THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

Proteomic Analysis of Histone Mark Crosstalk at Bivalent Domains

Elana Bryan

Doctor of Philosophy
The University of Edinburgh
September 2019

Lay Summary

All cells within an organism - including humans - contain the same DNA which is approximately 2 meters in length. This DNA contains all the information required for an organism to function. A human body has a large number of specialised cells all of which have very specific roles, allowing the development of specialised organs, for example a cardiac cell does a completely different job to that of a liver cell. Each specialised cell uses (expresses) different and specific sections of DNA called genes to allow it to perform its role in the organism. It is the specific genes used by the cell which determine its role and function. How does each cell select the relevant information from DNA in order to do their specific job?

All multicellular organisms develop from a single cell. These cells divide and develop into an embryo; each cell still contains the same DNA yet expresses only a small portion at any one time. These cells, called embryonic stem cells, can become any type of cell in the organism. They become more specialised, a process termed differentiation, as the embryo continues to develop and form different tissues and organs. As the cells become more specialised they lose the ability to become any type of cell and their differentiation potential becomes restricted. The set of genes the cell can express becomes restricted to the differentiation potential of the cell.

DNA is stored within the nucleus of the cell, because of the length of DNA compared to the size of the nucleus, it needs to be compacted to fit. The DNA is compacted by winding it around proteins, like thread on a spool, allowing organised packaging of the DNA. The packing of the DNA helps determine whether that section of DNA is used by the cell, as the cells need access to the DNA in order to express it. Tightly packed DNA is not accessible by the cell and can't be used. Chemical modifications of the DNA and the proteins on which it is stored help mark the DNA, determining whether it is used by the cell. These chemical modifications (marks) bind (recruit) specific factors which help unwind

the DNA and result in the cell accessing the DNA. These modifications and the factors recruited can allow (active) or prevent (repressive) the cell's access to the DNA. It is the combination of chemical modifications and the factors recruited to the DNA that determine whether the DNA is able to be utilised by the cell.

This project focusses on a particular combination of modifications that contain both an active and a repressive mark on the same location of DNA called bivalency or bivalent domains. The function of this combination of marks is unknown, although they share characteristics of both active and repressed DNA. However, they are present in large amounts in embryonic stem cells and signpost genes important in the cell's differentiation. During the organism's development and the cell's differentiation, genes need to be timely regulated to allow correct cellular specialisation. The combination is resolved after the cell has specialised, with either the active or repressive mark being removed. The combination of both marks, in addition to their presence at developmentally significant genes suggests that they are important in allowing quick regulation of the gene during development.

We think that by identifying the factors recruited (bound) by bivalent domains we will be able to determine their function. I found the factors recruited by bivalency, some of which were previously known to interact with either the active or repressive mark. However, some factors are more likely to be recruited to the bivalent state. Removal of one of these factors resulted in the cells being unable to differentiate, potentially due to its role at bivalent domains. I also demonstrated that the symmetry of the chemical modifications helps determine the factors recruited. Further work needs to be done to investigate the other factors identified at bivalent domains and isolate their function.

Abstract

The combination and interaction of histone marks and DNA-associated proteins are critical in the regulation of gene transcription. Individual histone marks have been associated with different gene expression states - histone H3 lysine 4 trimethylation (H3K4me3) is associated with "open" chromatin and active transcription while H3K27me3 is associated with "closed" chromatin and a repressive transcriptional state. In certain cases, these marks have been shown to co-localise at genomic loci, and on the same nucleosome. The co-localisation of active H3K4me3 and repressive H3K27me3 marks at CpG promoters is a hallmark of bivalent domains.

Bivalent domains have been implicated in priming developmental genes for timely activation. However, the complex network of proteins that bind to these domains to regulate and mediate their influence on transcription is unknown. This study has developed tools to enable the characterisation of the protein networks bound to bivalent domains and other specifically modified nucleosomes. *In vitro* synthesised specifically modified nucleosomes were utilised in pulldown assays with embryonic stem cell (ESC) nuclear extract to isolate the specific protein binders for different combinations of histone marks. This assay was validated by comparison of proteins bound to symmetrically modified nucleosomes with previously identified protein binders.

A comparison of symmetrically and asymmetrically modified nucleosomes has elucidated new binding preferences for known proteins. Analysis of proteins bound to asymmetrically modified nucleosomes showed previously unknown binding affinities and conformational preferences. TAF3, a known H3K4me3 mark binder, prefers to bind symmetrically rather than asymmetrically modified nucleosomes, even when the same amount of the mark is present. Therefore, this preference is not solely dependent on the amount of modification, but also due to the conformation of the marks on the nucleosomes. We have identified multiple

key proteins that prefer binding to this specific mark (H3K4me3/K27me3) conformation, including the acetyltransferase KAT6B. Work in mouse ESCs confirmed KAT6B binding to bivalent domains and showed a pronounced differentiation defect in KAT6B$^{-/-}$ cells due to mis-regulation of genes important in development. Further characterisation of these proteins and their interactions will help to clarify bivalent domain function.

iv

Declaration

I declare that this thesis is an original report of my research, has been written by me and has not been submitted for any previous degree. The experimental work is almost entirely my own work; the collaborative contributions have been indicated clearly and acknowledged. Due references have been provided on all supporting literatures and resources.

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

(Elana Bryan, January 2020)

Acknowledgements

A significant number of people have helped me during my PhD, if I were to thank each individually then the acknowledgements would be longer than the entirety of the PhD thesis. In the interests of keeping things brief I shall contain my thanks to the main players and if I fail to thank you specifically, my utmost apologies but be assured that I could not have done it without you.

I would like to thank the Wellcome Trust for their funding and allowing me to complete my PhD in such a supportive research center.

My thanks to my supervisor Philipp Voigt - although times were long and varied, we made it! I couldn't have done it without your supply of kit kats and bad puns. Thank you for all your advice and the many interesting discussions that helped me progress with my PhD. Many thanks to my lab mates for helpful advice and words of encouragement when science decided to misbehave.

Heartfelt thanks go to all the friends that supported me during my thesis and made the journey a much more enjoyable roller coaster than it would otherwise have been. A special thank you to Therese, Christos, Marie, Stefania, Jussi, Lucy, Bethany, Natalia, and Ilaria for their support whenever I needed it. To my Imperial mates for making me keep things in perspective - thank you. To the Smiths, my family away from home. I have never met a more generous, gracious or welcoming group of people. Clint, you are the kindest person I know and I cannot thank you enough.

My deep gratitude and utmost admiration goes to my family, for enabling and encouraging me to do this PhD. The cliché is true, a PhD is a roller coaster and you made the troughs seem less lonely, dark or permanent, while keeping those highs shining brighter. Sincere thanks go to my sister who was always there for a chat or a work buddy to keep me company. To my parents, you were always there for me, via phone calls or packages from home. Each was deeply appreciated and I will never forget. I could not have done it without you.

Abbreviations

PcG	Polycomb Group
PRC2	Polycomb Repressive Complex 2
PRC1	Polycomb Repressive Complex 1
TRX	Trithorax Group
EZH2	Enhancer of Zeste Homolog 2
EZH1	Enhancer of Zeste Homolog 1
SUZ12	Suppressor of Zeste-12 Protein
RbAp46/48	Retinoblastoma protein associated protein 46/48
EED	Embryonic Ectoderm Development
PCL 1/2/3	Polycomb-Like Protein 1/2/3
EPOP	Elongin BC And Polycomb Repressive Complex 2 Associated Protein
JARID2	Jumonji And AT-Rich Interaction Domain Containing 2
AEBP2	AE1 Binding Protein 2
CHD1	Chromodomain-helicase-DNA-binding protein 1
SRCAP	Snf2-Related CREBBP Activator Protein
RNAPII	RNA Polymerase II
PHC	Polyhomeotic Homologs
PCGF	Polycomb Group Ring Fingers
RYBP	RING and YY1 binding protein
YAP2	YY1 Associated Factor 2
MLL	Mixed Lineage Leukemia
SET	Suppressor of variegation 3-9 (Su(var)3-9), Enhancer of zeste (E(z)) and Trithorax (Trx)
LEDGF	Lens Epithelium-Derived Growth Factor
CBX	Chromobox
SAGA	Spt-Ada-Gcn5 acetyltransferase
Cfp1	CXXC-type zinc finger protein 1
TFIID	Transcription Factor II D
G9A/EHMT2	Euchromatic histone-lysine N-methyltransferase 2
GLP	G9a like protein
PHD	Plant Homeodomains
TAF3	TBP Associated Factor 3

ING	Inhibitor of Growth Factor
ORC	Origin Recognition Complex
LRWD1	Leucine-rich repeat and WD repeat-containing protein 1
BRPF1	Bromodomain and PHD finger-containing protein 1
SRSF	Serine-Rich Splicing Factor
USF1	Upstream Transcription factor 1
Hox genes	Homeobox genes
MLAs	Methyl Lysine Analogs
NCL	Native Chemical Ligation
HAT Assay	Histone Acetyltransferase Assay
HATs	Histone Acetyltransferases
HDACs	Histone Deacetylases
HMT	Histone Methyltransferase
DNMTs	DNA Methyltransferases
KMT	Lysine Methyltransferase
DMAP1	DNA Methyltransferase 1 Associated Protein 1
NCP	Nucleosome Core Particle
TF	Transcription Factors
PTMs	Post Translation Modifications
H3K4	Histone H3 Lysine 4
H3K36	Histone H3 Lysine 36
H3K27	Histone H3 Lysine 27
me1,me2,me3	Mono-, di-. tri-methylation
H2AK119ub	Histone H2A Lysine 119 Monoubiquitination
H3K27me3/3	Nucleosome with Histone H3 modified with H3K27me3
H3K4me3/3	Nucleosome with Histone H3 modified with H3K4me3
H3K27me0/3	Nucleosome containing H3'K27me3-Strep and H3"K27me0- His
H3K4me0/3	Nucleosome containing H3'K4me3 - His and H3"K4me0-Strep
H3K4me3/H3K27me3	Nucleosome containing H3'K4me3 and H3"K27me3
H3K4me3/tailless	Nucleosome with H3'K4me3 and a truncated tailless H3
H3K27me3/tailless	Nucleosome with H3'K27me3 and a truncated tailless H3
UnM UnT	Unmodified nucleosome containing no tags
UnM H/S	Unmodified nucleosome with H3'-His and H3"-Step
KO	Knock out
ChIP	Chromatin ImmunoPreciptation
ChIP-Seq	Chromatin Immunoprecipitation-Sequencing
SHP	Sepharose High-Performance

ESCs	Embryonic Stem Cells
LF-MS	Label Free Mass Spectrometry
LFQ	Label Free Quantification
β–me	Beta–Mercaptoethanol
PMSF	Phenylmethanesulfonyl fluoride
DNA	Deoxyribonucleic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ssODN	Single stranded oligo deoxynucleotide
PCR	Polymerase Chain Reaction
Å	Angstrom

0.1 Amino Acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

0.2 Nucleotides

G	Guanine
C	Cytosine
A	Adenine
T	Thymine

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Chromatin Organisation and Structure

1.1.1 Nucleosome Structure

There are approximately 2 meters of DNA in each human cell. This presents several issues, primarily how to condense the DNA in the cell while keeping it accessible. Chromatin packages DNA in such a way that these 1.7-2 metres of DNA fit into a nucleus of approximately 6 µm diameter in eukaryotic cells. The basic repeating unit of chromatin is termed the nucleosome core particle (NCP), which consists of a 146 base pairs (bp) of negatively charged DNA wrapped around the positively charged surface of a highly evolutionary conserved histone octamer in 1.65 turns of a 'flat left-handed superhelix' (Fig. 1.1) (Richmond et al., 1984; Luger et al., 1997a). The histone octamer consists of dimers of four core histone proteins: H3, H4, H2A and H2A. A fifth histone—histone H1—is referred to as a linker histone as it binds DNA between nucleosomes, helping with the ordering and compaction of DNA. These NCPs are linked together with the addition of further proteins including linker histones and further compacted to form a solenoid fibre, which then undergoes higher-order packaging to form a chromosome with the help of additional proteins such as Condensin (Tremethick, 2007; Kalashnikova et al., 2013; Zheng and Hayes, 2003b; Moraru and Schalch, 2019; Hirano, 2016).

In the nucleosome, the core histones form a symmetrical disc-like structure around

which the DNA wraps. The histone octamer is composed of two H2A/H2B dimers and one H3/H4 tetramer and can be split into four histone fold pairs (Fig. 1.1). All four core histones share a very similar motif termed the histone fold, consisting of 3 alpha helices connected by 2 loops, giving rise to a $\alpha1$ -L1-$\alpha2$-L2-$\alpha3$ topology (Fig. 1.1). This motif is integral for the core histone interactions and for ensuring the binding and bending of the DNA around the octamer. The two H3/H4 dimers interact through a four-helix bundle (histone fold pair) formed by the H3-H3' histone folds to assemble the H3/H4 tetramer (Luger et al., 1997a; Richmond et al., 1984). The H2A/H2B dimers interact with the tetramer through a similar four helix bundle between H4 and H2B to form the octamer. Each histone fold pair binds 2.5 turns of DNA double helix, generating a 140° bend and fixing 36-38 base pairs (bp) of DNA in place. Histone fold extensions bind a further 13 bp of DNA (Luger et al., 1997a).

The DNA binds the octamer with a central bp at the particle two-fold pseudo-symmetrical axis (Flaus et al., 1996; Luger et al., 1997a). The helices and loops of the histone fold create two DNA binding sites: one composed of the two alpha helices and the other of L1L2 and the termini of the $\alpha2$ helices at the end of the histone pair crescent. Hydrogen bonds between the main chain amide nitrogen atoms and the phosphates of the DNA backbone help secure the DNA backbone in place (Luger et al., 1997a; Moudrianakis and Arents, 1993). An arginine side chain from the histone is inserted into the minor groove of the wrapped DNA when the backbone faces the octamer, explaining why an A-T base pair is preferred in the minor groove of the nucleosomal DNA, as this creates a narrower minor groove allowing for stronger salt bridges with the arginine side chains. There are multiple interactions between the DNA and histone octamer, from non-polar interactions with deoxyribose groups, to salt bridges that help the histone octamer bind and wrap the DNA (Davey et al., 2002; Luger et al., 1997a; Richmond and Davey, 2003; Richmond et al., 1984). This wrapping of DNA around the octamer causes the phosphate backbones of the two DNA strands to be closer together on the side of the DNA closest to the histone octamer, narrowing the minor and major grooves but widening the grooves on the outside of the wrapped DNA. This bending of the DNA is energetically unfavourable but is maintained by multiple interactions between the DNA and the arginine and lysine sidechains of the histones (Luger et al., 2012; Kalashnikova et al., 2013; Zheng and Hayes, 2003b; Zheng and Hayes, 2003a; Luger et al., 1997a; Moudrianakis and Arents, 1993; Flaus et al., 1996).
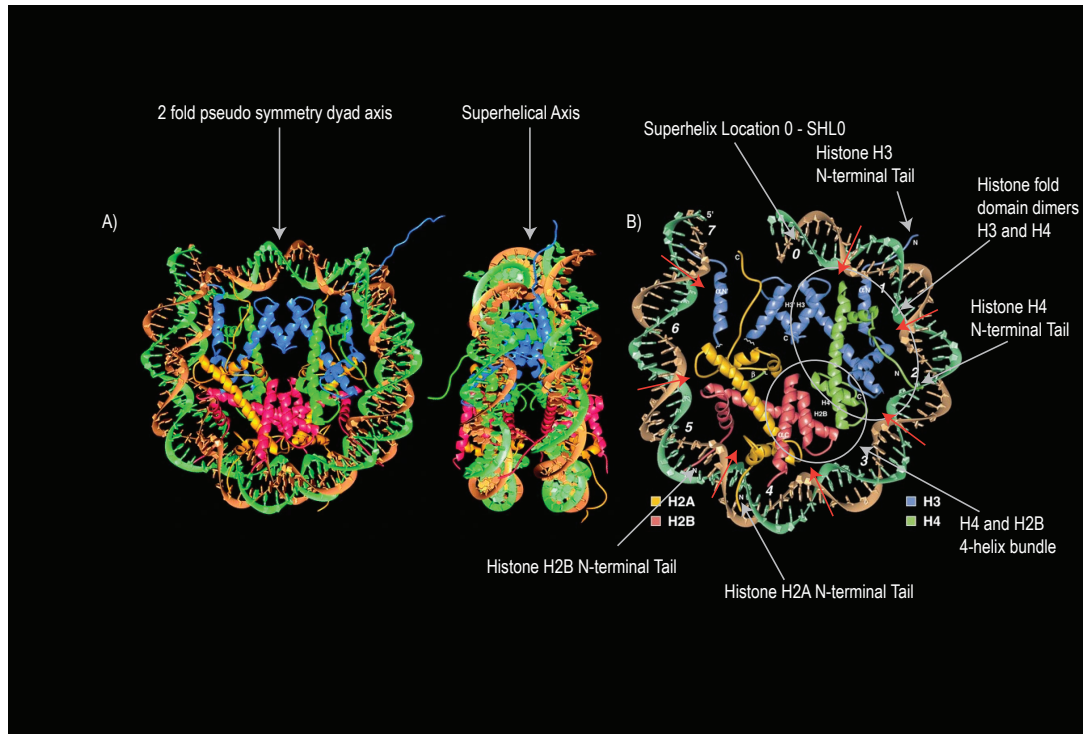
Figure 1.1 *The Crystal Structure of the Nucleosome Core Particle (NCP). The crystal structure of the NCP is composed of bacterially expressed Xenopus laevis core histone proteins and human alpha satellite DNA at 2.7Åresolution (Luger et al., 1997a). The 2-fold pseudo symmetry dyad axis and superhelical axis are indicated. A) NCP crystal structure: H2A is yellow, H2B is red, H4 is green and H3 is blue. The 146 bp of DNA wrapped around the octameric core is shown in green and gold to clearly delineate the different strands of the DNA. Two different viewpoints of the NCP are shown: down the superhelix axis and perpendicular to it. B) A blown-up view of the NCP particle. Half of the NCP is shown in order to be able to clearly see interactions between the proteins of the histone core. The histone fold domain dimer between H3 and H4 is indicated. The 4-helix bundles are labelled as H3' H3 and H2B H4; histone-fold extensions of H3 and H2B are labelled as αN', αN and αC, respectively. The 4-helix bundle between H4 and H2B is highlighted. The N terminus and C terminus of each protein is labelled N and C, respectively. The number of superhelix turns is shown indicating the rotational orientation of the DNA relative to the central base pair, with superhelix location zero (SHL0) defined as where the major groove faces the octamer and indicated on the figure, each further SHL label indicates one more double helix turn away from SHL0. Figure adapted from Luger et al., 1997a. Red arrows indicate sites of interaction between the minor groove of the DNA and the arginine side chain from the histone core.*

Approximately 25% of each core histone is composed of its N-terminal (NT) tail sequence. These N-terminal tails are flexible, resulting in poor definition within crystal or electron-microscopy (EM) structures. The flexible tail regions can be defined as the parts of the histone proteins that extend past or through the DNA superhelix and are not an integral part of the histone octamer. The core histones can still form an octameric complex when lacking their N-terminal flexible tails (Biswas et al., 2011; Freeman et al., 1996). Both H3 and H2B N-terminal tails have random coil regions that pass through superhelix channels formed from the minor grooves of the DNA, spanning residues 39-43 and 26-34, respectively (Luger et al., 1997a). Not all of the tails protrude out via a minor groove and do not contact the octamer. The H2A and H4 N-terminal tails follow minor grooves to the exterior of the nucleosome, while the C-terminal tail of H2A along with its $\alpha 3$ helix forms a docking domain termed the acidic patch for the H4 C-terminal tail which is folded over its $\alpha 3$ helix (Kalashnikova et al., 2013; Luger et al., 1997a). The core histone N-Terminal tails are sites of post-translation modifications (PTMs), which are known to alter the accessibility of the DNA bound to the nucleosome, nucleosome dynamics and to recruit binding proteins and will be discussed later (Stützer et al., 2016; Luger, 2006; Li et al., 2005b; Fletcher and Hansen, 1995; Zentner and Henikoff, 2013).

Although the histone octamer is normally made up of the 4 main canonical core histones H2A, H2B, H3, and H4, there are many non-canonical variants, which can be substituted for one or more of the canonical histones (Weber and Henikoff, 2014; Talbert and Henikoff, 2017). Each variant has a slightly different sequence, recruiting different proteins, potentially altering the nucleosome structure and affecting its biophysical properties (Chakravarthy et al., 2005; Suto et al., 2000; Abbott et al., 2001; Park et al., 2004). The majority of the non-canonical histone variants are replacements for canonical H3 and H2A (Weber and Henikoff, 2014; Talbert and Henikoff, 2010). These variants vary in sequence from only a four amino acid ($\alpha\alpha$) difference between H3.1 and H3.3 to a 50% change in sequence between H3.1 and H3 CENPA (a centromere specific H3 variant) (Sharma et al., 2019).

Packaging the DNA into the nucleosome allows for efficient compaction and storage of the DNA; however, one major consequence of the wrapping of DNA around the histone octamer is the occlusion of the DNA (Luger et al., 1997a; Luger et al., 1997b; Lutter, 1979). There are multiple mechanisms that regulate access to DNA including the nucleosome and modifications to it. The nucleosome

4

is dynamic, with very few interactions between the DNA and the histones needing to be broken in order for the nucleosome to partially unwrap (Li et al., 2005b; Polach and Widom, 1995; Li and Widom, 2004). The DNA sequence incorporated into the nucleosome can alter its stability (Hartley and Madhani, 2009; Lowary and Widom, 1998; Gottesfeld and Luger, 2001).

As a consequence of the DNA having to be held in an energetically unfavourable state the sequence of the DNA affects the positioning, and likelihood of nucleosome formation (Meersseman et al., 1992; Widom, 2006; Gottesfeld and Luger, 2001; Becker, 2002; Lowary and Widom, 1998; Widom, 2001; Yang and Ullah, 2007; Flaus, 2011). This is demonstrated by CpG-rich, GC-rich unmethylated DNA sequences, usually covering promoters, that destabilises nucleosomes (Flaus, 2011). The position of the nucleosome is not fixed and the histone octamer is able to slide along the DNA strand, exposing different DNA sequences as it moves position, which can be facilitated by chromatin remodellers *in vivo* (Meersseman et al., 1992; Hamiche et al., 1999; Park et al., 2004; Park et al., 2005; Becker, 2002; Pennings et al., 1991; Tyagi et al., 2016). The placement of the histone variants into the histone octamer can also alter the dynamicity of the nucleosome, affecting the accessibility of the DNA (Fan et al., 2004; Zhang et al., 2005). The histones' N-terminal tails including their modifications also modulate access to the DNA (Biswas et al., 2011; Lee et al., 1993; Polach et al., 2000; Yang et al., 2005). The modifications present on the tails (Allan et al., 1982; Fletcher and Hansen, 1995; Zheng and Hayes, 2003b; Pepenella et al., 2014) and in the histone core (Fenley et al., 2018) and the presence of histone variants can also prevent the formation of higher order chromatin structures (Luger et al., 2012; Shogren-Knaak, 2006; Tremethick, 2007; Kebede et al., 2015; Abbott et al., 2001) or affect the stability of the nucleosome itself (Fan et al., 2004; Jin and Felsenfeld, 2007).

1.2 Chromatin Readers, Writers and Modifiers - The Histone Code Hypothesis

Multicellular organisms require cellular specialisation and therefore differential gene expression that arises during development and cellular differentiation. As every cell in the human body contains the same DNA, the distinct expression patterns between, for example, a pancreatic cell and a cardiac muscle cell are

controlled through epigenetics. Although the genetic information is contained purely within the DNA, gene expression is controlled not only by regulatory factors such as histones or DNA-binding transcription factors (TFs) but also by a wide array of posttranslational modifications (PTMs) on the DNA and on the histones the DNA is wrapped around. The DNA sequence can have chemical groups added, affecting its transcription state without affecting the information contained. The addition of methyl groups to DNA, by DNA methyltransferases (DNMTs), especially to the CpG-rich promoters of genes effectively silences the transcription of the modified DNA sequence (Curradi et al., 2002; Bird, 2002; Jones, 2012; Klose and Bird, 2006). The addition of methyl groups to DNA occurs predominately at CpG dinucleotides (in mammals) and the majority of CpGs are methylated (Jang et al., 2017; Siegfried and Cedar, 1997). CpG-rich regions that make up regulatory regions including promoters of genes are protected from the addition of methylation. The methylated DNA can recruit methyl-binding proteins that protect the inactive DNA state (Bird, 2002; Jones, 2012; Jang et al., 2017). Histones are also subject to a vast array of modifications, including – but not restricted to – acetylation, phosphorylation, sumoylation and varying degrees of methylation (see Fig. 1.2 (Vaquero et al., 2003; Bannister and Kouzarides, 2011; Campos and Reinberg, 2009; Kouzarides, 2007). These marks are mainly located on the N-terminal tails, which protrude outwards from the nucleosome core. Protein recruitment, DNA accessibility and the histone tail's movement are affected by the placement of PTMS.

The placement of these modifications can affect nucleosome dynamics and the formation of higher order chromatin structures (see Section 1.1.1). Acetylation of H4 lysine 16 (H4K16ac) not only inhibits the formation of more compact structures (30-nanometer fibres) but also prevents interactions between fibers (Shogren-Knaak, 2006). The acetylation of the lysine residue neutralises the positive charge of the lysine, reducing the charge-dependent interactions with the negatively charged DNA phosphate backbone, thereby increasing the accessibility of the DNA to surrounding proteins (Zentner and Henikoff, 2013). Many acetylated lysines recruit proteins that then translate their modification into a biological function. Modifications not only to the histone tail but also the histone core can have significant effects: phosphorylation of residues on or close to the histone core near to the nucleosome DNA entry/exit point, such as H3Y41ph, have been shown to increase the accessibility of the DNA (Brehove et al., 2015; Lawrence et al., 2016). This phosphorylation can be combined with acetylation within the same region such as H3K56ac, leading to an synergistic increase in
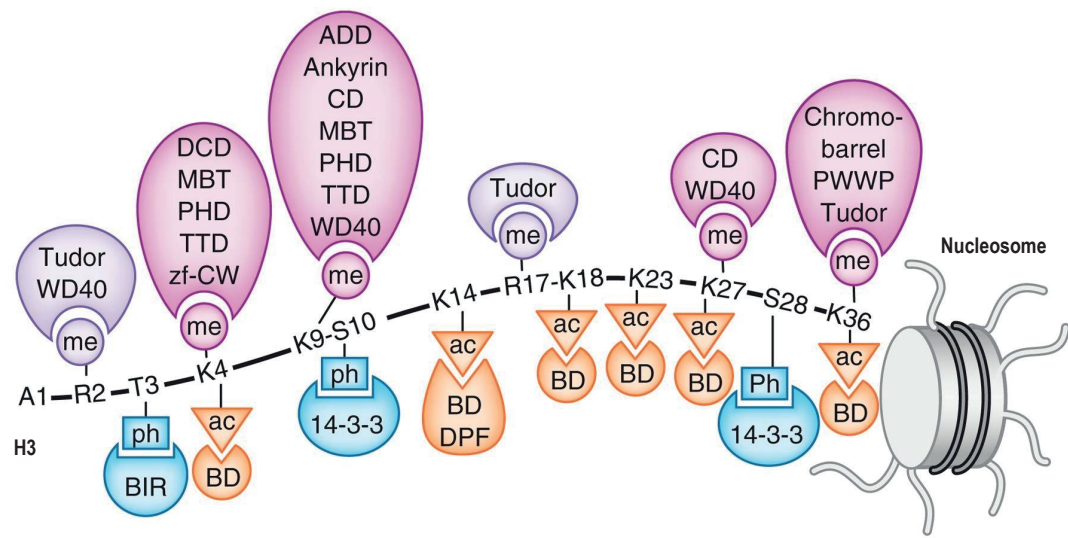
Figure 1.2 *Histone Post–Translational Modifications and the Proteins That Bind Them. Schematic of protein reader domains binding the PTMs on the residues of the histone H3 N-terminal tail. Figure taken from Musselman et al., 2012. Recognition of methylated (me) lysines, methylated (me) arginines, acetylated (ac) lysines, phosphorylated (ph) serines and theorines. BD - bromodomain, CD - chromodomain, DPF - double PHD finger, DBD - double bromodomain, MBT - Malignant brain tumour, TTD - tandem tudor domain, PHD - plant homeodomain, BAH - bromo-adjacent homology.*

accessibility (Brehove et al., 2015).

It has been posited that the combination of the PTMs on the histones' N-terminal tails, in addition to the histone variants included in the nucleosome, recruit proteins that bind these modifications and interpret them – providing a "histone code" for proteins to read (Jenuwein and Allis, 2001). These modifications can lead to the recruitment of proteins, resulting in alteration of chromatin structures, inheritable transcriptional states, or higher order structures (see Fig. 1.3). The PTMs are placed by chromatin "writers", removed by chromatin "erasers" and functions are regulated by chromatin "readers".

Chromatin "readers" are proteins that contain protein domains that recognise a particular PTM on a histone or a combination of marks and histone variants, and direct a specific biological outcome. They can be further characterised by the marks they bind. Both chromatin "writers" and "erasers" are enzymes that are classified based on the specific modification they place or remove, respectively. Histone methyltransferases (HMTs) place methyl groups while histone acetyltransferases (HATs) catalyse the formation of acetylation. These modifications can be placed on multiple residues on different histones, the specificity of placement is determined by the catalytic site of the chromatin "writer". This results in the overall "write" classification, for example HMTs, of

7
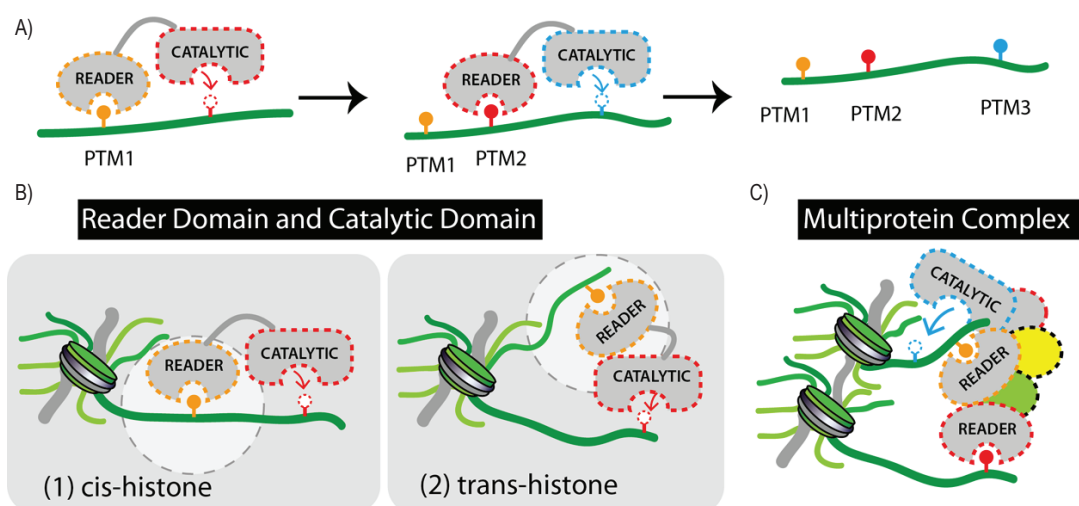
Figure 1.3 *Combinatorial Histone Modifications* - *Schematics of proteins binding one or more histone PTMs. Figure from Su and Denu, 2016. Histone tail in green, protein reader domain in orange and the catalytic domain in red. A) Schematic of the progressive nature of histone modifications. B) Binding of a chromatin modifier and placement of a subsequent mark in cis or trans C) Potential interplay between proteins of a multi-subunit complex. The recruitment of the complex by one subunit via a specific PTM, in order to modify a residue on a neighbouring nucleosome via a different catalytic subunit.*

enzymes being further divided by the residues they can modify. Both SETD1A, MLL2 and Polycomb Repressive Complex 2 (PRC2) are all HMTs but SETD1A and MLL2 place H3K4me3 whereas PRC2 places H3K27me3. Although the modification may be the same, the position of the residue modified and the chromatin "reader" it recruits determines its function. Chromatin "erasers" are divided similarly with enzymes that remove methylation or acetylation referred to as histone demethylases (HDMTs) and deacetylases (HDACs), respectively.

Numerous enzymes catalyse the placement of each type of possible histone modification (chromatin "writers"). These enzymes can be clustered into larger groups which are clearly evolutionarily related, have conserved mechanisms of function, and structure - protein families. Acetylation is placed by enzymes belonging to three protein families GNAT, MYST, and CBP/p300 while methylation is placed by histone methyltransferases (HMTs) that can be divided by those that modify lysines (HKMTs) and those that modify arginines (HRMTs) (Sterner and Berger, 2000; Blanc and Richard, 2017; Hyun et al., 2017). All HKMTs with the exception of DOT1 contain a SET (suppressor of variegation, enhancer of zeste, and trithorax) domain and belong to the large SET family of methyltransferases (see Fig. 1.4)(Schubert et al., 2003; Cheng et al., 2005). All chromatin "writers" catalyse the exchange of chemical groups from a co-substrate, for example HKMTs utilise SAM (S-Adenosyl Methionine) as a co-substrate, to catalyse the transfer

A)

Complex	Subunit
MLL and SET1	ASH2L
	RBBP5
	WDR5
	DPY30
SET1	WDR82
	CFP1
	HCF1
MLL	MLL1/2/3/4
SET1	SET1A/B
MLL1/2	Menin
MLL3/4	KDM6A
	PTIP
	PA1
	NCOA6

B)

Complex	Complex Makeup	Subunit
PRC1	Core	RING1A/B
	Canonical	CBX 1/2/3/4/5/6/7
	Canonical	PHC1/2/3
	Non-canonical	PCGF1/2/3/4/5/6
	Non-canonical	RYBP/YAF2
PRC2	Core	EZH1/2
		EED
		SUZ12
		RBAP46/48
	Accessory	JARID2
		AEBP2
		PCL1/2/3
		EPOP
		PALI

Figure 1.4 *Subunits of Trx and PRC Complexes* *Subunits of MLL, SET1 and PRC complexes detailed. A) Figure detailing the subunits of the MLL and SET1 complexes. ASH2L, DPY30, WDR5 and RBBP5 are shared between the different MLL and SET1 complexes. The catalytic subunit determines the specific complex. SET1 also contains the subunits WDR82, CFP1 and HCF1. B) Figure detailing PRC complex subunits. There are multiple variants of the PRC1 and 2 complexes. The core subunits for PRC2 are showed and the accessory subunits labelled. Canonical or non-canonical PRC1 complex subunits are shown.*

of a methyl group from SAM to the histone substrate resulting in a methylated histone substrate.

Each PTM placed by chromatin "writers" can be bound by a specific protein "reader" domain contained in a chromatin "reader". This allows the recruitment of specific readers to particular histone marks and chromatin locations. Each histone modification can be bound by one or more specific protein domains designated reader domains: methylated lysines can be bound by plant homeodomains (PHDs) (Ali et al., 2014; Champagne and Kutateladze, 2009; Sanchez and Zhou, 2011) or chromodomains depending on the residue modified (Flanagan et al., 2005; Yap and Zhou, 2011; Fischle et al., 2003; Pearce et al., 1992), while acetylated residues are bound by bromodomains (see Fig. 1.2) (Fujisawa and Filippakopoulos, 2017; Tamkun et al., 1992; Meslamani et al., 2016). H3K4me3 modification is bound by the PHD domains of multiple proteins including both TBP associated factor 3 (TAF3) (Ingen et al., 2008; Vermeulen et al., 2007) and inhibitor of growth 2 (ING2) (Peña et al., 2006), while H3K27me3 is bound by the chromodomain contained in the chromobox (CBX) proteins (Pearce et al., 1992; Morey et al., 2012; Min et al., 2003). Acetylated H3 and H4 act to recruit BRD4 through its bromodomain (Dey et al., 2003). Even the lack of a modification on a residue can recruit specific reader domains for example the PHD finger in BRPF1

is recruited to H3K4me0 (Qin et al., 2011).

The divisions between "writers", "readers" and "erasers" can become unclear. Chromatin modifiers (writers) need to be selectively recruited so that they modify only specific sites. Often the enzymes that modify the histone residues are recruited to the nucleosome by a combination of domains including DNA binding domains and histone modification reader domains; therefore, chromatin "writers" usually interact with or are also chromatin "readers". While the "readers" may contain a chromatin modifying function of their own, either catalysing a placement of a mark or erasing one (Farcas et al., 2012; Tachibana et al., 2001; Collins et al., 2008; Margueron et al., 2009). The SAGA complex contains both chromatin "readers", "writers" and "erasers" and therefore can be recruited to multiple histone PTMs and perform multiple functions. SAGA can be recruited by H4K16ac, as it contains multiple bromodomains, and then acetylates histone H3. SAGA also contains a deubiquitylation (DeUb) module which is able to remove ubiquitin from H2B, therefore the SAGA complex functions as a chromatin "reader", "writer" and "eraser" (Morgan et al., 2016; Koutelou et al., 2010; Owen, 2000)

Certain marks are primarily associated with specific chromatin states because they have been detected at these locations including active or repressed promoters, enhancers, or exons by Chromatin Immunoprecipitation-Sequencing (ChIPseq). Further investigation showed that these marks are recognised by particular reader domains, and analysis of the function of these chromatin "readers" enabled explanation of how their presence at those chromatin states might bring about functional consequences of those states.

Chromatin can be split into two major categories based on how condensed or "open" the chromatin is. Euchromatin is defined as "open" or transcriptionally active chromatin that usually contains highly transcribed genes and is associated with H3K4me1/2/3, H3K36me3, H3K14ac, H3K9ac, H3K27ac, and H3K79me3 (Bannister and Kouzarides, 2011; Kouzarides, 2007). Heterochromatin is generally defined as more compact chromatin and generally repressive; however, there are two accepted classes: facultative and constitutive heterochromatin. Facultative heterochromatin consists of genes that are differentially expressed during development and is usually enriched for H3K27me3 and Polycomb Repressive Complexes 1 and 2 (PRC1/2) (Trojer and Reinberg, 2007; Wutz, 2011). Constitutive heterochromatin contains permanently silenced genes and also includes pericentromeres and telomeres (Bannister and Kouzarides, 2011;

10

Kouzarides, 2007; Saksouk et al., 2015). Although the correlation of histone marks with transcription has been thoroughly researched, the functions of the majority of histone marks are not completely understood and the placement of PTMs does not always correlate with their known role.

The recruitment of a chromatin "reader" to a PTM or histone variant can lead to direct alterations to chromatin and potential alteration of the chromatin state. The modification of one histone residue could lead to a cascade in protein recruitment and subsequent alteration of the chromatin environment, altering the transcriptional state of the associated DNA. Unmethylated CpG-rich DNA sequences (CpG islands) can lead to the recruitment of H3K4 methyltransferases such as SET1 complexes via the Cfp1 – CXXC-type zinc finger protein 1 – subunit binding the DNA with its CXXC domain (Clouaire et al., 2012; Brown et al., 2017). The placement of H3K4me3 can recruit multiple proteins including ING2 (Peña et al., 2006; Shi et al., 2006) and TAF3 (Vermeulen et al., 2007; Ingen et al., 2008). H3K4me3 is bound by a variety of factors mostly involved in activating transcription including: the TAF3 subunit of the general transcription factor TFIID, and JMJD2A (KDM4A) (Pedersen et al., 2014; Ingen et al., 2008; Lauberth et al., 2013). The binding of TAF3 recruits the TFIID complex, helping form the RNAPII transcription pre-initiation complex (PIC) (Nuland et al., 2013; Lauberth et al., 2013; Vermeulen et al., 2007). The binding of an ING family protein via its PHD domain such as ING2, ING5 or ING3, can lead to a variety of effects on the local chromatin due to their incorporation into multiple complexes. ING5's or ING2's recruitment by H3K4me3 can recruit histone acetyltransferases (HAT), such as the MORF, NuA4 and HBO1 complex, or histone deacetylases (HDACs) such as SIN3A (Peña et al., 2006; Shi et al., 2006; Champagne et al., 2008; Ullah et al., 2008; Doyon et al., 2006).

Multiple residues can be modified simultaneously with different chemical groups, thereby adding complexity to the histone code. In the simplest case of histone combinatorial marks with only two PTMs, these could be present on the same histone tail, in the same nucleosome, or on neighbouring nucleosomes. With the number of histone residues available to be modified and the variety of possible different modifications the amount of combinatorial possibilities is large. The placement of one histone modification can affect the subsequent modifications placed on the nucleosome. In the simplest terms, the chemical modification of a histone residue will affect the chemical environment around it, potentially affecting the placement of other marks. The recruitment of a protein to a histone

PTM could decrease the likelihood of other modifications simply due to steric hindrance as a result of the protein being bound.

There is evidence that this cross talk between marks and their placement is more widespread and developed than purely alteration of the environment immediately surrounding the modification. The active site of multiple different chromatin modifiers binds to more than just the target residue, causing modifications to nearby residues to prevent the modifier binding or conversely increase the enzyme's catalytic activity, this effect can be mediated by the interaction of other subunits with the catalytic subunit (Voigt et al., 2012; Schmitges et al., 2011; Liokatis et al., 2012). PRC2 is inhibited from placing H3K27me3 on the same H3 N-terminal tail that carries H3K4me3 (Schmitges et al., 2011; Voigt et al., 2012), but its catalytic activity is increased when bound to H3K27me3 (Margueron et al., 2009). G9A/GLP is unable to place H3K9me3 if residues H3S10 or H3T11 are phosphorylated (ph) or H3R8 methylated (Rathert et al., 2008; Chin et al., 2005; Collins et al., 2008). This cross talk may extend to modifications present on different histone tails in the same nucleosome or even on neighbouring nucleosomes. Cross talk between residues that are not situated on the same histone tail may require the presence of proteins to form a bridge between these modifications to mediate the cross talk. PRC2 can bind to H3 tails from neighbouring nucleosomes, H3K27me3 binds PRC2 via one subunit and the catalytic subunit places H3K27me3 on the unmodified tail of a neighbouring nucleosome (Poepsel et al., 2018).

Chromatin modifiers can contain multiple and different reader domains for example chromatin modifiers: UHRF1 and BPTF contain a combination of tandem tudor domains, PHD domains and bromodomains. Not only can they be recruited by multiple histone PTMs, ranging from H3K4me3, to H3K9me2/3, to H3K4me0 to H4ac (Ruthenburg et al., 2011; Rothbart et al., 2013; Taverna et al., 2007) but they may also require a combination of these marks to be present in order to bind. The combination of histone marks can recruit different proteins, potentially amplifying the effect desired or allowing one histone PTM to contribute to multiple functions. As these marks can recruit multiple proteins, it may be the combination of multiple histone marks that help select a specific protein. Not all of these protein domains need to be bound in order for the whole protein or protein complex to be recruited, however, it could be argued that the more protein domains bound the stronger the recruitment of the complex to chromatin.

12

1.2.1 Polycomb and Trithorax Group Proteins

Polycomb group (PcG) and Trithorax group (Trx) proteins are mainly known for their mutually antagonistic roles in regulating gene expression. These two groups of proteins were initially discovered in *Drosophila melanogaster* but have similar composition in mammals, although with more diverse and a greater number of subunits (Slifer, 1942; Lewis, 1978; Kassis et al., 2017; Schuettengruber et al., 2017; Grossniklaus and Paro, 2014; Bajusz et al., 2018; Poynter and Kadoch, 2016; Geisler and Paro, 2015). They form a variety of complexes that act to either repress or activate transcription (Aloia et al., 2013; Geisler and Paro, 2015; Schuettengruber et al., 2017). PcG genes were first discovered in *Drosophila melanogaster* as regulators of Hox gene expression (Lewis, 1978). PcG proteins form a wide variety of complexes that are subdivided into 2 main classes: Polycomb Group Complex 1 (PRC1) (see Fig. 1.5 and Figure 1.4) and Polycomb Group Complex 2 (PRC2) (see Fig. 1.6 and Figure 1.4).

1.2.2 The Composition of Polycomb Repressive Complex 2 (PRC2) and its Accessory Factors

PRC2 mediates its function by the di- and trimethylation of histone H3 lysine 27 (H3K27me2/3), a hallmark of repressive chromatin (Margueron and Reinberg, 2011; Ferrari et al., 2014; Pasini et al., 2007; Pasini et al., 2010a). H3K27me3 may mediate chromatin repression by recruitment of PRC1 and compaction of chromatin or by prevention of the placement of active marks (Kundu et al., 2017; Francis et al., 2004; Grau et al., 2011; Lau et al., 2017). PRC2's major components are EZH2, SUZ12, EED, and RbAp46/48 (Højfeldt et al., 2019). Deletion of any of the core subunits is embryonically lethal, and results in ESCs with differentiation defects (Pasini et al., 2007; Pasini et al., 2004; Shen et al., 2008; Chamberlain et al., 2008; Faust et al., 1995; O'Carroll et al., 2001; Collinson et al., 2016). EZH2, EED, and SUZ12 are required for the stability and activity of PRC2 (Fig. 1.6) (Pasini et al., 2007; Pasini et al., 2004; Cao and Zhang, 2004a; Obier et al., 2015; Montgomery et al., 2005; Margueron et al., 2009; Wang et al., 2002; Faust et al., 1995; O'Carroll et al., 2001; Ketel et al., 2005).

EZH2 and its homologue EZH1 contain the catalytic methyltransferase SET domain and are mutually exclusive in the PRC2 complex (Wu et al., 2013; Antonysamy et al., 2013; Lewis, 1978; Cao and Zhang, 2004b; Margueron et al.,
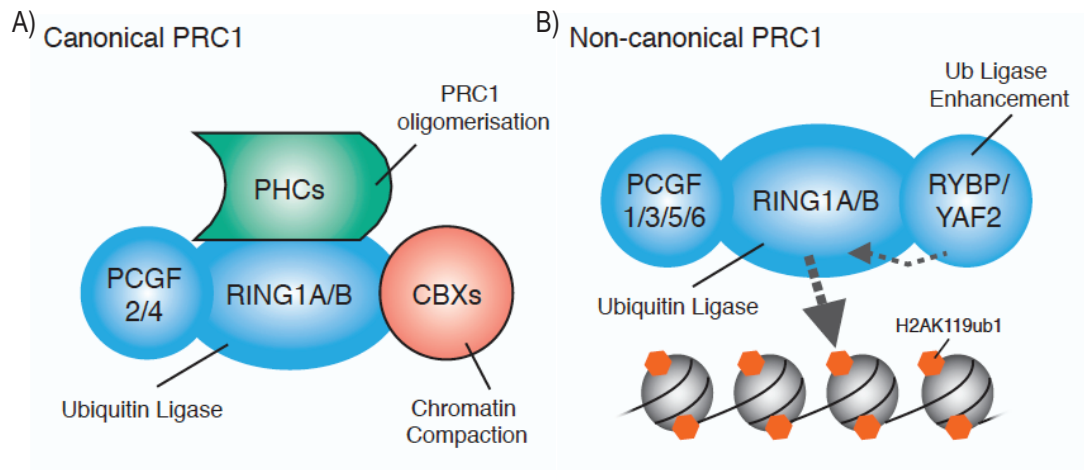
Figure 1.5 *Schematic of the Possible PRC1 Complex Variants.* *PRC1 complexes consist of the RING1A/B subunits with varying types of PCGF proteins. The additional subunits including CBX and PHC variants, or RYBP/YAF2 determine whether the complex is considered canonical or non-canonical. A) Canonical PRC1 variants shown. Subunits involved in chromatin compaction and oligermisation of PRC1 show in red and green respectively. B) Non-canonical PRC1 variants shown. All subunits labelled. The function of each subunit is detailed. Although canonical PRC1 complex has slight ubiquitylation activity, non-canonical PRC1 places the majority of H2AK119ub. Figure taken from Illingworth, 2019.*
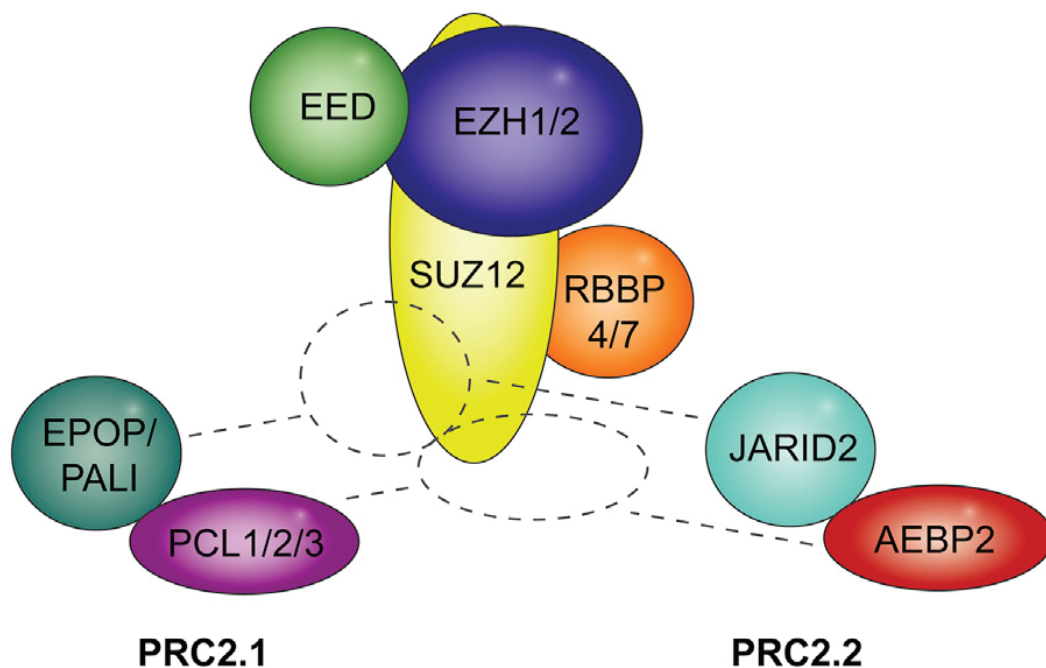
Figure 1.6 *Schematic Drawing of the Core PRC2 complex and Potential sites of Accessory Subunit Binding.* *The core PRC2 complex shown in orange (RbAp46/48 referred to as RBBP4/7 respectively), yellow (SUZ12), blue (EZH1/2) and green (EED). The subsidiary subunits and their interaction sites with the core complex are indicated. JARID2 (turquoise), and EPOP/PALI (dark green) both interact with SUZ12 at similar locations explaining their exclusivity. AEBP2 (red) and PCL1/2/3 (purple) proteins also interact with the SUZ12 subunit, albeit at a different location. Both AEBP2 and PCL1/2/3 complexes interact at a similar site of the SUZ12 protein explaining their incorporation into different variant PRC2 complexes. Figure from Laugesen et al., 2019.*

14

2008; Shen et al., 2008). EZH2 is the predominant H3K27me3 methyltransferase in ESCs, and is specific to proliferating cells while EZH1 is the dominant subunit in terminally differentiated cells (Margueron et al., 2008). These mutually exclusive subunits have overlapping functions and partially compensate for the removal of the other. The global levels of H3K27me3 decrease in EZH2$^{(-/-)}$ cells, however, some genes retain H3K27me3, suggesting compensatory mechanisms potentially relying on EZH1 or unknown maintenance factors (Shen et al., 2008), although there are some distinct nonredundant features of EZH1 and EZH2 (Hidalgo et al., 2012; Mochizuki-Kashio et al., 2012; Shen et al., 2008; Margueron et al., 2008). Deletion of EZH2 is embryonically lethal (O'Carroll et al., 2001), and EZH2$^{(-/-)}$ ESCs demonstrate differentiation defects, albeit viable (Shen et al., 2008; Collinson et al., 2016). Some of the genes that lose H3K27me3 in EZH2$^{(-/-)}$ cells continue to be repressed, potentially due to the absence of their activating transcription factors in ESCs (Margueron et al., 2008; Shen et al., 2008).

SUZ12 is a core component of PRC2 and is essential for stability and activity of the PRC2 complex (Pasini et al., 2007). SUZ12 contains a zinc finger domain and together with RbAp46/48 forms a binding site for the unmodified N-terminal tail of histone H3 (Schmitges et al., 2011). SUZ12 has multiple functions; it is involved in RNA binding, and integral for the formation and activation of PRC2 (Schmitges et al., 2011; Beltran et al., 2016; Jiao and Liu, 2015; Davidovich et al., 2013; Antonysamy et al., 2013). PRC2 is unable to place H3K27me3 on tails previously modified with H3K4me3. This inhibition of catalytic activity is mediated by both SUZ12 and Rbbp4 (RbAp46/48) (Schmitges et al., 2011). The presence of the active histone mark H3K4me3 also inhibits PRC2 binding to the H3 N-terminal tail further impairing PRC2's placement of H3K27me3 (Schmitges et al., 2011; Voigt et al., 2012).

EED interacts with both SUZ12 and EZH2, helping modulate the catalytic activity of PRC2. EED binds H3K27me3, helping recruit PRC2, and stimulating PRC2's catalytic activity, thus setting up a positive feed forward loop (Margueron et al., 2009). This helps spread the H3K27me3 mark, explaining the large domains of H3K27me3 seen and may also help the maintenance of the H3K27me3 mark either by binding it and preventing access to H3K27 demethylases or by replacement of the mark as it is diluted during DNA replication and cell division. RbAp46/48, although not always considered a core subunit, is important for PRC2 nucleosomal recognition and enzymatic inhibition upon binding of H3K4me3 (Nowak et al., 2011; Vizán et al., 2015; Nekrasov et al.,

2005; Murzina et al., 2008; Schmitges et al., 2011). As confirmed by recent structures of catalytically active subcomplexes of PRC2, EZH2's SET domain is in an autoinhibited conformation and requires interactions with both SUZ12 and EED to stabilise the active conformation of the SET domain necessary for methyltransferase activity (Jiao and Liu, 2016; Jiao and Liu, 2015; Cao and Zhang, 2004a; Chen et al., 2018; Wu et al., 2013; Justin et al., 2016). Recent protein structures have shown how H3K27me3 binding by EED affects PRC2's activity or binding (Margueron et al., 2009; Poepsel et al., 2018). These interactions allow conformational changes from one subunit to be communicated to the whole enzyme complex resulting in alterations to PRC2 binding or to its catalytic activity.

PRC2 has multiple subsidiary components that are not necessary for the core function of the complex, for example AEBP2 or PHF1, but that help to specify its function and recruitment. These accessory subunits may only be required for a subset of PRC2 activity and some are mutually exclusive in their binding to the PRC2 core complex. The core components of PRC2 can interact with accessory factors such as JARID2, AEBP2, EPOP, PALI and the polycomb-like (PCL) subunits (PHF1, PHF19 and MTF) (see Fig. 1.6). There are two main types of PRC2 complexes, PRC2.1 is defined by its binding to one of the PLC accessory subunits and EPOP or PALI while PRC2.2 is demarcated by its binding to AEBP2 and JARID2 (Kasinath et al., 2018; Gao et al., 2012; Wong et al., 2016; Laugesen et al., 2019; Grijzenhout et al., 2016; Pasini et al., 2010a; Chen et al., 2018; Beringer et al., 2016; Liefke et al., 2016; Højfeldt et al., 2019; Alekseyenko et al., 2014; Conway et al., 2018; Kim et al., 2009). These subunits have roles in recruitment of PRC2 to specific genomic locations or in alteration of the complex's activity.

JARID2 and AEBP2 have been implicated in PRC2's recruitment to genomic loci (Peng et al., 2010; Kim et al., 2009; Vizán et al., 2015). In PRC2.2 JARID2 helps recruit PRC2 to locations containing H2AK119ub which is placed by PRC1. JARID2 recruitment of PRC2 to chromatin is dependent on the presence of H2AK119ub (Cooper et al., 2016; Kalb et al., 2014). PRC2 can methylate K116 of JARID2, which is bound in turn by the core PRC2 subunit EED, mimicking a H3K27me3 methylated tail, and resulting in an increase in catalytic activity (Sanulli et al., 2015; Kasinath et al., 2018). PRC2 may initially methylate JARID2 before the methylated product binds and stimulates PRC2's activity. Knockout of JARID2 results in differentiation defects and a slight decrease in

16

H3K27me3 placement (Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010a; Peng et al., 2010; Shen et al., 2009). The slight decrease of H3K27me3 upon JARID2 KO may be due to the presence of redundant recruitment mechanisms. JARID2 interacts with SUZ12 as well as with EED when methylated, while AEBP2 binds to RbAp48 mimicking an unmodified H3 N-terminal tail (Kasinath et al., 2018; Justin et al., 2016; Vann and Kutateladze, 2018; Cooper et al., 2016). JARID2's interactions with EED stabilise a region of SUZ12 named the stimulatory response motif (SRM), which is normally disordered and interacts with the active site located in EZH2, stabilising the active conformation (Kasinath et al., 2018; Justin et al., 2016). The C-terminus of AEBP2 binds to RbAp48, positioning its flexible N-terminal domain next to the SET domain, potentially explaining the stimulation of activity seen when AEBP2 is part of the PRC2 complex (Cao and Zhang, 2004a).

The sites of interaction between PRC2 and the accessory subunits overlap, causing the incorporation of these subunits to be mutually exclusive. JARID2, EPOP and Pali all bind to the same area of SUZ12 while the PCL subunits binding site overlaps with AEBP2 (Chen et al., 2018; Laugesen et al., 2019; Mierlo et al., 2019). The PRC2.1 complex variant includes one of the three PCL subunits and either EPOP or PALI (Alekseyenko et al., 2014; Beringer et al., 2016; Grijzenhout et al., 2016; Liefke et al., 2016). The PCL subunits are PHF1, PHF19 and MTF named PCL1, 2, and 3, respectively (Boulay et al., 2011; Perino et al., 2018; Qin et al., 2013). The expression profiles of the PCL proteins dynamically alter throughout development. The PCL proteins contain a winged helix motif which specifically binds to unmethylated CpG motifs, helping recruit PRC2.1 to chromatin (Li and Reinberg, 2011). Knockout of any of the PCL proteins results in a global decrease of H3K27me3, due to the failure of PRC2.1 to be recruited to chromatin (Li and Reinberg, 2011).

PHF1 and PHF19 bind to H3K36me3 via their tudor domains with redundant functions (Cai et al., 2013; Brien et al., 2012; Musselman et al., 2012). PHF19 recruits both PRC2 and NO66 to H3K36me3 modified nucleosomes, resulting in demethylation of H3K36 and placement of H3K27me3 (Ballaré et al., 2012; Brien et al., 2012). PHF1 modulates PRC2's catalytic activity, increasing the overall catalytic activity but also mediating its inhibition when bound to H3K36me3 (Musselman et al., 2012; Li and Reinberg, 2011; Cai et al., 2013; Sarma et al., 2008; Qin et al., 2013). PRC2 containing PHF1 mutants unable to bind H3K36me3 exhibit greater methyltransferase activity (Musselman et al.,

2012) than PRC2 complexes containing wild-type PHF1. MTF regulates PRC2 activity in X-chromosome inactivation, pluripotency and recruitment of PRC2 to unmethylated CpG islands (CGIs) via its DNA binding activity (Casanova et al., 2011; Walker et al., 2010; Perino et al., 2018). The PCL subunits may act to recruit PRC2 to H3K36me3 containing sites and mediate their demethylation and the de novo placement of H3K27me3.

The PCL proteins are usually found in a PRC2 complex with either EPOP or PALI. EPOP helps mediate the interaction between PRC2 and Elongin BC complex (Beringer et al., 2016). The PRC2 complex containing EPOP is abundant in ESCs but is downregulated during differentiation (Beringer et al., 2016; Liefke et al., 2016). The Elongin BC complex helps promote the elongation of transcriptional engaged RNAPII (Bradsher et al., 1993; Liefke et al., 2016). The interaction between Elongin BC and PRC2, mediated by EPOP, may help to sustain transcriptionally capable RNAPII at specific PRC2 target genes and prevent the binding of JARID2 containing PRC2, as deletion of either protein increases the presence of the other at target locations (Beringer et al., 2016). PALI, which is mutually exclusive with both JARID2 and EPOP, stimulates PRC2 activity *in vitro* and a knockout results in a slightly decreased level of H3K27me3 (Alekseyenko et al., 2014; Hauri et al., 2016; Conway et al., 2018).

1.2.3 Polycomb Repressive Complex 1 Variants

The PRC1 group of complexes is highly heterogeneous due to the existence of multiple homologs for each subunit as well as unique accessory factors. So-called canonical PRC1 complexes contain one each of multiple chromodomain proteins (CBX), ubiquitinating enzymes RING1A/B, three polyhomeotic homologs (PHCs) and one of six Polycomb group ring fingers (PCGF) (Fig. 1.5). Combinatorial association of different PCGF proteins leads to functionally distinct PRC1 complexes (Gao et al., 2012; Vandamme et al., 2011; Gil and O'Loghlen, 2014; Aranda et al., 2015; Fursova et al., 2019; Blackledge et al., 2014; Scelfo et al., 2019). Both RING and YY1 binding protein (RYBP) and YY1 associated factor 2 (YAF2) are only found in complexes without both CBXs and PHCs, and therefore help to define non-canonical PRC1 complexes.

Each distinct complex occupies distinct genomic loci and thus may have individual functions within the cell. PRC1 complexes have been shown to catalyse the monoubiquitination of histone H2A at Lysine 119 (H2AK119ub) and to

be involved in compaction of chromatin (Vidal, 2009; Napoles et al., 2004). PRC1 and PRC2 are linked together in repression of genes as CBX-containing PRC1 complexes recognise H3K27me3 deposited by PRC2, while H2AK119ub has recently been shown to help recruit PRC2 through interaction with JARID2 and AEBP2 resulting in a feedforward mechanism (Simon and Kingston, 2009; Blackledge et al., 2015; Kalb et al., 2014; Cao et al., 2002; Min et al., 2003; Mujtaba et al., 2008; Wang et al., 2004b). Although both canonical and non-canonical PRC1 complexes have been shown to ubiquitinate H2AK119 in vitro, non-canonical complexes (Fig. 1.5) have a markedly higher level of catalytic activity (Gao et al., 2012; Scelfo et al., 2019; Blackledge et al., 2014; Wang et al., 2004a; Tavares et al., 2012; Rose et al., 2016; Fursova et al., 2019). Global levels of H2AK119ub were unaffected by *in vivo* depletion of EED, suggesting that non-canonical PRC1 recruited independently of H3K27me3 is responsible for the majority of H2AK119ub genome-wide (Tavares et al., 2012).

1.2.4 PcG Complexes and Repression

PRC1 and PRC2 act to reinforce and maintain repression. They largely share target genomic loci, but also have unique genomic targets and thus likely partially distinct roles in genomic regulation (Ku et al., 2008; Blackledge et al., 2015). PRC1 is also able to be recruited to chromatin in the absence of H3K27me3. KDM2B has been implicated in having a major role in recruitment of PRC1 to unmethylated CGIs (Farcas et al., 2012; Wong et al., 2016). Accessory subunits can also help PRC1's recruitment to chromatin for example PCGF4 interacts with H3 and H4 histone folds. PRC1 can interact with the H2A/H2B acidic patch of nucleosome via an arginine anchor mechanism, mediated by arginines and lysines from RING1B inserting into the acidic patch (Leung et al., 2014).

The recruitment of PRC2 to chromatin is a complicated mechanism that is still not fully understood. The dynamic composition of PRC2 complexes complicates understanding the recruitment of PRC2. PRC2 is recruited to specific DNA sequences in *Drosophila* called PREs; however, no such sequences have been found in mammals (Du et al., 2017; Kassis and Brown, 2013; Bauer et al., 2016; Chan et al., 1994). Artificial sequences with high CpG content and no conserved sequence motifs recruit PRC2 *in vivo* partially explained by PRC2's accessory subunits such as PCLs affinity for CpGs (Li and Reinberg, 2011; Perino et al., 2018; Lynch et al., 2012; Mendenhall et al., 2010; Wachter et al., 2014). PRC2 is

recruited to chromatin by multiple interactions of its subunits with chromatin: from JARID2 with H2AK119ub and EED with H3K27me3 to PCL based DNA binding. Together with PRC1's recruitment by H3K27me3, this helps to explain the co-occupancy of PRC2 and PRC1 at specific genomic locations (Cooper et al., 2016; Cooper et al., 2014; Blackledge et al., 2014; Kalb et al., 2014; Laugesen et al., 2019).

PRC2's potential recruitment to chromatin by RNA is still debated. RNA has been implicated in the recruitment of PRC2 to areas needing repression including X-chromosome inactivation (Kohlmaier et al., 2004; Rinn et al., 2007). However, RNA has been shown to be dispensable for PRC2 recruitment and repression (Rocha et al., 2014; Portoso et al., 2017). Multiple studies have demonstrated that PRC2 binds non-specifically to RNA by interaction sites on EZH2, JARID2 and SUZ12 (Cifuentes-Rojas et al., 2014). This has been posited to recruit and tether PRC2 to areas of active transcription to repress them (Kanhere et al., 2010; Davidovich et al., 2013; Kaneko et al., 2013; Kaneko et al., 2014; Zhao et al., 2010). Others have suggested that the binding of RNA and DNA to PRC2 is mutually exclusive and inhibits PRC2's catalytic activity (Kaneko et al., 2014; Cifuentes-Rojas et al., 2014; Beltran et al., 2016; Wang et al., 2017). Previous studies have shown PRC2 bound to nascent RNA at essentially all active genes suggesting that RNA prevents PRC2 recruitment to chromatin at active genes (Beltran et al., 2016). The dissimilar findings of PRC2's interaction with RNA suggests that different RNAs may have different functions, with some leading to the recruitment of PRC2 and subsequent repression while PRC2's non-specific binding to RNA prevents PRC2's recruitment to active genes.

PRC2 is affected by nucleosome density and it prefers to methylate nucleosomal arrays compared to mononucleosomes (Wang et al., 2017; Yuan et al., 2012). PRC2 genomic targets are usually unmethylated at the DNA level but in areas with low CpG density DNA methylation and PRC2 can co-exist (Brinkman et al., 2012; Murphy et al., 2013; Li et al., 2018). PRC2 and DNA methylation help to co-regulate each other with EED knockout resulting in elevated DNA methylation and DNA methylation preventing the spread of H3K27me3 (Li et al., 2018; Brinkman et al., 2012; Reddington et al., 2013; King et al., 2016; Neri et al., 2013; Rose and Klose, 2014; Hagarman et al., 2013). PRC2 binds poorly to methylated nucleosomes partially due to AEBP2 and EZH2's affinity for unmethylated CpG sequences (Bartke et al., 2010; Boyer et al., 2006; Lynch et al., 2012). PRC2 has been shown to be recruited by H2AK119ub, placed by

PRC1, which can bind to and be recruited by unmethylated CpG-rich DNA (Wu et al., 2013; Farcas et al., 2012).

There are two main methods by which Polycomb complexes repress chromatin; compaction, and placement of histone modifications H2AK119ub and H3K27me3, which inhibit transcription (Eskeland et al., 2010; Zhou et al., 2008; Wang et al., 2004a; Dellino et al., 2004; Francis et al., 2004; Grau et al., 2011; Lau et al., 2017; Collinson et al., 2016; Cao and Zhang, 2004a; Pasini et al., 2007). H2AK119ub represses the placement of H3K4me2/3 and inhibits the placement of H3K36me3 (Yuan et al., 2013; Nakagawa et al., 2008). The ability of H2AK119ub to repress chromatin by the inhibition of transcription is separate to PRC1's ability to compact chromatin (Kundu et al., 2017). The placement of H3K27me3 reduces H3K4me3 deposition by methyltransferases and limits the recruitment and activity of RNAPII (Dellino et al., 2004; Brien et al., 2012; Kim et al., 2013). Removal of H2AK119ub and H3K27me3 results in an increase in gene transcription, although the mechanism of this is debated (Stock et al., 2007; Endoh et al., 2012). Both H2AK119ub and H3K27me3 may help to establish a repressive chromatin state by prevention of H3K4me3 establishment and therefore a reduction in transcription initiation or by ensuring direct recruitment of Polycomb complexes (Schuettengruber and Cavalli, 2009; Zhou et al., 2008; Wang et al., 2004a; Napoles et al., 2004; Agger et al., 2007; Burgold et al., 2008; Nakagawa et al., 2008; Farcas et al., 2012; Wang et al., 2017; Margueron et al., 2009; Schwartz and Pirrotta, 2014; Morey et al., 2012; Wang et al., 2004b). PRC2 can also recruit H3K36me3 demethylases helping to repress the gene (Abed and Jones, 2012; Brien et al., 2012).

PRC1 can compact chromatin potentially restricting access of transcription factors to the compacted region (Eskeland et al., 2010; Francis et al., 2004; Illingworth, 2019; Grau et al., 2011; Lau et al., 2017; Simon and Kingston, 2013; Kundu et al., 2017). This mechanism involves the compaction region of CBX2 and a dimerization domain of PHC2 (Isono et al., 2013; Lau et al., 2017; Wani et al., 2016; Kundu et al., 2017). However, although PRC1 compacts chromatin and removal of this function leads to derepression of target genes, whether compaction represses genes via restriction of access is debated (Hodges et al., 2018; King et al., 2018; Isono et al., 2013; Lau et al., 2017; Wani et al., 2016; Kundu et al., 2017). Although Polycomb complexes have been mostly associated with transcriptional repression, there is some suggestion of its involvement in transcription activation (Creppe et al., 2014; Kondo et al., 2014), potentially via

helping maintain contacts between the promoter and enhancer of target genes (Kondo et al., 2014; Gentile et al., 2019; Cruz-Molina et al., 2017).

1.2.5   H3K4 methyltransferases of the SET1 and MLL families

Members of the Trx protein family form multiple important complexes in mammalian cells that are named after their catalytic subunits, including MLL1/2/3/4 and SET1A/B (Fig. 1.4). SET1A/B and MLL1/2 complexes place methylation at histone H3 lysine 4 (H3K4me3), a hallmark of active chromatin (Woo et al., 2017; Geisler and Paro, 2015; Poynter and Kadoch, 2016). SET1A/B and MLL complexes share many subunits including ASH2, WDR5, ASH2L, DPY30 and RBBP5, which form a stable core complex (the WRAD complex) independent of the catalytic subunit, resulting in a complex that can associate with any of MLL1/2 and SET1A/B (Dillon et al., 2005; Herz et al., 2013; Qian and Zhou, 2006; Ernst and Vakoc, 2012). The C-terminus of the SET1 and MLL proteins is conserved and contains the WDR5 interaction motif as well as the SET domain responsible for their histone methyltransferase (HMT) activity (Patel et al., 2009). Although the RBBP5 and ASH2L heterodimer is the minimum that is required for activation of the MLL proteins, WDR5 is important for maximal catalytic activity and proper targeting of the MLL complexes (Li et al., 2016). Components of the WRAD complex also interact with other transcriptions factors outside the Trx family such as OCT4 (Ang et al., 2011) and MYC (Ullius et al., 2014; Bochyńska et al., 2018; Ernst and Vakoc, 2012). This large range of interactions and roles of the WRAD subunits both with SET/MLL complexes and TFs may explain why these subunits are essential for mouse embryogenesis and development (Stoller et al., 2010; Yang et al., 2016; Yang and Ernst, 2017).

SETD1A/B are responsible for the global levels of H3K4me3 while MLL1/2 are more specific in H3K4me3 placement. Both SETD1A/B contain the CFP1 (CxxC zinc finger protein 1) subunit (Fig. 1.4) which helps the recruitment of all SET1 complexes to unmethylated CpG islands and links them to H3K4me3 via its recruitment to H3K4me3 via its PHD domain (Clouaire et al., 2012; Xu et al., 2011; Thomson et al., 2010). Deletion of this protein mainly affects H3K4me3 placed at highly expressed active genes (Hui Ng et al., 2003; Clouaire et al., 2012). MLL1 and MLL2 are homologous and can share two subunits in addition to those that form the WRAD complex: MENIN and LEDGF (FitzGerald and

Diaz, 1999; Hughes et al., 2004; Yokoyama and Cleary, 2008). MLL1 and MLL2 are both targeted to Hox genes and may be able to partially compensate for each other although they have distinct functions (Glaser et al., 2009; Denissov et al., 2014; Yokoyama and Cleary, 2008; Hughes et al., 2004; Murai et al., 2014).

The MLL2 protein (600 kDa) contains a CXXC zinc finger, three PHD zinc fingers, a bromodomain and a SET domain. MLL2 is recruited to chromatin through multiple mechanisms including interactions with sequence specific transcription factors (TFs) (Ang et al., 2011; Mo et al., 2006; Demers et al., 2007) and histone variants including H2A.Z (Hu et al., 2013b). The CXXC domain contained within MLL2 is important for its targeting to chromatin while the SET domain is required for catalytic activity (Risner et al., 2013; Hu et al., 2017; Birke et al., 2002; Bach et al., 2009). The third PHD finger is able to bind H3K4me3 and therefore may be able to recruit MLL2 to H3K4me3 initiating a feedbackward loop (Ali et al., 2014). MLL2$^{(-/-)}$ ESCs are viable but display proliferation defects, however the phenotype is more severe in a double knockout of MLL1$^{(-/-)}$ and MLL2$^{(-/-)}$ (Glaser et al., 2009; Denissov et al., 2014; Lubitz et al., 2007; Hu et al., 2013b). This suggests at least partial redundancy between MLL2 and MLL1. Knockout of MLL2 results in embryonic lethality suggesting its function is important in early development (Glaser et al., 2009). Although a knockout of the MLL2 does not affect genome-wide levels of H3K4me3 in ESCs, due to the presence of multiple of H3K4 methyltransferases, it does affect the expression of a large group of genes (Denissov et al., 2014) and is required for global H3K4 methylation in oocytes.

1.2.6   SET-domain Family Methyltransferases

There are a large variety of PTMs including methylation that can be placed on a number of residues on multiple core histones. Methylated residues are a significant player in the epigenetic control of gene expression. A number of enzymes that place these methyl groups are part of the SET-domain family methyltransferases including SETD1A/B, MLL1-4, and PRC2. The SET domain is approximately 130 residues long, and proteins containing SET domains are present in all eukaryotic organisms (Cheng, 2014; Dillon et al., 2005; Qian and Zhou, 2006; Herz et al., 2013). Both PRC2 and MLL2 contain SET domains and transfer a methyl group from the co-factor SAM to the ϵ amino group of the target lysine of either H3K4 or H3K27 on the substrate. The SET domain is composed

of β strands folding into three sheets that surround a knot-like structure formed by the C-terminus of the SET domain (see Fig. 1.7).

The SET domain contains the core SET domain which is usually flanked by both a N-terminal SET region (nSET) and a C-terminal SET region (cSET) that are not part of the SET domain itself. The nSET domain helps stabilise the active site, while the cSET domain forms part of the active site, usually by providing an aromatic residue to pack against the SET active site helping form a hydrophobic channel required for the enzymatic activity of the SET domain (Qian and Zhou, 2006; Zhang et al., 2002). Some SET domains have a third component of the active site referred to as I-SET which is inserted into the active site. This inserted region varies in length and sequence between enzymes that contain the I-SET, when combined with the structure of the SET domain it potentially helps to confer substrate specificity to the enzyme (Qian and Zhou, 2006). The residues either side of the substrate lysine contribute to substrate specificity. A narrow hydrophobic channel, which contains an invariant tyrosine residue implicated in the enzymatic mechanism, links the substrate lysine and the cofactor SAM, has been seen in all crystal structures of SET domains so far (Qian and Zhou, 2006).

The substrate and cofactor are placed at opposite sides of the active site. The sulphur atom of the SAM cofactor is positioned at the ϵ-amino group of the target lysine. There are a few mechanisms that would allow the transfer of the methyl group from the cofactor to the substrate. A theoretical study on SET7/9 supports the methyl transferase from the SAM to the target lysine proceeding by a direct in-line S_N2 nucleophilic attack (Qian and Zhou, 2006). However, this mechanism has not been confirmed and there are several alternatives including deprotonation of the substrate lysine by the active site before methyl transfer. The tyrosine may provide a charge-dipole interaction between its phenoxyl and positive charged sulphur of SAM or may interact with and align the amino nitrogen of the lysine to the methyl carbon of the SAM to facilitate the S_N2 methyl transfer (Dillon et al., 2005; Qian and Zhou, 2006; Herz et al., 2013).

The SET domain of PRC2 is contained within the EZH2 subunit and has many of the features associated with the classic SET domain structure (see Fig. 1.7). The SET domain has the classic beta sheet harbouring the active site and the typical pseudoknot (Wu et al., 2013). The I-SET contains a β-hairpin, which helps to form the substrate binding cleft, while the methyl-lysine channel is formed from aromatic residues including the invariant tyrosine. Nevertheless, the PRC2 SET domain does contain some divergent elements. The N-terminal region of

24
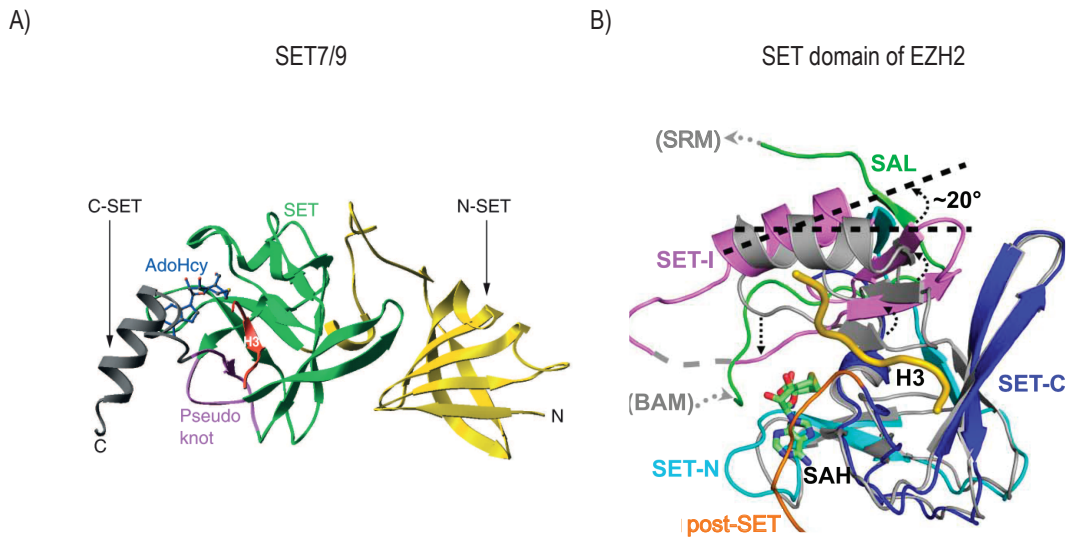
Figure 1.7 *Various SET domains Structures.* *A) The structure of human SET7/9 enzyme's classic SET domain. NSET in yellow, pseudoknot in purple, SET domain in green and CSET in grey. The H3 peptide is in red, and cofactor byproduct S-adenosyl-L-homocysteine (AdoHcy) in blue. Figure taken from Dillon et al., 2005. B) The crystal structure of active and inactive EZH2's SET domain are overimposed to demonstrate the 22° rotation of the I-SET. All domains are labelled. I-SET is depicted in purple, Sal in green, nSET in light blue, cSET in dark blue and the H3 peptide in yellow. Figure taken from Jiao and Liu, 2015.*

the catalytic domain comprises of two Zn_3Cys_9 motifs (CxC), whereas the SET domain is part of the C-terminal region of EZH2 with a I-SET that partially forms the active site cleft and a cSET motif which diverges from the classic SET structure (see Fig. 1.7). The CxC domains interacts with the DNA of the nucleosome, helping position the nucleosome and H3K27 for modification (Wu et al., 2013; Poepsel et al., 2018) potentially helping regulate PRC2's response to varying nucleosome density.

In the active conformation of SET domains, the cSET contributes to the formation of both the substrate and cofactor binding site, in EZH2 the cSET is positioned towards the CxC domains, partially blocking the active site and preventing formation of the co-factor site (see Fig. 1.7). When EZH2 is in the autoinhibited conformation the I-SET is also close to the cSET, causing the closure of the methyl-lysine binding channel (Jiao and Liu, 2015; Wu et al., 2013; Antonysamy et al., 2013). In addition, the cofactor binding cleft in the active PRC2 complex that is formed by the I-SET and the cSET domains, does not form fully in the autoinhibited conformation, this autoinhibition is relieved upon EZH2 binding to SUZ12 and EED (Wu et al., 2013; Jiao and Liu, 2015; Jiao and Liu, 2016).

There are multiple interaction points between EED, EZH2 and SUZ12. The interaction of the SET activation loop (SAL) of EZH2 with EED and the VEFS domain of SUZ12 is critical for the stabilisation of the active conformation of the SET domain and results in the movement of the I-SET 20 degrees anti-clockwise, away from the co-factor binding site and towards the substrate site, see Figure 1.7 (Jiao and Liu, 2015). This allows entry of the substrate into the active site and catalytic activity is restored. The SAL of EZH2 extrudes from between the SET domain and the VEFS to interact with the stimulatory recognition motif (SRM) of EZH2. PRC2 binds methylated H3K27 or JARID2 K116 in a sandwich like assembly between the EED WD40 aromatic cage and the SRM. SRM marks extensive contacts with the I-SET of the SET domain, in addition to the contacts SRM makes with the SAL potentially helping communicate structural changes from one to the other. Mutations to the SAL, I-SET, SEM and VEFS domains decrease or completely obliterate H3K27me3-meditated simulation of catalytic activity (Jiao and Liu, 2016; Wu et al., 2013; Antonysamy et al., 2013; Justin et al., 2016; Jiao and Liu, 2015). Both the EED and the SRM bind the H3K27me3 peptide stabilising the position of the I-SET stimulating the SET domain (Jiao and Liu, 2015).

1.3 Bivalent Domains

1.3.1 Discovery and Composition

H3K4me3 and H3K27me3 are largely segregated throughout the genome, marking different transcriptional states. However, these marks were found to co-occur on the majority of developmental gene's transcription start sites (TSS) in embryonic stem cells (ESCs). Their sites of coexistence were termed bivalent domains (see Fig. 1.8) (Bernstein et al., 2006a; Azuara et al., 2006). Bivalent domains are located at CpG islands and the genes on which they are located are expressed at very low levels in ESCs (Bernstein et al., 2006a; Mikkelsen et al., 2007; Azuara et al., 2006). H3K27me3 at bivalent domains is placed by PRC2 (Laugesen et al., 2019; Sharif and Koseki, 2018; Bernstein et al., 2006a; Azuara et al., 2006), and H3K4me3 is placed by MLL2 (Hu et al., 2013b; Denissov et al., 2014; Clouaire et al., 2012). The initial discovery of bivalency was obtained by the use of chromatin immunoprecipitation (ChIP), re-ChIP, and hybridisation arrays and therefore could have been present only in a fraction of the cell population

or the marks present on neighbouring nucleosomes. Several studies used various re-ChIP techniques to address these areas of concern (Geisberg and Struhl, 2004; Truax and Greer, 2012; Kinkley et al., 2016; Brookes et al., 2012) and confirmed the presence of both marks on the same nucleosome. The genome-wide re-ChIP-seq used by Kinkley et al. (2016), Vastenhouw et al. (2010), Mikkelsen et al. (2007), and Sen et al. (2016) confirmed wide-spread bivalency at hypomethylated CpG islands coinciding with inactive promoters of developmental regulators. Although some previously discovered bivalent promoters may have been due to previously undiscovered heterogenous bivalency of the cellular population further work utilising multiple techniques confirmed the existence of bivalency within a cell (Shema et al., 2016; Voigt et al., 2012; Kinkley et al., 2016; Brookes et al., 2012; Mikkelsen et al., 2007).

The existence of both H3K4me3 and H3K27me3 in the same nucleosome but on separate histone H3 tails was established by mass spectrometry analysis of H3 histone N-terminal tails and reChIP on purified mononucleosomes, however these results do not place asymmetric nucleosomes in the context of the genome (Young et al., 2009; Voigt et al., 2012). This asymmetric nucleosome conformation is supported by *in vitro* data from multiple studies. PRC2 is inhibited from placing H3K27me3 when nucleosomes are symmetrically but not asymmetrically modified with H3K4me3 (Voigt et al., 2012). This inhibition is reciprocated by histone-lysine N-methyltransferases (KMT2) enzymes which place H3K4me3, the placement of H3K27me3 reduces their catalytic activity (Kim et al., 2013). The combination of these techniques both *in vitro* and *in vivo* suggest that the bivalent domains present in cells consist of asymmetrically modified nucleosomes. ChIP techniques rely on populations of the cells, and sequential ChIP does not distinguish between nucleosomes modified on separate H3 tails or on the same H3 tail, while re-ChIP distinguishes between nucleosomes from different alleles. MS techniques only compares PTMs on the same histone peptides or within the same nucleosome but does not give any information on the genomic location of these marks. Shema et al. (2016) gave the definitive evidence of the bivalent asymmetric nucleosomes in combination with their genomic location *in vivo*. The distance between the different histone PTMs distinguishes between marks present on the same H3 tail or those on separate H3 tails, while sequencing of the DNA in the nucleosomes allows placement of the bivalent asymmetric nucleosome in specific genomic location.

Bivalent domains were shown not to be an artefact of *in vitro* cultured ESCs by
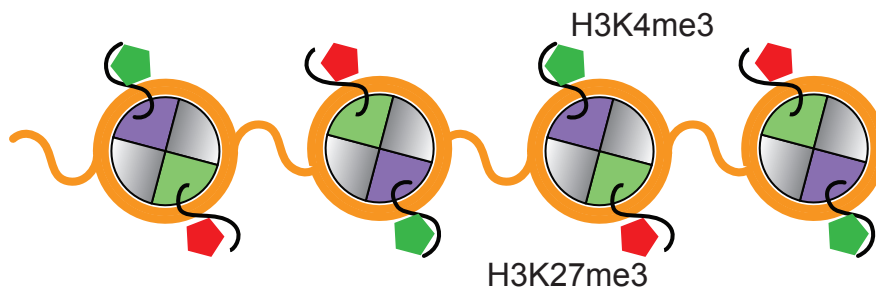
Figure 1.8 *Schematic of bivalent domain composition.* *Bivalent domains are composed of asymmetrical nucleosomes with one H3 modified with H3K4me3, and the sister H3 in the same nucleosome modified with the H3K27me3 mark. One H3 copy in green, the sister H3' copy in purple, H3K4me3 modification in green, and H3K27me3 modification in red.*

confirmation of their existence in both early pre and post implantation embryos (Rugg-Gunn et al., 2010; Alder et al., 2010; Dahl et al., 2010). The existence of bivalent domains *in vivo* was further confirmed by their presence in developing zebrafish (Vastenhouw and Schier, 2012; Lindeman et al., 2011). There is a strong degree of conservation in bivalent domains between human and mouse ESCs (Mikkelsen et al., 2007; Pan et al., 2007; Zhao et al., 2007) suggesting a link between bivalency and CpG islands and some maintenance of bivalent domains during evolution. Bivalent domains are not universal in all organisms, as studies on *Xenopus* embryos found very few true bivalent domains composed of nucleosomes asymmetrically modified with H3K4me3 and H3K27me3 (Akkers et al., 2009). Investigation into bivalent domains in *Drosophila* initially suggested that they may have bivalent domains consisting of the combination of H3K4me3 and H3K9me3 (Saha et al., 2019). Although recent research has shown evidence of the co-occurance of H3K4me3 and H3K27me3 forming bivalent domains (Akmammedov et al., 2019) where their function has been posited to be the same as in mammals, helping to poise the gene for quick regulation during development.

Bivalent domain composition may have been solved but the function of bivalent domains has yet to be ascertained, although, multiple hypotheses have been posited. As these marks were initially thought to be segregated in the genome due to their association with completely different transcriptional states, the overlap of both active and repressive marks is generally thought to 'poise' the gene for quick regulation during development (Bernstein et al., 2006a; Azuara et al., 2006). The low level of transcription of bivalent genes supports this juxtaposition of the marks ready for quick regulation. This hypothesis has been supported by the discovery that the majority of these bivalent domains found in ESCs are resolved during differentiation (Bernstein et al., 2006a; Mikkelsen et al., 2007). If a bivalent

28

domain is activated upon differentiation then it retains the H3K4me3 and loses H3K27me3, while other bivalent domains lose the H3K4me3 mark and gain more H3K27me3 to become completely silenced.

There are some bivalent domains that remain bivalent in terminally differentiated cells but they are a much lower number than that seen in ESCs. Approximately 8% of genes bivalent in ESCs remain bivalent in neuronal progenitor cells (NPCs) and may be due to the lack of expression of any transcription factors (TFs) that would activate them, therefore the loss of H3K4me3 is not required (Mikkelsen et al., 2007). Some bivalent domains lose both marks during differentiation and are silenced by DNA methylation (Sharov and Ko, 2007). The formation of bivalent domains may alter with different cell types as the PcG recruitment mechanism alters (Oliviero et al., 2016; Kloet et al., 2016), therefore the function of bivalent domains may change between cell types.

The theory that bivalent domains are important in differentiation is supported by the phenotypes observed in cells deficient in MLL2 and PRC2 subunits. Knockout of MLL2 results in removal of H3K4me3 specifically from bivalent domains and severe defects in differentiation (Lubitz et al., 2007; Hu et al., 2013b; Denissov et al., 2014). Surprisingly, if DPY30 is removed some studies have seen an alteration of H3K4me3 placement at bivalent domains, suggesting a function of DPY30 specific to bivalency (Jiang et al., 2011). The number of bivalent domains in ESCs vary slightly depending on the method used to quantify them, approximately 2500 bivalent domains are found in ESCs. Mikkelsen et al. (2007) stated that of the CGIs with H3K4me3, 22% of were bivalent. Voigt et al. (2012) estimated that 15% of all nucleosomes modified with H3K4me3 were co-modified with H3K27me3, while Shema et al. (2016) suggested the presence of approximately 0.5% of all nucleosomes are bivalent which decreased upon ESC differentiation. As the number of bivalent nucleosomes that make up a domain is unknown, and likely variable, a comparison of bivalent nucleosomes and domains is impossible. As a result of the different techniques and cell lines used, the percentages of bivalent domains or nucleosomes found in different studies are unable to be directly compared.

Bivalent domains are resolved differently depending on the differentiation pathway activated, further supporting the premise that bivalent domains are important during development (Mikkelsen et al., 2007). The number of bivalent domains present in terminally differentiated cells was proposed to correlate with their differentiation potential. The majority of bivalent domains in ESCs

are repressed during differentiation, while the majority of bivalent domains overlapping with developmentally important transcription factor promoters lose H3K27me3 (Pan et al., 2007; Zhao et al., 2007; Cui et al., 2009). If bivalent domains include developmentally important genes, for example TF and lineage specific genes, requiring quick regulation during differentiation then only a small subset of these would need to be activated during development as the vast majority would not be required for that specific lineage.

Nevertheless, there are non-bivalently modified genes that are activated quickly upon differentiation, indicating multiple methods of poising genes for regulation (Pan et al., 2007). As bivalent domains are found in multiple cell types and there are several genes that become bivalently modified during differentiation including pluripotency factors such as SOX2, this suggests that bivalent domains are not unique to ESCs and do not solely poise gene requiring quick expression (Mikkelsen et al., 2007; Mohn et al., 2008; Pan et al., 2007; Roh et al., 2006; Cui et al., 2009; Barski et al., 2007). It could be argued that the pluripotency genes need to be switched off quickly to allow for successful differentiation and the quickest way to do this, is to establish a bivalent domain as this requires only the removal and replacement of one mark and decreases expression of that gene to negligible levels (Bernstein et al., 2006a; Mikkelsen et al., 2007; Pan et al., 2007; Zhao et al., 2007).

Some cell types retain bivalent domains on genes required for rapid responses, such as immune response, even if they are not developmental genes. In CD8$^+$ memory T cells, genes required for their activation are in a bivalent chromatin state (Weng et al., 2012; Kinkley et al., 2016; Roh et al., 2006; Araki et al., 2010). However, some bivalent genes remain in terminally differentiated cells that are not important in development or rapid response, the reason why these genes remain bivalent is not fully understood. These genes may remain bivalent as the TF that would activate them is not expressed or they could have a different function. A consequence of bivalency is the presence of poised RNAPII at bivalent domains (Pan et al., 2007; Dao et al., 2016; Stock et al., 2007). Studies have suggested that bivalency and the RNAPII poising may mark the gene ready for or silenced after an expression burst. This still suggests a role for bivalency in quick regulation of gene expression and a plausible answer for the formation of bivalency on pluripotent genes after differentiation.

Artificial CpG-rich sequences introduced into cells obtain both H3K4me3 and H3K27me3, indicating that bivalency may be the default position of inactive

CpG-rich promoters (Mendenhall et al., 2010; Thomson et al., 2010; Woo et al., 2010; Lynch et al., 2012). Unmethylated CGIs recruit H3K4 methyltransferases such as SET1 and MLL1/2 (Clouaire et al., 2012; Bach et al., 2009; Birke et al., 2002; Risner et al., 2013; Hu et al., 2017). The presence of H3K4me3 prevents de novo DNA methylation, while the absence of activating transcription factors leads to the establishment of H3K27me3 and thus bivalency (Ku et al., 2008; Morselli et al., 2015). An extension of the existing model, which still links bivalent domains with transcriptional regulation, is posited by Kinkley et al. (2016); bivalent domains may be a dual repression control mechanism. The presence of H3K4me3 prevents DNA methylation of the promoter, while H3K27me3 prohibits productive elongation (Ferrai et al., 2017; Rada-Iglesias, 2017; Brookes et al., 2012; Aranda et al., 2015). The behaviour of H3K27me3 with DNA methylation is more complicated with most evidence suggesting the antagonistic role of H3K27me3 and DNA methylation (Reddington et al., 2013; Lynch et al., 2012; Xie et al., 2013; Li et al., 2018; Neri et al., 2013; Hagarman et al., 2013), however some studies have shown that a depletion of DNA methylation results in a decrease of H3K27me3 peaks and there is evidence for H3K27me3 and DNA methylation coexisting (Brinkman et al., 2012; Viré et al., 2006; Wang et al., 2017), although only at very low levels of both PRC2 and DNA methylation. This could be the point at the exchange between the different types of repression either caused by PRC2 or by DNA methylation. If H3K27me3 excludes DNA methylation then this extension of the model fails to explain H3K4me3's role at bivalent domains. Ku et al. (2008) suggested that bivalency may buffer the pluripotent state of ESCs by reinforcing the repression of factors that induce differentiation. Although the removal of either functional PRC2 or MLL2 affects the differentiation of stem cells the maintenance of EZH2$^{(-/-)}$ cells is possible making bivalency the only method of controlling self-renewal unlikely (Chamberlain et al., 2008; Shen et al., 2008).

1.3.2 Further Modification at Bivalent Domains

Not all bivalent domains are created equal; there are further marks and modifications that co-exist with bivalency (see Fig. 1.9), potentially leading to differences in the proteins that interact with them. Multiple studies have looked into which chromatin modifications co-localise with bivalent domains hoping to elucidate bivalent domain function and discover why certain bivalent domains are resolved differently. The distribution of H3K27me3 vs H3K4me3 varies between
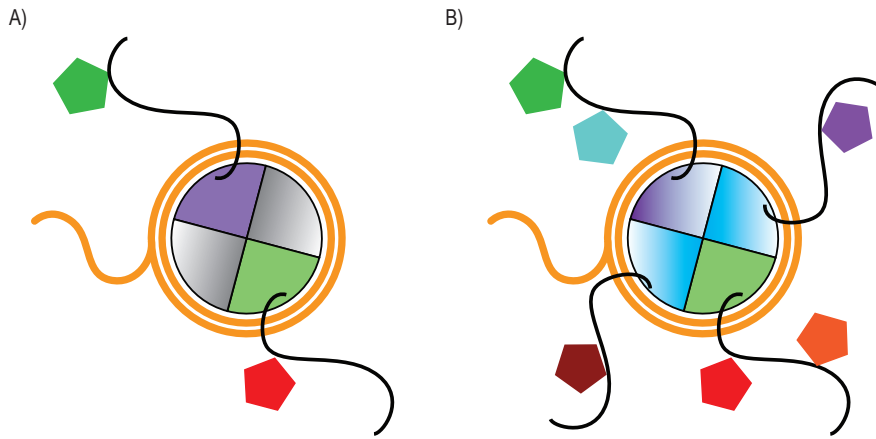
Figure 1.9 *Diagram of Bivalent Nucleosomes and Potential Further Modifications. A) Simplified bivalent nucleosome. One H3 copy with H3K4me3 and the sister H3 copy with H3K27me3. B) A schematic of a more complex and potentially more biologically accurate bivalent nucleosome. One H3 is modified with H3K4me3K14ac, while the sister H3' copy contains H3K9ac27me3. This nucleosome contains H2A.Z variants modified with ubiquitin and acetylation (H2A.Zub/ac). One of the 2 canonical core H3 copy (H3.1) has been replaced with the H3.3 variant.*

bivalent domains resulting in some studies classifying bivalent domains by the either wide or narrow H3K27me3 distribution (Mantsoki et al., 2018; Ku et al., 2008). Bivalent promoters containing a wide spread of H3K27me3 tend to coincide with the presence of PRC1 on the domain (Ku et al., 2008). Bivalent domains can also be classified by the presence or absence of PRC1. The presence of PRC1 at bivalent domains has been suggested to be important for multiple reasons: PRC1[+] bivalent domains are more likely to retain H3K27me3 during differentiation, they are better evolutionary conserved than PRC1[-] domains, and they affect the levels of poised RNAPII present (Ku et al., 2008; Min et al., 2011). 90% of PRC1 target sites correspond to PRC2 bivalent regions, suggesting a large overlap of PRC1 and PRC2 also at bivalent domains. The wider spread of H3K27me3 may make it easier for PRC1 to be recruited by H3K27me3 by its CBX subunits (Wang et al., 2004b; Morey et al., 2012). The presence of PRC1 has been shown to compact chromatin (as discussed earlier), and may help reinforce the repressive effect of H3K27me3.

Bivalent domains contain poised RNA polymerase II (RNAPII), are enriched for developmental genes and are dynamically regulated during development (see Fig. 1.10). The function of poised RNAPII at any gene, including bivalent domains, is debated. The presence of poised RNAPII may help to explain the low level of transcripts of bivalent genes. The presence of poised RNAPII at bivalent domains (Fig. 1.10) was first discovered by Stock et al. (2007). They classify bivalent

32

domains into two groups, based on the presence of PRC1. They attributed the presence of poised RNAPII to the presence of PRC1 and H2AK119ub. They find that the bivalent domains bound by PRC1 have a higher amount of poised RNAPII. The deletion of core PRC1 subunits resulted in the loss of poised RNAPII and an increase of gene repression. Min et al. (2011), however, found less RNAPII at bivalent domains where PRC1 was present.

Mantsoki et al. (2018) found poised RNAPII at bivalent genes, however, they divide bivalent genes into classes by different levels of RNAPII. Poised RNAPII is an accepted feature of bivalent domains although the mechanism of its placement and function is debated (Ku et al., 2008; Min et al., 2011; Mantsoki et al., 2018; Cui et al., 2009; Vastenhouw et al., 2010). The poised RNAPII along with the presence of both PTMs could contribute to the timely regulation of the bivalent genes during development. The class with higher RNAPII poising (1) is enriched for genes involved in transcriptional and cell cycle control while the other class (2) are more tissue specific and developmentally important genes. Class one is more likely to be bivalent across cell types as they consistently change from poised to not poised upon cell activation. Dao et al. (2016) suggested that poised RNAPII marks genes that are silenced after or before a quick burst of transcription supporting theory of bivalent domains being used for quick regulation. Mantsoki et al. (2018) and Xu et al. (2017) suggest another possible function of bivalent domains in achieving mono-allelic expression in different cell types. Poised RNAPII is an accepted feature of bivalent domains although the mechanism of its placement and function is debated (Ku et al., 2008; Min et al., 2011; Mantsoki et al., 2018; Cui et al., 2009; Vastenhouw et al., 2010). The poised RNAPII along with the presence of both PTMs could contribute to the timely regulation of the bivalent genes during development.

There are certain chromatin signatures that further characterise subsets of bivalent domains (see Fig. 1.9 and Fig. 1.10) (Cui et al., 2009; Roh et al., 2006; Ku et al., 2012). H2A.Z is a variant of the core histone H2A, enriched at active and bivalent genes and essential for development although its function is still controversial (Cui et al., 2009; Sen et al., 2016; Hu et al., 2013b; Ku et al., 2012; Margueron and Reinberg, 2010). It has been implicated in nucleosome stability, dynamics (see section 1.1.1), histone exchange, transcriptional activation and Polycomb repression. Its function may therefore depend on the surrounding environment that it is placed in, in the genome (Hu et al., 2013b; Giaimo et al., 2019; Guillemette et al., 2005; Jin and Felsenfeld, 2007; Bönisch and

Hake, 2012; Park et al., 2004; Li et al., 1993; Abbott et al., 2001; Suto et al., 2000; Zhang et al., 2005). H2A.Z is found flanking nucleosome deficient sites in nucleosomes containing histone variant H3.3, both of which have been associated with transcriptional repression and activation in a context dependent manner (Shi et al., 2017; Sen et al., 2016; Ku et al., 2012; Jin et al., 2009; Hardy et al., 2009; Szenker et al., 2011). The combination of both H3.3 and H2A.Z has been thought to render nucleosomes labile, and the chromatin containing them susceptible to nuclease digestion and high ionic conditions (Ku et al., 2012; Suto et al., 2000; Jin and Felsenfeld, 2007). The incorporation of H3.3 has been suggested as necessary for the proper establishment of bivalency (Banaszynski et al., 2013).

H2A.Z has been implicated in the recruitment of RNAPII and it can facilitate the association of both MLL2 and PRC2 to chromatin, potentially explaining its inclusion at bivalent domains (Hu et al., 2013b; Weber et al., 2014; Guillemette et al., 2005; Li et al., 2005a; Creyghton et al., 2008). The H2A.Z incorporated into bivalent promoters contains the H2A.Zub/ac mark (Ku et al., 2012). The ubiquitylation is placed by PRC1 on K119 and is associated with silencing, while the acetylation is associated with activation (Millar et al., 2006; Morales and Richard-Foy, 2000; Draker et al., 2012; Valdes-Mora et al., 2012; Sarcinella et al., 2007; Endoh et al., 2012; Napoles et al., 2004). Bivalent domains contain H2A.Z which is itself dually modified with both ubiquitylation and acetylation; potentially helping maintain bivalency and poised RNAPII until the gene is regulated during differentiation (Ku et al., 2012).

The PTMs of bivalent domains are not curtailed to the H3K4me3 and H3K27me3 marks (Fig. 1.10). In addition to H2A.Zub/ac marks, further PTMs may be placed on the nucleosomes. Cui et al. (2009) showed that the bivalent domains that are most likely to be activated are associated with higher levels of H2A.Z, H3K4me1, H4K20me1 and RNAPII. Additional studies performed on bivalency in differentiated cells including T cells demonstrated the presence of H3K14ac and H3K9ac at bivalent promoters (Roh et al., 2006; Barski et al., 2007). The presence of all these PTMs and chromatin modifications allow an insight into how bivalent domains functions and emphasise the complexity of the histone code and the difficulty of isolating the function of any specific chromatin state.

In addition to the transcription factors and chromatin modifiers that have been shown to localise to bivalent domains two proteins, NONO and ASF1A, were discovered that bind to a subset of bivalent domains. Both of these proteins are localised to other genomic regions in addition to bivalent genes. Nono, a

34
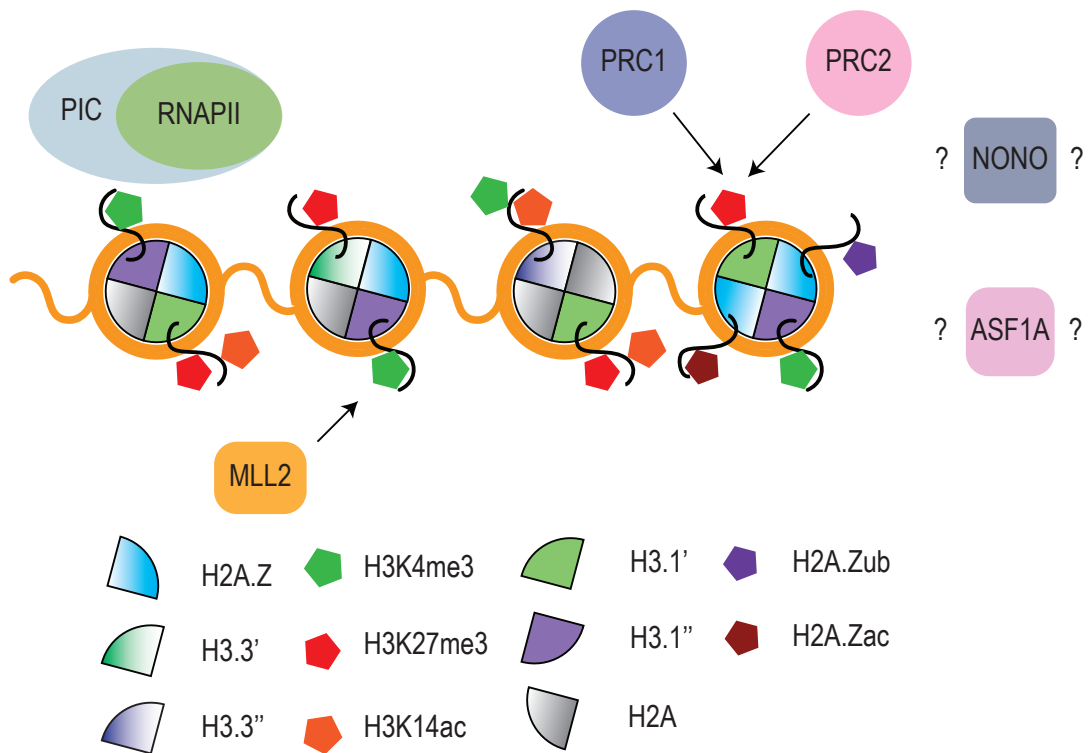
Figure 1.10 *Schematic of Bivalent Domain Composition with Additional Modifications and Potential Protein Interactions.* *Bivalent domains are composed of more than H3K4me3, and H3K27me3 co-occurrence. Details of added modifications depicted in the diagram. Possible recruitment of PRC2, PRC1, RNAPII and MLL2 is shown. Nono and Asf1a bivalent domain interactors are depicted, however, as yet their potential recruitment mechanism is unknown.*

component of para-speckles, interacts with ERK at bivalent domains with a high level of poised RNAPII (Ma et al., 2016). ERK helps keep RNAPII in a paused state (Tee and Reinberg, 2014). Culturing of mouse ESCs in 2i media containing MEK inhibitors results in a loss of ERK and NONO from chromatin and altered RNAPII activity at the NONO positive bivalent genes, potentially causing a switch to a more robust self-renewing phenotype (Ma et al., 2016). The increased pluripotency of the 2i ESCs may be partially due to the altered regulation of a subgroup of bivalent domains marked by NONO (Silva et al., 2008; Sim et al., 2017; Marks et al., 2012; Ma et al., 2016). Removal of NONO at these genes may partly explain the lower numbers of bivalent domains found in ESCs in 2i media. Asf1 is a histone chaperone involved in nucleosome assembly and disassembly (Mousson et al., 2007). Removal of Asf1 isoforms compromises the activation of certain lineage specific genes that are bivalent in ESCs (Gao et al., 2018) and impairs cellular proliferation in human cells (Corpet et al., 2011). Asf1a has been suggested to bind to H3.3-H4 dimers, removing bivalent nucleosomes and therefore helping to resolve the bivalent domain (Gao et al., 2018). How and why Nono and Asf1 are recruited to only a subset of bivalent domains is unknown (Fig. 1.10), this partial recruitment suggests that their function and importance is not applicable to all bivalent domains.

1.4 Aims

The roles of chromatin states are mediated by the proteins that bind them, for example H3K4me3 correlates with active transcription as it is bound by proteins that stimulate transcription (Lauberth et al., 2013). Bivalency is more complicated due to the presence of both activating and repressive marks, however, its role should be communicated in a similar manner. We hypothesize that the biological function of bivalent domains will be elucidated by the identification and analysis of the proteins that bind them. The primary aim of this study was to identify the proteins that bind bivalent asymmetric nucleosomes in order to clarify the nature of bivalent domains. Although the majority of previous studies hypothesize that bivalent domains act to poise the gene for timely regulation during differentiation there is a deficiency of functional data to support this. A secondary aim of this study was to elucidate the functional difference between the proteins recruited to a symmetrically modified and asymmetrically modified nucleosome. An *in vitro* approach was selected to ensure control over the factors

present and parameters tested. This requires the generation and purification of specifically modified histones and assembly of modified nucleosomes. The use of mononucleosomes avoids the complications of proteins binding to a specific combination and orientation of nucleosomes in an array. Specifically modified nucleosomes will be generated and incubated with nuclear extract from mouse embryonic stem cells. Removal of non-specific protein binding will then allow identification of proteins bound specifically to the differently modified nucleosomes, and consequently insight into the nucleosome's function. Once initial findings have been analysed, more complex *in vitro* studies can be completed with additional modifications or with nucleosomal arrays. Mouse embryonic stem cells (ESCs) were chosen for the experiment as the vast majority of information on bivalency has been obtained with them, helping with the interpretation of results. Once proteins have been identified as binding to bivalent nucleosomes *in vivo* and *in vitro*, further characterisation of their function will be required to identify their role at bivalent domains.

The overview of the complex nature of bivalent domain chromatin signatures given above illustrates the complicated nature of chromatin. As the simplest of bivalent domains are composed of both H3K4me3 and H3K27me3, and no chromatin modification acts in isolation. Each histone modification recruits specific proteins and in turn may affect the recruitment of proteins to nearby modifications and so on, therefore understanding how each protein functions in combination to help interpret the biological function of bivalent domains will be extensively complicated, even if the biological function is simple. Each different modification present at bivalent domains may alter the specific combination of proteins recruited, possibly resulting in slightly different outcomes.

Though multiple studies have been completed on bivalent domains and their coexisting marks, more modifications are being continuously found. Although the presence of additional marks at bivalent domains helps shed light on potential functions none elucidates the universal role of bivalent domains. The purpose of most of the modifications found to co-exist with bivalency are context dependent, for example, H2A.Z deposition is suggestive of both transcriptional activation and repression. This versatility of function depending on the context of deposition is not helpful in the investigation of the general role of bivalent domains. The simplest bivalent state was therefore chosen for analysis (mononucleosomes asymmetrically modified for H3K4me3 and H3K27me3) as it will isolate proteins that prefer to bind to the bivalent asymmetric mark conformation which is present

at all bivalent domains. This will hopefully elucidate a universal role of bivalent domains or a backbone of proteins that are potentially recruited to most bivalent domains.

Chapter 2

Materials and Methods

2.0.1 General Lab Chemicals

β–Mercaptoethanol (Sigma), Bovine Serum Albumin (Sigma), IPTG (Melford Biolaboratories Limited), Bambanker (Lymphotech Ltd), Bradford Reagent (BioRad), Ethylenediaminetetraacetic acid (EDTA) (Sigma), HEPES (Sigma), NaCl (Fisher Chemical), KCl(Sigma Aldrich), Glycine (ITW Reagents), Glycerol (Fisher Chemical), IGEPAL (Sigma), Formaldehyde (Thermo Scientific), Gelatin (Sigma), Tris (Thermo Fisher Scientific), $MgCI_2$ (BDH Laboratory Supplies), Sucrose (Fisher Chemical), Dithiothreitol (DTT) (Melford Biolaboratories Limited), TEMED (Tetramethylethylenediamine) (Sigma), Acrylamide (Severn Biotech ltd), Ammonium persulfate (APS) Acros Organic, Sodium dodecyl sulfate (SDS) (Severn Biotech ltd), DMEM (Gibco), Pyruvate (Gibco), Trypsin (Gibco), Glutamine (Gibco), Penicillin Streptomycin (Pen/Strep) (Gibco), Non-essential amino acids (Gibco), Foetal Bovine Serum (Gibco), Triton-X (Sigma), HCl (Fisher Chemical), N-Lauryl-Sarcosine (Sigma), Imidazole (Sigma), Dimethyl Sulfoxide (DMSO) (Sigma-Aldrich), Stripping buffer (Thermo Scientific), Tween (VWR Life Science), Imperial Protein Stain (Thermo Scientific), NaOH (Fisher Chemical), Agarose (Fisher Chemical), Urea (Sigma), Sodium Azide (Scientific Lab Supplies), KOH (Fisher), Bromophenol Blue (Sigma-Aldrich).

2.0.2 Antibodies

Antibodies for Western botting are listed in Table 2.1 and Table 2.2. While antibodies used for chromatin immunoprecipitation (ChIP) are listed in Table 2.7.

Target	Host	Catalogue Number	Company
H3K27me3	Rabbit	C36B11	Cell Signalling
EZH2	Mouse	612667	BD Transduction Laboratories
SRCAP	Rabbit	ESL104	Kerafast
H3.1	Rabbit	ABE154	Millipore
H3.3	Rabbit	09838	Millipore
Lamin A/C	Rabbit	SC20681	Santa Cruz Biotechnology
H4	Rabbit	D2X4V	Cell Signalling
RING1B	Rabbit	D22F2	Cell Signalling
PHF2	Rabbit	D45A2	Cell Signalling
H3	Rabbit	AB1791	Abcam
H3K4me3	Rabbit	AB8580	Abcam
CBX7	Rabbit	AB21873	Abcam
TAF3	Rabbit	AB188332	Abcam
HA-Tag	Rabbit	AB9110	Abcam
HA-Tag	Mouse	3724S	Cell Signalling
EZH1	Rabbit	ABE281	Abcam
H2A.Z	Rabbit	AB4174	Abcam
ING2	Rabbit	AB109594	Abcam
His-Tag	Mouse	H1029	Sigma
Strep-Tag	Mouse	21507001	IBA
Flag-Tag	Mouse	F3165	Sigma

Table 2.1 *Primary Antibodies used for Western blotting.*

Target	Host	Catalogue Number	Company
Rabbit	Donkey	711-035-152	Jackson Immunoresearch
Mouse	Donkey	715-635-150	Jackson Immunoresearch

Table 2.2 *Secondary Antibodies used for Western blotting.*

2.0.3 Vectors and Primers

Plasmids and primers for cloning and genotyping used are listed in Tables 2.3 and 2.5, respectively. Primers used in ChIP and qPCR experiments are listed in Tables 2.4 and 2.6.

Vector Backbone	Size	Comments
pGEM-3Z	approx 2.7 Kbp	Ampicillin Resistant
p177	approx 2.7 Kbp	Ampicillin Resistant
pET3A	approx 4.7 Kbp	Ampicillin Resistant
pX458	9288 bp	Cas9 from S. pyogenes with 2A-EGFP, and cloning backbone for sgRNA Ampicillin Resistant
pET3b	approx 4.7 Kbp	Ampicillin Resistant

Table 2.3 *Plasmid vectors used for cloning.*

Primer Name	Primer Sequence
HoxC5 forward	GTACTGCTACGGCGGATTGG
HoxC5 reverse	TACCCCGTGGAGAGAGTTGG
Olig1 forward	GGGTTACAGGCAGCCACCTA
Olig1 reverse	ATGCGGTGGAAGAGGATGAG
Polm forward	TGACGGGCACAATTACACCA
Polm reverse	AAAGGCTTCCGCGTCCTAGA
Gapdh forward	GGGTTCCTATAAATACGGACTGC
Gapdh reverse	CTGGCACTGCACAAGAAGATG
Fgf5 forward	GCGACGTTTTCTTCGTCTTC
Fgf5 reverse	ACGAAACCCTACCGGACTCT
Pou5f1 (Oct4) forward	GGCTCTCCAGAGGATGGCTGAG
Pou5f1 (Oct4) reverse	TCGGATGCCCCATCGCA
HoxA7 forward	GAGAGGTGGGCAAAGAGTGG
HoxA7 reverse	CCGACAACCTCATACCTATTCCTG
Fabp7 promoter f	TGAGCAAATCACAAGGAGGA
Fabp7 promoter reverse	TGGAGGAACTCGGGTCTTAC
Bcor CGI forward	GTAAAACCGAAAGCGAGCAA
Bcor CGI reverse	GAGGGTTTCTCCTCCGACTT
Bcor body forward	GGGGGTAACTGTGGGAATCT
Bcor body reverse	TCTCCACACCTTCAGCCTCT
Nanog promoter	TGGCCTTCAGATAGGCTGAT
Nanog promoter	CAAGAAGTCAGAAGGAAGTGAGC
Fzd1 CGI	ACATGAGCCCGTAAACCTTG
Fzd1 CGI	GGTGCCCTCCTACCTCAACT

Table 2.4 *Primers used in ChIP experiments*

Number	Primer Name	Primer Sequence
1	1x 601 209bp forward Biotin	Biotin-GCTTCACCTCGTGACCC
2	1x 601 209bp reverse	CGCTCTAGACCATGATGC
3	H3MT32C forward	ATGCGATCATCATATGTGCGGCGGAGTCAAGAAACC
4	H3MT32C STOP reverse	CTGATCGGTACCCTAAGCCCTCTCGCCTC
5	H3MT32C tagged reverse	CTGATCGGTACCGCCAGCCCTCTCGCCTCGG
6	H3T45C forward	CACCGTTACCGGCCCGGCTGTGTCGCTCTCCGCGAGATC
7	H3T45C reverse	GATCTCGCGGAGAGCGACACAGCCGGGCCGGTAACGGTG
8	Human H2A forward	GCATGATCTACATATGTCCGGCC
9	Human H2A reverse	GATCATGGATCCTTATTTGCC
10	Human H2B forward	GCATGATCTACATATGCCGGAAC
11	Human H2B reverse	GATCATGGATCCTTACTTTGC
12	EZH2 KO exon 3 mut spec forward	CAAGTCATCCCGTAAAACACCTA
13	EZH2 KO exon 3 mut spec reverse	AGGCGAATGCATTTAGGTGTTTT
14	EZH2 KO exon 3 forward	AAACGTGCTACAGATCAGGCT
15	EZH2 KO exon 3 reverse	TGGCCATCAGGCAACTAAGAA
16	EZH1 KO exon 8 forward	GTCCCAGTGCCAGTTGTTCT
17	EZH1 KO exon 8 reverse	GGAGTCTGTTGACAAGCCCT
18	EZH1 KO exon 8 mut spec forward	GATCCTAGTAACAGTTTCCAAACCAT
19	EZH1 KO exon 8 mut spec reverse	CAATGGCGCTGAATATCATATGG
20	KAT6B NT tag FLAG forward	GACTACAAGGACCACGACGG
21	KAT6B NT tag FLAG reverse	CCTTATAATCGCCGTCGTGGT
22	KAT6B NT tag HA forward	ACGATGTGCCAGATTACGCT
23	KAT6B NT tag HA reverse	CAGCGTAATCTGGCACATCG
24	KAT6B NT tag geno reverse	ATCATTACACGGCGGTCCTTT
25	KAT6B NT tag geno forward	AGCCTGTTTGACAGATGTGGT
26	KAT6B KO geno forward	AAGACAACTCTCCACGTTGC
27	KAT6B KO geno reverse	ATATCTCCAGGGCGTCGATAC
28	KAT6B NT KO mut spec reverse	CGTAGATCTTAGCATGGCTATCA
29	KAT6B NT KO mut spec forward	GGTCCTAATGATAGCCATGCTAA
30	MLL2 Y3944A 696 geno forward	GAAGGTAAGTGGCGTGAAGG
31	MLL2 Y3944A 696 geno reverse	CAACCACAGTAATAAGGCACG
32	MLL2 Y3944A 744 geno forward	CTCTGGCATTGTCATTCGCT
33	MLL2 Y3944A 744 geno reverse	TAATAAGGCACGGAAAGACCACC
34	MLL2 Y3944A mut spec forward	TGAGGAGCTGACATATGACGC
35	MLL2 Y3944A mut spec reverse	CATCTTCAATGGGAAACTTGGC
36	KAT6A exon3 KO 753 geno forward	GTTCCCATCCTCCGAGAATCC
37	KAT6A exon 3 KO 753 geno rev	CGTGAACGTGACTCCTATGCT
38	KAT6A exon 3 KO 413 mut spec forward	CTTCCTAAACCTCGGAACCATTGAT
39	KAT6A exon 3 KO 401 mut spec forward	CGGAACCATTGATAATAGGATACTTAGTG
40	KAT6A exon 3 KO 389 mut spec reverse	CGACACTTCACTAAGTATCCTATTA

Table 2.5 *Primers used in cloning and genotyping PCRs*

Primer Name	Primer Sequence
Oct4 forward	AGATCACTCACATCGCCAATCA
Oct4 reverse	CGCCGGTTACAGAACCATACTC
Otx2 forward	CATGATGTCTTATCTAAAGCAACCG
Otx2 reverse	GTCGAGCTGTGCCCTAGTA
Nanog forward	AGGCTTTGGAGACAGTGAGGTG
Nanog reverse	TGGGTAAGGGTGTTCAAGCACT
Fgf5 forward	AAACTCCATGCAAGTGCCAAAT
Fgf5 reverse	TCTCGGCCTGTCTTTTCAGTTC
Mash1 forward	TAACTCCCAACCACTAACAGGC
Mash1 reverse	TGAGGAAAGACATCAACCCAG
GAPDH forward	CATGGCCTTCCGTGTTCCT
GAPDH reverse	GCGGCACGTCAGATCCA
HoxB13 forward	CATTCTGGAAAGCAGCGTTTG
HoxB13 reverse	TGTTGGCTGCATACTCCCG
Pax6 forward	GATTGAGGCTCTGGAGAAAG
Pax6 reverse	CCATTTGGCCCTTCGATTAG
Fabp7 forward	TGTAAGTCTGTGGTTCGGTTG
Fabp7 reverse	AGCAACGATATCCCCAAAGG

Table 2.6 *Primers used in qPCR experiments*

2.1 Cell Culture

Basic Cell Culture Media : DMEM (Gibco 41965-039) supplemented with 15% FBS (Gibco A3160802), 1 mM Sodium Pyruvate (Gibco 11360-039),1% L-glutamine (Gibco 01853), 1% MEM non-essential amino acids (Gibco 11140-035), 0.5% Penicillin Streptomycin (Gibco 15140-122), $0.2\,\text{mM}$ BME (β–me) and $30\,\text{ng/ml}$ heterologously expressed recombinant LIF (made in-house and batch-tested for maintenance of self-renewal)

E14 mouse Embryonic Stem Cells (ESCs) were grown in basic cell culture media. E14 ESCs were thawed at $37\,°\text{C}$ and plated on pre-warmed $10\,\text{cm}$ plates coated with gelatin. After 4 hours media was changed. ESCs were grown on $10\,\text{cm}$ plates coated in gelatin (Sigma G1393), and routinely split 1:15 every two days. During passaging, cells are washed in PBS (Phosphate Buffered Saline) and trypsinised with 0.05% Trypsin (Gibco 25300-062), quenched with equal amounts of media and replated.

Each embryonic stem cell line was cryo-preserved for storage in liquid nitrogen (LN2). These cells were grown to the required number of plates for the required number of vials. The cells are treated as above, once quenched, the cells are counted and resuspended in Bambanker - $500\,\text{µl}$ per $2\,000\,000$ cells - in cryogenic vials (Corning) and placed in freezing containers (Nalgene) in a $-80\,°\text{C}$ before transferring to liquid nitrogen for long time storage.

2.2 DNA Extraction

2.2.1 Cultured Cells

10×10^5 of ESCs were resuspended with $100\,\text{µl}$ of QuickExtract DNA (Epicentre) solution, and extraction followed the manufacturer's instructions.

2.2.2 Agarose Gel Extraction

The desired DNA is run on an 1-2% agarose (manufacturer) gel at 100V in TBE buffer ($100\,\text{mM}$ Tris-HCl, $100\,\text{mM}$ boric acid,$2\,\text{mM}$ EDTA pH8) and stained with

44

SYBER safe (Invitrogen 1:10,000 dilution) to check the identity and purity of the DNA and exercised from the gel. The gel fragment is weighed and then the DNA is extracted with the E.Z.N.A. gel extraction kit (Omega Bio-tek) following manufacturer's instructions.

2.3 Nuclear Extraction

TMSD Buffer : 20 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl$_2$, 250 mM sucrose, 0.5 mM DTT, 0.2 mM PMSF

BC420 Buffer : 20 mM HEPES pH 7.9, 420 mM KCl, 20% (w/v) glycerol, 1.5 mM MgCl$_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF

BC150 Buffer : 20 mM HEPES pH 7.9, 150 mM KCl, 20% (w/v) glycerol, 1.5 mMMgCl$_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF

E14 ES cell nuclear extract (NE) was prepared based on the protocol by (Conaway et al., 1996; LeRoy et al., 1998). E14 ESCs were washed in PBS (Gibco) and harvested by scraping into PBS containing 1 mM EDTA. All subsequent steps were performed at 4 °C. E14 ES cells were collected by centrifugation at 500 g for 8 min and washed with 100 ml of TMSD and collected by centrifugation as before. The cells were resuspended in 10 cell pellet volumes (CPV) TMSD containing 0.01% NP40 and incubated on ice for 5 min to allow for cell lysis to occur. After successful cell lysis was confirmed by staining an aliquot of the cell suspension with Trypan Blue (Gibco), nuclei were pelleted by centrifugation for 10 min at 1000 g. The supernatant (cytosolic fraction) was removed and the nuclei pellet resuspended in 6 nuclei pellet volumes (NPV) of BC150 before slowly adding approximately 3.5 NPV of BC1000 dropwise until the conductivity is equal to 420 mM KCl and incubated at 4 °C with gentle rotation for 1 h. After a 15 s vortex, the suspension was spun for 30 min at 25.000 g. The supernatant was dialysed three times against 50 volumes of BC150 buffer for 1 h. The dialysate was centrifuged at 25.000 g for 20 min. The pellet contains DNA and protein that is insoluble in 150 mM salt. The resultant supernatant is the nuclear extract (NE) and contains proteins normally residing in the nucleus, although proteins strongly bound to the chromatin will not have been extracted. A higher salt extraction or solubilisation of the remaining proteins via the use of ammonium sulphate would help to extract more strongly bound proteins. For all NE used

45

in nucleosome pulldowns, three independent extractions were performed and NE combined to compensate for potential differences in any individual preparation. The resultant NE was flash frozen and stored at $-80\,°C$ until use.

2.4   Purification of endogenous tagged KAT6B

Kat6B was endogenously tagged using CRISPR as described below. Roughly 1.625 billion cells were grown as described above. Cells were washed in PBS (Gibco) and were harvested by scraping cells into PBS containing 1 mM EDTA before collecting to 250 ml centrifuge tubes (Corning). All subsequent steps were performed at $4\,°C$. The cells were collected by centrifugation at 500 g for 8 min. Cell pellets were washed with 10 cell pellet volumes (CPV) of TMSD + 0.1% NP40 and incubated on ice for 5 min to allow for cell lysis to occur. An aliquot of the cell suspension was mixed with Trypan Blue (Life Technologies, T10282) to visually confirm cell lysis. Nuclei were collected by centrifugation for 10 min at 1000 g. The supernatant (cytosolic fraction) was removed and the nuclei pellet resuspended in 6 nuclei pellet volumes (NPV) of BC150 before slowly adding approximately 3.5 NPV of BC1000 dropwise until the conductivity is equal to 420 mM KCl was obtained. This was then incubated at $4\,°C$ with gentle rotation for 1 h. After a 15 s vortex, the suspension was spun for 30 min at 25.000 g. The supernatant was collected and KCl concentration determined as 300 mM by conductivity (Mettler Toledo). ANTI-FLAG® M2 affinity agarose beads (Sigma) were treated with 100 mM glycine, washed with 1 M HEPES pH 7.9, and equilibrated with BC300 buffer. The treated anti-flag beads were added to the supernatant and incubated for 2 h, after which they were collected by centrifugation at 800 g. The anti-flag beads were washed twice with BC300 buffer and then once with BC100 buffer. KAT6B was eluted using 3X FLAG® tag peptide (Sigma) at a concentration of $0.3\,µg\,µL^{-1}$ in 120 µl of BC100 Buffer for 1 h with gentle rotation. No dialysis was performed. The purified KAT6B was frozen with liquid nitrogen and stored at $-80\,°C$.

TMSD Buffer : 20 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 250 mM sucrose, 0.5 mM DTT, 0.2 mM PMSF

BC420 Buffer : 20 mM HEPES pH 7.9, 420 mM KCl, 20% (w/v) glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF

46

BC300 Buffer : 20 mM HEPES pH 7.9, 300 mM KCl, 20% (w/v) glycerol, 1.5 mM MgCl$_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF

BC100 Buffer : 20 mM HEPES pH 7.9, 100 mM KCl, 20% (w/v) glycerol, 1.5 mM MgCl$_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF

2.5   Histone Acetyltransferase Assays

Recombinant nucleosomes were reconstituted as described below using plasmid p177 DNA (see Fig. 2.1). Nucleosomes and KAT6B enzyme were combined in HAT buffer. Reactions were initiated by the addition of 3H-acetyl coA (90 kBq; Hartmann) and the reactions incubated at 30 °C for 2 h. Reactions were stopped by boiling for 5 min at 95 °C with 3x SDS sample buffer before running on SDS page gel, transferred to PVDF membrane and stained with coomassie. Activity was imaged using Carestream® Kodak® BioMax® MS film (Sigma) and the signal was amplified using Carestream® BioMax® Transcreen LE (Sigma). The films were incubated with the PVDF membrane and transcreen at −80 °C for 24 h to 72 h, dependent upon signal strength, before developing.

HAT Buffer : 250 mM Tris-Cl pH 7.5, 50% (w/v) glycerol, 0.5 mM EDTA, 200 mM DTT
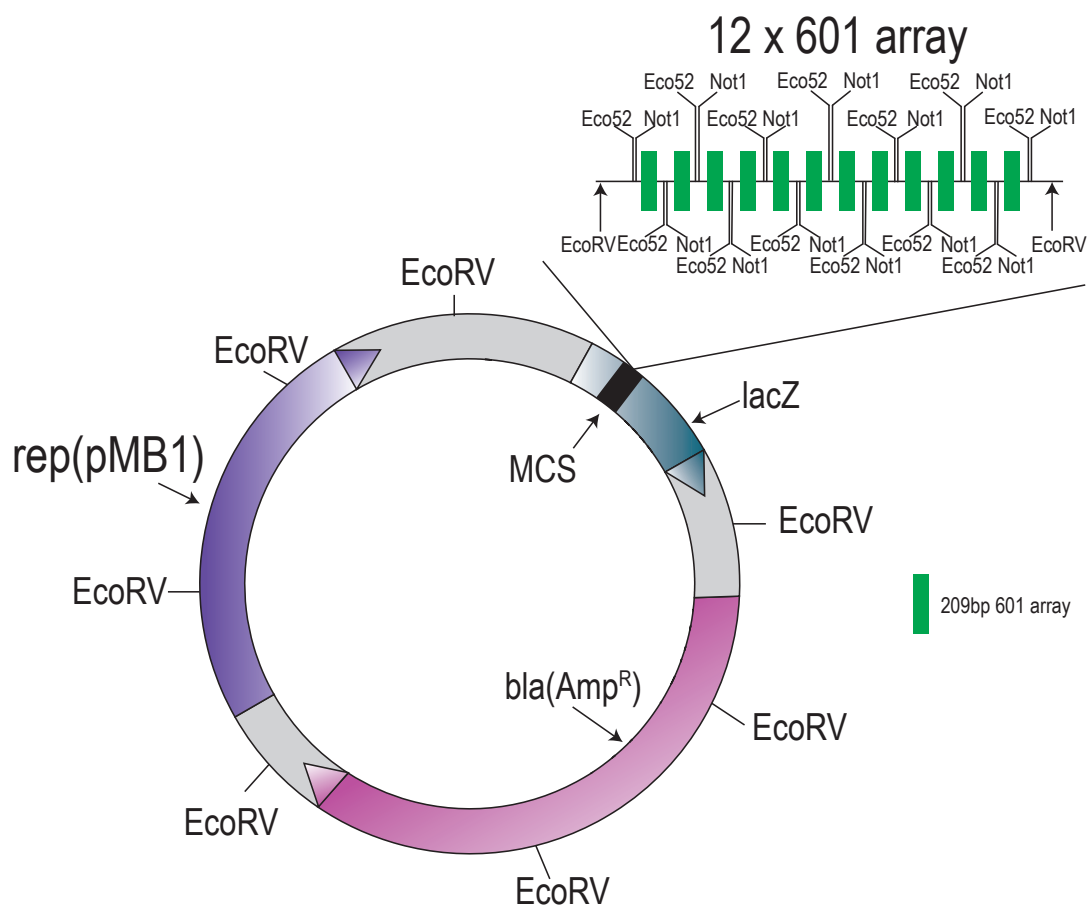
Figure 2.1 *p177 Plasmid Map - Plasmid map of p177. The plasmid contains the pMB1 replicon rep, responsible for the replication of the plasmid, while the bla gene encodes for beta-lactamase that confers resistance to amplicillin. An array of 12 X 601 209bp sequences are cloned into the multiple cloning site (MCS) located in the lacz part of the lac operon from E.coli. Each 601 sequence is separated by two enzyme restriction sites Eco52 and Not1 allowing the digestion of the 12X array into individual 209bp sequences. The 601 array can be separated from the vector by use of EcoRV enzyme restriction sites, spaced approximately 200-400bp around the vector backbone.*

2.6 CRISPR

CRISPR is a technique that allows introduction of precise double strand breaks of DNA at a specific genomic location (Cong et al., 2013; Gasiunas et al., 2012; Martin Jinek et al., 2012; Sander and Joung, 2014). It is highly efficient and has a relatively low number of off-target effects. The Cas9 nuclease combined with a guide RNA, containing both crRNA and tracrRNA, allows targeting of the Cas9 nuclease to cleave the complementary target-DNA sequence if adjacent to a PAM (protospacer adjacent motif). DNA repair by NHEJ (non-homologous end joining) lends itself to creation of insertion/deletions or indels, which can be used to create knockout of genes, while HDR (homology directed repair) can generate specific modifications at the target locus in the presence of a repair template. Both strategies were used in creation of the desired specific cell lines.

To perform genome editing in E14 ES cells, sgRNAs specific to the gene of interest were cloned into pX458 (addgene 48138, (Ran et al., 2013)). For ease of description an example of CRISPR editing will be described (EZH2 KO), for further details about CRISPR strategy and primers used in specific cases refer to the Table 2.5 and the Figure 6.1. For homology repair, single stranded oligonucleotides were designed encoding the desired mutations flanked by homology regions (synthesized by IDT, E14 mESCs were transfected with pX458 plasmid (Ran et al., 2013) encoding a guide RNA targeting exon 7 of the mouse Ezh2 gene (20-bp target sequence CAGCAGGAAATTTCCGAGGT), along with a single-stranded DNA oligonucleotide for homology-directed repair to introduce three consecutive stop codons at the 5' end of exon 7 (resulting sequence ATtAAtAAgCTTGatCACCTC, mutated bases in lower case). After fluorescence-based sorting for GFP-positive transfected single cells, single cell colonies were expanded, genotyped, and analysed for target expression by Western Blot. Correct genotypes were confirmed by Sanger sequencing on PCR-amplified genomic material.

2.6.1 Transfection and Flow Activated Cell Sorting (FACS)

Low passage number ESCs were used for CRISPR/gene editing. 400 000 cells were plated in non-antibiotic containing media. 5 µl of Lipofectamine was added to 100 µl of OptiMEM while 1 µg of pX458 plasmid containing the desired construct

was combined with 2 µl of 10 µM of single stranded oligodeoxoynucleotide (ssODN) donor template in 100 µl OptiMEM. These solutions were combined, incubated for 30 minutes and added to the freshly plated cells. The media was changed 24 hours after transfection.

Forty-eight hours after transfection, the cells were trypsinised and passed through a 70 µm cell strainer into FACS sorting media (basic cell culture media supplemented with 0.1% FBS, 25 mM HEPEs, and 1.5% Pen/Strep). These cells underwent FACS in FACS collection media (basic cell culture media supplemented with 20% (v/v) FBS, 1.5% Pen/Strep and 25 mM HEPEs) with an argon laser 488 nm, 15 000 GFP+ cells were sorted for plating on a 15 cm in cell culture media supplemented with 1.5% Pen/Step and 100 000 cells were sorted to confirm/check transfection and PCR screening. GFP is co-expressed with Cas9 from pX458 as a 2A fusion.

One week after plating single cell colonies were picked manually by dislodging and aspirating colonies, separated by trypsin and plated on duplicate 96-well plates. Allowing one plate for genotyping and one for maintenance of cell lines. After 2 days to 3 days growth single cell colonies from the genotype plate were screened by PCR and sequenced before verification by Western blot. Initial PCR verification uses primers one of which anneals to the genomic sequence adjacent to the targeted sequence and one which recognised the homology repair template see Figure 6.1. PCR positive cell lines were sequenced across the entire target sequence to check for indels, missense mutations and correct homology directed recombination. The protein expression levels of cell lines with the correct repaired DNA sequence were confirmed by western blot analysis. All verified cell lines were frozen and kept in liquid nitrogen storage.

2.7   Western Blot

Protein samples were prepared by boiling protein at an appropriate concentration at 95 °C for 5 min in SDS sample buffer before loading on an 1.5 mm acrylamide gel of an appropriate percentage. Gels were run at 120 V through the stacking gel and then 200 V for the separation gel. SDS-PAGE gels were transferred to nitrocellulose membrane (Biorad), via the semi-dry Trans-blot turbo transfer system (Biorad), and blocked with 5% milk in TBST for 1 h before incubation with primary antibody in 2% BSA in TBST overnight. Membranes were then

washed 3 x 5 min in TBST and incubated with 1:5000 dilution of HRP conjugated secondary antibody in 2% BSA in TBST for one hour. Membranes were washed 3 x 5 min in TBST and developed with Clarity ECL reagents (Biorad) and imaged with Chemidoc MP imaging system (Biorad).

TBST : 20 mM TRIS-HCl pH 7.6, 137 mM NaCI, 0.1% Tween 20

3XSDS Sample buffer: 190 mM Tris-HCl, 30% (w/v) glycerol, 6%SDS, 150 mM DTT, 0.3% bromophenol blue.

2.8 Chromatin Immunoprecipitation (ChIP)

Fixation Buffer : 10 mM HEPES pH 7.6, 15 mM NaCl, 1% formaldehyde, 0.15 mM EDTA, 0.075 mM EGTA in DMEM (Gibco 41965-039)

Lysis Buffer 1 : 50 mM HEPES pH 7.6, 140 mM NaCl, 1 mM EDTA, 10% (w/v) glycerol, 0.5% (v/v)Igepal, 0.25% (v/v) Triton-X

Lysis Buffer 2 : 10 mM Tris-HCl pH 8.0,200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA

Lysis Buffer 3 : 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 0.5% N-lauryl sarcosine

TE Buffer : 10 mM Tris pH8, 1 mM EDTA

2XIP Buffer : 300 mM NaCI, 2%(v/v) Triton X-100 in TE buffer

RIPA Buffer : 10 mM HEPEs pH 7.6, 1 mM EDTA, 500 mM LiCI, 1% (v/v) NP-40, 0.1% N-lauroyl Sarcosine

TEN Buffer : 10 mM NaCl in TE buffer.

Elution Buffer : 100 mM NaHCO$_3$, 1% SDS, 200 mM NaCI,0.4 µg µg^{-1} Proteinase K, 0.4 µg µg^{-1} RNAse A

2.8.1 Chromatin Preparation

For chromatin immunoprecipitation cells were grown as described above, approximately 130 million of cells were grown on four 15 cm plates coated with 0.1% (v/v)

gelatin. Media was removed and 15 ml of Fixation Buffer was added per plate for 10 min at RT (room temperature 25 °C) with gentle rocking. Fixation was stopped by quenching with 2.5 M glycine at 0.125 M final concentration for 3 min at RT with gentle rocking. Fixation buffer was removed, cells were washed with 10 mL cold PBS, and 3 mL more cold PBS was added. Cells were scraped off the plate and centrifuged at 2500 g for 5 min at 4 °C and the supernatant removed. Cell pellets were weighed, snap frozen in LN2 and stored at −80 °C. All subsequent steps were performed at 4 °C. The cell pellets were thawed and resuspended in Lysis Buffer 1 to obtain a suspension of 75 µg ml^{-1} and placed on a rotator for 10 min. Cells were centrifuged at 3000 g for 5 min and the supernatant removed before resuspending in Lysis Buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) to the same volume as used for Lysis Buffer 1. These samples were rotated for 10 min and then centrifuged at 3000 g for 5 min before removing the supernatant. Pellets were resuspended in Lysis Buffer 2 to obtain 75 µg ml^{-1}. The suspension was transferred to polystyrene tubes (BD Falcon, cat: 352099) and sonicated using a Bioruptor (Diagenode) at maximal output using 30 s ON, 30 s OFF cycles. Number of cycles was experimentally determined to yield 1.5 kb sized fragments of cross-linked DNA - approximately 250-300 bp when decrosslinked. The size of the DNA fragment was confirmed by 1% agarose gel analysis. A sample was decrosslinked overnight, purified and analysed on high sensitivity DNA chip in a 2100 Agilent bioanalyser - as below. Sonicated suspensions were centrifuged at 20.000 g for 10 min. DNA concentration was determined by Nanodrop ND-1000 (Thermo Fisher Scientific) and aliquots were snap frozen in LN.

2.8.2 Immunoprecipitation

Crosslinked chromatin was thawed and spun at 25.000 g for 10 min. The thawed chromatin is combined with an equal volume of 2XIP buffer, 10% by volume of each ChIP is removed for an input sample and stored at −20 °C. The desired antibody is added, mixed and incubated at 4 °C rotating overnight. Protein A (10002D) and Protein G (10004D) beads are prepared separately and all spins are at 800 g for 30 s at 4 °C. Beads are washed 2x in 1 ml PBS (supplemented with 0.5% (w/v) BSA) and then blocked with 1 ml PBS (supplemented with 0.5% (w/v) BSA) overnight.

The blocked beads are washed five times with 1 mL TE buffer and resuspended in

TE buffer. The ChIP samples are spun at full speed at 10 min and the supernatant of sample is added to 50 µl of bead suspension. The samples are incubated on the rotatory machine for 2 h to 4 h. Beads are washed 5 X with 1 ml of ChIP RIPA buffer for 2 min. Beads are transferred to a new eppendorf tube and washed with TEN buffer. 200 µl of elution buffer is added to the beads and thawed input sample. These are incubated in a thermocycler at 800 rpm at 37 °C for 1 h and left at 65 °C overnight to decrosslink.

Samples are collected by centrifugation and the supernatants were used for subsequent steps. The decrosslinked DNA is purified with NEB monarch Kit following manufacturer's instructions and eluted in 50 µl of elution buffer. The ChIP samples were analysed by qPCR and excess eluted DNA was stored at −20 °C.

2.8.3 qPCR

Eluted DNA was diluted 1:30 before use in qPCR experiments. qPCR experiments were carried out on a Lightycler 480 (Roche) using a 10 µl reaction volume with 2 µl of DNA, 5 µM primers and 1x SyGreen Blue Mix (PCRBiosystems). Enrichments were calculated as percentage of Input.

Antibody	Company	Catalogue No.	$\mu l\,\mu g^{-1}$(antibody/chromatin)
H3K4me3	CST	C42D8	4 µL/40 µg
H3K27me3	CST	C36B11	4 µL/40 µg
H3K23ac	Abcam	AB177275	2 µL/40 µg
H3K14ac	Abcam	AB52946	4 µL/40 µg
HA-tag	CST	C3724	4 µL/200 µg

Table 2.7 *Antibodies used in ChIP Experiments*

2.9 Expression and purification of histones

Xenopus H3 and H4 and human H2A and H2B were expressed from pET-3a or pET-3d vectors in BL21 (DE3) pLysS for H3, H2A, and H2B or BL21 (DE3) for H4 through induction with 0.2 mM IPTG for 4 h at 37 °C. Histones were purified from inclusion bodies and solubilised in unfolding buffer. Extracted histones were dialysed against three changes of urea dialysis buffer, this and all subsequent histone dialysis steps were carried out at 4 °C, and then purified further by

passing over a HiTrap Q column (GE Healthcare) before binding to and NaCl gradient elution from HiTrap SP cation exchange chromatography columns (GE Healthcare). The progression and quality of the purification was continuously evaluated by measuring the absorbance of the solution at 280 nm (A_{280}) and 260 nm (A_{260}) as demonstrated in the chromatogram (see Fig. 3.6). The sample's A_{280} can be used in conjunction with Beer's law and the known histone sequence to calculate the protein concentration, this was combined with Bradford assays to estimate protein concentrations. Fractions containing histones were pooled and dialysed three times against water containing 5 mM β–me and lyophilised for long-term storage at $-80\,°$C.

To express histones for native chemical ligation (NCL), constructs encoding truncated *Xenopus* histone H3 were generated in pET-3a. For generation of H3K4me3-modified histones, truncated H3 lacking residues 1-31 after the initiator methionine, with a threonine-to-cysteine substitution at position 32 of *Xenopus* H3 and a cysteine-to-alanine substitution at position 110 (H3Δ1–31 MT32C C110A) was expressed in BL21 (DE3) pLysS and purified as above, except for the final dialysis, which was carried out as two rounds of dialysis against 1 mM DTT in H2O and one round against 0.5 mM TCEP before lyophilisation and storage. For generation of H3K27me3-modified histones, a similarly truncated *Xenopus* H3 construct was used, either lacking the first 44 or the first 32 residues and carrying a threonine-to-cysteine mutation at residue 45 (H3Δ1–45 MT45C C110A or H3Δ1–32 MT32C C110A) For the mass spectrometry experiments H3Δ1–32 MT32C C110A was used, for all other experiments including nucleosome titrations H3Δ1–45 MT45C C110A was used .

Urea Dialysis Buffer : 10 mM Tris HCl pH 8, 7 M urea, 100 mM NaCl, 1 mM EDTA, 5 mM β–me

2.9.1 Generation of Modified Histones by Native Chemical Ligation (NCL)

8 mg of purified truncated histone is dissolved in 150 µl of Methoxyamine buffer and incubated at 25 °C ovenight which helps resolve any N-terminal adducts that may have formed during its expression in *E.coli*, masking the N-terminal cysteine and preventing the ligation reaction. The treated truncated histone is dialysed against the NCL buffer for 1 hour at 4 °C. After dialysis 6 µL of 0.5 M TCEP is

added and 5 M NaOH added to keep the pH of the solution at pH 6.5. 2 mg of the modified peptide is resuspended in 100 µL of NCL buffer containing 200 mM mercaptophenylacetic acid (MPPA) and combined with the dialysed histone. This is incubated for 72 h at room temperature with constant agitation with 3 µl of 0.5 M TCEP being added after 24 h. Reactions were then dialysed three times against urea dialysis buffer (see above, but with 1 mM DTT instead of 5 mM beta-mercaptoethanol). Ligated full-length modified histones were separated from unligated histone through cation exchange chromatography on a HiTrap SP column (GE) and then dialysed against three changes of water containing 5 mM beta-mercaptoethanol before lyophilization and storage at −80 °C until use. For H3K4me3-modified histone, H3Δ1–31 MT32C C110A was reacted with a synthetic peptide spanning residues 1–31 of histone H3.1 containing trimethylated lysine at position 4 and a C-terminal benzyl thioester (Peptide Protein Research Ltd., Fareham, UK). For H3K27me3-modified histones, H3Δ1–44 MT45C C110A was reacted with a synthetic peptide spanning H3.1 residues 1–44 including trimethylated lysine at position 27 and a C-terminal benzyl thioester.

Methyoxyamine Buffer : 6 M Guanidine HCl, 200 mM Methoxyamine, 20 mM TCEP.

Native Chemical Ligation (NCL) Buffer : 6 M Guanidine HCl, 250 mM sodium phosphate buffer pH 7.2, 150 mM 4-mercaptophenylacetic acid (MPAA), 50 mM TCEP

2.9.2 Reconstitution of Recombinant Nucleosomes

To reconstitute histone octamers, the four core histones were resuspended in unfolding buffer (see above), mixed in a mass ratio of 1:1:1.2:1.2 (H3:H4:H2A:H2B), and dialysed against three changes of refolding buffer at 4 °C. After centrifugation to remove precipitate formed during dialysis, correctly assembled histone octamers were purified by size exclusion chromatography in refolding buffer on a S200 column (GE Healthcare), using an Akta PURE system (GE Healthcare).

Asymmetrically modified octamers required further purification; the different copies of H3 incorporated into asymmetric octamers were differentially tagged with either His or Strep tag. Affinity chromatography was performed on the combined fractions from the S200 containing the required asymmetric protein complexes by first using a Ni-sepharose column (GE) and then a Streptactin

column (GE). Proteins were eluted from the Ni-sepharose column with a gradient (0 mM to 250 mM) of imidazole in elution buffer (2 M NaCl, 25 mM Tris-HCl pH 8). Elution fractions containing His/Strep- and His/His-tagged histone H3-H4 tetramers were combined and loaded onto a HiTrap Streptavidin HP column (GE). Proteins were eluted in elution buffer containing 2.5 mM desthiobiotin. After each chromatography step, elution profiles and sample purity were confirmed by Western blot.

Refolding Buffer : 10 mM Tris HCl pH 8, 2 M NaCl, 1 mM EDTA, 5 mM β–me.

Unfolding Buffer : 20 mM Tris-HCl pH 7.5, 7 M guanidine HCl, 10 mM DTT

2.9.3 Generation of 601 DNA

DNA template for mononucleosome assembly was generated by PCR with a biotinylated forward primer, amplifying a 209-bp fragment centered around the 147-bp 601 nucleosome positioning sequence, followed by PCR purification and elution into TE buffer. To reconstitute recombinant mononucleosomes, DNA and histone octamers were combined in refolding buffer supplemented with 5 M NaCl to compensate for reduction in NaCl concentration due to introduction of TE buffer with the DNA, followed by gradient dialysis against TE buffer down to 400 mM NaCl and then a step dialysis against TE buffer at 4 °C. Optimal ratios of DNA and histone octamer were determined so that at least 95% of DNA was complexed, but without over-assembly and unspecific DNA binding of histones. Assemblies were routinely checked by native gel electrophoresis on 6% acrylamide gels in TGE buffer.

TGE buffer: 12.5 mMTris HCl pH8 , 96 mM Glycine, 0.5 mM

2.10 Pulldowns

For pulldown assays with recombinant modified nucleosomes and E14 ES cell NE, streptavidin sepharose high performance beads (GE Healthcare, 10.5µl of slurry per pulldown) were briefly washed three times with pulldown buffer. All centrifugation steps were carried out at 1500 g for 2 min at 4 °C. All incubation steps were carried out at 4 °C. After washes, beads were incubated overnight with 10.5 µg of assembled recombinant nucleosomes (in TE buffer diluted with

pulldown buffer and adjusted to a final concentration of 0.1% NP-40 and 150 mM NaCl by addition of 10% NP-40 and 5 M NaCl, respectively) with constant rotation in the cold room. For pulldown titrations with increasing amounts of recombinant nucleosomes, nucleosomes were assembled in 4 batches of 10 μg each, combined, and incubated with beads in the amounts given in the Figure 5.4 and Figure 5.5. The amount of beads used was scaled with the amount of nucleosomes (1 μl slurry per 1 μg of nucleosomes). Beads were then collected by centrifugation and washed briefly with three changes of pulldown buffer. Bead-bound nucleosomes were incubated with 500 μg of NE for 2 h under constant agitation in the cold room. Beads were then washed with pulldown buffer by 5 min incubations under rotation first with two washes of pulldown buffer containing NP-40 followed by three washes of pulldown buffer without NP-40. After washes, bound proteins were eluted from the beads.

For Western Blot analysis, elution was performed by boiling for 5 min at 95 °C with 1.5x SDS sample buffer (95 mM Tris HCl pH 6.8, 15% (w/v) glycerol, 3% SDS, 75 mM DTT, 0.15% bromophenol blue). 30% of bound sample was loaded for western blot analysis.

Pulldown Buffer : 20 mM HEPES KOH pH 7.9, 150 mM NaCl, 10% (w/v) glycerol, 1 mM EDTA, 1 mM DTT, 0.2 mM PMSF, 0.1% (v/v) NP-40

2.11   Mass Spectrometry Analysis and Sample Preparation

2.11.1   Preparation of Protein Samples for Mass Spectrometry

For analysis of pulldown samples by LC-MS/MS, beads were resuspended in elution buffer (2 M Urea, 100 mM Tris pH7.5, 10 mM DTT) and incubated for 20 min on a shaker. All shaker steps were carried out at 1,000 rpm at RT (25 °C). After incubation Iodoacetamide was added to a final concentration of 55 mM and incubated for 10 min on the shaker. The eluted proteins were digested with 0.3 μg of trypsin for 2 h whilst shaking and then spun at 1500 g for 2 min. The supernatant was removed and 50 μl of elution buffer incubated with the beads for 5 min on the shaker. The beads were spun for 2 min at 1500 g and the two

supernatant elutions combined. 0.15 µg of trypsin were added and the solution was digested overnight at RT before digestion was stopped by the addition of 10% TFA. Following digestion, half of each sample was diluted with an equal volume of 0.1% TFA and spun onto StageTips as described by (Rappsilber et al., 2003). Peptides were eluted in 40 µl of 80% acetonitrile in 0.1% TFA and concentrated down to 1 µl by vacuum centrifugation (Concentrator 5301, Eppendorf, UK). Samples were then prepared for LC-MS/MS analysis by diluting them to 6 µl with 0.1% TFA.

Purified KAT6B complexes were boiled at 95 °C in 4X NuPAGE LDS Sample buffer (Invitrogen) and 5% of 1 M DTT for 5 min before running on a NuPAGE 4-12% Bis-Tris Protein Gel (Invitrogen) at 200 V with MOPS buffer (Invitrogen) for 5 min. Gels were washed three times with deionized water for 5 min and visualised using Imperial protein stain (Sigma-Aldrich, UK) for 1 h. Gels were washed again three times with deionized water for 5 min each to remove excess stain.

Gel bands were excised and de-stained with 50mM ammonium bicarbonate (Sigma Aldrich, UK) and 100% (v/v) acetonitrile (Sigma Aldrich, UK) and proteins were digested with trypsin, as previously described (Shevchenko et al., 1996). Briefly, proteins were reduced in 10 mM dithiothreitol (Sigma Aldrich, UK) for 30 min at 37 °C and alkylated in 55 mM iodoacetamide (Sigma Aldrich, UK) for 20 min at ambient temperature in the dark. They were then digested overnight at 37 °C with 12.5 ng µl^{-1} L-1 trypsin (Pierce, UK).

2.11.2 Mass Spectrometry Parameters and Analysis

LC-MS-analyses were performed on a Q Exactive mass spectrometer (Thermo Fisher Scientific, UK) (for samples that were digested on beads) and on Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific, UK) (for samples that were subjected to in-gel digestion) both coupled on-line to Ultimate 3000 RSLCnano Systems (Dionex, Thermo Fisher Scientific, UK). Peptides were separated on a 50 cm EASY-Spray column (Thermo Scientific, UK) assembled in an EASY-Spray source (Thermo Scientific, UK) and operated at 50 °C. In both cases, mobile phase A consisted of 0.1% formic acid in water while mobile phase B consisted of 80% acetonitrile and 0.1% formic acid. Peptides were loaded onto the column at a flow rate of 0.3 µl/ min and eluted at a flow rate of 0.25 µl/ min according to the following gradient: 2 to 40% buffer B in 150 min,

then to 95% in 11 min. For Q Exactive, FTMS spectra were recorded at 70,000 resolution (scan range 350-1400 m/z) and the ten most intense peaks with charge *geq*2 of the MS scan were selected with an isolation window of 2.0 Thomson for MS^2 (filling 1.0E6 ions for MS scan, 5.0E4 ions for MS^2, maximum fill time 60 ms, dynamic exclusion for 50 s). For Orbitrap Fusion™ Lumos™, survey scans were performed at 120,000 resolution (scan range 350-1400 m/z) with an ion target of 4.0e5. MS^2 was performed in the ion trap at a rapid scan rate with ion target of 2.0E4 and HCD fragmentation (Olsen et al., 2007) with normalized collision energy of 27. The isolation window in the quadrupole was 1.4. Only ions with charge between 2 and 7 were selected for MS^2.

2.11.3   Visualisation of Mass Spectrometry Data

The MaxQuant software platform (Cox and Mann, 2008) version 1.6.1.0 was used to process raw files and searches were conducted against the Mus musculus reference proteome (Uniprot, released in July, 2017), using the Andromeda search engine (Cox et al., 2011). The first search peptide tolerance was set to 20 ppm while the main search peptide tolerance was set to 4.5 pm. Isotope mass tolerance was 2 ppm and maximum charge to 7. A maximum of two missed cleavages was allowed. Carbamidomethylation of cysteine was set as fixed modification. Oxidation of methionine and acetylation of the N-terminal were set as variable modifications. Label-free quantitation analysis was performed by employing the MaxLFQ algorithm as described by Cox et al., 2014. Peptide and protein identifications were filtered to 1% FDR.

Data was analysed and visualised in Perseus (version 1.6.5.0) (Tyanova et al., 2016). Data was filtered to remove proteins identified with less than 3 peptides, false positives and potential contaminants before missing values were filtered to retain only proteins quantified in at least two replicates across all nucleosome pulldown conditions and replicates analysed. Proteins were considered enriched in a condition if their Log2 fold change was greater than 1.5 – approximately a 3 fold enrichment - and their p-value was lower than 0.01.

Data was analysed and visualised in R version 3.5.3. DEP 1.4.1 package (Zhang et al., 2018) was used to determine differential enrichment of proteins between nucleosome pulldowns. Data was filtered on reverse hits and potential contaminants, before missing values were filtered to retain only proteins quantified in at least two replicates across all nucleosome pulldown conditions and replicates analysed.

The data was normalized using vsn, and imputation was carried out using the MinProb function in DEP. Proteins were considered differentially enriched when the log2 fold change was greater than 1.5, and the p-value was lower than 0.01. Single protein bar plots were generated in DEP using the Shiny package. Volcano plots were generated in ggplot2. Proteins that were significantly altered (FDR=0.01, pvalue=0.05 and fold change $\geq$ 3) bound to the bivalent nucleosome in comparison to one of the other modified nucleosomes were visualised via a heatmap. Hierarchical clustering was performed by the Multiple Experiment Viewer tool (MeV), the Pearson correlation metric and average linkage clustering as linkage method.

2.12   Neuronal Differentiation Assays

Neuronal differentiation was performed following the protocol described by (Bibel et al., 2007). Embryoid body formation was induced by removal of LIF, on day 0 ES cells were plated onto non-adherent 10 cm bacterial plates (Greiner) in 10 ml EB medium (basic cell culture media without LIF and with FBS reduced to 10%) at a density of 4×10^6 cells per 10-cm plate. Embryoid bodies began to form after one day. On day 2 the media was changed by gently transferring to a 15 ml falcon tube, allowing EBs to settle for 5 min then aspirating the supernatant. 13 ml of fresh media was then added to EBs before gently resuspending and transferring them to a fresh plate using a 5 ml serological pipette (Sarstedt). On day 4 media was changed and increased to 15 ml of EB medium containing 10 M all-trans retinoic acid (Sigma). On day 6, media was changed again maintaining the same volume of EB medium + retinoic acid. On day 8 the EBs were transferred to a 50 ml falcon tube (pooling 2 plates per tube) and EBs were allowed to settle for 5 min before aspirating the supernatant. EBs were then washed 3 x in 20 ml PBS allowing EBs to settle between each wash. Trypsin (Sigma) was prepared at 0.5 µg µl^{-1} in 0.05% v/v EDTA/PBS. After the final wash EBs were trypsinised in 1 ml trypsin per 50 ml falcon, for approximately 3 min at 37 °C with agitation and then quenched in 10 ml EB medium. Cells were then centrifuged for 5 min at 300 g, resuspended in 10 ml EB medium and passed through a 40 µm cell strainer to remove clumped cells and undigested EBs. Cells were then counted, centrifuged again at 300 g for 5 min and resuspended in Advanced DMEM/F12 (Gibco) with 1x N2 supplement (Gibco) before being plated onto poly-D-lysine (Sigma) and laminin (Sigma) coated 6 cm dishes at a density of 3x106 cells/plate. On day 9

half the medium was replaced with Neurobasal medium (Gibco) containing 1x
B27 supplement (Gibco), this was repeated on day 12.

2.13 RNA preparation and qRT-PCR

BTE Buffer : 10 mM bis-tris pH6.7, 1 mM EDTA

Cells were resuspended in 1ml TriPure isolation reagent (Sigma) per 10 cm plate
(or per 2 x 6 cm plate for plated NPCs) and RNA extracted by isopropanol
precipitation. The resulting pellet was resuspended in 44 µl BTE buffer before
treating with Turbo DNAse (Thermo Fisher) for 30 min at 37 °C. 1 ml TriPure
was then added and isopropanol precipitation repeated before resuspending the
final RNA pellet in 50 µl BTE. Concentration was then measured on a Nanodrop
and the quality of RNA samples further assessed using RNA 6000 Nano chip on
an Agilent Bioanalyser. Samples were stored at −80 °C.

cDNA samples were prepared from starting material of 1 µg RNA to which was
added 2.5 µM oligodT and 0.5 mM dNTP mix. Samples were incubated at 65 °C
for 5 min to anneal primers to the RNA then returned to ice. To the annealed
RNA/primer solution was added 1 µL SuperScript IV reverse transcriptase and
4 µL 5X Superscript IV buffer with RNAase inhibitor and 100 µM DTT. Samples
were transferred to thermocycler for 10 min each at 42 °C, 50 °C, 55 °C and 80 °C
in succession. The resulting cDNA samples were then diluted at least 1/30 in
water before being used in qPCR experiments and stored at −20 °C.

2.13.1 qRT-PCR

qRT-PCR experiments were carried out on a Lightycler 480 (Roche) using a
10 µl reaction volume with 2 µl of DNA, 5 µM primers and 1x SyGreen Blue Mix
(PCRBiosystems). Results were analysed using the 2-DDCT method with Gapdh
as the reference gene.

2.14 Immunofluorescence Analysis

Cells were grown on 16 mm glass coverslips in 12 well plates. In preparation for plating coverslips were soaked overnight in ethanol before drying and washing twice in PBS before coating. Coverslips for mESCs were coated with 0.1% gelatin, coverslips for neuronal cells were coated with PDL and lamin. Before fixation, cells were washed 3 x 10 min in PBS, and then fixed with 4% PFA for 20 min. To carry out immunofluorescence, the cells fixed on coverslips were blocked in 10% donkey serum (Sigma) in 0.1% Triton X-100 for 1 h and incubated o/n with primary antibodies at 4 °C. The following day, cells were incubated with Alexa fluor secondary antibodies (Invitrogen) at a dilution of 1:1000 in 1% donkey serum in 0.1% Triton X-100 for 1 h at room temperature in the dark. The cells were then washed 3x with PBS for 10 min. Nuclei were counterstained with 50 nM DAPI for 5 min, and washed 2x 10 min PBS before being mounted on slides using vectashield. Imaging was carried out using a Zeiss Axio imager with 40x objective.

Chapter 3

Expression and Purification of Modified Histones and Reconstitution of Nucleosomes

3.1 Introduction

The aim of this study is to elucidate the biological function of bivalent domains. We hypothesise that the proteins that bind to bivalent domains mediate its function. Identification of these proteins will help to elucidate the biological role of bivalency. An *in vitro* approach was chosen to investigate bivalent domain function due to its ability to remove the complicating factors present *in vivo*. In order to investigate which proteins bind to bivalent domains *in vitro*, a reliable method of creating purified specifically modified nucleosomes is required. Once this has been established, proteins binding bivalently modified nucleosomes can be compared against those bound to differently modified (control) nucleosomes - detailed in later chapters. Thus the function of each of these modified nucleosomes can been seen by virtue of the proteins that bind them; therefore insight can be gained about bivalent domain function.
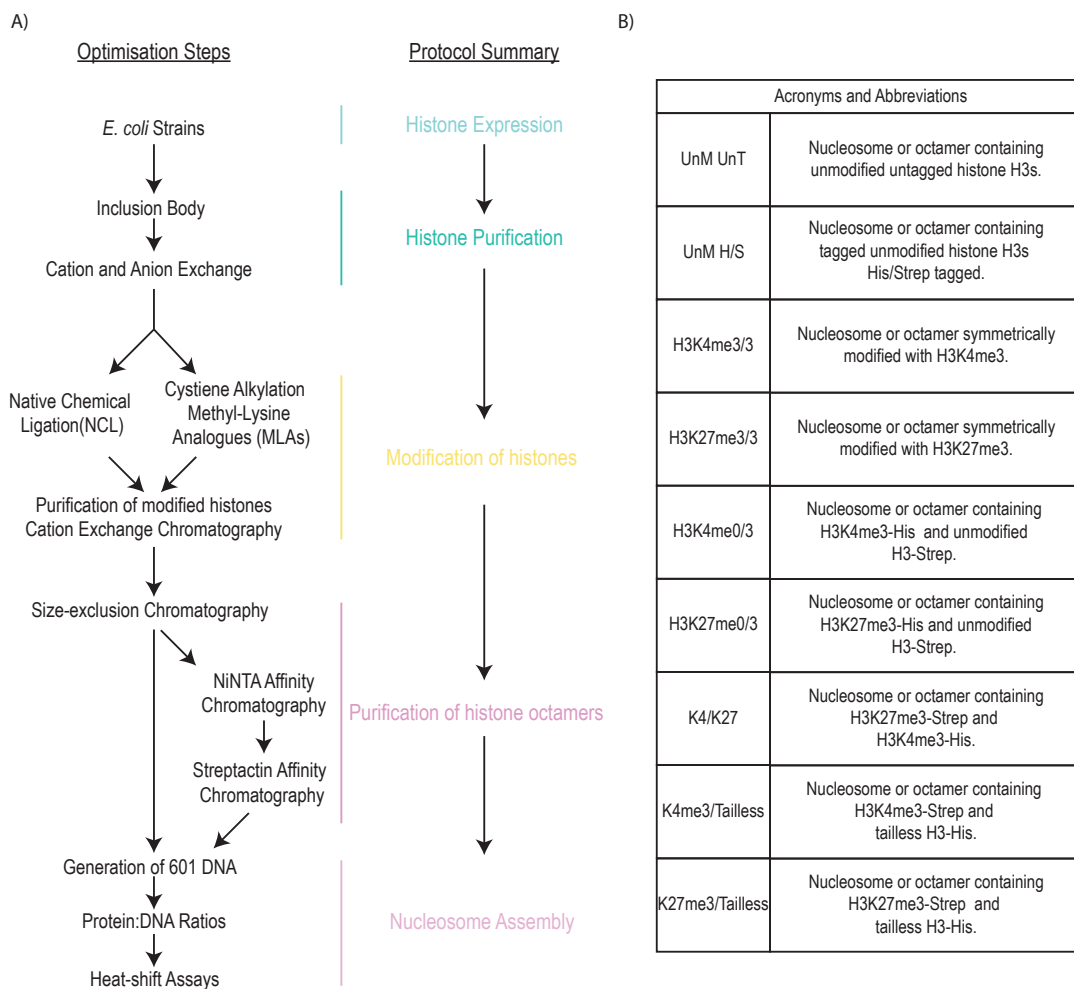
Figure 3.1 Steps Required for Histone Purification and Nucleosome Assembly
A) Flowchart giving an overview of the experimental steps required to purify histones and create specifically modified nucleosomes. The potential optimisation steps for each experimental technique are detailed to the left. B) Notations and detailed explanations for the octamers and nucleosomes used in this thesis.

3.2 Aims

The aim of this chapter is to detail the development and establishment of the generation of specifically modified nucleosomes, both symmetrically and asymmetrically modified. These will be used to investigate how the presence of certain modifications effects the presence of proteins bound. This will allow analysis of the proteins bound to bivalent asymmetric nucleosomes and therefore insight into their function. In order to use an *in vitro* approach recombinant bivalent asymmetric nucleosomes including controls need to be generated (see Fig. 3.1). To produce specifically modified nucleosomes there are a number of steps required - detailed in Figure 3.1 - including histone expression, modification, octamer assembly and nucleosome reconstitution. Each of these steps requires establishment and optimisation which will be detailed in this chapter using various core histones as examples. This chapter will focus on the establishment and optimisation of the techniques required for modification and purification of histones, and reconstitution of nucleosomes is demonstrated.

3.3 Histone Expression and Purification

To generate modified nucleosomes, the four core histones needed to be expressed along with the histone constructs required for the generation of modified histones. *Xenopus laevis* histone sequences were used for histones H3 and H4 because *Mus musculus* and *Xenopus laevis* histones H4 and H3.1 are 100% identical. The expression of *Xenopus Laevis* histones for use in biochemical assays has long been established (Luger et al., 1999a; Luger et al., 1997a; Dyer et al., 2004) and does not necessitate further codon adaption. Human H2A and H2B sequences were used as they are identical to mouse while the *Xenopus laevis* H2A and H2B sequences vary considerably - 94% sequence identity.

The histone constructs were expressed in bacterial cultures and the effect of different *E. coli* strains on protein expression levels was tested (see Fig. 3.2). BL21 strains are deficient in Lon protease (cytoplasm) and OmpT protease (outer membrane) and are thus more amenable to production of protein from the transformed/inserted plasmid. DE3 bacteria contain the DE3 lysogen that carries the gene for T7 RNA polymerase under control of the lacUV5 promoter. Inducible and maximal expression of a protein under the T7 promoter requires

65
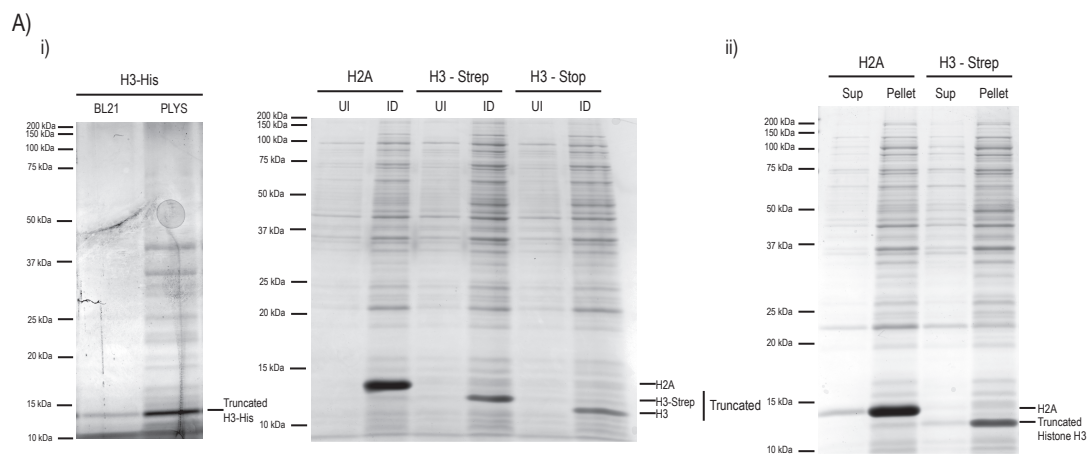
Figure 3.2 *Expression and solubility test of truncated H3 and core histones. A) i) SDS-PAGE analysis of expression tests of tagged H3-His in both BL21 (DE3) and BL21 (DE3) (PLYS) E.coli. H3-His expression is toxic to BL21, the yield is better from BL21 (DE3) (PLYS) as there is tighter control of expression and therefore better survival rate. Truncated and core histone expression upon induction of protein production by the addition of IPTG is shown. UI – uninduced, ID - induced. ii) Solubility test of tagged histone H3 (H3-Strep) and H2A supernatant and pellet - containing bacterial inclusion bodies.*

the addition of IPTG. PLYS strains contain a pLYS plasmid which produces T7 lysozyme which degrades T7 polymerase. This reduces background expression of the recombinant gene under the T7 promoter but does not affect protein expression induced by the addition of IPTG. PLYS therefore allows for tight control of target gene expression which is useful when expressing proteins which are toxic to the bacteria. The pLYS plasmid also contains a chloramphenicol resistance gene, the addition of chloramphenicol to the overnight culture ensures that BL21 DE3 pLYS strain retains the PLYS plasmid. All the core histones were expressed and mostly secluded in inclusion bodies by BL21 strains, although histone H4 was more soluble and thus yielded a lower amount of protein from the inclusion body purification.

There are two main and well-established ways of generating modified histone H3, through native chemical ligation (NCL) which generates modified histones with a cysteine point mutation or through alklyation chemistry which generates a methyl-lysine analogue (MLAs) (see Fig. 3.7). The NCL method was chosen to create modified histones due to MLAs not giving satisfactory results (see Chapter 4). The NCL method relies on ligation between a truncated histone with an N-terminal cysteine and a N-terminal tail peptide thioester carrying the modification of interest. In order to produce specifically modified histones via the NCL method, H3Δ1-31T32C truncated histones are expressed and purified, before ligation to a modified H3 N-terminal tail peptide resulting in histone H3

66

with a threonine to cysteine mutation at residue 32 (T32C). The T32C mutation is not expected to disrupt nucleosome structure as residue threonine 32 is well outside the globular domain of the nucleosome (Luger et al., 1997a), additionally Thr32 is not well conserved in eukaryotes (Wells and Brown, 1991).

To allow purification of asymmetrically modified octamers, constructs containing a C-terminally tagged (either with His or Strep) truncated H3 were cloned. The truncated histones are tagged via their C-terminus to reduce the likelihood of the tag affecting proteins binding to the H3's N-terminal tail. Tagged histones assemble into octamers and nucleosomes indicating that the tag's presence does not affect their assembly with other histones (see Fig. 3.9 & Fig. 3.11).

Truncated H3 expression constructs from the lab of Till Bartke (Bartke et al., 2010) were expressed in bacterial cultures (Fig. 3.3). The plasmid containing these histone constructs contains a start codon followed by a modified TEV cleavage site (ENLYFQ↓C) followed by the H3.1 sequence starting from glycine 33 (Fig. 3.3). The presence and cleavage of the TEV protease site is to help expose the cysteine in front of the histone core sequence in *E. coli* required for NCL. The desired outcome is a fully cleaved truncated histone H3 peptide, with an exposed N-terminal cysteine, ready for purification by size exclusion and ion affinity chromatography. However, these constructs were not fully cleaved by the TEV protease (expressed from the same plasmid pET28a under the control of a AraC/PBAD-promoter) in the bacteria (Fig. 3.4). The cleavage of these truncated histones was optimised by testing a range of different arabinose concentrations. As the TEV protease is under the control of the P_{BAD}, the expression of the TEV protease is induced by the addition of arabinose. The higher the amount of arabinose the more TEV protease is expressed. However, even at high concentrations of arabinose cleavage at the TEV cleavage site was still incomplete (Fig. 3.4). All histones expressed and purified are detailed in Figure 3.3.

As a result of the incomplete cleavage of these constructs, in order to maximise yield of the differently tagged truncated H3 we decided to generate the truncated histone by relying on bacterial cleavage of the initiator methionine followed directly by the T32C mutation or T45C mutation - if the truncated histone is destined for H3K27me3 (Fig. 3.2). The T45C mutation and truncation site was chosen for the creation of H3K27me3 histones as this places the site of the mutation further away from the site of modification reducing the likelihood that the mutation will influence protein binding. Suitable expression plasmids were

67

A)

B)

Expressed Histones	Species
H4	*Xenopus Laevis*
H3	*Xenopus Laevis*
H2A	*Homo Sapien*
H2B	*Homo Sapien*
Tev-Δ1-31T32C H3	*Xenopus Laevis*
MΔ1-31T32C H3	*Xenopus Laevis*
MΔ1-31T32C H3-Strep	*Xenopus Laevis*
MΔ1-31T32C H3-His	*Xenopus Laevis*
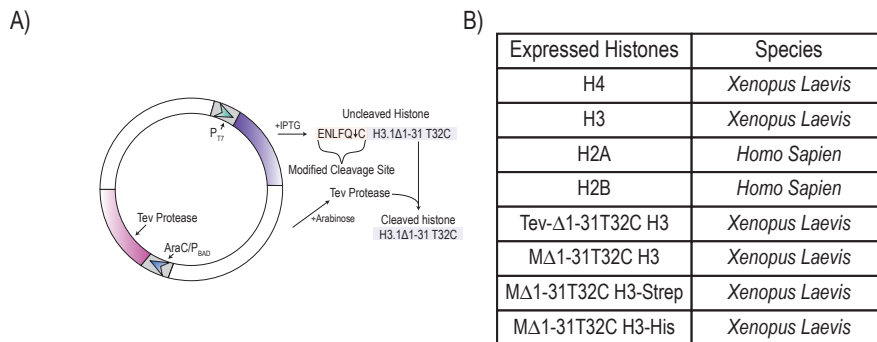
Figure 3.3 *Plasmid map and table of expressed histones.* *A) Plasmid map of pET28a(+)-AraC-PBAD-His6TEV/pro-H3.1Δ1-31T32C obtained from Till Bartke's lab containing Tev cleavage site H3.1Δ1-31 T32C,and the His tagged Tev protease. B) Table of all the expressed histones and the species their sequence comes from.*
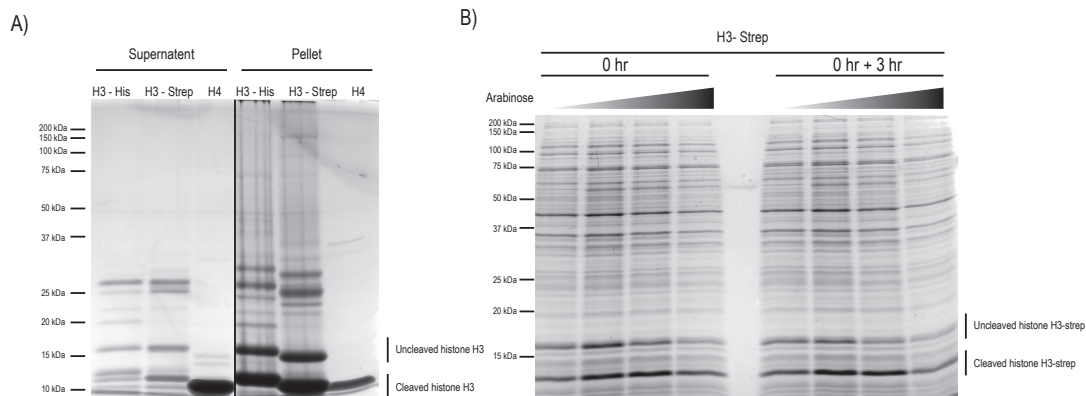
Figure 3.4 *Optimisation of the Yield of H3 T32C Cleaved Histones A) Solubility test of tagged histone H3 (H3-His H3-Strep) used for asymmetric histone octamer formation and H4 supernatant and pellet - containing bacterial inclusion bodies. B) Optimisation of TEV protease cleavage of the expressed histone. A range (0–1% (w/v)) of arabinose was added to optimise cleavage of the histone. Arabinose was added once (at the 0 h timepoint) or twice (3 h timepoint) - when target protein expression was induced.*

68
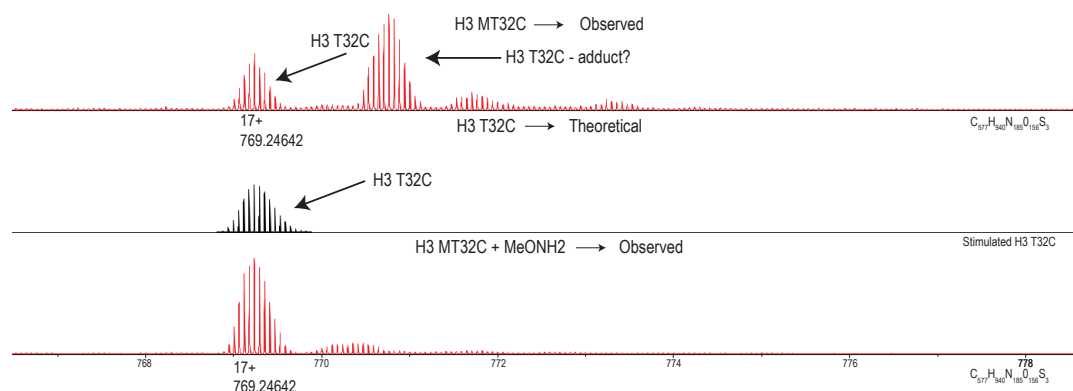
Figure 3.5 *Initiator Methionine Residue Cleavage. Spectrum for theoretical mass charge ratio of the truncated histone protein and the actual spectra of the truncated histones purified from the E.coli culture. The initial spectrum is obtained from the purified T32C truncated histone with the peaks identities labelled. There are two main peaks, the first is of the truncated histone H3 T32C protein and the second peak is prominent suggesting the presence of a truncated histone adduct. The second spectrum is the theoretical spectra of the truncated histone H3 T32C with cleaved methionine and no adduct. The third is the spectrum of the purified truncated histone H3 T32C in the presence of methoxylamine hydrochloride having resolved the adduct.*

generated and successful expression of all of the histones was confirmed (Fig. 3.2). The initiator methionine was cleaved for all truncated H3 histones and consequently these constructs could be used for NCL reactions and the TEV cleavage approach could be abandoned (see Fig. 3.5).

After inclusion body purification (as described in materials and methods), although relatively pure, the histones were further purified by tandem ion exchange in a urea-based buffer (Fig. 3.6). Further histone purification will be illustrated using H2B as an example. Ion exchange chromatography is used to separate histone proteins from other proteins present in the solution. The anion exchange column removes proteins with a negative net charge resulting in a solution that contains only positively charged proteins. This is further purified by running the flow-through through the cation exchange column which binds overall positively charged proteins. Histones have a net positive charge, largely due to their positively charged tail, this allows them to bind to the negatively charged resin of the cation exchange column - stationary phase. After the proteins are bound, a 0.1 M to 1 M NaCl gradient (Fig. 3.6) elutes the proteins bound to the cation column. As the concentration of NaCl ions increase, the proteins bound to the stationary phase of the cation column are released. Proteins with weaker ionic interactions will be released at lower ion concentrations compared with proteins with a higher ion interactions and consequently a higher affinity for the stationary phase. To perform ion exchange chromatography with truncated

69

histones, the running buffer contained 50 or 75 mM NaCl to allow binding of the N-terminally truncated and thus less positively charged histones to the cation exchange column. This combination of both anion and cation exchange columns allow purification of both full length and truncated histone proteins from the solubilised inclusion body preparation.

In this way, all required core histones were successfully purified (see Fig. 3.6 for a sample purification for histone H2B). The sample bound to the monoQ (anion) column and flowthrough of the ion exchange columns both have low $A_{280/260}$ ratios, indicating that they have high DNA contamination. Contrary to the fractions eluted from the cation column which have a high $A_{280/260}$ ratio indicating not only the presence of protein shown in the SDS-PAGE analysis (Fig. 3.6) but also very little DNA contamination. Separation of the histone from the majority of the contaminants is clear (Fig. 3.6). On average 24 1-ml fractions were pooled, with regards to the example shown (Fig. 3.6) fractions B3 to D5 were considered to contain sufficiently pure histone and thus pooled and dialysed against 3 mM beta-mercaptoethanol in water before lyophilisation. Once purified, the truncated histones are ready for native chemical ligation while other purified core histones can be used immediately for octamer assembly.

Prior to the start of this project H3 expression constructs used for the methyl-lysine analog approach were expressed ready for cysteine alkylation. However, these were not used further in this project and so their expression will not be detailed.

A)

B)

Elution from Cation Column

HiTrap - 225 - 460 mM NaCl
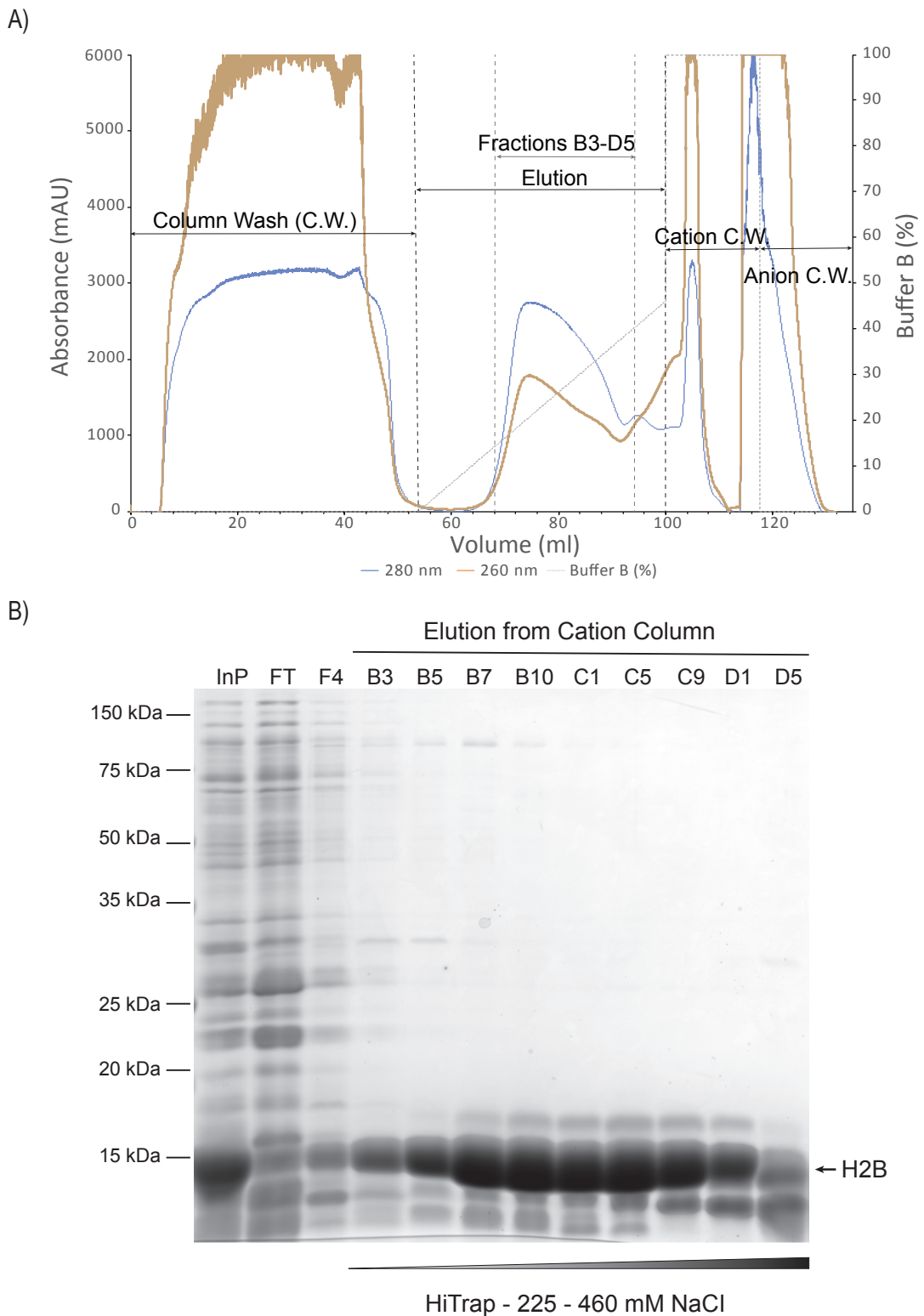
Figure 3.6 *Full-length Histone H2B Purification.* *A) A chromatogram of the ion exchange purification of histone H2B. The elution was monitored by A_{260} and A_{280}. The solution's $A_{280/260}$ ratio helps to assess DNA contamination of the solution. The flowthrough (FT) (0 ml to 53 ml) and monoQ (118 ml to 130 ml) mainly consists of DNA. Fractions B3 to D5 were pooled and lyophilised. B) SDS-PAGE analysis of ion exchange purification of H2B. InP - Input. FT- Flowthrough. Fraction F4 (115 ml)- Analysis of proteins bound to monoQ column. Fractions B3-D5 protein eluted from HiTrap column, mainly consisting of H2B.*

3.4   Methyl-lysine Analogs

Cysteine alkylation chemistry as first performed by the group of Kevan Shokat to site-specifically modify histones (Simon, 2010; Simon et al., 2007; Simon and Shokat, 2012; Chen et al., 2014; Chatterjee and Muir, 2010) resulting in the generation of methyl-lysine analogs (MLAs). MLAs (Fig. 3.7) are widely used as proxies for specifically modified histones in studies investigating nucleosome binders and modifiers (Lauberth et al., 2013; Francis et al., 2009; Margueron et al., 2009; Simon et al., 2007; Shen et al., 2009). Histone sequences are well conserved between eukaryotes (Wells and Brown, 1991) and many species' histone H3 contain only one cysteine (position C110) which can be mutated to alanine without noticeable effects on histone or nucleosome structure or function (Simon et al., 2007). Thus the site-specific mutation to cysteine at the desired modification site allows the use of a variety of methods including cysteine alklyation to create specifically modified histones. The cysteine is alkylated to a N-methylated aminoethylcysteine (K_c) which is an analog of methylated lysine with sulphur replacing the γcarbon of the lysine (Fig. 3.7) (Simon et al., 2007). It allows the production of modified histones with multiple sites of methylation - N-methylated aminoethylcysteine - although it does result in all the sites being equally methylated.

In spite of aminoethylcysteine's similarity to lysine, these early pulldown experiments showed an inability of proteins including PRC2 present in the nuclear extract to distinguish between differently modified nucleosomes (see Chapter 4 Fig. 4.9). EZH2 is used as a indicator of PRC2 binding as it is the catalytic subunit and an integral member of the protein complex. Due to the inability of PRC2 to distinguish between different MLA substrates in pulldown experiments, it was decided to change the method of *in vitro* synthesis of specifically modified nucleosomes to native chemical ligation.
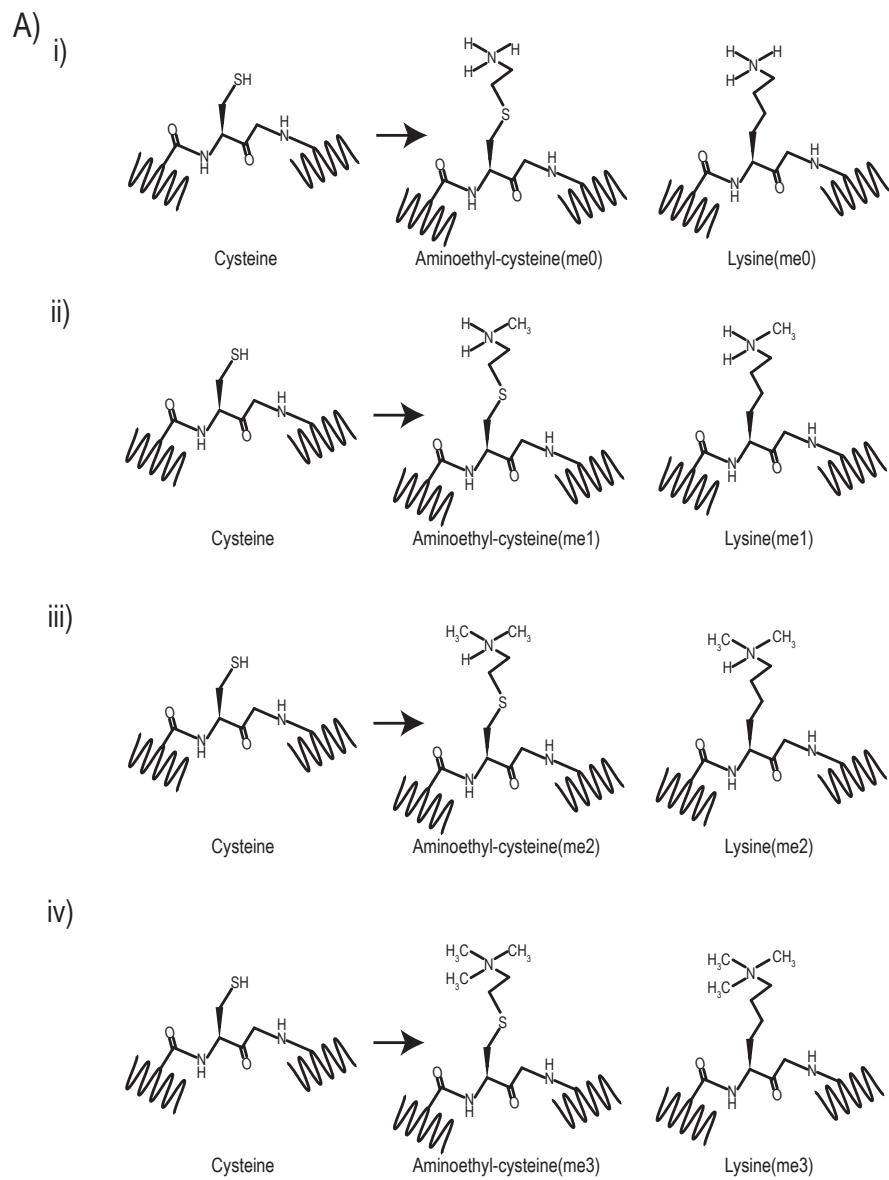
Figure 3.7 *Aminoethylcysteine and Lysine residues.* *Structure comparison of cysteine, and variously modified methyl-lysine analogs and lysines A) i) unmethylated residues ii) monomethylated residues iii) dimethylated residues and v) trimethylated residues.*

3.5 Native Chemical Ligation

Native Chemical Ligation (NCL) was developed by the Kent lab (Dawson et al., 1994) and developed for use on histone proteins in the Peterson lab, among others (Shogren-Knaak et al., 2003; Chen et al., 2014). In the NCL method, 2 peptides, one containing a C-terminal thioester (peptide-R-thioester) and one with a N-terminal cysteine, are joined together via a three-step process. Firstly, a thiol-thioester exchange between the thioester containing peptide and an externally added alkyl which results in a new thioester containing peptide and then a transthioesterifcation reaction between the new thioester and the peptide containing an N-terminal cysteine to produce a new thioester linked intermediate (Fig. 3.8) - these steps are reversible. The third step - a spontaneous S to N acyl shift resulting in an native amide bond joining the two peptides - is irreversible. The product polypeptide resulting from the NCL reaction is chemically identical to the endogenous protein except with a cysteine at the site of ligation. In the case of the NCL method, as the reaction is between an N-terminal cysteine and a peptide-R-thioester, the C110A mutation in histone H3 that is necessary for the MLA reaction is not obligatory. The cysteine at the site of ligation can be resolved by desulfurization to an alanine if required (Dawson et al., 1994; Dawson, 1997; Dawson and Kent, 2000; Kent, 2006; Thapa et al., 2014; Wang et al., 2011; Chatterjee and Muir, 2010). We did not perform this step as both threonine and cysteine are both polar amino acids with relatively long side chains capable of forming hydrogen bonds while alanine is hydrophobic; therefore the desulfurization step would result in a greater alteration to the sequence than leaving the T32C mutation in place.

To create specifically modified histones via this method (Fig. 3.8), a truncated histone peptide lacking residues 1-31 or 1-44 with a threonine to cysteine substitution/mutation at residue 32 or 45 respectively, is reacted with a peptide consisting of the histone H3.1 N-terminal tail and a C-terminal thioester. This mutation is not expected to cause any disruption to the histone or nucleosome structure as the residue mutated is outside the structured domain of the nucleosome (Luger et al., 1997a; Bartke et al., 2010). The tail peptide can include any desired modification at any preferred location for example, H3K4me3 or H3K4me3K27ac. Initially the truncated histone peptide is reacted with methyoxyamine to help resolve any N-terminal adducts that may have formed during its expression in *E.coli*, masking the N-terminal cysteine and preventing
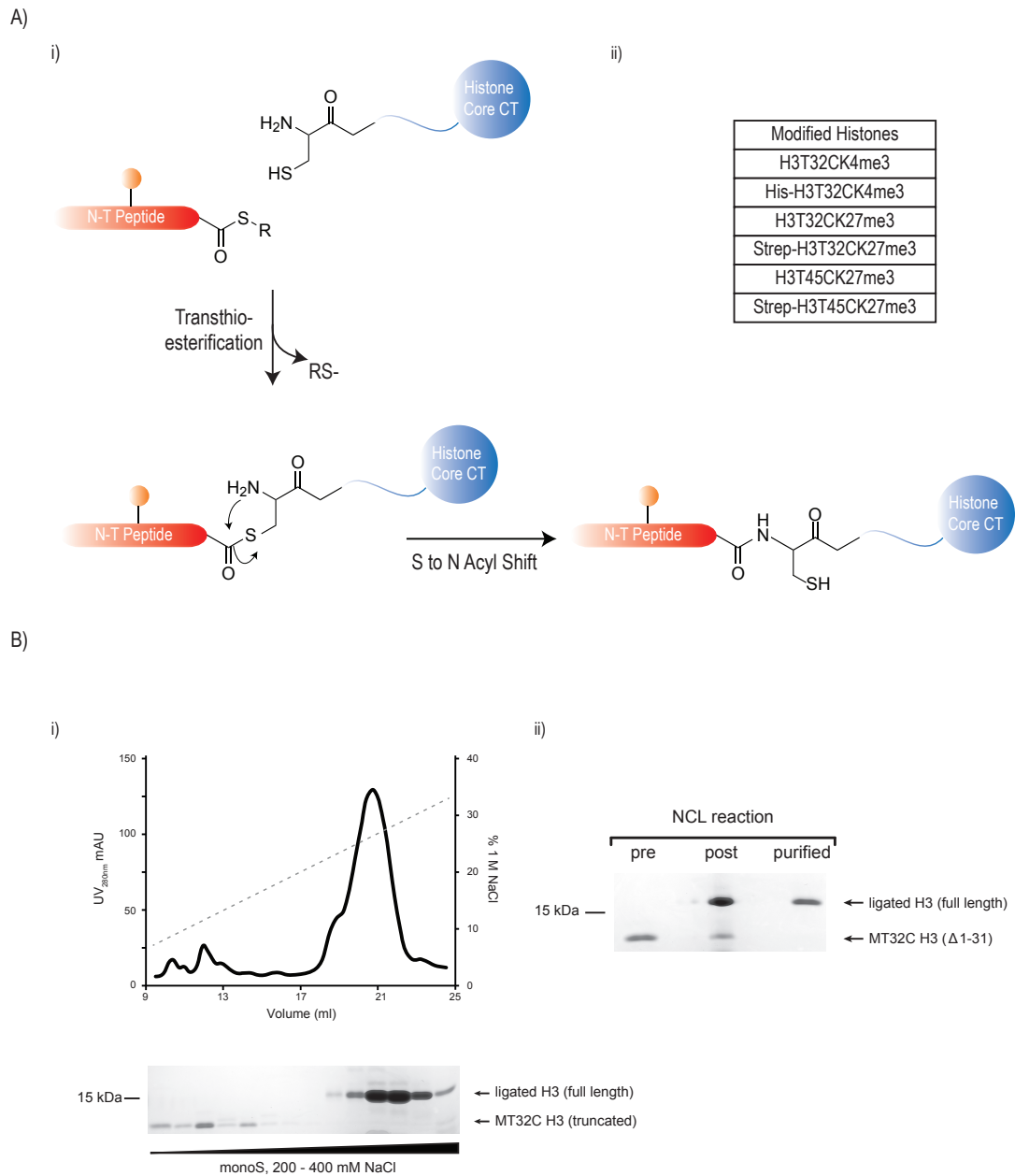
Figure 3.8 *Native Chemical Ligation (NCL) reaction scheme.* *A) Native chemical ligation mechanism Steps 2 (transthioesterification) and 3 (S to N acyl shift) shown. B) i) Cation exchange chromatography profile of ligated histone H3 purification. Ligated H3 eluted around 330 mM NaCl. SDS-PAGE analysis of purification shown below, with gradient of NaCl indicated. ii) Coomassie stained SDS-PAGE analysis of the native chemical ligation: pre-ligation reaction, post-ligation reaction and after cation exchange purification.*

the ligation reaction (Virdee et al., 2011; Chen et al., 2010; Chen et al., 2014) (Fig. 3.5). The NCL reaction is carried out in the presence of MPAA, an aryl thiol catalyst which reacts with the thioester containing peptide to create a more active thioester on the C-terminus of the peptide, increasing the speed of the transthioesterifcation reaction (Johnson and Kent, 2006).

As can be seen from the chromatogram and corresponding SDS-PAGE gel, the ligated H3 histone is purified away from the unligated histone by cation exchange chromatography (Fig. 3.8). The ligated histone contains a positively charged tail which allows it to bind to the monoS resin with greater affinity than the unligated histone. A gradient of NaCl helps separate out these proteins. The ligated histone elutes from approximately 325 mM NaCl onwards. The coomassie shows the progress of the NCL reaction, yield of ligated H3 and quality of the purification (Fig. 3.8).

3.6 Assembly of Histone Octamers

Once the histones have been successfully expressed and purified they need to be combined in order to assemble a histone octamer. The histones are mixed together in a mass ratio of H3:H4:H2A:H2B (1:1:1.2:1.2) and assembled into octamers through salt gradient deposition removing the guanidinium chaotrope while dialysing into 2 M NaCl (Luger et al., 1997a; Luger et al., 1999a). The excess of H2A and H2B ensures optimal octamer assembly of the H3-H4 tetramer, reducing contamination with hexamer particles. The (H3-H4)$_2$ tetramer and the (H2A-H2B) dimer is thought to form separately before forming the complete histone octamer (Flaus, 2011). The dialysed solution contains both assembled histone octamers, unassembled histones and potential aggregates (Fig. 3.9). The histone octamers are purified by size-exclusion chromatography (Fig. 3.9).

As can be seen from Figure 3.9 when purifying histone octamers by size-exclusion chromatography (S200) a characteristic five peak chromatogram is obtained. The initial peak in the void volume (<8 ml) contains aggregates, the second peak around 13 ml contains histone octamers and the third (15 ml), fourth (16.5 ml) and fifth (17 ml) peak contains tetramers, dimers and unassembled histones. If only one type of histone H3 is used then the purified symmetrically modified octamers are ready for assembly into nucleosomes.
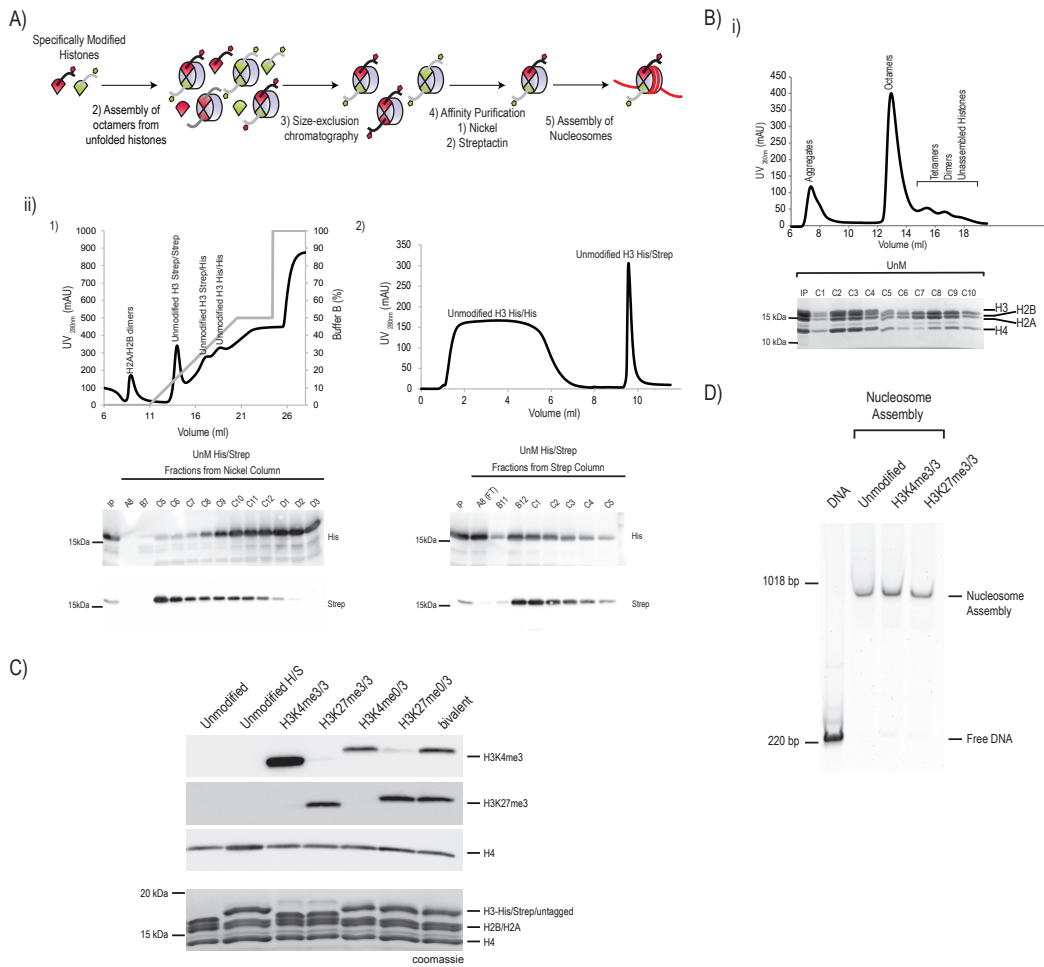
Figure 3.9 *Histone Octamer Assembly and Purification A) Scheme of nucleosome assembly B) i) S200 Size-exclusion purification of Unmodified untagged histone octamer. Chromatogram peaks labelled. Aggregated protein elutes in the void volume. Octamers elute first followed by: hexamers, tetramers, dimers and unassembled histones. ii) IMAC (NiNTA) of Unmodified His/Strep tagged octamer. The absorbance at $A_{(280)}$ steadily increases along with the concentration of Buffer B due to the absorbance at $A_{(280)}$ by the contaminates present in the imidazole used in Buffer B. iii) Affinity (Streptactin) purification of Unmodified His/Strep tagged octamer. C) Analysis of symmetrically and asymmetrically modified nucleosomes. Presence of H3K4me3, H3K27me3 and H4 shown by western blot. H3K4me3 antibody binds better to the untagged than tagged H3K4me3 modified histone. Loading and ratio of histones in octamer demonstrated by SDS-PAGE. D) Assembly of Unmodified, H3K4me3/3 and H3K27me3/3 untagged nucleosomes.*

To create asymmetrically modified octamers two types of histone H3 are used with different modification states as described by (Voigt et al., 2012). Each of these differently modified H3 histones are tagged with either a 6XHis ($_H$)or Strep ($_S$) tag. They are mixed together with other core histones in a mass ratio H3$_S$:H3$_H$:H4:H2A:H2B (0.5:0.5:1:1.2:1.2). Formation of octamers is unaffected by the presence of the tags (Voigt et al., 2012). The purified octamer fractions from size-exclusion chromatography contain a mixture of 3 octameric species (Fig. 3.9), only one of which contains the correct combination of the desired modifications.

These octameric species are separated out by two rounds of affinity purification. First the solution is passed through a Nickel column, which binds H3$_H$, removing octamers containing only H3$_S$ and thus symmetric for the His tagged H3 species (Fig. 3.9). During nickel affinity purification the H2A/H2B dimers are stripped from the H3/H4 tetramer. Secondly, the mixture is passed over a streptactin column which binds tetramers containing H3$_S$ and consequently separates out the tetramers which only contain H3$_H$ (Fig. 3.9). This leaves a solution containing only histone tetramers comprising of both histone H3$_S$ and H3$_H$ and are therefore asymmetrically modified.

3.7   Optimisation of Nucleosome Assembly

A nucleosome consists of approximately 147 base pairs (bp) of DNA wrapped around a histone octamer. Additional base pairs can be added to allow various lengths of overhangs on either side of the octamer. The DNA used to make octamers in this study consists of 209 bp of DNA - approximately 31 bp overhang on each side of the 601 Widom Positioning Sequence (Lowary and Widom, 1998). The polymerase chain reaction (PCR) generating the desired 209 bp 601 DNA fragment was optimised to improve its yield, higher primer and MgCl$_2$ concentrations increased the DNA yield by threefold, from 180 µg from a 96-well plate reaction to 540 µg (Fig. 3.10).

The assembly of the nucleosome from DNA and histone octamer is completed via salt gradient dialysis (Luger et al., 1999a; Luger et al., 1999b). As the concentration of salt decreases this allows the DNA to wrap around the histone octamers. When assembling asymmetrically modified nucleosomes, the specifically modified tetramers are mixed with the H2A/H2B dimers in a 1:1 ratio. The protein to DNA ratio is determined on a individual protein basis by
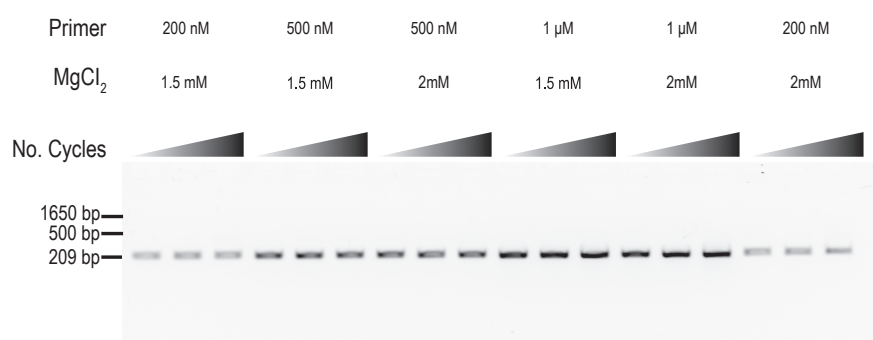
Figure 3.10 *601 Nucleosome DNA PCR Optimisation. An 1% agarose gel showing the amount and purity of DNA from each of the 209bp PCR conditions. Optimisation conditions including a range of PCR (30-40) cycles and different MgCI$_2$, and primer concentrations are indicated above the gel.*

titrating the protein against a constant amount of DNA (Fig. 3.11).This ratio was kept constant for all subsequent nucleosome assemblies from the same octamer purification and re-adjusted when a new octamer or DNA purification was used. In asymmetric nucleosome assembly the tetramer and dimer bind the DNA once the solution has reached different ionic strengths, the tetramer binds the DNA at approximately 0.85 M, and the dimer at around 0.65 M before reaching an equilibrium at 0.5 M and below (Flaus, 2011).

The reconstituted nucleosomes are run on a 6% TGE gel to verify their assembly and stained with Sybr Safe to ensure both the free DNA and DNA incorporated into the nucleosomes is visible. The incorporation of the DNA into the nucleosome causes a shift of the DNA on the gel, from free un-assembled DNA running at approximately 200 bp to fully assembled DNA running equivalent to 1 kb free DNA.

Although the method by which Sybr Safe binds to DNA is unclear, as it is a cyanine dye and a deriviative of thiazole orange it is reasonable to assume it intercalates between the base pairs of minor groove of DNA (Zipper et al., 2004; Bunkenborg et al., 2000; Petersen et al., 1999). The decreased efficiency of its binding to DNA incorporated into nucleosomes in comparison to free double stranded (ds)DNA could be due to two possible reasons. As the DNA is wound around the histone octamer this will obstruct intercalation sites, thus reducing the total amount of Sybr Safe able to bind the DNA and therefore decreasing the signal. Intercalation requires DNA to open up space between the DNA base pairs to allow the intercalating molecule to bind (Lerman, 1961), this movement may be restricted by the preceding co-ordination of the DNA by the histone octamer. This explains why the 100 ng free DNA run as a control is always brighter than
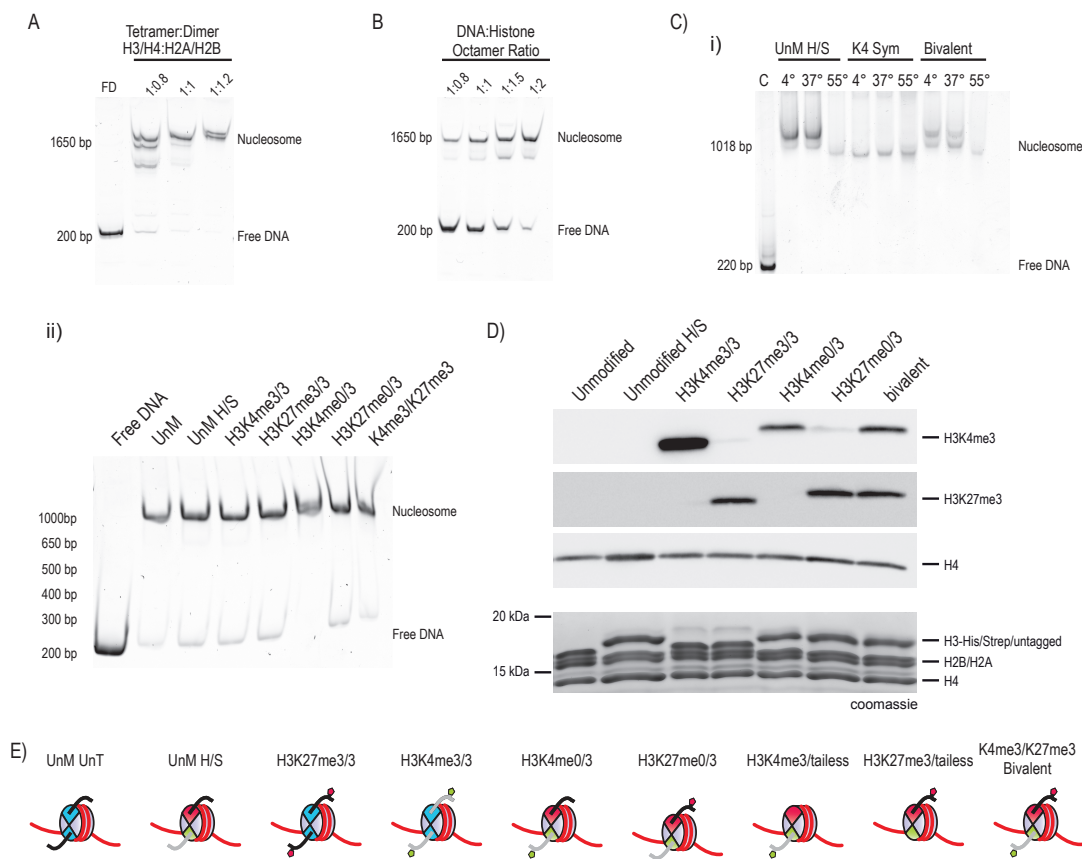
Figure 3.11 *Nucleosome Assembly Optimisation* shown on native 6% TGE gels
A) Nucleosome assembly of unmodified H/S tagged H3 histone octamer at a range of
H3/H4:H2A/H2B ratios B) Nucleosome assembly of unmodified untagged H3 histone octamer at
a range of DNA:Histone Octamer ratios C)i)Heat-shift assay at 4°C to 55°C on symmetrically
modified and asymmetrically modified nucleosomes. UnM H/S - Histone octamer containing
unmodified His/Strep tagged histone H3, H3K4me3/3 - Histone octamer containing untagged
H3K4me3, K4/K27 - Histone asymmetrically modified octamer containing one H3 modified with
H3K4me3-His and one modified with H3K27me3-Strep ii) Optimised nucleosome assembly.
unmodified - Histone octamer containing unmodified untagged histone H3, UnM H/S -
Histone octamer containing unmodified His/Strep tagged histone H3, H3K4me3/3 - Histone
octamer containing untagged H3K4me3, H3K27me3/3 - Histone octamer containing untagged
H3K27me3, H3K4me0/3- Histone octamer containing H3K4me3-His and unmodified H3-
Strep , H3K27me0/3 - histone octamer containing H3K27me3-Strep and Unmodified H3-His,
K4me3/K27me3 (bivalent)- histone octamer containing H3K4me3-His and H3K27me3-Strep D)
Analysis of symmetrically and asymmetrically modified nucleosomes. Presence of H3K4me3,
H3K27me3 and H4 shown by western blot. H3K4me3 antibody binds better to the untagged
than tagged H3K4me3 modified histone. Loading and ratio of histones in octamer demonstrated
by SDS-PAGE. E) Simplified depiction of the various modified nucleosomes generated. The
presence of the tags is indicated by the colour of the H3 copies - green (His) or red (Strep),
while the blue indicates the absence of tags. Red modification H3K27me3 and green modification
H3K4me3.

the DNA incorporated into the nucleosome even if there is no unassembled DNA in that sample. This effect also means that Sybr Safe will skew the assessment of unbound:bound DNA ratios by exaggerating the amount of unbound DNA.

Due to the presence of the DNA overhangs, the assembled nucleosomes can be heterogeneous with respect to the position of the octamer on the DNA. As the TGE gel separates based on the sample's size and shape this is visible as two slightly separated bands around 1000 kb (Fig. 3.11). This difference in nucleosome placement is resolved by one hour incubation at 40 °C to 55 °C (Fig. 3.11).

3.8 Conclusion

The yield of all of the histone constructs were optimised for expression and purification from *E. coli* BL21 (DE3) bacteria cultures. The yield of the truncated histones was originally reliant on the expression and activity of the TEV protease. Resulting in incomplete cleavage of the TEV cleavage site, and a low yield of the fully truncated histone. Therefore the method of expression was altered to rely on bacterial cleavage of the initiator methionine residue, generating a higher yield of fully truncated histone.

Successful purification protocols for all histones were established, consisting of inclusion body purification followed by ion-exchange purification. Purification of truncated histone H3 required alterations to the established purification protocols. Various methods of creation of specifically modified histones were tried and native chemical ligation was chosen. Production of specifically modified H3 by native chemical ligation followed similar parameters as optimised in Chen et al., 2014. Ligated H3 was separated from unligated H3 by cation exchange chromatography.

Purification of modified octamers was established, with three rounds of chromatography: size-exclusion chromatography then two rounds of affinity chromatography - strep-tactin and nickel. Affinity chromatography purifies the asymmetrically modified octamers by relying on differential binding of the histones containing a His or Strep tag. Resulting in the elution of purified octamers monitored by absorbance at A_{280}, A_{260} and $A_{280/260}$ ratios.

Nucleosome assembly was optimised for each specific modified octamer. The general assembly of the nucleosomes was optimised by titration of the H3/H4:H2A/H2B

ratio and Octamer:DNA ratio. Once assembled the placement of the octamer on the DNA was made homogenous by heat shift assay at 40 °C to 55 °C.

The generation of specifically modified histones and nucleosomes was successfully established and optimised. The specifically modified nucleosomes can be used in subsequent experiments to isolate the proteins that bind them.

Chapter 4

Establishment and Optimisation of Pulldown Approach

4.1   Introduction

Multiple studies have established recruitment of specific proteins to histone modifications, using modified peptides or nucleosomes *in vitro* or a combination of techniques. The *in vitro* techniques which show direct binding are often coupled with ChIP assays showing correlation with certain histone modifications *in vivo* (Musselman et al., 2012; Taverna et al., 2007; Bartke et al., 2010; Vermeulen et al., 2010; Vermeulen et al., 2007). Recently, a number of labs have shown how multiple modifications on the same nucleosome or histone tail either in peptide form or as part of a nucleosome, affects protein binding (Nuland et al., 2013; Brown et al., 2017; Ruthenburg et al., 2011). All of these studies have focused on either one histone tail by using peptides or a nucleosome containing the same modifications on both sister H3 copies. However, asymmetrical nucleosomes do exist *in vivo* and are likely present in multiple cell types (Bernstein et al., 2006a; Azuara et al., 2006; Roh et al., 2006; Pan et al., 2007; De Gobbi et al., 2011; Voigt et al., 2012; Mikkelsen et al., 2007).

The reason for using an *in vitro* approach to investigate the proteins that bind to bivalent nucleosomes was simple; it allows a non-biased view and comparison of all proteins bound to the differently modified nucleosomes without any of the additional complicating factors present *in vivo*. The use of a nucleosome rather

than simply modified peptides allows us to determine how the conformation of the nucleosome and its histone tails might influence binding of proteins, as they would *in vivo* (Chen et al., 2014; Nuland et al., 2013; Chen et al., 2018). This gives a much more accurate picture of the proteins' binding and allows sites on the histone modifiers that might bind to nucleosomal DNA, neighbouring histones, or parts of the histone apart from the modification in question to contribute to protein binding affinity. Using nucleosomes allows influences of all core histone tails and their modifications to be accounted for and quantified. Mononucleosomes allow a clear read out of binding rather than arrays where it is difficult to ensure a specific arrangement of nucleosomes and to attribute the result to a single nucleosome. Once proteins that bind a specifically modified mononucleosome are known, the effect of nucleosomal arrays could be assessed in future work.

4.2 Aims

In the previous chapter the generation of the specifically modified nucleosomes was detailed. In this chapter, the optimisation and establishment of the pulldown assay using these modified nucleosomes will be shown. H3K4me3/3 and H3K27me3/3 nucleosomes will be used to validate the method, as the proteins known to bind either the H3K4me3- or H3K27me3-modified peptides or nucleosomes are well established (Lauberth et al., 2013; Ingen et al., 2008; Clouaire et al., 2012; Kim et al., 2016; Chang et al., 2010; Peña et al., 2006; Simon, 2010; Bernstein et al., 2006b; Xu et al., 2010; Eberl et al., 2013; Vermeulen et al., 2010; Bartke et al., 2010). Validation of the assay will be accomplished by comparison of the proteins bound to the modified nucleosomes used here and proteins known to bind them in the current literature.

4.3 Optimisation of Pulldown Approach

Once the specifically modified nucleosomes have been made and purified (see Chapter 3) they can be used in pulldown experiments. The theory of the pulldown assay is simple (Fig. 4.1); the bait protein, in this case the specifically modified nucleosome, is immobilised on beads, and incubated with nuclear extract (NE) that contains potential binding partners. The proteins bound non-specifically or with low affinity are removed through multiple washes, and finally the proteins
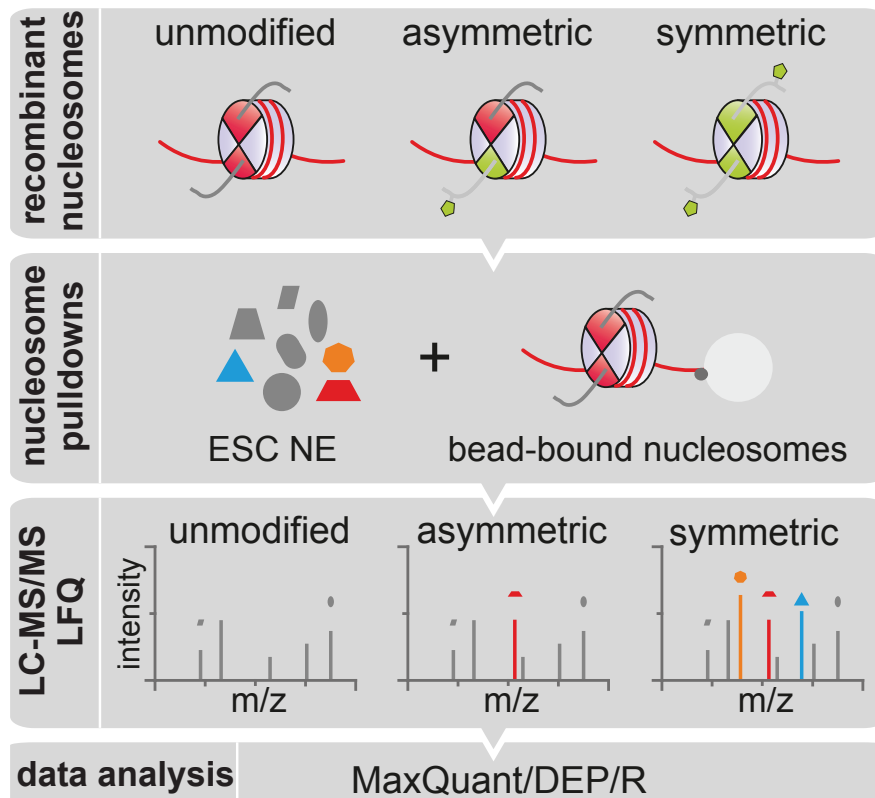
Figure 4.1 *Pulldown Theory Diagram. The purified recombinant specifically modified nucleosomes are immobilised on beads and incubated with nuclear extract (NE) from mouse ESCs. Unbound or unspecific protein binding is removed via multiple wash steps and then the proteins bound to the nucleosomes are eluted and analysed by label-free mass spectrometry (MS). The data obtained was analysed by three different programmes: MaxQuant, DEP and R.*

bound to the nucleosome are eluted.

In order for the nucleosomes to be immobilised for the pulldown, the DNA used to wrap around the histone octamer was tagged with biotin. To facilitate detection of specific binding proteins, unspecific binding of contaminating proteins must be minimised. A variety of beads coated with streptavidin (which binds biotin) were assessed for suitability of use in pulldowns (Fig. 4.2). Bovine serum albumin (BSA) blocked beads were removed from consideration as the presence and abundance of BSA could mask MS-based detection of some proteins present in the experiment - this removed M-280 and MyOne beads from consideration (Fig. 4.2) as BSA (66.5 kDa) was eluted in both methods used to elute proteins from the beads (see below). Streptavidin (13-14 kDa) was present in all elutions, although in greater amounts when the beads were boiled compared to a 37 °C elution. Based on the results of the initial optimisation, MyOne C1 beads were chosen, as they are not blocked with BSA, have a small diameter (same amount of streptavidin but smaller diameter means less extraneous surface for
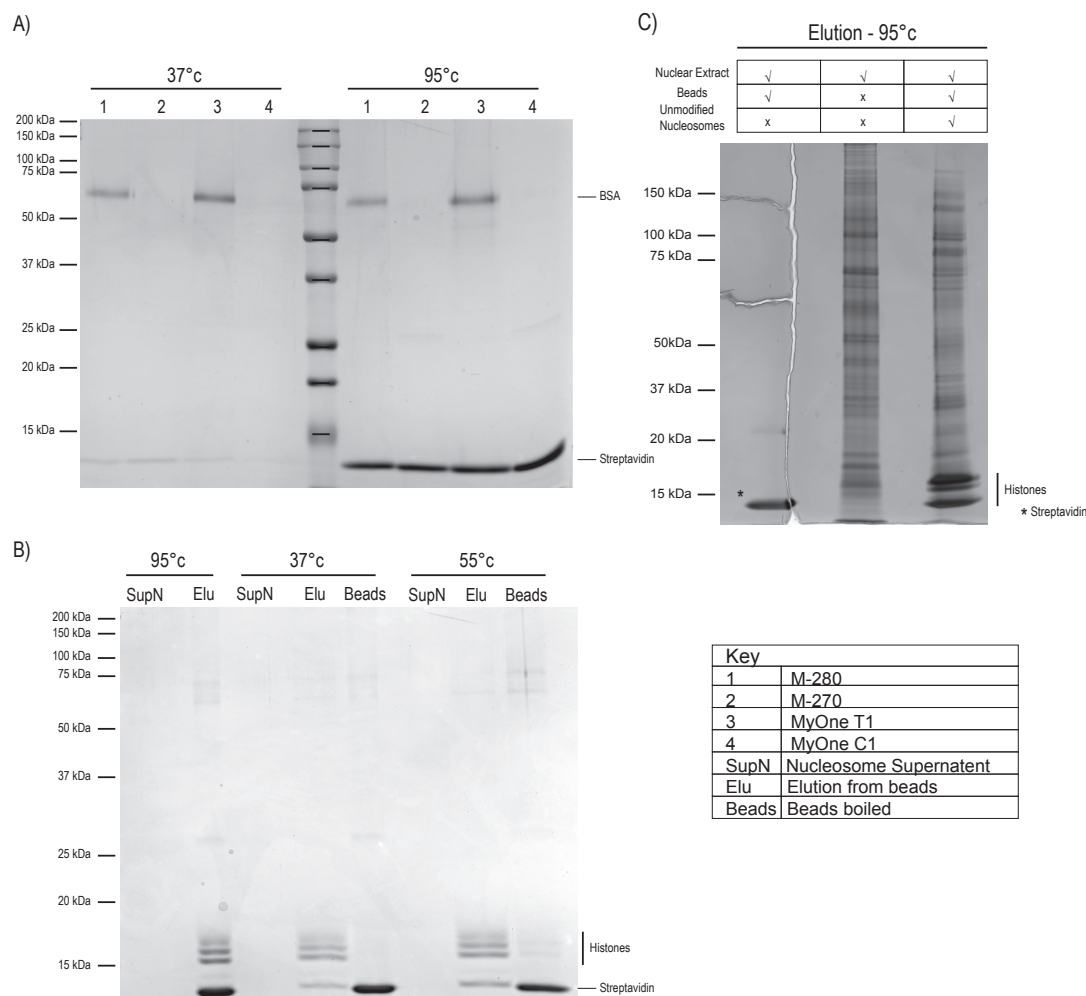
Figure 4.2 *Bead optimisation A) SDS-PAGE analysis of the elution from four different streptavidin beads - streptavidin (13-14 kDa) was present in all elutions. M-280 and MyOne beads are blocked with BSA (66.5 kDa). Key for beads type contained in the figure. B) A comparison of different elution methods on Myone C1 beads, two different elution buffers used at two different temperatures. SupN - supernatant Elu - elution Beads - boiled beads C) Silver-stained SDS-PAGE gel analysis of non-specific protein binding to beads.*

non-specific binding), are magnetic, have a greater binding capacity, and eluted less streptavidin in the 37 °C elution than the other beads tested. Magnetic beads allow for a cleaner sample, MS preparation, and enhance reproducibility.

To minimise the presence of contaminating proteins in the bead elution, three elution methods were compared. A gentler elution with FASP elution buffer containing (100 mM Tris-HCl pH8, 0.1% SDS, 25 mM DTT) at 37 °C or with FASP elution buffer (supplemented with 0.2% SDS) at 55 °C and a harsher elution in the presence of SDS sample loading buffer at 95 °C (Fig. 4.2). The 37 °C elution eluted the vast majority of the nucleosomes bound with minimal elution of streptavidin and BSA from the beads (Fig. 4.2), therefore this elution method

was selected for further pulldown experiments.

MyOne C1 beads were tested for non-specific protein binding to the proteins in the nuclear extract (NE) (Fig. 4.2). Beads were incubated with NE in pulldown conditions to reveal any non-specific binding that would be encountered in pulldown experiments with modified nucleosomes. The enrichment of different protein bands in the elution containing nucleosomes compared to the NE input sample illustrates the binding of specific proteins from the NE by the unmodified nucleosomes.

Minimal non-specific protein binding to the beads was preferred, as the presence of excess proteins may mask proteins specifically bound to the nucleosomes and affect the enrichment seen between conditions. As can be seen from the silver stained SDS-PAGE gel, MyOne C1 beads themselves non-specifically bound very few proteins from the NE, especially in comparison to the beads once nucleosomes were bound (Fig. 4.2). However, when these beads were evaluated using Mass Spectrometry (MS), a large number (1382) of previously unseen proteins are present bound to the beads (Fig. 4.3) even in the absence of nucleosomes, albeit at low levels. Some of the proteins bound to MyOneC1 beads included proteins that bind to the histone PTMs H3K27me3 or H3K4me3 including EZH2. The presence of proteins bound to the beads themselves may decrease visible enrichment of proteins between differently modified nucleosomes, making these beads less suitable for the pulldown experiments planned. Till Bartke's lab published a similar pulldown assay in 2018 using Streptavidin Sepharose High Performance (SHP) beads (Makowski et al., 2018). We compared these beads to MyOne C1 beads using the same method as before and discovered that the SHP beads bound non-specifically to fewer proteins (281 proteins, Fig. 4.3). As SHP are not blocked with BSA and have the least amount of non-specific binding of the beads tested, they were used for the subsequent experiments.

The amount of SHP beads required to bind 10.5 µg of nucleosomes was optimised via a bead titration for the pulldown assay. The optimal amount of beads would bind all the nucleosomes present with minimal excess surface area for non-specific protein binding. Western blot analysis of histone H3 bound to the beads and present in the supernatant showed that approximately 8 µL bead bed volume (10.5 µL slurry) was optimal for 10.5 µg nucleosome binding. This volume of beads was the minimal amount that bound all of the nucleosomes. Nucleosome concentration is determined by the amount of nucleosome DNA present as nucleosome assembly has been optimised to reduce free DNA, further

C)

Nucleosome	Magnetic Beads			Sepharose Beads		
	N/A	UnModified	H3K4me3 Sym	N/A	UnModified	H3K4me3 Sym
Protein						
KMT2B	-	-	-	-	8.39E+06	7.05E+06
TAF3	-	-	-	-	-	3.41E+07
Spin1	-	-	2.15E+08	-	-	2.49E+08
BAP18	5.51E+07	3.91E+07	8.34E+08	-	1.03E+08	1.76E+09
PHF2	-	-	2.43E+07	-	1.56E+07	6.83E+07
ING2	-	-	2.44E+07	-	-	8.92E+07
EZH2	5.36E+07	8.01E+08	1.09E+08	-	4.17E+07	1.41E+07
CBX7	-	-	-	-	-	-
KDM5B	2.71E+07	2.61E+07	1.79E+08	-	2.39E+06	1.06E+07

D)
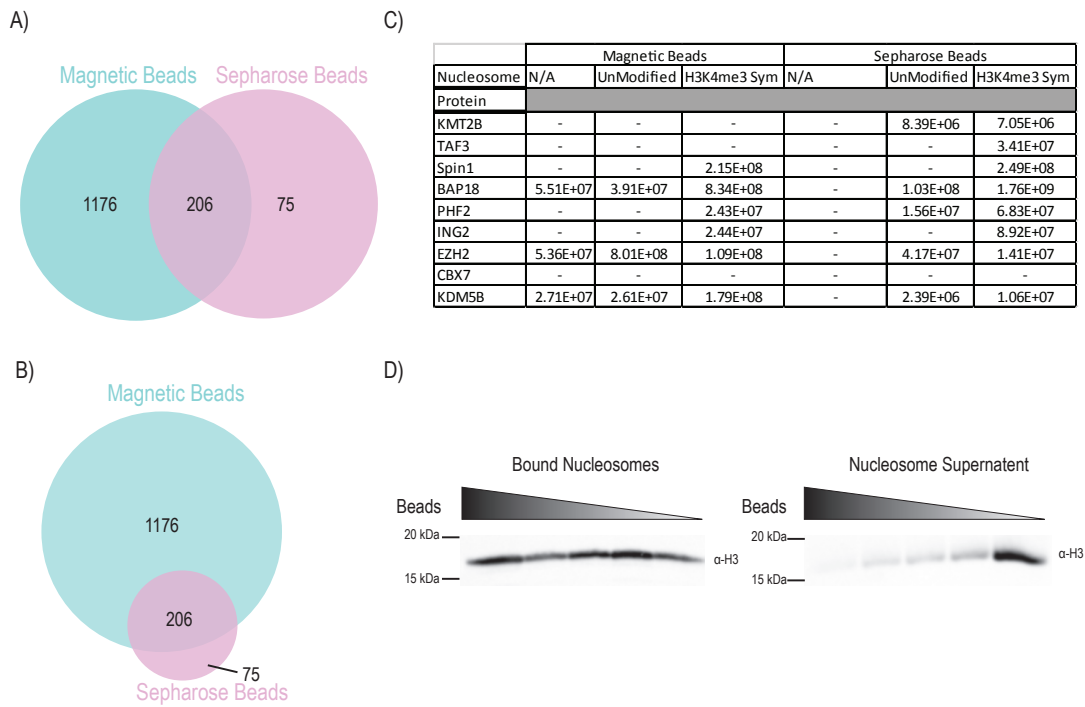
Figure 4.3 *Comparison of Myone C1 and Streptavidin Sepharose High Performance beads.* *Representational (A) and proportional (B) Venn Diagrams showing the overlap of proteins bound to both types of beads. Proteins identified with a high confidence and peptides ≥ 2 included in the analysis. C) Table detailing LFQ values for proteins bound to unmodified and H3K4me3/3 nucleosomes bound to the different beads. D) Western blots (WB) using the presence of H3 bound to the beads and present in the supernatant to indicate the amount of bound and unbound nucleosomes, respectively. Range of bead bed volume used, 4 μL to 12 μL.*

reducing the amount of proteins present in the elution that are bound to the DNA only. The use of the minimal amount of beads necessary for binding the nucleosomes results in fewer bead surface area being available for non-specific protein interactions.

4.4 Label-Free Quantification of Pulldown Assays via Mass Spectrometry

There are two main classes used in mass spectrometry (MS) based quantitative proteomics: label-free and label-incorporated MS. There are multiple well-established protein labelling techniques from TMT (Tandem Mass Tag), SILAC to iTRAQ (Isobaric Tags for Relative and Absolute Quantitation) protocols (Bantscheff et al., 2007; Patel et al., 2009). Both TMT and iTRAQ rely on isobaric labelling of molecules. In isotope-labelling approaches, the differentially labelled samples are combined after immunoprecipitation or purification, eluted, combined, and taken through the sample preparation steps for MS together. Thus, any differences in sample handling after combining individual samples affect all the samples equally. Chemical labelling occurs later on in sample handling and thus has reduced robustness compared to that seen in isotope-labelling. Although protein labelling techniques are well established and have definite benefits, they also have limitations including increased complexity of sample preparation, increased sample preparation time, increased cost per sample, and issues associated with incomplete labelling.

Label free (LF) mass spectrometry is a method that aims to determine the amount of a protein in two or more samples (in comparison with each other) without using a label that is chemically incorporated into the proteins of interest (Cox et al., 2014; Zhu et al., 2010; Wierer and Mann, 2016; Old et al., 2005). A drawback of LF MS is that as no label is present and each sample is prepared separately and run sequentially, differences in sample handling have greater effects on the samples compared to label-incorporated techniques. Label-free MS is a faster technique and allows a greater number of sample comparisons than label incorporated techniques, as it is not limited by the number of available labels. Label-free mass spectrometry was employed here as it requires no labelled media – the incorporation of amino-acid labels could affect protein expression in mouse embryonic stem cells (ESCs) affecting the results seen.

89

In addition, LF-MS allowed comparison of pulldowns performed with all the modified nucleosomes in the study against each other. Label-free MS has been shown to outperform TMT and iTRAQ techniques in terms of proteome coverage and protein sequence coverage respectively (Megger et al., 2014; Latosinska et al., 2015). SILAC and TMT labelling would only allow for triplet and ≥ 8 sample comparisons, respectively, while increasing the cost of each experiment and the time required for analysis. Label-free MS results can be evaluated by relative or absolute (iBAQ) protein quantification techniques. We utilised relative protein quantification techniques to assess the change in protein levels between the samples.

Protein quantification from LF-MS can be completed one of two ways: spectral counting or peak intensity. Peak intensity was used for relative protein quantification in this study. Previous studies have shown that peak area linearly correlates with the abundance of protein (R_2=0.991) (Chelius and Bondarenko, 2002). The data obtained from the LFQ MS was analysed for relative protein quantification between all the samples by the MaxLFQ algorithm as described by Cox et al. (2014). This algorithm relies on delayed normalisation using a dominant population of proteins that change nominally between samples. This allows the MS spectra of the different samples to be aligned and peptides to be quantified with high confidence. All samples must contain the majority of the same proteins to allow correct alignment of spectra between samples. The software takes all the information from peptide ratios between samples without resorting to arbitrarily assigning signal values to a peptide if its signal is not detected.

Pulldown samples underwent LC-MS on a Q-Exactive mass spectrometer. The in-line chromatography column separates the peptides in a sample by their chemical properties resulting in hydrophilic peptides eluting earlier in the gradient. The eluted peptides are ionized before entry into the MS and the m/z ratio of the peptides is determined via the Oribtrap (MS^1). The peptides are further fragmented in the collision cell. These peptide fragments re-enter the Orbitrap and their new m/z ratio is obtained as MS^2. Peptide and protein identification was based on MS^2 scans.

The proteins eluted from the beads are analysed by relative protein quantification of label-free mass spectrometry (LFQ-MS) samples (Cox et al., 2014; Cox et al., 2011; Cox and Mann, 2008). All MS experiments were performed in triplicate and we only considered proteins identified with a high level of confidence (number
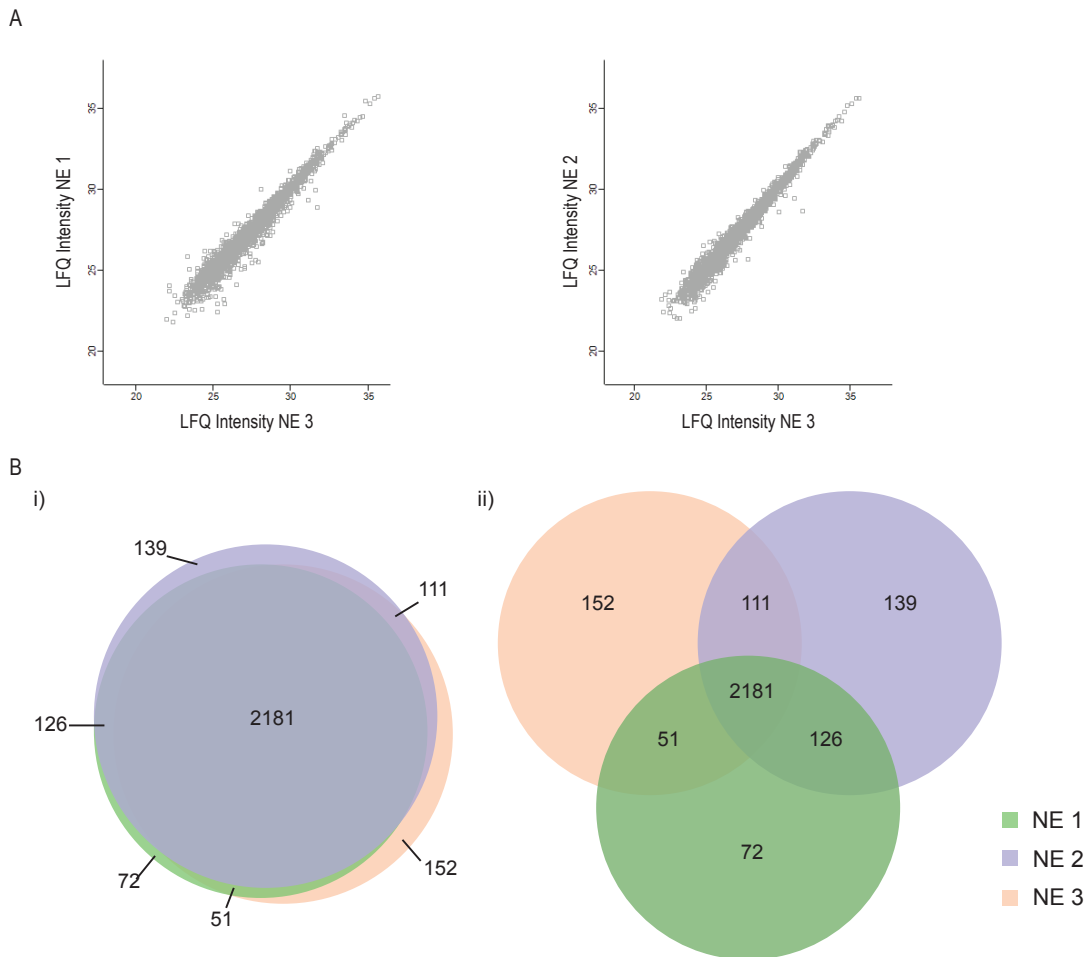
Figure 4.4 *Comparison of Nuclear Extract Biological Replicates. A) Scatter graphs generated by the Perseus program (version 1.6.5.0) comparing the enrichment seen for each protein shared between biological replicates. B) Venn diagram comparing the proteins present in all three replicates of NE preparations i) A proportional Venn diagram. ii) A representational Venn diagram. Venn diagrams generated online by Biovenn and Jvenn respectively. Proteins identified with a high confidence ≥ 2.*

of peptides greater than two).

The Nuclear Extract (NE) used for the pulldowns was combined from three replicates using a high salt extraction protocol (Conaway et al., 1996; LeRoy et al., 1998). The NE replicates were combined in order to average out differences between the replicates and obtain a mixture of proteins more representative of those present in an ideal nuclear extraction *in vivo*. The variation between NE replicas was assessed via LFQ-MS (Fig. 4.4). The relative protein intensities between NE samples was visualised with scatter graphs and at lower LFQ intensities the proteins present in the compared NE samples diverge more, due to the LFQ intensities approaching the lower limit of detection. The majority (87%) of proteins in the NE samples were present in at least two of the three biological

replicates showing high reproducibility. The MS analysis of the NE showed the presence of some mitochondrial proteins in the samples. This was likely due to some mitochondria co-sedimenting with the nuclei and thus mitochondrial proteins being extracted along with the proteins present in the nucleus. As these proteins would not normally be present in the nucleus, unless the cell is under stress or undergoing differentiation (Lionaki et al., 2016; Cardamone et al., 2018), they represent a potential source of false positives and were discounted from further analysis.

Technical replicates LFQ showed minimal change between samples run sequentially, and at least 94% of all proteins were shared between two of the three and 86% in all three technical replicates of pulldowns with unmodified nucleosomes (Fig. 4.5). This reproducibility remained similar between the technical replicates of the modified nucleosomes - H3K4me3/3 pulldowns are shown as an example. The variance of protein abundance increases at the lower range of abundance. The high reproducibility of technical replicates demonstrates that little protein variation should occur due to technical variation and therefore subsequent pulldown samples were only analysed as single LC-MS runs without technical replicates. Replicates using the same combined NE, but with fresh preparation of histones and nucleosomes, with experiments performed on different days are referred to as biological replicates. Biological replicates of the same pulldown conditions shared fewer proteins between all replicates - 92% - although at least 82% of proteins were shared between two of the three replicates (Fig. 4.6). The difference between the protein abundance between the biological replicates was greater than that of the technical replicates. This was expected due to the presence of technical variation along with biological and experimental variation. A comparison of the proteins bound between unmodified and H3K4me3/3 nucleosomes showed that the vast majority (84%)of proteins identified were the same between the differently modified nucleosomes (Fig. 4.7) and thus relative protein quantification of LFQ MS methodology can be used in conjunction with these pulldown experiments.

Once the data had been obtained and analysed using the MAXLFQ algorithm, further quality control measures were taken. Peptides that were only identified by MS_1 were removed. Potential contaminants, such as keratin, and false positives were removed after applying the false discovery rate (FDR) algorithm from MaxQuant. Only proteins identified with a high confidence by MS were included in the analysis (peptides $\geq$ 2). Proteins not present in 2 of the 3 replicates
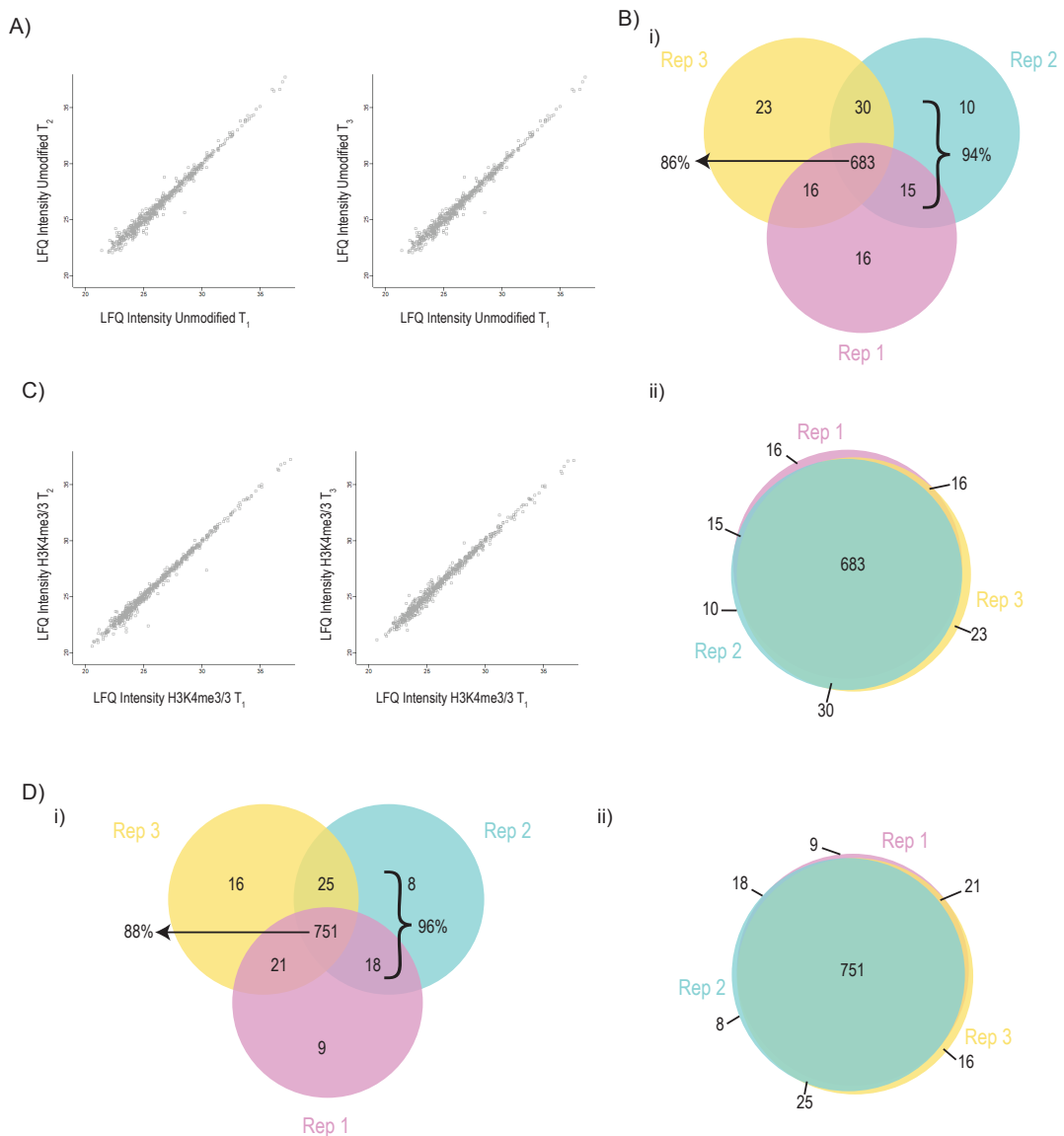
Figure 4.5 *Analysis of Mass Spectrometry Technical Replicate Runs A) Scatter plots comparing LFQ intensity of each protein between technical replicate runs of a pulldown with unmodified nucleosomes. B) Proportional (ii) or representational (i) Venn diagram of proteins present in each technical replicate. C) Scatter plots comparing the LFQ intensity of proteins shared between technical replicates of a pulldown with H3K4me3/3 nucleosomes. D) Venn diagrams comparing the total number of proteins shared between H3K4me3/3 replicates either proportional (ii) or representational (i). Venn diagrams generated online by Biovenn and Jvenn respectively. All scatter plots generated by Perseus (version 1.6.5.0). Proteins identified with a high confidence (peptides$\geq$ 2) are shown.*
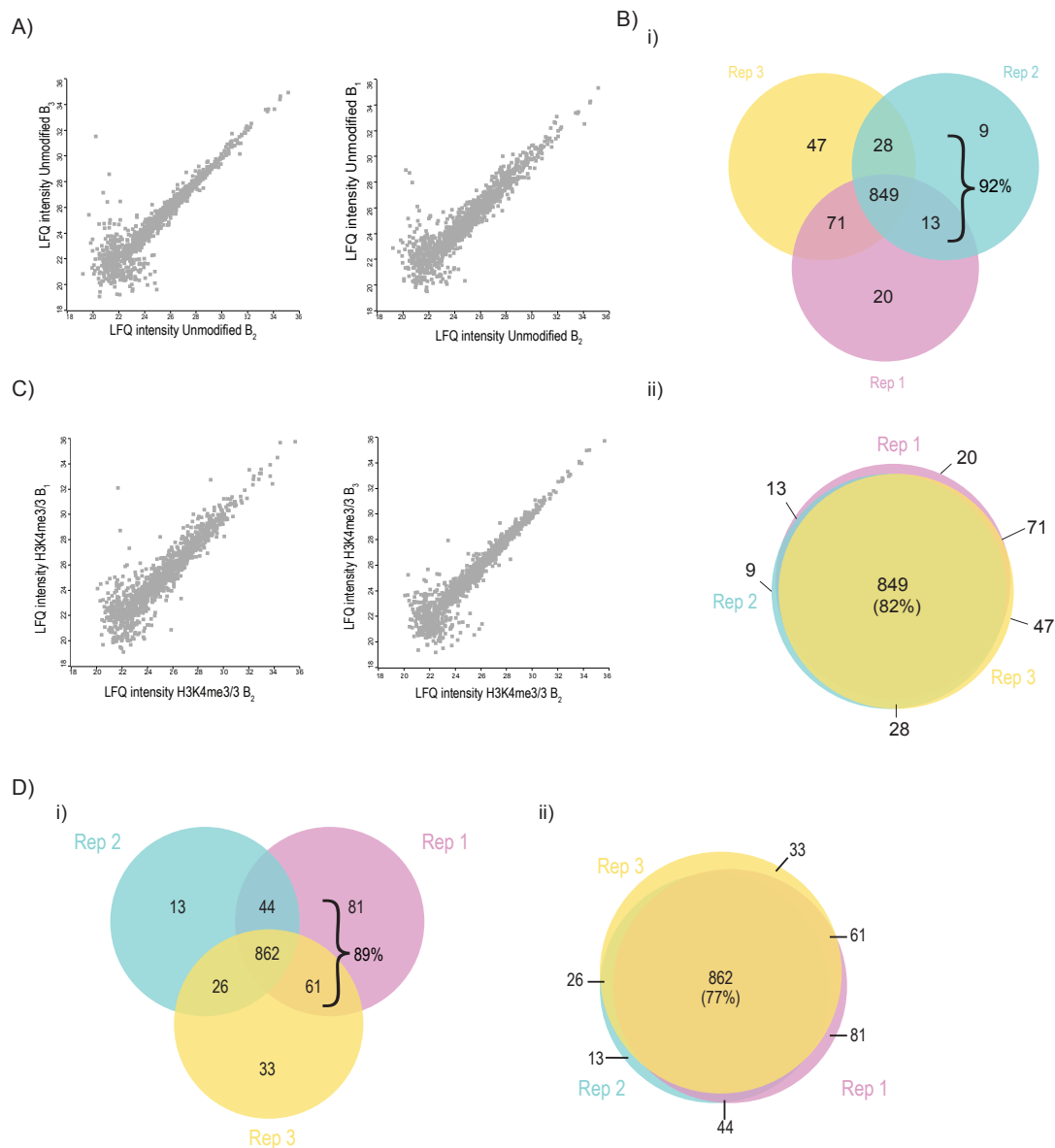
Figure 4.6 *Analysis of Mass Spectrometry Biological Replicates* A) Scatter plots comparing LFQ intensity of each protein between biological replicates of a pulldown with unmodified nucleosomes. B) Venn diagram of proteins present in each biological replicate either proportional (ii) or representational (i) diagrams. C) Scatter plots comparing the LFQ intensity of proteins shared between biological replicates of the pulldowns with the H3K4me3/3 nucleosomes. D) Venn diagrams comparing the total number of proteins shared between H3K4me3/3 nucleosome replicates either proportional (ii) or representational(i). Venn diagrams generated online by Biovenn and Jvenn respectively. All scatter plots generated by generated by the Perseus program (version 1.6.5.0). Proteins identified with a high confidence and with (peptides≥ 2) are included in the analysis.
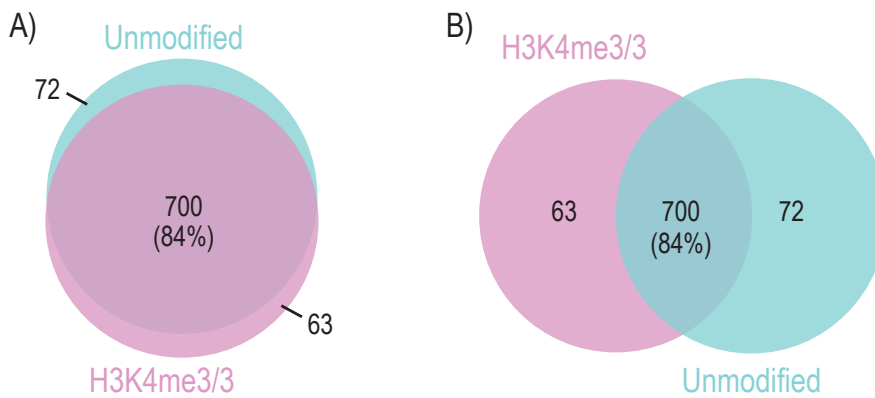
Figure 4.7 *Comparison Between Differently Modified Nucleosomes.* *A) Proportional and B) representational Venn diagram of proteins present in pulldowns with both unmodified and K4me3/3 modified nucleosomes. 84% of proteins were present bound to both types of nucleosomes, allowing identification and analysis of protein abundance and comparison between nucleosome pulldowns using LFQ methodology. Venn diagrams generated online by Biovenn and Jvenn respectively. Proteins identified with a high confidence (peptides ≥ 2) are shown.*

performed were removed to exclude proteins present due to sample handling variation.

Additionally, digestion methods were compared in terms of protein identifications. The different on-bead digestion protocols affected the proteins' abundance negligibly. Protein abundance was most altered between FASP digestion and any on-bead digestion duration tested (Fig. 4.8), nevertheless, approximately 78-82% of the proteins identified were common between digestion methods. The similarity of protein abundance between the differently digested samples is depicted via scatter plots (Fig. 4.8). Although the shared proteins between FASP and on-bead digestion samples exhibited a greater variation in abundance than other methods, there is a positive correlation and the majority of proteins present overlap between samples. On-bead digestion was selected as the optimal digestion method as it detected the presence of several proteins at higher levels, enabling a more accurate quantification compared to the FASP digestion method. This should remove a number of possible false positives from the dataset. On-bead digestion does not require SDS to elute proteins bound, thus reducing the likelihood of MS contamination. Analysis of varying incubation times with trypsin in on-bead digestion protocol showed very little effect, with 2 hours of incubation being chosen due to being more time efficient. FASP digestion method (Wiśniewski et al., 2009) was compared to the on bead digestion method as FASP allows the optimal removal of contaminants such as detergents, however, the on-bead digestion did not result in a significant amount of contaminates and could be used instead of FASP.

95
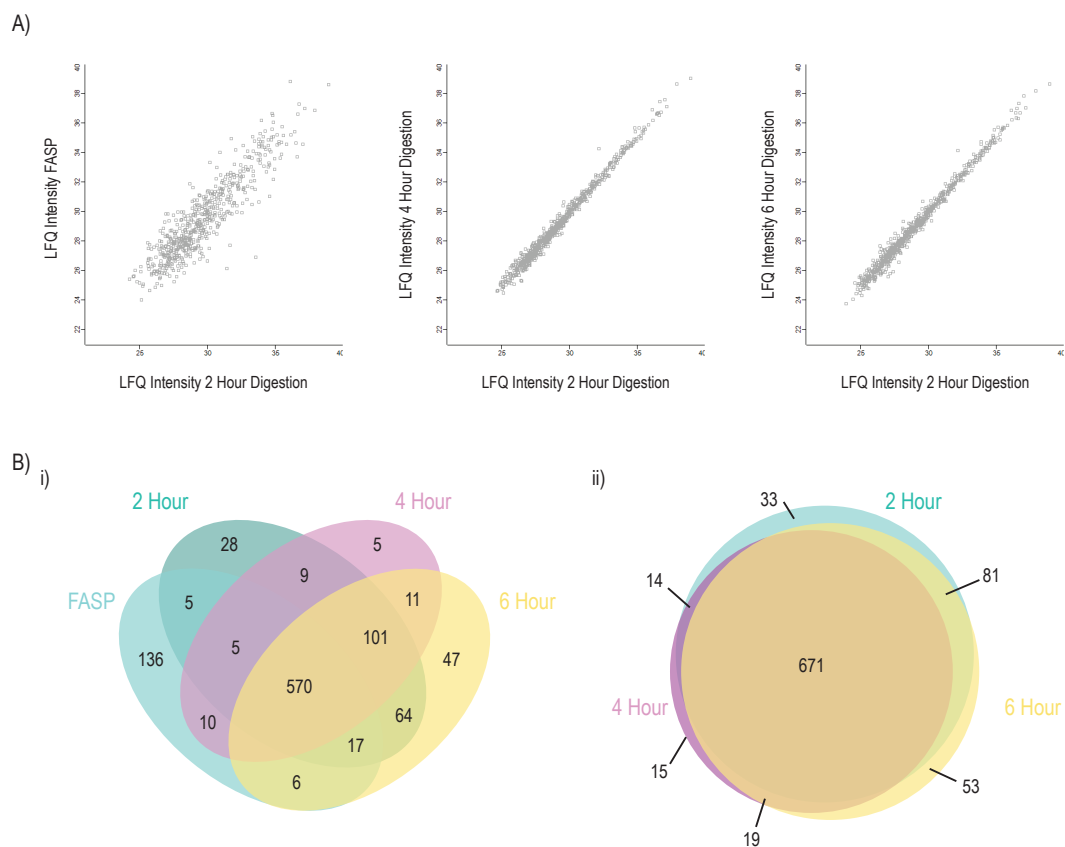
Figure 4.8 *Mass Spectrometry Digestion Optimisation. A) Scatter graphs comparing protein abundance in various MS digestion techniques. B) Venn diagram showing overlap of proteins present in all digestion methods (i) and a comparison of only the on-bead methods (ii). Venn diagrams generated online by Jvenn. All scatter graphs generated by generated by the Perseus program (version 1.6.5.0). Proteins identified with a high confidence (peptides≥ 2) are shown.*

4.5 Pilot Pulldown Experiments with Histones Containing Methyl-lysine Analogues (MLAs)

Initially methyl-lysine analogs (MLAs) were used as substitutes for histones containing specifically modified lysines. MLAs are widely used in place of endogenously modified histones. They contain a N-methyl aminoethylcysteine in place of the methylated lysine - and behave similarly to methylated lysines (Kenyon and Bruice, 1977; Simon et al., 2007; Lauberth et al., 2013; Francis et al., 2009; Margueron et al., 2009). In spite of aminoethylcysteine's similarity to lysine, as outlined below, early pulldown experiments showed an inability of proteins present in the nuclear extract to distinguish between differently modified nucleosomes (see Fig. 4.9).

PRC2 places the H3K27me3 mark on histone H3 and has been shown to exhibit a preference for H3K27me3/3 modified nucleosomes (Bartke et al., 2010). EZH2 is the catalytic subunit of this complex and was used as a proxy for PRC2 binding. The binding of SUZ12-RbAp48/46 submodule to the H3 N-terminal tail is decreased by 100 fold in the presence H3K4me3, thus decreasing PRC2's affinity for H3K4me3 modified N-terminal tails (Schmitges et al., 2011; Juan et al., 2016). If the complex was able to distinguish between H3K_c27me3 and H3K_c4me3 containing nucleosomes in this assay, EZH2 should be present in the highest quantities bound to H3$_c$K27me3 symmetrically modified, then unmodified nucleosomes, and depleted in H3$_c$K4me3 symmetrically modified nucleosomes. However, as shown in Figure 4.9, EZH2 bound equally well to all modified nucleosomes investigated, including those containing H3K_c4me3. The difference in PRC2's binding compared to previous studies could be due to use of MLAs. Although N-methyl aminoethylcysteines are comparatively structurally and chemically similar to methyl lysine they are slightly different, substituting the lysine -methylene with a sulphide, resulting in a slight lengthening of the side chain (0.28 Å). Moreover, the electron withdrawing effect of the thioether causes a small increase in the acidity (-1.1 pKa unit) of the ammonium protons (Gloss and Kirsch, 1995; Simon et al., 2007; Hopkins et al., 2005).

Native methyl-lysines have a 5-13 fold tighter binding in comparison to MLAs, quantitatively demonstrating the difference between these substrates and indicating the potential issues with utilisation of MLAs instead of native chemical lysines (Munari et al., 2012; Seeliger et al., 2012). Some proteins including

97
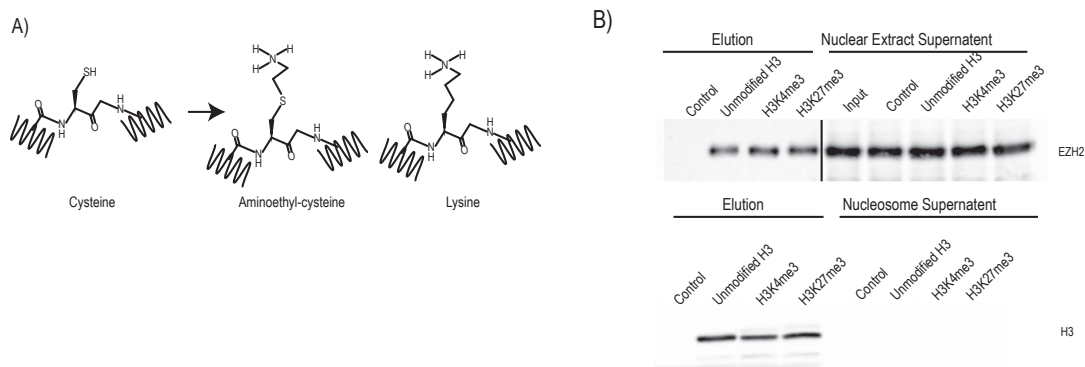
Figure 4.9 *Methyl-lysine Analogs A) Methyl lysine analogs are made by alkylating unique cysteine residues to form aminoethylcysteines. Comparison between cysteine, aminoethyl-cysteine, and lysine B) Western blot analysis showing EZH2 bound to three of the four different nucleosomes. Control - agarose beads with no nucleosomes bound. Unmodified - beads with unmodified (H3K$_c$4me0 H3K$_c$27me0) nucleosomes bound. H3K$_c$4me3 - beads with nucleosomes symmetrically modified with H3K$_c$4me3 bound. H3K$_c$27me3 - beads with nucleosomes symmetrically modified with H3K−c27me3 bound. Equal amounts of nucleosomes were bound to beads as shown by Western blotting against histone H3. Comparison of H3 present in nucleosome supernatant to beads to show complete binding of nucleosomes to beads in all conditions. EZH2 in the nuclear extract used as a loading control as it is still present in the nuclear extract supernatant.*

LEDGF are unable to bind the MLAs with equivalent affinity as the native modified nucleosome (Chen et al., 2014). The preference of PRC2 for H3K27me3 modified nucleosomes was established using modified nucleosomes generated by native chemical ligation (NCL)(Bartke et al., 2010). Due to the inability of PRC2 to distinguish between different MLA substrates in pulldown experiments (See Fig. 4.9), it was decided to change the method of *in vitro* synthesis of specifically modified nucleosomes to native chemical ligation (NCL). All further pulldowns were completed using specifically modified histones generated via the NCL method as described in Chapter 3.

4.6 Validation of Pulldown Assays with Symmetrically Modified Nucleosomes

Once the NCL method of generating specifically modified histones had been selected and the nucleosomes made, the *in vitro* pulldown approach had to be validated. To validate the robustness of the chosen approach, a comparison of proteins bound to T32C unmodified, H3K4me3/3, H3K27me3/3 nucleosomes was undertaken. The proteins that bind to these modifications or are unable to bind them have been previously demonstrated by multiple studies (Makowski et al.,

2018; Bartke et al., 2010; Vermeulen et al., 2010; Eberl et al., 2013). Initially proteins that have known binding preferences such as TAF3, CHD1, EED, CBX7 were analysed via western blot to confirm the assays validity before proceeding to the MS analysis (Vermeulen et al., 2007; Ingen et al., 2008; Lauberth et al., 2013; Flanagan et al., 2005; Xu et al., 2010; Morey et al., 2012).

LFQ-MS was utilised to analyse differential binding of proteins bound to the differently modified nucleosomes. Algorithms from Perseus (version 1.6.5.0) and R (version 3.5.3) were utilised to create volcano plots comparing the proteins bound to unmodified, H3K4me3/3 and H3K27me3/3 nucleosomes (Fig. 4.10). Proteins that have a 3-fold or greater change with a p-value of p<0.01 were considered significantly enriched and highlighted in Fig. 4.10. However, some proteins are missing from the comparison either due to not being present in the NE or only being present in one of the three pulldown replicates. To allow comparison of proteins present in one condition which are absent from another (due to their inability to bind in the other condition), missing LFQ intensity values were imputed so a pairwise analysis of the two nucleosome conditions could be completed. This was accomplished via two ways: initially from normal distribution in Perseus and subsequently when re-analysed, using the minprob function in DEP 1.4.1 package (Zhang et al., 2018).

The majority of the proteins significantly enriched bound to H3K4me3/3 (H3K4me3 symmetrically modified) nucleosomes consists of proteins already known to bind H3K4me3 including: CHD1, Spin1, CFP1 and BPTF (Flanagan et al., 2005; Li et al., 2006; Ruthenburg et al., 2011; Clouaire et al., 2012; Brown et al., 2017; Eberl et al., 2013; Su et al., 2014; Wang et al., 2011; Yang et al., 2012). The proteins enriched in H3K4me3/3 nucleosome pulldowns are known to bind to H3K4me3 directly (TAF3) or are part of complexes which do so (DPY30) (Vermeulen et al., 2007; Ingen et al., 2008; Lauberth et al., 2013; Tremblay et al., 2014; Bochyńska et al., 2018).

Although there are numerous protein complexes known to bind to H3K4me3 including MLL2, TFIID and SETD1A which are present in the NE, are not enriched on the H3K4me3/me3 nucleosome. Many of these complexes contain binding sites for other histone modifications and share subunits with multiple complexes (Sze et al., 2017; Clouaire et al., 2012; Bledau et al., 2014; Vermeulen et al., 2007; Ingen et al., 2008; Lauberth et al., 2013; Aloia et al., 2013; Denissov et al., 2014; Eberl et al., 2013; Ali et al., 2014). This may result in the complexes being recruited to nucleosomes with a variety of modifications and therefore not
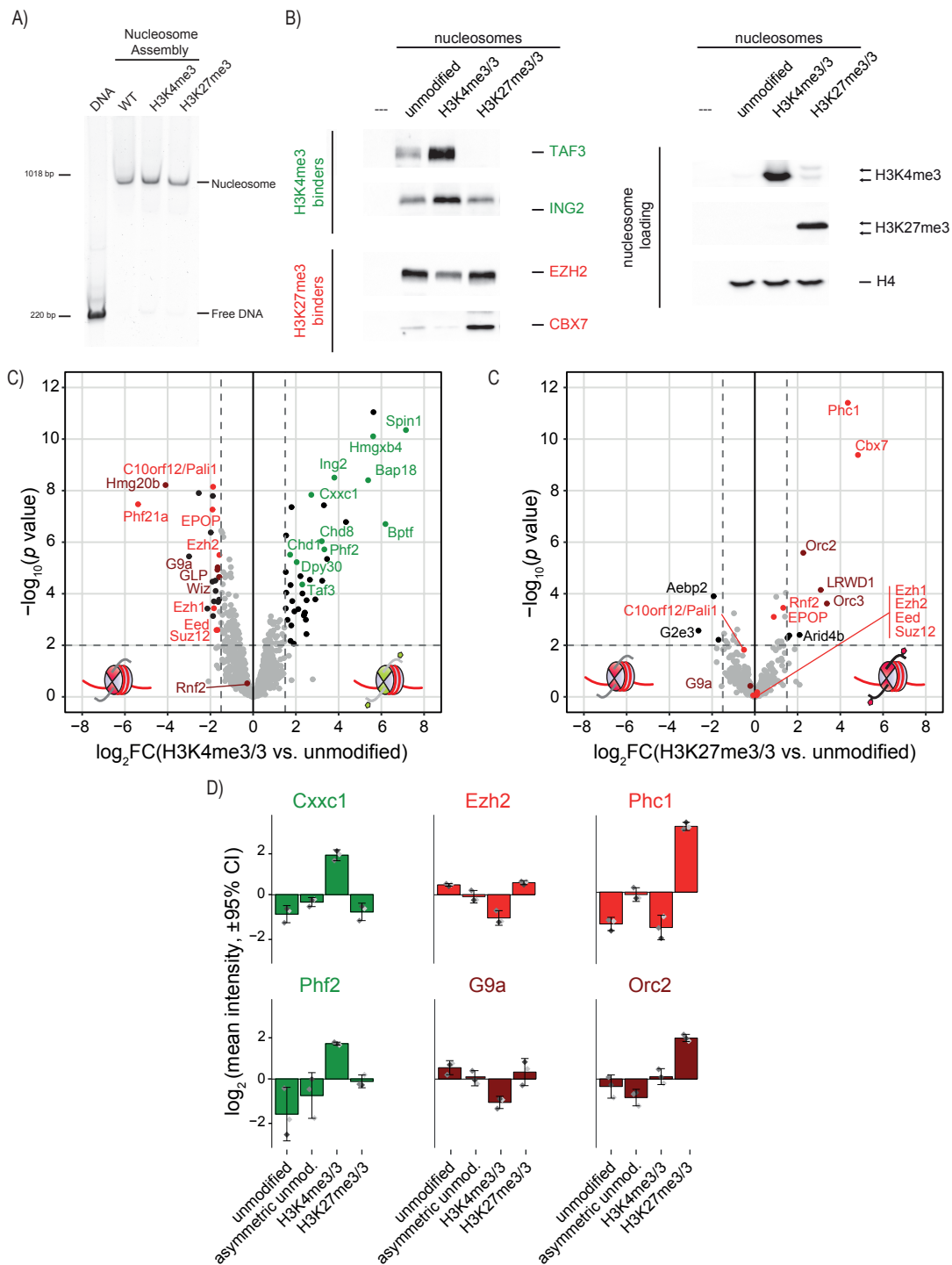
Figure 4.10 *Validation of Pulldown Approach using Symmetrically Modified Nucleosomes* A) 6% TGE showing the assembly of nucleosomes unmodified, H3K4me3/3, and H3K27me3/3 nucleosomes. B) Western blot analysis of proteins bound to unmodified, H3K4me3/3, and H3K27me3/3 nucleosomes. Proteins known to bind to H3K4me3 labelled in green, proteins known to bind H3K27me3 in red. H3K4me3, H3K27me3 and H3 shown to confirm loading order and as loading controls. C) Volcano plots showing proteins enriched in pulldowns with modified nucleosomes compared to unmodified nucleosomes. D) Bar graphs of individual protein enrichment in various modified nucleosome conditions. Proteins known to bind to H3K4me3 shown in green, proteins known to bind H3K27me3 in red. Volcano plots were created via R version 3.5.3 and individual protein enrichment was calculated with the DEP 1.4.1 package.

being classified as enriched on any one nucleosome. These complexes may not be able to stably bind the nucleosome in the pulldown conditions used such as the use of mononucleosomes instead of arrays or due to the longer length of washes utilised in this study - this is discussed in more detail later. The presence of subunits shared between complexes means that the presence of their shared subunit such as DPY30 or CFP1 does not give a clear indication of which specific complex is being recruited. Indeed, proteins that directly bind a singular modification may still be recruited to nucleosomes containing other modifications due to interactions with subunits that have different binding preferences.

The proteins enriched in the H3K27me3/me3 nucleosomes consist of proteins previously identified as H3K27me3 interactors including: CBX7, Polyhomeotic Homolog 1 (PHC1), G9a, ORC proteins and LRWD1 (Bartke et al., 2010; Vermeulen et al., 2010; Eberl et al., 2013). These are proteins that bind directly to H3K27me3 such as CBX7 (Bernstein et al., 2006b; Morey et al., 2012; Min et al., 2003), or are proteins that interact with proteins that do so (PHC1, G9A, and LRWD1) (Mozzetta et al., 2014; Shinkai and Tachibana, 2011; Tachibana et al., 2001; Tachibana et al., 2002; Illingworth, 2019; Giri and Prasanth, 2015; Vermeulen et al., 2010). Polyhomeotic Homolog 1 (PHC1) is expressed in ESCs and is part of the CBX7 canonical ESC PRC1 complex that mediates H2AK119Ub. CBX7 recruits PRC1 to areas of chromatin containing H3K27me3, therefore its strong enrichment in H3K27me3 modified condition similar to CBX7 is expected (Bernstein et al., 2006b; Morey et al., 2012). The other CBX variant expressed in ESCs, although a low levels, CBX6 is able to be recruited to chromatin independently of H3K27me3 (Santanach et al., 2017; Bernstein et al., 2006b) and therefore its lack of enrichment is expected. There were some proteins enriched in the H3K27me3 condition compared to the unmodified condition such as ORC proteins and LRWD1 which has been shown in previous studies to be enriched at H3K27me3 modified tails (Bartke et al., 2010; Vermeulen et al., 2010; Eberl et al., 2013). G9A places H3K9me2 which often co-localises with H3K27me3 *in vivo* and has been shown to directly interact with PRC2 although it does not bind directly to H3K27me3 itself (Mozzetta et al., 2014; Shinkai and Tachibana, 2011; Tachibana et al., 2001; Tachibana et al., 2002). Unsurprisingly G9A's behaviour in the nucleosome pulldowns closely follows that of the PRC2 subunits.

The comparison of H3K27me3 modified nucleosomes with unmodified nucleosomes show fewer proteins enriched in either condition and to a lower extent

compared to the H3K4me3/3, unmodified nucleosome comparison (Fig. 4.10). Many of the proteins associated with H3K27me3, although potentially excluded from binding H3K4me3, are part of complexes that can bind and be recruited to multiple histone marks. PRC2 is able to bind both H3K27me3 and H3K27me0, therefore, PRC2 subunits such as EZH2 and EED are present bound to both H3K27me3/3 and unmodified nucleosomes (see Fig. 4.10). Any differences in PRC2 bound to H3K27me3/3 and unmodified nucleosomes is small under the conditions used in these pulldown experiments.

Western blot analysis was used to validate the LFQ-MS results. Several proteins were chosen that are enriched on each nucleosome type to act as controls for protein binding. TAF3 was shown to be enriched bound to H3K4me3/3 nucleosomes via LFQ-MS analysis. TAF3 is the subunit of general transcription factor complex TFIID that binds directly to H3K4me3 and recruits TFIID, an indicator for a transcriptional permissive chromatin state (Ingen et al., 2008; Lauberth et al., 2013; Vermeulen et al., 2007). TAF3 (140 kDa) binds directly to H3K4me3 and is suitable for western blot analysis (Vermeulen et al., 2007; Ingen et al., 2008; Lauberth et al., 2013). TAF3 was therefore used as a positive control for proteins binding solely to H3K4me3. Western blot analysis shows TAF3 bound highly to H3K4me3 modified nucleosomes confirming results seen in the LFQ-MS analysis. The presence of some TAF3 bound to unmodified nucleosomes could be due to its recruitment by its various interaction partners, for example other subunits of the TFIID complex.

An additional positive control for H3K4me3 binding was used – ING2. ING2 binds H3K4me3 directly (Peña et al., 2006; Shi et al., 2006). However, it is part of multiple protein complexes, for example variants of the Sin3A complex, which contain multiple binding sites for all core histones H2A, H2B, H3 and H4, potentially allowing for some recruitment to other nucleosomes (Viiri et al., 2009; Yoon et al., 2004; Murzina et al., 2008). ING2 is clearly enriched bound to the H3K4me3/3 nucleosomes via western blot analysis confirming the enrichment seen in LFQ-MS analysis (see Fig. 4.10). The binding of ING2 seen at differently modified nucleosomes may be due to its interactions with multiple protein complexes and their recruitment to core histones.

CBX7 binds to H3K27me3/3 modified nucleosomes (Fig. 4.10) and was therefore selected as a positive control for protein binding selectivity for H3K27me3 (Bernstein et al., 2006b; Morey et al., 2012). CBX7 is a subunit of the PRC1 canonical complex and helps recruit PRC1 to H3K27me3 sites in chromatin

102

(Wang et al., 2004b). It contains a chromodomain which binds H3K27me3 and Morey et al. (2012) and Min et al. (2003) have shown its specific binding to H3K27me3. CBX7 is one of the most highly expressed CBX variants in ESCs, is essential for maintaining stem cell pluripotency and has been demonstrated to be the main CBX in the PRC1 complex in ESCs (Santanach et al., 2017; Morey et al., 2012). As CBX7 (28 kDa) has well characterised antibodies it is therefore easily identified via western blot. The western blot shows CBX7 bound highly to H3K27me3/3 nucleosomes in comparison to unmodified nucleosomes confirming previously established results.

EZH2 was depleted from H3K4me3/3 nucleosomes (Fig. 4.10). Both Voigt et al., 2012; Schmitges et al., 2011 demonstrated that PRC2 is inhibited from modifying a H3 N-terminal tail previously modified with H3K4me3. H3K4me3 inhibits the binding of H3 N-terminal tails to RBAP48/NURF55, this may affect EZH2's catalytic activity at H3 N-terminal tails via an allosteric mechanism communicated via conformational changes in SUZ12, shown by Schmitges et al. (2011). There are three H3 N-terminal tail binding sites in the core PRC2 complex: EZH2's SET domain, EED's aromatic cage and RBAP48 (Margueron et al., 2009; Schmitges et al., 2011; Jiao and Liu, 2015). PRC2 can still bind to H3K4me3/3 nucleosomes; however, its binding to two of the three binding sites is decreased. H3K4me3 binding to the RBAP48 binding site is decreased by 100-fold (Schmitges et al., 2011; Chen et al., 2018), and H3K4me3 peptide shows no appreciable binding to the aromatic cage of EED (Margueron et al., 2009; Xu et al., 2010), resulting in decreased PRC2 binding to H3K4me3/3 nucleosomes (Wu et al., 2013; Jiao and Liu, 2015; Kasinath et al., 2018). Wang et al., 2017 also observed a decrease of PRC2 and thus EZH2 binding to H3K4me3/3 nucleosomes, thus EZH2 works as a negative control for binding to H3K4me3 nucleosomes.

The decrease of EZH2 in H3K4me3 nucleosomes was not seen in Bartke et al., 2010, and this difference in EZH2 binding may be explained by the different beads used and different cell types utilised in both experiments. From previous optimisation trials it was clear that the magnetic beads bound to more unspecific proteins, including EZH2, than the sepharose beads used in this experiment. The presence of EZH2 bound unspecifically to the beads could cause any slight decrease in protein binding to be classified as not significant. In addition, Bartke et al., 2010 used NE from HeLa cells while this study used NE from mouse ESCs, these different cell types may contain different PRC2 variants resulting in different binding affinities to different modifications (Kloet et al., 2016; Oliviero

et al., 2016).

EZH2's binding to differently modified nucleosomes was investigated with both LFQ-MS and western blot analysis. Examination of all nucleosomes showed a similar enrichment of EZH2 at H3K27me3/3 and unmodified nucleosomes and a decrease at H3K4me3/3 nucleosomes via both types of analysis. The similar abundance of EZH2 bound to both unmodified and H3K27me3/3 nucleosomes can be explained by the presence of both H3K27me3 and H3K27me0 binding sites within the PRC2 protein complex as previously discussed (Wang et al., 2017; Wu et al., 2013; Margueron and Reinberg, 2011; Laugesen et al., 2019; Holoch and Margueron, 2017). PRC2 and thus EZH2 can be recruited to both unmodified and H3K27me3 tails.

Another possible cause for the varying results between this study and Bartke et al., 2010 is the difference in the experimental protocols. While the protocol used in the Bartke et al., 2010 paper used flick washes, this study used 5 minute washes. The length of washes in this protocol results in a longer time between binding and elution of proteins from the beads. PRC2 is a large multisubunit complex which binds nucleosomes due to multiple low affinity interactions, this may result in a relatively large Koff. Koff is the first-order rate constant for the dissociation of the protein-ligand complex and is inversely related to the affinity of the protein for the ligand. Longer washes may result in the loss of proteins or protein complexes from the nucleosomes with larger Koff constants. Therefore the enrichment of PRC2 on H3K27me3/3 nucleosomes compared to unmodified nucleosomes seen in Bartke et al., 2010 may be lost via this protocol. Moreover, different mass spectrometry techniques were utilised in these studies, LFQ-MS compared to SILAC. SILAC may be better adapted technique to pick up smaller alterations in protein binding as it has more reliable protein quantification and reduction in sampling handling variance. Due to the increased effect of sample handling variances created in the label-free methodology, it is less sensitive to small protein changes and more likely to only identify larger protein alterations. There could also be issues with steric hindrance of large complexes so that they cannot bind to the modifications potentially due to beads being saturated with nucleosomes. PRC2 has been shown to bind to dinucleosomes (Poepsel et al., 2018), the additional nucleosome provides extra sites of interaction with the PRC2 complex potentially reducing the complexes' overall Koff.

Validation of the experimental approach by confirmation of known protein interactions with symmetrically modified nucleosomes via two different analyses

allows investigation into protein interactions with asymmetrically modified nucleosomes.

4.7 Conclusion

Initially the pulldowns used MLAs as they can be made in large quantities (Simon and Shokat, 2012; Simon, 2010). However, the candidate histone PTM binding proteins tested were unable to distinguish between the different modification states of MLA-containing nucleosomes, likely due to the differences between N-methyl aminoethyl-cysteine and methyl lysine residues. They were therefore discarded in favour of histones made via the NCL method, and these were subsequently used in all further pulldowns. Modified histones made using the NCL method are chemically identical to endogenously occurring histones apart from the T32C mutation.

The pulldown approach was optimised to maximise nucleosome assembly, minimise non-specific protein binding, and enhance experimental reproducibility. The protocol for MS was optimised to minimise detergent contamination, and to improve identification and quantification of proteins. Label-free quantification was successfully employed to compare the relative protein abundance between MS samples. The pulldowns with the symmetrically modified nucleosomes, H3K4me3/3 and H3K27me3/3, enriched for known H3K4me3 and H3K27me3 proteins binders respectively. The enrichment of proteins that are known to bind these marks from the previous studies helped demonstrate that the pulldown approach worked. The results clearly show that our experimental approach is successful, validating previously identified protein interactions, and enables further investigation into nucleosome binding proteins.

Chapter 5

Identification of Proteins that Bind to Asymmetrically Modified Nucleosomes

5.1 Introduction

Asymmetrically modified nucleosomes exist in different species and cells types (Voigt et al., 2012; Shema et al., 2016; Barski et al., 2007; Choudhury et al., 2019). Asymmetry within the cell can arise for multiple reasons: recruited histone modifying enzymes compete for binding sites resulting in asymmetric modifications and once one sister histone is modified it can be excluded from modification by other enzymes until that mark is removed, resulting in further asymmetry (Voigt et al., 2012; Schmitges et al., 2011; Liokatis et al., 2012; Pasini et al., 2010b). In fact, some enzymes prefer (*in vitro*) to modify only one copy of histone H3 in the nucleosome (Liokatis et al., 2012), making asymmetric nucleosomes by default. The relative abundance of asymmetric nucleosomes does not contain any information as to whether they have an important biological function, are present due to no significant recruitment of a single type of histone modifier, or are an intermediate state before becoming symmetrically modified.

The ability of nucleosomes to be asymmetrically modified opens up a larger number of combinatorial possibilities for histone modification readout and function. There are numerous protein complexes large enough to span more than

one histone N-terminal tail and that contain multiple binding sites for various histone modifications (Taverna et al., 2007; Rothbart et al., 2013; Ruthenburg et al., 2011). Protein complexes such as SAGA have been shown to be bind both H3 N-terminal tails from the same nucleosome (Ruthenburg et al., 2011; Li and Shogren-Knaak, 2009; Li and Shogren-Knaak, 2008), and some of those complexes involved in the placement of bivalent domains (PRC2) can bind one H3 N-terminal tail of one nucleosome and modify the tail of a neighbouring nucleosome potentially spreading asymmetry (Poepsel et al., 2018). The existence and abundance of asymmetric nucleosomes *in vivo*, and the effect on multiple enzymes' catalysis suggests that they result in a specific biological function.

H3K27me0/3 and H3K4me0/3 nucleosomes may be sufficient to recruit their cognate binding proteins and maintain the repressive or active state, respectively, potentially with altered binding to PRC2 or TFIID subunits, or they may be no different to unmodified nucleosomes in their protein recruitment. Although these asymmetric nucleosomes are known to exist (Voigt et al., 2012; Sen et al., 2016; Shema et al., 2016) how the singular asymmetric modification alters their biological function and the identity of proteins that bind them is still unknown.

5.2   Aims

The pulldown approach using specifically modified *in vitro* nucleosomes has been validated using proteins previously identified to bind H3K4me3/3 and H3K27me3/3. The next phase of the project aimed to discover how the asymmetry of histone modifications affects protein binding and the biological function of asymmetric nucleosomes. The effect of asymmetric modification on nucleosomes and subsequent protein binding will be investigated in this chapter. All abbreviations used refer to the nucleosomes as in Figure 3.1.

5.3   Comparison of Proteins Bound to Asymmetrically Modified Nucleosomes

Each H3 sister histone within the asymmetric octamer is tagged at its C terminus to allow purification (Fig. 4.1) and generation of asymmetrically modified nucleosomes. A comparison of the unmodified nucleosome to the unmodified
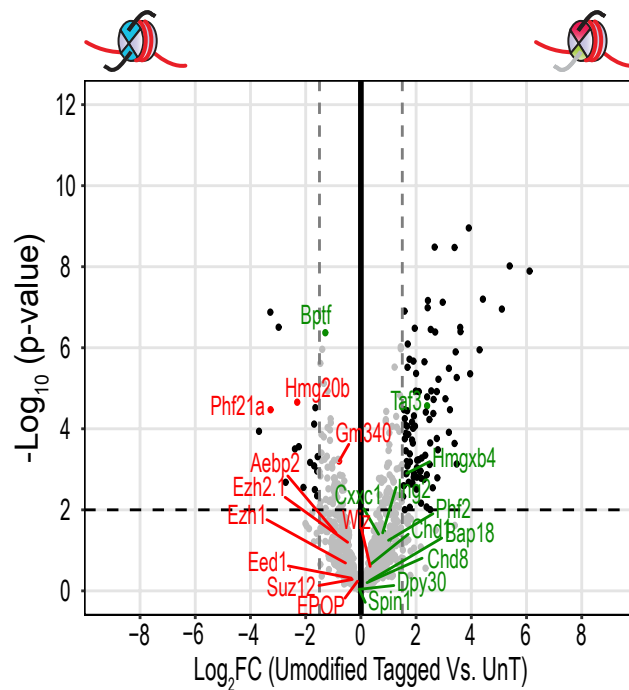
Figure 5.1 *The Effect of C-terminal Tags on Protein Binding.* *Volcano Plot comparing unmodified untagged (UnM UnT) and unmodified tagged (UnM H/S) conditions. Proteins enriched bound to Unmodified tagged on the left of the graph, proteins enriched bound to unmodified untagged on the right. Proteins significantly enriched (fold change ≤1.5 and adjusted p≤0.01) are shown in black, known H3K4me3 mark binders in green and repressive mark binders in red. Data from DEP suite analysis in R. Volcano plots were generated in ggplot2 in R (3.5.3.).*

tagged nucleosome allows proteins that are recruited specifically to the tags themselves or have altered binding due to the presence of the tags to be identified. Apart from TAF3, none of the proteins that bound to either H3K4me3 or H3K27me3 show altered binding in the presence of the tags (Fig. 5.1), verifying that the asymmetric nucleosomes with C-terminal tags on histone H3 are suitable tools to determine the difference in binding between symmetric and asymmetric nucleosomes. TAF3 showed increased binding to unmodified tagged nucleosome indicating that making comparisons of protein enrichment between samples the relative control must be taken into consideration.

The comparison of the asymmetrically modified nucleosomes with unmodified tagged nucleosomes reveals reduced protein enrichment compared to symmetrically modified nucleosomes in comparison with unmodified untagged nucleosome. Asymmetric conditions (Fig. 5.2) show lower overall enrichment of all proteins compared to symmetric samples. Greater protein enrichment for known mark binders is seen in the symmetric conditions compared to all other conditions (Fig. 5.2 & Fig. 5.3). Proteins known to be recruited to H3K4me3 or H3K27me3 marks
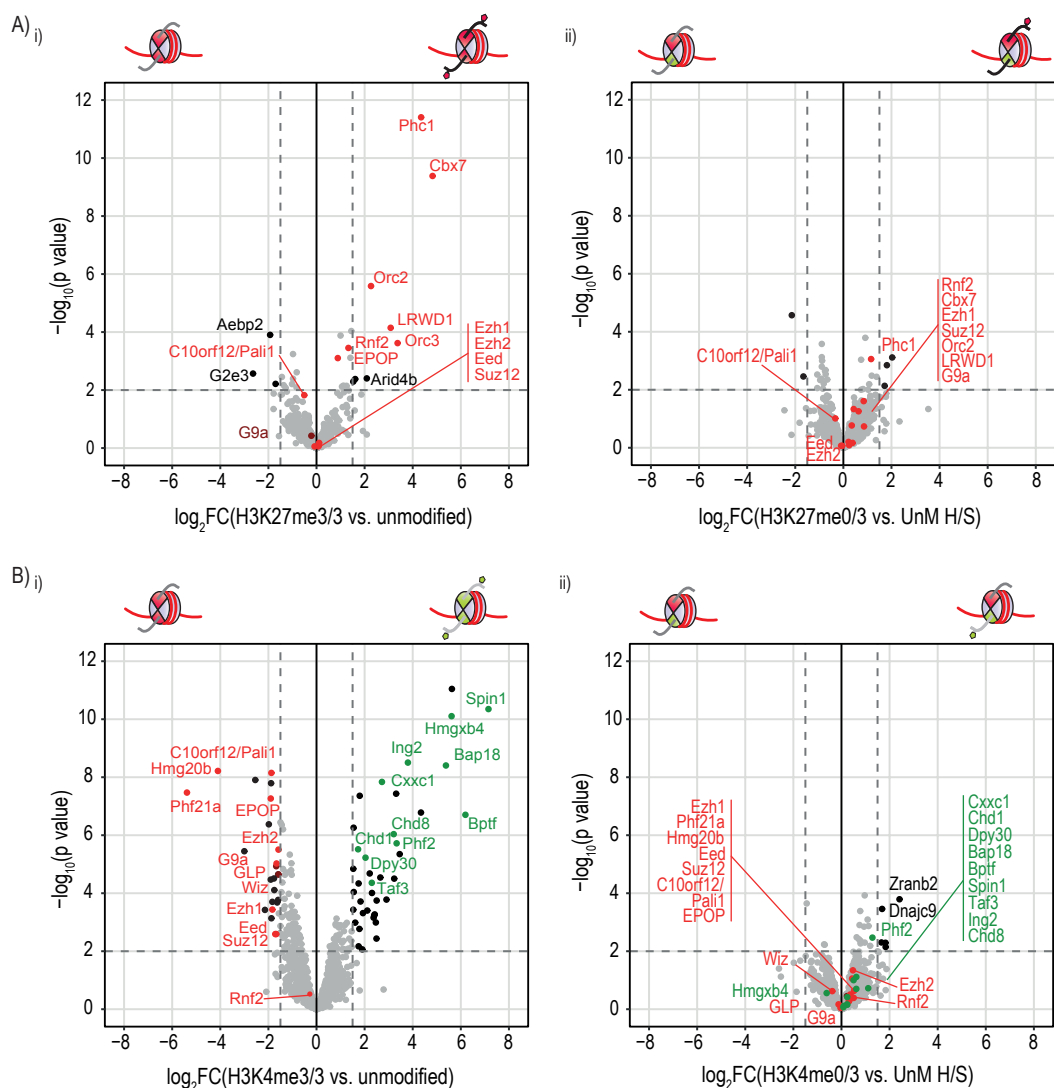
Figure 5.2 *Comparison of Symmetric and Asymmetric Nucleosomes. Comparison of symmetric H3K4me3/3 and H3K27me3/3 nucleosomes with unmodified untagged nucleosomes and H3K4me0/3 and H3K27me0/3 nucleosomes with unmodified tagged (UnM H/S) nucleosomes. Volcano plots comparing A)i) H3K27me3/3 and ii) H3K27me0/3 with their respective controls B)i) H3K4me3/3 and ii) H3K4me0/3 modified nucleosomes with their respective controls. Significant in black (fold change ≤1.5 and adjusted p ≤0.05). Proteins of interest labelled - "repressive" in red "active" in green. Data from DEP suite analysis in R version 3.5.3. Volcano plots were generated in ggplot2 in R (3.5.3.).*

110

are recruited to the symmetric nucleosomes in comparison to the unmodified nucleosomes, however, there is a significant decrease in enrichment on the asymmetric H3K4me0/3 and H3K27me0/3 nucleosomes (Fig. 5.3). This suggests that the presence of a singular asymmetric modification on the nucleosomes is not sufficient to recruit H3K4me3 or H3K27me3 binders.

The proteins enriched on both H3K4me3/3 and H3K4me0/3 nucleosomes compared to their respective controls indicates a large alteration in binding that is not due to the presence of tags. TAF3 is affected by the presence of the tags - it is enriched on the UnM H/S control relative to the unmodified untagged (UnM UnT) control, indicating that the difference in enrichment of TAF3 on H3K4me3/3 nucleosomes compared to H3K4me0/3 nucleosomes is probably greater in reality due to TAF3's increased affinity for nucleosomes carrying the tags present on the asymmetric histones. In summary, singular asymmetrically modified nucleosomes do not enrich for any specific protein mark binders compared to unmodified or symmetrically modified nucleosomes.

The lack of protein enrichment in the asymmetric samples may have multiple causes. Firstly, the proteins may require the conformation of marks present in a symmetric nucleosome to bind. Additionally, asymmetric conditions have half the amount of the mark compared to symmetric conditions – the resulting reduced concentration may not be sufficient for the protein to bind. Finally, the unmodified N-terminal tail present in the asymmetric nucleosome may disguise or compete out the enrichment of protein caused by the presence of the single mark. To investigate what combination of these potential causes is correct, nucleosomes containing one modified H3 and one tailless H3 copy were prepared. A comparison of proteins binding these truncated (tailless) nucleosomes, which contain one modified H3 and one tailless H3 in the same nucleosome, and the asymmetrically modified nucleosomes will demonstrate whether this decrease in protein binding is due to the presence of the unmodified N-terminal tail in the asymmetrically modified nucleosome. If this is the case, then the truncated asymmetric nucleosome will show similar protein binding as the symmetric sample. If the concentration of the modification linearly affects protein binding, then the truncated condition will bind proteins half as well as the symmetric condition. A titration of the amount of symmetric and asymmetric nucleosomes will elucidate whether the difference in protein enrichment is due to the concentration of the marks present in the differently modified nucleosome samples. Assuming that this is true, and the marks linearly affect protein binding,

111
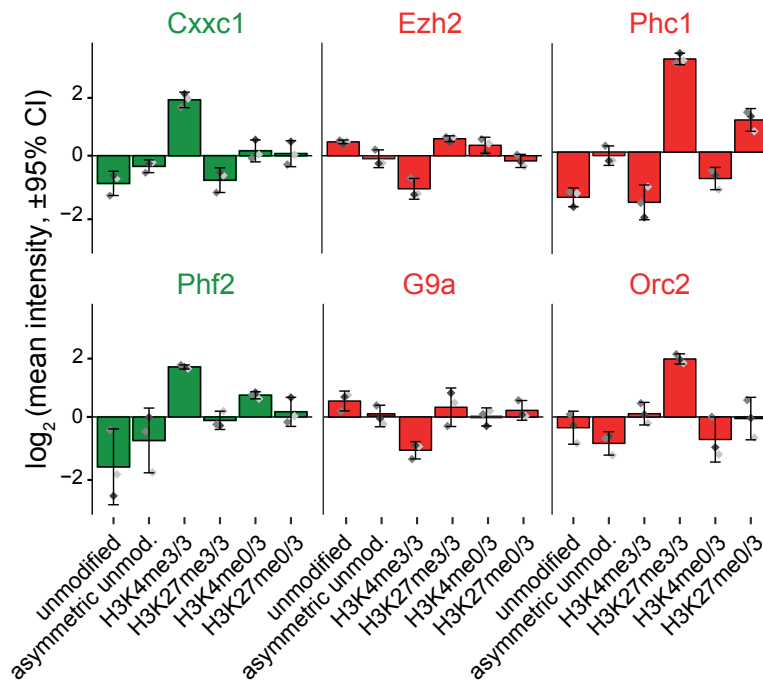
Figure 5.3 *Bar graphs showing enrichment of proteins in asymmetric and symmetric nucleosome conditions. Cxxc1 and Phf2 are known to be recruited to H3K4me3 and are labelled in green, while EZH2, Phc1, G9a and Orc2 are known to be recruited to H3K27me3 and are labelled in red. Neither H3K4me3 nor H3K27me3 mark binders are enriched on asymmetrically modified nucleosomes compared to symmetrically modified nucleosomes. Bar graphs are created by the DEP package (1.4.1) in R (3.5.3.).*

then the asymmetric nucleosomes should bind the same amount of protein as half the amount of symmetric nucleosomes.

Analysis of proteins bound to H3K4me3/3, H3K4me0/3 and H3K4me3/Tailless nucleosomes was used to investigate how active mark binders are affected by the presence of free H3 N-terminal tails, the conformation and concentration of marks/modifications present on the nucleosome. ING2 (Peña et al., 2006), TAF3 (Vermeulen et al., 2007; Ingen et al., 2008) and PHF2 (Wen et al., 2010), are known to bind to the H3K4me3 mark. PHF2 and TAF3 bound H3K4me0/3 in equal amounts to H3K4me3/Tailless (Fig. 5.4); therefore, the free unmodified H3 N-terminal tail in the asymmetric condition does not contribute to the binding of H3K4me3 binding proteins to H3K4me3/3 nucleosomes. ING2 binds H3K4me0/3 better than H3K4me3/tailless suggesting that the presence of the unmodified H3 N-terminal tail helps ING2 binding.

ING2 still binds the H3K4me3/Tailless albeit with reduced affinity compared to the H3K4me0/3 nucleosome, indicating that the free unmodified H3 N-terminal tail is partially responsible for the ING2 binding seen in the asymmetric
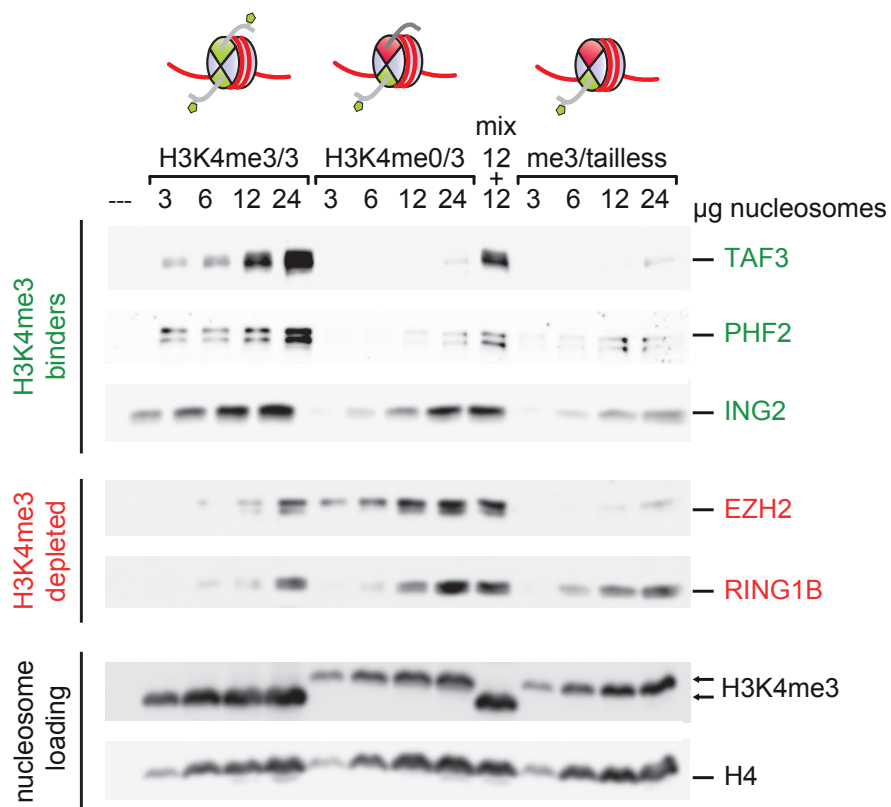
Figure 5.4 *Western Blot Analysis of Asymmetry.* *Western blots of proteins bound to the titrations of differently modified H3K4me3 nucleosomes (H3K4me3/3, H3K4me0/3 and H3K4me3/Tailless). TAF3, PHF2, and ING2 were used as representative H3K4me3 mark binders (labelled in green), and EZH2 and RING1B for behaviour of H3K27me3 or unmodified H3 N-terminal tail binders in red. H3K4me3 and H4 were probed as loading controls (labelled in black).*

condition. ING2 is part of multiple complexes and may therefore be recruited to the unmodified tail by those interactions (this will be examined further in the discussion (see Chapter7)).

Active mark binding proteins were still present in greater amounts on the symmetric compared to the asymmetric nucleosomes, suggesting the majority of protein recruited is due to the presence of H3K4me3 compared to H3K4me0. The concentration of the H3K4me3 mark in asymmetric nucleosomes did not linearly affect either TAF3 or PHF2 binding (Fig. 5.4). 24 µg of H3K4me0/3 did not bind the same amount of TAF3 than even 3 µg of H3K4me3/3 nucleosome. A mixture of 12 µg of unmodified and 12 µg of H3K4me3/3 nucleosome equates to the amount of mark present in 24 µg H3K4me0/3 and in 12 µg of H3K4me3/3. The mix of both 12 µg unmodified and 12 µg H3K4me3/3 bound TAF3 and PHF2 similarly to 12 µg of symmetrically modified nucleosome and better than 24 µg of asymmetrically modified nucleosome. These data indicate that it is the conformation of the H3K4me3 marks that is important in binding both of these proteins as the increase in mark does not compensate for the lack of the symmetric H3K4me3 conformation. It is possible that the local concentration of modification - present on one nucleosome - is important for protein binding. This could be investigated via pulldowns with asymmetrically modified dinucleosomes in specific conformations - discussed later.

EZH2 and RING1B are seen bound to lower concentrations of the H3K4me0/3 than any other condition; therefore, the free unmodified H3 N-terminal tail is responsible for the majority of their binding. EZH2 and RING1B were only detectable at the higher amounts of the tailless nucleosome, suggesting that these proteins bound with much lower affinity to nucleosomes solely containing the H3K4me3 modified tail. A decreased amount of EZH2 binds 24 µg of H3K4me3/tailless compared to 24 µg of both the asymmetric or symmetric condition, showing the contribution of both the H3K4me3 modified and unmodified H3 tail to EZH2 binding.

The effect of a singular asymmetrical modification on the binding of repressive mark binder to nucleosomes was further investigated by titration of H3K27me3/3, H3K27me0/3 and H3K27me3/tailless nucleosomes. EZH2 bound asymmetric and symmetric H3K27me3 nucleosomes equally well (see Fig. 5.5, Fig. 5.2 & Fig. 5.3); therefore binding to both unmodified and H3K27me3 modified H3 N-terminal tails is similar (see Fig. 5.5 & Fig. 5.3) while it is depleted from binding H3K4me3 tails (Fig. 5.2 & Fig. 5.4). EZH2 is part of the PRC2 complex and its binding

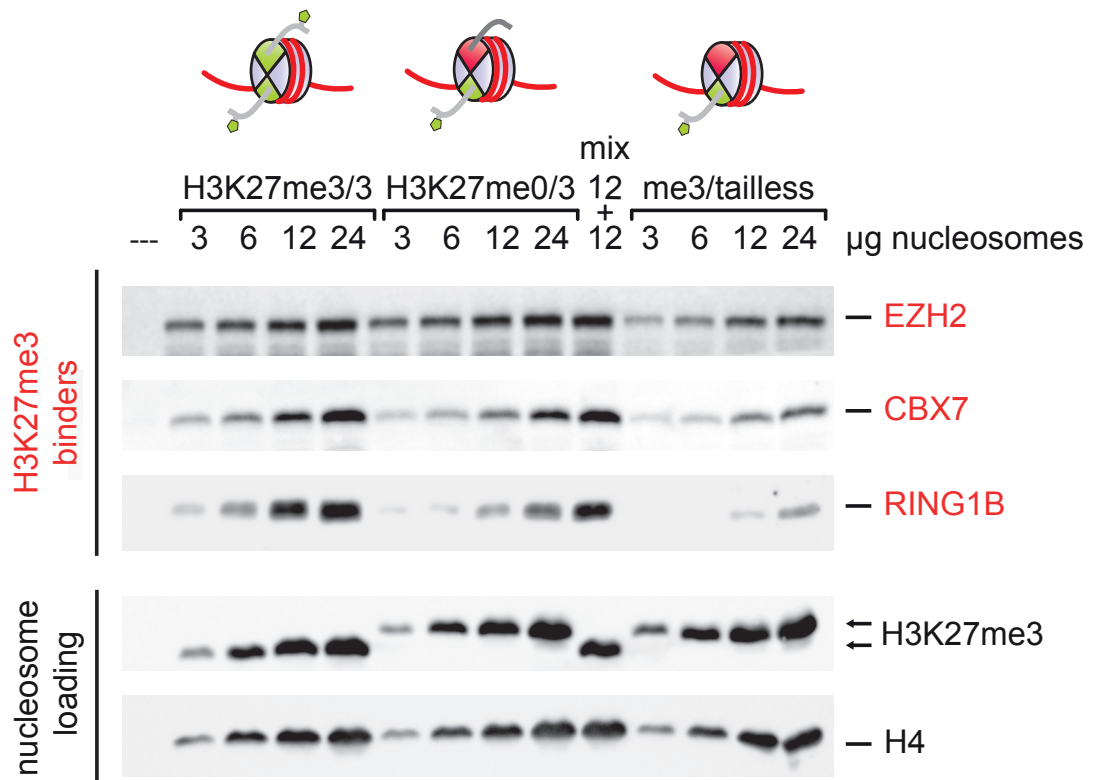**Figure 5.5** *Analysis of H3K27me3 Asymmetry.* *Western blots of proteins bound to the titrations of differently modified H3K27me3 nucleosomes H3K27me3/3, H3K27me0/3 and H3K27me3/tailless. EZH2, CBX7, and RING1B were used as representative H3K27me3 mark binders (labelled in red). The presence of H3K27me3 and H4 are shown as controls.*

to both H3K27me3 and unmodified H3 N-terminal tails can be explained by the presence of multiple binding sites for modified and unmodified H3 N-terminal tails in the PRC2 complex (Justin et al., 2016; Jiao and Liu, 2016; Jiao and Liu, 2015; Margueron et al., 2009; Schmitges et al., 2011), although it may be due to the technical reasons previously mentioned. The reduction in EZH2 binding to H3K4me3/3 nucleosomes may be due to a reduction in the number of sites available in the PRC2 complex for H3K4me3 modified tail to bind (Schmitges et al., 2011).

CBX7 binding to asymmetrically modified nucleosomes is linearly affected by the presence of the H3K27me3 modification. CBX7 is bound in similar amounts to 24 µg asymmetric nucleosomes as to the 12 µg symmetrical nucleosomes. If you double the amount of H3K27me0/3 nucleosomes you get equivalent binding of CBX7 to H3K27me3/3 nucleosomes. RING1B binding is also affected by the amount of H3K27me3 present in the sample (Fig. 5.5) although not in a linear manner. 3 µgs of symmetric nucleosomes binds the same amount of RING1B as 12 µg of asymmetric nucleosomes. This ratio seems to continue, with a similar amount of RING1B binding 6 µg of the symmetric nucleosome and the 24 µg asymmetric nucleosome. RING1B's binding could be a result of both the concentration of H3K27me3 present and the conformation of the marks on the nucleosome.

The binding of CBX7, EZH2, RING1B in the H3K27me3 nucleosome titration mimics that seen in the H3K27me3/3 and H3K27me0/3 nucleosomes via MS analysis. CBX7, EZH2 and RING1B are all part of multi-subunit complexes, therefore the binding seen is as a result of all the interactions in the complex. Both PRC1 and PRC2 have multiple variants that are recruited to different genomic sites and therefore the subunits that are present in more than one of these variant complexes will display binding that is a result of these interactions with different multi-subunit complexes.

Asymmetric nucleosomes, both H3K4me0/3 and H3K27me0/3, did not enrich for proteins bound to the H3K4me3/3 and H3K27me3/3 nucleosomes or proteins previously identified in the literature to bind to either H3K4me3 or H3K27me3. This result was investigated further by a titration of differently modified nucleosomes. All proteins chosen to be investigated displayed different binding behaviour, however, there was a division between the representative active mark and repressive mark binders – the potential significance of which will be discussed later (see Chapter 7).

116

5.4 Characterisation of Proteins That Bind Bivalent Nucleosomes

5.4.1 Proteins Enriched at Bivalent Domains

Although bivalent nucleosomes have been shown to exist *in vivo*, their function and the biological importance of nucleosomes asymmetrically modified with both H3K4me3 and H3K27me3 is not known. The individual modifications that make up bivalency have been individually assessed for their effect on protein binding both symmetrically and asymmetrically (see sections 4.6 and 5.3). The techniques used to assess how singular asymmetric modifications affect protein binding can also be used to evaluate the more complex nature of bivalent nucleosomes. How does the bivalent asymmetric conformation affect the biological function of the marked nucleosome?

The asymmetrical bivalent modified nucleosome was compared to the unmodified tagged (H/S) nucleosome (Fig. 5.6) so that the binding of proteins affected by the tags can be discounted. The binding of a large number of proteins is altered between these nucleosomes indicating that bivalent nucleosomes affect protein binding and this is not due to the presence of tags. The bivalent asymmetric nucleosome was compared to all other nucleosomes tested, in order to isolate those proteins that prefer to bind to the bivalent asymmetric conformation rather than any other combination or conformation of H3K4me3 or H3K27me3 (see Fig. 5.6). The proteins previously mentioned to bind to H3K4me3 or H3K27me3 are shown on these graphs - none of these are consistently significantly enriched on the bivalent condition.

Neither H3K4me0/3 or H3K27me0/3 enriched for H3K4me3 or H3K27me3 mark binders compared to the unmodified nucleosomes. Bivalent nucleosomes are enriched for H3K27me3 marks binders such as CBX7 and EPOP compared to the unmodified nucleosomes (see Fig. 5.6 & Fig. 5.14), therefore bivalent asymmetric nucleosomes recruit proteins differently to predicted based on the protein recruitment by H3K4me0/3 and H3K27me0/3 nucleosomes. The proteins enriched bound to H3K4me3/H3K27me3 are not simply an amalgamation of proteins enriched bound to H3K4me0/3 and H3K27me0/3 nucleosomes. The differences in protein enrichment between the bivalent asymmetric and the singular asymmetric nucleosomes can be clearly seen in Figures 5.6 and 5.7.

117
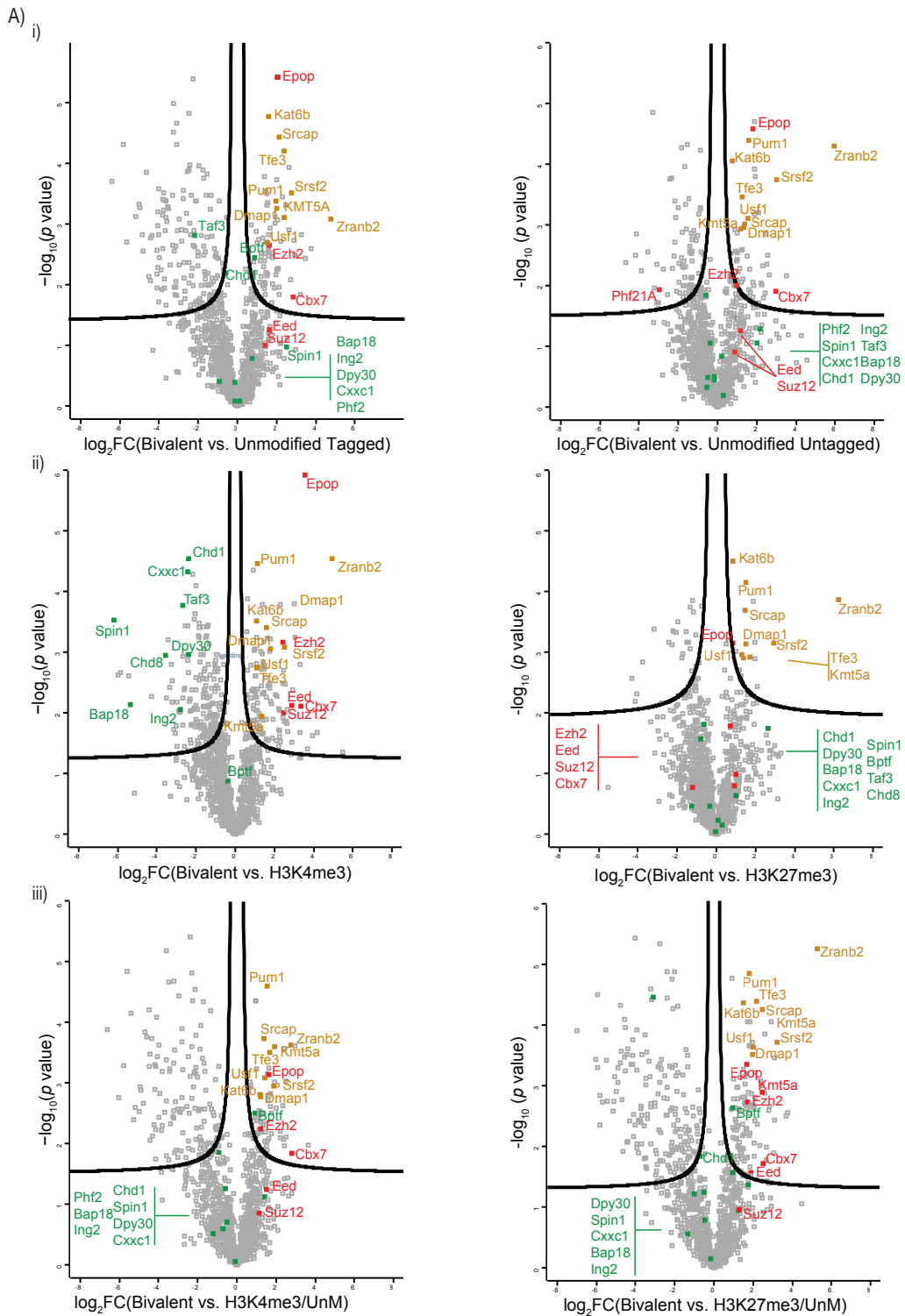
Figure 5.6 *Bivalent Nucleosomes Bind Specific Proteins. Volcano plots of bivalent asymmetrically modified nucleosomes against all other modified nucleosomes tested here in a pairwise comparison FDR=0.01. Proteins are classified as significant if Log2 fold change $\geq$ 1.5 and with a p-value of $\leq$ 0.01. A)i) The bivalent condition compared to both the untagged and tagged unmodified conditions. ii) Bivalent condition compared to nucleosomes symmetrically modified with either H3K4me3 or H3K27me3. ii) Bivalent condition compared to asymmetrically modified conditions, H3K4me0/3 and H3K27me0/3. Gold labels - proteins enriched in the bivalent condition (candidate proteins). Red labels - known H3K27me3 binders. Green labels - known H3K4me3 interactors. All data analysed using MaxQuant (version 1.6.1.0) and Perseus (version 1.6.5.0).*

118
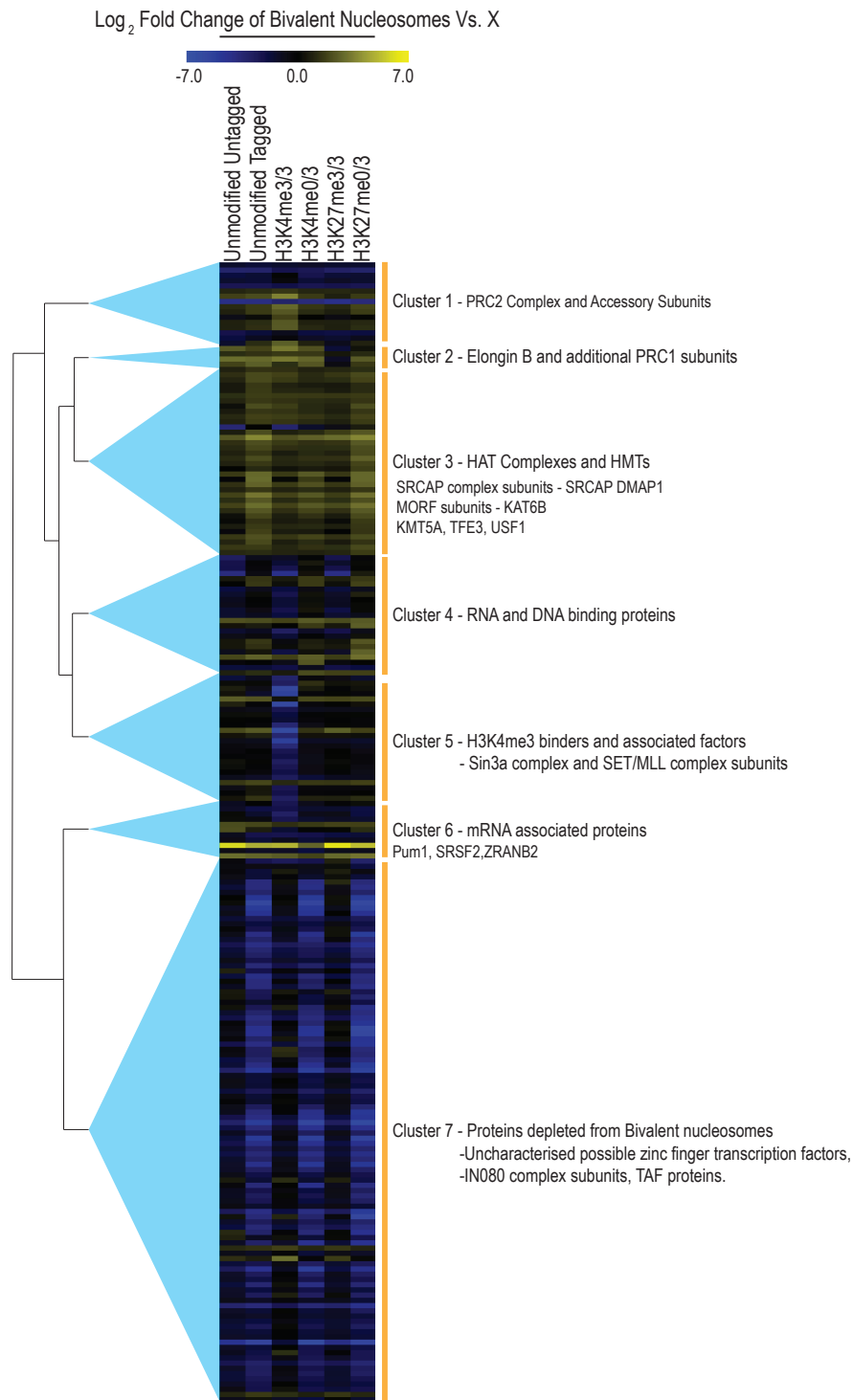
Figure 5.7 *Heatmap of Proteins Significantly Altered on Bivalent Nucleosomes. Hierarchical clustering of the 217 proteins based on their Log2 fold change in enrichment on bivalent nucleosomes compared to all other modified nucleosomes. Log2 fold change heatmap of bivalent nucleosomes vs. X where X is a differently modified nucleosome. This heatmap details proteins which are significantly altered (FDR=0.01, Log2 fold change $\geq$ 1.5 and with a p-value of $\leq$ 0.01) in at least one of the comparisons between bivalent nucleosomes and differently modified nucleosomes. Not all the proteins shown are significantly altered in all comparisons. Proteins grouped in each cluster are shown. Candidate proteins grouped in each cluster are named below each cluster. Candidate proteins are proteins that are enriched significantly on bivalent nucleosomes in all comparisons.*
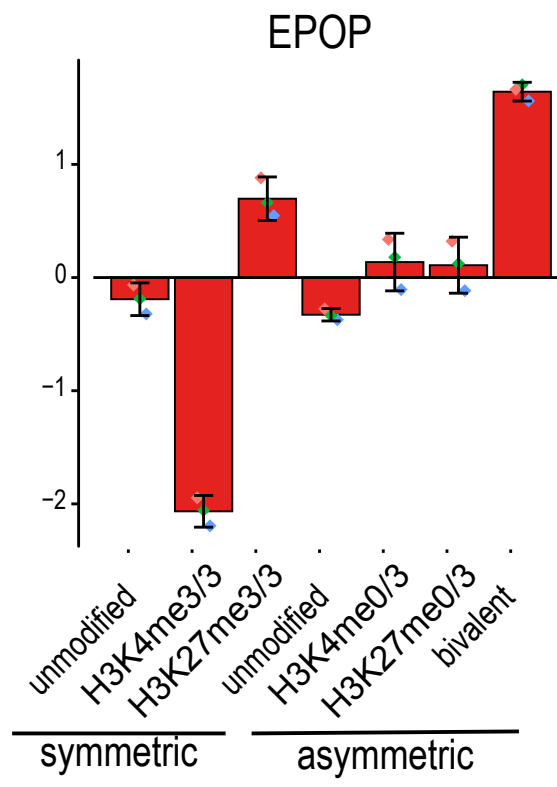
Figure 5.8 *EPOP Enrichment on Differently Modified Nucleosomes.* *Bar graph showing the enrichment of EPOP on Unmodified untagged, Unmodified tagged, H3K4me3/3, H3K27me3/3, H3K4me0/3, H3K27me0/3 and bivalent nucleosomes. EPOP is depleted from H3K4me3/3 nucleosomes, and enriched on H3K27me3/3 and bivalent nucleosomes. The presence of a singular asymmetric mark (H3K4me0/3 or H3K27me0/3) does not seem to affect EPOP binding. Bar plots were generated in the R (version 3.5.3) DEP (1.4.1) using the Shiny package.*

Compared to the singularly modified asymmetric nucleosomes, the bivalent nucleosome consistently enriched for more proteins: including members of the PcG complexes, as well as EPOP and BPTF. However, none of the previously mentioned H3K4me3 or H3K27me3 mark binders are enriched on bivalent nucleosomes compared to the H3K4me3/3 and H3K27me3/3 nucleosomes respectively. The H3K4me3 mark binders may require the conformation of the H3K4me3 marks on the symmetrically modified nucleosome in order to bind. The respective H3K4me3 or H3K27me3 mark binders prefer to bind to the conditions symmetrically modified for their preferred mark.

EPOP is the only protein associated with PRC2 that is significantly enriched in the bivalent condition compared to almost all other nucleosomes (see Fig. 5.6 & Fig. 5.8). EPOP bound to the H3K27me3/3 nucleosome, potentially preventing the enrichment of EPOP seen at bivalent nucleosomes to be classified as significant in comparison. EPOP is depleted from H3K4me3/3 nucleosomes and enriched in the H3K27me3/3 nucleosomes (see Fig. 5.6). However its binding is not affected by the presence of a singular asymmetric mark H3K4me0/3 or H3K27me0/3 (Fig. 5.2). Therefore its enrichment at bivalent nucleosomes is not solely due to the presence of a singular H3K27me3 present at the nucleosome as shown by Fig. 5.6, indicating a potential method of recruitment specific to bivalent nucleosomes. EPOP can form a variant PRC2 complex which has decreased H3K27me3 catalytic activity (Alekseyenko et al., 2014; Smits et al., 2014; Liefke et al., 2016). This variant PRC2 complex may be enriched bound to bivalent nucleosomes.

The log2 fold change of proteins that are significantly altered in at least one of the Bivalent Nucleosome Vs. X comparisons (where X is any other modified nucleosome) are detailed in Figure 5.7. Hierarchical clustering groups the proteins based upon their relative enrichment in each comparison, allowing proteins with a similar pattern of enrichment to be visualised easily. These are typically proteins that are part of the same protein complex or have similar functions. The proteins that show altered binding in bivalent nucleosome comparisons form seven distinct clusters based on their Log2 fold change in each bivalent nucleosome comparison.

Cluster 1 contains proteins that are slightly enriched on bivalent nucleosomes in comparison to the majority of nucleosomes. They are highly enriched on the bivalent nucleosome in comparison to the H3K4me3/3 nucleosome, whilst unaltered compared to the H3K27me3/3 nucleosome. This cluster contains PRC2 complex subunits plus accessory factors (described earlier). The behaviour of

the proteins contained in Cluster 2 – containing PRC1 subunits - follows a similar pattern although the enrichment seen on bivalent nucleosome compared to the majority of modified nucleosomes is greater, while depleted on the bivalent nucleosome in comparison to H3K27me3/3 nucleosome. The behaviour of these clusters supports their biological role (previously discussed) and previous analysis. Cluster 3 contains proteins that are enriched on the bivalent nucleosome compared to all other comparisons and are therefore candidates for proteins that specifically bind to bivalent nucleosomes. Proteins in Cluster 4 are enriched on all tagged nucleosomes compared to the bivalent nucleosome suggesting that their behaviour is affected by the presence of the tags. This cluster is enriched for RNA/DNA binding proteins. Proteins depleted on bivalent nucleosomes compared to H3K4me3/3 nucleosomes, and with unaltered binding on bivalent nucleosomes compared to the other modified nucleosomes are grouped in Cluster 5. This cluster consists of H3K4me3 binding proteins and associated factors for example members of the SIN3A or SET/MLL complexes. mRNA associated proteins form a separate cluster – number 6. This cluster contains several proteins that are enriched on the bivalent nucleosome in all comparisons. The largest cluster is Cluster 7, this is composed of proteins that are depleted from the bivalent nucleosome in all comparisons. This cluster contains INO80 complex subunits and many uncharacterised possible zinc finger transcription factors. However, their behaviour is affected by the presence of the tags on the nucleosome and therefore whether their depletion is biologically significant is unknown.

Bivalent binders are proteins that are enriched on the bivalent asymmetric nucleosome in comparison to all other nucleosomes - shown in yellow in Figure 5.7, and Figure 5.9. The vast majority group in Cluster 3, however, Cluster 6 and Cluster 4 also have some potential candidates. However, not all the proteins shown in the heatmaps such as Figure 5.7 are statistically significant in all bivalent nucleosome comparisons. None of the proteins enriched on the bivalent nucleosome in all comparisons in Cluster 4 were statistically significant and the behaviour of the proteins contained seems to be affected by the presence of the tags on the nucleosomes, therefore an enlarged view of this cluster was not shown. As both Cluster 3 and 6 contain proteins that are statistically significantly enriched on the bivalent nucleosome in all comparisons, further investigation into these proteins was undertaken (see Fig. 5.9). Cluster 3 and 6 will be named Bivalent Cluster 1 and 2, respectively.

Investigation of the Bivalent Cluster 1 shows it is enriched for certain HAT and
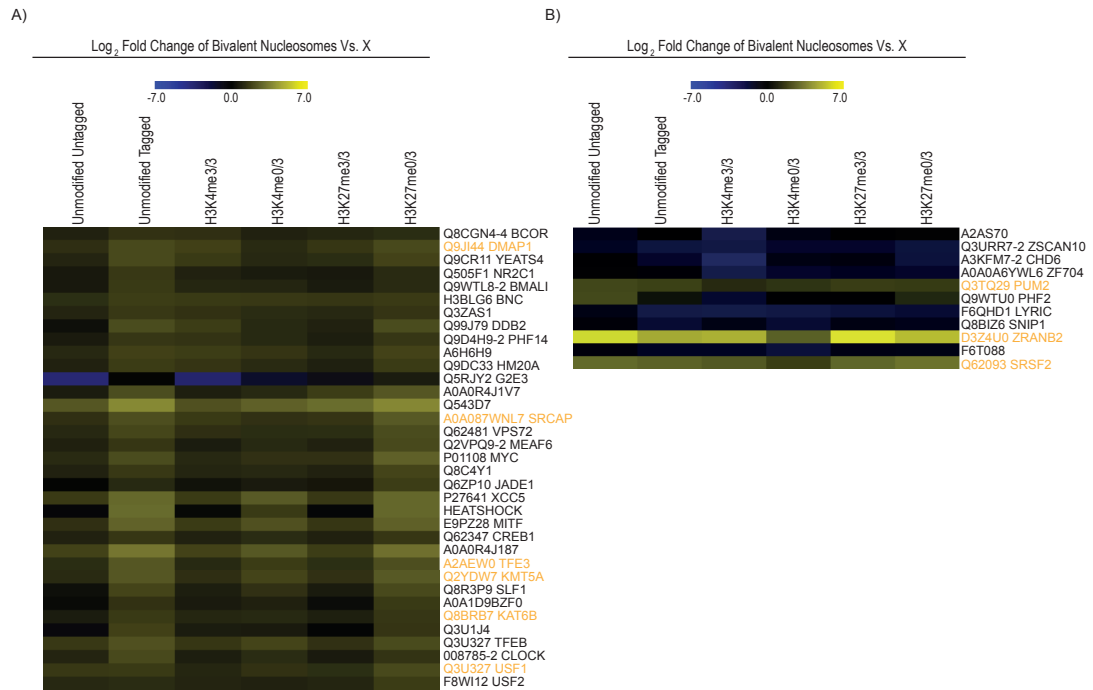
Figure 5.9 *Enlarged View of Bivalent Clusters. Log2 fold change heatmap of bivalent nucleosomes vs. X where X is a differently modified nucleosome. Not all the proteins shown are significantly altered in all comparisons. A) Enlarged view of Cluster 3 from Figure 5.7 named Bivalent Cluster 1. B) Enlarged view of Cluster 6 from Figure 5.7 named Bivalent Cluster 2. Each protein contained within the cluster is labelled with its UniProt I.D. and candidate proteins grouped shown in gold.*

HMT complexes (see Fig. 5.9). This cluster contains proteins from both SRCAP, KAT6B, and HBO1 complexes and certain transcription factors. The identity of catalytic protein subunits present in this cluster may indicate protein complexes recruited to bivalent nucleosomes and help isolate bivalent domain function. Investigation of the Bivalent Cluster 2 shows that it is enriched mostly for mRNA, splicing or uncharacterised proteins (see Fig. 5.9). This cluster contains some transcription factors, and various proteins involved in the establishment of the premRNA spliceosome such as SRSF2 and ZRANB2.

Proteins that are statistically significantly enriched bound to the bivalent nucleosome in all comparisons were isolated via volcano plot analysis (see Fig. 5.6). This was done to reduce the list of potential bivalent binders and isolate the high confidence bivalent binders. There are nine proteins that are consistently significantly enriched in the bivalent asymmetric condition compared to all other nucleosomes: PUM1, USF1, DMAP1, ZRANB2, SRCAP, KMT5A, SRSF2, TFE3, and KAT6B (Fig. 5.6, Fig. 5.7 & Fig. 5.9 - gold labels). There were no proteins that were consistently significantly depleted from the bivalent condition via this analysis. These consistently enriched proteins were termed "candidate proteins" and prefer to bind to the bivalent asymmetric conformation present in bivalent domains.

These enriched candidate proteins contain RNA splicing or binding proteins (SRSF2, PUM1, and ZRANB2), bHLHzip class of transcription factors which bind E-box sequences (TFE3, and USF1), and proteins which are part of complexes that can modify histones (KMT5A, DMAP1, KAT6B and SRCAP). They may contain proteins that are important in the establishment, maintenance or biological function of bivalent domains. To confirm and verify these results, the data was reanalysed using the R package DEP (Differential Enrichment of Protein Data), which provides a workflow for robust analysis of proteomics data (Zhang et al., 2018). Although the p-value and fold changes of the candidate proteins were altered slightly with these new parameters, the majority of the candidate proteins identified via the Perseus software remained significantly enriched in the bivalent sample see Figure 5.11, and Figure 5.10, showing the reliability of the analysis methods used. The candidate proteins still enriched after the second analysis are: PUM1, SRSF2, USF1, TFE3, ZRANB2, DMAP1, SRCAP, and KAT6B. The binding of these proteins is unaltered by the presence of the tags as seen in Figure 5.11 and Figure 5.10.

The copy numbers (number of protein molecules per cell) of the candidate proteins

124
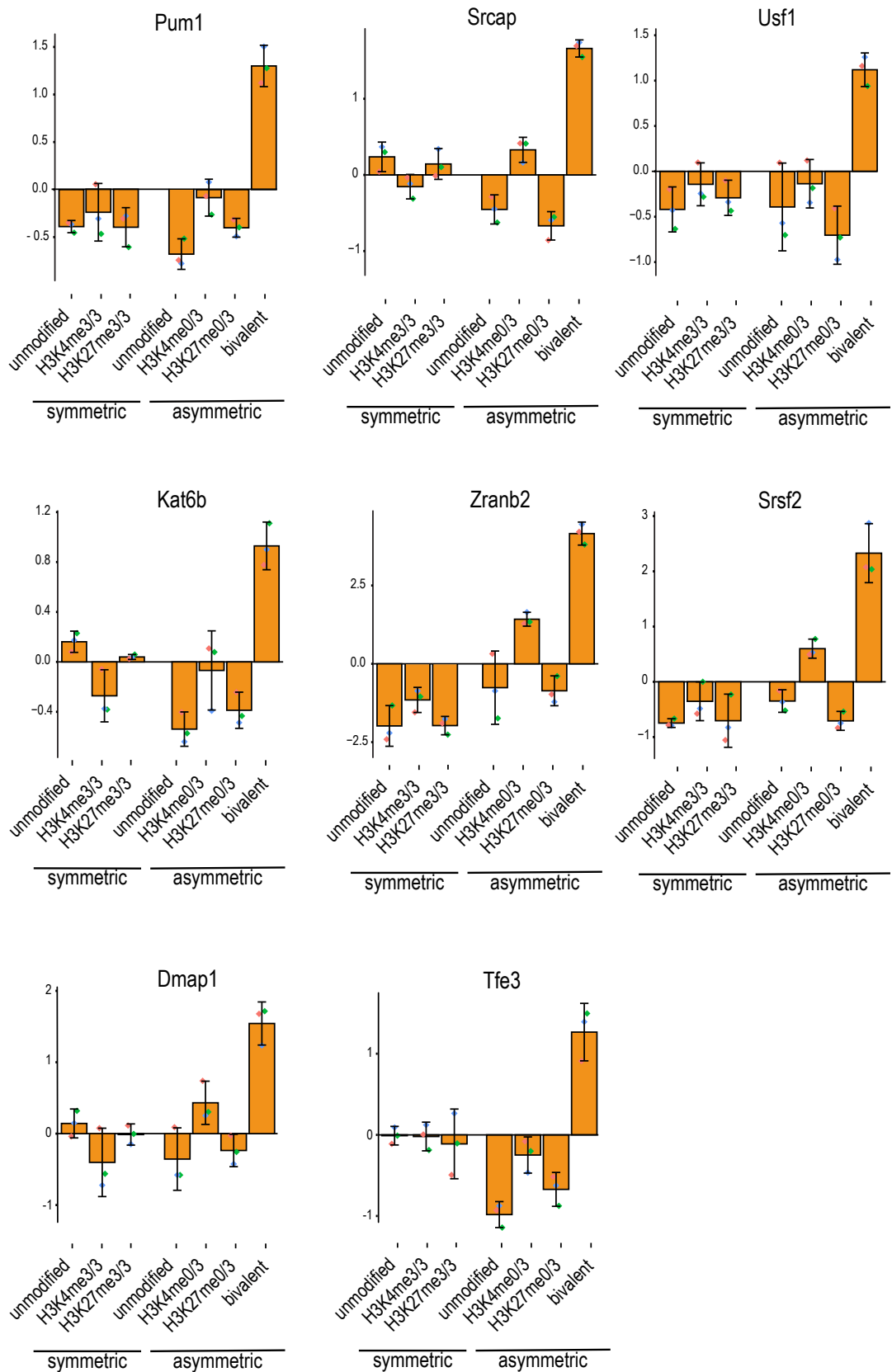
Figure 5.10 Candidate Protein Enrichment. *Bar graphs showing individual enrichment for the candidate proteins at each nucleosome condition. None of the candidate proteins is significantly affected by the presence of the tags on the asymmetrically modified nucleosomes. All candidate proteins are enriched in the bivalent asymmetrically modified condition. Bar plots were generated in the R (version 3.5.3) DEP (1.4.1) using the Shiny package.*

vary, with the proteins present in multiple complexes or in ubiquitous processes such as RNA splicing having higher copy numbers. KAT6B, DMAP1 and SRCAP all have copy numbers several orders of magnitude lower (see Fig. 5.12) (data from Zhang et al. (2017)), this may be due to their presence in fewer complexes and more specific biological functions. The fold change seen in a protein with low copy numbers may be a greater proportion of the total number of protein molecules in the cell compared to the same fold change in a protein with a high copy number. Therefore the same fold change in a low copy number protein may be more likely to result in a biological effect. The proteins with lower copy numbers in the cell may have more specific binding as there are fewer molecules to communicate their function. The copy number of proteins may have implications for binding, as proteins with lower affinity for bivalent nucleosomes may still show appreciable binding if present in large amounts in the cell (high copy number).

SRSF2 is a serine-arginine rich splicing factor which is required for the formation of the pre-mRNA spliceosome and regulates alternative splicing (Edmond et al., 2011; Pellagatti et al., 2018; Yoshida et al., 2011; Liu et al., 2000). Several other serine-arginine rich proteins such as SRSF1 and SRSF3 have been shown recently to bind to the H3 tail to control cell cycle progression and be targeted by chromatin remodelling complexes (Loomis et al., 2009), potentially indicating a mechanism for SRSF2 function. Pum1 is an RNA binding protein which has multiple functions, including regulating the stability and function of certain mRNAs (Van Etten et al., 2012). ZRANB2 is a relatively uncharacterised splicing protein which is involved in regulating pre-mRNA alternative splicing (Mangs and Morris, 2008; Adam et al., 2001). None of the proteins these candidate proteins are known to interact with are enriched on bivalent nucleosomes.

As bivalent genes contain poised RNAPII, the selective recruitment of SRSF2 to bivalent nucleosomes suggests a method for eventual RNAPII release into productive elongation when the gene becomes activated. SRSF2 may be pre-emptively loaded at bivalent domains to decrease the time between activation of bivalent domains and transcription, or its presence may result in the small level of transcription seen at bivalent genes (Bernstein et al., 2006a; Guenther et al., 2007; Brookes et al., 2012). The co-occurrence of PUM1 and ZRANB2 in addition to SRSF2 indicates a role for co-transcriptional, alternative splicing in the function of bivalency. Further studies will be required to determine how these proteins are recruited to bivalent domains, how they help interpet bivalen domain function and if they affect RNAPII poising.

126
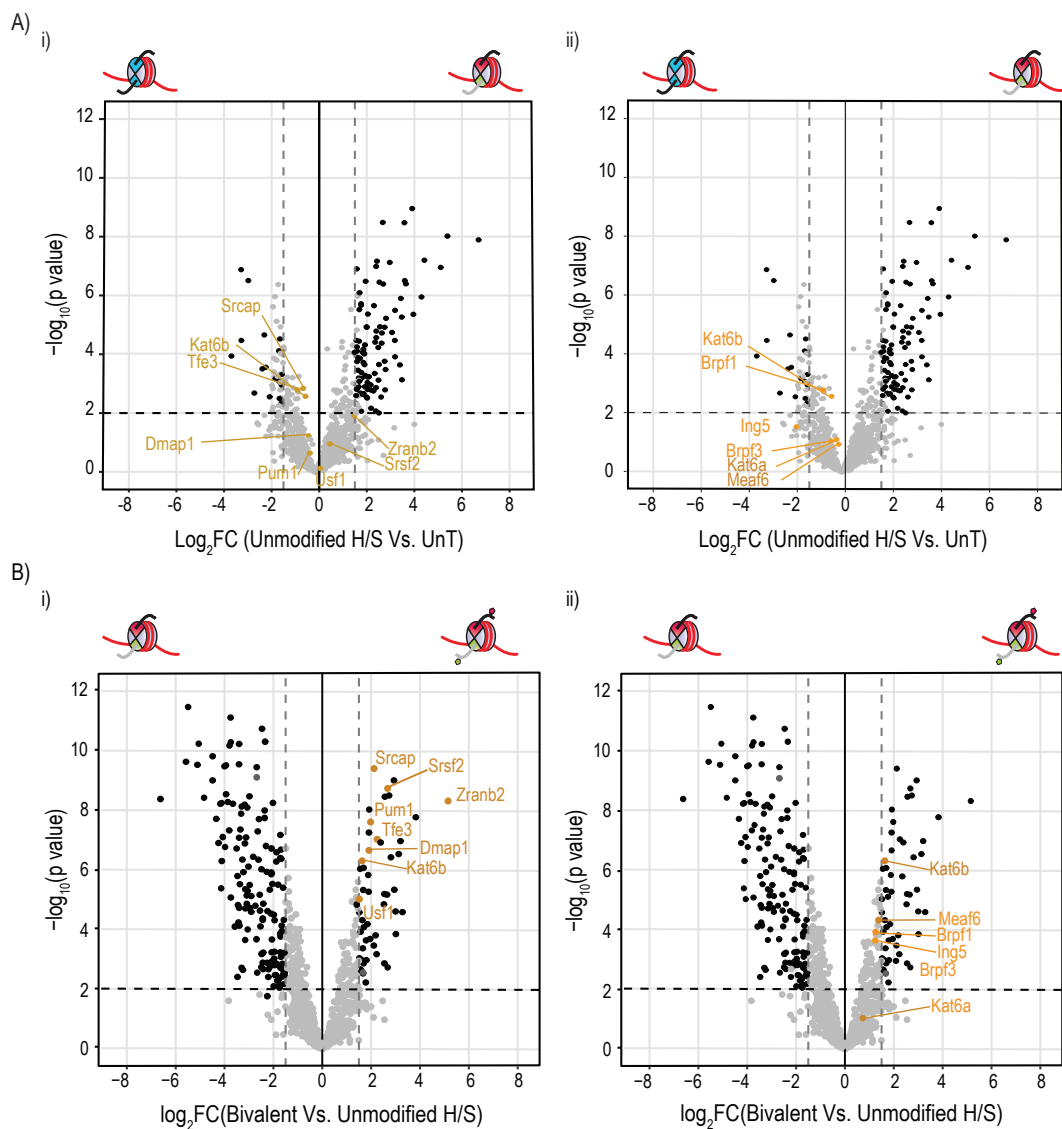
Figure 5.11 *Candidate Proteins are Unaffected by Tags and Enriched in the Bivalent Condition. Volcano plots of A) Untagged control compared to tagged control i) candidate proteins from Perseus labelled - gold. ii) Components of the MORF complex labelled - gold. None of the proteins are significantly affected by the presence of the tags. B)Volcano plots of the bivalent asymmetrically modified nucleosome condition against the unmodified tagged nucleosome condition with i) candidate proteins shown ii) components of MORF complex shown in gold. Proteins are classified as significant if they were enriched by Log2 $\geq$ 1.5 and have a p value lower than 0.01. Volcano plots were generated in ggplot2 in R (3.5.3.).*

Protein	Copy Number 1	Copy Number 2	Copy Number 3	Average C. No.
KAT6A	67.794	28.605	194.088	96.829
KAT6B	4391.983	4651.667	4792.506	4612.052
SRCAP	359.174	97.464	115.426	190.688
YEATS4	61694.600	36470.300	43901.600	47355.500
MEAF6	54872.854	56550.661	46603.308	52675.608
ING5	33397.300	31625.500	68289.200	44437.33
DMAP1	5901.400	3581.882	7368.934	5617.405
BPTF	1324.640	1234.632	2520.014	1693.096
BRPF1	257.612	198.686	410.115	288.807
VPS72	4439.275	9892.573	12229.134	8853.661
SRSF2	5976033.952	4028687.847	6024419.111	5343046.970
PUM1	145444.649	114902.654	133863.528	131403.610
ZRANB2	351985.293	328568.190	526108.419	402220.634
USF1	13618.896	14573.388	14705.778	14299.354
TFE3	3050.278	2772.463	3390.483	3071.075

Figure 5.12 *Copy Numbers of Candidate Proteins and Complexes. Candidate proteins are labelled in pink. Components of SRCAP complex are shown in yellow. Components of MORF/MOZ complex shown in green. Copy numbers are the number of protein molecules per cell. Data displayed from Zhang et al. (2017).*

The transcription factors recruited to bivalent nucleosomes include USF1 and TFE3. USF1 is present at bivalent domains and although dispensable for ESC renewal is required for ESC differentiation (Deng et al., 2013). TFE3 has diverse functions. It supports ESC pluripotency and is excluded from the nucleus during differentiation (Betschinger et al., 2013; Kalkan et al., 2017). This combination of transcription factors at the same location required both for the pluripotency and differentiation of ESCs supports the theory of bivalent domain importance in priming developmental genes for timing regulation during differentiation.

KMT5A places H4K20me1 which is present at bivalent domains (Cui et al., 2009) although its significance at bivalent domains is not known. H4K20me1's function is still unclear with some attributing its role to gene expression (Talasz et al., 2005; Vakoc et al., 2006; Barski et al., 2007), cell cycle control (Jorgensen et al., 2013; Beck et al., 2012), RNAPII pausing (Kapoor-Vazirani and Vertino, 2014) and others to gene repression (Nishioka et al., 2002; Kohlmaier et al., 2004; Karachentsev et al., 2005; Congdon et al., 2010).

DMAP1 is part of multiple complexes including NuA4 and the SRCAP complex (Doyon et al., 2004; Cai et al., 2003; Mohan et al., 2011), of which the SRCAP protein is the catalytic component. The SRCAP complex along with the p400 complex deposits H2A.Z in chromatin by catalysing the exchange of H2A/H2B for H2A.Z/H2B in the nucleosome (Wong et al., 2007; Liang et al., 2016; Ruhl et al., 2006; Gévry et al., 2007; Cai et al., 2005). H2A.Z is essential for embryo development (Faast et al., 2001). Its role in transcription is controversial, having

been linked to both repression (Rangasamy et al., 2003; Fan et al., 2004) and activation (Barski et al., 2007; Zhang et al., 2005; Meneghini et al., 2003; Wan et al., 2009; Mavrich et al., 2008; Schones et al., 2008). Its role in transcription may therefore be context dependent (Fan et al., 2002; Li et al., 2005a; Millar et al., 2006; Raisner et al., 2005; Swaminathan et al., 2005; Zhang et al., 2005) as it allows both MLL and PRC2 complexes to bind to chromatin (Hu et al., 2013b). In mouse ESCs it is enriched at active and bivalent promoters but excluded from repressed genes (Ku et al., 2012). A knockdown of H2A.Z leads to de-repression of bivalent genes (Hu et al., 2013b; Creyghton et al., 2008). SRCAP and its product H2A.Z has been posited to poise inactive promoters on which it is located for timely regulation (Wan et al., 2009; Ku et al., 2012; Guillemette et al., 2005; Zhang et al., 2005; Giaimo et al., 2019). GAS41 (also referred to as YEATS4) is a subunit of SRCAP that reads H3 acetylation of K27 and K14, recruiting SRCAP and promoting the deposition of H2A.Z. A knockdown on GAS41/YEATS4 results in a decrease of H2A.Z and H3K27me3 at bivalent domains (Hsu et al., 2018a). Other subunits of the SRCAP complex were enriched on bivalent nucleosomes in comparison to unmodified nucleosomes including YEATS4 and VPS72 suggesting formation of this complex is enriched at bivalent nucleosomes (see Fig. 5.13).

KAT6B is part of the multi-subunit complex MORF and acetylates H3 at K9, K14 and K23 (Ullah et al., 2008; Voss et al., 2009; Sheikh and Akhtar, 2019; Simó-Riudalbas et al., 2015; Qin et al., 2011; Ali et al., 2012). H3K14ac and H3K23ac are both found at bivalent domains (Voigt et al., 2012; Karmodiya et al., 2012). Investigation into KAT6B's biological function via any one of these marks is difficult as they are placed by a multiple enzymes, including KAT6B, KAT6A and KAT7 (Sapountzi and Côté, 2011; Sheikh and Akhtar, 2019). KAT6B's placement of H3K14ac found at bivalent domains and SRCAP's recruitment to bivalent domains via GAS41 suggests a link between these complexes and their function. The enzymatic activity of and potential link between SRCAP and KAT6B may be necessary for bivalent domain maintenance and function.

All of these candidate proteins may help the cell mediate bivalent domain function; however, only a few could be investigated further within the scope of this thesis. SRCAP and KAT6B are the catalytic subunit of their respective complexes and analysis of the heatmap (Fig. 5.9) shows the presence of additional complex subunits in the same cluster and therefore exhibiting similar binding behaviour. The non-catalytic subunits from these complexes are not significantly
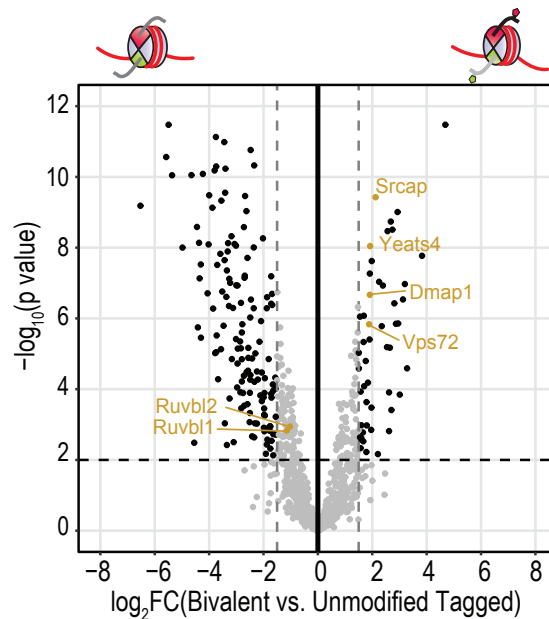
Figure 5.13 *Enrichment of SRCAP Complex Subunits.* *Volcano plot of bivalent nucleosome compared to unmodified tagged nucleosome. Four core SRCAP components are enriched on the bivalent nucleosomes compared to the control. SRCAP complex subunits shown in gold. Data from DEP suite analysis in R version 3.5.3. Volcano plots were generated in ggplot2 in R (3.5.3.). Proteins are classified as significant if they were enriched by Log2≥1.5 and have a p value lower than 0.01.*

enriched on bivalent nucleosomes in all comparisons and therefore are not shown in Figure 5.6. Therefore KAT6B and SRCAP (Snf2 Related CREBBP Activator Protein) (Fig. 5.14) were chosen for further investigation as they are significantly enriched after both analyses and are the catalytic subunits of multi-protein complexes, MORF and SRCAP respectively, both of which containing multiple protein binding domains. The other subunits in these complexes exhibit similar binding behaviour to KAT6B and SRCAP (see Figure 5.7 and 5.9) enriched bound to the bivalent nucleosome supporting a role for these complexes at bivalent domains. In addition, the histone PTMs H3K14ac, H3K23ac and H2A.Z that are placed by MORF and SRCAP respectively, are found at bivalent domains (Voigt et al., 2012; Karmodiya et al., 2012).

KAT6B and SRCAP are components of large multisubunit complexes giving plausible possibilities for recruitment, modification, maintenance to bivalent domains. Both of these proteins have low copy numbers in the cell (Fig. 5.12) and are only present in one complex, allowing easier dissection and analysis of their function although this increases the difficulty of visualisation of localisation *in vivo* via ChIP. The low copy number of both these proteins in ESCs, in addition to their importance in differentiation and embryonic development (Ku et al.,

130
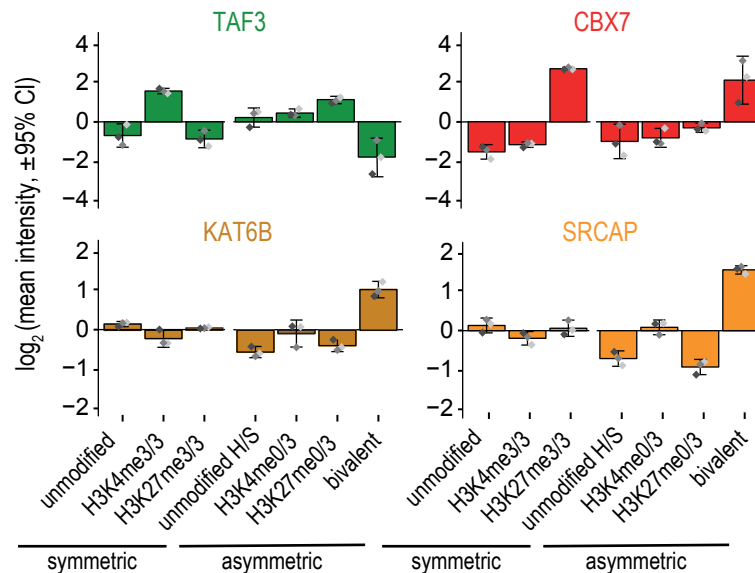
Figure 5.14 *Selected Protein Enrichment in all Conditions.* *Plots showing the enrichment of selected active mark binder TAF3 (green), repressive mark binder CBX7 (red) and candidate proteins SRCAP (yellow) and KAT6B (gold) in all conditions. Bar plots were generated in the R (version 3.5.3) using the DEP suite (1.4.1) and the Shiny package.*

2012), indicates that these proteins are involved in distinct biological mechanisms occurring at specific locations in the genome that are integral to differentiation and embryonic development. The enrichment of these proteins and complex subunits, as well as the presence of their marks at bivalent domains indicated that SRCAP and KAT6B are good candidates for proteins bound to bivalent domains *in vivo* and possible candidates for mediating their function. To resolve their possible role in bivalent domain function we generated mESCS using CRISPR-Cas9 expressing tagged versions of both SRCAP and KAT6B or knockouts of these proteins. However, because we failed to obtain meaningful enrichment in chromatin immunoprecipitation experiments with tagged SRCAP, precluding analysis of its binding to bivalent domains in mESCs, we focused on investigating KAT6B function.

KAT6B is a paralogue of KAT6A (Champagne et al., 1999) and shares 60% sequence identity, while their acetyltransferase MYST domain shares 88% sequence identity (Thomas et al., 2000). KAT6B has higher copy numbers in mESCs than KAT6A, suggesting a higher expression level and potentially a larger role in mESCS (Fig. 5.12) (Zhang et al., 2017). The similarity of KAT6B and KAT6A initially made separation of their functions difficult, as most early commercial antibodies could not distinguish between the two proteins. Because of their similarity, generalisations can be made from one complex to the other. More recent studies have managed to untangle some of their function. The

MORF complex consists of 4 proteins: KAT6B, MEAF6, ING5, and BRPF1/2/3 (Ullah et al., 2008). These subunits are enriched to similar, albeit slightly lower, levels as KAT6B in the bivalent pulldown condition compared to the unmodified nucleosome, see Figure 5.11, suggesting KAT6B is present in this complex when bound to bivalent nucleosomes. The MORF complex contains multiple histone PTM reader domains: 5 PHD domains, one PWWP domain, and one Bromodomain, potentially allowing the MORF complex to integrate the signals of multiple histone modifications.

The N-terminal portion of KAT6B including its MYST domain shares high similarity with Drosophila ENOK (Yang, 2004), which also assembles with Br140, Ing5, and Eaf6 which are the Drosophila orthologs of BRPF1/2/3, Ing4/5, and MEAF6, respectively. KAT6 complex formation and binding domains are evolutionary conserved from yeast to humans with some slight alterations/differences, for example Yng1 is the ortholog of ING1-2 rather than ING5 but still contains a PHD domain which binds H3K4me3 (Huang et al., 2016; Gordon et al., 2008). This evolutionary conservation between yeast and humans suggests that KAT6 complexes play an important and crucial role in a wide variety of organisms. Mutations, deletions and translocations of KAT6B result in multiple diseases including leukemia and several developmental disorders (Yang, 2015).

KAT6B consists of an N-terminal domain referred to either as NEMM (N-terminal part of ENOK) due to its similarity to ENOK's N-Terminal domain or H15 domain due to its sequence similarity to H1 and H5 histones. This is followed by a double/tandem PHD domain, the MYST domain, and a C-Terminus consisting of a glutamate/aspartate rich (ED) and a serine/methionine rich (SM) domain (Klein et al., 2014; Ullah et al., 2008; Yang and Ullah, 2007; Scott et al., 2001; Ullah et al., 2008). The ED and SM domains present in the C-Terminal tail of KAT6B help MORF's transcriptional activity via interaction with Runt-domain transcription factor 2 (Runx2) (Pelletier et al., 2002; Coffman, 2003). The tandem double PHD domain (DPD) of KAT6B is required for KAT6B association with acetylated H3 N-terminal tails. The DPD recognises the unmethylated H3 N-terminal tail, its binding is increased 3-4 fold in the presence of H3K14ac or H3K9ac, but inhibited by the presence of H3K4me3 (Ali et al., 2012). The presence of H3K14/9ac partially mitigates the effect of H3K4me3 on binding (Klein et al., 2017). This specificity is conserved for the DPD of KAT6A/MOZ. MOZ complexes, when recruited to H3K14ac, cannot modify lysines in the

same peptide (Ali et al., 2012), suggesting a potential mechanism of H3K14ac propogation by MORF/MOZ complexes. Once H3K14ac is bound by MORF's double plant homeodomain finger (DPF), its MYST domain may target lysines in the H3 in the same or adjacent nucleosomes (Klein et al., 2017).

KAT6B has a wide range of substrates and its MYST domain *in vitro* is able to acetylate H2A, H3 and H4; however, the association of BRPF1 and ING5 helps narrow its acetyltransferase activity to nucleosomal H3 (Yang, 2015; Doyon et al., 2006; Ullah et al., 2008; Qin et al., 2011). MORF's HAT targets are still unclear but MOZ/MORF has been suggested to place H3K14ac (Ullah et al., 2008; Ali et al., 2012), H3K9ac (Voss et al., 2009), and H3K23ac (Sheikh and Akhtar, 2019; Simó-Riudalbas et al., 2015; Mi et al., 2017). BRPF1 ($\alpha\alpha$ 150-245) interacts with KAT6B's MYST domain ($\alpha\alpha$ 761-782) changing MORF's specificity from H4 to H3 and increasing its acetyltransferase activity (Ullah et al., 2008; Doyon et al., 2006). BRPF1 has two PHD fingers separated by a Zinc finger (PZP) and a Bromodomain. The PZP domain binds in 2:1 stoichiometry to the nucleosome, shifting the equilibrium to unwrapping DNA from the nucleosomes, increasing accessibility (Klein et al., 2016). The first PHD finger binds the unmodified H3 tail, however this binding is prevented by methylated H3K4. The second PHD finger interacts non-specifically with DNA and both PHD fingers are required for BRPF1/MORF's binding to H3 *in vivo* (Lalonde et al., 2013; Liu et al., 2012). BRPF1's bromodomain ($\alpha\alpha$ 679-766) recognises and binds to acetylated histone lysine residues with a preference for H3K14ac, H4K12ac, H2AK5ac while its PWWP ($\alpha\alpha$ 1126-1226) binds both H2A/H2B and H3K36me3 and is necessary for BRPF1/MORF association with condensed chromatin (Poplawski et al., 2014; Ruthenburg et al., 2011; Laue et al., 2008; Qin et al., 2011; Klein et al., 2016).

BRPF1 and ING5 are essential for the formation of the MORF complex, and BRPF1 bridges the interaction between hEAF6 and ING5 by its M domain (Ullah et al., 2008). ING5 is part of the ING family of tumour suppressor proteins and contains a PHD domain which binds to methylated H3K4, with a binding preference of H3K4me3>H3K4me2>H3K4me1>H3K4me0 (Champagne et al., 2008). The inclusion of ING5 is essential for the MORF complex's acetyltransferase activity (Champagne et al., 2008) and causes MORF to acetylate methylated H3K4 peptides. This allows MORF to be recruited to and modify areas of chromatin containing H3K4me3. hEAF6 (191 $\alpha\alpha$) is the least characterised of the MORF complex subunits. It contains a leucine zipper motif and its expression helps the nuclear localisation of BRPF1 (Klein et al., 2014).
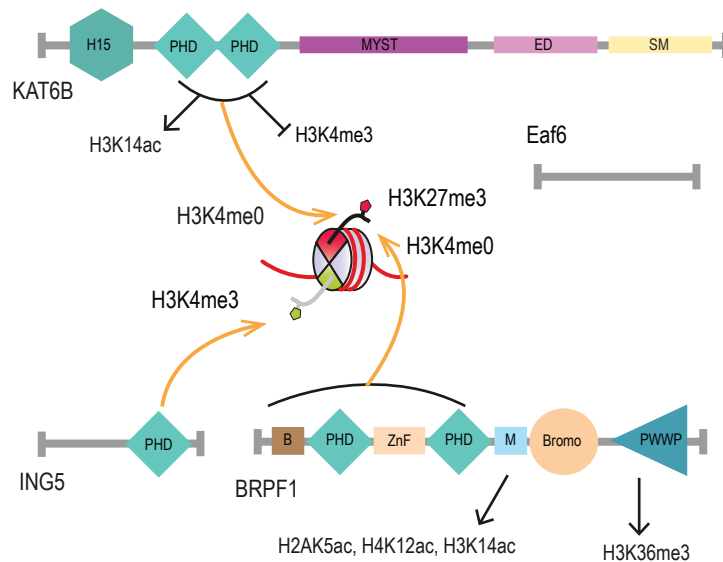
Figure 5.15 *Schematic Structures of MORF Complex Subunits and Protein Domains. Domains are labelled as follows: PHD - Plant homeodomain zinc finger, ZnF - Zinc finger, H15 - Sequence similarity to H1 and H5 histone domain, also referred to as NEMM, MYST - Acetyltransferase domain, ED - acidic (glutamate/aspartate rich) region, SM - Serine/methionine rich stretch, Bromo - bromodomain, PWWP - proline-tryptophan-tryptophan-proline domain, B - domain that interacts with KAT6B's MYST domain, M - domain that mediates BRPF1's interaction with ING5 and hEAF6.*

Due to the number of protein domains contained within the MORF complex there are a multitude of possible mechanisms for its recruitment to bivalent domains (Fig. 5.15). The PHD finger of ING5 interacts with H3K4me3 helping recruit the MORF complex while the DPF of KAT6B and PZP domain of BRPF1 can recruit the MORF complex to H3K4me0 - present on the H3K27me3 tail of an asymmetric bivalent nucleosome. The bromodomain of BRPF1 and DPF of KAT6B can recognise the marks placed by the MOZ/MORF complex, resulting in a feedforward loop and potential spreading or amplification of MORF's signals. MORF-like complex subunits in Drosophila, ENOK and EAF6, interact with Polycomb (Pc) a PRC1 complex member (Strubbe et al., 2011). If this interaction is maintained in mice and KAT6B and hEAF6 interact with CBX proteins it provides a further mechanism for MORF's recruitment to bivalent nucleosomes and one that is directly dependent on H3K27me3.

5.4.2 Proteins Depleted from Bivalent Nucleosomes

We primarily focused on the proteins bound by bivalent domains as they may mediate their biological function. However, the proteins depleted from bivalent nucleosomes may also shed light on the importance of bivalent domains. The
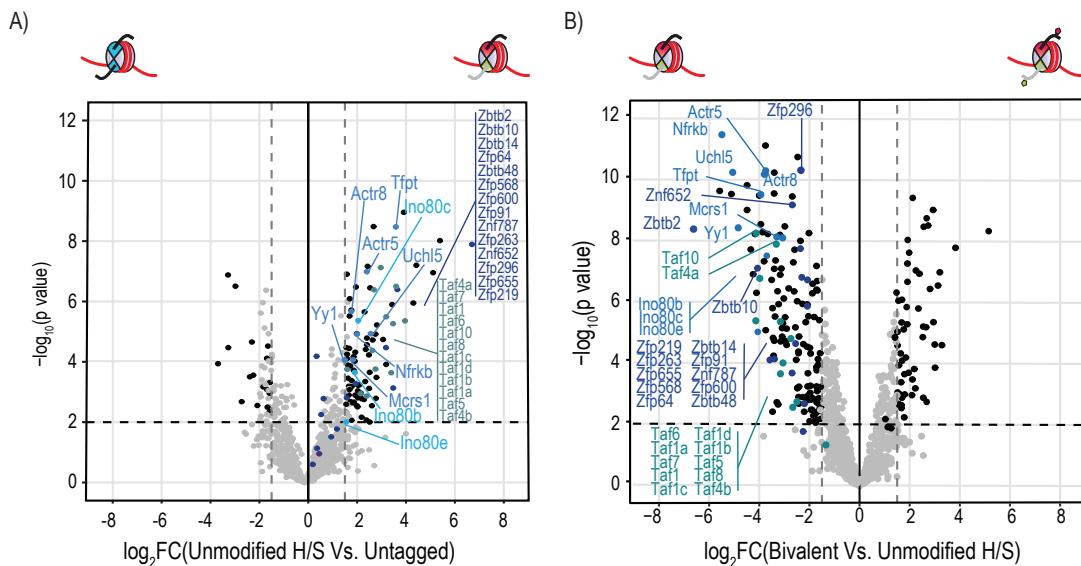
Figure 5.16 *Proteins Depleted from the Bivalent Nucleosome. Volcano plots showing the proteins depleted A) from the unmodified untagged control versus the unmodified H/S control B) from the bivalent condition in comparison to the unmodified H/S control. INO80 complex subunits light blue, TAFs - green and BTB domain containing proteins - dark blue. Data obtained from the DEP analysis suite in R.*

initial analysis using the Max-Quant suite of programmes including Perseus did not find any proteins that were depleted from the bivalent condition relative to all other conditions investigated. Nevertheless, a comparison of the unmodified H/S nucleosome versus the bivalent nucleosome, via both analyses used, shows a large number of proteins depleted (Fig. 5.16). The proteins previously used as controls for binding to H3K4me3 or H3K27me3 have been demonstrated not to be affected by the presence of the tags required for the purification of asymmetrically modified nucleosomes (see Fig. 5.1).

There are various proteins that are depleted from the bivalent condition relative to the tagged control. These can be split into 3 groups: subunits of the INO80 complex (light blue), proteins containing a BTB domain or zinc finger (darker blue) and a variety of transcription associated proteins (TAFs) (green) (Fig. 5.16). Unfortunately, the proteins depleted from bivalent nucleosomes compared to unmodified tagged nucleosomes are all affected by the presence of the tags on the octamer (Fig. 5.16). Both TAFs and BTB domain containing grouped proteins are depleted from bivalent nucleosomes, but enriched at unmodified tagged nucleosomes with approximately the same fold change, albeit in opposite direction.

While IN080 complex subunits (IN080 B, C, D, E and YY1 etc) are enriched
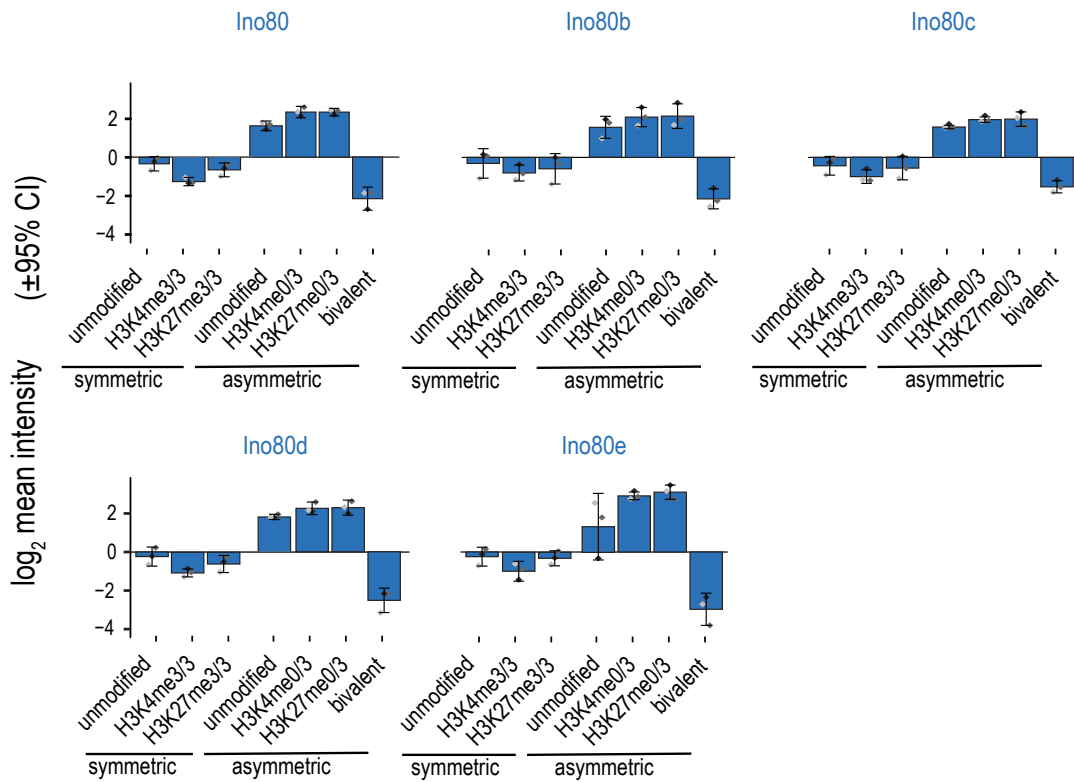
Figure 5.17 *INO80 Complex Subunit Enrichment.* *DEP plot of IN080 complex subunits presence in all conditions. Symmetrically modified nucleosomes have a slight reduction in the presence of INO80 subunits compared to the control. INO80 subunits are enriched at asymmetrically singularly modified nucleosomes due to the tags present on histone H3, however they are depleted from bivalent asymmetrically modified nucleosomes. The depletion of the subunits from bivalent nucleosomes is generally greater than the enrichment seen in the other asymmetric nucleosomes.*

approximately 2−fold bound to nucleosomes containing the tags (see Fig. 5.17), they are depleted, around 4−fold, from the bivalent condition. This indicates that the decrease is due to the presence of the bivalent marks. Due to their increased affinity towards the tagged nucleosomes we cannot say whether the depletion seen in the bivalent condition compared to unmodified H/S is biologically significant. IN080 is a part of the SWR1 family of ATP-dependent chromatin remodelling complexes (Bao and Shen, 2007; Brahma et al., 2017; Papamichos-Chronakis et al., 2011) and the depletion of IN080 from bivalent nucleosomes may reduce the rate of exchange of H2A.Z/H2B for H2A/H2B from bivalent domains, keeping them poised for quick transcription and thus biologically relevant. Further pulldown experiments with bivalent untagged nucleosomes would allow investigation into this depletion - detailed later.

The enrichment of proteins on the tagged versus the untagged control indicates that these proteins may interact with the surface of the nucleosome that contains the tags as well as the histone marks. As a result of these complicating variables it cannot be known whether the depletion of the proteins from the bivalent nucleosome in comparison to the tagged control is biologically significant. However, all the proteins that were depleted from bivalent nucleosomes are also enriched in the tagged control vs untagged control, suggesting that they are unlikely to be significant. The lack of certainty about the actual state of binding of these proteins meant they were excluded from further immediate study. In order to assess whether the proteins depleted from the bivalent condition are biologically relevant and not due to the presence of tagged histone H3, generation of untagged bivalent H3K4me3/H3K27me3 and comparison with unmodified untagged nucleosomes would be required.

5.5 Conclusion

Although multiple previous studies have investigated the biological function of symmetrically modified nucleosomes and the proteins that bind them, the full effect of an asymmetric, singular modification has yet to be investigated. How the asymmetric state affects identity and kinetics of protein binding was unknown. The use of singularly asymmetrically modified nucleosomes showed the lack of protein enrichment in comparison to unmodified nucleosomes. The presence of a single H3K4me3 mark is not sufficient to recruit active mark binders including TAF3 and PHF2 at amounts similar to H3K4me3/3 nucleosomes.

Additionally, the inhibition of EZH2 binding to H3K4me3/3 is reduced with H3K4me0/3; therefore it is unlikely that H3K4me0/3 nucleosomes are able to maintain the active state *in vivo*. H3K27me0/3 recruits PRC2 components similarly to H3K27me3/3 although it is less efficient in recruiting components of the PRC1 complex such as CBX7. The reduction in CBX7 binding is H3K27me3 concentration dependent. The recruitment of proteins to singularly modified asymmetrically modified nucleosomes may allow prediction of the transcriptional state of these nucleosomes *in vivo*.

The difference in the behaviour of proteins binding symmetrically modified, asymmetrically modified or unmodified nucleosomes raises the question whether the effects seen are due to the local concentration of histone PTMs, or the conformation of the PTMs in a symmetric nucleosome. The titration of both symmetrically and asymmetrically nucleosomes showed that with certain proteins including TAF3 and PHF2, the overall concentration of the nucleosomes and the marks does not result in a linear increase in protein binding. However, the local concentration of marks on the beads or the nucleosomes themselves may affect the protein binding. Electrophoretic Mobility Shift Assays (EMSAs) with purified TAF3 binding to nucleosomes in solution without immobilisation on beads would allow the effect of the beads to be seen clearly. Increasing the amount of beads relative to the concentration of nucleosomes (data not shown) showed an increase in TAF3 binding when the nucleosomes were more dilute and not saturated on the beads. This indicates that there may be steric hindrance between proteins binding the nucleosomes and the present of either other nucleosomes on the same bead or other proteins binding the adjacent nucleosome. This may also explain the lack of other TFIID subunits present bound to H3K4me3/3 nucleosome. Pulldowns with dinucleosomes may help proteins with high Koff constants or large protein complexes to bind due to the additional interaction surfaces provided by the additional nucleosome. Pulldowns with asymmetrically modified dinucleosomes in various conformations would allow investigation whether it is the specific conformation of PTMs present on a symmetric nucleosome that is required for protein binding or whether the local concentration of PTMs present on the nucleosome or dinucleosome is sufficient for the protein to bind.

Once the effect of the singular modifications was known, investigation into bivalent nucleosomes containing H3K4me3 on one H3 N-terminal tail and H3K27me3 on the other sister H3 could begin. Due to complicating presence of the tags no proteins solely depleted from bivalent asymmetric modified

138

nucleosomes could be determined. There are multiple methods that have been recently developed in order to generate asymmetrically modified nucleosomes, without the use of tags including the XY (Ichikawa et al., 2017) or 'link and cut' tag ("lnc-tag") (Lechner et al., 2016) methodology. The lnc-tag methodology relies on a transient crosslinking strategy between two H3 molecules each carrying distinct PTMs, forcing the assembly of them into the same histone octamer and later on the creation of an asymmetrical nucleosome. This lnc-tag contains a TEV cleavage site, allowing cleavage of the tag and resulting in an asymmetric nucleosome without tags and the crosslink removed (Lechner et al., 2016). The XY methodology generates asymmetric histones octamers, via the use of a pair of obligate H3 histone dimers termed H3X and H3Y. These obligate dimers were created via the mutation of a total of 7 mutated residues located at the site of the H3/H3 interface - H3X (C96S, C110A, L126A, and I130V) and H3Y (C96S, L109I, and C110W) (Ichikawa et al., 2017). The use of either of these approaches would allow the generation of asymmetric nucleosomes both bivalently modified and asymmetric mono-modified nucleosomes without the complications caused by the presence of the tags.

The proteins that bound to the bivalent nucleosome were not simply an amalgamation of the proteins bound to H3K4me0/3 and H3K27me0/3. EPOP is enriched at bivalent nucleosomes, along with CBX7 and EZH2, suggesting the enrichment of PRC2/EPOP complex at bivalent domains as well as a canonical PRC1 complex. I identified several proteins including KAT6B that prefer to bind to this bivalent asymmetric conformation. These are proteins that are enriched bound to bivalent nucleosomes in comparison to any other conformation or combination of marks. KAT6B was chosen for further investigation into its role and function at bivalent domains.

Chapter 6

Recruitment of KAT6B to Bivalent Domains *in vivo* and its Role in Neuronal Differentiation

6.1 Introduction

There are relatively few studies that took an unbiased approach to find novel binders of bivalent domains (Beringer et al., 2016; Gao et al., 2018; Ma et al., 2016; Yang et al., 2018; Hsu et al., 2018a; Ku et al., 2012). This study aimed to identify previously unrecognised proteins bound to bivalent domains and determine their function, thereby helping to elucidate bivalent domain function within cells. The techniques used in previous chapters have identified several proteins that preferentially bind to bivalently modified nucleosomes. If the identified proteins are involved in interpretation of the bivalent domain *in vivo*, we would expect to see them at bivalent domains *in vivo* and a knockout (KO) of the protein to affect bivalent domain function. If the protein is involved in regulating bivalent domain function, a KO would potentially result in the aberrant expression of genes that are bivalent; either in ESCs or during differentiation while having not effect on the self-renewal of ESCs. The previous chapter showed that KAT6B was enriched bound to bivalent nucleosomes and this chapter explores KAT6B's role within the cell and its possible function at bivalent domains.

6.2 Aims

In this chapter I aim to explore the role of KAT6B within the cell and its possible role in regulating bivalent domains. This investigation aims to discover proteins that bind bivalent domains and from their identification and characterisation elucidate bivalent domain function. Previous chapters identified proteins bound to bivalently modified nucleosomes *in vitro*; however, their function and role *in vivo* with respect to bivalent domains has yet to be characterised. Time constraints limited the scope of investigation to a singular protein - KAT6B. The utilisation of several techniques including CRISPR ("Clustered Regularly Interspaced Short Palindromic Repeats") genome editing, chromatin immunoprecipitation (ChIP), and neuronal cell differentiation helped characterise KAT6B function with ESCs and throughout differentiation. ChIP and differentiation assays were performed in collaboration with several Voigt lab members. Cell lines including HA-FLAG-KAT6B and KAT6B[(-/-)] were generated with CRISPR (type II) allowing deeper analysis of KAT6B function and importance within ESCs. ChIP was used to analyse the distribution of KAT6B within the genome, how KAT6B localisation is altered upon removal of bivalency and the effect on a selection of histone marks upon removal of KAT6B. Neuronal differentiation of KAT6B[(-/-)] ESCs allowed examination of KAT6B's importance and function throughout differentiation.

6.3 KAT6B Binds to Bivalent Domains *In Vivo*

This study has shown that KAT6B binds to bivalent nucleosomes *in vitro* however, its binding to bivalent domains has not yet been demonstrated *in vivo*. Confirmation of KAT6B's binding to bivalent domains is required before we can closely examine KAT6B and its role in cells. None of the previous studies into KAT6B function have managed to directly detect KAT6B binding *in vivo*, instead they relied on the indirect readout of the marks it places, for example H3K14ac and H3K23ac. Since multiple enzyme complexes place the same marks, KAT7 places H3K23ac and KAT6A or GCN5 place H3K14ac (Lo et al., 2000; Sheikh and Akhtar, 2019; Baell et al., 2018) these readouts are consequently open to confusion. The majority of available commercial antibodies fail to distinguish between KAT6B and KAT6A due to their homology, making

142

them inadvisable for experimental use (data not shown). In order to distinguish between KAT6A and KAT6B we tagged endogenous KAT6B via CRISPR in order to allow identification of genomic locations bound by KAT6B, this also enabled us to purify the MORF complex via the tagged KAT6B. KAT6B was tagged with a 3 X FLAG 1 X HA tag (Fig. 6.1) to enable its purification and allow the use of high affinity antibodies to increase sensitivity of KAT6B detection *in vivo* and *in vitro*. The general method of screening for positive CRISPR clones and confirmation of successful generation of desired cell lines is shown in Figure 6.1 illustrated with the generation of tagged *KAT6B*.

The CRISPR strategy used to tag KAT6B is detailed in Figure 6.1. PCR Screening of positive single cell-derived colonies was accomplished using various selected primers and allowed identification of homozygous or heterozygous cells containing tagged *Kat6b*. The initial PCR allows the selection of cells containing the tag. These then undergo a second PCR that allows determination of whether the cells are homozygous or heterozygous for tagged *Kat6b*. These results were confirmed by sequencing and at the protein level by analysis via western blot (Fig. 6.1). All other cell lines generated by CRISPR and used in this thesis underwent a similar PCR screening, sequencing and western blot confirmation process. The combination of low copy numbers of KAT6B in ESCs (Zhang et al., 2017) and low sensitivity or insufficient specificity of commercial antibodies for KAT6B meant that western blot confirmation of untagged KAT6B binding in the *in vitro* pulldown experiments was ineffective (data not shown); therefore, the nucleosome pulldowns were repeated using nuclear extract from the tagged-*Kat6b* cell lines, enabling the use of higher sensitivity antibodies. In this way, the MS results were confirmed by western blot analysis (Fig. 6.2). The tagged KAT6B (HA-KAT6B) is present in greater amounts bound to bivalent nucleosomes (Fig. 6.2).

As the tagged cell lines had enabled detection of KAT6B with the use of HA antibodies which are viable for use in ChIP experiments, the distribution of KAT6B in the genome was investigated. A comparison of H3K4me3 and H3K27me3 in both wild-type and HA-KAT6B cell lines was done in order to ensure that the tagging of KAT6B did not affect levels of these modifications (data not shown). KAT6B's binding to a selection of genes including active, inactive/repressed, and bivalent genes was analysed (Fig. 6.3). KAT6B binds in similar amounts to both active and bivalent gene promoters but is depleted from inactive promoters *in vivo*. KAT6B's presence at active genes is expected due to
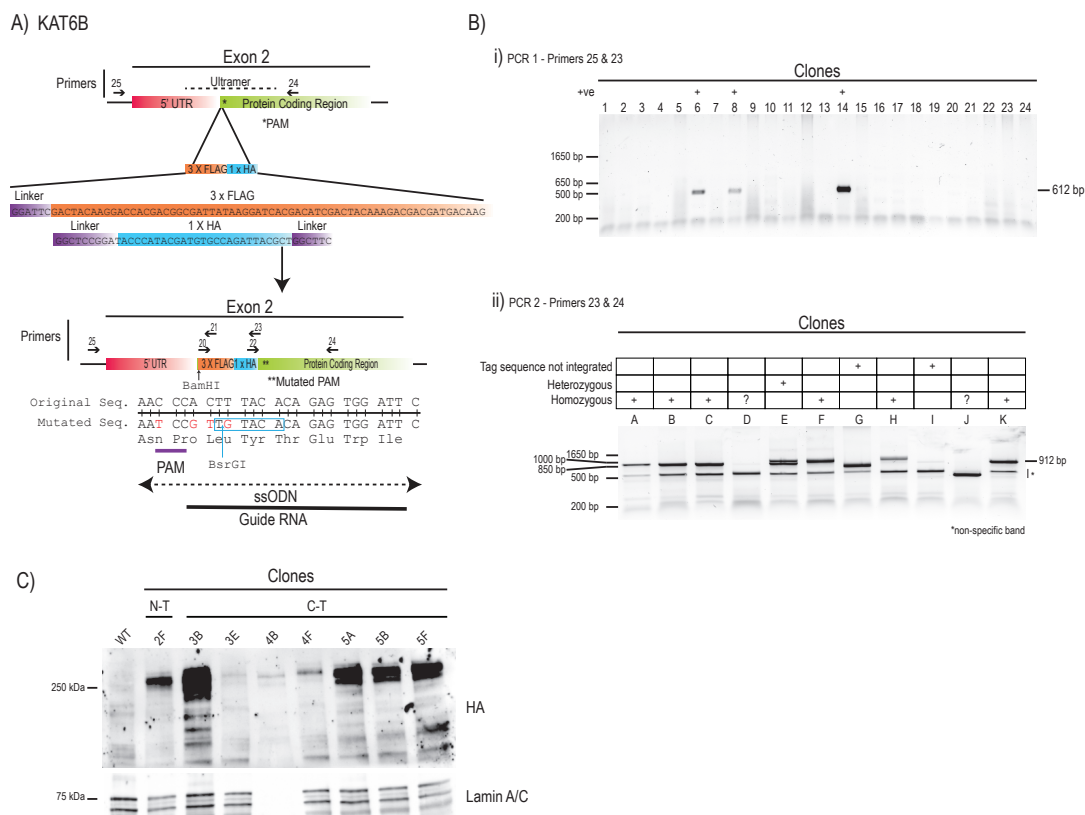
Figure 6.1 *CRISPR Strategy for the Generation of Tagged-KAT6B A) Schematic of CRISPR strategy for introducing a 1 X HA 3 X FLAG tag into the Kat6b genomic locus - not to scale - including the sequence of the guide RNA and PAM (highlighted in purple). The location and sequence of the tag is detailed along with the location of the PAM in relation to the target sequence. Location of the single stranded oligo deoxynucleotide (ssODN) donor homology is shown along with PCR primer positions. The (BsrGI) enzyme restriction site created by the desired mutations for potential use in Southern blotting is shown in blue. B) PCR screening for positive single cell-derived colonies. i) PCR 1 amplifies the part of the tag allowing identification of cells with alleles containing tagged-Kat6b. Only the cell colonies that were positive in PCR1 were then assessed via PCR2. ii) PCR 2 amplifies a region of the KAT6B gene that includes the complete sequence encoding the DNA sequence for the tag allowing identification of homozygous or heterozygous cells. A non-specific DNA band is present around 550 bp. The genotype of each clone is detailed in the table above. Each band from PCR 2 was gel-extracted and sequenced. C) Western blot analysis of previously identified (via PCR genotyping) positive cell lines. The cell lines illustrated in part B do not directly correspond to those seen in part C. Approximately hundred cell colonies were screened by PCR, only those positive via PCR were then screened via western blot. The presence of the tagged KAT6B is verified by the HA antibody. Numbered primers equate to the primer sequence detailed in Table 2.5*
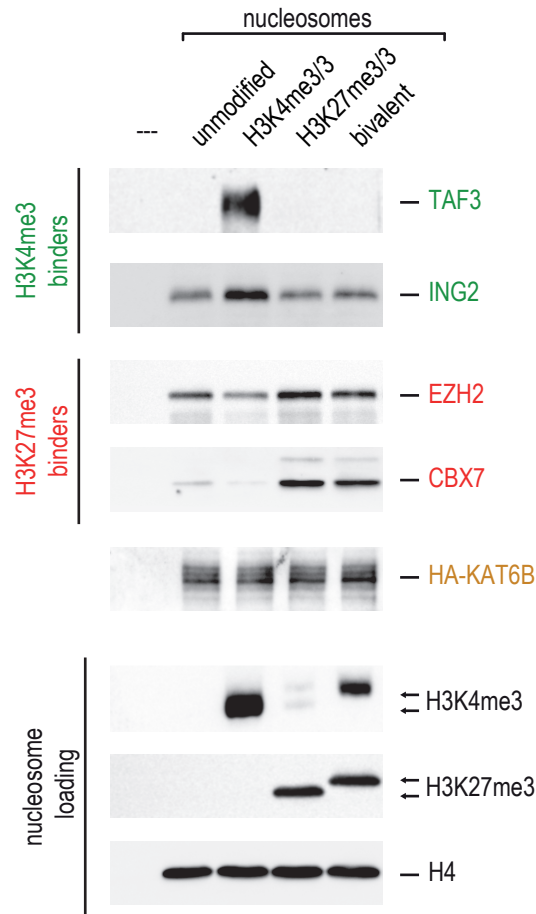
Figure 6.2 *Confirmation of KAT6B Binding.* *Western blot analysis of proteins bound to beads only (−), unmodified, H3K4me3/3, H3K27me3/3 and bivalent nucleosomes. H3K4me3 binders labelled in green, H3K27me3 binders labelled in red, KAT6B labelled in gold. H3K4me3, H3K27me3 and H4 antibodies serve as confirmation of sample identity and as loading controls.*
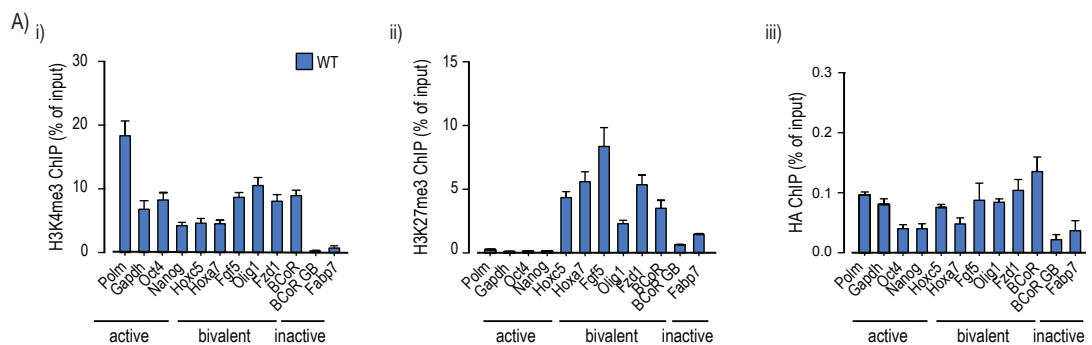
Figure 6.3 *ChIP of HA-KAT6B.* *A) Real-time PCR ratios reflect the enrichment of indicated sites when ES cells are subjected to ChIP with a selection of genes covering active, inactive and bivalent genomic regions are shown. Results for i) H3K4me3 antibody, ii) H3K27me3 antibody, iii) HA-KAT6B antibody. WT - HA-KAT6B cells. All experiments were performed in triplicate with the standard error of the experiments shown. H3K4me3 is present at both active promoters and bivalent domains. H3K27me3 is present at repressed promoters, gene bodies and bivalent domains. HA-KAT6B is present bound to both active and bivalent promoters. ChIP was completed by Dr Philipp Voigt with help from Kim Webb, Viktória Major and myself.*

145

its recruitment by H3K4me3, its mark's association with active transcription and placement of H3K14ac and H3K23ac both which are found at active promoters (Sheikh and Akhtar, 2019; Klein et al., 2017). The presence of KAT6B at bivalent domains is in agreement with the *in vitro* MS data. The presence of KAT6B bound to bivalent domains *in vivo* (Fig. 6.3) could be due to one of two reasons: KAT6B is specifically recruited to the combination of histone PTMs at bivalent domains, or the KAT6B is recruited by the H3K4me3 mark and is not affected by the presence of H3K27me3. ENOK a KAT6B homolog in Drosophila has been shown to interact with PRC1 (Scott et al., 2001), if this interaction remains in mammals, KAT6B could be recruited to H3K27me3 marked domains via interaction with PRC1. However, KAT6B is not recruited to repressed domains *in vivo* (Fig. 6.3), showing that this interaction, if present in mice, is not sufficient for KAT6B recruitment to repressed domains. However, the lack of KAT6B bound at repressed domains and a decrease of H3K14ac and H3K23ac is in accordance with current literature, showing a decrease of H3K14ac modification at repressed genes (Karmodiya et al., 2012).

KAT6B is bound in equivalent amounts at both active and bivalent domains *in vivo*, this is different to the prediction made from the *in vitro* MS results. That may be ascribed to a variety of reasons: the dynamic modification state of nucleosomes *in vivo*; proteins present *in vivo* not present in the nuclear extract obtained, which might modify KAT6B's recruitment; or a lower concentration of bivalent nucleosomes at individual genomic locations in comparison to the *in vitro* experiment. In addition, there are marks present on nucleosomes *in vivo* that are not present in the *in vitro* experiment such as KAT6B's own marks H3K14ac and H3K23ac, which can affect KAT6B binding.

Based on the observation of KAT6B at bivalent domains we generated tagged KAT6B ESCs with either catalytically inactive MLL2 or EZH2 knockout (Fig. 6.4). All of these cell lines used to analyse KAT6B distribution across the genome were generated from N-terminally tagged KAT6B cell line designated 2F. Removal of EZH2 and MLL2 enzymatic activity should remove H3K27me3 and H3K4me3 from bivalent domains (Denissov et al., 2014; Lee et al., 2006; Boyer et al., 2006; Hu et al., 2013a) thereby determining whether KAT6B'S recruitment to bivalent domains is due to the presence of H3K4me3, H3K27me3, or the combination of both marks.

Catalytically inactive (CI) MLL2 dramatically reduced the presence of H3K4me3 specifically at bivalent domains (Fig. 6.5) (Denissov et al., 2014). Although

146
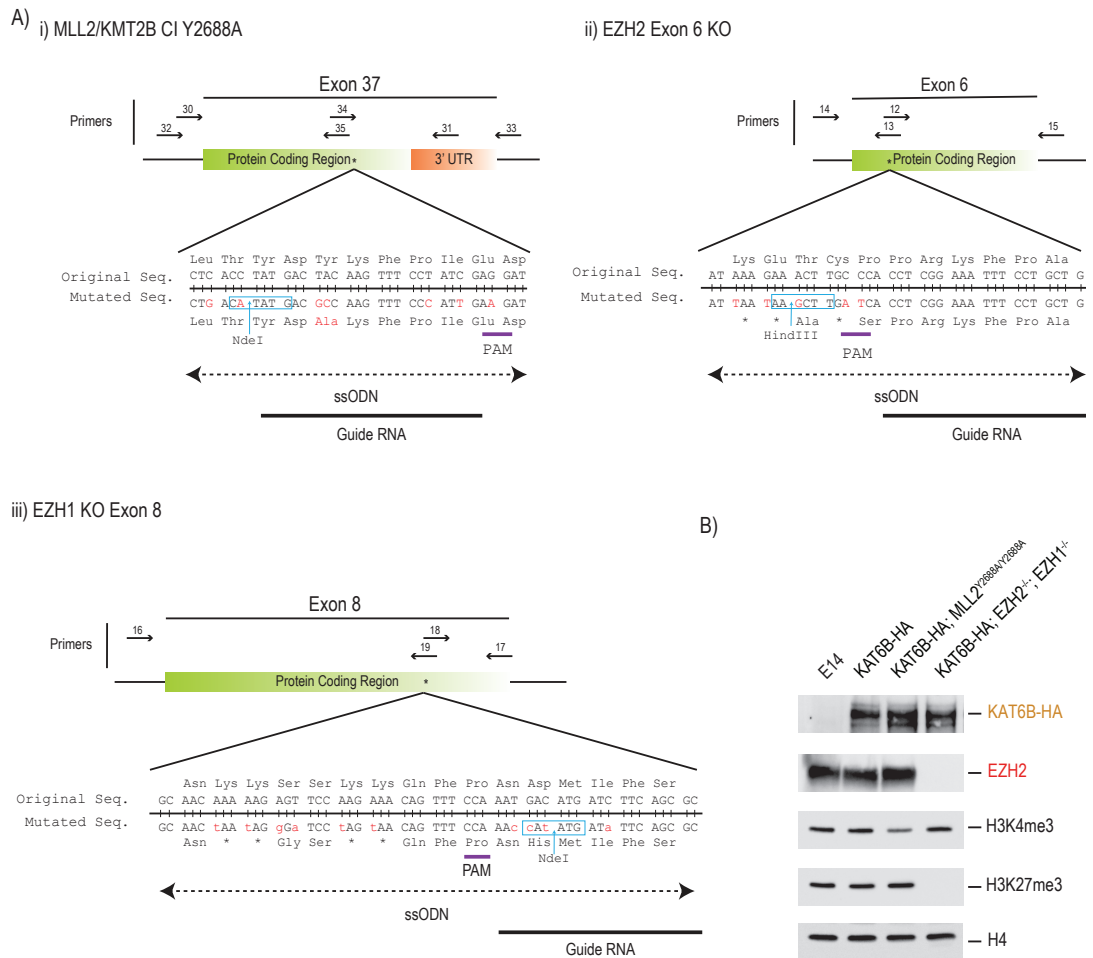
Figure 6.4 CRISPR Strategy and Confirmation of Cell Lines. *CRISPR Strategy and Confirmation of MLL2 CI, EZH1 and EZH2 KO Cell Lines A) Schematic of CRISPR strategy for generation of i) catalytic inactive MLL2 (CI) ii) EZH2 KO and iii) EZH1 KO cell lines - not to scale. The location of the desired mutations are detailed along with the location of the PAM (highlighted in purple) in relation to the target sequence. Placement of the ssODN donor homology is shown along with PCR primer positions. The sequence of the guide RNA is detailed, and enzyme restriction sites created for potential use in Southern blotting by the desired mutation are shown in blue. Numbered primers equate to the primer sequence detailed in Table 2.5. B) Western blot (WB) analysis of previously identified positive cell lines. The presence or absence of the proteins are shown by either an antibody specific to the protein itself or by reduction of the mark the enzyme places (an indirect readout). The cell lines shown were used in later ChIP and MS experiments.*
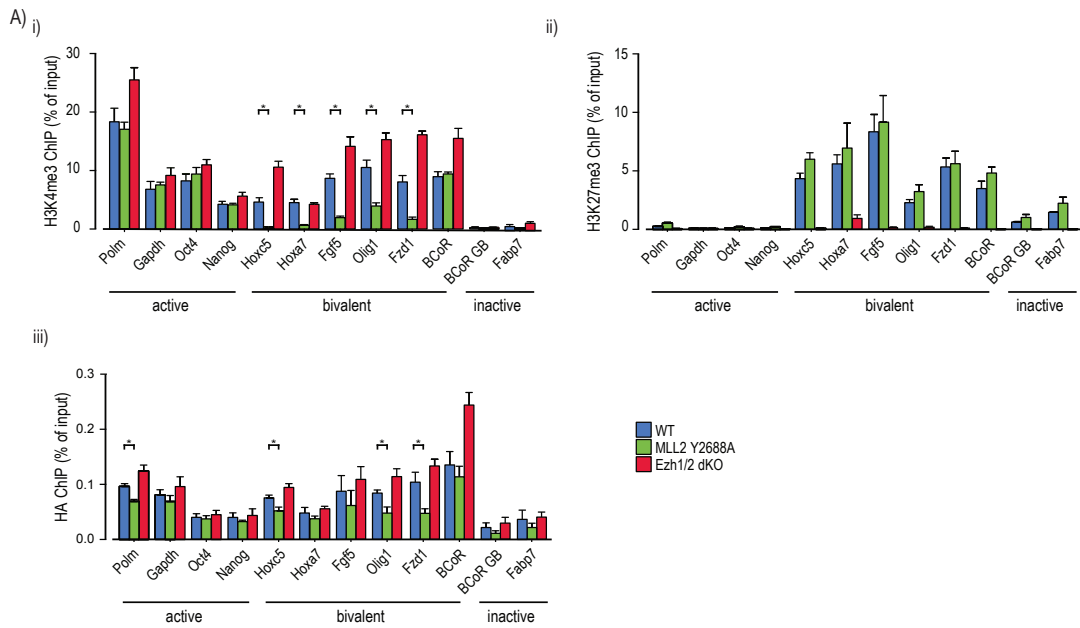
Figure 6.5 *Confirmation of KAT6B Binding In Vivo.* *ChIP analysis of HA-KAT6B binding and the presence of H3K4me3 and H3K27me3 modifications at selected promoters. i) H3K4me3 ChIP at active, bivalent and inactive promoters/gene bodies. ii) H3K27me3 ChIP at active, bivalent, and inactive promoters or gene bodies. iii) HA-ChIP (KAT6B) at active, bivalent, and inactive promoters or gene bodies. WT - HA-KAT6B, MLL2 Y2688A - HA-KAT6B and MLL2 CI, EZH1/2 dKO - HA-KAT6B EZH1/2 knockout. All experiments were performed in triplicate with the standard error of the experiments shown. ChIP was completed by Dr Philipp Voigt with help from Kim Webb, Viktória Major and myself.*

tagged-KAT6B EZH2 $^{-/-}$ decreases the amount of H3K27me3 present at bivalent domains there was residual H3K27me3 remaining (data not shown). It was therefore decided to create a double knockout (dKO) of EZH2 and EZH1 to ensure complete removal of H3K27me3, yielding Kat6b tagged, EZH1/2 dko ESCs (see Fig. 6.4). This combination of mutations removed H3K27me3 present at all genomic loci analysed including bivalent domains (Fig. 6.5).

Removal of EZH1/2 and consequent reduction in H3K27me3 levels resulted in increased levels of H3K4me3 particularly at bivalent domains, while removal of MLL2 did not result in an increase of H3K27me3 levels at bivalent domains. Removal of H3K4me3 at bivalent domains resulted in a decrease of KAT6B, while KAT6B was unaffected by the reduction in H3K27me3 at bivalent domains (see Fig. 6.5), this result is slightly different than predicted from the *in vitro* MS results. This could be due to different modification states of nucleosomes *in vivo* – containing a greater variety of marks – which may affect KAT6B recruitment, for example H3K23ac. KAT6B may be recruited to bivalent domains by multiple mechanisms *in vivo* including by proteins absent from the nuclear extract but present *in vivo*. KAT6B may be recruited to bivalent domains by H3K4me3 and

is unaffected by the co-localisation of H3K27me3. KAT6B may be recruited by the specific bivalent conformation *in vivo* however, the effect of H3K27me3 on KAT6B recruitment is masked by unknown interactions.

Does H3K27me3 play a role in the function or recruitment of KAT6B to bivalent domains? KAT6B's similarity with the *Drosophila melanogaster* protein ENOK led to the hypothesis that KAT6B may be recruited to bivalent domains by interaction with PRC1 components. PRC1 can be recruited by multiple methods including KDM2B and H3K27me3/CBX proteins; not all of these pathways are reliant on the presence of H3K27me3. KAT6B is not effected by the decrease of H3K27me3 at bivalent domains *in vivo*, suggesting that either KAT6B does not interact with PRC1 in ESCs, or that it interacts with PRC1 and is recruited by mechanisms not dependent on H3K27me3. Purification of the tagged KAT6B allowed analysis of interacting proteins (Fig. 6.6). Examination of the proteins that were co-purified with HA-KAT6B showed no PRC1 components. Consequently, it is unlikely that KAT6B interacts with subunits of PRC1 in mouse ESCS and may be recruited by a different method to bivalent domains.

The presence of H3K27me3 is not required to recruit KAT6B *in vivo*; however, its effect on KAT6B function has not been investigated. The potential effect of H3K27me3 on KAT6B was investigated by HAT (Histone AcetylTransferase) assays. H3K27me3 stimulates the catalytic activity of KAT6B (Fig. 6.6). This stimulation could help the placement of KAT6B marks/modifications and therefore increase recruitment of the enzyme, stabilise binding and help amplify its function. The increased catalytic activity seen at H3K27me3 modified nucleosomes may be due to H3K27me3 interaction with KAT6B's MYST domain, which binds close to H3K27me3 in order to place the H3K23ac. It is possible the H3K4me3 is required for recruitment while H3K27me3 simultaneously stabilises the binding interactions and increases KAT6B's catalytic activity; however, currently no crystal structure has been attempted with a bivalent nucleosome.

KAT6B is recruited mostly to bivalent nucleosomes *in vitro*, and is bound to both active and bivalent domains *in vivo* however, in the HAT assay the greatest catalytic activity seen is on H3K27me3/3 nucleosomes. These *in vitro* assays measure different events: the pulldown assay measures enrichment or binding of a protein, while the HAT assay measures catalytic activity. The catalytic activity of KAT6B on H3K27me3/3 nucleosomes does not require stable binding, explaining the lack of enrichment of KAT6B at H3K27me3/3 *in vitro* and at repressed domains *in vivo* (see Fig. 5.6, Fig. 5.10 & Fig. 6.5). *In vitro* KAT6B's catalytic

A)

protein	uniprot	MW (kDa)	peptides	unique peptides
KAT6B	Q8BRB7	208.52	54/53/45	54/4/44
BRPF1	A0A0N4SVG7 B2RRD7	137.99 137.31	23/29/– 23/29/24	1/1/– 1/1/2
BRPF2	Q3UE68	119.71	16/25/25	15/25/24
BRPF3	B2KF05	135.29	6/7/5	5/6/4
ING5	Q9D8Y8	27.798	4/5/4	4/5/4
EAF6	V9GX42	20.518	2/3/4	2/3/4
RNF2	Q9CQJ4	37.623	–/–/–	–/–/–
CBX7	Q8VDS3	18.109	–/–/–	–/–/–
EED	Q921E6-3	48.765	–/3/3	–/3/3
EZH2	Q61188	85.291	–/–/–	–/–/–
SUZ12	Q80U70	83.025	–/–/–	–/–/–

B)
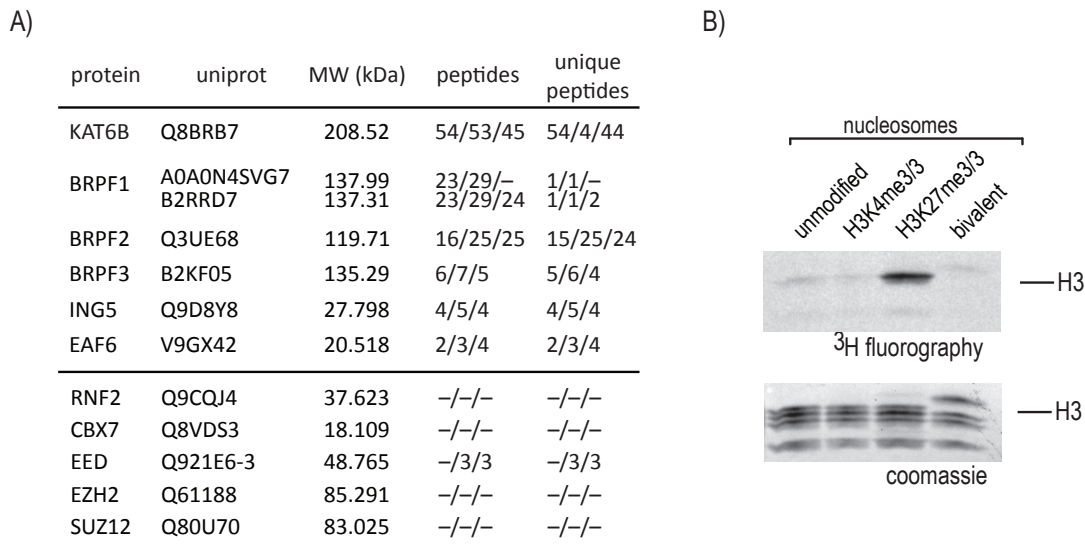
Figure 6.6 *KAT6B Purification and HAT Assay.| A) Mass spectrometry analysis of the purified tagged HA-KAT6B and co-purified proteins. B) Histone acetyltransferase (HAT) assay. Purified KAT6B is incubated with four different types of nucleosomes: unmodified, symmetrically modified with H3K4me3 or H327me3 and bivalent asymmetrically modified nucleosomes. The activity of the purified HA-KAT6B is evaluated by the intensity of the ^{3}H band. KAT6B exhibits most activity when incubated with nucleosomes symmetrically modified with H3K27me3. The coomassie of each condition shows equal amounts of nucleosomes in each condition. HAT assays were completed by Kim Webb.*

activity is stimulated by the presence of symmetrically modified H3K27me3/3 nucleosomes; however, *in vivo*, at repressed domains the presence of KAT6B's marks, H3K14ac and H3K23ac, are depleted in comparison to both active and bivalent domains. Additional factors in the cell and potentially the NE may regulate KAT6B's binding and catalytic activity *in vivo*, reducing activity at inactive domains. Proteins present *in vivo* and in the NE may prevent KAT6B binding to inactive domains or H3K27me3/3 nucleosomes, respectively – either by selection of H3K4me3 modified nucleosomes or by competition for H3K27me3/3 nucleosomes.

6.3.1 The Function of KAT6B

The previous section confirmed KAT6B's recruitment to bivalent domains *in vivo* but its function still remains obscure. Although KAT6B is a known histone acetyltransferase, it was only identified in 2000 (Thomas et al., 2000). Due to its similarity with KAT6A its function in ESCs has been difficult to isolate and previously unknown functions, including its role in placement of H3K23ac have been identified as recently as 2015 (Simó-Riudalbas et al., 2015). A paper

published in 2019 showed its importance in ESCs for regulation of NANOG and OCT4 binding and neuronal differentiation (Cosentino et al., 2019).

To fully investigate KAT6B's function in the cell and its importance at bivalent domains we generated KAT6A and KAT6B individual and double knockouts (Fig. 6.7). Due to the similarity of the proteins and overlap of their marks, we predicted a redundancy in function and therefore to fully remove the marks and function of KAT6B a double knockout (dKO) would be required. Although initial recruitment and placement of the histone PTMs may be specific to each enzyme, for example KAT6B or KAT6A, once the marks (H3K14ac) are placed then multiple enzymes can be recruited in a feedback mechanism to maintain it. As a result the knockout of a singular protein will not be sufficient to remove the mark from that genomic location. This was confirmed by ChIP of cell lines with a singular KO (Fig. 6.7 & Fig. 6.8). Knockout of either KAT6A or of KAT6B does not significantly reduce H3K14ac or H3K23ac marks placed by both KAT6A and KAT6B, however, removal of both enzymes does decrease the prevalence of the mark. KAT6B KO results in a definite increase in both H3K4me3 and H3K27me3 placement especially at bivalent domains, while this effect is not seen in the KAT6A or double KO. The singular knockout could cause a specific effect that is ameliorated with removal of KAT6A, or this may be due to a specific clonal effect. The increase of H3K4me3 at active promoters from KAT6B KO does not seem to result in an increase in mRNA expression (Fig. 6.8).

To gain a more complete picture of KAT6B's importance in whole cell and in development, the cell lines generated by CRISPR were differentiated into glutamatergic neurons. E14 ESCs were differentiated using a protocol adapted from Bibel et al., 2007 (Fig. 6.9). As can be seen from images taken during the differentiation process there are clear phenotypic differences between the cell lines (Fig. 6.9). The KO cells were slightly flatter and more spread out than E14 WT cells in their undifferentiated state. Their differences in the genotype resulted in an obvious phenotypic difference by day 15 of the differentiation protocol. The knockout cells are unable to form dendrites, and KAT6B KO or dKO cells do not express the neuronal marker – Tuj1 (Fig. 6.9). KAT6B and KAT6A KO cells are unable to differentiate into neurons.

Analysis of gene expression throughout differentiation enabled insight into the cause of the differentiation defect. The general markers for ESCs (pluripotency factors) including Oct4 and Nanog are expressed in slightly lower amounts in the mutant cells suggesting a low level disruption of gene expression in ESCs.
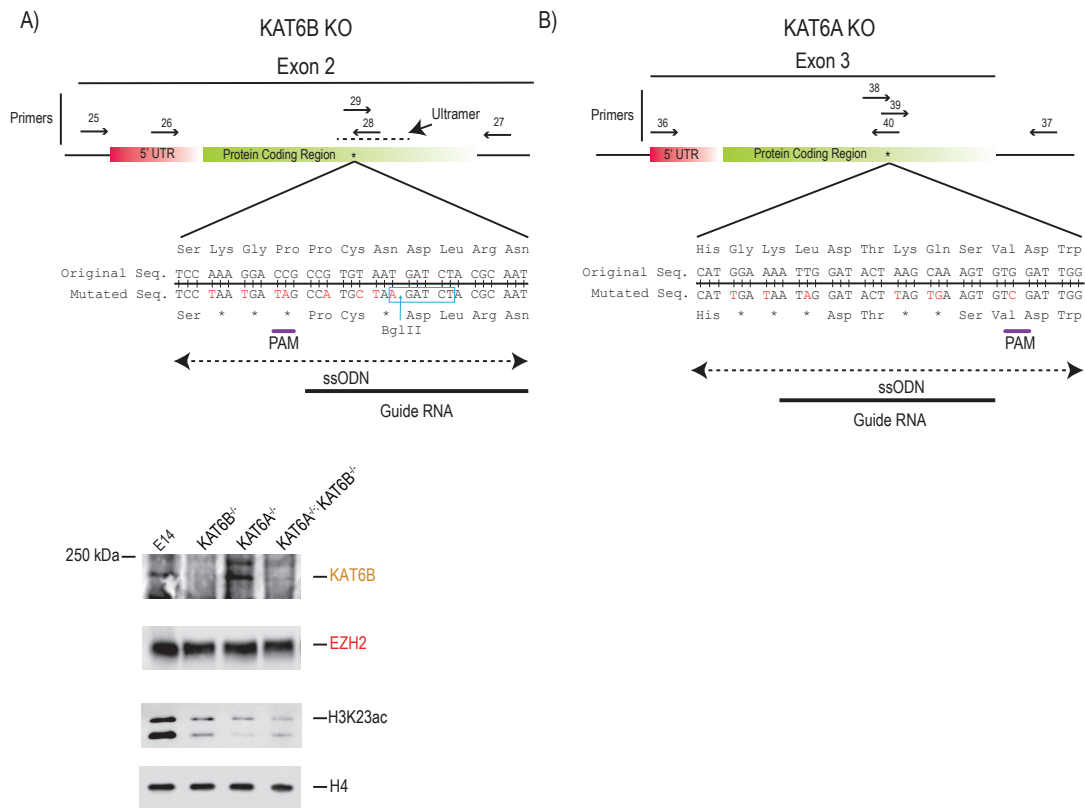
Figure 6.7 *CRISPR Strategy and Confirmation of KAT6A and KAT6B KO Cell Lines.* *CRISPR strategy (not to scale) for A) KAT6B KO and B) KAT6A KO. KAT6A/B double KO cells lines are both strategies combined. Both the original and mutated sequences including the guide RNA are shown with the PAM highlighted in purple. Any enzyme restriction sites created for potential use in Southern blotting by the designed mutations are shown in blue. Placement of the ssODN donor homology is shown along with PCR primer positions. Numbered primers equate to the primer sequence detailed in Table 2.5. C) Western blot confirmation of the knockout cell lines. KAT6A KO was unable to be confirmed by western blot at the protein level as all antibodies tested showed insufficient specificity. KAT6A KO was instead confirmed by a reduction in a histone PTM that it places - H3K23ac.*
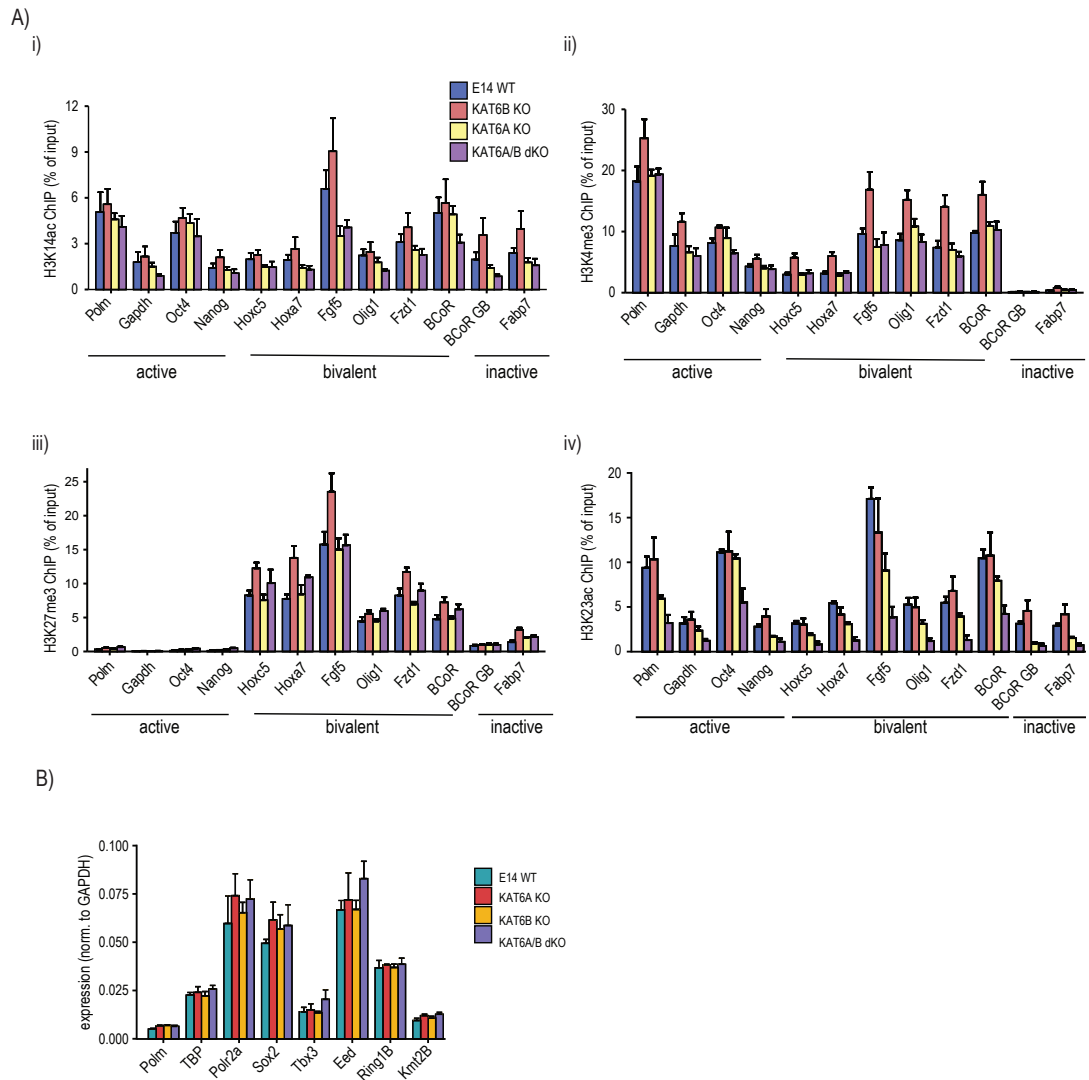
Figure 6.8 ***KAT6A and KAT6B KO ChIP*** *A) ChIP quantitative real time PCR detection of i) H3K14ac, ii) H3K4me3, iii) H3K27me3, iv) H3K23ac in E14 WT in blue, KAT6A KO in pink, KAT6B KO in yellow and KAT6A/B dKO in purple, cell lines. B) RT-qPCR expression analysis, normalised to GAPDH, of selected genes in E14 WT in turquoise, KAT6A KO in red, KAT6B KO in orange and KAT6A/B dKO in blue, cell lines. All experiments were performed in triplicate with the standard error of the experiments shown. ChIP was completed by Dr Philipp Voigt with help from Kim Webb, and Viktória Major.*
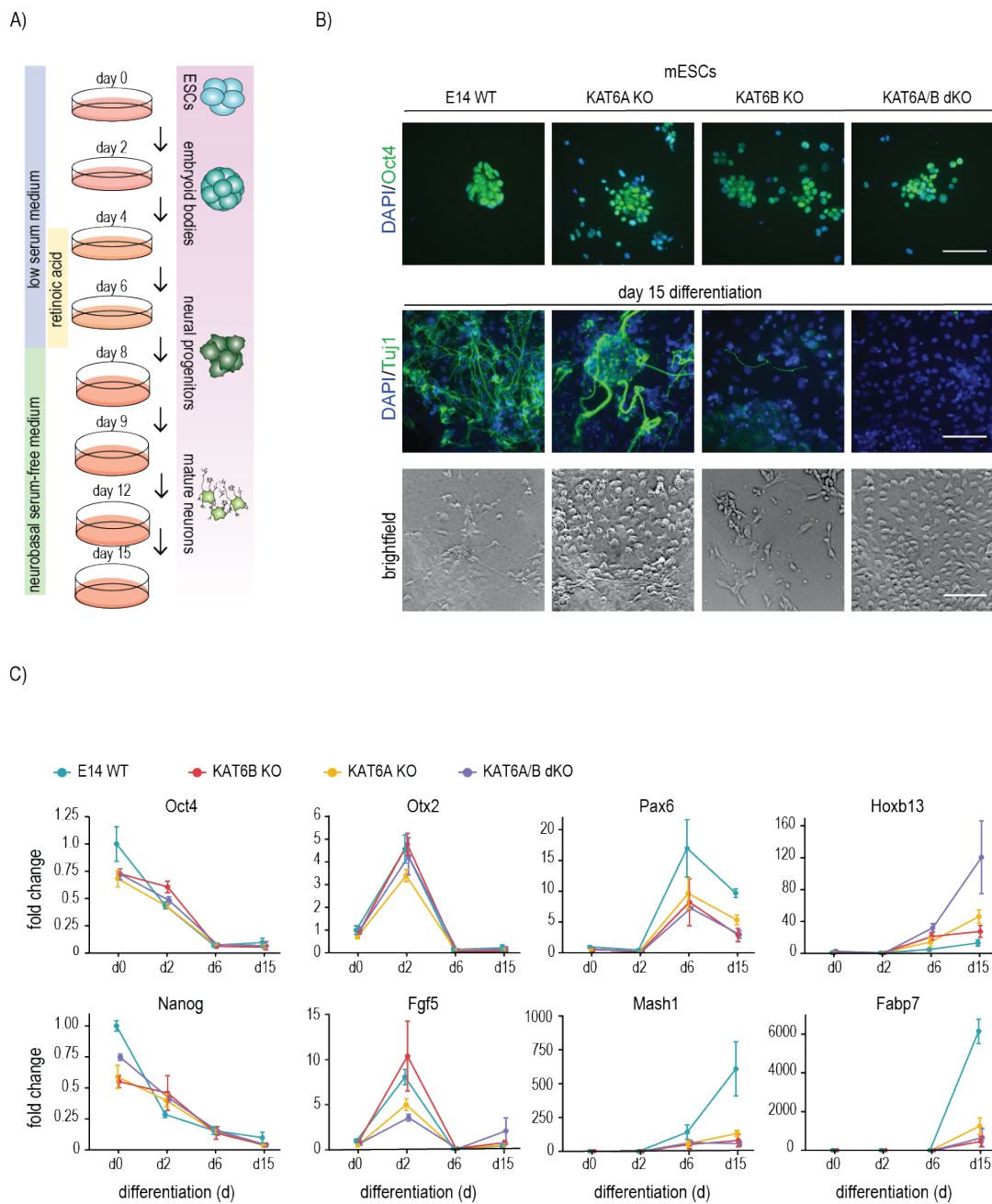
Figure 6.9 *Differentiation of Cell Lines. A) Schematic representation of the neuronal differentiation protocol. B) Cells at various stages of neuronal differentiation were grown on glass coverslips, permeabilized, fixed, and processed for in situ staining using antibodies as indicated. C) Quantitative real time PCR was used to follow the expression and regulation of a selection of genes known to be important in pluripotency and neuronal differentiation. All experiments were performed in triplicate with the standard error of the experiments shown. Neuronal differentiation assays were completed by Marie Warburton, with the RT-PCR being completed by Dr Philipp Voigt. The immunofluorescence microscopy was performed Dr Katy McLaughlin.*

The slight decrease in Nanog and Oct4 expression seen in KAT6B$^{-/-}$ ESCs was not observed by Costino 2019. This slight decrease may explain the slight visual differences seen in the undifferentiated cells. The defect in differentiation becomes more evident by day 6 with neuronal genes such at Pax6 and Mash1 failing to be properly upregulated. KAT6B and KAT6A are needed for proper differentiation of ESCs. The KAT6B KO ESC defect in neural differentiation was previously observed by the Guberman lab in 2019 (Cosentino et al., 2019).

6.4 Conclusion

CRISPR was used to tag *KAT6B* allowing purification of the protein, identification of interacting proteins and detection of binding sites by ChIP *in vivo*. No PRC1 subunits were co-purified with HA-KAT6B suggesting that KAT6B does not strongly interact with any PRC1 components in mouse ESCs and is therefore not recruited to bivalent domains by interaction with PRC1. KAT6B was confirmed to bind to both active and bivalent domains *in vivo* and to be recruited by H3K4me3. Although H3K27me3 did not seem to directly recruit KAT6B, its presence was shown to increase KAT6B's catalytic activity *in vitro*. Mutations or truncations of protein domains contained within the MORF complex such as the KAT6B's MYST domain or ING5's PHD domain would help isolate the mechanisms of H3K27me3's stimulation of enzymatic activity and of KAT6B's recruitment to bivalent domains.

KAT6A and KAT6B have redundant functions and can be recruited to the histone PTMs placed by the other enzyme, this is demonstrated by the requirement of a KAT6A/B dKO in order to decrease the level of H3K14ac or H3K23ac at specific locations. KAT6B$^{(-/-)}$, KAT6A $^{(-/-)}$ and dKO cells have defects in neuronal differentiation, suggesting that the marks they place or the complexes themselves are important in development. Generation of cell lines with catalytically inactive KAT6B or KAT6A would help distinguish between the importance of the MORF/MOZ complex binding, and the histone PTMs placed by the MORF/MOZ complex. Both singular KOs have defects in differentiation, with KAT6B$^{(-/-)}$ having the most severe phenotype, indicating that the function of these enzymes are important and both KAT6A and KAT6B have essential, non-redundant, roles in differentiation. Pluripotency genes were downregulated in all cell lines; however, the KAT6B and KAT6A deficient cell lines had difficulty upregulating neuronal specific genes. Genome-wide RNA expression analysis at

each stage of differentiation would give a more complete and thorough analysis of the genome-wide effects of the KAT6B knockout.

There are slight differences between the results obtained by *in vitro* and *in vivo* experiments (discussed further in Chapter 7), which can be attributed to multitude of differences between these types of experiments. The pulldown experiments investigated protein binding to the simplest type of bivalent nucleosome, this helps to simplify the experiment and isolate protein binding events that are solely reliant on the bivalent mark conformation. However, the simplified experiment lacks the multitude of complicating factors found in the nucleus such as a range of different histone PTMs, the dynamics of nucleosome modification or interactions with nucleosomal arrays, to name but a few. As a result of the *in vitro* experiment not fully encapsulating the interactions occurring *in vivo*, there are some difference between the *in vivo* and *in vitro* results. The results from an *in vitro* experiment gives results that have clear, defined causes but may not fully explain the *in vivo* system due to the system's lack of complexity. However, they do provide some clarification of biological mechanisms and isolate events, that when combined, help to explain the more complicated events occurring within the cell. The *in vitro* approach can be applied to more complex combinations of marks at bivalent nucleosomes, and also investigate the effects on protein binding of an array of bivalent nucleosomes.

Chapter 7

Discussion

In this thesis, I aimed to isolate the proteins bound to bivalent asymmetrically modified nucleosomes and identify how the asymmetry of nucleosome modifications can effect protein binding. Establishment, optimisation and validation of an *in vitro* pulldown approach, utilising specifically modified nucleosomes, demonstrated the effect histone modifications have on protein recruitment. This technique allowed identification of proteins recruited to bivalent nucleosomes and determined the effect a singular asymmetric modification can have on protein binding. KAT6B was identified bound to bivalent nucleosomes *in vitro* and was demonstrated to bind bivalent domains *in vivo* via the generation of KAT6B tagged cell lines. The role of KAT6B in ESCs and during differentiation was investigated via differentiation of KAT6B KO ESCs. The identification of proteins that bind to bivalent nucleosomes *in vitro*, and bivalent domains *in vivo* further elucidates the potential biological function of bivalent domains. The effects of asymmetric marks on protein recruitment will help in the identification and classification of the transcriptional state of chromatin *in vivo*.

7.1 Histone Mark Binders in Embryonic Stem Cells

To validate the experimental approach used, the proteins enriched on the symmetrically modified H3K4me3/3 and H3K27me3/3 nucleosomes were investigated and compared to findings in the literature. The proteins bound to symmetrically modified nucleosomes *in vitro* largely correlate with previously described binding

preferences seen by Bartke et al., 2010, Vermeulen et al., 2007 and Vermeulen et al., 2010. Proteins containing reader domains such as PHD fingers and Tudor domains known to bind the H3K4me3 mark on H3 N-terminal tail peptides or modified nucleosomes also bound to H3K4me3/3 nucleosomes here, including ING2, TAF3, Spin1 and BPTF (Ali et al., 2014; Shi et al., 2006; Yang et al., 2012).

The H3K4me3/3 nucleosome enriched for proteins that bind directly to the H3K4me3 mark and for proteins that interact with H3K4me3 binders. BAP18 and DPY30 are enriched bound to H3K4me3/3 nucleosomes however, they do not bind H3K4me3 directly, they interact with complexes which contain subunits that bind directly to H3K4me3; such as BPTF in the NURF complex and CFP1 in the SET1 complex resulting in the recruitment of both BAP18 and DPY30 to H3K4me3/3 respectively (Tremblay et al., 2014; Vermeulen et al., 2010; Sun et al., 2016; Eberl et al., 2013; Mahadevan and Skalnik, 2016). These proteins bind to the active H3K4me3 mark, mediating a range of functions from recruitment of proteins that enable the recruitment of the pre-initiation complex (PIC) via TAF3 binding to H3K4me3 (Lauberth et al., 2013; Vermeulen et al., 2007) or that repress active genes as in the case of ING2 and SIN3a repression (Shi et al., 2006; Grzenda et al., 2009) which can lead to the removal of activating marks such as histone acetylation and H3K4me3.

The presence of all these proteins that differ in function at the same genomic location reinforces the idea that although H3K4me3 is correlated with active gene expression and helps initiate RNAPII transcription, it needs to be reinforced by the presence of other marks and maintained at that genomic location. There is a balance between active and repressive marks and chromatin states and there are always proteins that can bind and cause the chromatin state to change unless there is reinforcement and maintenance of a specific mark. Without reinforcement and maintenance of H3K4me3, it would be removed by multiple demethylases including KDM5B and the gene not expressed (Kidder et al., 2014). The chromatin state is therefore always composed of multiple chromatin marks and it is the interplay of these marks and their cumulative effects that results in the chromatin state and its transcriptional output.

The proteins known to bind specifically to H3K27me3 such as CBX7 (Morey et al., 2012), which contains a chromodomain, are enriched bound to symmetrically modified H3K27me3 nucleosomes. CBX7 is a canonical PRC1 complex subunit along with PHC1, also required for the canonical PRC1 formation. PHC1 is also

enriched bound to H3K27me3/3 and is required for PRC1 mediated compaction of DNA (Isono et al., 2013; Wani et al., 2016; Kundu et al., 2017). PHC1 does not contain a H3K27me3 binding site and is likely recruited to H3K27me3/3 via interaction with CBX7.

Other members of the Polycomb family are not enriched bound to the nucleosome symmetrically modified for H3K27me3. As previously mentioned, the other CBX proteins are mutually exclusive and have different binding preferences and are therefore recruited to different genomic environments. The other PRC1 subunits such as RING1 and PCGF variants are present in multiple PRC1 complexes. Therefore, if some PRC1 variant complexes are present bound to multiple differently modififed nucleosomes then their enrichment on H3K27me3/3 nucleosomes will be decreased and classified as not significant. This data suggests that there is a variant of PRC1 recruited *in vitro* specifically to H3K27me3/3 nucleosomes, a canonical PRC1 complex containing both CBX7 and PHC1.

As both CBX7 and PHC1 are the most highly expressed CBX and PHC protein variant in ESCs, the majority of the ESC canonical PRC1 complex will contain them. This indicates that the vast majority of canonical PRC1 complex in ESCs are recruited specifically to H3K27me3/3 nucleosomes. Components of the origin recognition complex (ORC) including LRWD1 are seen enriched bound to the H3K27me3/3 nucleosome replicating earlier results shown in Bartke et al., 2010 and Vermeulen et al., 2010. These results confirm previously known interactions and validate the experimental approach used in this study.

EZH2, the catalytic subunit of PRC2, was depleted from H3K4me3/3 nucleosomes. PRC2 is inhibited from placing the H3K27me3 modification on a tail previously modified with H3K4me3, and has been shown by ChIP-Seq to be associated with CpG islands and genomic regions containing H3K27me3 (Voigt et al., 2012; Schmitges et al., 2011; Mendenhall et al., 2010; Li and Reinberg, 2011). PRC2 contains H3K27me3 binding sites but does not contain any H3K4me3 binding sites, suggesting that PRC2 has a preference to be bound at chromatin containing H3K27me3 rather than H3K4me3, although this has yet to be robustly demonstrated *in vitro*. Bartke et al., 2010 did not see any inhibition in EZH2 binding to H3K4me3 modified nucleosomes. Although in more recent results Bartke's lab does see this inhibition in EZH2 binding to H3K4me3/3 nucleosomes. PRC2 components including EZH2 are not seen enriched in the H3K27me3/3 compared to the unmodified nucleosome, opposing results seen in Bartke et al. (2010). The potential reasons for this have been discussed previously in Chapter

5.

In contrast to the H3K4me3/3 nucleosome, the presence of H3K27me3 did not prevent recruitment of any proteins that would bind normally to the unmodified H3 N-terminal tail.

7.2   The Effect on Protein Binding of a Singular Asymmetric Mark

The proteins that have been identified as binding symmetrically modified nucleosomes did not demonstrate the same behaviour in the context of asymmetrically modified nucleosomes. Neither H3K4me3 nor H3K27me3 asymmetrically modified nucleosomes enriched for any known H3K4me3 or H3K27me3 mark protein binders in comparison to the unmodified nucleosome. This could result from two possible causes: a requirement for a higher concentration of either mark to bind or the protein requires a symmetric nucleosome to bind. These potential reasons were investigated via a nucleosome titration (see Chapter 5). A selection of proteins that bind differently to symmetrically modified nucleosomes and asymmetrically nucleosomes were investigated. The binding preferences of TAF3, PHF2, and ING2 for H3K4me3 modified nucleosomes were examined while EZH2, CBX7, and RING1B were examined for the differences in binding to both H3K4me3 or H3K27me3 modified nucleosomes. Each protein examined displayed different recruitment to the tested nucleosomes.

TAF3, a component of the basal transcription factor complex TFIID, has a PHD finger which binds to H3K4me3 (Ingen et al., 2008; Vermeulen et al., 2007), however, TAF3 is bound to H3K4me3/3 but not to H3K4me3/0 nucleosomes. Only at the highest concentration of asymmetric nucleosomes did TAF3 bind, the level of binding seen is less than the amount of TAF3 bound to the lowest concentration of H3K4me3/3 nucleosomes. The difference in TAF3 binding to symmetric and asymmetric nucleosomes is unlikely to be due to concentration of the mark present and likely to do with the conformation of H3K4me3 marks presented on a symmetrically modified nucleosome; however, the mechanism by which this is communicated is unknown. Pulldowns with the symmetrically modified nucleosomes have a higher local concentration of modifications i.e. on the individual nucleosome. This may result in an increase in TAF3 binding,

160

consistent with allovalent model of binding - discussed below.

There is no evidence for the dimerization of TAF3 *in vivo*, suggesting that this affect may be communicated by the wide variety of proteins that interact with TAF3. TAF3's binding behaviour may be partially explainable by allovalency. Allovalency refers to an interaction between two molecules where you have more than one identical binding sites in tandem on a disordered region of a ligand and one receptor site (Klein et al., 2003; Locasale, 2008; Olsen et al., 2017). These two identical binding sites compete for the single receptor site as only one binding site can interact at one time. The presence and competition of the two binding sites increases affinity of the receptor for the ligand. In terms of TAF3 binding, the PHD domain of TAF3 is the receptor site, while the two H3K4me3 are the identical ligand binding sites. This mode of binding would explain the requirement for symmetrically modified nucleosome for TAF3 to bind - due to both ligand binding sites needing to be on the same ligand, in this case the nucleosome, for allovalent binding to occur. The difference in TAF3 binding is not linearly proportional to the concentration of H3K4me3 modification present indicating co-operative binding – fitting with the allovalent model. However, when considering the larger protein complexes involved, that is TFIID, and the multiple interaction sites present in TFIID and on the nucleosome itself, the mechanism of binding becomes more complex. TAF3 could be recruited to H3K4me3/3 nucleosomes via the interaction of other TFIID subunits and their non-identical binding sites with the rest of the nucleosome via multivalency or via fuzzy binding (Olsen et al., 2017).

PHF2 is a lysine demethylase that is recruited to H3K4me3 by its PHD domain and mediates H3K9me2 demethylation leading to gene activation (Baba et al., 2011). PHF2 displays similar behaviour to TAF3, indicating that for both of these H3K4me3 binding proteins the conformation of the symmetrically nucleosomes is important in regulating binding. These results are indicative of a wider binding behaviour that is seen for all H3K4me3 binders bound to the H3K4me3/3 nucleosome condition in the mass spectrometry experiment. This indicates that a singular H3K4me3 modification is not sufficient to maintain the presence of these active mark binders potentially due to the requirement of a symmetric conformation in order to bind and therefore H3K4me0/3 nucleosomes may not be adequate to maintain an active chromatin state.

ING2 contains a PHD finger which binds to H3K4me3 and is recruited to H3K4me3/3 nucleosomes (Peña et al., 2006). ING2 is able to recruit both

deacetylases (HDACs) and histone methyltransferases (HMTs). Its interaction with Sin3a-HDAC1, helps to remove activating acetylation marks leading to gene repression (Shi et al., 2006; Doyon et al., 2006; Peña et al., 2006) while helping modulate histone methylation (Goeman et al., 2008). Although, ING2 binds H3K4me3, and is enriched at H3K4me3/3 nucleosomes, it is not enriched at asymmetrically modified H3K4me0/3 nucleosomes. The titration experiment showed that ING2 has highest affinity for H3K4me3/3 nucleosomes. The asymmetric H3K4me3/me0 nucleosomes exhibited higher binding than the asymmetric H3K4me3/- tailless nucleosomes, signifying that ING2 is partially recruited by unmodified H3 N-terminal tail. Goeman et al. (2008) demonstrated ING2's preference for binding H3K4me3 and ability to bind unmodified H3 albeit weakly. This weak binding affinity could be the cause of the increased ING2 recruitment to H3K4me3/me0 compared to the H3K4me3/- (tailless) nucleosome. While the behaviour of TAF3 and PHF2 in the titration experiment recapitulates the binding preference seen in the MS experiment, ING2 shows more subtle binding preferences in the titration experiment. As a consequence of ING2's decreased recruitment to H3K4me0/me3 and its ability to bind H3K4me0, the difference between binding may not be classified as significant via the MS technique.

The Sin3/HDAC corepressor complex can be recruited by multiple transcription factors - it is possible that a number of these can bind H3K4me0 and help ING2 recruitment (Chandru et al., 2018; Knoepfler and Eisenman, 1999; Grzenda et al., 2009). RbAp48/46's interaction with histones is important for PRC2's nucleosomal recognition and may help with Sin3a recruitment (Shi et al., 2017; Grzenda et al., 2009). RbAp48/46's interaction with H2A and H4 is not the cause of ING2's recruitment to H3K4me0 tail as this interaction is present in the H3K4me3/- nucleosomes; however, it may help stabilise ING2 at unmodified tails (Zhang et al., 1999; Grzenda et al., 2009).

ING2's behaviour is different to that of both TAF3 and PHF2, it is recruited in small amounts to unmodified H3 N-terminal tails: this difference in behaviour may be due to the difference in interacting proteins. The reader domains present in the proteins that interact with TAF3 and PHF2 are selective for active marks (Filippakopoulos et al., 2012; Su et al., 2016; Nuland et al., 2013), while ING2 interacts with a variety of complexes containing reader domains for repressive marks as well as other core histones (Murzina et al., 2008; Shi et al., 2006; Viiri et al., 2009; Grzenda et al., 2009). This liberal recruitment of ING2

162

and possibly the Sin3 complex allows removal of acetylation marks, potentially demethylation of H3K4me3, and continued binding of Sin3, resulting in spreading of deacetylation and potential establishment of a repressive chromatin state. The role of ING2's permissive recruitment may allow maintenance of ING2's presence and therefore HDAC, HMT and SIN3 complexes at active genes longer, recruitment to a more diverse range of chromatin targets, spreading of ING2's function and reinforcement of a repressive state.

EZH2, the catalytic subunit of PRC2, is recruited equally to both H3K27me3/3, H3K27me0/3 and unmodified nucleosomes. A clear depletion is seen when comparing H3K27me3/- tailless and H3K27me0/3, indicating that the unmodified H3 tail does contribute to EZH2 binding. It does not distinguish between the unmodified H3 N-terminal tail and one modified with H3K27me3, reproducing the binding pattern seen in the MS results.

RING1B, a component of the PRC1 complex, has higher binding affinity for H3K27me3/3, indicating that RING1B binds better to H3K27me3 than H3K27me0. This binding is not proportional to the H3K27me3 present, indicating that it is a combination of both concentration of mark and the conformation of the marks that affects RING1B binding. RING1B is poorly recruited by H3K27me3/tailless, suggesting that the presence of RING1B at the H3K27me0/me3 is partially recruited or stabilised by the free H3 N-terminal tail; however, the mechanism of this is unclear. The decrease in RING1B binding at H3K27me0/3 compared to H3K27me3/3 nucleosomes echoes the decrease in RING1B present between these conditions when analysed by MS. This decrease may prevent the enrichment of RING1B at H3K27me0/3 compared to unmodified nucleosomes being classified as significant.

CBX7 is a subunit of canonical PRC1 complexes and recruits PRC1 to sites of H3K27me3 by its chromodomain (Morey et al., 2012; Bernstein et al., 2006b; Min et al., 2003; Wang et al., 2004b). CBX7 is bound to H3K27me3 modified nucleosomes proportional to the amount of H3K27me3 present. There is a minor decrease of CBX7 bound to H3K27me3/tailless compared to H3K27me0/3 nucleosomes, indicating that the free H3 N-terminal tail may have a small role in stabilising CBX7 bound at nucleosomes, potentially via stabilisation of the PRC1 complex. There is a visible difference in both RING1B and CBX7 binding to symmetric, asymmetric and tailless nucleosomes in the titration experiment. However, MS analysis does not classify the enrichment of RING1B and CBX7 at H3K27me3/0 compared to unmodified nucleosomes as significant. The fold

changes indicated by the MS analysis may be lower than seen by other methods and decreased compared to *in vivo*.

The inhibition of H3K4me3 on EZH2 binding seen in the MS results is shared by RING1B and is replicated in the titration experiments. The inhibition of binding is partially relived by the presence of an unmodified H3 N-terminal tail in H3K4me0/3 nucleosomes. Although the inhibition of binding is clear, at higher concentrations of H3K4me3/3, binding of both EZH2 and RING1B to the nucleosomes is detectable. This indicates that RING1B and EZH2 may bind these nucleosomes but at low amounts and that given a high enough concentration of a reactant the binding is detectable, or that the inhibition of binding can be partially relieved by increased amounts of the nucleosome. The H3K4me3 inhibition of binding affects EZH2 to a greater extent than RING1B demonstrated in both the titration and MS experiments. RING1B binding to H3K4me3/tailless indicates RING1B's and therefore PRC1's ability to be recruited with low binding affinity to the H3K4me3 modified H3 N-terminal tail. As both RING1B and EZH2 are the subunits of PRC1 and PRC2 respectively the binding behaviour observed is not solely due to either protein but is instead due to multiple interactions between the Polycomb subunits in the relevant complex and the nucleosome.

The ability of EZH2 and RING1B to bind to both H3K27me3 and H3K27me0 tails will allow the spread of H3K27me3, H2AK119ub and transcriptional repression. This helps establish new repressive domains, encourage the switch of chromatin states from active to repressive and maintain repression via by modification and binding of unmodified, H3K4me0/3 or H3K27me3/3 nucleosomes respectively. CBX7's specific binding to H3K27me3 modified nucleosomes will help maintain already established repressive states, and initiate chromatin compaction - reinforcing repression.

Although all proteins investigated display different preferences for recruitment, there is one clear division. Proteins that bind and are associated with establishment of transcriptional repression have more permissive binding, while those associated with transcriptional activation have more limited binding parameters. TAF3 and PHF2 not only require H3K4me3 present to bind but also a symmetrically modified nucleosome, while ING2, EZH2, and RING1B can bind both modified and unmodified H3 N-terminal tails and can be recruited with different affinity to symmetrically, asymmetrically and unmodified modified nucleosomes. This finding is in agreement with previous studies that have

164

proposed a sampling mechanism for PcG recruitment, resulting in broad repressed domains (Klose et al., 2013).

All proteins examined are part of multiple complexes containing subunits with different reader domains, therefore the recruitment seen is potentially a result of all these binding affinities. Although many SET family proteins contain subunits, including CFP1, that can sample chromatin by binding to unmethylated CGIs, the number of different SET domains containing enzymes that bind to and enforce "open" chromatin or active transcription is much larger, resulting in specification of recruitment and potentially a reduction in sampling mechanisms (Yu et al., 2017; Xu et al., 2011). TAF3 and PHF2 have specific roles, which are not required at all "open" chromatin and therefore their recruitment is more restricted. Further experiments will be needed to investigate this reduced affinity to H3K4me0/3 mononucleosomes. Whether the inhibition of PHF2 and TAF3 binding to H3K4me3/0 mononucleosomes can be relieved by H3K4me0/3 dinucleosomes or more complex arrays requires further experiments. These can be completed in the same experimental set up, with dinucleosomes in a specific configuration or randomly assembled.

This permissive binding of EZH2 and RING1B to H3K4me0/3 may account for the overlap of PRC1/2 with H3K4me3 domains *in vivo*, while the exclusion of PRC1/2 from H3K4me3 domains, or areas where KDM2B recruitment is the dominant method of PRC1 recruitment may indicate regions consisting of symmetrically modified H3K4me3 nucleosomes (Ku et al., 2008; Farcas et al., 2012; Bernstein et al., 2006a). The specific binding of TAF3 to symmetrically modified H3K4me3 may allow identification of H3K4me3 marked regions composed of H3K4me3/3 nucleosomes and isolation of H3K4me0/3 domains. Previous studies have identified genes marked with H3K4me3, that have no discernible expression or RNAPII present (Vastenhouw and Schier, 2012). Our *in vitro* studies have indicated that these regions may be composed of nucleosomes asymmetrically modified for H3K4me3, resulting in TAF3 being unable to bind and recruit the pre-initiation complex to these genes. These asymmetric H3K4me3 nucleosomes may have an important role in restricting the establishment of a repressive domains and poising the genes for activation. While symmetric H3K4me3 modification prevents PRC2's placement of H3K27me3 *in vitro*, asymmetric H3K4me0/3 allows the placement of H3K27me3 on the opposing H3 tail (Voigt et al., 2012; Schmitges et al., 2011). PRC2 cannot place H3K27me3 on the same H3 N-Terminal tail containing H3K4me3, consequently

H3K4me3 demethylases would need to be recruited and H3K4me3 removed before a fully repressive state could be established. Asymmetric H3K4me3 modified nucleosomes may not recruit TAF3 and thus not form a productive PIC with RNAPII; however, they may poise the gene for activation by reducing the number tails needing to be modified before the PIC can form, or by allowing other TFs to bind – providing a platform to easily bind RNAPII (Vastenhouw and Schier, 2012; Grzybowski et al., 2015; Barski et al., 2009). The presence of H3K4me0/3 nucleosomes may act to prime the genes to switch from repressed to active transcription.

Asymmetric H3K4me3 nucleosomes have been identified in yeast, suggesting their conservation throughout evolution, indicating possible conservation of function and mechanism of regulation (Takahashi et al., 2011; Miller et al., 2001; Choudhury et al., 2019). The only H3K4 methyltransferase in yeast, Set1c, dimerises and thereby prefers to symmetrically methylate nucleosomes (Choudhury et al., 2019). The sole H3K4me3 demethylase in yeast, Jhdh2 is a member of the KDM5 protein family and contains a conserved PHD1 zinc finger that binds H3K4me0 and is found in mammalian KDM5 members (Zhang et al., 2014; Torres et al., 2015). Jhdh2 has a preference for demethylating H3K4me0/3 nucleosomes, suggesting a preference for symmetrically H3K4me3/3 nucleosomes or unmodified nucleosomes in yeast. SET1 is recruited to H3K4me3 by the SPP1 subunit, indicating a feedback mechanism, and a system for asymmetrically modified nucleosomes to become symmetric (He et al., 2019). Instead of SET1 acting on a singular nucleosome, SET1 could bind a H3 tail of one nucleosome and methylate the H3 tail of a neighbouring nucleosome helping to spread H3K4me3 asymmetry. As yet, there have been no studies looking at SET1's catalytic activity when bound to asymmetrically modified nucleosomes in comparison to unmodified nucleosomes. A catalytic study of Set1c bound to unmodified or asymmetrically modified nucleosomes would highlight the importance of H3K4me3 asymmetry *in vivo* and potentially predict the interaction of mammalian COMPASS members with H3K4me3 asymmetry.

The conservation of the SET domain-WRAD (composed of WDR5, RBBP5, ASH2L and DPY30) scaffold is one of the most conserved protein groups in eukaryotic chromatin modifiers (Ernst and Vakoc, 2012; Ali and Tyagi, 2017; Takahashi et al., 2011). The function of the WRAD complex and its requirement for catalytic activity, even without the catalytic domain, is maintained throughout evolution. The discovery of dimerization in SET1

166

indicates that members of the mammalian COMPASS family may be dimerised. The mammalian SET/COMPASS family although evolutionary well conserved has many more members and specialised functions than the SET1 complex in yeast. SET1A/B complexes are most closely related to Set1c in function and sequence, while the MLL complexes have more diverse subunits and more specific functions. It is possible that in mammals dimerised SET1A/B complexes may preferentially generate symmetrically modified nucleosomes, while the MLL complexes place the asymmetric H3K4me3. Structural studies of the mammalian COMPASS members via crystallisation or mutational experiments may elucidate whether this dimerization of COMPASS members is conserved in mammals and enzymatic assays would explain how this correlates with enzymatic function. Whether MLL2 prefers to place asymmetric or symmetric H3K4me3 may indicate control of bivalent domain formation and maintenance.

These results have highlighted the importance of examining individual protein binding using nucleosomes rather than peptides. Previous studies reliant on peptide binding assays or symmetrically modified nucleosomes have predicted that the presence of one mark, that binds the protein via peptide assays and replicates this binding in symmetrically modified nucleosomes, is sufficient for binding. The use of asymmetrically modified nucleosomes has demonstrated the importance of the nucleosome structure and the conformation of marks presented to the enzyme in determining protein recruitment and binding.

7.3   Differences Between Asymmetry and Bivalency

The H3K4me3/3 nucleosomes enriched for proteins which bind directly to or interact with proteins that bind directly to H3K4me3 and prevents enrichment of proteins that would bind to H3K27me3. The enrichment of proteins bound to the H3K27me3/3 nucleosome is lower than that seen in the H3K4me3/3 nucleosome, and fewer proteins are enriched in comparison to the unmodified nucleosome in total in comparison to H3K4me3/3. The decreased number of enriched proteins on H3K27me3/3 does not mean the consequences of recruitment are less. The direct H3K27me3 binders seem to be able to bind to a greater variety of sites although with varying affinity or be part of complexes that can recruit them to a great variety of modifications and are therefore more permissive in their recruitment. This could be due to a more permissive binding site of the proteins or resulting from the variety recruitment mechanisms as a consequence of multiple

reader domains within the repressive protein complex. The enrichment seen on symmetrically modified nucleosomes seems to be lost when investigating protein binding to singularly asymmetrically modified nucleosomes.

Bivalent nucleosomes enriched for a greater variety of proteins and with higher fold changes of enrichment than either of the singularly modified asymmetric nucleosomes. The bivalent condition is not enriched for H3K4me3 binders, consistent with data seen for H3K4me3 monovalently asymmetrically modified nucleosomes. Although H3K4me3 may not recruit binders by itself it can recruit proteins through interaction with other modifications. In addition, the presence of H3K4me3 prevents the placement of H3K27me3 on the tail that carries H3K4me3 and the formation of a repressive domain (Voigt et al., 2012; Schmitges et al., 2011). In contrast to H3K4me3 binders, proteins that bind to H3K27me3, although not enriched in the H3K27me0/3 asymmetric condition, are enriched on the bivalent nucleosome in comparison to unmodified nucleosomes. Although bivalent nucleosomes lose H3K4me3 mark binders they retain H3K27me3 mark binders, including CBX7 and EZH2. The retention of H3K27me3 protein binders was unexpected, as previous experiments have shown that they are inhibited from binding the H3K4me3 tail and are not enriched on H3K27me0/3. The difference seen in repressive mark binders between H3K27me3 containing symmetric, monovalent asymmetric, and bivalent asymmetric nucleosomes results not only from the presence of H3K27me3 and permissive binding but potentially a hitherto unknown consequence of having both H3K27me3 and H3K4me3 present on the same nucleosome. There may be a distinct recruitment pathway for H3K27me3 binders specific to the bivalent asymmetric conformation potentially by recruitment of different recruitment proteins or conformational changes in the the H3K27me3 protein binders themselves.

EPOP is the only protein associated with PRC2 that is significantly enriched in the bivalent condition compared to almost all other nucleosomes (see Fig. 5.6 and Fig. 5.8). EPOP is depleted from H3K4me3/3 nucleosomes and enriched in the H3K27me3/3 nucleosomes (see Fig. 5.6). However its binding is not affected by the presence of a singular asymmetric mark H3K4me0/3 or H3K27me0/3 (Fig. 5.2). EPOP's enrichment at bivalent nucleosomes is not solely due to the presence of a singular H3K27me3 present at the nucleosome as shown by Figure 5.6, indicating a potential method of specific recruitment to bivalent nucleosomes. EPOP has been previously shown to link PRC2 and Elongin BC complex which has been found to localise at bivalent domains and has been suggested to be

168

required to maintain low levels of transcription at bivalent genes (Beringer et al., 2016; Liefke et al., 2016). Cell lines depleted for EPOP have a decreased amount of RNAPII at the transcription start sites, and decreased transcription of bivalent genes (Beringer et al., 2016; Liefke et al., 2016), the presence of EPOP at bivalent domains may therefore help to recruit RNAPII. EPOP's presence at bivalent domains reduces the binding of JARID2-containing PRC2, reducing the deposition of H3K27me3 and helping to prevent the establishment of a repressive domain (Beringer et al., 2016; Liefke et al., 2016). EPOP has also been indicated to have a role in regulating the deposition and removal of H3K4me3, although the mechanism for this has yet to be shown (Liefke et al., 2016). The presence of EPOP binding to bivalent nucleosomes *in vitro* supports a role of EPOP in the function of bivalent domains *in vivo* (Fig. 5.6 and Fig. 5.8). However, Elongin BC is not seen enriched bound to bivalent domains, this may be due to few PRC2-EPOP complexes interacting with Elongin BC or as a result of both ElonginB and C being present in multiple complexes including multiple ubiquitin ligase complexes (Okumura et al., 2012) and precluding it from being enriched significantly even if it does interact with PRC2-containing EPOP (PRC2/EPOP) in the pulldowns. The interaction between with PRC2/EPOP may be more transitory and therefore the stringency of the washes may remove it from bivalent domains before analysis.

The lack of recruitment of H3K4me3 binders, although not surprising in light of what was observed for asymmetric H3K4me0/3 nucleosomes, does pose questions about how the poised RNAPII seen at bivalent domains *in vivo* can be recruited and modified. RNAPII is seen at both active and bivalent domains, although it is more enriched at active domains (Ku et al., 2012; Stock et al., 2007; Mantsoki et al., 2018). Only TAF3 of the TFIID subunits is enriched bound to the H3K4me3 symmetrically modified nucleosomes *in vitro*, and it is not enriched bound to the bivalent nucleosome. The lack of TAF3 recruitment to bivalent nucleosomes correlates with the *in vivo* data demonstrating a decreased presence of TFIID and RNAPII at bivalent domains (Ku et al., 2012; Ferrai et al., 2017; Stock et al., 2007). Many subunits that compose TFIID are shared with other complexes that are not recruited directly to H3K4me3, therefore these subunits may not be enriched on any one modified nucleosome due to their presence in multiple complexes that are recruited to various differently modified nucleosomes (Koutelou et al., 2010; Grant et al., 1998). TFIID's composition and structure is dynamic indicating the possibility for different recruitment mechanisms depending on the promoter in question (Mousson et al., 2008;

Timmers and Sharp, 1991; Maston et al., 2012). All TAF dependent recruitment mechanisms may influence TFIID's recruitment, however, the strength of this influence may alter depending on the composition of the promoter and the TFIID complex in question. Although no TAF or TF was enriched at the bivalent condition, the recruitment of TAF3 to H3K4me3 is not required for RNAPII recruitment as RNAPII can be recruited solely by TBP's interaction with the TATA box (Majello et al., 1998; Mencía et al., 2002). There was a lack of general transcription factor (GTF) binding to the bivalent asymmetric nucleosome; however, the bivalent nucleosome investigated was of the simplest possible configuration and composition.

The lack of other transcription machinery recruited to the H3K4me3/3 nucleosome may indicate the requirement of multiple low affinity interactions required for the assembly of such a large complex which is not stable in the *in vitro* conditions used due to technical effects such as the number, length and stringency of washes in the pulldown experiments, or the requirement of an array of nucleosomes to allow the PIC to assemble and bind. The length of the washes between protein binding and elution may result in the loss of protein complexes with low overall Koff value potentially including the TFIID complex. The saturation of beads with nucleosomes may result in steric hindrance resulting in large protein complexes being unable to bind the nucleosome causing subunits that are not directly bound to the modification to be unable to be recruited to the nucleosome. Indeed increasing the amount of beads whilst keeping the concentration of H3K4me3/3 nucleosomes the same, resulted in an increase of TAF3 binding (data not shown) although these experiments were not analysed via MS and therefore the behaviour of the other TFIID subunits is unknown.

TFIID recruitment may require more than just TAF3 binding to recruit other transcription factors. It is possible that bivalent domains recruit TFIID and RNAPII due to many low affinity interactions, however, for this to be replicated *in vitro* it may require a bivalent nucleosomal array. Moreover, the additional chromatin marks present at bivalent domains *in vivo* may help with TFIID and RNAPII recruitment.

There are multiple additional PTMs that are found associated with bivalent domains, including H3K14ac, H3K9ac, and histone variants H2A.Z and H3.3 (Ku et al., 2012; Cui et al., 2009; Roh et al., 2006; Barski et al., 2007; Sen et al., 2016; Banaszynski et al., 2013). The presence of both H3K4me3 and H3K14ac has been shown to have a synergistic effect on the recruitment of TFIID (Vermeulen et al.,

2007). This effect is unaltered by the presence of H3K27me3 on the same tail as H3K4me3 (Nuland et al., 2013). H3K9ac helps to stabilise the presence of TFIID at promoters (Krasnov et al., 2016; Vermeulen et al., 2007). The combination of H3K4me3 and H3K14ac at bivalent domains may result in recruitment of TFIID and subsequently the PIC, while H3K9ac helps stabilise TFIID binding helping maintain poised RNAPII at bivalent domains. The binding of both H3K14ac and H3K9ac not only helps with the recruitment of RNAPII but also, via interaction with TAFII250 of TFIID, to stabilise it (Agalioti et al., 2002).

The relationship between H2A.Z and RNAPII is complicated; the C-terminus of yeast H2A.Z interacts with RNAPII, promoting its recruitment, however, there is an inverse relation between the presence of H2A.Z and transcription (Li et al., 2005a; Adam et al., 2001). This may be as a result of the chromatin remodeller FACT, and its subunit SPT16, which remove H2A.Z in coding regions (Jeronimo et al., 2015). The depletion of H2A.Z in coding regions correlates with the lack of H3K36me3 on H2A.Z containing nucleosomes (Jeronimo et al., 2015; Chen et al., 2012). A knockdown of H2A.Z results in a decrease of the RNAPII recruitment at TSSs in both human and yeast cells, suggesting a evolutionarily conserved function (Adam et al., 2001; Hardy et al., 2009). The presence of H2A.Z may therefore help with the recruitment and poising of RNAPII at bivalent promoters. In addition to H2A.Z's role in RNAPII recruitment to bivalent domains, it has been shown to allow access of multiple chromatin modifiers, including MLL and PcG complexes (Hu et al., 2013b; Weber et al., 2014; Guillemette et al., 2005; Li et al., 2005a; Creyghton et al., 2008). This access can be used to maintain, modify or communicate bivalent domain function. The presence of PRC1 and PRC2 as well as their marks, H3K27me3 and H2AK119ub, is inhibitory for productive transcription (Lau and So, 2015; Francis et al., 2004; Aranda et al., 2015; Zhou et al., 2008; Wang et al., 2004a; Napoles et al., 2004; Agger et al., 2007; Burgold et al., 2008; Nakagawa et al., 2008) although the mechanisms by which they enforce repression are still debated. These combinations of factors may result in small amounts of RNAPII recruitment and stabilisation at bivalent promoters but no transcriptional elongation.

The ability of H3K27me3 to recruit PcG complexes PRC1 and PRC2 in the context of the bivalent asymmetric nucleosome means that not only are the bivalent domains poised for activation via poised RNAPII but also for repression. After H3K4me3 removal, the bound PRC2, potentially in a conformation similar to that seen in Poepsel et al. (2018), would be able to modify the H3 tail with

H3K27me3. The presence of PRC2 and PRC1 may help reinforce the presence of H3K27me3 at bivalent domains, counteracting removal of H3K27me3 catalysed by JMJD3 and UTX (Agger et al., 2007; De Santa et al., 2007; Lan et al., 2007; Lee et al., 2006).

The presence of H3K27me3 and its ability to recruit PRC1 components even when in a bivalent conformation explains the co-occurrence of bivalency and H2A.Zub (Ku et al., 2012). Bivalent domains are enriched for H2A.Z which often contains both acetylation and ubiquitination marks. This ubiquitination may help to poise RNAPII at bivalent domains by decreasing FACT recruitment, inhibiting active RNAPII elongation and blocking the release of RNAPII from initiation (Stock et al., 2007; Zhou et al., 2008). H2A.Zub inhibits the recruitment of BRD2 (Surface et al., 2016), a protein enriched at bivalent domains that encourages RNAPII transcription (LeRoy et al., 2008). Removal of H2A.Zub decreases the recruitment of PRC2 and leads to depletion of H3K27me3 while increasing BRD2 recruitment and productive elongation (Surface et al., 2016). H2A.Z found at bivalent domains is usually dually modified with both ubiquitination and acetylation (Ku et al., 2012). The combination of these two marks at bivalent domains suggests another method of poising the gene for regulation with H2A.Z containing opposing marks. The presence of both H3K27me3 and H2A.Z are incompatible with DNA methylation (Zilberman et al., 2008; Hagarman et al., 2013; Meehan and Pennings, 2017). The presence of these at bivalent domains may prevent the establishment of DNA methylation and the formation of a stable repressive domain.

In order to investigate the importance of these additional chromatin signatures present at bivalent domains the pulldown experiment could be extended to nucleosomes modified with these additional modifications or histone variants H2A.Z and H3.3. These experiments could show how more complex chromatin modifications affect bivalent domain function and subsequent mark deposition. Does the presence of H3K14ac combined with H3K4me3 allow H3K4me0/3 or H3K4me3/H3K27me3 nucleosomes to recruit RNAPII? Is RNAPII at bivalent domains recruited or poised by H2A.Zub/ac? Does a nucleosome containing H3K4me3K9acK14ac poise RNAPII in the context of a bivalent nucleosome or a nucleosomal array? The use of nucleosomal arrays, additional modifications and specific DNA sequences may act to recruit further proteins involved in transcriptional regulation and recruit a network of proteins more similar to that bound to bivalent domains *in vivo*. The added PTMs, and histone variants

172

initially with mononucleosomes and then with nucleosomal arrays will increase the intricacy of the biological readout but, if done systematically, will allow deeper insight to the complexity of bivalent domains.

7.4 Bivalent Nucleosomes Recruit Specific Proteins *In Vitro*

The proteins that bind bivalent nucleosomes are not simply a combination of proteins that bind the singularly modified asymmetric nucleosomes or even an amalgamation of proteins that bind the symmetrically modified nucleosomes. This study has uncovered proteins that prefer to bind to the bivalent asymmetric nucleosome conformation, including PUM1, TFE1, SRCAP, KAT6B, SRSF2, DMAP1, USF1, and ZRANB2. The presence of these proteins may help maintain and communicate the biological function of bivalency to the cell. These proteins have a diverse range of functions, some of which oppose each other, indicating once again the co-occurrence of opposing functions, marks, such as H2A.Zub/ac, and proteins at bivalent domains. Due to time constraints only SRCAP and KAT6B were investigated further.

A number of components of SRCAP are enriched on the bivalent nucleosome in comparison to the unmodified nucleosome. However, the two unique components of the SRCAP complex were enriched on bivalent nucleosomes compared to all other nucleosomes investigated; SRCAP and DMAP1. The SRCAP complex is critical for the deposition of H2A.Z/H2B dimers, found at bivalent domains (Wong et al., 2007). The recruitment of these proteins *in vitro* along with the co-occurrence of their mark at bivalent domains supports a biological significance for the proteins found bound to the bivalent asymmetric nucleosomes *in vitro*. Multiple mechanisms for recruitment of SRCAP to chromatin have been proposed, including the recruitment of GAS41 (YEATS4) to H3K14ac and H3K27ac marks (Hsu et al., 2018a; Hsu et al., 2018b). The enrichment of SRCAP and DMAP1 on bivalent nucleosomes suggests the presence of another previously unknown recruitment mechanism – dependent on H3K4me3 and H3K27me3 co-localisation. This mechanism may be direct although currently there is no evidence to suggest direct binding via the protein domains present in the SRCAP complex. It is therefore more likely to be through indirect recruitment through another protein either contained in the candidate list or a protein present in

multiple complexes, preventing them from being classified as enriched on bivalent domains. The presence of H2A.Z at bivalency has been previously remarked upon and its function is debated as mentioned previously.

KAT6B, the catalytic component of the MORF complex, a MYST histone acetyltransferase (HAT) is also found enriched in the bivalent condition. KAT6B catalyses the acetylation of histone H3 resulting in the placement of H3K14ac and H3K23ac (Voigt et al., 2012; Karmodiya et al., 2012; Ullah et al., 2008; Voss et al., 2009; Sheikh and Akhtar, 2019; Simó-Riudalbas et al., 2015; Qin et al., 2011; Ali et al., 2012). H3K14ac contributes to recruitment of the SRCAP complex, indicating a link between these two complexes and their functions (Hsu et al., 2018a). KAT6B is unable to modify the nucleosomes used in the *in vitro* experiment due to removal of cofactors such as acetyl-CoA via dialysis in NE preparation. SRCAP and KAT6B may be recruited independently to bivalent domains or they may interact through the deposition of H3K14ac and by protein-protein interactions. The presence of both the catalytic components of MORF and SRCAP on bivalent nucleosomes suggests the co-operation of these two complexes in mediating bivalent domain function. The placement of H3K14ac may have additional functions apart from SRCAP recruitment. BRD2 binds to acetylated H4K5, H4K12 and H3K14 and can help promote transcription (LeRoy et al., 2008; Hargreaves et al., 2009; Cheung et al., 2017; Draker et al., 2012; Hnilicová et al., 2013). The recruitment of KAT6B may not only help H2A.Z deposition but also help stabilise poised RNAPII via the placement of H3K14ac and H3K23ac. The interactions and potential synergy between KAT6B and SRCAP at bivalent domains needs be investigated further possibly via double knockout cell lines of both proteins to see how this affects bivalent domains.

The binding of KAT6B to bivalent domains was confirmed *in vivo*. Nevertheless the exact mechanism of recruitment to bivalent asymmetric nucleosomes remains partially unclear. The interaction between PRC1 and MORF seen in *Drosophila* (Strubbe et al., 2011) is not present in mouse ESCs preventing PRC1's interaction with either H3K27me3 or H2AK119ub to mediate recruitment of KAT6B/MORF. The MORF complex does not contain any known reader domains for H3K27me3, although, it does contain reader domains that recruit MORF to both H3K4me0 and H3K4me3 (Ullah et al., 2008; Klein et al., 2014; Dreveny et al., 2014; Champagne et al., 2008). KAT6B could be recruited to bivalent nucleosomes via a previously unknown interaction or the double PHD finger of KAT6B may recruit MORF via its interaction with H3K4me0 on the tail containing H3K27me3.

174

In order to investigate KAT6B's binding and function *in vivo*, cell lines with tagged-KAT6B were generated. KAT6B is present at both active and bivalent domains as suggested by previous studies that investigated H3K14ac placement (Karmodiya et al., 2012; Guenther et al., 2007; Hsu et al., 2018b). The removal of H3K4me3 by the inactivation of MLL2 decreased the recruitment of KAT6B mainly at bivalent domains. While the deletion of both EZH1 and EZH2 removed H3K27me3 genome-wide it did not decrease KAT6B recruitment at bivalent domains. KAT6B is still present at bivalent domains at lower levels when H3K4me3 is removed, indicating that KAT6B recruitment is only partially dependent on the presence of H3K4me3 and bivalency suggesting another method for KAT6B recruitment. The deposition of its own mark also helps further recruitment of the KAT6A/B complex (Ali et al., 2012; Klein et al., 2017) resulting in a feedback mechanism and may be partially responsible for the low levels of KAT6B remaining. H3K27me3 present at bivalent domains may not act to recruit KAT6B *in vivo* by any known mechanism.

Further research is required in order to fully isolate and determine the mechanism of KATB6's recruitment to bivalent domains. A sequence of protein truncations combined with mutations of important residues within protein domains will allow determination of which protein subunit or combination of subunits recruits KAT6B to bivalent domains. This may also elucidate the importance of H3K27me3 for recruitment, stability, or catalytic activity of KAT6B. Mutation of the ING5 PHD domain would allow confirmation of its role in recruiting KAT6B to both active and bivalent domains, while mutational analysis of KAT6B's double PHD domain would help isolate the domain responsible for MORF's recruitment to bivalent domains. It is possible that the bivalent nucleosomes could "trap" KAT6B in a specific conformation, potentially at a stage of the enzymatic reaction that prevents the transfer of the acetyl group. This would increase the amount of KAT6B bound while explaining the lack of catalytic activity seen *in vivo* and *in vitro*.

A different role for H3K27me3 and its importance for KAT6B's function was suggested by a HAT assay performed with the specifically modified nucleosomes. The purified MORF complex showed increased catalytic activity towards H3K27me3 symmetrically modified nucleosomes. The difference between the presence of KAT6B and its marks *in vivo*, analysed via ChIP, and the activity of the purified MORF complex towards H3K27me3/3 nucleosomes in the HAT assay suggests a further interaction or factor missing from the purified MORF complex. This

interaction may affect the activity of KAT6B at repressive domains *in vivo*.

The MYST family of acetyltransferases is poorly researched and so far the structure and catalytic mechanism of only one of the MYST family members has been solved – ESA1 (Yan et al., 2000). There are two different catalytic mechanisms of KAT6B proposed, but neither has been confirmed. This uncertainty of MYST binding and catalytic mechanism means it is difficult to predict how H3K27me3 may cause this increase in catalytic activity. The double PHD domain on MOZ twists an 11 residue stretch of the H3 N-terminal tail into an α helical structure, allowing summation of H3K4me3 and H3K14ac effects on the MYST domain (Dreveny et al., 2014). This demonstrates a mechanism for modified residues more than 11 $\alpha\alpha$ distant to the target residue to affect KAT6B's function. Currently no structure of a mammalian MYST domain bound to a defined peptide substrate has been solved, therefore the expanse of peptide that interacts with the domain and influences MYST binding and catalytic activity is unknown. The binding of a peptide to an enzyme is not solely due to the target residue, interactions with the residues alongside the substrate residue contribute to specificity of substrate-enzyme binding. Due to the proximity of residue H3K27 with H3K23, H3K27me3 may interact with MORF's active site, the MYST domain, when depositing the H3K23ac modification. It may be this interaction with the MYST domain that causes the increase in catalytic activity observed in the HAT assay. This interaction of H3K27me3 with the MYST domain may have multiple functions, not only increasing KAT6B'S catalytic activity but also stabilising the binding of the MORF complex at bivalent domains (see Fig. 7.1). H3K27me3's effect on MORF requires further investigation to elucidate its mechanism and biological function. Crystallisation of KAT6B or KAT6B's MYST domain bound to a peptide would allow determination of H3K27me3's role in MYST activity and binding. Once catalytic residues are determined, mutational analyses combined with *in vitro* catalytic assays would show the importance of each residue, elucidate KAT6B's catalytic mechanism and indicate the mechanism of H3K27me3 stimulation. Crosslinking mass spectrometry could be used to determine interactions between a bivalent nucleosome and the MORF complex. This would help identify the potential role H3K27me3 has in recruitment or stimulation of KAT6B's catalytic activity.

Generation and neuronal differentiation of KAT6B KO mouse ESCs cell lines allowed investigation into KAT6B's role in ESCs and at bivalent domains. There are no obvious phenotypic or gene expression differences between E14 control and

176
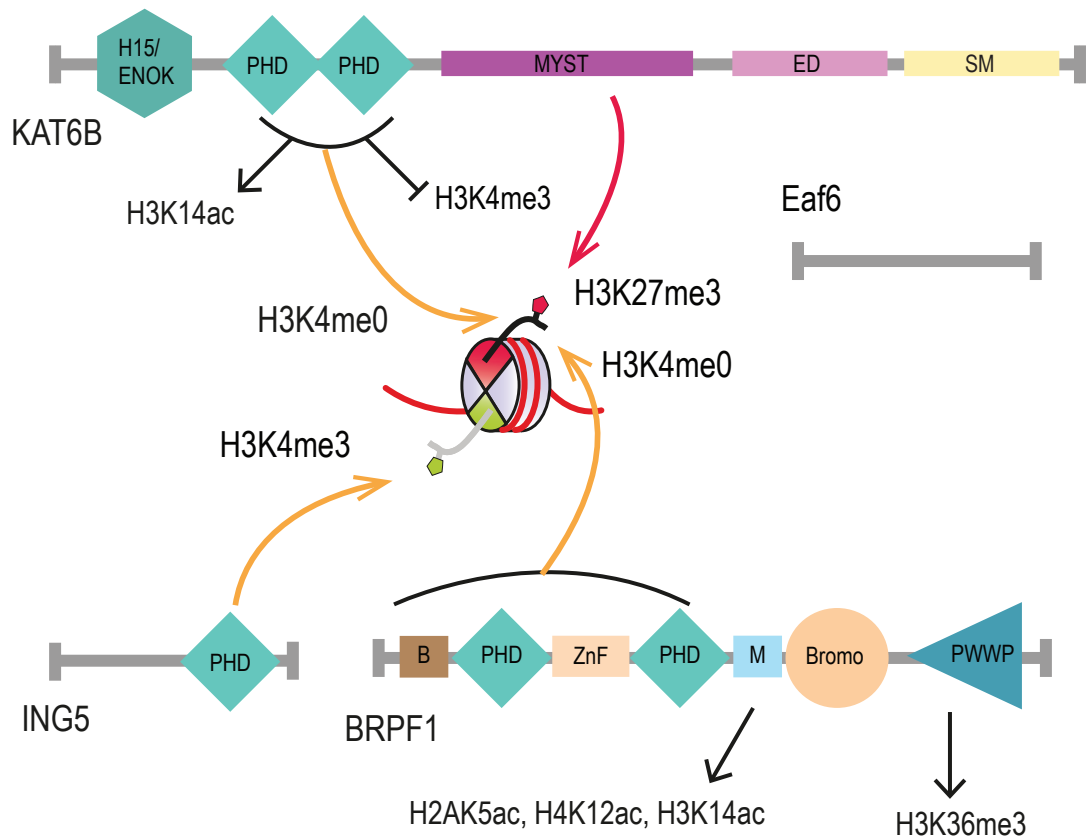
Figure 7.1 *Potential Interactions of the MORF Complex with Histone Modifications.* *Domains are labelled as follows: PHD - Plant homeodomain zinc finger, ZnF - Zinc finger, H15 - Sequence similarity to H1 and H5 histone domain, also referred to as NEMM, MYST - Acetyltransferase domain, ED - acidic (glutamate/aspartate rich) region, SM - Serine/methionine rich stretch, Bromo - bromodomain, PWWP - proline-tryptophan-tryptophan-proline domain, B - domain that interacts with KAT6B's MYST domain, M - domain that mediates BRPF1's interaction with ING5 and hEAF6. The potential interaction of KAT6B's MYST domain with H3K27me3 shown with a red arrow.*

KAT6A/B KO cell lines. This suggests that removal of KAT6B does not affect the expression of actively transcribed genes, even though previously obtained data demonstrates its presence at these genes. The differences between the cell lines and the importance of KAT6B becomes clearer as the cells start to differentiate. From day 2 onwards KAT6B KO cell lines fail to properly upregulate genes required for neuronal differentiation and for exit from pluripotency. This phenotype is replicated, although to a lesser extent, in KAT6A KO cell lines and amplified in the double KO. The importance of KAT6B and the marks it places to upregulate genes in differentiation is supported by the failure of KAT6B KO, KAT6A KO and KAT6A/B KO cell lines to form neuronal cells. The removal of KAT6B does not affect the expression of genes involved in pluripotency expressed in ESCs; however, it does affect the upregulation of genes that are required for differentiation, including genes that are bivalent in ESCs. Whether this result is representative for ESC differentiation in general or specific to neuronal linage specific genes is unknown. Further differentiation experiments are required to see whether KAT6B is essential for successful ESCs differentiation into other lineages as well.

7.5   Summary

This work has provided important advances into understanding the unique network of proteins that bind bivalent asymmetric nucleosomes *in vitro*. Further experiments are required to elucidate how the addition of further marks and substitution of histone variants found at bivalent domains affects this network. While investigating bivalency and histone PTM asymmetry, the inability of both H3K27me3 and H3K4me3 marks to enrich for proteins when present in asymmetric rather than symmetric fashion was discovered. This asymmetric effect may have important consequences *in vivo* and should be investigated further. Identification of sites consisting of asymmetric H3K4me3, potentially by construction of a recombinant protein that is able to bind this asymmetric state, would allow determination of their biological significance. The proteins identified bound to bivalent nucleosomes *in vitro* support the co-occurrence of chromatin modifications previously found at bivalent domains *in vivo*. The co-occurrence of H3K4me3, and H3K27me3 at bivalent domains seem to recruit proteins that involved in both activation and repression of transcription and reinforce the bivalent nature of the domain.

This combination of proteins and marks found occurring at bivalent domains, supports the theory of bivalent domains poising the genes for timely regulation during differentiation. One member of the network of proteins that prefers to bind the bivalent asymmetric nucleosome conformation is KAT6B. KAT6B binds bivalent nucleosomes both *in vitro* and *in vivo*, and is important for embryonic stem cell differentiation, although the mechanism of recruitment requires further investigation.

Bibliography

Abbott, D. W., V. S. Ivanova, X. Wang, W. M. Bonner, and J. Ausió (2001). "Characterization of the stability and folding of H2A.Z chromatin particles: Implications for transcriptional activation". In: *Journal of Biological Chemistry* 276.45, pp. 41945–41949.

Abed, J. A. and R. S. Jones (2012). "H3K36me3 key to Polycomb-mediated gene silencing in lineage specification". In: *Nature Structural & Molecular Biology* 19.12, pp. 1214–1215.

Adam, M., F. Robert, M. Larochelle, and L. Gaudreau (2001). "H2A.Z Is Required for Global Chromatin Integrity and for Recruitment of RNA Polymerase II under Specific Conditions". In: *Molecular and Cellular Biology* 21.18, pp. 6270–6279.

Agalioti, T., G. Chen, and D. Thanos (2002). "Deciphering the Transcriptional Histone Acetylation Code for a Human Gene positioning. Rather, it has been proposed that distinct patterns of histone modification function as a recogni-tion code (histone code) for the recruitment of chromatin". In: *Cell* 111, pp. 381–392.

Agger, K., P. A. Cloos, J. Christensen, D. Pasini, S. Rose, J. Rappsilber, I. Issaeva, E. Canaani, A. E. Salcini, and K. Helin (2007). "UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development". In: *Nature* 449.7163, pp. 731–734.

Akkers, R. C., S. J. van Heeringen, U. G. Jacobi, E. M. Janssen-Megens, K. J. Françoijs, H. G. Stunnenberg, and G. J. C. Veenstra (2009). "A Hierarchy of H3K4me3 and H3K27me3 Acquisition in Spatial Gene Regulation in Xenopus Embryos". In: *Developmental Cell* 17.3, pp. 425–434.

Akmammedov, A., M. Geigges, and R. Paro (2019). "Bivalency in Drosophila embryos is associated with strong inducibility of Polycomb target genes". In: *Fly* 0.0, pp. 1–9.

Alder, O., F. Lavial, A. Helness, E. Brookes, S. Pinho, A. Chandrashekran, P. Arnaud, A. Pombo, L. O'Neill, and V. Azuara (2010). "Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment". In: *Development* 137.15, pp. 2483–2492.

Alekseyenko, A. A., A. A. Gorchakov, P. V. Kharchenko, and M. I. Kuroda (2014). "Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.7, pp. 2488–2493.

Ali, A. and S. Tyagi (2017). "Diverse roles of WDR5-RbBP5-ASH2L-DPY30 (WRAD) complex in the functions of the SET1 histone methyltransferase family". In: *Journal of Biosciences* 42.1, pp. 155–159.

Ali, M., R. A. Hom, W. Blakeslee, L. Ikenouye, and T. G. Kutateladze (2014). "Diverse functions of PHD fingers of the MLL/KMT2 subfamily". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1843.2, pp. 366–371.

Ali, M., K. Yan, M.-E. Lalonde, C. Degerny, S. B. Rothbart, B. D. Strahl, J. Côté, X.-J. Yang, and T. G. Kutateladze (2012). "Tandem PHD Fingers of MORF/MOZ Acetyltransferases Display Selectivity for Acetylated Histone H3 and Are Required for the Association with Chromatin". In: *Journal of Molecular Biology* 424.5, pp. 328–338.

Allan, J., N. Harborne, D. C. Rau, and H. Gould (1982). "Participation of core histone "tails" in the stabilization of the chromatin solenoid". In: *Journal of Cell Biology* 93.2, pp. 285–297.

Aloia, L., B. Di Stefano, and L. Di Croce (2013). "Polycomb complexes in stem cells and embryonic development". In: *Development* 140.12, pp. 2525–2534.

Ang, Y.-S. et al. (2011). "Wdr5 Mediates Self-Renewal and Reprogramming via the Embryonic Stem Cell Core Transcriptional Network". In: *Cell* 145.2, pp. 183–197.

Antonysamy, S. et al. (2013). "Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain." In: *PloS one* 8.12, e84147.

Araki, Y. et al. (2010). "Genome-wide analysis of histone methylation reveals chromatin state-based complex regulation of differential gene transcription and function of CD8 memory T cells Yasuto". In: *Immunity* 30.6, pp. 912–925.

Aranda, S., G. Mas, and L. Di Croce (2015). "Regulation of gene transcription by Polycomb proteins". In: *Science Advances* 1.11, pp. 1–16.

Azuara, V. et al. (2006). "Chromatin signatures of pluripotent cell lines." In: *Nature cell biology* 8.5, pp. 532–8.

Baba, A. et al. (2011). "PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B". In: *Nature Cell Biology* 13.6, pp. 668–675.

Bach, C., D. Mueller, S. Buhl, M. P. Garcia-Cuellar, and R. K. Slany (2009). "Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2". In: *Oncogene* 28.6, pp. 815–823.

Baell, J. B. et al. (2018). "Inhibitors of histone acetyltransferases KAT6A/B induce senescence and arrest tumour growth". In: *Nature* 560.7717, pp. 253–257.

Bajusz, I., G. Kovács, and M. Pirity (2018). "From Flies to Mice: The Emerging Role of Non-Canonical PRC1 Members in Mammalian Development". In: *Epigenomes* 2.1, p. 4.

Ballaré, C. et al. (2012). "Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity". In: *Nature Structural and Molecular Biology* 19.12, pp. 1257–1265.

Banaszynski, L. A. et al. (2013). "Hira-Dependent Histone H3.3 Deposition Facilitates PRC2 Recruitment at Developmental Loci in ES Cells". In: *Cell* 155.1, pp. 107–120.

Bannister, A. J. and T. Kouzarides (2011). "Regulation of chromatin by histone modifications." In: *Cell research* 21.3, pp. 381–95.

Bantscheff, M., M. Schirle, G. Sweetman, J. Rick, and B. Kuster (2007). "Quantitative mass spectrometry in proteomics: A critical review". In: *Analytical and Bioanalytical Chemistry* 389.4, pp. 1017–1031.

Bao, Y. and X. Shen (2007). "INO80 subfamily of chromatin remodeling complexes". In: *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 618.1-2, pp. 18–29.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao (2007). "High-Resolution Profiling of Histone Methylations in the Human Genome". In: *Cell* 129.4, pp. 823–837.

Barski, A., R. Jothi, S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones, and K. Zhao (2009). "Chromatin poises miRNA- and protein-coding genes for expression". In: *Genome Research* 19.10, pp. 1742–1751.

Bartke, T., M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann, and T. Kouzarides (2010). "Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation". In: *Cell* 143.3, pp. 470–484.

Bauer, M., J. Trupke, and L. Ringrose (2016). "The quest for mammalian Polycomb response elements: are we there yet?" In: *Chromosoma* 125.3, pp. 471–496.

Beck, D. B., H. Oda, S. S. Shen, and D. Reinberg (2012). "PR-set7 and H4K20me1: At the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription". In: *Genes and Development* 26.4, pp. 325–337.

Becker, P. B. (2002). "Nucleosome sliding: Facts and fiction". In: *EMBO Journal* 21.18, pp. 4749–4753.

Beltran, M. et al. (2016). "The interaction of PRC2 with RNA or chromatin is mutually antagonistic". In: *Genome Research* 26.7, pp. 896–907.

Beringer, M. et al. (2016). "EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells". In: *Molecular Cell* 64.4, pp. 645–658.

Bernstein, B. E. et al. (2006a). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." In: *Cell* 125.2, pp. 315–26.

Bernstein, E., E. M. Duncan, O. Masui, J. Gil, E. Heard, and C. D. Allis (2006b). "Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin". In: *Molecular and Cellular Biology* 26.7, pp. 2560–2569.

Betschinger, J., J. Nichols, S. Dietmann, P. D. Corrin, P. J. Paddison, and A. Smith (2013). "Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3". In: *Cell* 153.2, pp. 335–347.

Bibel, M., J. Richter, E. Lacroix, and Y.-A. Barde (2007). "Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells." In: *Nature protocols* 2.5, pp. 1034–1043.

Bird, A. (2002). "DNA methylation patterns and epigenetic memory". In: *Genes & Development* 16.1, pp. 6–21.

Birke, M., S. Schreiner, M.-P. García-Cuéllar, K. Mahr, F. Titgemeyer, and R. Slan (2002). "The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation". In: *Nucleic Acids Research* 30.4, pp. 958–965.

Biswas, M., K. Voltz, J. C. Smith, and J. Langowski (2011). "Role of histone tails in structural stability of the nucleosome". In: *PLoS Computational Biology* 7.12, pp. 2–13.

Blackledge, N. P. et al. (2014). "Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation." In: *Cell* 157.6, pp. 1445–59.

Blackledge, N. P., N. R. Rose, and R. J. Klose (2015). "Targeting Polycomb systems to regulate gene expression: modifications to a complex story". In: *Nature Reviews Molecular Cell Biology* 16.11, pp. 643–649.

Blanc, R. S. and S. Richard (2017). "Arginine Methylation: The Coming of Age". In: *Molecular Cell* 65.1, pp. 8–24.

Bledau, A. S. et al. (2014). "The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation". In: *Development* 141.5, pp. 1022–1035.

Bochyńska, A., J. Lüscher-Firzlaff, and B. Lüscher (2018). "Modes of Interaction of KMT2 Histone H3 Lysine 4 Methyltransferase/COMPASS Complexes with Chromatin". In: *Cells* 7.3, p. 17.

Bönisch, C. and S. B. Hake (2012). "Histone H2A variants in nucleosomes and chromatin: More or less stable?" In: *Nucleic Acids Research* 40.21, pp. 10719–10741.

Boulay, G., C. Rosnoblet, C. Guérardel, P. O. Angrand, and D. Leprince (2011). "Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes". In: *Biochemical Journal* 434.2, pp. 333–342.

Boyer, L. A. et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells". In: *Nature* 441.7091, pp. 349–353.

Bradsher, J. N., S. Tan, H. J. McLaury, J. W. Conaway, and R. C. Conaway (1993). "RNA polymerase II transcription factor SIII. II. Functional properties and role in RNA chain elongation". In: *Journal of Biological Chemistry* 268.34, pp. 25594–25603.

Brahma, S., M. I. Udugama, J. Kim, A. Hada, S. K. Bhardwaj, S. G. Hailu, T.-H. Lee, and B. Bartholomew (2017). "INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers". In: *Nature Communications* 8.1, p. 15616.

Brehove, M., T. Wang, J. North, Y. Luo, S. J. Dreher, J. C. Shimko, J. J. Ottesen, K. Luger, and M. G. Poirier (2015). "Histone Core Phosphorylation Regulates DNA Accessibility". In: *Journal of Biological Chemistry* 290.37, pp. 22612–22621.

Brien, G. L. et al. (2012). "Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation". In: *Nat Struct Mol Biol* 19.12, pp. 1273–1281.

Brinkman, A. B. et al. (2012). "Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk". In: *Genome Research* 22.6, pp. 1128–1138.

Brookes, E. et al. (2012). "Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs". In: *Cell Stem Cell* 10.2, pp. 157–170.

Brown, D. A. et al. (2017). "The SET1 Complex Selects Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin". In: *Cell Reports* 20.10, pp. 2313–2327.

Bunkenborg, J., N. I. Gadjev, T. Deligeorgiev, and J. P. Jacobsen (2000). "Concerted intercalation and minor groove recognition of DNA by a homodimeric thiazole orange dye". In: *Bioconjugate Chemistry* 11.6, pp. 861–867.

Burgold, T., F. Spreafico, F. De Santa, M. G. Totaro, E. Prosperini, G. Natoli, and G. Testa (2008). "The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment". In: *PLoS ONE* 3.8.

Cai, L. et al. (2013). "An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting". In: *Molecular Cell* 49.3, pp. 571–582.

Cai, Y., J. Jin, L. Florens, S. K. Swanson, T. Kusch, B. Li, J. L. Workman, M. P. Washburn, R. C. Conaway, and J. W. Conaway (2005). "The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes". In: *Journal of Biological Chemistry* 280.14, pp. 13665–13670.

Cai, Y., J. Jin, C. Tomomori-Sato, S. Sato, I. Sorokina, T. J. Parmely, R. C. Conaway, and J. W. Conaway (2003). "Identification of New Subunits of the Multiprotein Mammalian TRRAP/TIP60-containing Histone Acetyltransferase Complex". In: *Journal of Biological Chemistry* 278.44, pp. 42733–42736.

184

Campos, E. I. and D. Reinberg (2009). "Histones: Annotating Chromatin". In: *Annual Review of Genetics* 43.1, pp. 559–599.

Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang (2002). "Role of histone H3 lysine 27 methylation in polycomb-group silencing". In: *Science* 298.5595, pp. 1039–1043.

Cao, R. and Y. Zhang (2004a). "SUZ12 Is Required for Both the Histone Methyl-transferase Activity and the Silencing Function of the EED-EZH2 Complex". In: *Molecular Cell* 15.1, pp. 57–67.

— (2004b). "The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3". In: *Current Opinion in Genetics & Development* 14.2, pp. 155–164.

Cardamone, M. D., B. Tanasa, C. T. Cederquist, J. Huang, K. Mahdaviani, W. Li, M. G. Rosenfeld, M. Liesa, and V. Perissi (2018). "Mitochondrial Retrograde Signaling in Mammals Is Mediated by the Transcriptional Cofactor GPS2 via Direct Mitochondria-to-Nucleus Translocation". In: *Molecular Cell* 69.5, 757–772.e7.

Casanova, M. et al. (2011). "Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells". In: *Development* 138.8, pp. 1471–1482.

Chakravarthy, S., S. Kumar, Y. Gundimella, C. Caron, P.-Y. Perche, J. R. Pehrson, S. Khochbin, and K. Luger (2005). "Structural Characterization of the Histone Variant macroH2A". In: *Molecular and Cellular Biology* 25.17, pp. 7616–7624.

Chamberlain, S. J., D. Yee, and T. Magnuson (2008). "Polycomb Repressive Complex 2 Is Dispensable for Maintenance of Embryonic Stem Cell Pluripotency". In: *Stem Cells* 26.6, pp. 1496–1505.

Champagne, K. S., N. Saksouk, P. V. Peña, K. Johnson, M. Ullah, X.-j. Yang, J. Côté, and T. G. Kutateladze (2008). "The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide". In: *Proteins: Structure, Function, and Bioinformatics* 72.4, pp. 1371–1376.

Champagne, K. and T. Kutateladze (2009). "Structural Insight Into Histone Recognition by the ING PHD Fingers". In: *Current Drug Targets* 10.5, pp. 432–441.

Champagne, N., N. R. Bertos, N. Pelletier, A. H. Wang, M. Vezmar, Y. Yang, H. H. Heng, and X. J. Yang (1999). "Identification of a human histone acetyltransferase related to monocytic leukemia zinc finger protein". In: *Journal of Biological Chemistry* 274.40, pp. 28528–28536.

Chan, C., L. Rastelli, and V. Pirrotta (1994). "A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression." In: *The EMBO Journal* 13.11, pp. 2553–2564.

Chandru, A., N. Bate, G. W. Vuister, and S. M. Cowley (2018). "Sin3A recruits Tet1 to the PAH1 domain via a highly conserved Sin3-Interaction Domain". In: *Scientific Reports* 8.1, pp. 1–10.

Chang, P.-Y., R. A. Hom, C. A. Musselman, L. Zhu, A. Kuo, O. Gozani, T. G. Kutateladze, and M. L. Cleary (2010). "Binding of the MLL PHD3 Finger to Histone H3K4me3 Is Required for MLL-Dependent Gene Transcription". In: *Journal of Molecular Biology* 400.2, pp. 137–144.

Chatterjee, C. and T. W. Muir (2010). "Chemical approaches for studying histone modifications". In: *Journal of Biological Chemistry* 285.15, pp. 11045–11050.

Chelius, D. and P. V. Bondarenko (2002). "Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry". In: *Journal of Proteome Research* 1.4, pp. 317–323.

Chen, J., A. Miller, A. L. Kirchmaier, and J. M. K. Irudayaraj (2012). "Single-molecule tools elucidate H2A.Z nucleosome composition". In: *Journal of Cell Science* 125.12, pp. 2954–2964.

Chen, J., Y. Ai, J. Wang, L. Haracska, and Z. Zhuang (2010). "Chemically ubiquitylated PCNA as a probe for eukaryotic translesion DNA synthesis". In: *Nature Chemical Biology* 6.4, pp. 270–272.

Chen, S., L. Jiao, M. Shubbar, X. Yang, and X. Liu (2018). "Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding". In: *Molecular Cell* 69.5, 840–852.e5.

Chen, Z., A. T. Grzybowski, and A. J. Ruthenburg (2014). "Traceless Semisynthesis of a Set of Histone 3 Species Bearing Specific Lysine Methylation Marks". In: *ChemBioChem* 15.14, pp. 2071–2075.

Cheng, X. (2014). "Structural and Functional Coordination of DNA and Histone Methylation". In: *Cold Spring Harbor Perspectives in Biology* 6.8, a018747–a018747.

Cheng, X., R. E. Collins, and X. Zhang (2005). "Structural and Sequence Motifs of Protein (Histone) Methylation Enzymes". In: *Annual Review of Biophysics and Biomolecular Structure* 34.1, pp. 267–294.

Cheung, K. L. et al. (2017). "Distinct Roles of Brd2 and Brd4 in Potentiating the Transcriptional Program for Th17 Cell Differentiation". In: *Molecular Cell* 65.6, 1068–1080.e5.

Chin, H. G., M. Pradhan, P. O. Estève, D. Patnaik, T. C. Evans, and S. Pradhan (2005). "Sequence specificity and role of proximal amino acids of the histone H3 tail on catalysis of murine G9a lysine 9 histone H3 methyltransferase". In: *Biochemistry* 44.39, pp. 12998–13006.

Choudhury, R., S. Singh, S. Arumugam, A. Roguev, and A. F. Stewart (2019). "The Set1 complex is dimeric and acts with Jhd2 demethylation to convey symmetrical H3K4 trimethylation". In: *Genes & development* 33.9-10, pp. 550–564.

Cifuentes-Rojas, C., A. J. Hernandez, K. Sarma, and J. T. Lee (2014). "Regulatory Interactions between RNA and Polycomb Repressive Complex 2". In: *Molecular Cell* 55.2, pp. 171–185.

Clouaire, T., S. Webb, P. Skene, R. Illingworth, A. Kerr, R. Andrews, J.-H. Lee, D. Skalnik, and A. Bird (2012). "Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells". In: *Genes & Development* 26.15, pp. 1714–1728.

Coffman, J. (2003). "Runx transcription factors and the developmental balance between cell proliferation and differentiation". In: *Cell Biology International* 27.4, pp. 315–324.

Collins, R. E., J. P. Northrop, J. R. Horton, D. Y. Lee, X. Zhang, M. R. Stallcup, and X. Cheng (2008). "The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules". In: *Nature Structural & Molecular Biology* 15.3, pp. 245–250.

Collinson, A., A. J. Collier, N. P. Morgan, A. R. Sienerth, T. Chandra, S. Andrews, and P. J. Rugg-Gunn (2016). "Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells". In: *Cell Reports* 17.10, pp. 2700–2714.

Conaway, R., D. Reines, K. P. Garrett, W. a. D. E. Powell, and J. Conaway (1996). "Purification of RNA Polymerase II General Transcription Factors from Rat Liver". In: *Methods in enzymology* 273.1993, pp. 194–207.

Cong, L. et al. (2013). "Multiplex Genome Engineering Using CRISPR/Cas Systems". In: *Science* 339.6121, pp. 819–823.

Congdon, L. M., S. I. Houston, C. S. Veerappan, T. M. Spektor, and J. C. Rice (2010). "PR-Set7-mediated monomethylation of histone H4 lysine 20 at specific genomic regions induces transcriptional repression". In: *Journal of Cellular Biochemistry* 110.3, pp. 609–619.

Conway, E. et al. (2018). "A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities". In: *Molecular Cell* 70.3, 408–421.e8.

Cooper, S. et al. (2014). "Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment". In: *Cell Reports* 7.5, pp. 1456–1470.

Cooper, S. et al. (2016). "Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2". In: *Nature Communications* 7.1, p. 13661.

Corpet, A. et al. (2011). "Asf1b, the necessary Asf1 isoform for proliferation, is predictive of outcome in breast cancer". In: *EMBO Journal* 30.3, pp. 480–493.

Cosentino, M. S. et al. (2019). "Kat6b Modulates Oct4 and Nanog Binding to Chromatin in Embryonic Stem Cells and Is Required for Efficient Neural Differentiation". In: *Journal of Molecular Biology* 431.6, pp. 1148–1159.

Cox, J., M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann (2014). "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ." In: *Molecular & cellular proteomics : MCP* 13.9, pp. 2513–26.

Cox, J. and M. Mann (2008). "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification". In: *Nature Biotechnology* 26.12, pp. 1367–1372.

Cox, J., N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann (2011). "Andromeda: A peptide search engine integrated into the MaxQuant environment". In: *Journal of Proteome Research* 10.4, pp. 1794–1805.

Creppe, C., A. Palau, R. Malinverni, V. Valero, and M. Buschbeck (2014). "A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation". In: *PLoS Genetics* 10.12. Ed. by R. Eskeland, e1004851.

Creyghton, M. P., S. Markoulaki, S. S. Levine, J. Hanna, M. A. Lodato, K. Sha, R. A. Young, R. Jaenisch, and L. A. Boyer (2008). "H2AZ Is Enriched at Polycomb Complex Target Genes in ES Cells and Is Necessary for Lineage Commitment". In: *Cell* 135.4, pp. 649–661.

Cruz-Molina, S. et al. (2017). "PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation". In: *Cell Stem Cell* 20.5, 689–705.e9.

Cui, K., C. Zang, T.-Y. Roh, D. E. Schones, R. W. Childs, W. Peng, and K. Zhao (2009). "Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation". In: *Cell Stem Cell* 4.1, pp. 80–93.

Curradi, M., A. Izzo, G. Badaracco, and N. Landsberger (2002). "Molecular Mechanisms of Gene Silencing Mediated by DNA Methylation". In: *Molecular and Cellular Biology* 22.9, pp. 3157–3173.

Dahl, J. A., A. H. Reiner, A. Klungland, T. Wakayama, and P. Collas (2010). "Histone H3 Lysine 27 Methylation Asymmetry on Developmentally-Regulated Promoters

Distinguish the First Two Lineages in Mouse Preimplantation Embryos". In: *PLoS ONE* 5.2. Ed. by R. Feil, e9150.

Dao, P., D. Wojtowicz, S. Nelson, D. Levens, and T. M. Przytycka (2016). "Ups and Downs of Poised RNA Polymerase II in B-Cells". In: *PLoS Computational Biology* 12.4, pp. 1–19.

Davey, C. A., D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond (2002). "Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution". In: *Journal of Molecular Biology* 319.5, pp. 1097–1113.

Davidovich, C., L. Zheng, K. J. Goodrich, and T. R. Cech (2013). "Promiscuous RNA binding by Polycomb repressive complex 2". In: *Nature Structural & Molecular Biology* 20.11, pp. 1250–1257.

Dawson, P., T. Muir, I. Clark-Lewis, and S. Kent (1994). "Synthesis of proteins by native chemical ligation". In: *Science* 266.5186, pp. 776–779.

Dawson, P. E. (1997). "Synthesis of chemokines by native chemical ligation". In: *Methods in enzymology*. Vol. 287. September. Academic Press Inc., pp. 34–45.

Dawson, P. E. and S. B. H. Kent (2000). "Synthesis of Native Proteins by Chemical Ligation". In: *Annual Review of Biochemistry* 69.1, pp. 923–960.

De Gobbi, M. et al. (2011). "Generation of bivalent chromatin domains during cell fate decisions". In: *Epigenetics & Chromatin* 4.1, p. 9.

De Santa, F., M. G. Totaro, E. Prosperini, S. Notarbartolo, G. Testa, and G. Natoli (2007). "The Histone H3 Lysine-27 Demethylase Jmjd3 Links Inflammation to Inhibition of Polycomb-Mediated Gene Silencing". In: *Cell* 130.6, pp. 1083–1094.

Dellino, G. I., Y. B. Schwartz, G. Farkas, D. McCabe, S. C. Elgin, and V. Pirrotta (2004). "Polycomb Silencing Blocks Transcription Initiation". In: *Molecular Cell* 13.6, pp. 887–893.

Demers, C., C.-P. Chaturvedi, J. A. Ranish, G. Juban, P. Lai, F. Morle, R. Aebersold, F. J. Dilworth, M. Groudine, and M. Brand (2007). "Activator-Mediated Recruitment of the MLL2 Methyltransferase Complex to the β-Globin Locus". In: *Molecular Cell* 27.4, pp. 573–584.

Deng, C. et al. (2013). "USF1 and hSET1A Mediated Epigenetic Modifications Regulate Lineage Differentiation and HoxB4 Transcription". In: *PLoS Genetics* 9.6. Ed. by S. M. Weissman, e1003524.

Denissov, S., H. Hofemeister, H. Marks, A. Kranz, G. Ciotta, S. Singh, K. Anastassiadis, H. G. Stunnenberg, and A. F. Stewart (2014). "Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant". In: *Development* 141.3, pp. 526–537.

Dey, A., F. Chitsaz, A. Abbasi, T. Misteli, and K. Ozato (2003). "The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.15, pp. 8758–8763.

Dillon, S. C., X. Zhang, R. C. Trievel, and X. Cheng (2005). "The SET-domain protein superfamily: Protein lysine methyltransferases". In: *Genome Biology* 6.8.

Doyon, Y., C. Cayrou, M. Ullah, A.-J. Landry, V. Côté, W. Selleck, W. S. Lane, S. Tan, X.-J. Yang, and J. Côté (2006). "ING Tumor Suppressor Proteins Are Critical Regulators of Chromatin Acetylation Required for Genome Expression and Perpetuation". In: *Molecular Cell* 21.1, pp. 51–64.

Doyon, Y., W. Selleck, W. S. Lane, S. Tan, and J. Côté (2004). "Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans." In: *Molecular and cellular biology* 24.5, pp. 1884–96.

Draker, R., M. K. Ng, E. Sarcinella, V. Ignatchenko, T. Kislinger, and P. Cheung (2012). "A Combination of H2A.Z and H4 Acetylation Recruits Brd2 to Chromatin during Transcriptional Activation". In: *PLoS Genetics* 8.11. Ed. by J. Workman, e1003047.

Dreveny, I., S. E. Deeves, J. Fulton, B. Yue, M. Messmer, A. Bhattacharya, H. M. Collins, and D. M. Heery (2014). "The double PHD finger domain of MOZ/MYST3 induces α-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification". In: *Nucleic Acids Research* 42.2, pp. 822–835.

Du, J., B. Kirk, J. Zeng, J. Ma, and Q. Wang (2017). "Three classes of response elements for human PRC2 and MLL1/2-trithorax complexes". In: *SSRN Electronic Journal* 46.July, pp. 1–17.

Dyer, P. N., R. S. Edayathumangalam, C. L. White, Y. Bao, S. Chakravarthy, U. M. Muthurajan, and K. Luger (2004). "Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA". In: *Methods in Enzymology* 375.970, pp. 23–44.

Eberl, H. C., C. G. Spruijt, C. D. Kelstrup, M. Vermeulen, and M. Mann (2013). "A Map of General and Specialized Chromatin Readers in Mouse Tissues Generated by Label-free Interaction Proteomics". In: *Molecular Cell* 49.2, pp. 368–378.

Edmond, V., E. Moysan, S. Khochbin, P. Matthias, C. Brambilla, E. Brambilla, S. Gazzeri, and B. Eymin (2011). "Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin". In: *EMBO Journal* 30.3, pp. 510–523.

Endoh, M. et al. (2012). "Histone H2A Mono-Ubiquitination Is a Crucial Step to Mediate PRC1-Dependent Repression of Developmental Genes to Maintain ES Cell Identity". In: *PLoS Genetics* 8.7. Ed. by D. Schübeler, e1002774.

Ernst, P. and C. R. Vakoc (2012). "WRAD: Enabler of the SET1-family of H3K4 methyltransferases". In: *Briefings in Functional Genomics* 11.3, pp. 217–226.

Eskeland, R. et al. (2010). "Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination". In: *Molecular Cell* 38.3, pp. 452–464.

Faast, R., V. Thonglairoam, T. C. Schulz, J. Beall, J. R. Wells, H. Taylor, K. Matthaei, P. D. Rathjen, D. J. Tremethick, and I. Lyons (2001). "Histone variant H2A.Z is required for early mammalian development". In: *Current Biology* 11.15, pp. 1183–1187.

Fan, J. Y., F. Gordon, K. Luger, J. C. Hansen, and D. J. Tremethick (2002). "The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states". In: *Nature Structural Biology* 9.3, pp. 172–176.

Fan, J. Y., D. Rangasamy, K. Luger, and D. J. Tremethick (2004). "H2A.Z alters the nucleosome surface to promote HP1α-mediated chromatin fiber folding". In: *Molecular Cell* 16.4, pp. 655–661.

Farcas, A. M. et al. (2012). "KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands". In: *eLife* 1.1, pp. 1–26.

Faust, C., A. Schumacher, B. Holdener, and T. Magnuson (1995). "The eed mutation disrupts anterior mesoderm production in mice". In: *Development* 121.2, pp. 273–285.

Fenley, A. T., R. Anandakrishnan, Y. H. Kidane, and A. V. Onufriev (2018). "Modulation of nucleosomal DNA accessibility via charge-altering post-translational modifications in histone core". In: *Epigenetics and Chromatin* 11.1, pp. 1–19.

Ferrai, C. et al. (2017). " RNA polymerase II primes Polycomb-repressed developmental genes throughout terminal neuronal differentiation ". In: *Molecular Systems Biology* 13.10, p. 946.

Ferrari, K. J., A. Scelfo, S. Jammula, A. Cuomo, I. Barozzi, A. St??tzer, W. Fischle, T. Bonaldi, and D. Pasini (2014). "Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity". In: *Molecular Cell* 53.1, pp. 49–62.

Filippakopoulos, P. et al. (2012). "Histone recognition and large-scale structural analysis of the human bromodomain family". In: *Cell* 149.1, pp. 214–231.

Fischle, W., Y. Wang, S. A. Jacobs, Y. Kim, C. D. Allis, and S. Khorasanizadeh (2003). "Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains". In: *Genes and Development* 17.15, pp. 1870–1881.

FitzGerald, K. T. and M. O. Diaz (1999). "MLL2: A new mammalian member of the trx/MLL family of genes." In: *Genomics* 59.2, pp. 187–192.

Flanagan, J. F., L. Z. Mi, M. Chruszcz, M. Cymborowski, K. L. Clines, Y. Kim, W. Minor, F. Rastinejad, and S. Khorasanizadeh (2005). "Double chromodomains cooperate to recognize the methylated histone H3 tail". In: *Nature* 438.7071, pp. 1181–1185.

Flaus, A. (2011). "Principles and practice of nucleosome positioning in vitro". In: *Frontiers in Life Science* 5.1-2, pp. 5–27.

Flaus, A., K. Luger, S. Tan, and T. J. Richmond (1996). "Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals". In: *Proceedings of the National Academy of Sciences of the United States of America* 93.4, pp. 1370–1375.

Fletcher, T. M. and J. C. Hansen (1995). "Core histone tail domains mediate oligonucleosome folding and nucleosomal DNA organization through distinct molecular mechanisms". In: *Journal of Biological Chemistry* 270.43, pp. 25359–25362.

Francis, N. J., N. E. Follmer, M. D. Simon, G. Aghia, and J. D. Butler (2009). "Polycomb Proteins Remain Bound to Chromatin and DNA during DNA Replication In Vitro". In: *Cell* 137.1, pp. 110–122.

Francis, N. J., R. E. Kingston, and C. L. Woodcock (2004). "Chromatin compaction by a polycomb group protein complex". In: *Science* 306.5701, pp. 1574–1577.

Freeman, L., H. Kurumizaka, and A. P. Wolffe (1996). "Functional domains for assembly of histones H3 and H4 into the chromatin of Xenopus embryos". In: *Proceedings of the National Academy of Sciences of the United States of America* 93.23, pp. 12780–12785.

Fujisawa, T. and P. Filippakopoulos (2017). "Functions of bromodomain-containing proteins and their roles in homeostasis and cancer". In: *Nature Reviews Molecular Cell Biology* 18.4, pp. 246–262.

Fursova, N. A., N. P. Blackledge, M. Nakayama, S. Ito, Y. Koseki, A. M. Farcas, H. W. King, H. Koseki, and R. J. Klose (2019). "Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression". In: *Molecular Cell* 74.5, 1020–1036.e8.

Gao, Y., H. Gan, Z. Lou, and Z. Zhang (2018). "Asf1a resolves bivalent chromatin domains for the induction of lineage-specific genes during mouse embryonic stem cell differentiation". In: *Proceedings of the National Academy of Sciences* 115.27, p. 201801909.

Gao, Z., J. Zhang, R. Bonasio, F. Strino, A. Sawai, F. Parisi, Y. Kluger, and D. Reinberg (2012). "PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes". In: *Molecular Cell* 45.3, pp. 344–356.

Gasiunas, G., R. Barrangou, P. Horvath, and V. Siksnys (2012). "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.39, pp. 2579–2586.

Geisberg, J. V. and K. Struhl (2004). "Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo." In: *Nucleic acids research* 32.19, pp. 1–8.

Geisler, S. J. and R. Paro (2015). "Trithorax and polycomb group-dependent regulation: A tale of opposing activities". In: *Development (Cambridge)* 142.17, pp. 2876–2887.

Gentile, C., S. Berlivet, A. Mayran, D. Paquette, F. Guerard-Millet, E. Bajon, J. Dostie, and M. Kmita (2019). "PRC2-Associated Chromatin Contacts in the Developing Limb Reveal a Possible Mechanism for the Atypical Role of PRC2 in HoxA Gene Expression". In: *Developmental Cell* 50.2, 184–196.e4.

Gévry, N., M. C. Ho, L. Laflamme, D. M. Livingston, and L. Gaudreau (2007). "p21 transcription is regulated by differential localization of histone H2A.Z". In: *Genes and Development* 21.15, pp. 1869–1881.

Giaimo, B. D., F. Ferrante, A. Herchenröther, S. B. Hake, and T. Borggrefe (2019). "The histone variant H2A.Z in gene regulation". In: *Epigenetics and Chromatin* 12.1, pp. 1–22.

Gil, J. and A. O'Loghlen (2014). "PRC1 complex diversity: where is it taking us?" In: *Trends in Cell Biology* 24.11, pp. 632–641.

Giri, S. and S. G. Prasanth (2015). "Association of ORCA/LRWD1 with repressive histone methyl transferases mediates heterochromatin organization". In: *Nucleus* 6.6, pp. 435–441.

Glaser, S. et al. (2009). "The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis". In: *Epigenetics & Chromatin* 2.1, p. 5.

Gloss, L. M. and J. F. Kirsch (1995). "Decreasing the Basicity of the Active Site Base, Lys-258, of Escherichia Coli Aspartate Aminotransferase by Replacement with γ-Thialysine". In: *Biochemistry* 34.12, pp. 3990–3998.

Goeman, F., K. Otto, S. Kyrylenko, O. Schmidt, and A. Baniahmad (2008). "ING2 recruits histone methyltransferase activity with methylation site specificity distinct from histone H3 lysines 4 and 9". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1783.10, pp. 1673–1680.

Gordon, P. M., M. A. Soliman, P. Bose, Q. Trinh, C. W. Sensen, and K. Riabowol (2008). "Interspecies data mining to predict novel ING-protein interactions in human". In: *BMC Genomics* 9.1, p. 426.

Gottesfeld, J. M. and K. Luger (2001). "Energetics and affinity of the histone octamer for defined DNA sequences". In: *Biochemistry* 40.37, pp. 10927–10933.

Grant, P. A., D. Schieltz, M. G. Pray-Grant, D. J. Steger, J. C. Reese, J. R. Yates, and J. L. Workman (1998). "A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation". In: *Cell* 94.1, pp. 45–53.

Grau, D. J., B. A. Chapman, J. D. Garlick, M. Borowsky, N. J. Francis, and R. E. Kingston (2011). "Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge". In: *Genes & Development* 25.20, pp. 2210–2221.

Grijzenhout, A., J. Godwin, H. Koseki, M. R. Gdula, D. Szumska, J. F. McGouran, S. Bhattacharya, B. M. Kessler, N. Brockdorff, and S. Cooper (2016). "Functional analysis of AEBP2, a PRC2 polycomb protein, reveals a trithorax phenotype in embryonic development and in ESCs". In: *Development (Cambridge)* 143.15, pp. 2716–2723.

Grossniklaus, U. and R. Paro (2014). "Transcriptional Silencing by Polycomb-Group Proteins". In: *Cold Spring Harbor Perspectives in Biology* 6.11, a019331–a019331.

Grzenda, A., G. Lomberk, J. S. Zhang, and R. Urrutia (2009). "Sin3: Master scaffold and transcriptional corepressor". In: *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1789.6-8, pp. 443–450.

Grzybowski, A. T., Z. Chen, and A. J. Ruthenburg (2015). "Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide". In: *Molecular Cell* 58.5, pp. 886–899.

Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch, and R. A. Young (2007). "A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells". In: *Cell* 130.1, pp. 77–88.

Guillemette, B., A. R. Bataille, N. Gévry, M. Adam, M. Blanchette, F. Robert, and L. Gaudreau (2005). "Variant Histone H2A.Z Is Globally Localized to the Promoters of Inactive Yeast Genes and Regulates Nucleosome Positioning". In: *PLoS Biology* 3.12. Ed. by M. Groudine, e384.

Hagarman, J. A., M. P. Motley, K. Kristjansdottir, and P. D. Soloway (2013). "Coordinate Regulation of DNA Methylation and H3K27me3 in Mouse Embryonic Stem Cells". In: *PLoS ONE* 8.1. Ed. by J. G. Knott, e53880.

Hamiche, A., R. Sandaltzopoulos, D. A. Gdula, and C. Wu (1999). "ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF". In: *Cell* 97.7, pp. 833–842.

Hardy, S., P.-É. Jacques, N. Gévry, A. Forest, M.-È. Fortin, L. Laflamme, L. Gaudreau, and F. Robert (2009). "The Euchromatic and Heterochromatic Landscapes Are Shaped by Antagonizing Effects of Transcription on H2A.Z Deposition". In: *PLoS Genetics* 5.10. Ed. by J. D. Lieb, e1000687.

Hargreaves, D. C., T. Horng, and R. Medzhitov (2009). "Control of Inducible Gene Expression by Signal-Dependent Transcriptional Elongation". In: *Cell* 138.1, pp. 129–145.

Hartley, P. D. and H. D. Madhani (2009). "Mechanisms that Specify Promoter Nucleosome Location and Identity". In: *Cell* 137.3, pp. 445–458.

Hauri, S., F. Comoglio, M. Seimiya, M. Gerstung, T. Glatter, K. Hansen, R. Aebersold, R. Paro, M. Gstaiger, and C. Beisel (2016). "A High-Density Map for Navigating the Human Polycomb Complexome". In: *Cell Reports* 17.2, pp. 583–595.

He, C., N. Liu, D. Xie, Y. Liu, Y. Xiao, and F. Li (2019). "Structural basis for histone H3K4me3 recognition by the N-terminal domain of the PHD finger protein Spp1". In: *Biochemical Journal* 476.13, pp. 1957–1973.

Herz, H.-M., A. Garruss, and A. Shilatifard (2013). "SET for life: biochemical activities and biological functions of SET domain-containing proteins". In: *Trends in Biochemical Sciences* 38.12, pp. 621–639.

Hidalgo, I., A. Herrera-Merchan, J. M. Ligos, L. Carramolino, J. Nuñez, F. Martinez, O. Dominguez, M. Torres, and S. Gonzalez (2012). "Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest". In: *Cell Stem Cell* 11.5, pp. 649–662.

192

Hirano, T. (2016). "Condensin-Based Chromosome Organization from Bacteria to Vertebrates". In: *Cell* 164.5, pp. 847–857.

Hnilicová, J., S. Hozeifi, E. Stejskalová, E. Dušková, I. Poser, J. Humpolíčková, M. Hof, and D. Staněk (2013). "The C-terminal domain of Brd2 is important for chromatin interaction and regulation of transcription and alternative splicing". In: *Molecular Biology of the Cell* 24.22, pp. 3557–3568.

Hodges, H. C., B. Z. Stanton, K. Cermakova, C. Y. Chang, E. L. Miller, J. G. Kirkland, W. L. Ku, V. Veverka, K. Zhao, and G. R. Crabtree (2018). "Dominant-negative SMARCA4 mutants alter the accessibility landscape of tissue-unrestricted enhancers". In: *Nature Structural and Molecular Biology* 25.1, pp. 61–72.

Højfeldt, J. W., L. Hedehus, A. Laugesen, T. Tatar, L. Wiehle, and K. Helin (2019). "Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity". In: *Molecular Cell*, pp. 1–14.

Holoch, D. and R. Margueron (2017). "Polycomb Repressive Complex 2 Structure and Function". In: *Polycomb Group Proteins*. Vol. 2. Elsevier, pp. 191–224.

Hopkins, C. E., G. Hernandez, J. P. Lee, and D. R. Tolan (2005). "Aminoethylation in model peptides reveals conditions for maximizing thiol specificity". In: *Archives of Biochemistry and Biophysics* 443.1-2, pp. 1–10.

Hsu, C. C. et al. (2018a). "Gas41 links histone acetylation to H2A.Z deposition and maintenance of embryonic stem cell identity". In: *Cell Discovery* 4.1.

Hsu, C. C. et al. (2018b). "Recognition of histone acetylation by the GAS41 YEATS domain promotes H2A.Z deposition in non-small cell lung cancer". In: *Genes and Development* 32.1, pp. 58–69.

Hu, D., A. S. Garruss, X. Gao, M. A. Morgan, M. Cook, E. R. Smith, and A. Shilatifard (2013a). "The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells". In: *Nature Structural & Molecular Biology* 20.9, pp. 1093–1097.

Hu, D. et al. (2017). "Not All H3K4 Methylations Are Created Equal: Mll2/COMPASS Dependency in Primordial Germ Cell Specification". In: *Molecular Cell* 65.3, 460–475.e6.

Hu, G., K. Cui, D. Northrup, C. Liu, C. Wang, Q. Tang, K. Ge, D. Levens, C. Crane-Robinson, and K. Zhao (2013b). "H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation". In: *Cell Stem Cell* 12.2, pp. 180–192.

Huang, F., S. M. Abmayr, and J. L. Workman (2016). "Regulation of KAT6 Acetyltransferases and Their Roles in Cell Cycle Progression, Stem Cell Maintenance, and Human Disease". In: *Molecular and Cellular Biology* 36.14, pp. 1900–1907.

Hughes, C. M. et al. (2004). "Menin associates with a trithorax family histone methyltransferase complex and with the Hoxc8 locus". In: *Molecular Cell* 13.4, pp. 587–597.

Hui Ng, H., F. ois Robert, R. A. Young, and K. Struhl (2003). "Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity genome-wide fashion (Kuo et al that is linked to subsequent events of mRNA production including capping, splicing". In: *Orphanides and Reinberg* 11, pp. 709–719.

Hyun, K., J. Jeon, K. Park, and J. Kim (2017). "Writing, erasing and reading histone lysine methylations". In: *Experimental and Molecular Medicine* 49.4.

Ichikawa, Y. et al. (2017). "A synthetic biology approach to probing nucleosome symmetry". In: *eLife* 6.

Illingworth, R. S. (2019). "Chromatin folding and nuclear architecture: PRC1 function in 3D". In: *Current Opinion in Genetics & Development* 55, pp. 82–90.

Ingen, H. van, F. M. van Schaik, H. Wienk, J. Ballering, H. Rehmann, A. C. Dechesne, J. A. Kruijzer, R. M. Liskamp, H. M. Timmers, and R. Boelens (2008). "Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3". In: *Structure* 16.8, pp. 1245–1256.

Isono, K., T. A. Endo, M. Ku, D. Yamada, R. Suzuki, J. Sharif, T. Ishikura, T. Toyoda, B. E. Bernstein, and H. Koseki (2013). "SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing". In: *Developmental Cell* 26.6, pp. 565–577.

Jang, H. S., W. J. Shin, J. E. Lee, and J. T. Do (2017). "CpG and non-CpG methylation in epigenetic gene regulation and brain function". In: *Genes* 8.6, pp. 2–20.

Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." In: *Science (New York, N.Y.)* 293.5532, pp. 1074–1080.

Jeronimo, C., S. Watanabe, C. D. Kaplan, C. L. Peterson, and F. Robert (2015). "The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations". In: *Molecular Cell* 58.6, pp. 1113–1123.

Jiang, H., A. Shukla, X. Wang, W.-y. Chen, B. E. Bernstein, and R. G. Roeder (2011). "Role for Dpy-30 in ES Cell-Fate Specification by Regulation of H3K4 Methylation within Bivalent Domains". In: *Cell* 144.4, pp. 513–525.

Jiao, L. and X. Liu (2015). "Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2". In: *Science* 350.6258, aac4383–aac4383.

Jiao, L. and X. Liu (2016). "Structural analysis of an active fungal PRC2". In: *Nucleus* 7.3, pp. 284–291.

Jin, C. and G. Felsenfeld (2007). "Nucleosome stability mediated by histone variants H3.3 and H2A.Z". In: *Genes and Development* 21.12, pp. 1519–1529.

Jin, C., C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao, and G. Felsenfeld (2009). "H3.3/H2A.Z double variant–containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions". In: *Nature Genetics* 41.8, pp. 941–945.

Johnson, E. C. B. and S. B. H. Kent (2006). "Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction". In: *Journal of the American Chemical Society* 128.20, pp. 6640–6646.

Jones, P. A. (2012). "Functions of DNA methylation: Islands, start sites, gene bodies and beyond". In: *Nature Reviews Genetics* 13.7, pp. 484–492.

Jorgensen, S., G. Schotta, and C. S. Sorensen (2013). "Histone H4 Lysine 20 methylation: key player in epigenetic regulation of genomic integrity". In: *Nucleic Acids Research* 41.5, pp. 2797–2806.

Juan, A. H. et al. (2016). "Roles of H3K27me2 and H3K27me3 Examined during Fate Specification of Embryonic Stem Cells". In: *Cell Reports* 17.5, pp. 1369–1382.

Justin, N. et al. (2016). "Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2". In: *Nature Communications* 7.1, p. 11316.

Kalashnikova, A. A., M. E. Porter-Goff, U. M. Muthurajan, K. Luger, and J. C. Hansen (2013). "The role of the nucleosome acidic patch in modulating higher order chromatin structure". In: *Journal of the Royal Society Interface* 10.82.

Kalb, R., S. Latwiel, H. I. Baymaz, P. W. T. C. Jansen, C. W. Müller, M. Vermeulen, and J. Müller (2014). "Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression". In: *Nature Structural & Molecular Biology* 21.6, pp. 569–571.

Kalkan, T. et al. (2017). "Tracking the embryonic stem cell transition from ground state pluripotency". In: *Development* 144.7, pp. 1221–1234.

Kaneko, S., R. Bonasio, R. Saldaña-Meyer, T. Yoshida, J. Son, K. Nishino, A. Umezawa, and D. Reinberg (2014). "Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin". In: *Molecular Cell* 53.2, pp. 290–300.

Kaneko, S., J. Son, S. S. Shen, D. Reinberg, and R. Bonasio (2013). "PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells". In: *Nature Structural and Molecular Biology* 20.11, pp. 1258–1264.

Kanhere, A. et al. (2010). "Short RNAs Are Transcribed from Repressed Polycomb Target Genes and Interact with Polycomb Repressive Complex-2". In: *Molecular Cell* 38.5, pp. 675–688.

Kapoor-Vazirani, P. and P. M. Vertino (2014). "A dual role for the histone methyltransferase PR-SET7/SETD8 and histone H4 lysine 20 monomethylation in the local regulation of RNA polymerase II pausing". In: *Journal of Biological Chemistry* 289.11, pp. 7425–7437.

Karachentsev, D., K. Sarma, D. Reinberg, and R. Steward (2005). "PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis". In: *Genes and Development* 19.4, pp. 431–435.

Karmodiya, K., A. R. Krebs, M. Oulad-Abdelghani, H. Kimura, and L. Tora (2012). "H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells". In: *BMC Genomics* 13.1.

Kasinath, V., M. Faini, S. Poepsel, D. Reif, X. A. Feng, G. Stjepanovic, R. Aebersold, and E. Nogales (2018). "Structures of human PRC2 with its cofactors AEBP2 and JARID2". In: *Science* 359.6378, pp. 940–944.

Kassis, J. A. and J. L. Brown (2013). "Polycomb Group Response Elements in Drosophila and Vertebrates". In: *Bone.* Vol. 23. 1, pp. 83–118.

Kassis, J. A., J. A. Kennison, and J. W. Tamkun (2017). "Polycomb and trithorax group genes in drosophila". In: *Genetics* 206.4, pp. 1699–1725.

Kebede, A. F., R. Schneider, and S. Daujat (2015). "Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest". In: *FEBS Journal* 282.9, pp. 1658–1674.

Kent, S. B. H. (2006). "Fundamental Insights into the Mechanism of the Native Chemical Ligation Reaction". In: 6, pp. 1–28.

Kenyon, G. L. and T. W. Bruice (1977). "Novel sulfhydryl reagents". In: *Methods in Enzymology.* Vol. 47. C, pp. 407–430.

Ketel, C. S., E. F. Andersen, M. L. Vargas, J. Suh, S. Strome, and J. A. Simon (2005). "Subunit Contributions to Histone Methyltransferase Activities of Fly and Worm Polycomb Group Complexes". In: *Molecular and Cellular Biology* 25.16, pp. 6857–6868.

Kidder, B. L., G. Hu, and K. Zhao (2014). "KDM5B focuses H3K4 methylation near promoters and enhancers during embryonic stem cell self-renewal and differentiation". In: *Genome Biology* 15.2.

Kim, D.-H. et al. (2013). "Histone H3K27 Trimethylation Inhibits H3 Binding and Function of SET1-Like H3K4 Methyltransferase Complexes". In: *Molecular and Cellular Biology* 33.24, pp. 4936–4946.

Kim, H., K. Kang, and J. Kim (2009). "AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2". In: *Nucleic Acids Research* 37.9, pp. 2940–2950.

Kim, S. et al. (2016). "Mechanism of Histone H3K4me3 recognition by the plant homeodomain of inhibitor of growth 3". In: *Journal of Biological Chemistry* 291.35, pp. 18326–18341.

King, A. D., K. Huang, L. Rubbi, S. Liu, C.-Y. Wang, Y. Wang, M. Pellegrini, and G. Fan (2016). "Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells". In: *Cell Reports* 17.1, pp. 289–302.

King, H. W., N. A. Fursova, N. P. Blackledge, and R. J. Klose (2018). "Polycomb repressive complex 1 shapes the nucleosome landscape but not accessibility at target genes". In: *Genome Research* 28.10, pp. 1494–1507.

Kinkley, S., J. Helmuth, J. K. Polansky, I. Dunkel, G. Gasparoni, S. Fröhler, W. Chen, J. Walter, A. Hamann, and H. R. Chung (2016). "ReChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4+ memory T cells". In: *Nature Communications* 7.

Klein, B. J., M.-E. Lalonde, J. Côté, X.-J. Yang, and T. G. Kutateladze (2014). "Crosstalk between epigenetic readers regulates the MOZ/MORF HAT complexes". In: *Epigenetics* 9.2, pp. 186–193.

Klein, B. J., J. Simithy, X. Wang, J. W. Ahn, F. H. Andrews, Y. Zhang, J. Côté, X. Shi, B. A. Garcia, and T. G. Kutateladze (2017). "Recognition of Histone H3K14 Acylation by MORF". In: *Structure* 25.4, 650–654.e2.

Klein, B. J. et al. (2016). "Bivalent interaction of the PZP domain of BRPF1 with the nucleosome impacts chromatin dynamics and acetylation". In: *Nucleic Acids Research* 44.1, pp. 472–484.

Klein, P., T. Pawson, and M. Tyers (2003). "Mathematical Modeling Suggests Cooperative Interactions between a Disordered Polyvalent Ligand and a Single Receptor Site". In: *Current Biology* 13.19, pp. 1669–1678.

Kloet, S. L., M. M. Makowski, H. I. Baymaz, L. Van Voorthuijsen, I. D. Karemaker, A. Santanach, P. W. Jansen, L. Di Croce, and M. Vermeulen (2016). "The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation". In: *Nature Structural and Molecular Biology* 23.7, pp. 682–690.

Klose, R. J. and A. P. Bird (2006). "Genomic DNA methylation: The mark and its mediators". In: *Trends in Biochemical Sciences* 31.2, pp. 89–97.

Klose, R. J., S. Cooper, A. M. Farcas, N. P. Blackledge, and N. Brockdorff (2013). "Chromatin Sampling-An Emerging Perspective on Targeting Polycomb Repressor Proteins". In: *PLoS Genetics* 9.8.

Knoepfler, P. S. and R. N. Eisenman (1999). "Sin meets NuRD and other tails of repression". In: *Cell* 99.5, pp. 447–450.

Kohlmaier, A., F. Savarese, M. Lachner, J. Martens, T. Jenuwein, and A. Wutz (2004). "A chromosomal memory triggered by Xist regulates histone methylation in X inactivation". In: *PLoS Biology* 2.7.

Kondo, T., K. Isono, K. Kondo, T. A. Endo, S. Itohara, M. Vidal, and H. Koseki (2014). "Polycomb Potentiates Meis2 Activation in Midbrain by Mediating Interaction of the Promoter with a Tissue-Specific Enhancer". In: *Developmental Cell* 28.1, pp. 94–101.

Koutelou, E., C. L. Hirsch, and S. Y. R. Dent (2010). "Multiple faces of the SAGA complex". In: *Current Opinion in Cell Biology* 22.3, pp. 374–382.

Kouzarides, T. (2007). "Chromatin Modifications and Their Function". In: *Cell* 128.4, pp. 693–705.

Krasnov, A. N., M. Y. Mazina, J. V. Nikolenko, and N. E. Vorobyeva (2016). "On the way of revealing coactivator complexes cross-talk during transcriptional activation". In: *Cell and Bioscience* 6.1, pp. 1–14.

Ku, M., J. D. Jaffe, R. P. Koche, E. Rheinbay, M. Endoh, H. Koseki, S. A. Carr, and B. E. Bernstein (2012). "H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions". In: *Genome Biology* 13.10.

Ku, M. et al. (2008). "Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains". In: *PLoS Genetics* 4.10.

Kundu, S., F. Ji, H. Sunwoo, G. Jain, J. T. Lee, R. I. Sadreyev, J. Dekker, and R. E. Kingston (2017). "Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation". In: *Molecular Cell* 65.3, 432–446.e5.

Lalonde, M. E. et al. (2013). "Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity". In: *Genes and Development* 27.18, pp. 2009–2024.

Lan, F. et al. (2007). "A histone H3 lysine 27 demethylase regulates animal posterior development". In: *Nature* 449.7163, pp. 689–694.

Landeira, D. et al. (2010). "Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators". In: *Nature Cell Biology* 12.6, pp. 618–624.

Latosinska, A. et al. (2015). "Comparative analysis of label-free and 8-plex iTRAQ approach for quantitative tissue proteomic analysis". In: *PLoS ONE* 10.9, pp. 1–25.

Lau, M. S. et al. (2017). "Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning". In: *Science* 355.6329, pp. 1081–1084.

Lau, P. N. I. and C. W. E. So (2015). "Polycomb and Trithorax factors in transcriptional and epigenetic regulation". In: *Epigenetic Gene Expression and Regulation.* Elsevier, pp. 63–94.

Lauberth, S. M., T. Nakayama, X. Wu, A. L. Ferris, Z. Tang, S. H. Hughes, and R. G. Roeder (2013). "H3K4me3 Interactions with TAF3 Regulate Preinitiation Complex Assembly and Selective Gene Activation". In: *Cell* 152.5, pp. 1021–1036.

Laue, K., S. Daujat, J. G. Crump, N. Plaster, H. H. Roehl, C. B. Kimmel, R. Schneider, and M. Hammerschmidt (2008). "The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity". In: *Development* 135.11, pp. 1935–1946.

Laugesen, A., J. W. Højfeldt, and K. Helin (2019). "Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation". In: *Molecular Cell* 74.1, pp. 8–18.

Lawrence, M., S. Daujat, and R. Schneider (2016). "Lateral Thinking: How Histone Modifications Regulate Gene Expression". In: *Trends in Genetics* 32.1, pp. 42–56.

Lechner, C. C., N. D. Agashe, and B. Fierz (2016). "Traceless Synthesis of Asymmetrically Modified Bivalent Nucleosomes". In: *Angewandte Chemie International Edition* 55.8, pp. 2903–2906.

Lee, D. Y., J. J. Hayes, D. Pruss, and A. P. Wolffe (1993). "A positive role for histone acetylation in transcription factor access to nucleosomal DNA". In: *Cell* 72.1, pp. 73–84.

Lee, T. I. et al. (2006). "Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells". In: *Cell* 125.2, pp. 301–313.

Lerman, L. S. (1961). "Structural considerations in the interaction of DNA and acridines". In: *Journal of Molecular Biology* 3.1, pp. 18–30.

LeRoy, G., G. Orphanides, and W. S. Lane (1998). "Requirement of RSF and FACT for transcription of chromatin templates in vitro". In: *Science* 282.5395, pp. 1900–1904.

LeRoy, G., B. Rickards, and S. J. Flint (2008). "The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription". In: *Molecular Cell* 30.1, pp. 51–60.

Leung, J. W., P. Agarwal, M. D. Canny, F. Gong, A. D. Robison, I. J. Finkelstein, D. Durocher, and K. M. Miller (2014). "Nucleosome Acidic Patch Promotes RNF168- and RING1B/BMI1-Dependent H2AX and H2A Ubiquitination and DNA Damage Signaling". In: *PLoS Genetics* 10.3. Ed. by N. Maizels, e1004178.

Lewis, E. B. (1978). "A gene complex controlling segmentation in Drosophila". In: *Nature* 276.5688, pp. 565–570.

Li, B., S. G. Pattenden, D. Lee, J. Gutierrez, J. Chen, C. Seidel, J. Gerton, and J. L. Workman (2005a). "Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling". In: *Proceedings of the National Academy of Sciences* 102.51, pp. 18385–18390.

Li, G., R. Margueron, M. Ku, P. Chambon, B. E. Bernstein, and D. Reinberg (2010). "Jarid2 and PRC2, partners in regulating gene expression". In: *Genes and Development* 24.4, pp. 368–380.

Li, G., M. Levitus, C. Bustamante, and J. Widom (2005b). "Rapid spontaneous accessibility of nucleosomal DNA". In: *Nature Structural and Molecular Biology* 12.1, pp. 46–53.

Li, G. and J. Widom (2004). "Nucleosomes facilitate their own invasion". In: *Nature Structural and Molecular Biology* 11.8, pp. 763–769.

Li, G. and D. Reinberg (2011). "Chromatin higher-order structures and gene regulation". In: *Current Opinion in Genetics & Development* 21.2, pp. 175–186.

Li, H., S. Ilin, W. Wang, E. M. Duncan, J. Wysocka, C. D. Allis, and D. J. Patel (2006). "Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF". In: *Nature* 442.7098, pp. 91–95.

Li, S. and M. A. Shogren-Knaak (2008). "Cross-talk between histone H3 tails produces cooperative nucleosome acetylation". In: *Proceedings of the National Academy of Sciences* 105.47, pp. 18243–18248.

Li, S. and M. A. Shogren-Knaak (2009). "The Gcn5 bromodomain of the SAGA complex facilitates cooperative and cross-tail acetylation of nucleosomes". In: *Journal of Biological Chemistry* 284.14, pp. 9411–9417.

Li, W., S. Nagaraja, G. P. Delcuve, M. J. Hendzel, and J. R. Davie (1993). "Effects of histone acetylation". In: *In Vivo* 744, pp. 737–744.

Li, Y. et al. (2016). "Structural basis for activity regulation of MLL family methyltransferases". In: *Nature* 530.7591, pp. 447–452.

Li, Y. et al. (2018). "Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys". In: *Genome Biology* 19.1, pp. 1–16.

Liang, X. et al. (2016). "Structural basis of H2A.Z recognition by SRCAP chromatin-remodeling subunit YL1". In: *Nature Structural and Molecular Biology* 23.4, pp. 317–323.

Liefke, R., V. Karwacki-Neisius, and Y. Shi (2016). "EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape". In: *Molecular Cell* 64.4, pp. 659–672.

Lindeman, L. C. et al. (2011). "Prepatterning of Developmental Gene Expression by Modified Histones before Zygotic Genome Activation". In: *Developmental Cell* 21.6, pp. 993–1004.

Liokatis, S., A. Stützer, S. J. Elsässer, F. X. Theillet, R. Klingberg, B. Van Rossum, D. Schwarzer, C. D. Allis, W. Fischle, and P. Selenko (2012). "Phosphorylation of histone H3 Ser10 establishes a hierarchy for subsequent intramolecular modification events". In: *Nature Structural and Molecular Biology* 19.8, pp. 819–823.

Lionaki, E., I. Gkikas, and N. Tavernarakis (2016). "Differential Protein Distribution between the Nucleus and Mitochondria: Implications in Aging". In: *Frontiers in Genetics* 7.

Liu, H.-X., S. L. Chew, L. Cartegni, M. Q. Zhang, and A. R. Krainer (2000). "Exonic Splicing Enhancer Motif Recognized by Human SC35 under Splicing Conditions". In: *Molecular and Cellular Biology* 20.3, pp. 1063–1071.

Liu, L., S. Qin, J. Zhang, P. Ji, Y. Shi, and J. Wu (2012). "Solution structure of an atypical PHD finger in BRPF2 and its interaction with DNA". In: *Journal of Structural Biology* 180.1, pp. 165–173.

Lo, W. S., R. C. Trievel, J. R. Rojas, L. Duggan, J. Y. Hsu, C. D. Allis, R. Marmorstein, and S. L. Berger (2000). "Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14". In: *Molecular Cell* 5.6, pp. 917–926.

Locasale, J. W. (2008). "Allovalency revisited: An analysis of multisite phosphorylation and substrate rebinding". In: *The Journal of Chemical Physics* 128.11, p. 115106.

Loomis, R. J., Y. Naoe, J. B. Parker, V. Savic, M. R. Bozovsky, T. Macfarlan, J. L. Manley, and D. Chakravarti (2009). "Chromatin Binding of SRp20 and ASF/SF2 and Dissociation from Mitotic Chromosomes Is Modulated by Histone H3 Serine 10 Phosphorylation". In: *Molecular Cell* 33.4, pp. 450–461.

Lowary, P. T. and J. Widom (1998). "New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning." In: *Journal of molecular biology* 276.1, pp. 19–42.

Lubitz, S., S. Glaser, J. Schaft, A. F. Stewart, and K. Anastassiadis (2007). "Increased Apoptosis and Skewed Differentiation in Mouse Embryonic Stem Cells Lacking the Histone Methyltransferase Mll2". In: *Molecular Biology of the Cell* 18.6. Ed. by W. Bickmore, pp. 2356–2366.

Luger, K., a. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond (1997a). "Crystal structure of the nucleosome core particle at 2.8 A resolution." In: *Nature* 389.6648, pp. 251–260.

Luger, K., T. J. Rechsteiner, and T. J. Richmond (1999a). "Expression and purification of recombinant histones and nucleosome reconstitution." In: *Methods in molecular biology (Clifton, N.J.)* 119.4, pp. 1–16.

Luger, K. (2006). "Dynamic nucleosomes". In: *Chromosome Research* 14.1, pp. 5–16.

Luger, K., M. L. Dechassa, and D. J. Tremethick (2012). "New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?" In: *Nature Reviews Molecular Cell Biology* 13.7, pp. 436–447.

Luger, K., T. J. Rechsteiner, A. J. Flaus, M. M. Waye, and T. J. Richmond (1997b). "Characterization of nucleosome core particles containing histone proteins made in bacteria". In: *Journal of Molecular Biology* 272.3, pp. 301–311.

Luger, K., T. J. Rechsteiner, and T. J. Richmond (1999b). "Preparation of nucleosome core particle from recombinant histones". In: *Nature*. Vol. 304. 1997, pp. 3–19.

Lutter, L. C. (1979). "Precise location of DNase I cutting sites in the nucleosome core determined by high resolution gel electrophoresis". In: *Nucleic Acids Research* 6.1, pp. 41–56.

Lynch, M. D. et al. (2012). "An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment." In: *The EMBO journal* 31.2, pp. 317–29.

Ma, C. et al. (2016). "Nono, a Bivalent Domain Factor, Regulates Erk Signaling and Mouse Embryonic Stem Cell Pluripotency". In: *Cell Reports* 17.4, pp. 997–1007.

Mahadevan, J. and D. G. Skalnik (2016). "Efficient differentiation of murine embryonic stem cells requires the binding of CXXC finger protein 1 to DNA or methylated histone H3-Lys4". In: *Gene* 594.1, pp. 1–9.

Majello, B., G. Napolitano, P. De Luca, and L. Lania (1998). "Recruitment of human TBP selectively activates RNA polymerase II TATA- dependent promoters". In: *Journal of Biological Chemistry* 273.26, pp. 16509–16516.

Makowski, M. M., C. Gräwe, B. M. Foster, N. V. Nguyen, T. Bartke, and M. Vermeulen (2018). "Global profiling of protein–DNA and protein–nucleosome binding affinities using quantitative mass spectrometry". In: *Nature Communications* 9.1, p. 1653.

Mangs, A. H. and B. J. Morris (2008). "ZRANB2: Structural and functional insights into a novel splicing protein". In: *International Journal of Biochemistry and Cell Biology* 40.11, pp. 2353–2357.

Mantsoki, A., G. Devailly, and A. Joshi (2018). "Dynamics of promoter bivalency and RNAP II pausing in mouse stem and differentiated cells". In: *BMC Developmental Biology* 18.1, pp. 1–16.

Margueron, R., G. Li, K. Sarma, A. Blais, J. Zavadil, C. L. Woodcock, B. D. Dynlacht, and D. Reinberg (2008). "Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms". In: *Molecular Cell* 32.4, pp. 503–518.

Margueron, R. and D. Reinberg (2010). "Chromatin structure and the inheritance of epigenetic information". In: *Nature Reviews Genetics* 11.4, pp. 285–296.

— (2011). "The Polycomb complex PRC2 and its mark in life." In: *Nature* 469.7330, pp. 343–349.

Margueron, R. et al. (2009). "Role of the polycomb protein Eed in the propagation of repressive histone marks". In: *Nature cell biology* 461.7265, pp. 762–767.

Marks, H. et al. (2012). "The transcriptional and epigenomic foundations of ground state pluripotency". In: *Cell* 149.3, pp. 590–604.

Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier (2012). "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity". In: *Science* 337.6096, pp. 816–821.

Maston, G. A., L. J. Zhu, L. Chamberlain, L. Lin, M. Fang, and M. R. Green (2012). "Non-canonical TAF complexes regulate active promoters in human embryonic stem cells". In: *eLife* 2012.1, pp. 1–19.

Mavrich, T. N. et al. (2008). "Nucleosome organization in the Drosophila genome". In: *Nature* 453.7193, pp. 358–362.

Meehan, R. R. and S. Pennings (2017). " Shoring up DNA methylation and H3K27me3 domain demarcation at developmental genes ". In: *The EMBO Journal* 36.23, pp. 3407–3408.

Meersseman, G., S. Pennings, and E. Bradbury (1992). "Mobile nucleosomes–a general behavior." In: *The EMBO Journal* 11.8, pp. 2951–2959.

Megger, D. A., L. L. Pott, M. Ahrens, J. Padden, T. Bracht, K. Kuhlmann, M. Eisenacher, H. E. Meyer, and B. Sitek (2014). "Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines". In: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1844.5, pp. 967–976.

Mencía, M., Z. Moqtaderi, J. V. Geisberg, L. Kuras, and K. Struhl (2002). "Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast". In: *Molecular Cell* 9.4, pp. 823–833.

Mendenhall, E. M., R. P. Koche, T. Truong, V. W. Zhou, B. Issac, A. S. Chi, M. Ku, and B. E. Bernstein (2010). "GC-rich sequence elements recruit PRC2 in mammalian ES cells". In: *PLoS Genetics* 6.12, pp. 1–10.

Meneghini, M. D., M. Wu, and H. D. Madhani (2003). "Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin". In: *Cell* 112.5, pp. 725–736.

Meslamani, J., S. G. Smith, R. Sanchez, and M. M. Zhou (2016). "Structural features and inhibitors of bromodomains". In: *Drug Discovery Today: Technologies* 19, pp. 3–15.

Mi, W. et al. (2017). "YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer". In: *Nature Communications* 8.1, p. 1088.

Mierlo, G. van, G. J. C. Veenstra, M. Vermeulen, and H. Marks (2019). "The Complexity of PRC2 Subcomplexes". In: *Trends in Cell Biology* 29.8, pp. 660–671.

Mikkelsen, T. S. et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." In: *Nature* 448.7153, pp. 553–560.

Millar, C. B., F. Xu, K. Zhang, and M. Grunstein (2006). "Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast". In: *Genes and Development* 20.6, pp. 711–722.

Miller, T., N. J. Krogan, J. Dover, H. Erdjument-Bromage, P. Tempst, M. Johnston, J. F. Greenblatt, and A. Shilatifard (2001). "COMPASS: A complex of proteins associated with a trithorax-related SET domain protein". In: *Proceedings of the National Academy of Sciences* 98.23, pp. 12902–12907.

Min, I. M., J. J. Waterfall, L. J. Core, R. J. Munroe, J. Schimenti, and J. T. Lis (2011). "Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells". In: *Genes and Development* 25.7, pp. 742–754.

Min, J., Y. Zhang, and R. M. Xu (2003). "Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at Lys 27". In: *Genes and Development* 17.15, pp. 1823–1828.

Mo, R., S. M. Rao, and Y. J. Zhu (2006). "Identification of the MLL2 complex as a coactivator for estrogen receptor α". In: *Journal of Biological Chemistry* 281.23, pp. 15714–15720.

Mochizuki-Kashio, M., Y. Mishima, and S. Miyagi (2012). "Dependency on the polycomb gene Ezh2 distinguishes fetal from adult hematopoietic stem cells (Blood (2011) 118, 25 (6553-6561))". In: *Blood* 120.4, p. 924.

Mohan, K. N., F. Ding, and J. R. Chaillet (2011). "Distinct Roles of DMAP1 in Mouse Development". In: *Molecular and Cellular Biology* 31.9, pp. 1861–1869.

Mohn, F., M. Weber, M. Rebhan, T. C. Roloff, J. Richter, M. B. Stadler, M. Bibel, and D. Schübeler (2008). "Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors". In: *Molecular Cell* 30.6, pp. 755–766.

Montgomery, N. D., D. Yee, A. Chen, S. Kalantry, S. J. Chamberlain, A. P. Otte, and T. Magnuson (2005). "The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation". In: *Current Biology* 15.10, pp. 942–947.

Morales, V. and H. Richard-Foy (2000). "Role of Histone N-Terminal Tails and Their Acetylation in Nucleosome Dynamics". In: *Molecular and Cellular Biology* 20.19, pp. 7230–7237.

Moraru, M. and T. Schalch (2019). "Chromatin fiber structural motifs as regulatory hubs of genome function?" In: *Essays in Biochemistry* 63.1, pp. 123–132.

Morey, L., G. Pascual, L. Cozzuto, G. Roma, A. Wutz, S. A. Benitah, and L. Di Croce (2012). "Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells". In: *Cell Stem Cell* 10.1, pp. 47–62.

Morgan, M. T., M. Haj-Yahya, A. E. Ringel, P. Bandi, A. Brik, and C. Wolberger (2016). "Structural basis for histone H2B deubiquitination by the SAGA DUB module". In: *Science* 351.6274, pp. 725–728.

Morselli, M. et al. (2015). "In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse". In: *eLife* 2015.4, pp. 1–21.

Moudrianakis, E. N. and G. Arents (1993). "Structure of the histone octamer core of the nucleosome and its potential interactions with DNA". In: *Cold Spring Harbor Symposia on Quantitative Biology* 58, pp. 273–279.

Mousson, F., A. Kolkman, W. W. Pijnappel, H. T. M. Timmers, and A. J. Heck (2008). "Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes". In: *Molecular and Cellular Proteomics* 7.5, pp. 845–852.

Mousson, F., F. Ochsenbein, and C. Mann (2007). "The histone chaperone Asf1 at the crossroads of chromatin and DNA checkpoint pathways". In: *Chromosoma* 116.2, pp. 79–93.

Mozzetta, C., J. Pontis, L. Fritsch, P. Robin, M. Portoso, C. Proux, R. Margueron, and S. Ait-Si-Ali (2014). "The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing". In: *Molecular Cell* 53.2, pp. 277–289.

Mujtaba, S., K. L. Manzur, J. R. Gurnon, M. Kang, J. L. Van Etten, and M.-M. Zhou (2008). "Epigenetic transcriptional repression of cellular genes by a viral SET protein." In: *Nature cell biology* 10.9, pp. 1114–1122.

Munari, F. et al. (2012). "Methylation of lysine 9 in histone H3 directs alternative modes of highly dynamic interaction of heterochromatin protein hHP1β with the nucleosome". In: *Journal of Biological Chemistry* 287.40, pp. 33756–33765.

Murai, M. J., J. Pollock, S. He, H. Miao, T. Purohit, A. Yokom, J. L. Hess, A. G. Muntean, J. Grembecka, and T. Cierpicki (2014). "The same site on the integrase-binding domain of lens epithelium-derived growth factor is a therapeutic target for MLL leukemia and HIV". In: *Blood* 124.25, pp. 3730–3737.

Murphy, P. J., B. R. Cipriany, C. B. Wallin, C. Y. Ju, K. Szeto, J. A. Hagarman, J. J. Benitez, H. G. Craighead, and P. D. Soloway (2013). "Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells". In: *Proceedings of the National Academy of Sciences* 110.19, pp. 7772–7777.

Murzina, N. V. et al. (2008). "Structural Basis for the Recognition of Histone H4 by the Histone-Chaperone RbAp46". In: *Structure* 16.7, pp. 1077–1085.

Musselman, C. A. et al. (2012). "Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1". In: *Nature Structural & Molecular Biology* 19.12, pp. 1266–1272.

Nakagawa, T. et al. (2008). "Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation". In: *Genes & Development* 22.1, pp. 37–49.

Napoles, M. de et al. (2004). "Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation". In: *Developmental Cell* 7.5, pp. 663–676.

Nekrasov, M., B. Wild, and J. Müller (2005). "Nucleosome binding and histone methyltransferase activity of Drosophila PRC2." In: *EMBO reports* 6.4, pp. 348–53.

Neri, F., A. Krepelova, D. Incarnato, M. Maldotti, C. Parlato, F. Galvagni, F. Matarese, H. G. Stunnenberg, and S. Oliviero (2013). "Dnmt3L Antagonizes DNA Methylation at Bivalent Promoters and Favors DNA Methylation at Gene Bodies in ESCs". In: *Cell* 155.1, pp. 121–134.

Nishioka, K. et al. (2002). "PR-Set7 Is a Nucleosome-Specific Methyltransferase that Modifies Lysine 20 of Histone H4 and Is Associated with Silent Chromatin The N-terminal tails of the core histone proteins protrude". In: *Molecular Cell* 9, pp. 1201–1213.

Nowak, A. J., C. Alfieri, C. U. Stirnimann, V. Rybin, F. Baudin, N. Ly-Hartig, D. Lindner, and C. W. Müller (2011). "Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding sites". In: *Journal of Biological Chemistry* 286.26, pp. 23388–23396.

Nuland, R. van, A. W. Schram, F. M. A. van Schaik, P. W. T. C. Jansen, M. Vermeulen, and H. T. Marc Timmers (2013). "Multivalent Engagement of TFIID to Nucleosomes". In: *PLoS ONE* 8.9. Ed. by P.-A. Defossez, e73495.

O'Carroll, D., S. Erhardt, M. Pagani, and S. C. Barton (2001). "The Polycomb -Group Gene Ezh2 Is Required for Early Mouse Development". In: *Molecular and cellular biology* 21.13, pp. 4330–4336.

Obier, N., Q. Lin, P. Cauchy, V. Hornich, M. Zenke, M. Becker, and A. M. Müller (2015). "Polycomb Protein EED is Required for Silencing of Pluripotency Genes upon ESC Differentiation". In: *Stem Cell Reviews and Reports* 11.1, pp. 50–61.

Okumura, F., M. Matsuzaki, K. Nakatsukasa, and T. Kamura (2012). "The Role of Elongin BC-Containing Ubiquitin Ligases". In: *Frontiers in Oncology* 2, pp. 1–13.

Old, W. M., K. Meyer-Arendt, L. Aveline-Wolf, K. G. Pierce, A. Mendoza, J. R. Sevinsky, K. A. Resing, and N. G. Ahn (2005). "Comparison of label-free methods for quantifying human proteins by shotgun proteomics". In: *Molecular and Cellular Proteomics* 4.10, pp. 1487–1502.

Oliviero, G. et al. (2016). "Dynamic Protein Interactions of the Polycomb Repressive Complex 2 during Differentiation of Pluripotent Cells." In: *Molecular & cellular proteomics : MCP* 15.11, pp. 3450–3460.

Olsen, J. V., B. Macek, O. Lange, A. Makarov, S. Horning, and M. Mann (2007). "Higher-energy C-trap dissociation for peptide modification analysis". In: *Nature Methods* 4.9, pp. 709–712.

Olsen, J. G., K. Teilum, and B. B. Kragelund (2017). "Behaviour of intrinsically disordered proteins in protein–protein complexes with an emphasis on fuzziness". In: *Cellular and Molecular Life Sciences* 74.17, pp. 3175–3183.

Owen, D. J. (2000). "The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p". In: *The EMBO Journal* 19.22, pp. 6141–6149.

Pan, G., S. Tian, J. Nie, C. Yang, V. Ruotti, H. Wei, G. A. Jonsdottir, R. Stewart, and J. A. Thomson (2007). "Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells". In: *Cell Stem Cell* 1.3, pp. 299–312.

Papamichos-Chronakis, M., S. Watanabe, O. J. Rando, and C. L. Peterson (2011). "Global Regulation of H2A.Z Localization by the INO80 Chromatin-Remodeling

Enzyme Is Essential for Genome Integrity". In: *Cell* 144.2. Ed. by H. Bönig, pp. 200–213.

Park, Y. J., J. V. Chodaparambil, Y. Bao, S. J. McBryant, and K. Luger (2005). "Nucleosome assembly protein 1 exchanges histone H2A-H2B dimers and assists nucleosome sliding". In: *Journal of Biological Chemistry* 280.3, pp. 1817–1825.

Park, Y.-J., P. N. Dyer, D. J. Tremethick, and K. Luger (2004). "A New Fluorescence Resonance Energy Transfer Approach Demonstrates That the Histone Variant H2AZ Stabilizes the Histone Octamer within the Nucleosome". In: *Journal of Biological Chemistry* 279.23, pp. 24274–24282.

Pasini, D., A. P. Bracken, J. B. Hansen, M. Capillo, and K. Helin (2007). "The polycomb group protein Suz12 is required for embryonic stem cell differentiation." In: *Molecular and cellular biology* 27.10, pp. 3769–3779.

Pasini, D., A. P. Bracken, M. R. Jensen, E. L. Denchi, and K. Helin (2004). "Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity". In: *The EMBO Journal* 23.20, pp. 4061–4071.

Pasini, D., P. A. C. Cloos, J. Walfridsson, L. Olsson, J.-P. Bukowski, J. V. Johansen, M. Bak, N. Tommerup, J. Rappsilber, and K. Helin (2010a). "JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells". In: *Nature* 464.7286, pp. 306–310.

Pasini, D. et al. (2010b). "Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes". In: *Nucleic Acids Research* 38.15, pp. 4958–4969.

Patel, A., V. Dharmarajan, V. E. Vought, and M. S. Cosgrove (2009). "On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex". In: *Journal of Biological Chemistry* 284.36, pp. 24242–24256.

Pearce, J. J. H., P. B. Singh, and S. J. Gaunt (1992). "The mouse has a Polycomb-like chromobox gene". In: *Development* 114.4, pp. 921–929.

Pedersen, M. T., K. Agger, A. Laugesen, J. V. Johansen, P. A. C. Cloos, J. Christensen, and K. Helin (2014). "The Demethylase JMJD2C Localizes to H3K4me3-Positive Transcription Start Sites and Is Dispensable for Embryonic Development". In: *Molecular and Cellular Biology* 34.6, pp. 1031–1045.

Pellagatti, A. et al. (2018). "Impact of spliceosome mutations on RNA splicing in myelodysplasia: Dysregulated genes/pathways and clinical associations". In: *Blood* 132.12, pp. 1225–1240.

Pelletier, N., N. Champagne, S. Stifani, and X.-J. Yang (2002). "MOZ and MORF histone acetyltransferases interact with the Runt-domain transcription factor Runx2." In: *Oncogene* 21, pp. 2729–2740.

Peña, P. V., F. Davrazou, X. Shi, K. L. Walter, V. V. Verkhusha, O. Gozani, R. Zhao, and T. G. Kutateladze (2006). "Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2". In: *Nature* 442.7098, pp. 100–103.

Peng, J. C., A. Valouev, T. Swigut, J. Zhang, Y. Zhao, and J. Wysocka (2010). "Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells". In: *Cell* 139.7, pp. 1290–1302.

Pennings, S., G. Meersseman, and E. Bradbury (1991). "Mobility of positioned nucleosomes on 5 S rDNA". In: *Journal of Molecular Biology* 220.1, pp. 101–110.

Pepenella, S., K. J. Murphy, and J. J. Hayes (2014). "Intra- and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure". In: *Chromosoma* 123.1-2, pp. 3–13.

Perino, M., G. Van Mierlo, I. D. Karemaker, S. Van Genesen, M. Vermeulen, H. Marks, S. J. Van Heeringen, and G. J. C. Veenstra (2018). "MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding". In: *Nature Genetics* 50.7, pp. 1002–1010.

Petersen, M., A. A. Hamed, E. B. Pedersen, and J. P. Jacobsen (1999). "Bis-intercalation of homodimeric thiazole orange dye derivatives in DNA". In: *Bioconjugate Chemistry* 10.1, pp. 66–74.

Poepsel, S., V. Kasinath, and E. Nogales (2018). "Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes". In: *Nature Structural & Molecular Biology* 25.2, pp. 154–162.

Polach, K. J., P. T. Lowary, and J. Widom (2000). "Effects of core histone tail domains the equilibrium constants for dynamic DNA site accessibility in nucleosomes". In: *Journal of Molecular Biology* 298.2, pp. 211–223.

Polach, K. J. and J. Widom (1995). "Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation". In: *Journal of Molecular Biology* 254.2, pp. 130–149.

Poplawski, A., K. Hu, W. Lee, S. Natesan, D. Peng, S. Carlson, X. Shi, S. Balaz, J. L. Markley, and K. C. Glass (2014). "Molecular insights into the recognition of N-terminal histone modifications by the BRPF1 bromodomain". In: *Journal of Molecular Biology* 426.8, pp. 1661–1676.

Portoso, M., R. Ragazzini, Ž. Brenčič, A. Moiani, A. Michaud, I. Vassilev, M. Wassef, N. Servant, B. Sargueil, and R. Margueron (2017). " PRC 2 is dispensable for HOTAIR -mediated transcriptional repression ". In: *The EMBO Journal* 36.8, pp. 981–994.

Poynter, S. T. and C. Kadoch (2016). "Polycomb and trithorax opposition in development and disease". In: *Wiley Interdisciplinary Reviews: Developmental Biology* 5.6, pp. 659–688.

Qian, C. and M. M. Zhou (2006). "SET domain protein lysine methyltransferases: Structure, specificity and catalysis". In: *Cellular and Molecular Life Sciences* 63.23, pp. 2755–2763.

Qin, S., Y. Guo, C. Xu, C. Bian, M. Fu, S. Gong, and J. Min (2013). "Tudor domains of the PRC2 components PHF1 and PHF19 selectively bind to histone H3K36me3". In: *Biochemical and Biophysical Research Communications* 430.2, pp. 547–553.

Qin, S., L. Jin, J. Zhang, L. Liu, P. Ji, M. Wu, J. Wu, and Y. Shi (2011). "Recognition of unmodified histone H3 by the first PHD finger of bromodomain-PHD finger protein 2 provides insights into the regulation of histone acetyltransferases monocytic leukemic zinc-finger protein (MOZ) and MOZ-related factor (MORF)". In: *Journal of Biological Chemistry* 286.42, pp. 36944–36955.

Rada-Iglesias, A. (2017). " Ready, Set…Poised!: Polycomb target genes are bound by poised RNA polymerase II throughout differentiation ". In: *Molecular Systems Biology* 13.10, p. 950.

Raisner, R. M., P. D. Hartley, M. D. Meneghini, M. Z. Bao, C. L. Liu, S. L. Schreiber, O. J. Rando, and H. D. Madhani (2005). "Histone variant H2A.Z Marks the 5 ends of both active and inactive genes in euchromatin". In: *Cell* 123.2, pp. 233–248.

Ran, F. A., P. D. P. P. D. Hsu, J. Wright, V. Agarwala, D. a. Scott, and F. Zhang (2013). "Genome engineering using the CRISPR-Cas9 system." In: *Nature protocols* 8.11, pp. 2281–308.

Rangasamy, D., L. Berven, P. Ridgway, and D. J. Tremethick (2003). "Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development". In: *EMBO Journal* 22.7, pp. 1599–1607.

Rappsilber, J., Y. Ishihama, and M. Mann (2003). "Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics". In: *Analytical Chemistry* 75.3, pp. 663–670.

Rathert, P., A. Dhayalan, M. Murakami, X. Zhang, R. Tamas, R. Jurkowska, Y. Komatsu, Y. Shinkai, X. Cheng, and A. Jeltsch (2008). "Protein lysine methyltransferase G9a acts on non-histone targets". In: *Nature Chemical Biology* 4.6, pp. 344–346.

Reddington, J. P. et al. (2013). "Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes". In: *Genome Biology* 14.3, R25.

Richmond, T. J., J. T. Finch, B. Rushton, D. Rhodes, and A. Klug (1984). "Structure of the nucleosome core particle at 7 resolution". In: *Nature* 311.5986, pp. 532–537.

Richmond, T. J. and C. A. Davey (2003). "The structure of DNA in the nucleosome core". In: *Nature* 423.6936, pp. 145–150.

Rinn, J. L. et al. (2007). "Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs". In: *Cell* 129.7, pp. 1311–1323.

Risner, L. E., A. Kuntimaddi, A. A. Lokken, N. J. Achille, N. W. Birch, K. Schoenfelt, J. H. Bushweller, and N. J. Zeleznik-Le (2013). "Functional specificity of CpG DNA-binding CXXC domains in mixed lineage leukemia". In: *Journal of Biological Chemistry* 288.41, pp. 29901–29910.

Rocha, S. T. da et al. (2014). "Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome". In: *Molecular Cell* 53.2, pp. 301–316.

Roh, T.-Y., S. Cuddapah, K. Cui, and K. Zhao (2006). "The genomic landscape of histone modifications in human T cells." In: *Proceedings of the National Academy of Sciences of the United States of America* 103.43, pp. 15782–15787.

Rose, N. R., H. W. King, N. P. Blackledge, N. A. Fursova, K. J. Ember, R. Fischer, B. M. Kessler, and R. J. Klose (2016). "RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes". In: *eLife* 5, pp. 1–29.

Rose, N. R. and R. J. Klose (2014). "Understanding the relationship between DNA methylation and histone lysine methylation". In: *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1839.12, pp. 1362–1372.

Rothbart, S. B., B. M. Dickson, M. S. Ong, K. Krajewski, S. Houliston, D. B. Kireev, C. H. Arrowsmith, and B. D. Strahl (2013). "Multivalent histone engagement by the linked tandem tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation". In: *Genes and Development* 27.11, pp. 1288–1298.

Rugg-Gunn, P. J., B. J. Cox, A. Ralston, and J. Rossant (2010). "Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.24, pp. 10783–10790.

Ruhl, D. D., J. Jin, Y. Cai, S. Swanson, L. Florens, M. P. Washburn, R. C. Conaway, J. W. Conaway, and J. C. Chrivia (2006). "Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes". In: *Biochemistry* 45.17, pp. 5671–5677.

Ruthenburg, A. J. et al. (2011). "Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions". In: *Cell* 145.5, pp. 692–706.

Saha, P., D. T. Sowpati, and R. K. Mishra (2019). "Epigenomic and genomic landscape of Drosophila melanogaster heterochromatic genes". In: *Genomics* 111.2, pp. 177–185.

Saksouk, N., E. Simboeck, and J. Déjardin (2015). "Constitutive heterochromatin formation and transcription in mammals". In: *Epigenetics and Chromatin* 8.1, pp. 1–17.

Sanchez, R. and M. M. Zhou (2011). "The PHD finger: A versatile epigenome reader". In: *Trends in Biochemical Sciences* 36.7, pp. 364–372.

Sander, J. D. and J. K. Joung (2014). "CRISPR-Cas systems for editing, regulating and targeting genomes." In: *Nature biotechnology* 32.4, pp. 347–55.

Santanach, A., E. Blanco, H. Jiang, K. R. Molloy, M. Sansó, J. LaCava, L. Morey, and L. Di Croce (2017). "The Polycomb group protein CBX6 is an essential regulator of embryonic stem cell identity". In: *Nature Communications* 8.1, p. 1235.

Sanulli, S. et al. (2015). "Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation". In: *Molecular Cell* 57.5, pp. 769–783.

Sapountzi, V. and J. Côté (2011). "MYST-family histone acetyltransferases: beyond chromatin". In: *Cellular and Molecular Life Sciences* 68.7, pp. 1147–1156.

Sarcinella, E., P. C. Zuzarte, P. N. I. Lau, R. Draker, and P. Cheung (2007). "Monoubiquitylation of H2A.Z Distinguishes Its Association with Euchromatin or Facultative Heterochromatin". In: *Molecular and Cellular Biology* 27.18, pp. 6457–6468.

Sarma, K., R. Margueron, A. Ivanov, V. Pirrotta, and D. Reinberg (2008). "Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27 Trimethylation In Vivo". In: *Molecular and Cellular Biology* 28.8, pp. 2718–2731.

Scelfo, A., D. Fernández-Pérez, S. Tamburri, M. Zanotti, E. Lavarone, M. Soldi, T. Bonaldi, K. J. Ferrari, and D. Pasini (2019). "Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities". In: *Molecular Cell* 74.5, 1037–1052.e7.

Schmitges, F. W. et al. (2011). "Histone methylation by PRC2 is inhibited by active chromatin marks." In: *Molecular cell* 42.3, pp. 330–41.

Schones, D. E., K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao (2008). "Dynamic Regulation of Nucleosome Positioning in the Human Genome". In: *Cell* 132.5, pp. 887–898.

Schubert, H. L., R. M. Blumenthal, and X. Cheng (2003). "Many paths to methyltransfer: A chronicle of convergence". In: *Trends in Biochemical Sciences* 28.6, pp. 329–335.

Schuettengruber, B., H. M. Bourbon, L. Di Croce, and G. Cavalli (2017). "Genome Regulation by Polycomb and Trithorax: 70 Years and Counting". In: *Cell* 171.1, pp. 34–57.

Schuettengruber, B. and G. Cavalli (2009). "Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice". In: *Development* 136.21, pp. 3531–3542.

Schwartz, Y. B. and V. Pirrotta (2014). "Ruled by Ubiquitylation: A New Order for Polycomb Recruitment". In: *Cell Reports* 8.2, pp. 321–325.

Scott, E. K., T. Lee, and L. Luo (2001). "Enok encodes a Drosophila putative histone acetyltransferase required for mushroom body neuroblast proliferation". In: *Current Biology* 11.2, pp. 99–104.

Seeliger, D., S. Soeroes, R. Klingberg, D. Schwarzer, H. Grubmüller, and W. Fischle (2012). "Quantitative assessment of protein interaction with methyl-lysine analogues by hybrid computational and experimental approaches". In: *ACS Chemical Biology* 7.1, pp. 150–154.

Sen, S., K. F. Block, A. Pasini, S. B. Baylin, and H. Easwaran (2016). "Genome-wide positioning of bivalent mononucleosomes". In: *BMC Medical Genomics* 9.1, p. 60.

Sharif, J. and H. Koseki (2018). "Rewriting the past: De novo activity of PRC2 restores global H3K27 methylation patterns". In: *Nature Structural and Molecular Biology* 25.3, pp. 197–199.

Sharma, A. B., S. Dimitrov, A. Hamiche, and E. Van Dyck (2019). "Centromeric and ectopic assembly of CENP-A chromatin in health and cancer: old marks and new tracks". In: *Nucleic Acids Research* 47.3, pp. 1051–1069.

Sharov, A. a. and M. S. Ko (2007). "Human ES Cell Profiling Broadens the Reach of Bivalent Domains". In: *Cell Stem Cell* 1.3, pp. 237–238.

Sheikh, B. N. and A. Akhtar (2019). "The many lives of KATs — detectors, integrators and modulators of the cellular environment". In: *Nature Reviews Genetics* 20.1, pp. 7–23.

Shema, E., D. Jones, N. Shoresh, L. Donohue, O. Ram, and B. E. Bernstein (2016). "Single-molecule decoding of combinatorially modified nucleosomes". In: *Science* 352.6286, pp. 717–721.

Shen, X., W. Kim, Y. Fujiwara, M. D. Simon, Y. Liu, M. R. Mysliwiec, G.-C. C. Yuan, Y. Lee, and S. H. Orkin (2009). "Jumonji Modulates Polycomb Activity and Self-Renewal versus Differentiation of Stem Cells". In: *Cell* 139.7. Ed. by H. Bönig, pp. 1303–1314.

Shen, X., Y. Liu, Y.-J. Hsu, Y. Fujiwara, J. Kim, X. Mao, G.-C. Yuan, and S. H. Orkin (2008). "EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency". In: *Molecular Cell* 32.4, pp. 491–502.

Shevchenko, A., M. Wilm, O. Vorm, and M. Mann (1996). "Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels". In: *Analytical Chemistry* 68.5, pp. 850–858.

Shi, L., H. Wen, and X. Shi (2017). "The Histone Variant H3.3 in Transcriptional Regulation and Human Disease". In: *Journal of Molecular Biology* 429.13, pp. 1934–1945.

Shi, X. et al. (2006). "ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression". In: *Nature* 442.7098, pp. 96–99.

Shinkai, Y. and M. Tachibana (2011). "H3K9 methyltransferase G9a and the related molecule GLP". In: *Genes & Development* 25.8, pp. 781–788.

Shogren-Knaak, M. (2006). "Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions". In: *Science* 311.5762, pp. 844–847.

Shogren-Knaak, M. a., C. J. Fry, and C. L. Peterson (2003). "A Native Peptide Ligation Strategy for Deciphering Nucleosomal Histone Modifications". In: *Journal of Biological Chemistry* 278.18, pp. 15744–15748.

Siegfried, Z. and H. Cedar (1997). "DNA methylation: A molecular lock". In: *Current Biology* 7.5, pp. 305–307.

Silva, J., O. Barrandon, J. Nichols, J. Kawaguchi, T. W. Theunissen, and A. Smith (2008). "Promotion of reprogramming to ground state pluripotency by signal inhibition". In: *PLoS Biology* 6.10, pp. 2237–2247.

Sim, Y.-J., M.-S. Kim, A. Nayfeh, Y.-J. Yun, S.-J. Kim, K.-T. Park, C.-H. Kim, and K.-S. Kim (2017). "2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms". In: *Stem Cell Reports* 8.5, pp. 1312–1328.

Simó-Riudalbas, L. et al. (2015). "KAT6B is a tumor suppressor histone H3 lysine 23 acetyltransferase undergoing genomic loss in small cell lung cancer". In: *Cancer Research* 75.18, pp. 3936–3944.

Simon, J. A. and R. E. Kingston (2013). "Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put". In: *Molecular Cell* 49.5, pp. 808–824.

Simon, J. a. and R. E. Kingston (2009). "Mechanisms of polycomb gene silencing: knowns and unknowns." In: *Nature reviews. Molecular cell biology* 10.10, pp. 697–708.

Simon, M. D. (2010). "Installation of Site-Specific Methylation into Histones Using Methyl Lysine Analogs". In: *Current Protocols in Molecular Biology*. SUPPL. 90. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 1–10.

Simon, M. D., F. Chu, L. R. Racki, C. C. de la Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar, and K. M. Shokat (2007). "The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones". In: *Cell* 128.5, pp. 1003–1012.

Simon, M. D. and K. M. Shokat (2012). *A method to site-specifically incorporate methyl-lysine analogues into recombinant proteins.* 1st ed. Vol. 512. Elsevier Inc., pp. 57–69.

Slifer, E. H. (1942). "A mutant stock of Drosophila with extra sex-combs". In: *Journal of Experimental Zoology* 90.1, pp. 31–40.

Smits, A. H., R. G. Lindeboom, M. Perino, S. J. Van Heeringen, G. J. C. Veenstra, and M. Vermeulen (2014). "Global absolute quantification reveals tight regulation of protein expression in single Xenopuseggs". In: *Nucleic Acids Research* 42.15, pp. 9880–9891.

Sterner, D. E. and S. L. Berger (2000). "Acetylation of Histones and Transcription-Related Factors". In: *Microbiology and Molecular Biology Reviews* 64.2, pp. 435–459.

Stock, J. K., S. Giadrossi, M. Casanova, E. Brookes, M. Vidal, H. Koseki, N. Brockdorff, A. G. Fisher, and A. Pombo (2007). "Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells". In: *Nature Cell Biology* 9.12, pp. 1428–1435.

Stoller, J. Z., L. Huang, C. C. Tan, F. Huang, D. D. Zhou, J. Yang, B. D. Gelb, and J. A. Epstein (2010). "Ash2l interacts with Tbx1 and is required during early embryogenesis". In: *Experimental Biology and Medicine* 235.5, pp. 569–576.

Strubbe, G., C. Popp, A. Schmidt, A. Pauli, L. Ringrose, C. Beisel, and R. Paro (2011). "Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners". In: *Proceedings of the National Academy of Sciences* 108.14, pp. 5572–5577.

Stützer, A., S. Liokatis, A. Kiesel, D. Schwarzer, R. Sprangers, J. Söding, P. Selenko, and W. Fischle (2016). "Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails". In: *Molecular Cell* 61.2, pp. 247–259.

Su, X., G. Zhu, X. Ding, S. Y. Lee, Y. Dou, B. Zhu, W. Wu, and H. Li (2014). "Molecular basis underlying histone H3 lysine-arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1". In: *Genes & Development* 28.6, pp. 622–636.

Su, Z. and J. M. Denu (2016). "Reading the Combinatorial Histone Language". In: *ACS Chemical Biology* 11.3, pp. 564–574.

Su, Z. et al. (2016). "Reader domain specificity and lysine demethylase-4 family function". In: *Nature Communications* 7, pp. 1–15.

Sun, J., Y. Zhao, R. McGreal, Y. Cohen-Tayar, S. Rockowitz, C. Wilczek, R. Ashery-Padan, D. Shechter, D. Zheng, and A. Cvekl (2016). "Pax6 associates with H3K4-specific histone methyltransferases Mll1, Mll2, and Set1a and regulates H3K4 methylation at promoters and enhancers". In: *Epigenetics and Chromatin* 9.1, pp. 1–18.

Surface, L. E., P. A. Fields, V. Subramanian, R. Behmer, N. Udeshi, S. E. Peach, S. A. Carr, J. D. Jaffe, and L. A. Boyer (2016). "H2A.Z.1 Monoubiquitylation Antagonizes BRD2 to Maintain Poised Chromatin in ESCs". In: *Cell Reports* 14.5, pp. 1142–1155.

Suto, R. K., M. J. Clarkson, D. J. Tremethick, and K. Luger (2000). "Crystal structure of a nucleosome core particle containing the variant histone H2A.Z." In: *Nature structural biology* 7.12, pp. 1121–4.

Swaminathan, J., E. Baxter, and V. Corces (2005). "The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin". In: *Genes & Development* 19.1, pp. 65–76.

Sze, C. C., K. Cao, C. K. Collings, S. A. Marshall, E. J. Rendleman, P. A. Ozark, F. X. Chen, M. A. Morgan, L. Wang, and A. Shilatifard (2017). "Histone H3K4 methylation-dependent and -independent functions of set1A/COMPASS in embryonic stem cell self-renewal and differentiation". In: *Genes and Development* 31.17, pp. 1732–1737.

Szenker, E., D. Ray-Gallet, and G. Almouzni (2011). "The double face of the histone variant H3.3". In: *Cell Research* 21.3, pp. 421–434.

Tachibana, M., K. Sugimoto, T. Fukushima, and Y. Shinkai (2001). "SET Domain-containing Protein, G9a, is a Novel Lysine-preferring Mammalian Histone Methyltransferase with Hyperactivity and Specific Selectivity to Lysines 9 and 27 of Histone H3". In: *Journal of Biological Chemistry* 276.27, pp. 25309–25317.

Tachibana, M. et al. (2002). "G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis". In: *Genes and Development* 16.14, pp. 1779–1791.

Takahashi, Y. H., G. H. Westfield, A. N. Oleskie, R. C. Trievel, A. Shilatifard, and G. Skiniotis (2011). "Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.51, pp. 20526–20531.

Talasz, H., H. H. Lindner, B. Sarg, and W. Helliger (2005). "Histone H4-Lysine 20 Monomethylation Is Increased in Promoter and Coding Regions of Active Genes and Correlates with Hyperacetylation". In: *Journal of Biological Chemistry* 280.46, pp. 38814–38822.

Talbert, P. B. and S. Henikoff (2010). "Histone variants ancient wrap artists of the epigenome". In: *Nature Reviews Molecular Cell Biology* 11.4, pp. 264–275.

— (2017). "Histone variants on the move: Substrates for chromatin dynamics". In: *Nature Reviews Molecular Cell Biology* 18.2, pp. 115–126.

Tamkun, J. W., R. Deuring, M. P. Scott, M. Kissinger, A. M. Pattatucci, T. C. Kaufman, and J. A. Kennison (1992). "brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2 SWI2". In: *Cell* 68.3, pp. 561–572.

Tavares, L. et al. (2012). "RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3". In: *Cell* 148.4, pp. 664–678.

Taverna, S. D., H. Li, A. J. Ruthenburg, C. D. Allis, and D. J. Patel (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers". In: *Nature Structural & Molecular Biology* 14.11, pp. 1025–1040.

Tee, W.-W. and D. Reinberg (2014). "Chromatin features and the epigenetic regulation of pluripotency states in ESCs". In: *Development* 141.12, pp. 2376–2390.

Thapa, P., R. Y. Zhang, V. Menon, and J. P. Bingham (2014). "Native chemical ligation: A boon to peptide chemistry". In: *Molecules* 19.9, pp. 14461–14483.

Thomas, T., A. K. Voss, K. Chowdhury, and P. Gruss (2000). "Querkopf, a MYST family histone acetyltransferase, is required for normal cerebral cortex development." In: *Development (Cambridge, England)* 127.12, pp. 2537–48.

Thomson, J. P. et al. (2010). "CpG islands influence chromatin structure via the CpG-binding protein Cfp1". In: *Nature* 464.7291, pp. 1082–1086.

Timmers, H. T. M. and P. A. Sharp (1991). "The mammalian TFIID protein is present in two functionally distinct complexes". In: *Genes and Development* 5.11, pp. 1946–1956.

Torres, I. O., K. M. Kuchenbecker, C. I. Nnadi, R. J. Fletterick, M. J. S. Kelly, and D. G. Fujimori (2015). "Histone demethylase KDM5A is regulated by its reader domain through a positive-feedback mechanism". In: *Nature Communications* 6.1, p. 6204.

Tremblay, V., P. Zhang, C.-P. Chaturvedi, J. Thornton, J. S. Brunzelle, G. Skiniotis, A. Shilatifard, M. Brand, and J.-F. Couture (2014). "Molecular Basis for DPY-30 Association to COMPASS-like and NURF Complexes". In: *Structure* 22.12, pp. 1821–1830.

Tremethick, D. J. (2007). "Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber". In: *Cell* 128.4, pp. 651–654.

Trojer, P. and D. Reinberg (2007). "Facultative Heterochromatin: Is There a Distinctive Molecular Signature?" In: *Molecular Cell* 28.1, pp. 1–13.

Truax, A. D. and S. F. Greer (2012). "ChIP and Re-ChIP Assays: Investigating Interactions Between Regulatory Proteins, Histone Modifications, and the DNA Sequences to Which They Bind". In: *Nature Communications.* Vol. 7. 1, pp. 175–188.

Tyagi, M., N. Imam, K. Verma, and A. K. Patel (2016). "Chromatin remodelers: We are the drivers!!" In: *Nucleus* 7.4, pp. 388–404.

Tyanova, S., T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, and J. Cox (2016). "The Perseus computational platform for comprehensive analysis of (prote)omics data". In: *Nature Methods* 13.9, pp. 731–740.

Ullah, M. et al. (2008). "Molecular Architecture of Quartet MOZ/MORF Histone Acetyltransferase Complexes". In: *Molecular and Cellular Biology* 28.22, pp. 6828–6843.

Ullius, A. et al. (2014). "The interaction of MYC with the trithorax protein ASH2L promotes gene transcription by regulating H3K27 modification". In: *Nucleic Acids Research* 42.11, pp. 6901–6920.

Vakoc, C. R., M. M. Sachdeva, H. Wang, and G. A. Blobel (2006). "Profile of Histone Lysine Methylation across Transcribed Mammalian Chromatin". In: *Molecular and Cellular Biology* 26.24, pp. 9185–9195.

Valdes-Mora, F., J. Z. Song, A. L. Statham, D. Strbenac, M. D. Robinson, S. S. Nair, K. I. Patterson, D. J. Tremethick, C. Stirzaker, and S. J. Clark (2012). "Acetylation

of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer". In: *Genome Research* 22.2, pp. 307–321.

Van Etten, J., T. L. Schagat, J. Hrit, C. A. Weidmann, J. Brumbaugh, J. J. Coon, and A. C. Goldstrohm (2012). "Human pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs". In: *Journal of Biological Chemistry* 287.43, pp. 36370–36383.

Vandamme, J., P. Völkel, C. Rosnoblet, P. Le Faou, and P. O. Angrand (2011). "Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells". In: *Molecular and Cellular Proteomics* 10.4, pp. 1–23.

Vann, K. R. and T. G. Kutateladze (2018). "Architecture of PRC2 Holo Complexes". In: *Trends in Biochemical Sciences* xx, pp. 2–4.

Vaquero, A., A. Loyola, and D. Reinberg (2003). "The constantly changing face of chromatin." In: *Science of aging knowledge environment : SAGE KE* 2003.14, RE4.

Vastenhouw, N. L., Y. Zhang, I. G. Woods, F. Imam, A. Regev, X. S. Liu, J. Rinn, and A. F. Schier (2010). "Chromatin signature of embryonic pluripotency is established during genome activation". In: *Nature* 464.7290, pp. 922–926.

Vastenhouw, N. L. and A. F. Schier (2012). "Bivalent histone modifications in early embryogenesis". In: *Current Opinion in Cell Biology* 24.3, pp. 374–386.

Vermeulen, M., K. W. Mulder, S. Denissov, W. W. Pijnappel, F. M. van Schaik, R. a. Varier, M. P. Baltissen, H. G. Stunnenberg, M. Mann, and H. T. M. Timmers (2007). "Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4". In: *Cell* 131.1, pp. 58–69.

Vermeulen, M. et al. (2010). "Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers". In: *Cell* 142.6, pp. 967–980.

Vidal, M. (2009). "Role of polycomb proteins Ring1A and Ring1Bin the epigenetic regulation of gene expression". In: *The International Journal of Developmental Biology* 53.2-3, pp. 355–370.

Viiri, K. M., J. Janis, T. Siggers, T. Y. K. Heinonen, J. Valjakka, M. L. Bulyk, M. Maki, and O. Lohi (2009). "DNA-Binding and -Bending Activities of SAP30L and SAP30 Are Mediated by a Zinc-Dependent Module and Monophosphoinositides". In: *Molecular and Cellular Biology* 29.2, pp. 342–356.

Virdee, S., P. B. Kapadnis, T. Elliott, K. Lang, J. Madrzak, D. P. Nguyen, L. Riechmann, and J. W. Chin (2011). "Traceless and site-specific ubiquitination of recombinant proteins". In: *Journal of the American Chemical Society* 133.28, pp. 10708–10711.

Viré, E. et al. (2006). "The Polycomb group protein EZH2 directly controls DNA methylation". In: *Nature* 439.7078, pp. 871–874.

Vizán, P., M. Beringer, C. Ballaré, and L. Di Croce (2015). "Role of PRC2-associated factors in stem cells and disease". In: *FEBS Journal* 282.9, pp. 1723–1735.

Voigt, P., G. LeRoy, W. J. Drury, B. M. Zee, J. Son, D. B. Beck, N. L. Young, B. a. Garcia, and D. Reinberg (2012). "Asymmetrically modified nucleosomes". In: *Cell* 151.1, pp. 181–193.

Voss, A. K., C. Collin, M. P. Dixon, and T. Thomas (2009). "Moz and Retinoic Acid Coordinately Regulate H3K9 Acetylation, Hox Gene Expression, and Segment Identity". In: *Developmental Cell* 17.5, pp. 674–686.

Wachter, E., T. Quante, C. Merusi, A. Arczewska, F. Stewart, S. Webb, and A. Bird (2014). "Synthetic CpG islands reveal DNA sequence determinants of chromatin structure." In: *eLife* 3, pp. 1–16.

Walker, E., W. Y. Chang, J. Hunkapiller, G. Cagney, K. Garcha, J. Torchia, N. J. Krogan, J. F. Reiter, and W. L. Stanford (2010). "Polycomb-like 2 Associates with PRC2 and Regulates Transcriptional Networks during Mouse Embryonic Stem Cell Self-Renewal and Differentiation". In: *Cell Stem Cell* 6.2, pp. 153–166.

Wan, Y., R. A. Saleem, A. V. Ratushny, O. Roda, J. J. Smith, C.-H. Lin, J.-H. Chiang, and J. D. Aitchison (2009). "Role of the Histone Variant H2A.Z/Htz1p in TBP Recruitment, Chromatin Dynamics, and Regulated Expression of Oleate-Responsive Genes". In: *Molecular and Cellular Biology* 29.9, pp. 2346–2358.

Wang, C., Q. X. Guo, and Y. Fu (2011). "Theoretical analysis of the detailed mechanism of native chemical ligation reactions". In: *Chemistry - An Asian Journal* 6.5, pp. 1241–1251.

Wang, H., L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R. S. Jones, and Y. Zhang (2004a). "Role of histone H2A ubiquitination in Polycomb silencing". In: *Nature* 431.7010, pp. 873–878.

Wang, J., J. Mager, E. Schnedier, and T. Magnuson (2002). "The mouse PcG gene eed is required for Hox gene repression and extraembryonic development". In: *Mammalian Genome* 13.9, pp. 493–503.

Wang, L., J. L. Brown, R. Cao, Y. Zhang, J. A. Kassis, and R. S. Jones (2004b). "Hierarchical recruitment of polycomb group silencing complexes". In: *Molecular Cell* 14.5, pp. 637–646.

Wang, X., R. D. Paucek, A. R. Gooding, Z. Z. Brown, E. J. Ge, T. W. Muir, and T. R. Cech (2017). "Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA". In: *Nature Structural & Molecular Biology* 24.12, pp. 1028–1038.

Wani, A. H., A. N. Boettiger, P. Schorderet, A. Ergun, C. Munger, R. I. Sadreyev, X. Zhuang, R. E. Kingston, and N. J. Francis (2016). "Chromatin topology is coupled to Polycomb group protein subnuclear organization". In: *Nature Communications* 7.

Weber, C. M., S. Ramachandran, and S. Henikoff (2014). "Nucleosomes are context-specific, H2A.Z-Modulated barriers to RNA polymerase". In: *Molecular Cell* 53.5, pp. 819–830.

Weber, C. M. and S. Henikoff (2014). "Histone variants: dynamic punctuation in transcription". In: *Genes & Development* 28.7, pp. 672–682.

Wells, D. and D. Brown (1991). "Histone update and and alignment". In: *Gene* 19.1990, pp. 461–473.

Wen, H., J. Li, T. Song, M. Lu, P. Y. Kan, M. G. Lee, B. Sha, and X. Shi (2010). "Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation". In: *Journal of Biological Chemistry* 285.13, pp. 9322–9326.

Weng, N.-p., Y. Araki, and K. Subedi (2012). "The molecular basis of the memory T cell response: differential gene expression and its epigenetic regulation". In: *Nature Reviews Immunology* 12.4, pp. 306–315.

Widom, J. (2001). *Role of DNA sequence in nucleosome stability and dynamics.* Vol. 34. 3. University of Edinburgh, pp. 269–324.

Widom, J. (2006). "A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells." In: *Proceedings of the National Academy of Sciences* 89.3, pp. 1095–1099.

Wierer, M. and M. Mann (2016). "Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes". In: *Human Molecular Genetics* 25.R2, R106–R114.

Wiśniewski, J. R., A. Zougman, N. Nagaraj, and M. Mann (2009). "Universal sample preparation method for proteome analysis". In: *Nature Methods* 6.5, pp. 359–362.

Wong, M. M., L. K. Cox, and J. C. Chrivia (2007). "The chromatin remodeling protein, SRCAP, is critical for deposition of the histone variant H2A.Z at promoters". In: *Journal of Biological Chemistry* 282.36, pp. 26132–26139.

Wong, S. J. et al. (2016). "KDM2B Recruitment of the Polycomb Group Complex, PRC1.1, Requires Cooperation between PCGF1 and BCORL1". In: *Structure* 24.10, pp. 1795–1801.

Woo, C. J., P. V. Kharchenko, L. Daheron, P. J. Park, and R. E. Kingston (2010). "A Region of the Human HOXD Cluster that Confers Polycomb-Group Responsiveness". In: *Cell* 140.1, pp. 99–110.

Woo, H., S. Dam Ha, S. B. Lee, S. Buratowski, and T. Kim (2017). "Modulation of gene expression dynamics by co-transcriptional histone methylations". In: *Experimental & Molecular Medicine* 49.4, e326–e326.

Wu, H. et al. (2013). "Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations". In: *PLoS ONE* 8.12, pp. 1–12.

Wutz, A. (2011). "Gene silencing in X-chromosome inactivation: Advances in understanding facultative heterochromatin formation". In: *Nature Reviews Genetics* 12.8, pp. 542–553.

Xie, W. et al. (2013). "Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells". In: *Cell* 153.5, pp. 1134–1148.

Xu, C. et al. (2010). "Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2)". In: *Proceedings of the National Academy of Sciences* 107.45, pp. 19266–19271.

Xu, C., C. Bian, R. Lam, A. Dong, and J. Min (2011). "The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain". In: *Nature Communications* 2.1, pp. 227–228.

Xu, J., A. C. Carter, A.-V. Gendrel, M. Attia, J. Loftus, W. J. Greenleaf, R. Tibshirani, E. Heard, and H. Y. Chang (2017). "Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells". In: *Nature Genetics* 49.3, pp. 377–386.

Yan, Y., N. A. Barlev, R. H. Haley, S. L. Berger, and R. Marmorstein (2000). "Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases". In: *Molecular Cell* 6.5, pp. 1195–1205.

Yang, N., W. Wang, Y. Wang, M. Wang, Q. Zhao, Z. Rao, B. Zhu, and R.-M. Xu (2012). "Distinct mode of methylated lysine-4 of histone H3 recognition by tandem tudor-like domains of Spindlin1". In: *Proceedings of the National Academy of Sciences* 109.44, pp. 17954–17959.

Yang, W. and P. Ernst (2017). "Distinct functions of histone H3, lysine 4 methyltransferases in normal and malignant hematopoiesis". In: *Current Opinion in Hematology* 24.4, pp. 322–328.

Yang, X.-J. and M. Ullah (2007). "MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells". In: *Oncogene* 26.37, pp. 5408–5419.

Yang, X. J. (2004). "The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases". In: *Nucleic Acids Research* 32.3, pp. 959–976.

Yang, X. J. (2015). "MOZ and MORF acetyltransferases: Molecular interaction, animal development and human disease". In: *Biochimica et Biophysica Acta - Molecular Cell Research* 1853.8, pp. 1818–1826.

Yang, Z., C. Zheng, C. Thiriet, and J. J. Hayes (2005). "The Core Histone N-Terminal Tail Domains Negatively Regulate Binding of Transcription Factor IIIA to a Nucleosome Containing a 5S RNA Gene via a Novel Mechanism". In: *Molecular and Cellular Biology* 25.1, pp. 241–249.

Yang, Z., K. Shah, A. Khodadadi-Jamayran, and H. Jiang (2016). "Dpy30 is critical for maintaining the identity and function of adult hematopoietic stem cells". In: *Journal of Experimental Medicine* 213.11, pp. 2349–2364.

Yang, Z. et al. (2018). "EBS is a bivalent histone reader that regulates floral phase transition in Arabidopsis". In: *Nature Genetics* 50.9, pp. 1247–1253.

Yap, K. L. and M.-M. Zhou (2011). "Structure and Mechanisms of Lysine Methylation Recognition by the Chromodomain in Gene Transcription". In: *Biochemistry* 50.12, pp. 1966–1980.

Yokoyama, A. and M. L. Cleary (2008). "Menin Critically Links MLL Proteins with LEDGF on Cancer-Associated Target Genes". In: *Cancer Cell* 14.1, pp. 36–46.

Yoon, H.-G., Y. Choi, P. A. Cole, and J. Wong (2004). "Reading and Function of a Histone Code Involved in Targeting Corepressor Complexes for Repression". In: *Molecular and Cellular Biology* 25.1, pp. 324–335.

Yoshida, K. et al. (2011). "Frequent pathway mutations of splicing machinery in myelodysplasia". In: *Nature* 478.7367, pp. 64–69.

Young, N. L., P. A. DiMaggio, M. D. Plazas-Mayorca, R. C. Baliban, C. A. Floudas, and B. A. Garcia (2009). "High throughput characterization of combinatorial histone codes". In: *Molecular and Cellular Proteomics* 8.10, pp. 2266–2284.

Yu, C. et al. (2017). "CFP1 Regulates Histone H3K4 Trimethylation and Developmental Potential in Mouse Oocytes". In: *Cell Reports* 20.5, pp. 1161–1172.

Yuan, G. et al. (2013). "Histone H2A ubiquitination inhibits the enzymatic activity of H3 lysine 36 methyltransferases". In: *Journal of Biological Chemistry* 288.43, pp. 30832–30842.

Yuan, W. et al. (2012). "Dense chromatin activates polycomb repressive complex 2 to regulate H3 lysine 27 methylation". In: *Science* 337.6097, pp. 971–975.

Zentner, G. E. and S. Henikoff (2013). "Regulation of nucleosome dynamics by histone modifications". In: *Nature Structural and Molecular Biology* 20.3, pp. 259–266.

Zhang, H., D. N. Roberts, and B. R. Cairns (2005). "Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss". In: *Cell* 123.2, pp. 219–231.

Zhang, X., A. H. Smits, G. B. van Tilburg, P. W. Jansen, M. M. Makowski, H. Ovaa, and M. Vermeulen (2017). "An Interaction Landscape of Ubiquitin Signaling". In: *Molecular Cell* 65.5, 941–955.e8.

Zhang, X., A. H. Smits, G. B. Van Tilburg, H. Ovaa, W. Huber, and M. Vermeulen (2018). "Proteome-wide identification of ubiquitin interactions using UbIA-MS". In: *Nature Protocols* 13.3, pp. 530–550.

Zhang, X., H. Tamaru, S. I. Khan, J. R. Horton, L. J. Keefe, E. U. Selker, and X. Cheng (2002). "Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase". In: *Cell* 111.1, pp. 117–127.

Zhang, Y. et al. (2014). "The PHD1 finger of KDM5B recognizes unmodified H3K4 during the demethylation of histone H3K4me2/3 by KDM5B". In: *Protein and Cell* 5.11, pp. 837–850.

Zhang, Y., H. H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, and D. Reinberg (1999). "Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation". In: *Genes and Development* 13.15, pp. 1924–1935.

Zhao, J., T. K. Ohsumi, J. T. Kung, Y. Ogawa, D. J. Grau, K. Sarma, J. J. Song, R. E. Kingston, M. Borowsky, and J. T. Lee (2010). "Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq". In: *Molecular Cell* 40.6, pp. 939–953.

Zhao, X. D. et al. (2007). "Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells". In: *Cell Stem Cell* 1.3, pp. 286–298.

Zheng, C. and J. J. Hayes (2003a). "Intra- and Inter-nucleosomal Protein-DNA Interactions of the Core Histone Tail Domains in a Model System". In: *Journal of Biological Chemistry* 278.26, pp. 24217–24224.

— (2003b). "Structures and interactions of the core histone tail domains". In: *Biopolymers* 68.4, pp. 539–546.

Zhou, W., P. Zhu, J. Wang, G. Pascual, K. A. Ohgi, J. Lozach, C. K. Glass, and M. G. Rosenfeld (2008). "Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation". In: *Molecular Cell* 29.1, pp. 69–80.

Zhu, W., J. W. Smith, and C.-M. Huang (2010). "Mass Spectrometry-Based Label-Free Quantitative Proteomics". In: *Journal of Biomedicine and Biotechnology* 2010, pp. 1–6.

Zilberman, D., D. Coleman-Derr, T. Ballinger, and S. Henikoff (2008). "Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks". In: *Nature* 456.7218, pp. 125–129.

Zipper, H., H. Brunner, J. Bernhagen, and F. Vitzthum (2004). "Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications." In: *Nucleic acids research* 32.12, e103.

216