Q5VTE0	EF1A3	172	KRYEEIVKEVSTYIK	N
KMBD-1	Kme1	Q5VTE0	EF1A3	165	TEPPYSQKRYEEIVK	Y
53BP1	Kme2	Q5VTE0	EF1A3	165	TEPPYSQKRYEEIVK	N
KMBD-1	Kme2	Q5VTE0	EF1A3	55	EMGKGSFKYAWVLDK	N
KMBD-1	Kme2	Q5VTE0	EF1A3	179	KEVSTYIKKIGYNPD	N
KMBD-1	Kme2	Q5VTE0	EF1A3	172	KRYEEIVKEVSTYIK	N
KMBD-1	Kme2	Q5VTE0	EF1A3	62	KYAWVLDKLKAERER	N
53BP1	kme3	Q5VTE0	EF1A3	80	GITIDISLWKFETSK	N
KMBD-1	Kme3	Q5VTE0	EF1A3	172	KRYEEIVKEVSTYIK	Y

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs (Continued)

KMBD-1	Kme3	Q5VTE0	EF1A3	79	TIDISLWKFETSKYY	N
KMBD-1	Kme1	Q5VUA4	ZNF318	1001	DSREPTEKPGKAEKS	N
KMBD-1	Kme1	Q5VUA4	ZNF318	1004	EPTEKPGKAEKSKSP	Y
53BP1	Kme1	Q5VUR7	A20A3	255	QQQILEHKKKILKKE	N
KMBD-1	Kme2	Q68DI1	ZNF776	122	GFGAYEKKLDDDANH	N
MPP8	Kme2	Q68E01	INT3	136	ILMEKYLKLQDTCRT	N
MPP8	Kme2	Q68E01	INT3	133	INQILMEKYLKLQDT	N
KMBD-1	Kme3	Q6NWY9	PRPF40B	813	AGKESDEKEQEQDKD	Y
KMBD-1	Kme1	Q6PJG2	ELMSAN1	186	VRPMMPQKVQLEVGR	N
KMBD-1	Kme2	Q6PJG2	ELMSAN1	186	VRPMMPQKVQLEVGR	N
KMBD-1	Kme1	Q6ZMT4	KDM7A	484	EENGKPVKSQGIPIV	N
KMBD-1	Kme1	Q6ZN16	MAP3K15	751	PMKEPTIKFYTKQIL	N
KMBD-1	Kme1	Q6ZN16	MAP3K15	755	PTIKFYTKQILEGLK	N
KMBD-1	Kme2	Q6ZN16	MAP3K15	751	PMKEPTIKFYTKQIL	N
KMBD-1	Kme3	Q6ZN16	MAP3K15	755	PTIKFYTKQILEGLK	N
53BP1	Kme2	Q6ZN92	DUTL	134	LTQEQKTKHCMFSLH	N
53BP1	Kme2	Q6ZN92	DUTL	132	SKLTQEQKTKHCMFS	N
53BP1	Kme1	Q6ZRR7	LRRC9	247	LADTTAMKKIMYYNM	N
KMBD-1	Kme1	Q6ZV73	FGD6	964	VKKGPYLKMYSTYIK	N
MPP8	Kme1	Q7L775	EPMIP	449	DELKQQNKEDEKIFD	N
MPP8	Kme1	Q7L775	EPMIP	445	REVVDELKQQNKEDE	N
MPP8	Kme3	Q7Z2K8	GRIN1	232	TYTVSPRKEDPGSLR	N
KMBD-1	Kme3	Q7Z401	DENND4A	1826	DAFDKEYKMAYDRLT	N
KMBD-1	Kme1	Q7Z699	SPRED1	312	PKDSVVFKTQPSSLK	N
KMBD-2	Kme2	Q7Z6G3	NECA2	265	LIGRLESKALWFDLQ	N
53BP1	Kme2	Q86T96	RN180	206	LSKASEPKYQLFVPQ	N
KMBD-2	Kme2	Q86UK7	ZN598	106	DIYFADGKVYALYRQ	N
KMBD-2	Kme2	Q86UK7	ZN598	97	HQLQHEKKYDIYFAD	N
KMBD-2	Kme2	Q86V81	THOC4	161	ERKADALKAMKQYNG	N
MPP8	Kme1	Q8IWJ2	GCC2	1196	EIKIQKQKQETLQEE	N
MPP8	Kme1	Q8IWJ2	GCC2	1194	EQEIKIQKQKQETLQ	N
KMBD-2	Kme2	Q8IXJ9	ASXL1	726	EDLPSLRKEESCLLQ	N
MPP8	Kme3	Q8IXJ9	ASXL1	1302	SNVTGQGKKLFGSGN	N
53BP1	Kme2	Q8IY18	SMC5	337	DIKEASQKCKQKQDV	N
53BP1	Kme2	Q8IY18	SMC5	339	KEASQKCKQKQDVIE	N
KMBD-1	Kme2	Q8IYJ2	C10orf67	154	DRILEIEKHYQQNED	N
KMBD-1	Kme1	Q8IYP9	ZDHHC23	8	MTQKGSMKPVKKKKT	N
53BP1	Kme2	Q8N1E6	FXL14	308	AQGLDGLKSLSLCSC	N
53BP1	Kme2	Q8N567	ZCHC9	108	EEIAVALKKDSRREG	N
53BP1	Kme2	Q8N567	ZCHC9	109	EIAVALKKDSRREGR	N
MPP8	Kme3	Q8N954	GPT11	38	EARRKEEKQQEANLK	N
MPP8	Kme3	Q8N954	GPT11	45	KQQEANLKNRQKSLK	N

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs (Continued)

KMBD-2	Kme2	Q8N9B5	JMY	134	RLRSPGSKGAESRLR	N
KMBD-1	Kme3	Q8NA19	L3MBTL4	159	PKGYRKDKFVWMDYL	N
53BP1	Kme2	Q8NB14	UBP38	995	LQVSWKYKLYLLKIL	N
53BP1	Kme2	Q8NB14	UBP38	993	LYLQVSWKYKLYLLK	N
MPP8	Kme2	Q8NCX0	CC150	1068	QEKDQDVKHDVMSNQ	N
MPP8	Kme3	Q8NCX0	CC150	1063	GEDRWQEKDQDVKHD	N
KMBD-1	Kme3	Q8ND07	CCDC176	76	KKQCKMEKDIMSVLS	N
KMBD-1	Kme3	Q8ND82	ZNF280C	311	GVTEKEPKTYTTFKC	Y
KMBD-1	Kme1	Q8NDH2	CCDC168	560	TREQEEEKVQKVKSG	N
KMBD-1	Kme2	Q8NDV7	TNRC6A	1279	ERNPYFDKDGIVADE	N
MPP8	Kme2	Q8NEN0	ARMC2	263	EILINLIKQINENIK	N
KMBD-1	Kme1	Q8NFC6	BOD1L1	786	TKLSSDDKTERKSKH	N
KMBD-1	Kme2	Q8NGA6	OR10H5	303	KELKVAMKKTCFTKL	N
KMBD-2	Kme1	Q8NGM1	OR4CF	311	LMVVSDEKENIKL__	N
KMBD-2	Kme2	Q8NGM1	OR4CF	311	LMVVSDEKENIKL__	N
KMBD-2	Kme1	Q8NHY3	GA2L2	479	LPLRDEAKGAFFQFR	N
KMBD-1	Kme2	Q8NI35	INADL	1187	QNQDTQEKKEKRQGT	N
KMBD-1	Kme1	Q8TAC1	RFESD	12	GSAQDPEKREYSSVC	N
KMBD-2	Kme2	Q8TBM8	DJB14	367	ALSMDNCKELERLTS	N
KMBD-2	Kme2	Q8TC76	F110B	359	IKWLYSIKQARESQK	N
MPP8	Kme2	Q8TCX1	DC2L1	119	VLVLDLSKPNDLWPT	N
KMBD-2	Kme2	Q8TF45	ZN418	255	HQRVHTGKRPYECGE	N
KMBD-2	Kme1	Q8WUM4	PDC6I	751	PRTMPPTKPQPPARP	Y
53BP1	Kme1	Q8WXA3	RUFY2	309	INIEMYQKLQGSEDG	N
53BP1	Kme1	Q8WXA3	RUFY2	318	QGSEDGLKEKNEIIA	N
KMBD-2	Kme1	Q8WXF1	PSPC1	519	GNFEGPNKRRRY___	Y
KMBD-2	Kme2	Q8WYA0	IFT81	637	GPNMKQAKMWRDLEQ	N
MPP8	Kme1	Q92628	K0232	683	SSANMLGKTQSRLLI	N
KMBD-2	Kme2	Q92734	TFG	155	SASDSSGKQSTQVMA	N
KMBD-2	Kme1	Q92804	RBP56	576	DRGGYGGKMGGRNDY	N
KMBD-2	Kme2	Q92804	RBP56	576	DRGGYGGKMGGRNDY	Y
KMBD-2	Kme1	Q969Q0	RL36L	53	SGYGGQTKPIFRKKA	Y
KMBD-2	Kme1	Q96E39	RMXL1	86	IKVEQATKPSFERGR	N
KMBD-1	Kme2	Q96E39	RBMXL1	86	IKVEQATKPSFERGR	N
KMBD-1	Kme1	Q96E52	OMA1	390	FNRPYSRKLEAEADK	N
53BP1	Kme2	Q96F24	NRBF2	179	ATKIADLKRHVEFLV	N
53BP1	Kme2	Q96F24	NRBF2	174	IIEEQATKIADLKRH	N
53BP1	Kme2	Q96F24	NRBF2	164	SKAPKDDKTIIEEQA	N
KMBD-1	Kme1	Q96IZ6	METTL2A	53	QAAAAERKVQENSIQ	N
KMBD-1	Kme1	Q96KQ7	EHMT2	200	GQPPVPEKRPPEIQH	N
KMBD-1	Kme1	Q96KQ7	EHMT2	189	RARKTMSKPGNGQPP	Y
KMBD-1	Kme2	Q96KQ7	EHMT2	189	RARKTMSKPGNGQPP	N

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs (Continued)

KMBD-2	Kme2	Q96KQ7	EHMT2	114	ILLGHATKSFPSSPS	N
KMBD-1	Kme3	Q96KQ7	EHMT2	189	RARKTMSKPGNGQPP	N
53BP1	Kme1	Q96PY6	NEK1	507	FNERQQIKAKLRGEK	N
MPP8	Kme1	Q96Q89	KI20B	1300	TDAKKQIKQVQKEVS	N
KMBD-1	Kme2	Q96T58	SPEN	902	PVEKLKAKLDNDTVK	N
KMBD-2	Kme1	Q99729	ROAA	232	EIKVAQPKEVYQQQQ	N
KMBD-2	Kme1	Q99729	ROAA	190	IELPMDPKLNKRRGF	N
KMBD-2	Kme2	Q99729	ROAA	232	EIKVAQPKEVYQQQQ	N
KMBD-2	Kme2	Q99729	ROAA	330	GGHQNNYKPY_____	N
KMBD-1	Kme2	Q99880	HIST1H2BL	58	PDTGISSKAMGIMNS	N
53BP1	Kme1	Q9BRV8	SIKE1	178	AQLELENKELRELLS	N
KMBD-1	Kme2	Q9BUP0	EFHD1	9	ASEELACKLERRLRR	N
KMBD-2	Kme1	Q9BYN0	SRXN1	61	PSVLDPAKVQSLVDT	N
KMBD-1	Kme1	Q9C0H9	SRCIN1	813	QRLDGLLKRCRGVTD	N
KMBD-1	Kme1	Q9H0D6	XRN2	858	RGQAQIPKLMSNMRP	Y
KMBD-1	Kme1	Q9H0V9	LMAN2L	150	PVFGNMDKFVGLGVF	N
MPP8	Kme2	Q9HAH1	ZN556	426	QIGQKPSKCEKCGKA	N
MPP8	Kme2	Q9HAH1	ZN556	423	VRTQIGQKPSKCEKC	N
KMBD-2	Kme2	Q9HAZ2	PRD16	824	GGGREPRKNHVYGER	N
KMBD-1	Kme1	Q9HCF6	TRPM3	1370	SFYSVNMKDKGGIEK	N
MPP8	Kme1	Q9NPF5	DMAP1	35	KDIINPDKKKSKKSS	N
KMBD-1	Kme1	Q9NPG3	UBN1	238	SKEKKKKKYSGALSV	N
KMBD-1	Kme1	Q9NPG3	UBN1	246	YSGALSVKEMLKKFQ	N
KMBD-2	Kme2	Q9NS66	GP173	166	VFDVGTYKFIREEDQ	N
MPP8	Kme2	Q9NS91	RAD18	462	EAWEASHKNDLQDTE	N
KMBD-2	Kme1	Q9NVQ4	FAIM1	42	DGKEEIRKEWMFKLV	N
KMBD-2	Kme2	Q9NVQ4	FAIM1	95	KYMEDRSKTTNTWVL	N
KMBD-2	Kme2	Q9NWH9	SLTM	476	MKKENDEKSSSRSSG	N
KMBD-1	Kme1	Q9NXG0	CNTLN	1289	TFVKALAKELQNDVH	N
KMBD-1	Kme3	Q9NXG0	CNTLN	720	EGNKKLMKENDFLKS	N
KMBD-1	Kme3	Q9NXG0	CNTLN	1003	SVLQNAKKTAELSVK	N
KMBD-2	Kme1	Q9NZM5	NOP53	410	PRRLGRLKYQAPDID	N
KMBD-1	Kme2	Q9P2H0	CEP126	65	QILLQQQKICRNRAR	N
53BP1	Kme1	Q9P2P6	STAR9	786	QRLLEAQKRLEKLTT	N
KMBD-1	Kme3	Q9UIF9	BAZ2A	695	TDNRPLKKLEAQETL	N
KMBD-1	Kme1	Q9UIH9	KLF15	319	NPAAELIKMHKCTFP	N
53BP1	Kme2	Q9UIV8	SPB13	214	IDKISPEKLVEWTSP	N
53BP1	Kme2	Q9UIV8	SPB13	209	WMNKIIDKISPEKLV	N
KMBD-2	Kme2	Q9UKV8	AGO2	381	DRQEEISKLMRSASF	N
53BP1	Kme2	Q9ULL0	K1210	1416	KKFSQGSKNPIKSIP	N
53BP1	Kme2	Q9ULL0	K1210	1420	QGSKNPIKSIPAPAT	N
KMBD-1	Kme1	Q9UPW6	SATB2	722	AEEENADKSKAAPAE	N

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs (Continued)

KMBD-1	Kme1	Q9UPW6	SATB2	724	EENADKSKAAPAEID	N
KMBD-1	Kme2	Q9UPW6	SATB2	722	AEEENADKSKAAPAE	N
KMBD-1	Kme2	Q9UPW6	SATB2	724	EENADKSKAAPAEID	N
KMBD-2	Kme2	Q9Y233	PDE10	754	DNLSQWEKVIRGEET	N
KMBD-2	Kme2	Q9Y233	PDE10	328	EIRFSIEKGIAGQVA	N
KMBD-2	Kme2	Q9Y233	PDE10	402	KTDENNFKMFAVFCA	N
KMBD-1	Kme1	Q9Y283	INVS	315	FLKHPSVKDDSDLEG	N
MPP8	Kme2	Q9Y2H2	SAC2	509	QVQQNELKKMFIQCQ	N
MPP8	Kme2	Q9Y2H2	SAC2	510	VQQNELKKMFIQCQT	N
MPP8	Kme3	Q9Y2H2	SAC2	509	QVQQNELKKMFIQCQ	N
MPP8	Kme3	Q9Y2H2	SAC2	510	VQQNELKKMFIQCQT	N
53BP1	Kme2	Q9Y2W1	TR150	202	DNQGDEAKEQTFSGG	N
53BP1	Kme1	Q9Y3R5	DOP2	886	VLWNQLNKETREHHV	N
MPP8	Kme2	Q9Y471	CMAH	494	TNEPNRNKFSVENKA	N
MPP8	Kme3	Q9Y6N7	ROBO1	51	EPATLNCKAEGRPTP	N

Table 3.2. List of all the enriched Kme sites from HEK293 cells by KMBDs. This table lists all the Kme sites enriched with KMBDs, all four experiments from HEK293 cells. The lysine modification is labelled under Mod. The table also has Uniprot ID, gene name where applicable. The sequence is ± 7 residue from the central modified lysine, where protein is at the end or beginning, "__" indicates the number of spaces. Found column signify whether this site has been previously reported, where Y means known site and N means novel site.

4 Discussion and future directions

4.1 General discussion and conclusions

4.1.1 Systematic characterization of KMBD binding profile

4.1.1.1 Probing KMBD binding via membrane peptide array

Membrane peptide array is a powerful tool for high throughput substrate screening [125]. In this project, three different peptide membranes were generated. First, designed by Dr. Kyle Biggar from his prediction experiments, and we picked the top 70 protein candidates and made four different variants of it with Kme0, 1, 2 and 3 (Fig 3.5). Each of the peptides was 15 amino acid residues with lysine at the center. The second peptide array used was the 140-cluster array from the phylogenetic-like clustering approach (Fig. 3.8). By grouping similar peptide sequences based on similarity and origin, we were able to pick out 140 clusters with four different modifications of the center lysine with some control peptides. These peptides were also 15 residues in length with center lysine. The third array was a permutation array of NHEJ1 peptide (Fig. 3.10). This permutation array was 11 residues in length with center lysine. Each of the amino acid residues was permutated to another amino acid residue, except for cysteine. Cysteines during reaction could result in sulphide bridges to complicate the synthesis [4,106]. The permutation array was used to assess whether any one type of residue can impact the affinity of the peptide for the protein, whether it can increase or decrease binding affinity [125].

There were several important factors to achieve maximum results from these membrane peptide arrays — first, the background binding to the membrane. Achieving the best data from the membrane peptide array requires the background binding to be kept to a minimum. The background should be white to minimize signal interference for positive binding. The higher the background, the harder to distinguish if a spot was binding or it was just background noise. One way to minimize the background was to block the membrane in 5% BSA in PBS at 4 °C overnight with mild shaking. Blocking the

membrane 12-24 hours achieved the best results to have minimal background noise. Second, the purity of the purified protein would determine the amount of binding and can increase the background noise if the protein sample was not pure. Even though the sample was 90% pure, it was best to run another column to make the purity even higher. With pure proteins, there will be less contamination, less undesired binding and cleaner background. Third, the protein concentration blotting onto the peptide array. Choosing the correct concentration of binding protein was important. Too much protein could result in over binding or over-saturation of a spot [126]. Too little protein could result in less than ideal binding or very weak signals from the spots. The best way to determine what protein concentration to use is by trial and error. Last, the chemiluminescence solution that was being used and the exposure time. Depending on the enhanced chemiluminescence (ECL) used, the incubation time and exposure time could be different. Too many substrates could result in very dark spots across all spots and make the background very dark as well. Too little substrates could result in underdeveloped spots. We used our house-made ECL and substrates to incubate our membrane for two minutes and manual exposure of one picture every five seconds for 60 seconds. With the continuous exposure, I could pick the picture with the most optimal exposure for all the spots and a clear background.

There were several controls that one needs to do before the start of the experiment. First, the freshly synthesized membrane was blotted with the antibody (or antibodies) to assess whether there were any interactions from the antibody with the peptides and the membrane. This step will ensure that all the interactions seen on the membrane happened between our protein of interest and the membrane peptides. Second, evaluate the antibody (or antibodies) binding to our protein of interest. Third, assess whether the membranes could be stripped and reused. If the proteins could be stripped off the membrane without a trace and reblotting yield similar results, the reuse of membranes will be cost-effective. I used two stripping buffers to strip the proteins of our membranes (Fig. S.1). After stripping the proteins, the membrane was exposed to ECL and substrate to see if there were any residual antibodies left on the membrane. Subsequently, I incubated the

membrane with antibody assess whether any proteins were left on the membrane. In figure S.1, it showed that the stripped membrane was protein-free and reblotting the membrane yielded similar data.

One important quality control (QC) was the sequence fidelity of the peptides on the membrane; how accurate were they? It was a very important question and a hard one to answer. I needed to make sure that each spot was what we intended, and no mutations occur during synthesis. However, it was difficult to do QC on every single membrane. There were groups using membrane peptide array to synthesize peptides and cleave them for MS studies [127]. They would design tryptic peptides and use trypsin to digest the peptides off the membrane. However, our designed peptides, most of the time, were not tryptic. Only peptide sequences with one lysine or arginine at the C-terminal could work. In the past, our group tried the tryptic digestion method for QC. However, the digestion was poor. Now, I included control peptides during our synthesis for each batch of synthesis. For example, I would include GGpYGG peptide, which can be bound by the 4G10 antibody, placing this peptide in a few positions on the membrane to make sure the dotting was robust. Furthermore, I included biotin on the N-terminal ends of peptides, with the random amino acid composition of various lengths, up to the maximum length of our peptide on the membrane. These biotinylated peptides were done in duplicates, one with biotin and the other without. When the synthesis was complete, the control peptides were blotted and evaluated. Through these control peptides, I assumed the rest of the synthesis was good.

One of the drawbacks of using peptide array was that stronger signal does not necessarily translate to a higher affinity. However, the stronger signal does suggest the interaction was strong, but we could not determine the actual K_D from comparing the signals of control to another peptide. We could determine that our protein had bound with this peptide, and they have a good interaction or a weak interaction. The actual K_D could not be determined by this method. The peptide arrays are a great screening tool to assess binding or no binding. There were several reasons why a stronger signal does not

translate to a stronger binding. First, we did not know the precise number of peptides that were on each spot; Intavis gave us a rough estimate of 1 µM of the peptide. The synthesis has a range of peptides being spotted into each spot, but the precise number of peptides per spot could not be determined easily. Thus, two same peptides with different quantities will result in one having a stronger signal. Second, K_{on} and K_{off} could not be determined since the membrane was already at equilibrium [128]. Last, signal saturation and over-exposure. There was a certain range where the signal has a linear response [129]. However, after a certain range, the signal did not respond linearly anymore. Thus, we could not infer any binding affinity from peptides with strong signals. Since the analysis program also has a maximum pixel density, by overexposing the image, it will not behave linearly [130]. However, strong signal, stronger than positive control would mean a positive binding, even though the binding affinity could not be determined.

4.1.1.2 Substrate binding affinity via fluorescence polarization

Fluorescence polarization (FP) is a powerful tool to test protein-substrate binding affinity [114]. The fluorescent-labelled peptides, when excited, it will rotate rapidly, and when bound to a protein, its rotation will be slowed, and the emitted light is polarized [131]. With our FP experiments, the control experiments with 53BP1-tudor and H4K20 peptides showed binding to Kme1/2 (Fig. 3.4). However, in some other experiments, the change in fluorescence polarization (ΔFP) was low. With the H4K20 experiment, I had ΔFP max around 200 mP. However, with the other ones, the ΔFP max was only 50 mP. There were experiments with max ΔFP less than 25 mP, where the data was ambiguous, hard to distinguish a positive binding or just background noise. Some known binding peptides to our proteins showed weak ΔFP and sometimes no ΔFP. There had been a few peptides that were giving me a hard time.

After all the peptides were synthesized, I sent all the peptides to MALDI to examine if the synthesis was successful. All in-solution peptides used in the project were validated via MALDI, and all the peptides were correctly synthesized. One important aspect of FP is the pH of the peptides and the proteins. Fluorescein is sensitive to pH, and its emission

will change upon different pH [132]. All the proteins and peptides were diluted with PBS. However, a lot of the methylated peptides have many basic residues, where their pI was above 12. These Kme peptides were diluted more than 1,000 times. Thus, their pH should be the same as PBS at pH 7.4. Before I began the experiments, I used a pH paper to test the pH of our peptide solution and the protein solution. Both were the same, had a pH of 7.4. However, this did not solve that several peptides were not giving good FP results. Resulting in elusive K_D in some interesting peptides that I wanted to know.

4.1.2 53BP1-tudor substrate prediction

The 53BP1-tudor domain had the best binding performance compared to the other eight KMBDs that I had tested. It had lots of interactions with various degrees of binding (Fig. 3.6). Based on that, I was able to make an enriched amino acid logo for the top 5% from the peptide array. As well as, taking into consideration of the amino acid composition of the top 5% Kme1s versus the bottom 5% of Kme1s (Fig. 3.7), I was able to find a pattern that acidic residues from -2 to +3 region will have poor interaction with the 53BP1-tudor domain. This binding pattern was supported by the structure of 53BP1-tudor binding pocket from crystallography data [78]. There was an acidic residue, D1521, in the binding pocket of the 53BP1-tudor domain.

From the predicted substrate list, by looking at the top-ranked candidates, and grouping them by biological processes, it gave us important hints to the 53BP1-tudor domain interactions with several major pathways. It was already known that 53BP1-tudor binds to p53 during cellular stress and during DNA damage events [25]. All the top predicted pathways were very important and related to 53BP1 and its tudor domain. 53BP1, its name is p53 binding protein 1, it binds to p53 during cellular stress [25]. Since it binds to p53, it is involved in apoptosis regulation, DNA damage pathway, cell division [3]. During DNA double-strand break (DSB), 53BP1-tudor recognizes histone markers H3K9, H3K79 and H4K20 and important in the DSB repair pathway [78,120]. I observed some of the best-predicted proteins were involved in DNA damage repair: NHEJ1, ETAA1 and SETX. 53BP1-tudor domain relation to RNA process could be led by tudor interacting

repair regulator protein (TIRR). TIRR inhibits the 53BP1-tudor domain by binding onto it during no stress conditions [133]. It prevents the 53BP1-tudor binding to H4K20 and other substrates. During DNA damage, TIRR interacts with RNA molecule and releases 53BP1, and it is recruited to DSB site [134]. Thus, the categories of predicted proteins fell in line with the function of 53BP1.

The amount of data used to generate the enriched logo was still a little less than ideal. By taking consideration of NHEJ1 permutation data, I was able to rank all the currently known Kme peptides. However, there were still several peptides that were tied in score, and this can be observed by the sudden dips and horizontal dots from Fig. 3.11. I am in the process of generating another set of arrays to validate the prediction list. Choosing the top 400 peptides, along with 50 peptides at 25%, 50%, 75% and 100% percentile. This second array is to validate the predicted substrates, and I also wanted to have more data points to refine the enrichment logo.

If I can get the data from our prediction array, I could potentially have enough data points to use a deep learning model to predict 53BP1-tudor substrate from all human proteins [135]. Based on its binding specificity, I could identify novel methylated sites. These sites will require *in vitro*, as well as *in vivo* validation. *In vitro* validation could be done using the peptide array. Whereas, *in vivo* validation would require using cell lines to enrich for the specific site. The *in vivo* validation could be achieved by using a Q-TRAP MS to pick out the peptide transitions and screen for peptides that match our predicted transitions [136].

4.1.3 Methyllysine peptide enrichment by KMBDs

Previously, many groups used KMBDs to enrich for methyllysine peptides [4,107]. However, they only used one domain for their enrichment. Here, I differ by using five different domains and five relatively less commonly used in Kme peptide enrichment. I first systematically characterized all the nine KMBDs that were available, then based on their complementary binding profile, I picked five KMBDs. These KMBDs were 53BP1-tudor, CBX7-CD, L3MBTL2-3xMBT, MPP8-CD, PHF1-tudor. Overall, the results of all

four experiments combined, the number of proteins found were 229 with 301 novel Kme sites. With lots of Kme1/2 sites as well as many Kme2 novel sites. Since the number of discovered Kme2 sites was small compared to Kme1, there are more out there to be discovered. Thus, pulling out several hundred novel Kme2 sites was not very surprising.

There were several factors to improve the enrichment of Kme peptides. One, optimize the washing step after peptide incubation. Unlike other AP approaches where their substrate affinities are in the nanomolar range [137]. The affinity of KMBDs to their substrates was much lower, sub-micromolar at best to high micromolar to even millimolar range [122]. Due to their lower binding affinity to their substrates, the amount of washing after incubation with trypsinized cell lysate needs to be specially optimized. In my current method, I washed the KMBDs twice, with a total of 400 times the dilution factor. I did not want to wash too much to dislodge the Kme substrates and those weak binding Kme substrates. At the same time, I wanted to remove as many unspecific binders as possible. Thus, I kept the washing at a minimum, but there were many non-modified peptides within the final sample. It could be the result of inadequate washing. More interestingly, it could also be the result of a non-modified form of the Kme peptides. From the binding data, I observed that some of the Kme0s had higher binding or higher affinity than Kme1/2/3 of another peptide. Since the KMBD recognition was mostly sequence-based, Kme0 of a high-affinity substrate could be higher affinity than other methyllysine peptides (Fig. 3.6). This, however, could potentially replace methylated substrate and leaving the non-modified peptide in the binding pocket. This theory should be considered in future experiments and check the whole MS spectra to see any of these non-methylated high-affinity peptides do exist in the enrichment.

Another factor to consider was that the peptides used for the test conditions and peptide arrays were not trypsinized peptides. It could affect how the KMBDs binds to their substrates and our characterization of their binding specificity. Trypsinized peptides will end with lysine or an arginine residue, and it is common to find these residues within ± 5 residues around the modified lysine. It could potentially cut off parts of the substrate and

making the KMBD binding to this truncated substrate weaker. There was also the possibility that the modified lysine was the last residue on the peptide. These truncated peptides could abolish half of the binding recognition, making it an even weaker substrate. However, trypsin did have less efficiency to cleave at modified lysines due to the methyl-groups [109]. But there will be some methylated peptides having C-terminus modified lysines. Truncated Kme peptides and C-terminus Kme peptides were not tested in this project. It would be an interesting experiment to assess the binding differences between the full-length peptide versus the truncated peptides.

Another crucial factor in having a good enrichment was the amount of conjugated KMBDs, and the starting digested cell lysates that were used. Previously, people used from 5 mg to upwards of 30 mg of starting cell lysate [138–140]. Here, in this project, I used 30 µL of packed streptavidin beads and 2 mg of starting cell lysate. I chose to use less cell lysate was particularly due to our MS machine injection limit of 500 ng. Since protein methylation events are low in the cells, it would be better to increase the amount of starting material. However, more starting material could potentially translate to more unspecific substrates. I kept the amount of KMBDs the same since more binders could result in more unspecific binding and more non-modified peptides into the MS. Also, the more abundant non-modified peptides could interfere by masking the lower abundant Kme peptides signals [141]. By using more starting material, and keeping KMBDs constant, with improved enrichment method, I believe I can achieve higher enrichment numbers.

During sample preparation for mass spectrometry, several key factors could determine the number of peptides going through for detection. One crucial factor was peptide loss through cleaning steps. I lost, on average, about 30% of our peptides during C18 desalting step [142]. By our current method using two C18 cleaning steps, I am losing about 50% of enriched peptides. I could use another method involves using a 10 kD concentrator method [143]. By switching to this method, I could remove all the trypsin and undigested peptides, which could cause problems in the MS. It will ensure only digested peptides passes through; I also would lose fewer peptides compared to a C18 cleanup.

And this leads to another issue in sample preparation for MS, which is the undigested or partially digested peptides. When digesting proteins overnight with trypsin, there may be some undigested proteins and partially digested peptides, due to the decreasing activity of trypsin over time [144]. However, acidifying the solution would precipitate trypsin but not partially digested peptides. And C18 columns bind hydrophobic peptides and partially digested could also bind to it. Thus, it creates a problem down the line and into the MS machine. By using the new method, the partially digested peptides would be sequestered, buffers changed without a C18 cleanup. It could potentially increase the quality of the MS sample, as well as more peptides retained. The last important issue was the sample was not clean, or some impurities were causing problems with the MS column [145]. It will reduce the MS sensitivity and reduce the usage of the column. Here, MS grade buffers and clean glassware were to be used all the time to reduce contamination. Plastic tubes, Eppendorf tubes were all prewashed with the harshest buffers to clean up any residual polymer contaminants. All reagents, buffers, proteins, lysates were handled with extreme care in MS grade buffers to ensure minimal contamination. However, there were contaminants in the final MS sample, which could be due to partially digested peptides or polymers from tubes or beads. One important factor mentioned above was the washing step after lysate and KMBDs incubate. This step could be optimized to minimize contamination.

4.1.4 Strong cation exchange column in methyllysine enrichment

Strong cation exchange column could potentially add another layer of filtering and purification before the KMBD domains. In my preliminary test for SCX, I used the non-methylated H4K20 and H4K20me1/2 to compare their binding to the SCX column. The mechanism of SCX is that Kme peptides get retained on the column longer than other peptides due to the extra lysine or arginine positive charge. Trypsin has less efficiency at cleaving modified lysine and arginine residues [109]. As mentioned before, methylated lysine or arginine did not suppress its charge. Thus, in a perfectly tryptic peptide should have only one positive charged residue and locate at the C-terminus of the peptide. The modified lysine or arginine will remain uncut and retain one extra positive charge. It was

different from my testing condition, where the number of charges for both modified and non-modified lysine peptides was the same. Thus, resulting in no enrichment for H4K20me1/2 (Fig. 3.16). This experiment could be redone using tryptic peptides with a non-modified peptide with exactly one basic residue and modified having two basic residues.

The SCX column should be used in conjunction with KMBDs enrichment by enriching Kme peptides using our KMBD super column first to extract as many Kme peptides as possible. After, the flow-through and the washed off peptides are enriched using the SCX column. Combining the SCX column will allow us to capture Rme peptides, as well as some escaped Kme peptides. Especially those Kme peptides having modified lysine residues near the C-terminus. From this addition, I could analyze the whole methylome inside the cell and comparing the differences between treatments.

4.1.5 Mass spectrometry identification and database search

MS database searching software is every important in obtaining reliable data, as well as novel sites. There was a huge difference when switching from PEAKS to Maxquant [146]. The 53BP1-tudor AP-MS was searched using PEAKS. I was able to get almost twice as many proteins and peptides as the Maxquant search of MPP8-CD. However, the databases the software used were different. PEAKS had a putative protein database, any peptides that "hit" will be a novel site since it is a putative, non-reviewed protein. I had gotten 23 of these proteins in my 53BP1-tudor experiment. Whereas Maxquant searched from the reviewed Uniprot list. However, this could miss a few actual putative proteins that exist in the cells.

With the combination of all four experiments, I was able to identify 229 NHKMPs, 404 Kme sites and 301 out of 404 were novel Kme sites. Even more interesting was the number of Kme1/2 sites I was able to identify; 129 Kme1 and 173 Kme2 sites. Moreover, 107 out of 129 Kme1 sites were novel, composing 83% of total Kme1s identified. Furthermore, 169 out of 173 Kme2 sites were novel, composing 97.7% of total Kme2

identified. Currently, the number of Kme2/3 sites are very low compared to Kme1 (Fig. 1.2). There were only 610 Kme2 sites known to date [24]. In my experiments, I identified 173 Kme2 sites, with 169 novel Kme2 sites. The amount of novel site that has been identified here was about 28% more than what was currently published. Kme2, as a methyl marker, is unique. It is in a transition modification with lots of KDM and KMTs that could act on them (Table 1.1). It can be built up or broken down, a very versatile middle modification.

The number of sites being reported could depend on how the search parameters were set up during the database search. With less stringent parameters, when a shift of 15 Da for Kme1, 30 Da for Kme2, 45 Da for Kme3 were found, the placement of the modified residue depends on the parameters set [147]. Due to trypsin inefficiency at digesting methylated lysine residues, it can result in missed cleavages around Kme residue resulting in multiple lysines or arginine residues within proximity to each other [109]. By having a loose search parameter, if there were two lysine residues, the Kme1 could be designated to either one or the other lysine, making this a two-site "hit." Kme2 could also follow the same lose parameter. It has even more ways of placing the Kmes, assigning one Kme1 to each lysine or assigning Kme2 to either one or the other lysine, thus making this a three-site "hit." Thus, the possibility of Kme3s is even greater. These site identifications could cause a problem due to poor search parameters and dumping lots of potential sites but not valid sites. Thus, the data processing must be stringent and robust to ensure the highest quality of data.

One interesting group of proteins that I thought were going to mask other Kmes due to their high abundance but was not very abundant in our enrichment were the histone proteins. Histones are full of Kme and other PTM modifications [27]. However, not many histone proteins were enriched. One reason was due to many basic residues within histone proteins [148]. When histones were digested by trypsin, it resulted in many small peptides that were too small for the MS detection [149]. Another possibility was the partially digested histone proteins that contained a few basic residues were cleaned up by

the C18 column [150]. By using 0.1% FA, all the side chains on these basic residues will be protonated and having a positive charge. There retention time on the C18 column was greatly reduced and washed off. Thus, resulting in very little histone proteins being enriched and detected via MS.

4.2 Future Directions

With the results achieved here, the immediate future would be to test two different conditions on a cell line. One condition could be using doxorubicin to treat HEK293 cells and compare the lysine methylome changes with and without treatment. Since a lot of proteins being pulled out in the DNA damage repair and RNA processing pathways, with doxorubicin disrupting DNA stability, would stimulate a response from the cell [151]. It could signal a starting point, and if there were important key lysine methylated proteins in play between the two conditions, I could use SILAC to quantify the amount of increase or decrease [109].

Another direction to take the project is to characterize all the KMBDs and, based on their binding specificity, make a prediction logo or algorithm and integrate deep learning into the prediction [135]. From this, I could predict their potential substrates from the current known lysine methylated database. But more importantly, to generate a database from all the human proteins ± 5 residues from a central lysine and predict the chances of this being methylated. And I could validate this approach via AP-MS from cell culture. Only the positive results could confirm our predictions.

Another direction in the future would be to optimize the purification, digestion, binding and washing conditions. By having pure proteins, to begin with, this would minimize contamination. Binding and washing conditions would need to be optimized to have the highest amount of peptide binding and washing away the unmodified peptides. Currently, the washing step was not great due to lots of non-modified peptides being found in the elution. For protein digestion, there may be a lot of undigested proteins or large undigested peptides. We can use another method of protein digestion to minimize the

number of undigested proteins and filter out large undigested peptides using a 10k spin column [143]. Instead of digesting proteins in diluted urea, we can treat the cell lysate before and add the trypsin onto precipitated proteins. This way, it is easier to monitor the digestion. Also, by using the 10k spin column, it can filter out all the trypsin and undigested proteins [143]. This step could potentially save one C18 cleaning and can achieve a higher yield of peptides for the KMBD enrichment.

The application of this research could be applied to identifying key regulatory markers between cancer and healthy cells by monitoring the changes in lysine methylome between different cells. This method can track the changes of protein methylation status that are different between the cell lines to identify novel regulatory proteins. Thus, this method could identify novel targets for therapeutic intervention.

5 Reference

1. Burkle, A. Posttranslational Modification Predator ± Prey and Parasite ± Host Interactions. in *Encyclopedia of Genetics* (2001). doi:10.1006/rwgn.2001.1022

2. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* (2011). doi:10.1016/j.cell.2011.02.013

3. Hamamoto, R., Saloura, V. & Nakamura, Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat. Rev. Cancer* **15**, 110–24 (2015).

4. Liu, H. *et al.* A method for systematic mapping of protein lysine methylation identifies functions for HP1β in DNA damage response. *Mol. Cell* **50**, 723–35 (2013).

5. Kholodenko, B. N. Cell-signalling dynamics in time and space. *Nature Reviews Molecular Cell Biology* (2006). doi:10.1038/nrm1838

6. Clarke, S. Protein methylation. *Curr. Opin. Cell Biol.* (1993). doi:10.1016/0955-0674(93)90080-A

7. Beaver, J. E. & Waters, M. L. Molecular Recognition of Lys and Arg Methylation. *ACS Chemical Biology* (2016). doi:10.1021/acschembio.5b00996

8. Ambler, R. P. & Rees, M. W. ε-N-methyl-lysine in bacterial flagellar protein. *Nature* (1959). doi:10.1038/184056b0

9. Murn, J. & Shi, Y. The winding path of protein methylation research: Milestones and new frontiers. *Nature Reviews Molecular Cell Biology* **18**, 517–527 (2017).

10. Duerre, J. A. & Lee, C. T. IN VIVO METHYLATION AND TURNOVER OF RAT BRAIN HISTONES. *J. Neurochem.* (1974). doi:10.1111/j.1471-4159.1974.tb06057.x

11. Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends in Biochemical Sciences* (2000). doi:10.1016/S0968-0004(99)01535-2

12. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* (2007). doi:10.1038/nature05918

13. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* (2000). doi:10.1038/47412

14. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase

homolog LSD1. *Cell* (2004). doi:10.1016/j.cell.2004.12.012

15. Henry, M. F. & Silver, P. A. A novel methyltransferase (Hmt1p) modifies poly(A)+-RNA-binding proteins. *Mol. Cell. Biol.* (1996). doi:10.1128/mcb.16.7.3668

16. Biggar, K. K. & Li, S. S.-C. Non-histone protein methylation as a regulator of cellular signalling and function. *Nat. Rev. Mol. Cell Biol.* **16**, 5–17 (2014).

17. Hamamoto, R., Toyokawa, G., Nakakido, M., Ueda, K. & Nakamura, Y. SMYD2-dependent HSP90 methylation promotes cancer cell proliferation by regulating the chaperone complex formation. *Cancer Lett.* **351**, 126–133 (2014).

18. Zee, B. M., Levin, R. S., Dimaggio, P. A. & Garcia, B. A. Global turnover of histone post-translational modifications and variants in human cells. *Epigenetics and Chromatin* (2010). doi:10.1186/1756-8935-3-22

19. Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences* (2010). doi:10.1016/j.tibs.2010.05.006

20. Greer, E. L. & Shi, Y. Histone methylation: A dynamic mark in health, disease and inheritance. *Nature Reviews Genetics* (2012). doi:10.1038/nrg3173

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

22. Neuberger, A. & Sanger, F. The availability of the acetyl derivatives of lysine for growth. *Biochem. J.* (1943). doi:10.1042/bj0370515

23. Neuberger, A. & Sanger, F. The metabolism of lysine. *Biochem. J.* **38**, 119–125 (1944).

24. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gku1267

25. Lu, R. & Wang, G. G. Tudor: A versatile family of histone methylation ‘readers’. *Trends in Biochemical Sciences* (2013). doi:10.1016/j.tibs.2013.08.002

26. Ren, C. *et al.* Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. *Chem. Biol.* **22**, 161–168 (2015).

27. Eissenberg, J. C. Structural biology of the chromodomain: Form and function. *Gene* (2012). doi:10.1016/j.gene.2012.01.003

28. Huang, J. *et al.* p53 is regulated by the lysine demethylase LSD1. **449**, 1–6 (2007).

29. Blanc, R. S. & Richard, S. Arginine Methylation: The Coming of Age. *Molecular Cell* (2017). doi:10.1016/j.molcel.2016.11.003

30. Poulin, M. B. *et al.* Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36. *Proc. Natl. Acad. Sci.* (2016). doi:10.1073/pnas.1521036113

31. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Experimental and Molecular Medicine* **49**, e324-22 (2017).

32. Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Molecular Cell* (2012). doi:10.1016/j.molcel.2012.11.006

33. Batista, I. de A. A. & Helguero, L. A. Biological processes and signal transduction pathways regulated by the protein methyltransferase SETD7 and their significance in cancer. *Signal Transduction and Targeted Therapy* (2018). doi:10.1038/s41392-018-0017-6

34. Xu, J. *et al.* MLL1 and MLL1 fusion proteins have distinct functions in regulating leukemic transcription program. *Cell Discov.* (2016). doi:10.1038/celldisc.2016.8

35. Anand, R. & Marmorstein, R. Structure and mechanism of lysine-specific demethylase enzymes. *Journal of Biological Chemistry* (2007). doi:10.1074/jbc.R700027200

36. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science (80-.).* (2017). doi:10.1126/science.aan2507

37. Xu, Y. *et al.* WERAM: A database of writers, erasers and readers of histone acetylation and methylation in eukaryotes. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkw1011

38. Tian, X. *et al.* Histone Lysine-Specific Methyltransferases and Demethylases in Carcinogenesis: New Targets for Cancer Therapy and Prevention. *Curr. Cancer Drug Targets* (2013). doi:10.2174/1568009611313050007

39. Lee, J.-H. & Skalnik, D. G. Wdr82 Is a C-Terminal Domain-Binding Protein That Recruits the Setd1A Histone H3-Lys4 Methyltransferase Complex to Transcription Start Sites of Transcribed Human Genes. *Mol. Cell. Biol.* (2008). doi:10.1128/mcb.01356-07

40. Keilhack, H. L07.04 * PRECLINICAL CHARACTERIZATION OF E7438, A POTENT, SELECTIVE EZH2 INHIBITOR WITH ROBUST ANTITUMOR ACTIVITY AGAINST EZH2 MUTATED NHL XENOGRAFTS IN MICE. *Ann. Oncol.* (2013). doi:10.1093/annonc/mdt042.33

41. Gulati, N., Béguelin, W. & Giulino-Roth, L. Enhancer of zeste homolog 2 (EZH2) inhibitors. *Leukemia and Lymphoma* (2018). doi:10.1080/10428194.2018.1430795

42. Yu, T. *et al.* The EZH2 inhibitor GSK343 suppresses cancer stem-like phenotypes and reverses mesenchymal transition in glioma cells. *Oncotarget* (2017). doi:10.18632/oncotarget.21311

43. Taplin, M.-E. *et al.* ProSTAR: A phase Ib/II study of CPI-1205, a small molecule inhibitor of EZH2, combined with enzalutamide (E) or abiraterone/prednisone (A/P) in patients with metastatic castration-resistant prostate cancer (mCRPC). *J. Clin. Oncol.* (2019). doi:10.1200/jco.2019.37.7_suppl.tps335

44. Zeng, D., Liu, M. & Pan, J. Blocking EZH2 methylation transferase activity by GSK126 decreases stem cell-like myeloma cells. *Oncotarget* (2017). doi:10.18632/oncotarget.13773

45. Konze, K. D. *et al.* An orally bioavailable chemical probe of the lysine methyltransferases EZH2 and EZH1. *ACS Chem. Biol.* (2013). doi:10.1021/cb400133j

46. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* (2008). doi:10.1038/sj.emboj.7601967

47. Fumasoni, I. *et al.* Family expansion and gene rearrangements contributed to the functional specialization of PRDM genes in vertebrates. *BMC Evol. Biol.* (2007). doi:10.1186/1471-2148-7-187

48. Di Zazzo, E., Porcile, C., Bartollino, S. & Moncharmont, B. Critical Function of PRDM2 in the Neoplastic Growth of Testicular Germ Cell Tumors. *Biology (Basel).* (2016). doi:10.3390/biology5040054

49. Pandzic, T. *et al.* Somatic *PRDM2* c.4467delA mutations in colorectal cancers control histone methylation and tumor growth. *Oncotarget* (2017). doi:10.18632/oncotarget.21713

50. Kim, K. C., Geng, L. & Huang, S. Inactivation of a Histone Methyltransferase by Mutations in Human Cancers. *Cancer Res.* (2003).

51. Porter, L. F. *et al.* A role for repressive complexes and H3K9 di-methylation in PRDM5-associated brittle cornea syndrome. *Hum. Mol. Genet.* (2015). doi:10.1093/hmg/ddv345

52. Ciechomska, I. A., Przanowski, P., Jackl, J., Wojtas, B. & Kaminska, B. BIX01294, an inhibitor of histone methyltransferase, induces autophagy-

dependent differentiation of glioma stem-like cells. *Sci. Rep.* (2016). doi:10.1038/srep38723

53. Pappano, W. N. *et al.* The Histone methyltransferase inhibitor A-366 uncovers a role for G9a/GLP in the epigenetics of leukemia. *PLoS One* (2015). doi:10.1371/journal.pone.0131716

54. Devkota, K. *et al.* Analogues of the natural product sinefungin as inhibitors of EHMT1 and EHMT2. *ACS Med. Chem. Lett.* (2014). doi:10.1021/ml4002503

55. Vedadi, M. *et al.* A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nat. Chem. Biol.* (2011). doi:10.1038/nchembio.599

56. Liu, F. *et al.* Protein lysine methyltransferase g9a inhibitors: Design, synthesis, and structure activity relationships of 2,4-diamino-7-aminoalkoxy-quinazolines. *J. Med. Chem.* (2010). doi:10.1021/jm100478y

57. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–73 (2010).

58. Bromberg, K. D. *et al.* The SUV4-20 inhibitor A-196 verifies a role for epigenetics in genomic integrity. *Nat. Chem. Biol.* (2017). doi:10.1038/nchembio.2282

59. Ferguson, A. D. *et al.* Article Structural Basis of Substrate Methylation and Inhibition of SMYD2. *Struct. Des.* **19**, 1262–1273 (2011).

60. Eggert, E. *et al.* Discovery and Characterization of a Highly Potent and Selective Aminopyrazoline-Based in Vivo Probe (BAY-598) for the Protein Lysine Methyltransferase SMYD2. *J. Med. Chem.* (2016). doi:10.1021/acs.jmedchem.5b01890

61. Nguyen, H. *et al.* LLY-507, a cell-active, potent, and selective inhibitor of protein-lysine methyltransferase SMYD2. *J. Biol. Chem.* (2015). doi:10.1074/jbc.M114.626861

62. Levy, D. *et al.* Lysine methylation of the NF-κ B subunit RelA by SETD6 couples activity of the histone methyltransferase GLP at chromatin to tonic repression of NF-κ B signaling. *Nat. Immunol.* (2011). doi:10.1038/ni.1968

63. Barsyte-Lovejoy, D. *et al.* (R)-PFI-2 is a potent and selective inhibitor of SETD7 methyltransferase activity in cells. *Proc. Natl. Acad. Sci.* (2014). doi:10.1073/pnas.1407358111

64. Yu, W. *et al.* Catalytic site remodelling of the DOT1L methyltransferase by selective inhibitors. *Nat. Commun.* (2012). doi:10.1038/ncomms2304

65. Daigle, S. R. *et al.* Selective Killing of Mixed Lineage Leukemia Cells by a Potent Small-Molecule DOT1L Inhibitor. *Cancer Cell* (2011). doi:10.1016/j.ccr.2011.06.009

66. Stein, E. M. *et al.* A Phase 1 Study of the DOT1L Inhibitor, Pinometostat (EPZ-5676), in Adults with Relapsed or Refractory Leukemia: Safety, Clinical Activity, Exposure and Target Inhibition. *Blood* (2015).

67. Schmidt, D. M. Z. & McCafferty, D. G. trans-2-phenylcyclopropylamine is a mechanism-based inactivator of the histone demethylase LSD1. *Biochemistry* (2007). doi:10.1021/bi0618621

68. Mimasu, S. *et al.* Structurally designed trans -2-phenylcyclopropylamine derivatives potently inhibit histone demethylase LSD1/KDM1. *Biochemistry* (2010). doi:10.1021/bi100299r

69. Huang, Y. *et al.* Inhibition of lysine-specific demethylase 1 by polyamine analogues results in reexpression of aberrantly silenced genes. *Proc. Natl. Acad. Sci.* (2007). doi:10.1073/pnas.0700720104

70. Chen, Y. K. *et al.* Design of KDM4 Inhibitors with Antiproliferative Effects in Cancer Models. *ACS Med. Chem. Lett.* (2017). doi:10.1021/acsmedchemlett.7b00220

71. MacKeen, M. M. *et al.* Small-molecule-based inhibition of histone demethylation in cells assessed by quantitative mass spectrometry. *J. Proteome Res.* (2010). doi:10.1021/pr100269b

72. Hamada, S. *et al.* Synthesis and activity of N-oxalylglycine and its derivatives as Jumonji C-domain-containing histone lysine demethylase inhibitors. *Bioorganic Med. Chem. Lett.* (2009). doi:10.1016/j.bmcl.2009.03.098

73. Feng, T. *et al.* Identification of novel JMJD2A inhibitor scaffold using shape and electrostatic similarity search combined with docking method and MM-GBSA approach. *RSC Adv.* (2015). doi:10.1039/c5ra11896d

74. Luo, X. *et al.* A selective inhibitor and probe of the cellular functions of jumonji C domain-containing histone demethylases. *J. Am. Chem. Soc.* (2011). doi:10.1021/ja201597b

75. Heinemann, B. *et al.* Inhibition of demethylases by GSK-J1/J4. *Nature* (2014). doi:10.1038/nature13688

76. Boswell, R. E. & Mahowald, A. P. tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. *Cell* (1985). doi:10.1016/0092-8674(85)90015-

7

77. Chen, C., Nott, T. J., Jin, J. & Pawson, T. Deciphering arginine methylation: Tudor tells the tale. *Nature Reviews Molecular Cell Biology* (2011). doi:10.1038/nrm3185

78. Botuyan, M. V. *et al.* Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair. *Cell* (2006). doi:10.1016/j.cell.2006.10.043

79. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* (2011). doi:10.1038/nature09784

80. Bernstein, E. *et al.* Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. *Mol. Cell. Biol.* (2006). doi:10.1128/mcb.26.7.2560-2569.2006

81. Hughes, R. M., Wiggins, K. R., Khorasanizadeh, S. & Waters, M. L. Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proc. Natl. Acad. Sci.* (2007). doi:10.1073/pnas.0610850104

82. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* (2001). doi:10.1038/35065138

83. Musselman, C. A., Lalonde, M. E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nature Structural and Molecular Biology* (2012). doi:10.1038/nsmb.2436

84. Shaw, N. *et al.* The multifunctional human p100 protein 'hooks' methylated ligands. *Nat. Struct. Mol. Biol.* (2007). doi:10.1038/nsmb1269

85. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* (2007). doi:10.1016/j.cell.2007.05.009

86. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* (2000). doi:10.1038/35020506

87. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* (2001). doi:10.1038/35065132

88. Loyola, A. *et al.* The HP1α-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Rep.* (2009). doi:10.1038/embor.2009.90

89. Woo, H., Ha, S. D., Lee, S. B., Buratowski, S. & Kim, T. S. Modulation of gene

expression dynamics by co-Transcriptional histone methylations. *Experimental and Molecular Medicine* (2017). doi:10.1038/emm.2017.19

90. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* (2007). doi:10.1038/ng1966

91. Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications. *Science (80-.).* **4428**, 1–8 (2016).

92. Jørgensen, S., Schotta, G. & Sørensen, C. S. Histone H4 Lysine 20 methylation: Key player in epigenetic regulation of genomic integrity. *Nucleic Acids Research* (2013). doi:10.1093/nar/gkt012

93. Rao, R. C. & Dou, Y. Hijacked in cancer: The KMT2 (MLL) family of methyltransferases. *Nature Reviews Cancer* (2015). doi:10.1038/nrc3929

94. Tenney, K. & Shilatifard, A. A COMPASS in the voyage of defining the role of trithorax/MLL-containing complexes: Linking leukemogensis to covalent modifications of chromatin. *Journal of Cellular Biochemistry* (2005). doi:10.1002/jcb.20421

95. Thiel, A. T. *et al.* MLL-AF9-Induced Leukemogenesis Requires Coexpression of the Wild-Type Mll Allele. *Cancer Cell* (2010). doi:10.1016/j.ccr.2009.12.034

96. Cierpicki, T. & Grembecka, J. Challenges and opportunities in targeting the menin-MLL interaction. *Future Medicinal Chemistry* (2014). doi:10.4155/fmc.13.214

97. Yamane, K. *et al.* PLU-1 Is an H3K4 Demethylase Involved in Transcriptional Repression and Breast Cancer Cell Proliferation. *Mol. Cell* (2007). doi:10.1016/j.molcel.2007.03.001

98. West, L. E. & Gozani, O. Regulation of p53 function by lysine methylation. *Epigenomics* (2011). doi:10.2217/epi.11.21

99. Huang, J. *et al.* G9a and Glp methylate lysine 373 in the tumor suppressor p53. *J. Biol. Chem.* (2010). doi:10.1074/jbc.M109.062588

100. Shi, X. *et al.* Article Modulation of p53 Function by SET8-Mediated Methylation at Lysine 382. 636–646 (2007). doi:10.1016/j.molcel.2007.07.012

101. Lohrum, M. A. E., Woods, D. B., Ludwig, R. L., Balint, E. & Vousden, K. H. C-Terminal Ubiquitination of p53 Contributes to Nuclear Export. *Mol. Cell. Biol.* (2001). doi:10.1128/mcb.21.24.8521-8532.2001

102. Yang, J. *et al.* Stat3 activation is limiting for reprogramming to ground state pluripotency. *Cell Stem Cell* (2010). doi:10.1016/j.stem.2010.06.022

103. Kim, E. *et al.* Phosphorylation of EZH2 Activates STAT3 Signaling via STAT3 Methylation and Promotes Tumorigenicity of Glioblastoma Stem-like Cells. *Cancer Cell* **23**, 839–852 (2013).

104. Yang, J. *et al.* Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. *Proc. Natl. Acad. Sci.* (2010). doi:10.1073/pnas.1016147107

105. Maere, S., Heymans, K. & Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* (2005). doi:10.1093/bioinformatics/bti551

106. Liu, H. *et al.* Systematic identification of methyllysine-driven interactions for histone and nonhistone targets. *J. Proteome Res.* (2010). doi:10.1021/pr100597b

107. Carlson, S. M., Moore, K. E., Green, E. M., Martín, G. M. & Gozani, O. Proteome-wide enrichment of proteins modified by lysine methylation. *Nat. Protoc.* **9**, 37–50 (2014).

108. Uhlmann, T. *et al.* A Method for Large-scale Identification of Protein Arginine Methylation. *Mol. Cell. Proteomics* (2012). doi:10.1074/mcp.m112.020743

109. Ong, S. E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* (2004). doi:10.1038/nmeth715

110. Kaneko, T., Joshi, R., Feller, S. M. & Li, S. S. C. Phosphotyrosine recognition domains: The typical, the atypical and the versatile. *Cell Communication and Signaling* (2012). doi:10.1186/1478-811X-10-32

111. Bian, Y. *et al.* Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder. *Nat. Chem. Biol.* (2016). doi:10.1038/nchembio.2178

112. Shi, S.-P., Xu, H.-D., Wen, P.-P. & Qiu, J.-D. Progress and challenges in predicting protein methylation sites. *Mol. BioSyst.* **11**, 2610–2619 (2015).

113. Carpino, L. A. & Han, G. Y. The 9-Fluorenylmethoxycarbonyl Function, a New Base-Sensitive Amino-Protecting Group. *J. Am. Chem. Soc.* (1970). doi:10.1021/ja00722a043

114. Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* (2011). doi:10.1038/nprot.2011.305

115. Qin, S. *et al.* Tudor domains of the PRC2 components PHF1 and PHF19 selectively bind to histone H3K36me3. *Biochem. Biophys. Res. Commun.* (2013). doi:10.1016/j.bbrc.2012.11.116

116. Guo, Y. *et al.* Methylation-state-specific recognition of histones by the MBT repeat protein L3MBTL2. *Nucleic Acids Res.* (2009). doi:10.1093/nar/gkp086

117. Kaustov, L. *et al.* Recognition and specificity determinants of the human Cbx chromodomains. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.191411

118. Sims, R. J. *et al.* Human but Not Yeast CHD1 Binds Directly and Selectively to Histone H3 Methylated at Lysine 4 via Its Tandem Chromodomains. *J. Biol. Chem.* **280**, 41789–41792 (2005).

119. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* (2004). doi:10.1101/gr.849004

120. Huyen, Y. *et al.* Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* (2004). doi:10.1038/nature03114

121. Palmer, H. M. Using Antibodies: A Laboratory Manual. *J. Antimicrob. Chemother.* (2002). doi:10.1093/jac/45.3.413

122. Miller, T. C. R. *et al.* Competitive binding of a benzimidazole to the histone-binding pocket of the pygo PHD finger. *ACS Chem. Biol.* (2014). doi:10.1021/cb500585s

123. Ma, B. *et al.* PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* (2003). doi:10.1002/rcm.1196

124. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* (2008). doi:10.1038/nbt.1511

125. Li, L. *et al.* Prediction of phosphotyrosine signaling networks using a scoring matrix-assisted ligand identification approach. *Nucleic Acids Res.* (2008). doi:10.1093/nar/gkn161

126. Kudithipudi, S., Kusevic, D., Weirich, S. & Jeltsch, A. Specificity Analysis of Protein Lysine Methyltransferases Using SPOT Peptide Arrays. *J. Vis. Exp.* (2014). doi:10.3791/52203

127. Hilpert, K., Winkler, D. F. H. & Hancock, R. E. W. Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nat. Protoc.* (2007).

doi:10.1038/nprot.2007.160

128. Corzo, J. & Santamaria, M. Time, the forgotten dimension of ligand binding teaching. *Biochem. Mol. Biol. Educ.* (2006). doi:10.1002/bmb.2006.494034062678

129. Wang, X. *et al.* Improved protein arrays for quantitative systems analysis of the dynamics of signaling pathway interactions. *Proteome Sci.* (2011). doi:10.1186/1477-5956-9-53

130. Taylor, S. C., Berkelman, T., Yadav, G. & Hammond, M. A defined methodology for reliable quantification of western blot data. *Mol. Biotechnol.* (2013). doi:10.1007/s12033-013-9672-6

131. Moerke, N. J. Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding. in *Current Protocols in Chemical Biology* (2009). doi:10.1002/9780470559277.ch090102

132. Zhang, X. X. *et al.* PH-sensitive fluorescent dyes: Are they really ph-sensitive in cells? *Mol. Pharm.* (2013). doi:10.1021/mp3006903

133. Drané, P. *et al.* TIRR regulates 53BP1 by masking its histone methyl-lysine binding function. *Nature* (2017). doi:10.1038/nature21358

134. Botuyan, M. V. *et al.* Mechanism of 53bp1 activity regulation by rna-binding tirr and a designer protein. *Nat. Struct. Mol. Biol.* (2018). doi:10.1038/s41594-018-0083-z

135. Chen, Z. *et al.* Integration of A Deep Learning Classifier with A Random Forest Approach for Predicting Malonylation Sites. *Genomics, Proteomics Bioinforma.* **16**, 451–459 (2018).

136. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci.* (2007). doi:10.1073/pnas.0608638104

137. Kaneko, T. *et al.* Superbinder SH2 domains act as antagonists of cell signaling. *Sci. Signal.* (2012). doi:10.1126/scisignal.2003021

138. Schiapparelli, L. M. *et al.* Direct detection of biotinylated proteins by mass spectrometry. *J. Proteome Res.* (2014). doi:10.1021/pr5002862

139. Pankow, S., Bamberger, C., Calzolari, D., Bamberger, A. & Yates, J. R. Deep interactome profiling of membrane proteins by co-interacting protein identification technology. *Nat. Protoc.* (2016). doi:10.1038/nprot.2016.140

140. Cao, X. J. & Garcia, B. A. Global proteomics analysis of protein lysine

methylation. *Curr. Protoc. Protein Sci.* (2016). doi:10.1002/cpps.16

141. Gingras, A. C., Aebersold, R. & Raught, B. Advances in protein complex analysis using mass spectrometry. *Journal of Physiology* (2005). doi:10.1113/jphysiol.2004.080440

142. Cunningham, R., Wang, J., Wellner, D. & Li, L. Investigation and reduction of sub-microgram peptide loss using molecular weight cut-off fractionation prior to mass spectrometric analysis. *J. Mass Spectrom.* (2012). doi:10.1002/jms.3069

143. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* (2009). doi:10.1038/nmeth.1322

144. FRASER, D. & POWELL, R. E. The kinetics of trypsin digestion. *J. Biol. Chem.* (1950).

145. Hodge, K., Have, S. Ten, Hutton, L. & Lamond, A. I. Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *J. Proteomics* (2013). doi:10.1016/j.jprot.2013.02.023

146. Välikangas, T., Suomi, T. & Elo, L. L. A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Brief. Bioinform.* (2017). doi:10.1093/bib/bbx054

147. Kim, M. S., Zhong, J. & Pandey, A. Common errors in mass spectrometry-based analysis of post-translational modifications. *Proteomics* (2016). doi:10.1002/pmic.201500355

148. Müller, M. M. & Muir, T. W. Histones: At the crossroads of peptide and protein chemistry. *Chemical Reviews* (2015). doi:10.1021/cr5003529

149. Sidoli, S., Bhanu, N. V., Karch, K. R., Wang, X. & Garcia, B. A. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *J. Vis. Exp.* (2016). doi:10.3791/54112

150. Bobály, B., Mikola, V., Sipkó, E., Márta, Z. & Fekete, J. Recovery of proteins affected by mobile phase trifluoroacetic acid concentration in reversed-phase chromatography. *J. Chromatogr. Sci.* (2015). doi:10.1093/chromsci/bmu169

151. El-Awady, R. A. *et al.* Modulation of DNA damage response and induction of apoptosis mediates synergism between doxorubicin and a new imidazopyridine derivative in breast and lung cancer cells. *DNA Repair (Amst).* (2016). doi:10.1016/j.dnarep.2015.10.004

6 Supplemental Data

Table S.1. List of peptides used in the 70-protein array.

Position	Peptide Sequence	Protein	Site	Modification
1	EETRGVLKVFLENVI	H4K59	k59	kme3
2	TSLARFLKKTLSEKK	MYO15A	k577	kme3
3	IKHLDFLKEIKWFAV	SAMD9L	k436	kme3
4	DRMTYAIKNFVEEKM	DNAH17	k3829	kme3
5	AQEREQIKVLNDKFA	KRT74	k145	kme3
6	AVLHSFQKQNVTIMD	iNOS	k422	kme3
7	TKRKMNLKIQELRRQ	CCDC150	k212	kme3
8	ILASKYLKMLKEEKR	CHRAC1	k105	kme3
9	KESKSGLKIIKLTDS	DNAH6	k3099	kme3
10	QEIKGEVKVLNNITN	MSR1	k245	kme3
11	KKQKEIIKKLIERKQ	PCAF	k671	kme3
12	RAHSSHLKSKKGQST	p53	k370	kme3
13	KGLGMQLKGPLGPGG	AVEN	k230	kme3
14	RTKQTARKSTGGKAP	H3K9	k9	kme3
15	EETRGVLKVFLENVI	H4K59	k59	kme2
16	TSLARFLKKTLSEKK	MYO15A	k577	kme2
17	IKHLDFLKEIKWFAV	SAMD9L	k436	kme2
18	DRMTYAIKNFVEEKM	DNAH17	k3829	kme2
19	AQEREQIKVLNDKFA	KRT74	k145	kme2
20	AVLHSFQKQNVTIMD	iNOS	k422	kme2
21	TKRKMNLKIQELRRQ	CCDC150	k212	kme2
22	ILASKYLKMLKEEKR	CHRAC1	k105	kme2
23	KESKSGLKIIKLTDS	DNAH6	k3099	kme2
24	QEIKGEVKVLNNITN	MSR1	k245	kme2

25	KKQKEIIKKLIERKQ	PCAF	k671	kme2
26	RAHSSHLKSKKGQST	p53	k370	kme2
27	KGLGMQLKGPLGPGG	AVEN	k230	kme2
28	RTKQTARKSTGGKAP	H3K9	k9	kme2
29	EETRGVLKVFLENVI	H4K59	k59	kme1
30	TSLARFLKKTLSEKK	MYO15A	k577	kme1
31	IKHLDFLKEIKWFAV	SAMD9L	k436	kme1
32	DRMTYAIKNFVEEKM	DNAH17	k3829	kme1
33	AQEREQIKVLNDKFA	KRT74	k145	kme1
34	AVLHSFQKQNVTIMD	iNOS	k422	kme1
35	TKRKMNLKIQELRRQ	CCDC150	k212	kme1
36	ILASKYLKMLKEEKR	CHRAC1	k105	kme1
37	KESKSGLKIIKLTDS	DNAH6	k3099	kme1
38	QEIKGEVKVLNNITN	MSR1	k245	kme1
39	KKQKEIIKKLIERKQ	PCAF	k671	kme1
40	RAHSSHLKSKKGQST	p53	k370	kme1
41	KGLGMQLKGPLGPGG	AVEN	k230	kme1
42	RTKQTARKSTGGKAP	H3K9	k9	kme1
43	EETRGVLKVFLENVI	H4K59	k59	kme0
44	TSLARFLKKTLSEKK	MYO15A	k577	kme0
45	IKHLDFLKEIKWFAV	SAMD9L	k436	kme0
46	DRMTYAIKNFVEEKM	DNAH17	k3829	kme0
47	AQEREQIKVLNDKFA	KRT74	k145	kme0
48	AVLHSFQKQNVTIMD	iNOS	k422	kme0
49	TKRKMNLKIQELRRQ	CCDC150	k212	kme0
50	ILASKYLKMLKEEKR	CHRAC1	k105	kme0
51	KESKSGLKIIKLTDS	DNAH6	k3099	kme0
52	QEIKGEVKVLNNITN	MSR1	k245	kme0
53	KKQKEIIKKLIERKQ	PCAF	k671	kme0

54	RAHSSHLKSKKGQST	p53	k370	kme0
55	KGLGMQLKGPLGPGG	AVEN	k230	kme0
56	RTKQTARKSTGGKAP	H3K9	k9	kme0
57	SSPTNFSKLISNGYK	LOC136288	k215	kme3
58	KHRLLFFKHRLQCMT	B4GALNT1	k524	kme3
59	MGKEDFTKIPHGVSG	CRMP5	k343	kme3
60	GPDEAKIKALLERTG	NSAP1	k125	kme3
61	ISTEDFGKLWLSFAN	AP4E1	k1028	kme3
62	CQKPSTSKVILRAVA	HILS1	k119	kme3
63	IRHGKFQKMTLKLIL	MRGPRB3	k288	kme3
64	GRLAVFTKATLTTVQ	PRLR iso 4	k369	kme3
65	NISKKEYKLLYSMKE	URG4	k371	kme3
66	KKVWIGIKKLLMLIE	NSF	k708	kme3
67	LYICDFHKNFIQSVR	SAP30L	k78	kme3
68	LQPEMHSKEQILELL	ZNF306	k82	kme3
69	QMNVYHFKKGTEICN	WIPI1	k89	kme3
70	QSTSRHKKLMFKTEG	P53	k382	kme3
71	SSPTNFSKLISNGYK	LOC136288	k215	kme2
72	KHRLLFFKHRLQCMT	B4GALNT1	k524	kme2
73	MGKEDFTKIPHGVSG	CRMP5	k343	kme2
74	GPDEAKIKALLERTG	NSAP1	k125	kme2
75	ISTEDFGKLWLSFAN	AP4E1	k1028	kme2
76	CQKPSTSKVILRAVA	HILS1	k119	kme2
77	IRHGKFQKMTLKLIL	MRGPRB3	k288	kme2
78	GRLAVFTKATLTTVQ	PRLR iso 4	k369	kme2
79	NISKKEYKLLYSMKE	URG4	k371	kme2
80	KKVWIGIKKLLMLIE	NSF	k708	kme2
81	LYICDFHKNFIQSVR	SAP30L	k78	kme2
82	LQPEMHSKEQILELL	ZNF306	k82	kme2

83	QMNVYHFKKGTEICN	WIPI1	k89	kme2
84	QSTSRHKKLMFKTEG	P53	k382	kme2
85	SSPTNFSKLISNGYK	LOC136288	k215	kme1
86	KHRLLFFKHRLQCMT	B4GALNT1	k524	kme1
87	MGKEDFTKIPHGVSG	CRMP5	k343	kme1
88	GPDEAKIKALLERTG	NSAP1	k125	kme1
89	ISTEDFGKLWLSFAN	AP4E1	k1028	kme1
90	CQKPSTSKVILRAVA	HILS1	k119	kme1
91	IRHGKFQKMTLKLIL	MRGPRB3	k288	kme1
92	GRLAVFTKATLTTVQ	PRLR iso 4	k369	kme1
93	NISKKEYKLLYSMKE	URG4	k371	kme1
94	KKVWIGIKKLLMLIE	NSF	k708	kme1
95	LYICDFHKNFIQSVR	SAP30L	k78	kme1
96	LQPEMHSKEQILELL	ZNF306	k82	kme1
97	QMNVYHFKKGTEICN	WIPI1	k89	kme1
98	QSTSRHKKLMFKTEG	P53	k382	kme1
99	SSPTNFSKLISNGYK	LOC136288	k215	kme0
100	KHRLLFFKHRLQCMT	B4GALNT1	k524	kme0
101	MGKEDFTKIPHGVSG	CRMP5	k343	kme0
102	GPDEAKIKALLERTG	NSAP1	k125	kme0
103	ISTEDFGKLWLSFAN	AP4E1	k1028	kme0
104	CQKPSTSKVILRAVA	HILS1	k119	kme0
105	IRHGKFQKMTLKLIL	MRGPRB3	k288	kme0
106	GRLAVFTKATLTTVQ	PRLR iso 4	k369	kme0
107	NISKKEYKLLYSMKE	URG4	k371	kme0
108	KKVWIGIKKLLMLIE	NSF	k708	kme0
109	LYICDFHKNFIQSVR	SAP30L	k78	kme0
110	LQPEMHSKEQILELL	ZNF306	k82	kme0
111	QMNVYHFKKGTEICN	WIPI1	k89	kme0

112	QSTSRHKKLMFKTEG	P53	k382	kme0
113	ESRRCFVKVRAYRSE	CILP	k755	kme3
114	LFIYIFTKISVDMYA	SLC5A11	k157	kme3
115	FLNDSYLKYVGWTLH	STAG1	k333	kme3
116	FSVKGHVKMLRLALT	CKLF	k22	kme3
117	KYIKSHYKVGENADS	H1F0	k59	kme3
118	VASASSIKQILLEWC	SMTNL2	k354	kme3
119	EISQRLLKLYSDKFG	DOCK9	k1834	kme3
120	QEIRYRSKLKLIRAK	PDGFRA	k378	kme3
121	EGKKWQAKIEGIRNK	CEP290	k1921	kme3
122	NMDNLIYKLLKPSTK	B3GALT1	k190	kme3
123	NEVTEFAKTCVADES	Albumin	k755	kme3
124	PMDMSTIKSKLEARE	BRD4	k404	kme3
125	YQELMNAKLGLDIEI	KRT84	k455	kme3
126	GGAKRHRKVLRDNIQ	H4K20	k20	kme3
127	ESRRCFVKVRAYRSE	CILP	k755	kme2
128	LFIYIFTKISVDMYA	SLC5A11	k157	kme2
129	FLNDSYLKYVGWTLH	STAG1	k333	kme2
130	FSVKGHVKMLRLALT	CKLF	k22	kme2
131	KYIKSHYKVGENADS	H1F0	k59	kme2
132	VASASSIKQILLEWC	SMTNL2	k354	kme2
133	EISQRLLKLYSDKFG	DOCK9	k1834	kme2
134	QEIRYRSKLKLIRAK	PDGFRA	k378	kme2
135	EGKKWQAKIEGIRNK	CEP290	k1921	kme2
136	NMDNLIYKLLKPSTK	B3GALT1	k190	kme2
137	NEVTEFAKTCVADES	Albumin	k755	kme2
138	PMDMSTIKSKLEARE	BRD4	k404	kme2
139	YQELMNAKLGLDIEI	KRT84	k455	kme2
140	GGAKRHRKVLRDNIQ	H4K20	k20	kme2

141	ESRRCFVKVRAYRSE	CILP	k755	kme1
142	LFIYIFTKISVDMYA	SLC5A11	k157	kme1
143	FLNDSYLKYVGWTLH	STAG1	k333	kme1
144	FSVKGHVKMLRLALT	CKLF	k22	kme1
145	KYIKSHYKVGENADS	H1F0	k59	kme1
146	VASASSIKQILLEWC	SMTNL2	k354	kme1
147	EISQRLLKLYSDKFG	DOCK9	k1834	kme1
148	QEIRYRSKLKLIRAK	PDGFRA	k378	kme1
149	EGKKWQAKIEGIRNK	CEP290	k1921	kme1
150	NMDNLIYKLLKPSTK	B3GALT1	k190	kme1
151	NEVTEFAKTCVADES	Albumin	k755	kme1
152	PMDMSTIKSKLEARE	BRD4	k404	kme1
153	YQELMNAKLGLDIEI	KRT84	k455	kme1
154	GGAKRHRKVLRDNIQ	H4K20	k20	kme1
155	ESRRCFVKVRAYRSE	CILP	k755	kme0
156	LFIYIFTKISVDMYA	SLC5A11	k157	kme0
157	FLNDSYLKYVGWTLH	STAG1	k333	kme0
158	FSVKGHVKMLRLALT	CKLF	k22	kme0
159	KYIKSHYKVGENADS	H1F0	k59	kme0
160	VASASSIKQILLEWC	SMTNL2	k354	kme0
161	EISQRLLKLYSDKFG	DOCK9	k1834	kme0
162	QEIRYRSKLKLIRAK	PDGFRA	k378	kme0
163	EGKKWQAKIEGIRNK	CEP290	k1921	kme0
164	NMDNLIYKLLKPSTK	B3GALT1	k190	kme0
165	NEVTEFAKTCVADES	Albumin	k755	kme0
166	PMDMSTIKSKLEARE	BRD4	k404	kme0
167	YQELMNAKLGLDIEI	KRT84	k455	kme0
168	GGAKRHRKVLRDNIQ	H4K20	k20	kme0
169	KKLADYLKVLIDNKH	RUFY1	k225	kme3

170	KGRTDFIKGMKKKSR	VWA5B1	k311	kme3
171	FCEESFVKHRSSVMK	DENND1B	k362	kme3
172	IQRELFSKLGELAVG	PLZF	k387	kme3
173	MLSSVFQKQFYRLGG	LYST	k979	kme3
174	ELYISIAKCLLEMTD	FOCAD	k1558	kme3
175	ERQKFCFKVFDVDRD	USP32	k274	kme3
176	VEMDWVLKHTGPNSP	hnRNP H2	k98	kme3
177	QRSRYFKKAIPINNK	CBWD2 iso3	k204	kme3
178	DSCPAVSKILERSLK	REV3	k2868	kme3
179	SRSVSLLKGFAQDSQ	C2CD4C	k237	kme3
180	YVNIYGLKIWQEEVS	KIAA0196	k768	kme3
181	RETKCMIKMKLDVPE	DNAH	k617	kme3
182	LESYLHAKKYLKPSG	CARM1	k276	kme3
183	KKLADYLKVLIDNKH	RUFY1	k225	kme2
184	KGRTDFIKGMKKKSR	VWA5B1	k311	kme2
185	FCEESFVKHRSSVMK	DENND1B	k362	kme2
186	IQRELFSKLGELAVG	PLZF	k387	kme2
187	MLSSVFQKQFYRLGG	LYST	k979	kme2
188	ELYISIAKCLLEMTD	FOCAD	k1558	kme2
189	ERQKFCFKVFDVDRD	USP32	k274	kme2
190	VEMDWVLKHTGPNSP	hnRNP H2	k98	kme2
191	QRSRYFKKAIPINNK	CBWD2 iso3	k204	kme2
192	DSCPAVSKILERSLK	REV3	k2868	kme2
193	SRSVSLLKGFAQDSQ	C2CD4C	k237	kme2
194	YVNIYGLKIWQEEVS	KIAA0196	k768	kme2
195	RETKCMIKMKLDVPE	DNAH	k617	kme2
196	LESYLHAKKYLKPSG	CARM1	k276	kme2
197	KKLADYLKVLIDNKH	RUFY1	k225	kme1
198	KGRTDFIKGMKKKSR	VWA5B1	k311	kme1

199	FCEESFVKHRSSVMK	DENND1B	k362	kme1
200	IQRELFSKLGELAVG	PLZF	k387	kme1
201	MLSSVFQKQFYRLGG	LYST	k979	kme1
202	ELYISIAKCLLEMTD	FOCAD	k1558	kme1
203	ERQKFCFKVFDVDRD	USP32	k274	kme1
204	VEMDWVLKHTGPNSP	hnRNP H2	k98	kme1
205	QRSRYFKKAIPINNK	CBWD2 iso3	k204	kme1
206	DSCPAVSKILERSLK	REV3	k2868	kme1
207	SRSVSLLKGFAQDSQ	C2CD4C	k237	kme1
208	YVNIYGLKIWQEEVS	KIAA0196	k768	kme1
209	RETKCMIKMKLDVPE	DNAH	k617	kme1
210	LESYLHAKKYLKPSG	CARM1	k276	kme1
211	KKLADYLKVLIDNKH	RUFY1	k225	kme0
212	KGRTDFIKGMKKKSR	VWA5B1	k311	kme0
213	FCEESFVKHRSSVMK	DENND1B	k362	kme0
214	IQRELFSKLGELAVG	PLZF	k387	kme0
215	MLSSVFQKQFYRLGG	LYST	k979	kme0
216	ELYISIAKCLLEMTD	FOCAD	k1558	kme0
217	ERQKFCFKVFDVDRD	USP32	k274	kme0
218	VEMDWVLKHTGPNSP	hnRNP H2	k98	kme0
219	QRSRYFKKAIPINNK	CBWD2 iso3	k204	kme0
220	DSCPAVSKILERSLK	REV3	k2868	kme0
221	SRSVSLLKGFAQDSQ	C2CD4C	k237	kme0
222	YVNIYGLKIWQEEVS	KIAA0196	k768	kme0
223	RETKCMIKMKLDVPE	DNAH	k617	kme0
224	LESYLHAKKYLKPSG	CARM1	k276	kme0
225	MDVFSFVKIAKLSSH	FRMPD4	k8	kme3
226	LKKYDQLKVYLEQNL	PTPN23	k613	kme3
227	NRLKSLMKILSEVTP	OIP5	k215	kme3

228	TKHLDFLKEIKWFAV	SAMD9	k432	kme3
229	TVRDSNLKLTLAFGI	ASCC3	k1609	kme3
230	LETINRIKLYSESLA	GDI1	k210	kme3
231	DAGKYILKLENSSGS	Titin	k25023	kme3
232	DLTRNKFKLLSGTEQ	DNAH7	k1934	kme3
233	TSRYWRIKSKNHAMS	BBS2	k239	kme3
234	SYTLQVIKLNLMSEE	CEACAM8	k126	kme3
235	SGLDMHTKPWVRARA	BHMT	k340	kme3
236	TLVLNRLKVGLQVVA	HSP60	k292	kme3
237	AKRTGMQKVLYSTAM	PIPK1B	k301	kme3
238	TSKNDFTKKESRAVS	MDM1	k573	kme3
239	MDVFSFVKIAKLSSH	FRMPD4	k8	kme2
240	LKKYDQLKVYLEQNL	PTPN23	k613	kme2
241	NRLKSLMKILSEVTP	OIP5	k215	kme2
242	TKHLDFLKEIKWFAV	SAMD9	k432	kme2
243	TVRDSNLKLTLAFGI	ASCC3	k1609	kme2
244	LETINRIKLYSESLA	GDI1	k210	kme2
245	DAGKYILKLENSSGS	Titin	k25023	kme2
246	DLTRNKFKLLSGTEQ	DNAH7	k1934	kme2
247	TSRYWRIKSKNHAMS	BBS2	k239	kme2
248	SYTLQVIKLNLMSEE	CEACAM8	k126	kme2
249	SGLDMHTKPWVRARA	BHMT	k340	kme2
250	TLVLNRLKVGLQVVA	HSP60	k292	kme2
251	AKRTGMQKVLYSTAM	PIPK1B	k301	kme2
252	TSKNDFTKKESRAVS	MDM1	k573	kme2
253	MDVFSFVKIAKLSSH	FRMPD4	k8	kme1
254	LKKYDQLKVYLEQNL	PTPN23	k613	kme1
255	NRLKSLMKILSEVTP	OIP5	k215	kme1
256	TKHLDFLKEIKWFAV	SAMD9	k432	kme1

257	TVRDSNLKLTLAFGI	ASCC3	k1609	kme1
258	LETINRIKLYSESLA	GDI1	k210	kme1
259	DAGKYILKLENSSGS	Titin	k25023	kme1
260	DLTRNKFKLLSGTEQ	DNAH7	k1934	kme1
261	TSRYWRIKSKNHAMS	BBS2	k239	kme1
262	SYTLQVIKLNLMSEE	CEACAM8	k126	kme1
263	SGLDMHTKPWVRARA	BHMT	k340	kme1
264	TLVLNRLKVGLQVVA	HSP60	k292	kme1
265	AKRTGMQKVLYSTAM	PIPK1B	k301	kme1
266	TSKNDFTKKESRAVS	MDM1	k573	kme1
267	MDVFSFVKIAKLSSH	FRMPD4	k8	kme0
268	LKKYDQLKVYLEQNL	PTPN23	k613	kme0
269	NRLKSLMKILSEVTP	OIP5	k215	kme0
270	TKHLDFLKEIKWFAV	SAMD9	k432	kme0
271	TVRDSNLKLTLAFGI	ASCC3	k1609	kme0
272	LETINRIKLYSESLA	GDI1	k210	kme0
273	DAGKYILKLENSSGS	Titin	k25023	kme0
274	DLTRNKFKLLSGTEQ	DNAH7	k1934	kme0
275	TSRYWRIKSKNHAMS	BBS2	k239	kme0
276	SYTLQVIKLNLMSEE	CEACAM8	k126	kme0
277	SGLDMHTKPWVRARA	BHMT	k340	kme0
278	TLVLNRLKVGLQVVA	HSP60	k292	kme0
279	AKRTGMQKVLYSTAM	PIPK1B	k301	kme0
280	TSKNDFTKKESRAVS	MDM1	k573	kme0

Table S.2. List of peptides used in the 140-cluster array.

Position	Peptide	Mod*	Gene Name	Site	Uniprot #
1	KEPTIKFYTKQ	0	MAP3K15	K755	Q6ZN16
2	KEPTIKFYTKQ	1	MAP3K15	K755	Q6ZN16
3	KEPTIKFYTKQ	2	MAP3K15	K755	Q6ZN16
4	KEPTIKFYTKQ	3	MAP3K15	K755	Q6ZN16
5	ALRRQKKKDTL	0	CDH12	K641	P55289
6	ALRRQKKKDTL	1	CDH12	K641	P55289
7	ALRRQKKKDTL	2	CDH12	K641	P55289
8	ALRRQKKKDTL	3	CDH12	K641	P55289
9	AVTYAKVKHSR	0	LILRB4	K362	Q8NHJ6
10	AVTYAKVKHSR	1	LILRB4	K362	Q8NHJ6
11	AVTYAKVKHSR	2	LILRB4	K362	Q8NHJ6
12	AVTYAKVKHSR	3	LILRB4	K362	Q8NHJ6
13	AKSEPKPGLPE	0	CMSS1	K88	Q9BQ75
14	AKSEPKPGLPE	1	CMSS1	K88	Q9BQ75
15	AKSEPKPGLPE	2	CMSS1	K88	Q9BQ75
16	AKSEPKPGLPE	3	CMSS1	K88	Q9BQ75
17	GNKMAKARQYP	0	SEPTIN11	K251	Q9NVA2
18	GNKMAKARQYP	1	SEPTIN11	K251	Q9NVA2
19	GNKMAKARQYP	2	SEPTIN11	K251	Q9NVA2
20	GNKMAKARQYP	3	SEPTIN11	K251	Q9NVA2
21	IIGQAKKHPSL	0	NDUFA4	K10	O00483
22	IIGQAKKHPSL	1	NDUFA4	K10	O00483
23	IIGQAKKHPSL	2	NDUFA4	K10	O00483
24	IIGQAKKHPSL	3	NDUFA4	K10	O00483
25	DGDDKKRPVII	0	TARSL2	K663	A2RTX5
26	DGDDKKRPVII	1	TARSL2	K663	A2RTX5
27	DGDDKKRPVII	2	TARSL2	K663	A2RTX5

28	DGDDKKRPVII	3	TARSL2	K663	A2RTX5
29	QSTSRHKKLMFKTEG	1	TP53BP1	K382	P04637
30	TKRKMNLKIQELRRQ	3	CCDC150	K227	Q8NCX0
31	GPDEAKIKALL	0	SYNCRIP	K123	O60506
32	GPDEAKIKALL	1	SYNCRIP	K123	O60506
33	GPDEAKIKALL	2	SYNCRIP	K123	O60506
34	GPDEAKIKALL	3	SYNCRIP	K123	O60506
35	DEGKLKAGQSV	0	CNN3	K158	Q15417
36	DEGKLKAGQSV	1	CNN3	K158	Q15417
37	DEGKLKAGQSV	2	CNN3	K158	Q15417
38	DEGKLKAGQSV	3	CNN3	K158	Q15417
39	EIDTIKHQNQE	0	ANKRD62	K579	A6NC57
40	EIDTIKHQNQE	1	ANKRD62	K579	A6NC57
41	EIDTIKHQNQE	2	ANKRD62	K579	A6NC57
42	EIDTIKHQNQE	3	ANKRD62	K579	A6NC57
43	EETGAKISVLG	0	KHDRBS1	K194	Q07666
44	EETGAKISVLG	1	KHDRBS1	K194	Q07666
45	EETGAKISVLG	2	KHDRBS1	K194	Q07666
46	EETGAKISVLG	3	KHDRBS1	K194	Q07666
47	VHSPQKSTKNH	0	AKAP10	K55	O43572
48	VHSPQKSTKNH	1	AKAP10	K55	O43572
49	VHSPQKSTKNH	2	AKAP10	K55	O43572
50	VHSPQKSTKNH	3	AKAP10	K55	O43572
51	EVEMLKGLQHP	0	WNK4	K223	Q96J92
52	EVEMLKGLQHP	1	WNK4	K223	Q96J92
53	EVEMLKGLQHP	2	WNK4	K223	Q96J92
54	EVEMLKGLQHP	3	WNK4	K223	Q96J92
55	GGEPSKKRKRR	0	POU6F1	K542	Q14863
56	GGEPSKKRKRR	1	POU6F1	K542	Q14863

57	GGEPSKKRKRR	2	POU6F1	K542	Q14863
58	GGEPSKKRKRR	3	POU6F1	K542	Q14863
59	EGKKWQAKIEGIRNK	1	CEP290	K1919	O15078
60	KKQKEIIKKLIERKQ	3	KAT2A	K674	Q92830
61	LGGAPKEPAKL	0	IRS1	K1161	P35568
62	LGGAPKEPAKL	1	IRS1	K1161	P35568
63	LGGAPKEPAKL	2	IRS1	K1161	P35568
64	LGGAPKEPAKL	3	IRS1	K1161	P35568
65	ATGEEKAATAP	0	TACC2	K1366	O95359
66	ATGEEKAATAP	1	TACC2	K1366	O95359
67	ATGEEKAATAP	2	TACC2	K1366	O95359
68	ATGEEKAATAP	3	TACC2	K1366	O95359
69	ATVNVKKDKED	0	PHF20	K177	P9BVI0
70	ATVNVKKDKED	1	PHF20	K177	P9BVI0
71	ATVNVKKDKED	2	PHF20	K177	P9BVI0
72	ATVNVKKDKED	3	PHF20	K177	P9BVI0
73	FYGTEKDKNSV	0	XRCC6	K92	P12956
74	FYGTEKDKNSV	1	XRCC6	K92	P12956
75	FYGTEKDKNSV	2	XRCC6	K92	P12956
76	FYGTEKDKNSV	3	XRCC6	K92	P12956
77	RIQEMKAKTTH	0	NKTR	K637	P30414
78	RIQEMKAKTTH	1	NKTR	K637	P30414
79	RIQEMKAKTTH	2	NKTR	K637	P30414
80	RIQEMKAKTTH	3	NKTR	K637	P30414
81	KPVSKKHGKLI	0	ARHGAP32	K338	A7KAX9
82	KPVSKKHGKLI	1	ARHGAP32	K338	A7KAX9
83	KPVSKKHGKLI	2	ARHGAP32	K338	A7KAX9
84	KPVSKKHGKLI	3	ARHGAP32	K338	A7KAX9
85	GKGLGKGGAKR	0	H4	K12	P62805

86	GKGLGKGGAKR	1	H4	K12	P62805
87	GKGLGKGGAKR	2	H4	K12	P62805
88	GKGLGKGGAKR	3	H4	K12	P62805
89	EGKKWQAKIEGIRNK	0	CEP290	K1919	O15078
90	TKRKMNLKIQELRRQ	1	CCDC150	K227	Q8NCX0
91	ARRHLKMMHIA	0	ASPM	K2260	Q8IZT6
92	ARRHLKMMHIA	1	ASPM	K2260	Q8IZT6
93	ARRHLKMMHIA	2	ASPM	K2260	Q8IZT6
94	ARRHLKMMHIA	3	ASPM	K2260	Q8IZT6
95	KDVRHKGKRGK	0	SPART	K382	Q8N0X7
96	KDVRHKGKRGK	1	SPART	K382	Q8N0X7
97	KDVRHKGKRGK	2	SPART	K382	Q8N0X7
98	KDVRHKGKRGK	3	SPART	K382	Q8N0X7
99	RDFLRKEYGGL	0	CBR1	K79	P16152
100	RDFLRKEYGGL	1	CBR1	K79	P16152
101	RDFLRKEYGGL	2	CBR1	K79	P16152
102	RDFLRKEYGGL	3	CBR1	K79	P16152
103	KMNPPKFSKVE	0	MYH10	K83	P35580
104	KMNPPKFSKVE	1	MYH10	K83	P35580
105	KMNPPKFSKVE	2	MYH10	K83	P35580
106	KMNPPKFSKVE	3	MYH10	K83	P35580
107	DKMHKKATKWM	0	CCDC125	K353	Q86Z20
108	DKMHKKATKWM	1	CCDC125	K353	Q86Z20
109	DKMHKKATKWM	2	CCDC125	K353	Q86Z20
110	DKMHKKATKWM	3	CCDC125	K353	Q86Z20
111	APPESKASTPL	0	NACA	K1768	E9PAV3
112	APPESKASTPL	1	NACA	K1768	E9PAV3
113	APPESKASTPL	2	NACA	K1768	E9PAV3
114	APPESKASTPL	3	NACA	K1768	E9PAV3

115	KDEMNKEIEAA	0	ODF2	K580	Q5BJF6
116	KDEMNKEIEAA	1	ODF2	K580	Q5BJF6
117	KDEMNKEIEAA	2	ODF2	K580	Q5BJF6
118	KDEMNKEIEAA	3	ODF2	K580	Q5BJF6
119	DLTRNKFKLLSGTEQ	1	DNAH7	K1892	Q8WXX0
120	ILASKYLKMLKEEKR	1	CHRAC1	K105	Q9NRG0
121	EKLQEKREKRR	0	KIF2A	K169	O00139
122	EKLQEKREKRR	1	KIF2A	K169	O00139
123	EKLQEKREKRR	2	KIF2A	K169	O00139
124	EKLQEKREKRR	3	KIF2A	K169	O00139
125	NDIHKKIALVK	0	DAG1	K626	Q14118
126	NDIHKKIALVK	1	DAG1	K626	Q14118
127	NDIHKKIALVK	2	DAG1	K626	Q14118
128	NDIHKKIALVK	3	DAG1	K626	Q14118
129	AEILLKKKIYK	0	IPO7	K445	O95373
130	AEILLKKKIYK	1	IPO7	K445	O95373
131	AEILLKKKIYK	2	IPO7	K445	O95373
132	AEILLKKKIYK	3	IPO7	K445	O95373
133	GGNRTKTPGPG	0	RPS14B	K107	P39516
134	GGNRTKTPGPG	1	RPS14B	K107	P39516
135	GGNRTKTPGPG	2	RPS14B	K107	P39516
136	GGNRTKTPGPG	3	RPS14B	K107	P39516
137	DNPGAKRILEL	0	XRCC6	K136	P12956
138	DNPGAKRILEL	1	XRCC6	K136	P12956
139	DNPGAKRILEL	2	XRCC6	K136	P12956
140	DNPGAKRILEL	3	XRCC6	K136	P12956
141	IHQLAKGDFGT	0	KPNA3	K379	O00505
142	IHQLAKGDFGT	1	KPNA3	K379	O00505
143	IHQLAKGDFGT	2	KPNA3	K379	O00505

144	IHQLAKGDFGT	3	KPNA3	K379	O00505
145	DFGTQKEAAWA	0	KPNA4	K386	O00629
146	DFGTQKEAAWA	1	KPNA4	K386	O00629
147	DFGTQKEAAWA	2	KPNA4	K386	O00629
148	DFGTQKEAAWA	3	KPNA4	K386	O00629
149	ARTKQTARK	0	H3	K4	P68431
150	KKQKEIIKKLIERKQ	1	KAT2A	K674	Q92830
151	AGFGSKGSSSS	0	DES	K43	P17661
152	AGFGSKGSSSS	1	DES	K43	P17661
153	AGFGSKGSSSS	2	DES	K43	P17661
154	AGFGSKGSSSS	3	DES	K43	P17661
155	ESGAGKTENTK	0	MYH11	K184	P35749
156	ESGAGKTENTK	1	MYH11	K184	P35749
157	ESGAGKTENTK	2	MYH11	K184	P35749
158	ESGAGKTENTK	3	MYH11	K184	P35749
159	GKVGRKDGQML	0	SF1	K191	Q15637
160	GKVGRKDGQML	1	SF1	K191	Q15637
161	GKVGRKDGQML	2	SF1	K191	Q15637
162	GKVGRKDGQML	3	SF1	K191	Q15637
163	EEMSEKPKKKK	0	NOP56	K478	O00567
164	EEMSEKPKKKK	1	NOP56	K478	O00567
165	EEMSEKPKKKK	2	NOP56	K478	O00567
166	EEMSEKPKKKK	3	NOP56	K478	O00567
167	AMFGPKGFGRG	0	CRIP1	K64	P50238
168	AMFGPKGFGRG	1	CRIP1	K64	P50238
169	AMFGPKGFGRG	2	CRIP1	K64	P50238
170	AMFGPKGFGRG	3	CRIP1	K64	P50238
171	KGANTKTYTLT	0	TTC7B	K302	Q86TV6
172	KGANTKTYTLT	1	TTC7B	K302	Q86TV6

173	KGANTKTYTLT	2	TTC7B	K302	Q86TV6
174	KGANTKTYTLT	3	TTC7B	K302	Q86TV6
175	GGGGGKIRTRR	0	NUP153	K16	P49790
176	GGGGGKIRTRR	1	NUP153	K16	P49790
177	GGGGGKIRTRR	2	NUP153	K16	P49790
178	GGGGGKIRTRR	3	NUP153	K16	P49790
179	ARTKQTARK	1	H3	K4	P68431
180	RTKQTARKSTGGKAP	1	H3	K9	P68431
181	EYVVEKVLDRR	0	CBX5	K24	P45973
182	EYVVEKVLDRR	1	CBX5	K24	P45973
183	EYVVEKVLDRR	2	CBX5	K24	P45973
184	EYVVEKVLDRR	3	CBX5	K24	P45973
185	AKRHRKVLRDN	0	H4	K20	P62805
186	AKRHRKVLRDN	1	H4	K20	P62805
187	AKRHRKVLRDN	2	H4	K20	P62805
188	AKRHRKVLRDN	3	H4	K20	P62805
189	ELKRAKENQKN	0	NEBL	K708	O76041
190	ELKRAKENQKN	1	NEBL	K708	O76041
191	ELKRAKENQKN	2	NEBL	K708	O76041
192	ELKRAKENQKN	3	NEBL	K708	O76041
193	DEWLMKNMDPL	0	MYH10	K594	P35580
194	DEWLMKNMDPL	1	MYH10	K594	P35580
195	DEWLMKNMDPL	2	MYH10	K594	P35580
196	DEWLMKNMDPL	3	MYH10	K594	P35580
197	IKADMKAARDI	0	FBXO38	K686	Q6PIJ6
198	IKADMKAARDI	1	FBXO38	K686	Q6PIJ6
199	IKADMKAARDI	2	FBXO38	K686	Q6PIJ6
200	IKADMKAARDI	3	FBXO38	K686	Q6PIJ6
201	PLAGNKDNKFS	0	PHF2	K6231	O75151

202	PLAGNKDNKFS	1	PHF2	K6231	O75151
203	PLAGNKDNKFS	2	PHF2	K6231	O75151
204	PLAGNKDNKFS	3	PHF2	K6231	O75151
205	AVEEEKGEELE	0	TP53BP1	K135	Q12888
206	AVEEEKGEELE	1	TP53BP1	K135	Q12888
207	AVEEEKGEELE	2	TP53BP1	K135	Q12888
208	AVEEEKGEELE	3	TP53BP1	K135	Q12888
209	ARTKQTARK	2	H3	K4	P68431
210	TKRKMNLKIQELRRQ	0	CCDC150	K227	Q8NCX0
211	FLRKLKADKEF	0	ZDHHC17	K296	Q8IUH5
212	FLRKLKADKEF	1	ZDHHC17	K296	Q8IUH5
213	FLRKLKADKEF	2	ZDHHC17	K296	Q8IUH5
214	FLRKLKADKEF	3	ZDHHC17	K296	Q8IUH5
215	FPNFDKQELRE	0	SMARCAD1	K268	Q9H4L7
216	FPNFDKQELRE	1	SMARCAD1	K268	Q9H4L7
217	FPNFDKQELRE	2	SMARCAD1	K268	Q9H4L7
218	FPNFDKQELRE	3	SMARCAD1	K268	Q9H4L7
219	ATGGVKKPHRY	0	H3	K40	P06352
220	ATGGVKKPHRY	1	H3	K40	P06352
221	ATGGVKKPHRY	2	H3	K40	P06352
222	ATGGVKKPHRY	3	H3	K40	P06352
223	NPMYQKERETP	0	SFPQ	K472	P23246
224	NPMYQKERETP	1	SFPQ	K472	P23246
225	NPMYQKERETP	2	SFPQ	K472	P23246
226	NPMYQKERETP	3	SFPQ	K472	P23246
227	GNSKSKVSSQF	0	EXOSC10	K835	Q01780
228	GNSKSKVSSQF	1	EXOSC10	K835	Q01780
229	GNSKSKVSSQF	2	EXOSC10	K835	Q01780
230	GNSKSKVSSQF	3	EXOSC10	K835	Q01780

231	HMPLLKAVLKE	0	CYP27A1	K390	Q02318
232	HMPLLKAVLKE	1	CYP27A1	K390	Q02318
233	HMPLLKAVLKE	2	CYP27A1	K390	Q02318
234	HMPLLKAVLKE	3	CYP27A1	K390	Q02318
235	IQSMLKRAPSY	0	ST5	K676	P78524
236	IQSMLKRAPSY	1	ST5	K676	P78524
237	IQSMLKRAPSY	2	ST5	K676	P78524
238	IQSMLKRAPSY	3	ST5	K676	P78524
239	ARTKQTARK	3	H3	K4	P68431
240	KKQKEIIKKLIERKQ	0	KAT2A	K674	Q92830
241	AEKLMKQIGVK	0	ATAD1	K60	Q8NBU5
242	AEKLMKQIGVK	1	ATAD1	K60	Q8NBU5
243	AEKLMKQIGVK	2	ATAD1	K60	Q8NBU5
244	AEKLMKQIGVK	3	ATAD1	K60	Q8NBU5
245	EEENSKVELKS	0	CASC3	K112	O15234
246	EEENSKVELKS	1	CASC3	K112	O15234
247	EEENSKVELKS	2	CASC3	K112	O15234
248	EEENSKVELKS	3	CASC3	K112	O15234
249	DVNSKKTLREV	0	APC	K559	P25054
250	DVNSKKTLREV	1	APC	K559	P25054
251	DVNSKKTLREV	2	APC	K559	P25054
252	DVNSKKTLREV	3	APC	K559	P25054
253	KELNGKQIYVG	0	PABPC1	K259	P11940
254	KELNGKQIYVG	1	PABPC1	K259	P11940
255	KELNGKQIYVG	2	PABPC1	K259	P11940
256	KELNGKQIYVG	3	PABPC1	K259	P11940
257	ESSPFKSGMSM	0	ADAD1	K500	Q96M93
258	ESSPFKSGMSM	1	ADAD1	K500	Q96M93
259	ESSPFKSGMSM	2	ADAD1	K500	Q96M93

260	ESSPFKSGMSM	3	ADAD1	K500	Q96M93
261	EGIMVKQPLSI	0	LIG4	K432	P49917
262	EGIMVKQPLSI	1	LIG4	K432	P49917
263	EGIMVKQPLSI	2	LIG4	K432	P49917
264	EGIMVKQPLSI	3	LIG4	K432	P49917
265	ISTSIKTARKS	0	LCOR	K450	Q96JN0
266	ISTSIKTARKS	1	LCOR	K450	Q96JN0
267	ISTSIKTARKS	2	LCOR	K450	Q96JN0
268	ISTSIKTARKS	3	LCOR	K450	Q96JN0
269	KQTARKSTGGK	0	H3	K9	P68432
270	CQKPSTSKVILRAVA	3	HILS1	K117	P60008
271	ALKTQKTSERK	0	Titin	K34143	A2ASS6
272	ALKTQKTSERK	1	Titin	K34143	A2ASS6
273	ALKTQKTSERK	2	Titin	K34143	A2ASS6
274	ALKTQKTSERK	3	Titin	K34143	A2ASS6
275	EWYRKKMPFSY	0	HS3ST5	K142	Q8IZT8
276	EWYRKKMPFSY	1	HS3ST5	K142	Q8IZT8
277	EWYRKKMPFSY	2	HS3ST5	K142	Q8IZT8
278	EWYRKKMPFSY	3	HS3ST5	K142	Q8IZT8
279	KSLLVKAEKRK	0	SAP130	K878	Q9H0E3
280	KSLLVKAEKRK	1	SAP130	K878	Q9H0E3
281	KSLLVKAEKRK	2	SAP130	K878	Q9H0E3
282	KSLLVKAEKRK	3	SAP130	K878	Q9H0E3
283	IAAKKKMKKHK	0	UPT20	K2746	O75691
284	IAAKKKMKKHK	1	UPT20	K2746	O75691
285	IAAKKKMKKHK	2	UPT20	K2746	O75691
286	IAAKKKMKKHK	3	UPT20	K2746	O75691
287	NDSYLKYVGWT	0	STAG1	K340	Q8WVM7
288	NDSYLKYVGWT	1	STAG1	K340	Q8WVM7

289	NDSYLKYVGWT	2	STAG1	K340	Q8WVM7
290	NDSYLKYVGWT	3	STAG1	K340	Q8WVM7
291	DEAKIKALLER	0	HNRNPR	K128	O43390
292	DEAKIKALLER	1	HNRNPR	K128	O43390
293	DEAKIKALLER	2	HNRNPR	K128	O43390
294	DEAKIKALLER	3	HNRNPR	K128	O43390
295	EEIKFKDRAVF	0	JAKMIP3	K218	Q5VZ66
296	EEIKFKDRAVF	1	JAKMIP3	K218	Q5VZ66
297	EEIKFKDRAVF	2	JAKMIP3	K218	Q5VZ66
298	EEIKFKDRAVF	3	JAKMIP3	K218	Q5VZ66
299	KQTARKSTGGK	1	H3	K9	P68432
300	IRHGKFQKMTLKLIL	3	MRGPRB3	K288	Q91ZC1
301	APGKKKAQYEE	0	SPAG17	K404	Q6Q759
302	APGKKKAQYEE	1	SPAG17	K404	Q6Q759
303	APGKKKAQYEE	2	SPAG17	K404	Q6Q759
304	APGKKKAQYEE	3	SPAG17	K404	Q6Q759
305	ALPGRKEQTPV	0	MSRA	K41	Q9UJ68
306	ALPGRKEQTPV	1	MSRA	K41	Q9UJ68
307	ALPGRKEQTPV	2	MSRA	K41	Q9UJ68
308	ALPGRKEQTPV	3	MSRA	K41	Q9UJ68
309	DEELNKLLGKV	0	H2A	K96	Q4R3X5
310	DEELNKLLGKV	1	H2A	K96	Q4R3X5
311	DEELNKLLGKV	2	H2A	K96	Q4R3X5
312	DEELNKLLGKV	3	H2A	K96	Q4R3X5
313	DSENIKHKNNI	0	DTHD1	K162	Q6ZMT9
314	DSENIKHKNNI	1	DTHD1	K162	Q6ZMT9
315	DSENIKHKNNI	2	DTHD1	K162	Q6ZMT9
316	DSENIKHKNNI	3	DTHD1	K162	Q6ZMT9
317	FKGFEKVKDVQ	0	TIMP4	K86	Q99727

318	FKGFEKVKDVQ	1	TIMP4	K86	Q99727
319	FKGFEKVKDVQ	2	TIMP4	K86	Q99727
320	FKGFEKVKDVQ	3	TIMP4	K86	Q99727
321	KKLAAKGLRDP	0	NDUFB3	K39	O43676
322	KKLAAKGLRDP	1	NDUFB3	K39	O43676
323	KKLAAKGLRDP	2	NDUFB3	K39	O43676
324	KKLAAKGLRDP	3	NDUFB3	K39	O43676
325	EANKEKNLEQY	0	SMC3	K429	Q9UQE7
326	EANKEKNLEQY	1	SMC3	K429	Q9UQE7
327	EANKEKNLEQY	2	SMC3	K429	Q9UQE7
328	EANKEKNLEQY	3	SMC3	K429	Q9UQE7
329	KQTARKSTGGK	2	H3	K9	P68432
330	LYICDFHKNFIQSVR	3	SAP30L	K78	Q9HAJ7
331	DSMVHKHGLEF	0	PRMT3	K79	O60678
332	DSMVHKHGLEF	1	PRMT3	K79	O60678
333	DSMVHKHGLEF	2	PRMT3	K79	O60678
334	DSMVHKHGLEF	3	PRMT3	K79	O60678
335	KQLATKAARKS	0	H3	K23	P68431
336	KQLATKAARKS	1	H3	K23	P68431
337	KQLATKAARKS	2	H3	K23	P68431
338	KQLATKAARKS	3	H3	K23	P68431
339	LLKNKKLKAHQ	0	PLCE1	K1522	Q9P212
340	LLKNKKLKAHQ	1	PLCE1	K1522	Q9P212
341	LLKNKKLKAHQ	2	PLCE1	K1522	Q9P212
342	LLKNKKLKAHQ	3	PLCE1	K1522	Q9P212
343	TDYMNKSDDFT	0	FILIP1L	K195	Q4L180
344	TDYMNKSDDFT	1	FILIP1L	K195	Q4L180
345	TDYMNKSDDFT	2	FILIP1L	K195	Q4L180
346	TDYMNKSDDFT	3	FILIP1L	K195	Q4L180

347	QKFLEKPYKHK	0	ZNF491	K73	Q8N8L2
348	QKFLEKPYKHK	1	ZNF491	K73	Q8N8L2
349	QKFLEKPYKHK	2	ZNF491	K73	Q8N8L2
350	QKFLEKPYKHK	3	ZNF491	K73	Q8N8L2
351	AYELEKRTSPQ	0	NUFIP2	K569	Q7Z417
352	AYELEKRTSPQ	1	NUFIP2	K569	Q7Z417
353	AYELEKRTSPQ	2	NUFIP2	K569	Q7Z417
354	AYELEKRTSPQ	3	NUFIP2	K569	Q7Z417
355	ESNKRKSNFSN	0	CBX5	K91	P45973
356	ESNKRKSNFSN	1	CBX5	K91	P45973
357	ESNKRKSNFSN	2	CBX5	K91	P45973
358	ESNKRKSNFSN	3	CBX5	K91	P45973
359	KQTARKSTGGK	3	H3	K9	P68432
360	CQKPSTSKVILRAVA	1	HILS1	K117	P60008
361	AARGIKRKMMQ	0	ZNF326	K238	Q5BKZ1
362	AARGIKRKMMQ	1	ZNF326	K238	Q5BKZ1
363	AARGIKRKMMQ	2	ZNF326	K238	Q5BKZ1
364	AARGIKRKMMQ	3	ZNF326	K238	Q5BKZ1
365	AILKHKKSLQK	0	SH3BP1	K142	Q9Y3L3
366	AILKHKKSLQK	1	SH3BP1	K142	Q9Y3L3
367	AILKHKKSLQK	2	SH3BP1	K142	Q9Y3L3
368	AILKHKKSLQK	3	SH3BP1	K142	Q9Y3L3
369	FESRLKKEIDS	0	LEKR1	K208	Q6ZMV7
370	FESRLKKEIDS	1	LEKR1	K208	Q6ZMV7
371	FESRLKKEIDS	2	LEKR1	K208	Q6ZMV7
372	FESRLKKEIDS	3	LEKR1	K208	Q6ZMV7
373	DISENKRAVRR	0	HSPA8	K257	P11142
374	DISENKRAVRR	1	HSPA8	K257	P11142
375	DISENKRAVRR	2	HSPA8	K257	P11142

376	DISENKRAVRR	3	HSPA8	K257	P11142
377	EEIDEKIGFRN	0	MSANTD4	K114	Q8NCY6
378	EEIDEKIGFRN	1	MSANTD4	K114	Q8NCY6
379	EEIDEKIGFRN	2	MSANTD4	K114	Q8NCY6
380	EEIDEKIGFRN	3	MSANTD4	K114	Q8NCY6
381	LLSGEKRPESE	0	URB1	K89	O60287
382	LLSGEKRPESE	1	URB1	K89	O60287
383	LLSGEKRPESE	2	URB1	K89	O60287
384	LLSGEKRPESE	3	URB1	K89	O60287
385	FKSIMKKSPFS	0	RELA	K314	Q04206
386	FKSIMKKSPFS	1	RELA	K314	Q04206
387	FKSIMKKSPFS	2	RELA	K314	Q04206
388	FKSIMKKSPFS	3	RELA	K314	Q04206
389	TKAARKSAPAT	0	H3	K27	P68431
390	IRHGKFQKMTLKLIL	1	MRGPRB3	K288	Q91ZC1
391	DKHKDKIQINS	0	ANKRD12	K1036	Q6UB98
392	DKHKDKIQINS	1	ANKRD12	K1036	Q6UB98
393	DKHKDKIQINS	2	ANKRD12	K1036	Q6UB98
394	DKHKDKIQINS	3	ANKRD12	K1036	Q6UB98
395	VTESVKNIVPG	0	NUP153	K49	P49790
396	VTESVKNIVPG	1	NUP153	K49	P49790
397	VTESVKNIVPG	2	NUP153	K49	P49790
398	VTESVKNIVPG	3	NUP153	K49	P49790
399	KQAVRKPLEAV	0	MTA1	K532	Q13330
400	KQAVRKPLEAV	1	MTA1	K532	Q13330
401	KQAVRKPLEAV	2	MTA1	K532	Q13330
402	KQAVRKPLEAV	3	MTA1	K532	Q13330
403	AKALDKRQAHL	0	RPS12	K44	P47840
404	AKALDKRQAHL	1	RPS12	K44	P47840

405	AKALDKRQAHL	2	RPS12	K44	P47840
406	AKALDKRQAHL	3	RPS12	K44	P47840
407	AQAKSKQAILA	0	EDF1	K25	O60869
408	AQAKSKQAILA	1	EDF1	K25	O60869
409	AQAKSKQAILA	2	EDF1	K25	O60869
410	AQAKSKQAILA	3	EDF1	K25	O60869
411	EDADRKYEEVA	0	TPM1	K161	P09493
412	EDADRKYEEVA	1	TPM1	K161	P09493
413	EDADRKYEEVA	2	TPM1	K161	P09493
414	EDADRKYEEVA	3	TPM1	K161	P09493
415	DVIRGKLGEKL	0	NAA25	K219	Q14CX7
416	DVIRGKLGEKL	1	NAA25	K219	Q14CX7
417	DVIRGKLGEKL	2	NAA25	K219	Q14CX7
418	DVIRGKLGEKL	3	NAA25	K219	Q14CX7
419	TKAARKSAPAT	1	H3	K27	P68431
420	LYICDFHKNFIQSVR	1	SAP30L	K78	Q9HAJ7
421	NQQSLKRSANQ	0	MED27	K134	Q6P2C8
422	NQQSLKRSANQ	1	MED27	K134	Q6P2C8
423	NQQSLKRSANQ	2	MED27	K134	Q6P2C8
424	NQQSLKRSANQ	3	MED27	K134	Q6P2C8
425	GETKWKPVGMA	0	DNAJC6	K849	O75061
426	GETKWKPVGMA	1	DNAJC6	K849	O75061
427	GETKWKPVGMA	2	DNAJC6	K849	O75061
428	GETKWKPVGMA	3	DNAJC6	K849	O75061
429	DSVDKKDAVFQ	0	HEPHL1	K258	Q6MZM0
430	DSVDKKDAVFQ	1	HEPHL1	K258	Q6MZM0
431	DSVDKKDAVFQ	2	HEPHL1	K258	Q6MZM0
432	DSVDKKDAVFQ	3	HEPHL1	K258	Q6MZM0
433	ETALAKTPQKR	0	PCDHB5	K7	Q9Y5E4

434	ETALAKTPQKR	1	PCDHB5	K7	Q9Y5E4
435	ETALAKTPQKR	2	PCDHB5	K7	Q9Y5E4
436	ETALAKTPQKR	3	PCDHB5	K7	Q9Y5E4
437	DREVRKIKQGL	0	ASS1	K309	P00966
438	DREVRKIKQGL	1	ASS1	K309	P00966
439	DREVRKIKQGL	2	ASS1	K309	P00966
440	DREVRKIKQGL	3	ASS1	K309	P00966
441	FSANPKELKGT	0	RGS1	K24	Q08116
442	FSANPKELKGT	1	RGS1	K24	Q08116
443	FSANPKELKGT	2	RGS1	K24	Q08116
444	FSANPKELKGT	3	RGS1	K24	Q08116
445	DEELPKRVKSR	0	KDM4A	K1012	O75164
446	DEELPKRVKSR	1	KDM4A	K1012	O75164
447	DEELPKRVKSR	2	KDM4A	K1012	O75164
448	DEELPKRVKSR	3	KDM4A	K1012	O75164
449	TKAARKSAPAT	2	H3	K27	P68431
450	CQKPSTSKVILRAVA	0	HILS1	K117	P60008
451	QAGPLKKDGPM	0	PRR11	K220	Q96HE9
452	QAGPLKKDGPM	1	PRR11	K220	Q96HE9
453	QAGPLKKDGPM	2	PRR11	K220	Q96HE9
454	QAGPLKKDGPM	3	PRR11	K220	Q96HE9
455	NLDDAKGLAES	0	FANCE	K272	Q9HB96
456	NLDDAKGLAES	1	FANCE	K272	Q9HB96
457	NLDDAKGLAES	2	FANCE	K272	Q9HB96
458	NLDDAKGLAES	3	FANCE	K272	Q9HB96
459	IHAPWKAIQKF	0	RNASE13	K80	Q5GAN3
460	IHAPWKAIQKF	1	RNASE13	K80	Q5GAN3
461	IHAPWKAIQKF	2	RNASE13	K80	Q5GAN3
462	IHAPWKAIQKF	3	RNASE13	K80	Q5GAN3

463	DKEKGKHDDGR	0	RNF20	K614	Q5VTR2
464	DKEKGKHDDGR	1	RNF20	K614	Q5VTR2
465	DKEKGKHDDGR	2	RNF20	K614	Q5VTR2
466	DKEKGKHDDGR	3	RNF20	K614	Q5VTR2
467	KIPKDKDGKPK	0	RBM7	K45	Q9Y580
468	KIPKDKDGKPK	1	RBM7	K45	Q9Y580
469	KIPKDKDGKPK	2	RBM7	K45	Q9Y580
470	KIPKDKDGKPK	3	RBM7	K45	Q9Y580
471	KEQVSKMASVR	0	TPR	K1105	P12270
472	KEQVSKMASVR	1	TPR	K1105	P12270
473	KEQVSKMASVR	2	TPR	K1105	P12270
474	KEQVSKMASVR	3	TPR	K1105	P12270
475	AWVLDKLKAER	0	EF1A2	K62	P27592
476	AWVLDKLKAER	1	EF1A2	K62	P27592
477	AWVLDKLKAER	2	EF1A2	K62	P27592
478	AWVLDKLKAER	3	EF1A2	K62	P27592
479	TKAARKSAPAT	3	H3	K27	P68431
480	IRHGKFQKMTLKLIL	0	MRGPRB3	K288	Q91ZC1
481	RRPGEKTYTQR	0	SFPQ	K291	P23246
482	RRPGEKTYTQR	1	SFPQ	K291	P23246
483	RRPGEKTYTQR	2	SFPQ	K291	P23246
484	RRPGEKTYTQR	3	SFPQ	K291	P23246
485	GFIKMKSYPSS	0	AVIL	K319	O75366
486	GFIKMKSYPSS	1	AVIL	K319	O75366
487	GFIKMKSYPSS	2	AVIL	K319	O75366
488	GFIKMKSYPSS	3	AVIL	K319	O75366
489	EFNIRKPNEGA	0	SERBP1	K310	Q8NC51
490	EFNIRKPNEGA	1	SERBP1	K310	Q8NC51
491	EFNIRKPNEGA	2	SERBP1	K310	Q8NC51

492	EFNIRKPNEGA	3	SERBP1	K310	Q8NC51
493	DNLIYKLLKPS	0	B3GALT1	K190	Q9Y5Z6
494	DNLIYKLLKPS	1	B3GALT1	K190	Q9Y5Z6
495	DNLIYKLLKPS	2	B3GALT1	K190	Q9Y5Z6
496	DNLIYKLLKPS	3	B3GALT1	K190	Q9Y5Z6
497	AQDHQKKETVV	0	BAZ1B	K175	Q9UIG0
498	AQDHQKKETVV	1	BAZ1B	K175	Q9UIG0
499	AQDHQKKETVV	2	BAZ1B	K175	Q9UIG0
500	AQDHQKKETVV	3	BAZ1B	K175	Q9UIG0
501	FGIEDKDKQII	0	RAD50	K20	Q92878
502	FGIEDKDKQII	1	RAD50	K20	Q92878
503	FGIEDKDKQII	2	RAD50	K20	Q92878
504	FGIEDKDKQII	3	RAD50	K20	Q92878
505	EAMRLKRANET	0	ZNF821	K331	O75541
506	EAMRLKRANET	1	ZNF821	K331	O75541
507	EAMRLKRANET	2	ZNF821	K331	O75541
508	EAMRLKRANET	3	ZNF821	K331	O75541
509	AKRHRKVLRDN	0	H4	K20	P62805
510	QEIRYRSKLKLIRAK	3	PDGFRA iso3	k403	P16234
511	AKKKSKISASR	0	TNNI3	K40	P19429
512	AKKKSKISASR	1	TNNI3	K40	P19429
513	AKKKSKISASR	2	TNNI3	K40	P19429
514	AKKKSKISASR	3	TNNI3	K40	P19429
515	DKSEDKVIAVY	0	HSPA9	K238	P38646
516	DKSEDKVIAVY	1	HSPA9	K238	P38646
517	DKSEDKVIAVY	2	HSPA9	K238	P38646
518	DKSEDKVIAVY	3	HSPA9	K238	P38646
519	DRQKTKETLQS	0	CD72	K196	P21854
520	DRQKTKETLQS	1	CD72	K196	P21854

521	DRQKTKETLQS	2	CD72	K196	P21854
522	DRQKTKETLQS	3	CD72	K196	P21854
523	IAENRKFKSLM	0	VAV3	K735	Q9UKW4
524	IAENRKFKSLM	1	VAV3	K735	Q9UKW4
525	IAENRKFKSLM	2	VAV3	K735	Q9UKW4
526	IAENRKFKSLM	3	VAV3	K735	Q9UKW4
527	DIAAKKKMKKH	0	UTP20	K2745	O75691
528	DIAAKKKMKKH	1	UTP20	K2745	O75691
529	DIAAKKKMKKH	2	UTP20	K2745	O75691
530	DIAAKKKMKKH	3	UTP20	K2745	O75691
531	GKKQNKKKVEE	0	CBX1	K7	P83916
532	GKKQNKKKVEE	1	CBX1	K7	P83916
533	GKKQNKKKVEE	2	CBX1	K7	P83916
534	GKKQNKKKVEE	3	CBX1	K7	P83916
535	EEAIRKIESER	0	DSP	K1880	P15924
536	EEAIRKIESER	1	DSP	K1880	P15924
537	EEAIRKIESER	2	DSP	K1880	P15924
538	EEAIRKIESER	3	DSP	K1880	P15924
539	AKRHRKVLRDN	1	H4	K20	P62805
540	EGKKWQAKIEGIRNK	3	CEP290	K1919	O15078
541	IIDQDKHALLD	0	TJP1	K701	Q07157
542	IIDQDKHALLD	1	TJP1	K701	Q07157
543	IIDQDKHALLD	2	TJP1	K701	Q07157
544	IIDQDKHALLD	3	TJP1	K701	Q07157
545	ALGKLKHSQDP	0	NOL11	K346	Q5RB52
546	ALGKLKHSQDP	1	NOL11	K346	Q5RB52
547	ALGKLKHSQDP	2	NOL11	K346	Q5RB52
548	ALGKLKHSQDP	3	NOL11	K346	Q5RB52
549	EIATDKLSFPL	0	PIEZO2	K2622	Q9H5I5

550	EIATDKLSFPL	1	PIEZO2	K2622	Q9H5I5
551	EIATDKLSFPL	2	PIEZO2	K2622	Q9H5I5
552	EIATDKLSFPL	3	PIEZO2	K2622	Q9H5I5
553	NKHGIKTVSQI	0	RSL1D1	K124	O76021
554	NKHGIKTVSQI	1	RSL1D1	K124	O76021
555	NKHGIKTVSQI	2	RSL1D1	K124	O76021
556	NKHGIKTVSQI	3	RSL1D1	K124	O76021
557	FSLVGKRAIST	0	COX4I1	K12	P13073
558	FSLVGKRAIST	1	COX4I1	K12	P13073
559	FSLVGKRAIST	2	COX4I1	K12	P13073
560	FSLVGKRAIST	3	COX4I1	K12	P13073
561	AKENQKNISNV	0	NEBL	K712	O76041
562	AKENQKNISNV	1	NEBL	K712	O76041
563	AKENQKNISNV	2	NEBL	K712	O76041
564	AKENQKNISNV	3	NEBL	K712	O76041
565	ASGSFKLNKKA	0	H1	K105	P10412
566	ASGSFKLNKKA	1	H1	K105	P10412
567	ASGSFKLNKKA	2	H1	K105	P10412
568	ASGSFKLNKKA	3	H1	K105	P10412
569	AKRHRKVLRDN	2	H4	K20	P62805
570	DSCPAVSKILERSLK	3	REV3L	k2868	O60673
571	AINNSKSFADI	0	NRAS	K88	P12825
572	AINNSKSFADI	1	NRAS	K88	P12825
573	AINNSKSFADI	2	NRAS	K88	P12825
574	AINNSKSFADI	3	NRAS	K88	P12825
575	EPDSAKNVQLK	0	SAMD15	K73	Q9P1V8
576	EPDSAKNVQLK	1	SAMD15	K73	Q9P1V8
577	EPDSAKNVQLK	2	SAMD15	K73	Q9P1V8
578	EPDSAKNVQLK	3	SAMD15	K73	Q9P1V8

579	DAWVYKRLVED	0	MELK	K631	Q14680
580	DAWVYKRLVED	1	MELK	K631	Q14680
581	DAWVYKRLVED	2	MELK	K631	Q14680
582	DAWVYKRLVED	3	MELK	K631	Q14680
583	AAETEKQVALA	0	FHOD1	K454	Q9Y613
584	AAETEKQVALA	1	FHOD1	K454	Q9Y613
585	AAETEKQVALA	2	FHOD1	K454	Q9Y613
586	AAETEKQVALA	3	FHOD1	K454	Q9Y613
587	KRMVEKSLPSK	0	MAGI3	K1364	Q5TCQ9
588	KRMVEKSLPSK	1	MAGI3	K1364	Q5TCQ9
589	KRMVEKSLPSK	2	MAGI3	K1364	Q5TCQ9
590	KRMVEKSLPSK	3	MAGI3	K1364	Q5TCQ9
591	AGKVTKSAQKA	0	EF1A2	K453	Q05639
592	AGKVTKSAQKA	1	EF1A2	K453	Q05639
593	AGKVTKSAQKA	2	EF1A2	K453	Q05639
594	AGKVTKSAQKA	3	EF1A2	K453	Q05639
595	IGDAAKNQVAM	0	HSPA1L	K608	P34931
596	IGDAAKNQVAM	1	HSPA1L	K608	P34931
597	IGDAAKNQVAM	2	HSPA1L	K608	P34931
598	IGDAAKNQVAM	3	HSPA1L	K608	P34931
599	AKRHRKVLRDN	3	H4	K20	P62805
600	KGRTDFIKGMKKKSR	1	VWA5B1	K311	Q5TIE3

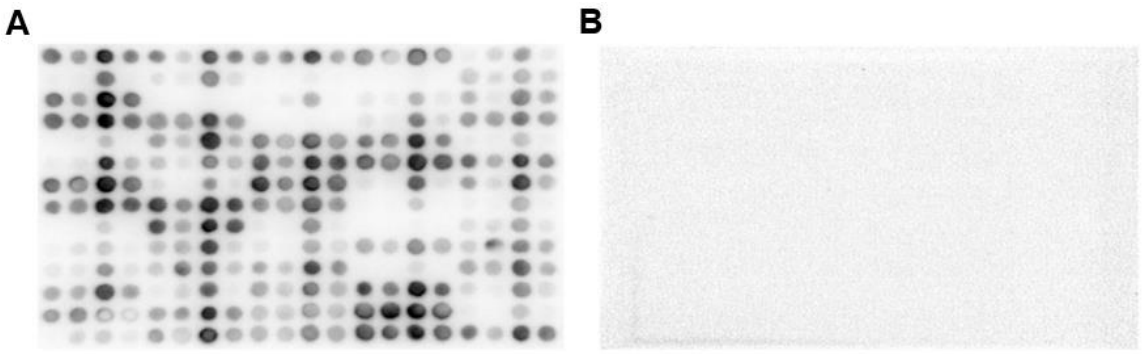

Figure 21

Figure S. 1. Stripped 140-cluster array. A) The original membrane blotted with the 53BP1-tudor domain. B) Stripped membrane blotted with streptavidin HRP to detect residual proteins left on the membrane.

Curriculum Vitae

Name:	Wen Tao Qin
Post-secondary Education and Degrees:	Western University London, Ontario, Canada 2012-2016 BMSc
	Western University London, Ontario, Canada 2016-2019 MSc Canadidate
Honours and Awards:	Biochemistry Undergraduate Student Research Award 2014
	Biochemistry Research Fellowship Award 2016
	Dean's Honour List 2012-2016
	Western Scholar 2012-2016
	Western Graduate Research Scholarship 2016-2018
	Canadian National Proteomics Network – Travel Award 2017
	CIHR – Cancer Research and Technology Transfer (CaRTT) 2017-2018
	CHRI – Epigenetics Trainee Award 2018-2019
Related Work Experience	Teaching Assistant Western University 2016-2018

Publications:

Chen, Z., He, N., Huang, Y., **Qin, W.T.**, Liu, X., Li, L. (2018). Integration of a deep learning classifier with a random forest approach for predicting malonylation sites. Genomics, Proteomics & Bioinformatics 16 (6), 451-459.

Zhang, A., Tessier, T.M., Galpin, K.J.C., King, C.R., Gameiro, S.F., Anderson, W.W., Yousef, A.F., **Qin, W.T.**, Li, S.S., Mymryk, J.S. (2018). The transcriptional repressor BS69 is a conserved target of the E1A proteins from several human adenovirus species. Virus 10(12), 662.

Liu, H., Huang, H., Voss, C., Kaneko, T., **Qin, W.T.**, Sidhu, S., Li, S.S. (2019). Surface loops in a single SH2 domain are capable of encoding the spectrum of specificity of the SH2 family. Molecular Cell Proteomics 18 (2), 372-382.

Wei, R., Liu, X., Voss, C., **Qin, W.T.**, Dagnino, L., Li, L., Vigny, M., Li, S.S. (2019). Numb regulates the endocytosis and activity of the anaplastic lymphoma kinase in an isoform-specific manner. Journal of Molecular Cell Biology, mjz003.