KDM2B links recognition of CpG islands to polycomb domain formation *in vivo*

Anca Madalina Farcas

St John's College

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Biochemistry

University of Oxford

MT 2013

Declaration of authorship

I declare that this thesis was fully composed by me and that the research presented here is my own work unless otherwise stated. I have clearly indicated the presence of material I have quoted from other sources. This thesis has not been submitted, either partially or in full, for another degree, diploma, certificate or other qualification at this University or at any other institution.

Anca Madalina Farcas

MT 2013

KDM2B links recognition of CpG islands to polycomb domain formation *in vivo*

Anca Madalina Farcas
St John's College
Doctor of Philosophy (DPhil) in Biochemistry
University of Oxford
MT 2013

Abstract

Mammalian genomes are characterised by global and pervasive DNA methylation and this modification is generally thought to be inhibitory to transcription. An exception to this widespread DNA modification are genomic elements called CpG islands (CGI), contiguous regions of non-methylated DNA which encompass the transcription start site of two thirds of mammalian genes. Although CGIs represent the most prominent feature of mammalian promoters, the contribution of these elements to promoter function remains unclear. Work in this study shows that the histone lysine demethylase KDM2B (FBXL10/ JHDM1B) is a nuclear protein which binds specifically to non-methylated CpG dinucleotides and associates with CGI elements genome-wide through its zinc-finger CxxC (ZF-CxxC) DNA binding domain. Furthermore, in mouse embryonic stem cells, biochemical investigation revealed that KDM2B associates with Polycomb group E3 ubiquitin ligase RING1B to form a variant Polycomb repressive complex 1 (PRC1) characterized by the PCGF1 subunit.

Considering that KDM2B has clear DNA-binding activity and that CGIs were reported to function as nucleation sites for polycomb repressive complexes, a potential role for KDM2B in mediating PRC1 recruitment to target genes was investigated. Stable depletion studies indicated that KDM2B is required for the normal targeting of RING1B to CGIs and the regulation of expression of a subset of Polycomb-occupied genes. By taking advantage of a genetic ablation system in which the DNA binding domain of KDM2B can be conditionally deleted, results in this thesis reveal that the ability of KDM2B to recognize

KDM2B links recognition of CpG islands to polycomb domain formation *in vivo*

Anca Madalina Farcas
St John's College
Doctor of Philosophy (DPhil) in Biochemistry
University of Oxford
MT 2013

non-methylated DNA is essential for polycomb domain formation and normal embryonic development. Finally, through the use of a *de novo* targeting assay, an unexpected PRC2 recruitment pathway was discovered which is dependent on PRC1-mediated H2AK119ub1 deposition. Together this work uncovers a novel mechanism linking KDM2B-dependent recognition of non-methylated DNA with recruitment of Polycomb proteins and provides the framework on which to further investigate the contribution of CGIs to formation of polycomb domains.

List of publications

1. N. P. Blackledge, J. C. Zhou, M. Y. Tolstorukov, **A. M. Farcas**, P. J. Park and R. J. Klose (2010): CpG Islands Recruit a Histone H3 Lysine 36 Demethylase, ***Molecular Cell***, 38, 179–190.

2. J.C. Zhou, N.P. Blackledge, **A.M. Farcas** and R.J. Klose (2011): Recognition of CpG island chromatin by KDM2A requires direct and specific interaction with linker DNA, ***Molecular and Cellular Biology***, 32(2): 479–489.

3. **A.M. Farcas**, N.P. Blackledge, I. Sudbery, H.K. Long, J.F. McGouran, N.R. Rose,S. Lee, D. Sims, A. Cerase, T.W. Sheahan, H. Koseki, N. Brockdorff, C.P. Ponting, B.M. Kessler and R.J. Klose (2012): KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands, ***eLIFE***, 1:e00205.

4. R.J. Klose, S. Cooper, **A.M. Farcas**, N.P. Blackledge and N. Brockdorff (2013): Chromatin Sampling – An Emerging Perspective on Targeting Polycomb Repressor Proteins, ***PLoS Genetics***, 9(8): e1003717.

5. N.P. Blackledge*, **A.M. Farcas***, T. Kondo*, H. King, J.F. McGouran, L. Hanssen, S. Cooper, K. Kondo, Y. Koseki, T. Ishikura, H.K. Long, T.W. Sheahan, N. Brockdorff, B.M. Kessler, H. Koseki and R.J. Klose (2014): Variant PRC1 complex dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation, ***Cell***, Jun 5;157(6):1445-59. (* these authors contributed equally)

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Rob Klose, for his constant support, patience as I always tend to disagree, experimental advice, guidance, motivation, mentoring and challenge. Simply put, neither this thesis nor the work described in this study would have been possible without him. I was passionate about chromatin biology when I started this DPhil. Now, 4 years later, I am even more inspired and dedicated through your example to find my own way in the field and continue to ask exciting scientific questions.

I would like to extend this gratitude to Dr. Neil Blackledge, for…well….everything really, starting with my introduction to ChIP (am not sure how grateful I should be for that!) to his constant support, experimental advice, his countless ChIP optimisations which allowed me to then carry out my work with ease. I can't believe you actually put up with me and my impressive set of random scientific questions for so long. Your presence in the lab and constant willingness to listen to me kept me sane in times of doubt and when I was ready to give up under the pressure of competition and elusive ChIP-qPCR results. I enjoyed sharing a mutual scientific goal for this last year and I hope to find someone like you in the groups I join next, although I am afraid you might just be a Klose lab blessing.

I would like to thank all members of the Klose lab (past and present) for their daily guidance and support. Jin, Shirley, Dave, Hannah, Nathan, Hamish, Emilia, you guys are amazing and you made and continue to make the Klose lab such a cool place to be in! Special thanks to Hannah for her fantastic bioinformatics work and contribution to the KDM2B story. And thank you to Jin for staying with me for a lot of the initial FPLC injections, when I was scared of destroying the columns (after Rob warned me that they're expensive)! Dave…you're a good friend, it's your year man! Nathan, thank you for all the last-minute thesis proofreading! Thomas, you joining our lab as a summer student and Part II project made such an incredible change in my life! You're so stubborn and determined and I am not sure if I can ever thank you enough for everything you've done and continue to do for me! I hope things turn out the way you want them to!

I would like to extend my thanks to Prof. Neil Brockdorff and members of the Brockdorff lab for stimulating scientific discussions. Deepest gratitude to Tatyana for making our joint lab space an enjoyable spotless working environment, Nick for support and Ligia, for advice and continual help. Thank you to Andrea, who has always been by my side when I needed unconditional friendship and wise advice. Am also grateful to Anna for her support, constant supply of German goodies and her calm attitude to life which she attempted in vain to impart on me.

Furthermore, I acknowledge scientific support from our multiple collaborators: Dr. Joanna McGouran and Dr. Benedikt Kessler for MS work, Dr. Lorna Gregory and the WTCHG for ChIP-Seq analysis, Drs. Ian Sudberry and David Sims from the group of Dr. Chris Ponting for bioinformatic help on the *eLIFE* KDM2B paper, Dr. Haruhiko Koseki and members of his lab for essential reagents and insightful experimental suggestions. I also would like to thank the Cancer Research Center UK (CRUK) for financially supporting my DPhil.

Outside of the lab (if such a thing exists!), I am grateful to Maeve for being my friend for over 4 years now and I hope to stay in touch once our time at St John's comes to an end. Thank you to Dragos for support, friendship, for installing Linux on all my computers (regardless of the fact that I still don't know how to use it) and for being there whenever I panicked. Hidden in here (hoping nobody reads this section)…I want to thank myself! I had such a rough start into this DPhil, I can't believe I've made it (well, almost) and that it was such an enjoyable and successful experience! I gave it my best, and I can't imagine having spent this time anywhere else but in this lab! Let this be a lesson….

Finally, I would like to thank my parents, who constantly encouraged me in my academic endeavours and were more than understanding and forgiving of my long absences from home. Vă mulţumesc din inimă pentru tot ceea ce faceţi pentru mine şi Bogdan! Nu aş fi ajuns atât de departe fără sprijinul şi iubirea voastră! Sper să vă pot face mândri de mine toată viaţa! Vă iubesc, tineretul muncitor, pupici.

Table of contents

List of figures

List of tables

List of abbreviations

aa	Amino acid
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
bp	Base pairs
Bio-CAP	Biotinylated CXXC Affinity Purification
BSA	Bovine serum albumin
°C	Degree centigrade
C-	Carboxyl
CFP1	CxxC finger protein 1
CGI	CpG Island
ChIP	Chromatin immunoprecipitation
ChIP-qPCR	ChIP followed by quantitative PCR
ChIP-Seq	ChIP followed by massively parallel sequencing
COMPASS	Complex of proteins associated with SET1
Da	Dalton
DAPI	4,6-diamino-2-phenylindole
dNTP	Deoxyribonucleoside triphosphate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNMT	DNA methyltransferase
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGS	Ethylene glycol bis[succinimidylsuccinate]
EMSA	Electrophoretic mobility shift assay
ES cell/ESC	Embryonic stem cell
FS2	Flag/2xStrepII tag
g	Relative centrifugal force
GFP	Green fluorescent protein
GO	Gene ontology
h	hour

HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEK 293T	Human embryonic kidney 293T
HEPES	N-2-hydroxyethylpeperazine-N'-2-ethanesulfonic acid
HKDM	Histone lysine demethylase
HKMT/HKMTase	Histone lysine methyltransferase
H2AK119ub1	Histone H2A lysine 119 monoubiquitylation
H3K4me1/2/3	Histone H3 lysine 4 mono/di/trimethylation
H3K27me3	Histone H3 lysine 27 trimethylation
H3K27ac	Histone H3 lysine 27 acetylation
H3K36me2	Histone H3 lysine 36 dimethylation
IF	Immunofluorescence
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	kilobase pairs
kDa	kilodalton
KDM2A	Lysine-specific demethylase 2A
KDM2B	Lysine-specific demethylase 2B
L	liter
LIF	Leukaemia Inhibitory Factor
M	molar
MBD	Methyl CpG binding domain
5mC/5hmC	5-methylcytosine/5-hydroxymethylcytosine
mg	milligram
min	minutes
MLL	Mixed lineage leukaemia
mol	mole
N-	amino-
NCE	Nuclear cell extract
ng	nanogram
nm	nanometre
nt	nucleotide
NuRD	Nucleosome remodelling and histone deacteylase complex
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction

PcG	Polycomb group
pH	-log10[H+]
PHD domain	Plant homeodomain
PMSF	phenylmethylsulfonyl fluoride
Pol II	RNA polymerase II
PRC1/PRC2	Polycomb Repressive Complex 1/ 2
PTM	post-translational modification
qPCR	quantitative real-time Polymerase Chain Reaction
RNA	ribonucleic acid
RNA-Seq	Whole-transcriptome sequencing
RNase	Ribonuclease
rpm	revolutions per minute
RT-PCR	Reverse transcriptase PCR
RT	room temperature
SAM	S-adenosyl-L-methionine
SDS	sodium dodecyl sulphate
SEM	standard error of the mean
TAE	Tris-acetate-EDTA
TCA	Trichloroacetic acid
TEMED	Tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
TrxG	Trithorax Group
TSS	Transcription start site
µg	microgram
µl	microliter
UTR	untranslated region
V	volts
WCE	Whole-cell extract
WT	Wild-type
ZF-CxxC domain	Zinc finger CxxC DNA binding domain
Xist	X inactive specific transcript

1. Chapter one - Introduction

1.1. Chromatin and epigenetics – general introduction

The term “chromatin” was coined by German cytogenetics pioneer Walther Flemming in his studies of cellular division to describe the fibrous network within the nucleus that is strongly stained by basic aniline dyes (Flemming 1879). Although Flemming had identified what was later defined as chromosomes, the term chromatin was employed to define the complex between DNA and associated proteins, of which the basic repeating unit is represented by nucleosomes (Kornberg 1974). Chromatin is usually divided into two classes, euchromatin and heterochromatin, originally distinguished by their distinct staining properties reflective of their differing degrees of chromatin condensation (Heitz 1928; Trojer and Reinberg 2007). While decondensed euchromatin is usually associated with gene rich and transcriptionally active regions of the genome, condensed constitutive heterochromatin is primarily composed of repetitive and non-coding sequences with limited transcriptional activity. Interestingly, there are also non-repetitive regions of facultative heterochromatin, which have the potential to revert to euchromatin despite their stable gene repression status, the best example being X-chromosome inactivation during mammalian development (Chow and Brown 2003; Arthold et al. 2011; Wutz 2011). These observations suggest that chromatin is a dynamic structure, which is able to undergo continuous structural transitions in response to changes in gene activity (Ausio 1992).

The term epigenetics was originally used in a developmental biology sense to define the “causal mechanisms by which the genes of the genotype bring about phenotypic effects” (Waddington 1942; Jablonka and Lamb 2002; Haig 2004), to distinguish it from preformationism ideas in neoclassical embryology. In later times epigenetics was redefined as “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” (Holliday 1994; Russo 1996). Nowadays, the field of epigenetics is primarily associated with the study of how chromatin and chromatin-related processes contribute to gene transcription regulation, by “registering,

signalling or perpetuating altered" activity states (Bird 2007). Importantly, this definition places epigenetic marks as "responsive" to a change already imposed by another event, and not as "proactive" in initiating the gene expression status. There remains, however, intense debate in the field regarding the degree of heritability or self-perpetuating nature of what is currently understood by epigenetic phenomena (Ptashne 2013).

Modifications to the chromatin template and their subsequent impact on gene expression are believed to occur in response to the following main pathways: modifications to the DNA template, modifications to canonical and variant histone proteins, the activity of ATP-dependent chromatin remodellers, RNA-interference and RNA-directed gene silencing events. Together, these processes control the assembly and regulation of the chromatin structure and contribute to genome integrity. The interplay between DNA methylation and specific histone-tail modifications represents the main focus of this study and will be discussed in greater detail throughout this thesis.

1.2. The nucleosome core particle and higher-order chromatin organization in the nucleus

Eukaryotic genomic DNA does not exist in free random coils in the nucleus, but is instead wrapped around histone proteins and folded into higher-order chromatin structures. This arrangement not only allows compaction of genomic DNA for packaging into the cell nucleus, but it also influences all processes which use DNA as a template, such as transcription, DNA repair or recombination. The nucleosome is the fundamental repeating unit of chromatin and is composed of 145–147 base pairs of DNA wrapped around a histone octamer core comprising two copies of each histone protein H2A, H2B, H3 and H4 (Kornberg 1974; Luger et al. 1997).

Nucleosomes are separated by a linker DNA region of variable length depending on species and tissue type (ranging between ~20–90 bp) (Szerlong and Hansen 2011). Linker histone protein H1 is generally located on the surface of the nucleosome, simultaneously binding to the nucleosome-linking DNA strands at the entry and exit points from the nucleosome core particle (Zhou et al. 1998; Izzo et al.

2008). The binding of histone H1 is thought to promote and stabilize folding of the nucleosomal array from the "beads-on-a-string" 11nm-diameter chromatin organizational unit into higher-order more compact 30nm-chromatin fibres. However, the *in vivo* relevance of the 30 nm structures observed *in vitro* has been disputed, with recent findings instead suggesting that mitotic chromosomes consist of irregularly arranged nucleosome fibres, which would permit a more dynamic and flexible genome organization, although principles governing such a folding remain insufficiently characterized (Maeshima et al. 2010; Nishino et al. 2012).

Importantly, by using *in vitro*-nucleosome assembly coupled with cell-free transcription systems, nucleosomes were shown to be inhibitory to basal transcription initiation (Knezetic and Luse 1986; Lorch et al. 1987; Kornberg and Lorch 1999). This early observation already indicated that there must be processes at regulatory elements which expose promoters or other essential regions of DNA to allow transcription, suggesting that the factors which can influence chromatin compaction and DNA accessibility could have an important contribution to the regulation of gene expression.

1.3. Post-translational histone modifications

The highly basic and evolutionarily conserved core histone proteins have a similar domain architecture, which comprises of a C-terminal structural motif known as the histone fold, and an extended N-terminal flexible unstructured tail which protrudes from the surface of the nucleosome (Luger et al. 1997). While the histone fold is essential for core octamer assembly, the exposed histone tails are subject to a variety of post-translational covalent modifications (PTMs), whose location and type can affect chromatin structure and influence transcriptional outcome. Furthermore, recent work indicates that, comparable to the flexible histone tails, the more structured globular domain of the octamer core particle can also be extensively modified, adding to the complexity of interpreting the functional outcomes of certain PTMs (Cosgrove and Wolberger 2005; Tropberger and Schneider 2010).

Some histone modifications can have *in-cis* effects by directly altering the physical properties of the nucleosome (i.e acetylation or phosphorylation), while others act *in-trans* through the recruitment of chromatin-modifying factors which can specifically "read" that histone mark (i.e. methylation) to induce further chromatin modifications and mediate the optimal functional consequences. Interestingly, the multitude of potential histone PTMs together with the proteins which can interpret them raised the question of whether PTMs act in isolation or as part of a more complex "combinatorial code" in which combinations of different modifications can specify alternative chromatin states (Jenuwein and Allis 2001; Dion et al. 2005; Wang et al. 2008). Although extensive ChIP-Seq analysis of histone modifications revealed some degree of correlation between specific histone marks at active or repressed promoters (Roh et al. 2006; Wang et al. 2008), the universality of a "histone code" is unclear and the question of whether histone modifications at promoter regions are a cause or consequence of transcriptional status remains debatable.

1.3.1. Main classes of histone tail modifications and histone-modifying enzymes

Following the discovery of lysine methylation on calf thymus histones (Murray 1964), the list of histone tail modifications has expanded to now include acetylation, phosphorylation, ubiquitylation, ADP-ribosylation (Strahl and Allis 2000), sumoylation (Shiio and Eisenman 2003), biotinylation (Camporeale et al. 2004), glycosylation (Zhang et al. 2011) and others. Interestingly, most if not all histone post-translational modifications are reversible, suggesting that the main function of these marks is not to lock a chromatin state but to enable a dynamic fine-tuning mechanism for regulating chromatin structure in response to various cellular stimuli, thereby adding another degree of plasticity to chromosomal processes.

Importantly, some histone modifications are proposed to be heritable through cell division, primarily by a semiconservative mechanism in which parental histones combine with newly synthesized histones within the same core nucleosome, thereby allowing transmission of the mark from the old to the new

histones of the daughter cell (Ray-Gallet and Almouzni 2010; Xu et al. 2010). However, contradicting evidence exists which shows absence of mixing between old or new H3-H4 dimers, thereby precluding transmission of histone marks based on pre-existing modifications within the same nucleosome but instead suggesting that propagation can occur between adjacent or nearby nucleosomes via chromatin-binding proteins specific for that mark (Katan-Khaykovich and Struhl 2011). Interestingly, recent evidence shows that it may not be the histone marks *per se* which are maintained through cellular division, but that the histone modifying enzymes persist through replication and re-establish the histone code on newly assembled histones (Petruk et al. 2012). Although this study was limited to only two histone methyltransferases and their respective marks, it indicates that more work is required to understand the degree to which histone PTMs can be considered epigenetic (i.e. heritable) modifications.

The functional relevance of histone acetylation, methylation and ubiquitylation in relation to gene transcription will be further discussed in greater detail.

1.3.1.1. Histone lysine acetylation

The first histone acetyltransferase (HAT) enzyme was isolated from macronuclei of the ciliate *Tetrahymena* by using an in-gel acetyltransferase assay based on the incorporation of [^{3}H]acetate from [^{3}H]acetyl-CoA into core histone substrates (Brownell and Allis 1995). Since this discovery, it became apparent that the HAT identified was homologous to the yeast transcriptional activator GCN5, which was speculated to function by facilitating the access of transcriptional activators to their DNA binding sites by an uncharacterized mechanism (Georgakopoulos and Thireos 1992). Three main HAT families have been currently defined: GNAT (Gcn5-related acetyltransferase), MYST and CBP/p300, which function by catalysing the transfer of an acetyl group from acetyl-coenzyme A (acetyl-CoA) to the ε-amino group of lysine side chains (Sterner and Berger 2000). Importantly, although HAT proteins can acetylate lysine residues on all four core histones, it is essential to consider that HATs also acetylate

prominent non-histone proteins (i.e p53, NFkB, GATA transcription factors), suggesting that acetylation plays an essential role in regulation of a vast array of cellular processes.

Histone lysine acetylation is associated with an open chromatin architecture (Vettese-Dadey et al. 1996; Bernstein et al. 2005; Ramsey et al. 2010) which was proposed to allow increased accessibility of transcription factors to their DNA binding sequence elements, thereby promoting transcriptional activation. The role of histone acetylation in promoting a transcriptionally-permissive chromatin state was suggested to be two-fold. Firstly, histone acetylation results in neutralization of the positively charged lysine residue, which reduces the strength of binding between histones and the negatively charged DNA molecule (Figure 1.1). Along the same lines, there is also evidence to suggest that H4K16 acetylation is sufficient to impair the formation of compact 30nm–like fibres and disrupts higher-order chromatin structures (Shogren-Knaak et al. 2006), although the functional mechanism by which this is achieved and the *in vivo* relevance remain poorly defined.

Secondly, histone acetylation can be recognized by proteins containing a specific histone lysine-binding domain known as the bromodomain (Dhalluin et al. 1999; Dyson et al. 2001), thereby allowing recruitment of factors to acetylated sites. Interestingly, this bromodomain is found in many chromatin-associated proteins, including ATP-dependent remodelling complexes SWI/SNF (mating type switching/ sucrose non fermenting) and RSC (remodels structure of chromatin) (Hassan et al. 2002; Rando and Winston 2012). It is therefore believed that recruitment of these factors via their bromodomains can alter nucleosome structure and facilitate transcription factor binding.

Histone acetylation is removed by deacetylases (HDACs), many of which are found within large multisubunit complexes containing co-regulators and other chromatin-modifying activities required for transcriptional repression (Kadosh and Struhl 1998; Kurdistani and Grunstein 2003). Particularly relevant for this thesis is the acetylation of H3K27 by CBP/p300 HATs (Tie et al. 2009; Pasini et al. 2010b). This modification was proposed to be required for maintenance of a transcriptionally active state by antagonizing repressive Polycomb-mediated H3K27 trimethylation. Importantly, deacetylation of H3K27 by the nucleosome remodelling and deacetylation complex (NuRD) facilitates Polycomb recruitment and subsequent H3K27me3 (Reynolds et al. 2012), suggesting that at some histone sites there is a

competition between multiple modifications, the balance of which could provide a new level of complexity in transcriptional regulation.

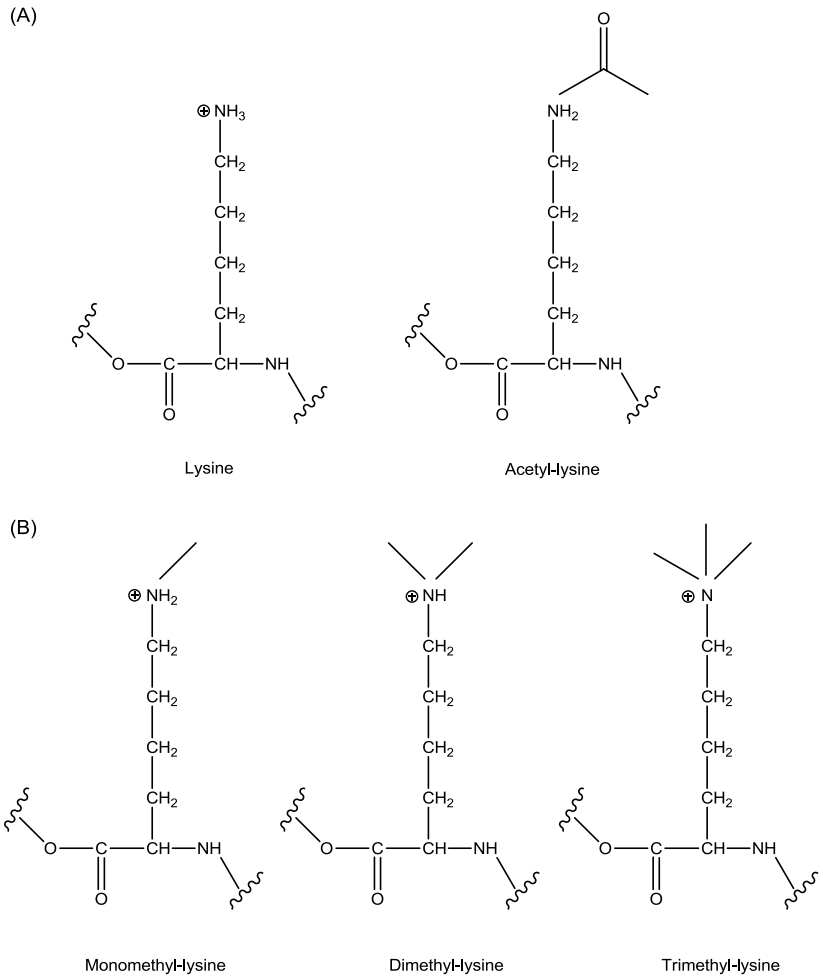


Figure 1.1. Common histone tail lysine modifications.

(A) Lysine acetylation neutralizes the positive charge on the NH3(+) form of the ε-amino group.

(B) Lysine can be mono-, di- or tri-methylated, with methylation retaining the positive charge.

1.3.1.2. Histone lysine methylation

Histones can be methylated on all basic residues (lysine, arginine and histidine), with lysine methylation representing the most studied modification (Greer and Shi 2012). Lysines can be mono- (me1), di- (me2) or tri- (me3) methylated on their ε-amine group mainly by SET (Su (var) 3-9–Enhancer of zeste–Trithorax)-domain containing histone lysine methyltransferases (HKMTs) using S-adenosylmethionine (SAM) as a cofactor (Rea et al. 2000; Dillon et al. 2005). As lysine methylation does not neutralize the

charge of the modified residue (Figure 1.1), it is believed to primarily function as a docking site for recruiting or blocking specific effector proteins which interpret the mark and regulate chromatin function (Taverna et al. 2007). Consistent with this mode of action, a multitude of domains recognizing methylated lysines has been identified, such as PHD, chromo, WD40, Tudor, double/tandem Tudor, MBT, ankyrin repeats, zinc finger-CW or PWWP domains (Lachner et al. 2001; Wysocka et al. 2006; Yun et al. 2011; Musselman et al. 2012). Interestingly, the same histone lysine mark can be recognized by different domains, suggesting that the potential impact of lysine methylation on gene expression has to be context-dependent.

Histone methylation was proposed to have lower turnover rate compared to other modifications (Byvoet et al. 1972) and for a long time it was considered to be an irreversible mark. However, the discovery that LSD1 (lysine-specific demethylase 1, also known as KDM1A) can specifically demethylate H3K4me1 and H3K4me2 via an FAD-dependent oxidation mechanism indicated that at least histone lysine mono and di-methylation was readily reversible (Shi et al. 2004). The subsequent identification of the Jumonji C (JmjC)-catalytic domain containing lysine demethylases (KDMs) provided a mechanism by which all three methylation states could be specifically removed. JmjC-domain enzymes catalyse lysine demethylation through an oxidative reaction that requires iron Fe(II) and alpha-ketoglutarate (alpha-KG) as cofactors and releases formaldehyde in the reaction (Klose et al. 2006; Tsukada et al. 2006). More than 30 human proteins were shown to possess JmjC domains, with 17 of these demonstrated to function as demethylases with activity towards H3K4, H3K9, H3K27, H3K36 and H4K20 (Kooistra and Helin 2012; Hojfeldt et al. 2013). Importantly, misregulation of histone lysine methylation or demethylation can have dramatic consequences for biological processes frequently resulting in tumour development, underlining the important role of these enzymes in regulating transcription and cell fate.

Of the many known methylated sites, four have been particularly well-studied: H3K4me3 and H3K36me2/me3, which have been linked to transcriptionally active chromatin states, and H3K9me3 and H3K27me3, which are generally associated with repression. Modifications to H3K4, H3K27 and H3K36 are particularly relevant for this thesis and are discussed in detail below.

1.3.1.2.1. H3K4me3

H3K4 methylation was first associated with *Tetrahymena* transcriptionally active macronuclei, where it was proposed to play a "facilitatory role in the transcription process" by an uncharacterized mechanism (Strahl et al. 1999). The sole H3K4 HKMTase in budding yeast is SET1, an enzyme which is capable of catalysing all three K4 methylation states (Roguev et al. 2001; Nagy et al. 2002). SET1 was purified as part of a complex termed COMPASS (Complex Proteins Associated with SET1), consisting of seven other polypeptides, including SPP1, the yeast homologue of the mammalian zinc-finger CxxC DNA-binding domain containing protein CFP1 (Miller et al. 2001; Roguev et al. 2001). Interestingly, *Set1*-null mutant *Saccharomyces cerevisiae* cells are viable, although they exhibit a slower growth phenotype (Miller et al. 2001), indicating that in yeast SET1 is not essential. However, it was reported that *Set1*- strains resulted in a relatively modest decrease in transcription of approximately 80% of genes as determined by microarray analysis (Boa et al. 2003), but were also defective in telomere, mating cassette loci HML/HMR and rDNA silencing (Briggs et al. 2001; Krogan et al. 2002; Fingerman et al. 2005). These results suggest that, at least in budding yeast, H3K4 methylation may play a more general role in transcriptional regulation.

High-resolution measurement of histone modifications using tiled microarrays showed that in yeast, most expressed open reading frames exhibit a striking pattern of H3K4 methylation, with H3K4me3 enriched at promoter regions, shifting to H3K4me2 within the 5' to middle region of genes, while H3K4me1 peaked towards the 3' region of longer genes (Santos-Rosa et al. 2002; Liu et al. 2005). This correlation between H3K4 tri-methylation and transcriptionally active promoters is consistent with COMPASS recruitment to chromatin via its association with the initiating serine-5 phosphorylated form of RNA polymerase II (RNAPII) and its interaction with the PAF1 complex, which associates with the initiating and early elongating RNAPII and can directly recruit SET1 to the transcriptional machinery (Hampsey and Reinberg 2003; Krogan et al. 2003a). Interestingly, this specific enrichment of H3K4me3 at promoters of active genes was not hypothesized to drive transcription, but was speculated to act

more as a “molecular memory” within the mRNA coding region for recent transcriptional activity (Ng et al. 2003).

A similar complex termed dSET 1 was found to be the main H3K4 trimethyltransferase in flies (Ardehali et al. 2011; Hallson et al. 2012), while other H3K4 methyltransferases like Trithorax (TRX) and Trithorax-related (TRR) were suggested to have more specific roles. TRX is believed to antagonize Polycomb-mediated silencing and act as positive regulator of *Hox* gene function (Klymenko and Muller 2004), while TRR co-localizes extensively with the ecdysone receptor (EcR) and seems to be primarily required to mediate H3K4 methylation at gene targets responsive to this insect nuclear hormone (Sedkov et al. 2003).

The situation is even more complicated in mammals, where at least ten known or predicted H3K4 methyltransferases exist (reviewed in (Ruthenburg et al. 2007)), with the most studied being the six COMPASS-like complexes SET1A/SET1B (related to dSET1), MLL1/MLL2 (related to TRX) and MLL3/MLL4 (related to TRR). Interestingly, SET1A/SET1B (via CFP1) and MLL1/MLL2 complexes contain ZF-CxxC DNA binding domains and have the capacity to specifically recognize non-methylated CpGs. Consistent with this observation, mammalian H3K4me3 is almost exclusively found at CpG-island promoters regardless of transcriptional status, although higher levels of H3K4me3 are generally observed at active genes (Guenther et al. 2007). Embryonic stem cells with a targeted disruption of *Cfp1* gene show that, despite a widespread association between CFP1 and CGIs (Thomson et al. 2010), only a fraction of H3K4me3-positive promoters are significantly affected by the absence of CFP1 and the most dramatic loss of H3K4me3 was observed at highly expressed genes (Clouaire et al. 2012), suggesting that CFP1 is not essential for basal H3K4me3 at CGIs but is important for transcriptional-dependent deposition of K4me3.

In contrast, MLL1 localizes to only a subset of CpG island promoters in human lymphoma cells, highly coincident with RNAPII and active transcription (Milne et al. 2005). This is consistent with reports using *Mll1*$^{-/-}$ mouse embryonic fibroblasts (MEFs), which show MLL1 is required for the H3K4me3 of less than 5% of promoters carrying this modification, but that loss of MLL1 leads to decreased levels of RNA polymerase II recruitment and expression at many of these genes, which include developmental

regulators such as *Hox* genes (Wang et al. 2009). Similar to MLL1, MLL2 was also shown to be essential for embryonic development (Yu et al. 1995; Glaser et al. 2009), although it was mostly believed to function in gametogenesis and also briefly in the zygote as a maternally derived factor (Andreu-Vieyra et al. 2010). Recent work however suggests that MLL2 is required to specifically trimethylate H3K4 at developmentally regulated promoters in mouse embryonic stem cells which are also marked by H3K27me3 (so-called bivalent promoters) (Hu et al. 2013b), although the significance of this discovery remains to be clarified. MLL3/MLL4 were found to be primarily responsible for placing H3K4me1 at enhancers (Hu et al. 2013a), suggesting that despite sharing a similar SET-catalytic domain, the H3K4 methyltransferases are not redundant.

H3K4 methylation is thought to affect gene expression by recruiting other effector proteins, including several chromatin remodelling and modifying complexes, the best studied examples being H3K4me3-dependent recruitment of the NURF nucleosome remodelling complex (Wysocka et al. 2006) and recruitment of the basal transcription factor TFIID via the PHD domain of TAF3 subunit (Vermeulen et al. 2007). Furthermore, H3K4me3 can block binding of other factors to chromatin. The interaction between DNMT3L, an enzymatically inactive partner of the *de novo* DNA methyltransferases DNMT3A and DNMT3B, and the H3 tail is inhibited by H3K4me3 (Ooi et al. 2007) and H3K4me3 was also shown to be inhibitory to binding of the repressive NuRD (nucleosome remodelling and deacetylase) complex (Musselman et al. 2009). In addition, H3K4me3 inhibits H3-binding of the Nurf55-Su(z)12 submodule of the PRC2 repressive complex and subsequent deposition of H3K27me3 *in vitro* (Schmitges et al. 2011), suggesting that the contribution of H3K4me3 to generating an accessible chromatin architecture at promoters is multilateral .

H3K4 methylation can be reversed by members of the JARID1 family (K4me3) (Klose et al. 2006; Christensen et al. 2007; Klose et al. 2007b), whilst LSD1 can demethylate H3K4me2 and H3K4me1 (Shi et al. 2004). Interestingly, the JARID1 Drosophila homolog LID (Little Imaginal Discs) was actually classified as a trithorax-group gene (Eissenberg et al. 2007; Lee et al. 2007c), and *Lid* fly mutants lead to loss of homeotic gene activity, further complicating the understanding of the extent to which H3K4me3 is required or causative for transcriptional activation.

1.3.1.2.2. H3K27me3

H3K27me3 is generally considered a repressive histone modification found at polycomb target promoters and the inactive X-chromosome in mammals or at polycomb-response elements (PREs) in the case of Drosophila. Additionally, in *Suv39h* H3K9 methyltransferase double-null cells, H3K27me3 can also be found enriched at pericentric heterochromatic regions, indicative of potential plasticity between the H3K9 and H3K27 methylation systems (Peters et al. 2003). H3K27 trimethylation is catalysed by the SET-domain containing Enhancer of Zeste (EZ) (in flies)/Enhancer of Zeste Homologue 1/2 (EZH1/2) (in mammals) subunit of the Polycomb repressive complex 2 (PRC2) (Cao et al. 2002; Muller et al. 2002). Importantly, the SET-subunit is insufficient for H3K27 trimethylation on its own and requires association with at least two other core PRC2 proteins, Suppressor of Zeste 12 (SUZ12) and Extra Sex Combs (ESC) for enzymatic activity (Cao and Zhang 2004b; Pasini et al. 2004).

In flies, Polycomb repressive complexes (PRCs) were shown to associate with specific DNA-elements called PREs in a dedicated transcription factor-dependent manner. However, it is still not understood what targets EZH2-containing PRC2 to target genes in mammals, as will be discussed in greater detail later. H3K27me3 deposition was proposed to act as a binding site for the CBX chromodomain-containing subunit of polycomb repressive complex 1 (PRC1) (Cao et al. 2002; Min et al. 2003). This in turn would function to establish polycomb repressive domains, for example by preventing access of chromatin-remodelling factors such as SWI/SNF (Francis and Kingston 2001), in a manner analogous to H3K9me3-dependent recruitment and reinforcement of silent heterochromatin via HP1. Furthermore, H3K27me3 was also shown to be directly bound by the carboxy-terminal domain of EED PRC2 subunit, thereby potentially enabling a self-propagating mechanism for the transmission of a histone modification from mother to daughter cells (Hansen et al. 2008; Margueron et al. 2009).

Interestingly, as mentioned previously, PRC2-mediated deposition of H3K27me3 is inhibited by activating marks such as H3K4me3 (Schmitges et al. 2011). Consistent with such an observation, PRC2 was reported to directly recruit the H3K4me3 demethylase RBP2 to its target genes in mouse embryonic stem cells, a mechanism suggested to be essential for the correct regulation of polycomb targets (Pasini

et al. 2008). The interplay and inter-dependence between H3K4 and H3K27 methylation is further complicated by the discovery that the H3K27me2/me3 JmjC-demethylase UTX/KDM6A is a component of a multiprotein complex containing the Trithorax-related protein MLL4 (Cho et al. 2007; Issaeva et al. 2007; Lee et al. 2007b), indicative of an elegant mechanism of counteracting polycomb-mediated silencing during gene activation. In addition to UTX, which is constitutively expressed and was proposed to act as a "housekeeping" demethylase for surveillance of H3K27methylation levels , another JmjC-demethylase specific for H3K27me2/me3 identified was JMJD3/KDM6B (Agger et al. 2007; De Santa et al. 2007; Lan et al. 2007). The activity of JMJD3 was shown to be rapidly induced in macrophages in response to the inflammatory stimulus lipopolysaccharide (LPS), enabling the activation of bivalently marked inflammation-responsive genes.

As H3K27me3 deposition is inhibited by the activating mark H3K4me3, it was postulated that these two modifications cannot coexist on nucleosomes. However, as early as 2006, chromatin immunoprecipitation (ChIP) and DNA tiling array-type studies in mouse ES cells discovered the existence of so-called "bivalent domains" at developmental regulatory loci, which show co-enrichment of the two opposing histone modifications (Bernstein et al. 2006). Despite the presence of activating H3K4me3, bivalent genes are expressed only at low levels, but upon differentiation resolve to monovalency: genes that become expressed lose H3K27me3, while those that become silenced lose H3K4me3. This discovery suggested that bivalent domains silence developmental genes in ES cells while keeping them poised for activation upon differentiation. Importantly, it was also reported that bivalency is not unique to pluripotent cells, but that it is retained to a lower extent in differentiated cells, with approximately 4% of CpG-rich promoters in mouse embryonic fibroblasts (MEFs) showing a bivalent chromatin status (Mikkelsen et al. 2007). One solution to the co-existence of these two opposing histone marks was provided by combining mononucleosome ChIP with mass spectrometry (MS)-based quantitative profiling of histone modifications, which showed that H3K27me3 and H3K4me3 reside on opposite tails in asymmetrically-modified nucleosomes (Voigt et al. 2012; Voigt et al. 2013). Despite this discovery, the biological significance of bivalent domains and their contribution to developmental plasticity remain to be investigated.

EZH2 is responsible for catalysing all three H3K27 methylation states. Although H3K27me3 has been extensively studied, the functions of the mono- and di-methylation states are still not clear. Interestingly, results quantifying the relative abundance of several histone H3 PTMs in mouse ESCs indicated that 70% of H3K27 is present in its di-methylated form (H3K27me2), whereas 7% and 4% of total H3 is in the H3K27me3 and H3K27me1 forms, respectively (Jung et al. 2010). Although intriguingly abundant, H3K27me2 was attributed to a 'hit-and-run' un-recruited activity of PRC2 (Schwartz and Pirrotta 2013a). Recent work however indicates that the function of PRC2 may go beyond its association with Polycomb-mediated silencing. In a study by Ferrari et al., H3K27me1 was shown to specifically accumulate within transcribed regions in a H3K36me3-dependent manner, while H3K27me2 was suggested to prevent unspecific H3K27 acetylation and serve as a protective mechanism from firing non-cell-type-specific enhancers (Ferrari et al. 2013). These exciting observations clearly underline the need to further clarify the mechanisms by which the activity of PRC2 contributes to transcriptional regulation and differentiation.

1.3.1.2.3. H3K36me2/me3

In yeast, SET-domain containing 2 (SET2) catalyses all three methylation events at H3K36 (Strahl et al. 2002). In mammals, however, there is a division of labour between mono- and di-methylation and Set2-type H3K36me3 (Wagner and Carpenter 2012), with nuclear receptor SET domain-containing 1/2 (NSD1/NSD2) responsible for H3K36me1/me2 (Rayasam et al. 2003; Li et al. 2009b; Lucio-Eterovic et al. 2010), while HYPB/SETD2 mediates all global and transcription-dependent H3K36 trimethylation (Edmunds et al. 2008). Interestingly, H3K36me3 is highly enriched over the transcribed region of genes with peaks towards the 3' end in a manner that correlates with active transcription. This chromatin profile was shown to be the result of SET2 association with the elongating serine-2 and to a lesser extent serine-5 phosphorylated form of RNAPII (Krogan et al. 2003b; Li et al. 2003; Kizer et al. 2005). The interaction between SET2 and the hyperphosphorylated C-terminal domain (CTD) of RNAPII occurs via the SET2 Rpb1-domain (SRI) within the methyltransferase (Kizer et al. 2005). Interestingly, deletion of

this domain not only ablates SET2-RNAPII association, but it also abolishes H3K36 methylation *in vivo* and leads to a transcription elongation defect, suggesting that the interaction between SET2-RNAPII is an essential contributor to the fidelity of the elongation process. Although in yeast H3K36me2 is also enriched over coding regions, studies indicate that, unlike H3K36me3, it is not proportional to transcriptional levels but instead an indication of the on/off transcriptional status of the respective gene (Pokholok et al. 2005; Rao et al. 2005), although how this differential methylation is achieved requires further clarification.

In mammals, H3K36me3 is also deposited in the core and towards the 3' end of active genes (Mikkelsen et al. 2007) in a manner dependent on the association between SETD2 and Ser2-phosphorylated RNAPII (Yoh et al. 2008). H3K36me2 however was found to pervasively coat intra- and intergenic regions as placed by the histone methyltransferase NSD2 (Kuo et al. 2011), while being generally depleted at CpG island elements via the action of two ZF-CxxC containing JmjC-demethylases, KDM2A and KDM2B, as discussed later.

The most studied function of H3K36methylation (di-and tri-) is its ability to suppress spurious initiation of transcription from cryptic start sites within coding regions (Carrozza et al. 2005; Joshi and Struhl 2005; Keogh et al. 2005; Li et al. 2009a). Briefly, transcriptional elongation is coupled to extensive nucleosome acetylation (Winkler et al. 2002), which in turn creates a permissive chromatin architecture potentially unmasking otherwise inaccessible cryptic transcription start sites. H3K36 methylation was shown to be recognized by the chromodomain of EAF3 subunit of the Rpd3S HDAC complex, whose deacetylase activity removes histone acetylation associated with elongation therefore supressing internal transcription initiation.

In mammals, H3K36-dependent repression of cryptic transcription seems to be two-fold. Firstly, H3K36me3 was suggested to recruit histone H3K4me2/3 demethylase KDM5B, which in turn would represses cryptic transcription by removing intragenic H3K4me3 (Xie et al. 2011). Secondly, SETD2-mediated repression of intragenic transcription was shown to be attributable to an association with the FACT histone chaperone and direct reassembly of nucleosomes in the wake of RNAPII elongation (Carvalho et al. 2013). H3K36 methylation has also been implicated in alternative splicing, although the

cause-versus-effect relationship between the two has not been unambiguously established (Kolasinska-Zwierz et al. 2009; de Almeida et al. 2011; Pradeepa et al. 2012).

Interestingly, H3K36methylation also appears to be inhibitory to PRC2-mediated H3K27me3 deposition *in vitro* (Yuan et al. 2011), antagonism which seems to be a conserved feature. In *Caenorhabditis elegans* embryos lacking MES-4, the major H3K36 methyltransferase, there is a striking redistribution of PRC2 from the X chromosome, where it normally binds, to autosomes, suggesting that H3K36me3 repels H3K27me3 from germline genes to concentrate it on other regions of the genome (Gaydos et al. 2012). This inhibitory role of H3K36methylation on Polycomb function is further reinforced by studies showing that mammalian Absent, small, or homeotic discs 1-like (ASH1L), a member of the Trithorax group proteins, specifically di- and tri-methylates H3K36 independently of transcriptional elongation to counteract Polycomb silencing and allow establishment of *Hox* gene expression (Miyazaki et al. 2013). The antagonism between activating histone marks such as H3K4/H3K36 methylation and PRC2-mediated H3K27me3 will be referred to extensively throughout this study.

1.3.1.3. Histone lysine ubiquitylation

The 76 amino acid (~8.5kDa) small ubiquitin regulatory protein can be conjugated to substrates in a process consisting of three steps, namely activation, conjugation and ligation, carried out by E1, E2 and E3 enzymatic activities (Handley et al. 1991; Weake and Workman 2008; Zhou et al. 2009). The first step in the ubiquitylation reaction involves ATP-dependent activation of ubiquitin to a high-energy intermediate and is catalysed by E1, the ubiquitin-activating enzyme. Ubiquitin is then transferred to a cysteine residue in E2, the ubiquitin-conjugating enzyme (also known as ubiquitin carrier protein or UBC), via a reactive thioester bond. E3 ubiquitin ligases catalyse the final step of the conjugation reaction, by transferring ubiquitin from the E2 enzyme to a lysine residue in the target protein either directly or through a covalent E3-ubiquitin thioester intermediate. There are two major types of E3 ligases in mammals, characterized by either a HECT (homologous to the E6AP carboxyl terminus) or a

RING (really interesting new gene) domain (Freemont et al. 1991; Deshaies and Joazeiro 2009), with RING-based E3 ligases representing the most abundant family, being specified by over 600 human genes.

Substrates can be mono- or poly-ubiquitylated, with polyubiquitylation through Ub lysine 48 residue (K48) generally targeting proteins for 26S proteasomal degradation, while monoubiquitylation or polyubiquitylation via Ub K63-linked chains is gaining increasing understanding as a post-translational modification mediating diverse cellular functions, such as DNA repair, signal transduction, subcellular localization or gene expression (Passmore and Barford 2004). The two most studied histone ubiquitylation events in mammals are represented by monoubiquitylation of H2AK119 and H2BK120. While H2BK120ub1 is closely associated with transcriptionally active regions and was shown to be required for both H3K4me3 by COMPASS and H3K79me3 by DOT1 in a histone crosstalk pathway conserved from yeast to humans (Lee et al. 2007a; Nakanishi et al. 2009; Kim et al. 2013), H2AK119ub1 was instead suggested to be repressive to transcription.

1.3.1.3.1. H2AK119ub1

H2AK119ub1 is primarily mediated in mammals by the RING1A (RING1)/RING1B (RNF2) E3 ubiquitin ligase subunit of the Polycomb repressive complex 1 (PRC1) (Wang et al. 2004a), the enzymatic activity of which is stimulated *in vitro* by the incorporation of fly PSC homologues PCGF2 (Cao et al. 2005) or PCGF4 (Elderkin et al. 2007). Knockdown of RING1B in HeLa cells (Wang et al. 2004a) or gene knockout in mice (de Napoles et al. 2004) resulted in a striking depletion of H2AK119ub1 levels, suggesting that this represents the major H2AK119 E3 ubiquitin ligase, although some lower-level redundancy was suggested to exist with RING1A-catalyzed ubiquitylation. In agreement with this, single-knockout mouse ES cells lacking RING1B retain ES-like morphology and can be cultured for >20 passages, while double deletion of RING1A/RING1B results in halted cell proliferation and loss of typical ES cell morphology (Endoh et al. 2008).

In addition to inclusion in the PRC1 complex, RING1A/RING1B was also found associated with other distinct polypeptides as part of the E2F-6.com-1 (Ogawa et al. 2002) and the FBXL10-BCOR (Gearhart et al. 2006; Sanchez et al. 2007) complexes. Although preliminary studies suggest that the E2F-6.com-1 complex contributes to silencing of E2F- and Myc-responsive genes in quiescent cells, it is still not clear to what extent this is related to the E3 ubiquitin ligase activity of RING1B. In contrast, the FBXL10 (KDM2B)-BCOR complex was shown to catalyse the addition of Flag-tagged ubiquitin onto H2A *in vitro* in a concentration-dependent manner, suggesting that, although identified only in transformed cells, this complex may play essential roles in regulating global H2AK119ub1 levels.

Interestingly, another RING domain-containing H2AK119ub1 E3 ligase was identified in mammals, 2A-HUB, which was shown to be specifically recruited by the nuclear receptor co-repressor N-CoR/HDAC1/3 complex in macrophages to mediate ubiquitylation-dependent repression of a subset of regulated chemokine gene promoters, through a mechanism suggested to negatively regulate transcriptional elongation by preventing FACT recruitment to those sites (Zhou et al. 2008). As 2A-HUB depletion does not seem to affect global H2AK119ub1 levels, it may be that the function of this ligase is to specifically regulate this localized rather than general histone modification.

In Drosophila, Lagarou et al. (2008) showed that the dRING protein forms an additional complex distinct from the core PRC1 complex, which was termed dRAF (dRING-associated factors). Importantly, this complex comprises the single KDM2 homologue in flies, together with the Posterior Sex Combs (PSC) classical polycomb member, in a manner reminiscent of the FBXL10 (KDM2B)-BCOR complex purified from mammals. Furthermore, using dKDM2 knockdown studies and *in vitro* reconstitution reactions, dRAF was shown to not only mediate H3K36me2 demethylation, but was also suggested to be the main H2A ubiquitylating complex in cells, although the potential interplay between PRC1 and dRAF in flies remains to be fully characterized.

H2AK119ub1 is found to be present on the inactive X chromosome, as well as at promoters of developmentally-regulated polycomb target genes. Although it represents a relatively abundant modification comprising between 5-15% of available H2A (Goldknopf et al. 1975), the mechanisms by which H2A ubiquitylation contributes to gene regulation remain to be fully identified, especially as the

majority of studies carried out fail to unambiguously distinguish between H2AK119ub1-dependent or independent functions of PRC1. One of the main silencing mechanisms attributed to PRC1-mediated H2AK119ub1 deposition is nevertheless suggested to be through inhibition of transcriptional elongation by restraining poised RNA polymerase II at bivalent genes in mouse ES cells (Stock et al. 2007). This observation is in agreement with results from Ring1A/B-dKO ESCs, which show increased RNAPII loading at PRC1 target gene promoters in a manner-dependent on disrupting the E3 ligase catalytic activity (Endoh et al. 2012).Nevertheless, Sce (dRING) knockout allele flies show de-repression of only a subset of genes upon loss of H2A ubiquitylation, while other polycomb targets become misexpressed only in absence of Psc-Su(z)2 and Polyhomeotic (Ph) subunits (Gutierrez et al. 2012), questioning whether PRC1 repression mechanisms are dependent on this modification or whether it is primarily mediated by H2A ubiquitylation-independent mechanisms such as chromatin compaction by other PRC1 subunits.

One potential mechanism by which H2AK119ub1 could be exerting its function in gene regulation could be related to specific recruitment of effector proteins, similar to other histone modifications. However, to date, few definite histone-specific H2AK119ub1-binding proteins have been identified. One report suggested that the non-canonical PRC1 subunit RYBP harbours an ubiquitin-binding domain (UBD) responsible for mediating its interaction with H2AK119ub1 (Arrigoni et al. 2006), report which was later called into question by experiments showing that MG132 treatment-dependent depletion of ubiquitylated H2A does not displace RYBP from chromatin (Tavares et al. 2012). In addition, a study by Richly et al. (2010) suggested that, in human cell lines, ZRF1 (zuotin-related factor 1) specifically binds H2AK119ub1 to compete with Polycomb occupancy and facilitate transcriptional activation of PcG-repressed targets through a potential cooperation with deubiquitinase enzymes. However, it is not clear how ZRF1 competes with RING1B for H2AK119ub1 binding, as the PcG E3 ubiquitin ligase was not reported to directly bind its own histone modification.

Recently, by employing semisynthetic H2Bub1 reconstituted chromatin on a biotinylated DNA template, Shema-Yaacoby et al. (2013) identified factors from SILAC-labelled nuclear extracts which preferentially bound chromatin containing H2BK120ub1. Interestingly, among the proteins identified was the SWI/SNF chromatin remodeling complex, subsequently shown in the same study to be required for optimal

transcription of H2BK120ub1-containing genes. One can envisage that such a method could in principle also be applied to identify H2AK119ub1 “readers”, thereby allowing a better understanding of the functions of this PTM in mammalian gene regulation.

A number of mammalian deubiquitinating enzymes (DUBs) which catalyse the removal of ubiquitin from H2A have recently been characterized, including UBP-M, 2A-DUB, USP21 and USP22 (Joo et al. 2007; Zhu et al. 2007; Nakagawa et al. 2008; Zhao et al. 2008b). Although these DUBs have been linked with gene activation, for example UBP-M was suggested to regulate *Hox* gene expression through H2A deubiquitination while 2A-DUB associates with the p300/CAF histone acetyl transferase to mediate transcriptional regulation events in androgen receptor-dependent gene activation, it is still not clear how these enzymes are regulated nor how they are recruited to specific chromosomal loci.

Interestingly, work in Drosophila identified an association between the polycomb protein ASX (Additional Sex Combs) and the histone deubiquitinase CALYPSO (dBAP1) to form the Polycomb Repressive Deubiquitinase (PR–DUB) complex specifically bound at polycomb targets to deubiquitylate the histone mark placed by PRC1 (Scheuermann et al. 2010) (detailed in section 1.6.4). Disruption in flies of the CALYPSO catalytic activity or absence of ASX subunit results in failed *Hox* gene repression, although the reasons for why efficient Polycomb-repression might require a H2A deubiquitylating activity remain unclear.

1.4. DNA methylation in higher eukaryotes and its role in gene expression

1.4.1. Discovery of DNA methylation and its distribution in mammals

DNA methylation is a widespread chemical modification in both prokaryotes and eukaryotes (Iyer et al. 2011). In bacteria, methylation can occur on both cytosine (C-5 or N-4 positions) and adenine (N-6 position) residues as part of the restriction-modification system important for protecting hosts from foreign DNAs such as transposons and viral DNAs (Meselson et al. 1972; Kobayashi et al. 1999). In mammals, methylation occurs predominantly at the C-5 position of cytosine residues in the context of

CpG dinucleotides (Figure 1.2, panel A), with approximately 70 to 80% of all CpGs being methylated (Ehrlich et al. 1982; Bird and Wolffe 1999). Non-CpG context DNA methylation was also reported to occur primarily in embryonic stem cells at CpA and, to a lesser extent, at CpT dinucleotides (Ramsahoye et al. 2000), although whether this is biologically relevant or a consequence of DNA methyltransferase promiscuity remains unclear.

DNA methylation was initially associated with inactivation of one X chromosome during mammalian dosage compensation, following initial observations that treatment of cells with the demethylating agent 5-azacytidine (5-AzaC) can reactivate genes on the inactive human X chromosome (Mohandas et al. 1981; Venolia et al. 1982). Importantly, at a time when the identity of mammalian methyltransferases was not known, work on X inactivation and imprinting already suggested that the self-complementary nature of CpG dinucleotide methylation could enable maintenance of the DNA methylation pattern following semi-conservative DNA replication, thus acting as a heritable regulatory mark (Holliday and Pugh 1975; Riggs 1975).

CpG dinucleotides are significantly depleted in mammalian genomes where the observed frequency of CpG dinucleotides is only 0.20–0.25 of that expected from the G and C base composition (Gardiner-Garden and Frommer 1987). This observation was suggested to be a consequence of the increased mutability of methylated cytosine residues, which spontaneously deaminate to thymine (Figure 1.2, panel B) (Coulondre et al. 1978; Duncan and Miller 1980). While cytosine residues can spontaneously deaminate to uracil, this unnatural mismatched DNA base is recognized and repaired by uracil glycosylases (Krokan et al. 2002) . In contrast, thymine resulting from deamination of 5 methyl-cytosine is a naturally occurring DNA base and generates a T:G mismatch which is less efficiently repaired by either of the two mismatch-specific T glycosylases, TDG (Neddermann and Jiricny 1993) and MBD4 (Hendrich et al. 1999; Millar et al. 2002). If not repaired to cytosine before the following round of DNA replication, this T:G mismatch will be converted to TpG (or CpA in the complementary strand) (Antequera 2003), resulting in gradual depletion of CpG dinucleotides from the genome.

Although in human somatic cells, 5mC accounts for ~1% of total DNA bases (Bird 2002), DNA methylation is unevenly distributed in the mammalian genome, with heterochromatic regions,

transposons and repetitive sequences generally associated with hypermethylation as a mechanism proposed to be essential for maintaining genome integrity (Chen et al. 1998; Robertson and Wolffe 2000). In addition, DNA methylation was found to be preferentially enriched in exons relative to introns, consistent with the targeting of DNA methylation to nucleosomes (Chodavarapu et al. 2010), in agreement with reports supporting a role for intragenic DNA methylation in regulating alternative pre-mRNA splicing (Maunakea et al. 2013). Furthermore, DNA methylation was also detected in intergenic regions (Rakyan et al. 2004), suggesting that mammalian genomes are dominated by 5mC.

An exception to this widespread DNA methylation is represented by regions of the genome called CpG islands (CGIs), topic addressed in greater detail below.

1.4.2. De novo and maintenance DNA methyltransferases

DNA methyltransferases (DNMTs) catalyse the transfer of a methyl group from *S*-adenosylmethionine (SAM) to the C5-position on the cytosine pyrimidine ring. Methylation does not interfere with the Watson/Crick base pairing properties of cytosine, being positioned instead in the major groove of the DNA for potential interrogation by DNA-binding proteins (Jeltsch 2002). Due to the chemistry of the covalent modification mechanism, DNA cytosine methyltransferases requires access to both sides of the aromatic cytosine ring, providing an explanation for why these enzymes act through a base-flipping mechanism whereby cytosine is rotated out of the DNA helix during the reaction (Cheng et al. 1993; Klimasauskas et al. 1998; O'Gara et al. 1998).

DNMT1 was the first eukaryotic methyltransferase to be purified and cloned (Bestor and Ingram 1983). In agreement with its *in vitro* 30-40 fold higher preference for hemi-methylated versus non-methylated DNA, DNMT1 was shown to be responsible for the propagation of DNA methylation by maintenance of the symmetrical DNA methylation pattern following replication (Jeltsch 2006). DNMT1 is a large protein of 1620 amino acid residues, which can be subdivided into a regulatory N-terminal part and the C-terminal part of the enzyme, which encodes the catalytic activity.

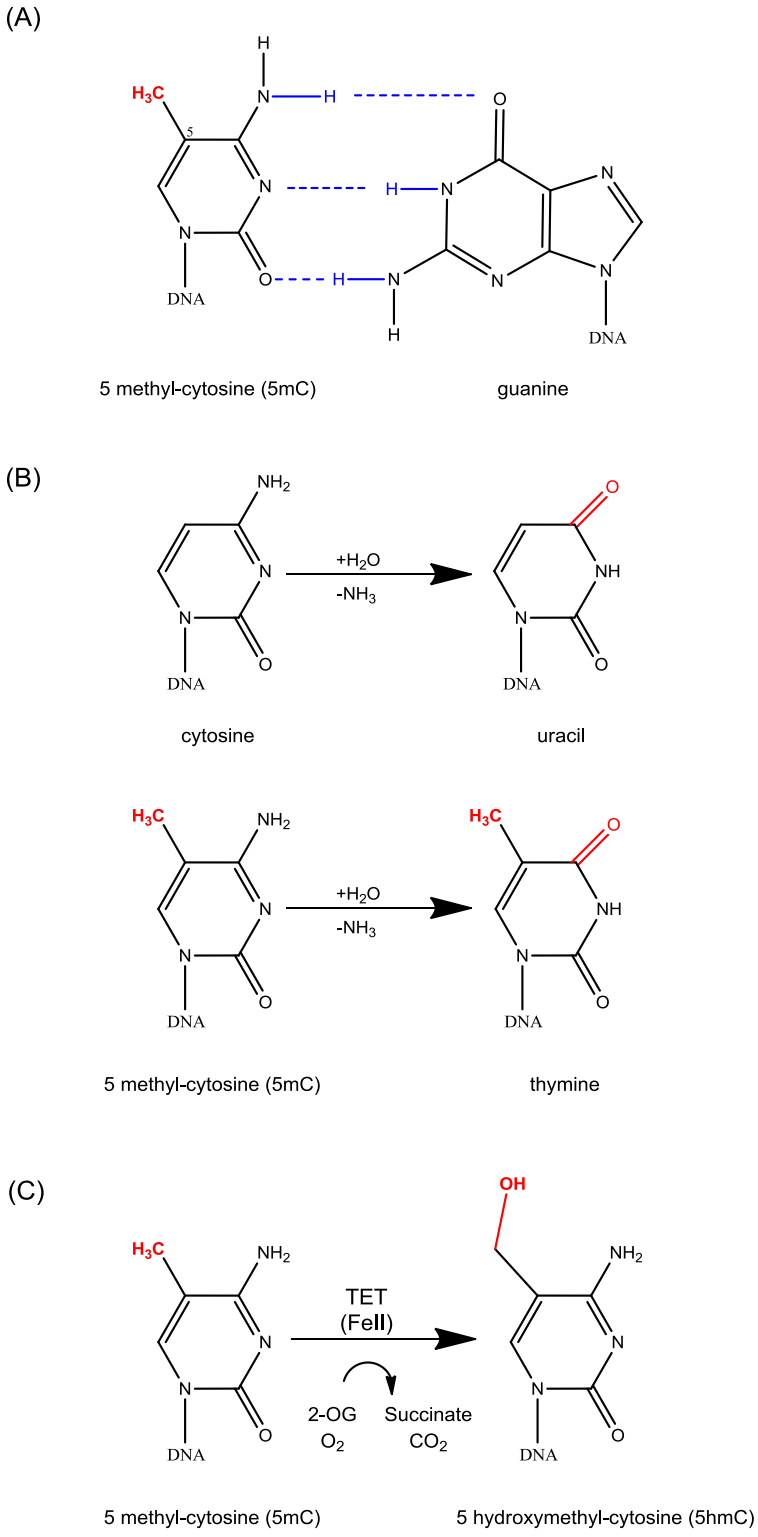


Figure 1.2. Cytosine base modifications in mammalian DNA.

(A) Addition of a methyl group (shown in red) at the C5 position of the cytosine pyrimidine ring by DNA methyltransferases does not sterically hinder GC base pairing.

(B) Cytosine and 5 methyl-cytosine are subject to spontaneous deamination, resulting in uracil and thymine, respectively.

(C) Tet family proteins (TET1/2/3) oxidize 5mC into 5hmC in the presence of 2-oxoglutarate and iron (FeII).

Interestingly, the N-terminal domain harbours a functional ZF-CxxC DNA binding domain, which is surprising considering that the preferred substrate of DNMT1 is hemi-methylated DNA. Recent crystallographic work on mouse DNMT1 (650–1602) and human DNMT1 (646–1600) bound to duplex DNA-containing non-methylated CpG probes suggested that the ZF-CxxC domain plays an essential role in a DNMT1 auto-inhibitory mechanism (Song et al. 2011), although later reports conflicted with this finding (Takeshita et al. 2011).

DNMT1 was shown to be concentrated at replication foci in early and mid-S phase as a result of an interaction between the RFTS (replication foci targeting sequence) of the maintenance DNMT and PCNA (proliferating cell nuclear antigen), the sliding clamp of eukaryotic DNA polymerase (Leonhardt et al. 1992; Schermelleh et al. 2007). However, this interaction was detected to be not only transient, but also insufficient to explain the maintenance of DNA methylation, as PCNA-binding mutant DNMT1 retained the capacity to localize at DNA and was catalytically engaged at hemi-methylated sites following replication. An alternative mechanism to ensure faithful propagation of the DNA methylation pattern following DNA replication was shown to be mediated by UHRF1 (Ubiquitin-like with PHD and ring finger domains 1; also known as NP95/ICBP90)-dependent recognition of hemi-methylated DNA via its SRA (SET and RING finger associated) domain and subsequent recruitment of DNMT1 (Bostick et al. 2007; Sharif et al. 2007). However, no direct interaction was detected between UHRF1 and DNMT1. Instead, recent work indicated that UHRF1 can specifically ubiquitylate H3K23 and this mark is a prerequisite for DNMT1 recruitment to DNA replication sites (Nishiyama et al. 2013). The importance of the maintenance methyltransferase for development is demonstrated by the observation that knockout of *Dnmt1* gene in mice results in genome-wide loss of DNA methylation, abnormal expression of imprinted genes, ectopic X-chromosome inactivation due to demethylation and activation of *Xist* and a recessive embryonic lethal phenotype (E9.5) (Li et al. 1992; Li et al. 1993; Panning and Jaenisch 1996).

The retention of significant amounts of 5mC in the DNA of the homozygous Dnmt1 mutant ES cells (~5% of wild type levels) (Li et al. 1992) suggested that alternative uncharacterized DNMTs may be present in cells. Indeed, by searching the EST database using full-length bacterial type II cytosine-5 methyltransferase sequences as queries, DNMT3A and DNMT3B were shown to contain the highly

conserved cytosine-5 methyltransferase motifs and shown to possess comparable catalytic activity towards hemi- and non-methylated DNA (Okano et al. 1998). Gene targeting-based inactivation of both *Dnmt3a* and *Dnmt3b* resulted in embryonic lethality at ~8.5dpc, with global demethylation of their genomes and inability to methylate newly integrated retroviral DNA, suggesting that these two enzymes are responsible for *de novo* DNA methylation. However, they also seem to have some non-overlapping functions during early development, with DNMT3A important in germ cells for proper establishment of imprints, while loss of DNMT3B results in demethylation of centromeric minor satellite repeats (Okano et al. 1999). Despite intense research, it is still not clear how *de novo* methylation patterns are established or how DNMT3A/3B is targeted to chromatin in early embryogenesis. Interestingly, both enzymes harbour a PWWP domain, which was suggested to be essential for targeting the *de novo* DNMTs to chromatin through its ability to recognize H3K36me3 (Dhayalan et al. 2010).

A third DNMT3 member, DNMT3L, is highly expressed in embryonic stem (ES) cells but down-regulated in differentiated embryoid bodies (Hata et al. 2002). Although DNMT3L lacks catalytic activity, it was shown to physically associate with DNMT3A/3B (Suetake et al. 2004), stimulate their catalytic activity (Gowher et al. 2005) and play an essential role in maternal methylation imprinting and mouse spermatogenesis (Bourc'his et al. 2001; Hata et al. 2002). Importantly, structural data has revealed that methylation at H3K4 prevents interaction of the mammalian DNMT3L with the histone H3 tail (Ooi et al. 2007), potentially as a mechanism to inhibit *de novo* deposition of DNA methylation at transcriptionally active promoter regions.

1.4.3. Interpretation of DNA methylation and its role in transcriptional silencing

Cytosine methylation does not affect the integrity of the GC base pairing or disrupt the chromatin structure directly, but it can exert its biological function in at least two ways. Firstly, DNA methylation could impair binding of transcriptional factors to their target genes, thereby directly interfering with gene activation, as is the case for methylation-responsive CTCF binding in regulating imprinting of the

H19/Igf2 locus in mice (Watt and Molloy 1988; Kass et al. 1997; Bell and Felsenfeld 2000). Secondly, DNA methylation can be specifically recognized by methyl-CpG binding domain (MBD) proteins, which influence gene activity by their association with different co-repressor complexes (Hendrich and Bird 1998; Ballestar and Wolffe 2001; Klose and Bird 2006).

MeCP2 (methyl CpG binding protein 2) was the first MBD-containing protein to be identified (Meehan et al. 1992) and was found to be associated with the transcriptional repressor mSIN3A and histone deacetylases (Nan et al. 1998), as well as interacting with the NCoR/SMRT co-repressor complexes (Lyst et al. 2013) as identified from studies of *Mecp2* gene mutations in the severe neurological disorder Rett syndrome (RTT). Homology searches for other proteins sharing the MECP2 MBD domain resulted in the identification of MBD1-4 (Hendrich and Bird 1998). While MBD3 has lost the capacity to bind cytosine-methylated DNA, MBD1 forms a stable complex with histone H3K9 HKMTase SETDB1 to facilitate replication-coupled methylation of H3K9 (Sarraf and Stancheva 2004), MBD2/MBD3 have been identified as core subunits of the Mi-2/NuRD (Nucleosome Remodeling Deacetylase) complex (Zhang et al. 1999; Le Guezennec et al. 2006). MBD4 was shown to function primarily in DNA repair by removing thymines from T:G mismatches resulting from deamination of 5mC (Hendrich et al. 1999), although an association with SIN3 and HDAC1 to mediate transcriptional repression was also reported (Kondo et al. 2005).

In addition, KAISO protein was shown to recognize methylated CpG dinucleotides in the absence of an MBD-domain, employing instead zinc-finger domains (Daniel and Reynolds 1999; Prokhortchouk et al. 2001). Interestingly, KAISO was also reported to mediate repression of methylated genes by associating with the histone-deacetylase-containing NCoR co-repressor complex (Yoon et al. 2003), suggesting a fundamental role of 5mC in mediating transcriptional silencing.

The importance of promoter DNA methylation in transcriptional repression is underlined by cancer studies, where aberrant methylation of otherwise non-methylated tumour suppressor promoters results in gene inactivation (Esteller 2002; Jones and Baylin 2007). However, the observation that a strong transcriptional activator can drive transcription through a heavily methylated template (Thompson et al. 1986; Thompson et al. 1988), together with the fact that ES cells devoid of all three DNMTs retain stem

cell properties and chromosomal stability in the absence of detectable CpG methylation (Tsumura et al. 2006), suggests that further work is required to precisely dissect the role of 5mC in gene regulation.

1.4.4. CpG islands (CGI) are characterised by an unique chromatin architecture

Despite the prevalence of CpG methylation, some regions of the genome remain free of this modification. These contiguous regions of non-methylated DNA were first identified over 30 years ago by digestion of the mouse genome with restriction endonucleases Hpa II and Hha I which are blocked by 5mC (Cooper et al. 1983; Bird et al. 1985). While bulk vertebrate DNA is generally a poor substrate for these enzymes due to high levels of CpG methylation, approximately 1% of the genome was frequently cut and shown to correspond to islands of non-methylated DNA. These regions, originally termed HTFs (HpaII tiny fragment) and later renamed as CpG islands (CGIs) (Bird 1987; Gardiner-Garden and Frommer 1987), were shown to generally encompass the promoters of housekeeping but also many tissue-specific genes (Larsen et al. 1992; Saxonov et al. 2006; Weber et al. 2007; Zhu et al. 2008).

CpG islands have on average a length of 1 kb and display elevated CpG and GC content compared to bulk genomic DNA, being characterised by ≥ 50% GC content and ≥ 0.6 observed over expected (obs/exp) ratio of CpG frequency as compared to the average genome GC content of ~ 41% with an obs/exp CpG ratio of only 0.2-0.25 (Josse et al. 1961; Gardiner-Garden and Frommer 1987; Takai and Jones 2002). Approximately 50% of the ~23,000 CGIs mapped in the mouse and human genomes are associated with currently annotated TSSs while the remaining fraction, denoted as “orphan” CGIs, was evenly distributed across intragenic and intergenic regions but appear to retain promoter-like characteristics displaying a more tissue restricted expression profile (Illingworth et al. 2010). Despite their extensive association with promoters (approximately 60-70% of mammalian genes carrying a CGI-containing promoter), it is still not clear how CpG islands contribute to gene regulation or the mechanisms by which they are kept free from DNA methylation.

One potential mechanism by which CGIs could be exerting their function is by acting as platform for recruitment of different chromatin-modifying activities or additional regulatory factors, whose activities in turn would generate a unique chromatin architecture at CGIs to distinguish them from the surrounding genome (Blackledge and Klose 2011). Importantly, CpG islands are insufficient to drive transcriptional output *per se*, as is suggested by the fact that CGIs are found at promoters of both active and inactive genes. Instead, one can envisage that CGIs play a role in facilitating the transcriptional process by providing a permissive and more accessible chromatin configuration favouring binding of the basal transcription machinery.

Several lines of evidence suggest that this might be the case. First, as early as 1990, experiments analysing the composition of HpaII-released fragments indicated that CpG island nucleosomes are depleted of the linker histone H1 and are hyperacetylated (Tazi and Bird 1990), properties suggested to favour an open chromatin conformation. Secondly, extensive bioinformatic analysis of genome-wide Pol II accumulation by ChIP-seq and nucleosome density by micrococcal nuclease digestion (MNase-seq) in mouse primary T-cells showed that CGIs and GC content dictate nucleosome depletion in an essentially transcription-independent manner, supporting the notion that CGIs are characterised by an open chromatin conformation (Fenouil et al. 2012). In agreement with these previous results, Ramirez-Carrozzi et al. (2009) demonstrated that lipopolysaccharide (LPS)-dependent induction of CGI-containing immediate early genes occurs more rapidly than their non-CGI counterparts and without the need for SWI/SNF nucleosome remodeling complexes, suggesting that the chromatin environment of CpG islands is more permissive to transcriptional initiation.

Another factor potentially contributing to this open chromatin conformation is represented by the CGI characteristic set of histone modifications. ChIP-Seq studies indicated that H3K4me3 is a signature histone mark of CpG island promoters, regardless of the transcriptional status of the associated gene, in a manner partially dependent on the ZF-CxxC domain containing protein CFP1 (Mikkelsen et al. 2007; Thomson et al. 2010). Moreover, CGIs are specifically depleted of the histone mark H3K36me2, demethylation catalysed by the ZF-CxxC domain containing proteins KDM2A/2B, as discussed in greater detail below. Therefore, CGIs are enriched in H3K4me3, modification shown to support transcriptional

initiation, and also depleted in H3K36me2, a mark which is otherwise inhibitory to transcriptional initiation (see Section 1.3.1.2.3), further contributing to the accessible chromatin architecture of these genomic elements. In agreement with the transcriptionally permissive chromatin conformation found at CGIs, systematic 5'-end analysis of the mouse and human transcriptomes using cap analysis of gene expression (CAGE) approach indicated that, while TATA-box promoters show sharp TSSs, CGIs utilize dispersed transcription start sites and are generally associated with broad TSS regions (Carninci et al. 2006), as well as bidirectional promoter activity producing short sense and antisense RNAs, often in the absence of productive elongation and mature mRNA transcripts (Trinklein et al. 2004; Core et al. 2008; Seila et al. 2008). Interestingly, Bio-CAP profiling of non-methylated DNA across seven vertebrate species revealed that NMIs (non-methylated DNA islands) are a highly conserved feature of vertebrate promoters, further reinforcing a central role for CGIs in gene regulation (Long et al. 2013b).

1.4.5. DNA demethylation

One possibility by which CpG islands would be kept free from cytosine methylation is through the activity of DNA demethylating enzymes nucleating at these genomic elements. However, DNA methylation has long been considered a relatively stable and heritable epigenetic modification. Nevertheless, waves of global demethylation are observed at two developmental stages, namely in pre-implantation embryos and in developing primordial germ cells (PGCs), and this alteration can take place either actively or passively through cellular division dilution (Surani 1998; Oswald et al. 2000; Hajkova et al. 2008; Seisenberger et al. 2013) . The discovery of 5 hydroxymethyl-cytosine (5hmC) in differentiated Purkinje neuronal cells where it is most abundant (Kriaucionis and Heintz 2009), as well as in ES cells (Pastor et al. 2011), together with the identification of all three mammalian 5mC dioxygenases belonging to the ten-eleven translocation (Tet) family responsible for the conversion of 5mC to 5hmC (Tahiliani et al. 2009; Ito et al. 2010) , pointed towards 5-hydroxymethyl-C as a possible route to actively remove DNA methylation (Figure 1.2, panel C).

All three mammalian TET proteins, namely TET1, TET2 and TET3, belong to a distinct family of 2-oxoglutarate (2OG) and Fe (II)-dependent dioxygenases which use molecular oxygen as a substrate to oxidize 5mC to 5hmC, but also to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al. 2011; Pastor et al. 2013). Interestingly, although all three TET enzymes share the core catalytic dioxygenase domain, TET1 and TET3 also encode an N-terminal ZF-CxxC DNA binding domain which in other proteins was shown to have high affinity for non-methylated CpGs. However, there are conflicting reports on whether TET1/TET3 DNA binding domain can recognize non-methylated DNA, especially considering that the ZF-CxxC domains found in TET1 and TET3 differ from that found in classical CGI-binders such as CFP1 and exhibit a truncated linker region and a divergent DNA-binding loop (Long et al. 2013a). Although TET2 lacks a ZF-CxxC domain itself, it was recently shown to physically associate with the independent neighbouring IDAX (CXXC4) ZF-CxxC containing protein for potential recruitment to promoters and CpG islands in genomic DNA (Ko et al. 2013).

Despite controversy regarding their ability to recognize non-methylated DNA, ChIP-Seq profiling revealed that TET1 is preferentially enriched in gene-rich regions, with the highest preference for CGI-promoters of both active and repressed genes (in a manner which positively correlates with the GC density of CGIs) and less intense binding of TET1 throughout gene bodies (Williams et al. 2011; Wu et al. 2011; Xu et al. 2011b). Surprisingly however, only ~ 32% of all CGIs in the genome were detected to overlap with TET1 peaks, suggesting that the ZF-CxxC DNA binding domain may not be the sole determinant of TET1 chromatin binding profile. When the distribution of 5hmC was assessed by using a 5hmC antibody-based DNA immunoprecipitation (hMeDIP) approach, it was evident that 5hmC also localizes to gene-rich regions, but shows a negative correlation with the GC content in CGIs, being instead enriched in CGIs with low or medium GC content, pattern which is surprising considering the association of TET1 with non-methylated GC-rich regions.

Importantly, as 5mC is generally associated with transcriptional repression, 5hmC could therefore be linked to potential transcriptional activation. However, TET1 knockdown experiments in mouse ES cells showed instead that more genes become upregulated than downregulated in the absence of TET1, including many repressed polycomb-occupied developmental regulators, suggesting that the role of this

DNA demethylase is still far from being understood (Wu et al. 2011). Interestingly, in support of TET1 potentially harbouring repressive functions, TET1 was found to recruit the repressive MBD3–NURD (nucleosome remodelling and deacetylase) (Yildirim et al. 2011) and SIN3A (switch-independent 3A) (Williams et al. 2011) complexes to promoters, further complicating the debate of whether it mediates transcriptional activation or silencing.

Assessing the biological relevance of TET proteins is also complicated by the relatively mild phenotype observed for TET1/TET2 double knockout mice, which display reduced 5hmC and abnormal methylation at various imprinted loci, but are viable and otherwise overtly normal (Dawlaty et al. 2013). Although redundancy with TET3 was suggested to explain this result, it is important to consider that TET3 expression seems mostly restricted to the oocyte and zygote (Inoue and Zhang 2011). TET3-null mice on the other hand exhibit maternal dominant lethality due to the role of TET3 in demethylation of the paternal pronucleus at the zygotic developmental stage (Gu et al. 2011). Together, these observations clearly suggest that further work is required to understand the biological relevance of TET-mediated DNA demethylation and to clarify whether 5hmC is only an intermediate in the demethylation reaction or whether it can act like a signalling molecule interpreted by different effector molecules to direct a transcriptional output.

1.5. ZF-CxxC-domain containing proteins recognize non-methylated CpG dinucleotides

1.5.1. Discovery of ZF-CxxC DNA binding domain and ZF-CxxC-containing protein families

In a manner analogous to how methyl-CpG is interpreted by MBD-domain containing proteins, it was speculated that non-methylated CpGs could also be specifically “read” by a distinct set of proteins potentially sharing a recognition domain. The ZF-CxxC DNA binding domain was originally described in mammalian DNMT1 (Bestor and Verdine 1994), in the methyl-CpG binder MBD1/MeCP1 (Cross et al. 1997) and in the human trithorax protein (HRX) (also known as MLL or ALL-1) (Ma et al. 1993).

Interestingly, although its precise function was not clearly understood, it was already speculated to play a role in allowing the proteins harbouring this domain to distinguish the state of methylation of a given target gene. Using a phage-based ligand screen for proteins which can bind to a CpG-motif containing probe, Voo et al. (2000) identified a novel CpG binding protein, hCGBP (human CpG binding protein) which can recognize non-methylated CpGs in a ZF-CxxC dependent manner, and whose binding to the DNA probe is abrogated by cytosine methylation (Lee et al. 2001). Subsequent bioinformatics analysis identified multiple ZF-CxxC domain-containing proteins, the majority of which are associated with chromatin modifying activities, as discussed in greater detail throughout this chapter when the individual proteins were mentioned (Iyer et al. 2011).

The ZF-CxxC DNA binding domain is characterized by eight cysteine residues arranged in two clusters with the consensus sequence of $CxxCxxCX_{4-5}CGxCxxCX_nCxxRxC$ motif (where x is any amino acid, Figure 1.3). These eight cysteine residues bind two zinc ions to form two C4-type zinc fingers, adopting an extended crescent-like structure, as determined by crystallographic and NMR spectrometry studies of the MLL1, DNMT1 and CFP1 ZF-CxxC DNA binding domains either free or in complex with CpG-containing probes (Allen et al. 2006; Cierpicki et al. 2010; Xu et al. 2011a). The linker region between these two cysteine-rich clusters contains a conserved four peptide "KFGG" motif required for protein folding and rigidity of the ZF-CxxC domain fold, followed by a hydrophilic positively charged DNA-binding loop which interfaces the DNA major groove in a wedge-like manner mediating specific contacts via hydrogen-bonding with the CpG-dinucleotide through a KQ/RQ dipeptide motif.

As specificity for CpG recognition is achieved by means of intermolecular hydrogen bonds, methylation of cytosine results in steric clashes with the ZF-CxxC containing protein backbone and loss of hydrogen bonding. Regions flanking the ZF-CxxC domain reach the phosphate backbone in the minor groove, contributing to the overall stability of the binding complex (Figure 1.3, panel A). Therefore, although a small and compact DNA binding module, the ZF-CxxC domain requires access to both major and minor groove for efficient DNA binding, observation which indirectly implies that ZF-CxxC containing proteins cannot access DNA occluded by nucleosomes, but can only interface chromatin in the accessible linker DNA regions, as was demonstrated to be the case for KDM2A (Zhou et al. 2012).

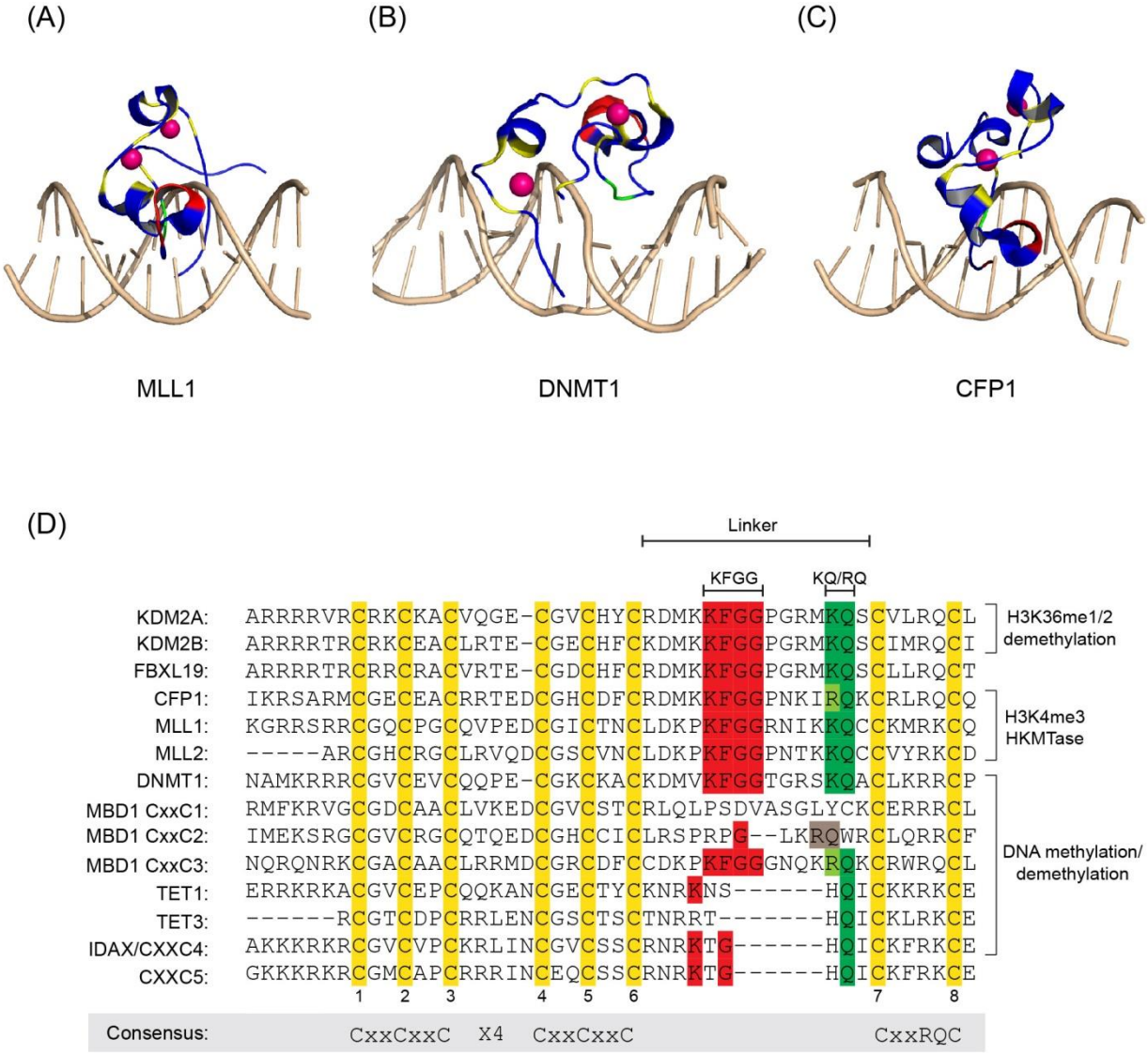


Figure 1.3. The ZF-CxxC domain is employed by multiple proteins to recognize non-methylated CpG dinucleotides.

(A) – (C) Cartoon representation of the crystal structure of mammalian MLL1, DNMT1 and CFP1 ZF-CxxC binding domain in complex with a non-methylated CpG-containing DNA probe. The DNA double helix and protein are shown in wheat and blue, respectively. The colour coding is identical to the one used in panel D, with the KFGG motif shown in red, the RQ/KQ motif in green and the eight conserved cysteine residues in yellow. The two zinc ions (pink) are coordinated by the cysteines within the ZF-CXXC domain. Note the loop from the domain inserting into the major groove of the DNA helix and mediating recognition of the CG base pairs. The crystal structures were obtained from the Protein Data Bank (PDB), with the following accession codes: 2KKF (human MLL1) (Cierpicki et al. 2010), 3PTA (human DNMT1) (Song et al. 2011) and 3QMG (human CFP1) (Xu et al. 2011a), and manipulated using PyMol molecular visualization software.

(D) Multiple sequence alignment of all the ZF-CxxC domains identified in mouse. All sequences were obtained from Uniprot protein database, with the exception of TET3 (isoform HQ423151, GeneBank) and MLL2 (isoform NM_029274, NCBI). The eight conserved cysteine residues (labelled 1 to 8) involved in zinc ion coordination are highlighted in yellow, while the KFGG motif shown to be important for the DNA binding activity of ZF-CxxC domain containing proteins is shown in red, where present. The conserved RQ/KQ motif (highlighted in green, where

identified) makes specific contacts with the CpG-dinucleotide sequence in the DNA substrate. The ZF-CxxC containing proteins differ in the degree of conservation across the linker region, which results in their classification as either type 1, type 2 or type 3 ZF-CxxC domains, as described in the text. Importantly, the majority of ZF-CxxC domain-containing proteins are associated with chromatin-modifying activities, with the exception of FBXL19 and CXXC5, whose biological relevance remains poorly defined.

Differences in the linker region between the two cysteine-rich clusters divide the ZF-CxxC domains into three classes and define their ability to specifically recognize non-methylated DNA (Long et al. 2013a). Currently, 12 ZF-CxxC domain encoding proteins were identified in the mouse genome (Figure 1.3, panel B), with non-methylated CpG DNA binding activity detected *in vitro* and *in vivo* for type-1 ZF-CxxC domain containing proteins KDM2A, KDM2B, CFP1, MLL1/MLL2, DNMT1, MBD1 ZF-CxxC 3, while the ZF-CxxC domain of FBXL19 is only presumed to be functional but has not yet been reported. MBD1 ZF-CxxC 1 and 2 domains lack the conserved KFGG motif as well as the functional DNA binding loop and belong to type-2 ZF-CxxC domains which appear to have lost any CpG-dependent DNA binding activity. In contrast, type-3 ZF-CxxC domain-containing proteins such as TET1/3 and IDAX/CXXC4 also lack a canonical KFGG motif but retain some remnant of the KQ/RQ dipeptide in the form of an "HQ" motif, which presumably may explain why proteins of this class were reported to have less specificity in their substrate binding. The binding activity of type-3 ZF-CxxC domain containing protein CXXC5 remains to be established.

Interestingly, not all of the proteins containing functional ZF-CxxC DNA binding domains occupy all CGI elements in the genome at a given time, raising the question of the complexity of factors which regulate their chromatin localization profile. For example, although harbouring a functional non-methylated DNA binding domain, MLL1 is restricted to binding a subset of CpG islands (Guenther et al. 2005), while CFP1 occupancy appears reduced at Polycomb-occupied CGIs (Thomson et al. 2010) (detailed in Section 1.3.1.2.1). It is intriguing that the majority of ZF-CxxC containing proteins are associated with chromatin state modifying activities, reinforcing the possibility that CGIs function as platforms for assembly of these factors which then influence CGI chromatin architecture potentially as a means of impinging upon gene regulation. The two highly homologous ZF-CxxC-domain encoding proteins KDM2A and KDM2B represent a major focus of this thesis and will be discussed in greater detail in the following section.

1.5.2. The lysine demethylase 2 (KDM2) protein family

1.5.2.1. KDM2A

Using an *in vitro* assay based on detection of [^{3}H]-formaldehyde, one of the predicted release products from demethylation of 1-methyladenine (1meA) and 3-methylcytosine (3meC) in DNA by bacterial AlkB enzyme family, Tsukada et al. (2006) first detected FBXL11/KDM2A/ JHDM1A histone demethylase activity in HeLa cell extracts. This enzyme was previously identified in a search for F-box motif containing components known to interact with the protein SKP1 (S-phase kinase-associated protein 1A) and form E3-ubiquitin ligase complexes called SCFs (Skp, Cullin, F-box containing complex), which generally function in phosphorylation-dependent ubiquitylation of hundreds of substrates primarily for subsequent proteasomal degradation (Cenciarelli et al. 1999; Craig and Tyers 1999; Winston et al. 1999). The Fe (II) and 2OG-dependent enzymatic activity of KDM2A is attributed to its JmjC-catalytic domain and was shown to specifically demethylate *in vitro* and *in vivo* H3K36me2 and to a lesser extent H3K36me1 (Tsukada et al. 2006). In addition to the N-terminal JmjC-catalytic domain and the F-box motif, KDM2A also contains several other domains, such as a ZF-CxxC DNA binding domain, a PHD (Plant Homeo) domain and six C-terminal leucine-rich repeats (LRRs) known to be involved in mediating protein-protein interactions (Kobe and Deisenhofer 1994; Enkhbayar et al. 2004) (Figure 1.4, panel A). Importantly, although the PHD domain has been reported to act as a histone binding module, the contribution of this domain to mammalian KDM2A function is not clear as it does not seem to bind to any known histone modification (Zhou et al. 2012).

Members of the KDM2 protein family are found in organisms from budding yeast to human, with dKDM2/CG11033 (flies), 3H549 (worms), EPE1 (fission yeast) and JHD1 (budding yeast) currently characterized (Klose et al. 2006). Interestingly, the JmjC-domain is highly conserved in terms of catalytic residues across these orthologues, with the exception of EPE1, where a mutation in one of the Fe (II)-binding residues renders the enzyme inactive *in vitro* (Tsukada et al. 2006), although it seems to play an essential role in regulating chromatin organization by antagonizing heterochromatinization in a JmjC-dependent manner (Ayoub et al. 2003; Wang et al. 2013). While the fly orthologue (discussed more in

the next section) has an almost identical domain architecture, with the exception of the PHD domain, the budding yeast JHD1 protein only retains the PHD and JmjC-catalytic domain and was shown to function with the RPH1 H3K36me3-demethylase to increase RNA pol II occupancy at transcribed regions by antagonizing SET2-mediated H3K36me2/me3 and subsequent RPD3S-dependent transcriptional repression (Fang et al. 2007; Kim and Buratowski 2007; Klose et al. 2007a).

The function of KDM2A in mammals remains insufficiently characterized, especially due to lack of mouse knock-out models. However, reports have implicated this protein in acting as a negative regulator of replicative senescence and promoting immortalization of mouse embryonic fibroblasts (MEFs) in culture in a manner dependent on its JmjC-enzymatic activity, although the mechanisms behind this are unclear (Pfau et al. 2008). Furthermore, KDM2A activity was proposed to be stimulated by starvation to repress transcription of ribosomal RNA genes by binding to rDNA promoters and demethylating H3K36me1/2 (Tanaka et al. 2010). Nevertheless, KDM2A does not encode a canonical nucleolar localization signal (NoLS) and the signalling pathways by which the enzymatic activity of KDM2A is regulated remain undefined. Interestingly, another report suggested that KDM2A associates with heterochromatin and interacts with HP1-α and HP1-γ to mediate silencing of satellite RNAs that are embedded within centromeric repeats in a JmjC-domain dependent manner (Frescas et al. 2008). Consistent with this, KDM2A was reported to bind H3K9me3-modified nucleosomes through HP1 (Bartke et al. 2010), although the domain responsible for this interaction was not dissected. In addition, recruitment of KDM2A to H3K9me3 nucleosomes in the presence of HP1 was abolished by DNA methylation, suggesting that the association between KDM2A and HP1 may not be the major targeting activity for the demethylase.

KDM2A was also suggested to demethylate substrates other than the canonical H3 histone tail, showing specificity for K218/K221 residues on the p65 subunit of the transcription factor NF-κB, thereby acting as a potent negative regulator of NF-κB–dependent gene expression (Lu et al. 2013).

Despite this multitude of potential functions, the question of how KDM2A is recruited to its target genes remained unanswered. Considering that it harbours a ZF-CxxC DNA binding domain known in other proteins to enable recognition of non-methylated CpGs, Blackledge et al. (2010) showed that the ZF-

CxxC domain of KDM2A is functional *in vitro* and then analysed the genome-wide localization of the demethylase. ChIP-Seq studies revealed that KDM2A binds to over 90% of CpG island elements genome-wide, regardless of the transcriptional state of the CGI-associated gene, in a manner similar to the chromatin binding profile of another ZF-CxxC containing protein, CFP1 (Thomson et al. 2010). In agreement with KDM2A acting as an H3K36-specific demethylase, knockdown of the enzyme in HeLa cells resulted in a significant increase in H3K36me2 levels specifically at CpG island elements.

Interestingly, H3K36me1/me2 is a highly abundant mark in mammalian cells, found on approximately 30-50% of histone H3 and was shown to coat pervasively both intra- and inter-genic regions (Garcia et al. 2008; Kuo et al. 2011). Given the transcriptionally inhibitory function associated with H3K36 methylation in yeast (see Section 1.3.1.2.3), it is therefore tempting to speculate that KDM2A nucleation at CpG islands functions as a mechanism to specifically remove this repressive mark from these genomic elements, thereby creating a permissive CGI chromatin architecture to facilitate transcription initiation. However, KDM2A depletion by shRNA-based approach in transformed cells caused only minor changes in gene expression (Blackledge et al. 2010), suggesting that KDM2A occupancy *per se* does not directly affect the transcriptional output, but that it potentially plays more subtle roles in facilitating a normal transcriptional profile. A clearer understanding of the biological relevance of KDM2A would be greatly facilitated by a genetic ablation system in which to check whether deletion of KDM2A affects the gene regulatory functions of CpG islands.

1.5.2.2. KDM2B

In addition to KDM2A, the highly related protein KDM2B (also called FBXL10/JHDM1B) was identified to be an H3K36 demethylase with specific enzymatic activity towards H3K36me2 and H3K36me1 modification states *in vitro* and *in vivo* (Tsukada et al. 2006; He et al. 2008). KDM2A and KDM2B have a highly similar domain architecture, with the latter protein also harbouring a conserved ZF-CxxC DNA binding domain shown in KDM2A to mediate genome-wide recognition of non-methylated CGIs (Figure

1.4, panel A). Interestingly, both the KDM2A and KDM2B genes also contain alternative transcription start sites downstream of their catalytic JmjC domain, giving rise to short forms of these proteins which retain an intact ZF-CxxC domain, although the biological relevance of these short isoforms is unclear.

Multiple lines of evidence support a ZF-CxxC-dependent DNA binding activity of KDM2B. For example, Koyama-Nasu et al. (2007) reported that KDM2B specifically co-immunoprecipitates with the C-JUN transcription factor and localizes to the *c-jun* gene promoter in a ZF-CxxC dependent manner, where it potentially functions to silence *c-jun* transcription by recruitment of a SIN3A/HDAC repressor complex. In addition, Yamagishi and colleagues (2008) suggested that KDM2B may play a role in *Hox* gene regulation by specifically binding to the *Hoxb9* promoter via the ZF-CxxC DNA binding domain. Moreover, KDM2B has been extensively associated with the control of cellular proliferation and replicative senescence, with KDM2B binding at the $p15^{Ink4b}$ tumour-suppressor gene promoter in primary mouse embryonic fibroblasts (MEFs) to silence $p15^{Ink4b}$ expression in a manner dependent on its H3K36 demethylase activity (He et al. 2008; Pfau et al. 2008). These reports are particularly interesting, considering that the KDM2B gene targets identified in these studies are associated with CpG island elements, suggesting that the ZF-CxxC domain of KDM2B can specifically recognize non-methylated DNA *in vivo*.

The nuclear function of KDM2B was questioned by Frescas et al. (2007), who used protein overexpression studies in HeLa cells to show that KDM2B localizes to the nucleolus through a putative nucleolar localization signal (NoLS), where it contributes to repression of ribosomal RNA genes by demethylating H3K4me3. However, multiple other studies fail to observe this nucleolar localization staining pattern (He et al. 2008; Pfau et al. 2008; Janzer et al. 2012), although some reports indicate a potential H3K4me3-specific demethylase activity for KDM2B at least *in vitro* towards bulk histones (Janzer et al. 2012).

Interestingly, two independent studies have identified KDM2B in association with Polycomb group proteins in transformed mammalian cells (Gearhart et al. 2006; Sanchez et al. 2007). In HeLa and HEK293 cells, epitope-tagging and purification of the BCL6 co-repressor protein BcoR identified a BcoR complex that contains KDM2B and PRC1 members RING1A/RING1B, PCGF1 and RYBP (Gearhart et al.

2006). Importantly, this complex was distinct from those formed by the other BCL6 co-repressors NCOR and SMRT, which were reported to silence transcription of BCL6 target genes via their association with HDAC3 and the JmjC-containing H3K9/H3K36me3 lysine demethylase KDM4A/JMJD2A (Li et al. 2000; Wen et al. 2000; Berry and Janknecht 2013).

As the BcoR-KDM2B complex was shown to possess E3 ligase activity for histone H2A, it was hypothesized that the complex is recruited to BCL6 targets where it mediates silencing of the associated genes by the combination of Polycomb-dependent H2AK119ub1 and the H3K36 demethylase activity of KDM2B. However, when members of this complex were found to co-localize with mono-ubiquitylated H2A at several BCL6 targets, the contribution of the ZF-CxxC binding domain of KDM2B was not investigated. Instead, the direct interaction between BCOR and sequence-specific DNA-binding protein BCL6 was considered to be responsible for the DNA targeting activity. Using a single-step purification of *in vivo* biotinylated RING1B, Sanchez et al. (2007) also identified KDM2B and BCOR as major partners of RING1B in a murine erythroleukemia (MEL) cell line and showed that PCGF1 and KDM2B enhanced the H2A ubiquitylating activity of RING1B.

A similar complex containing the fly KDM2B homologue, dKDM2, in association with the sole PRC1 E3 ligase dRING and a PCGF1 homologue PSC, together with dRAF2 (CG4877) and two nuclear pore associated proteins MTOR/ULP1 was purified from Drosophila embryos and called dRAF (dRING-associated factors complex) (Lagarou et al. 2008). Although the contribution of MTOR/ULP1 and CG4877 to the function of this complex remains to be established, it is clear that dRAF is distinct from the classical PRC1 complex containing the PC (Polycomb) and PH (Polyhomeotic) subunits. Furthermore, RNAi-mediated depletion of dKDM2 in S2 cells resulted in a dramatic loss of histone H2AK118ub1, suggesting that in addition to H3K36 demethylase activity, dRAF is more critical for bulk cellular histone H2A ubiquitylation levels than classical PRC1. Finally, two out of three mutant fly lines harbouring distinct P-element insertions in the *dkdm2* gene were homozygous lethal, while the third was a homozygous viable hypomorph. Although none displayed obvious homeotic transformations, *dkdm2* mutant alleles significantly increased the frequency of homeotic transformations in *Pc* heterozygotes flies, underlining the contribution of dKDM2 to silencing of the homeotic loci *in vivo*.

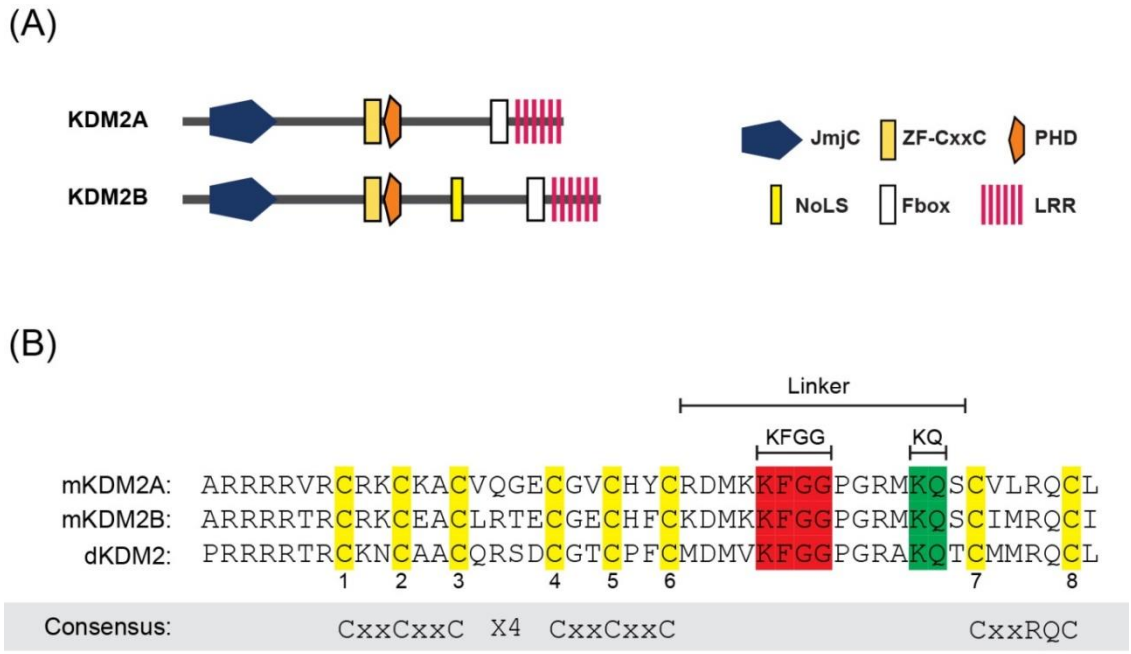


Figure 1.4. KDM2A and KDM2B are highly homologous histone lysine demethylases.

(A) Schematic representation of human KDM2A and KDM2B illustrating their similar domain architecture (JmjC: Jumonji C catalytic domain; ZF-CxxC: CxxC zinc finger; PHD: plant homeodomain; FBOX: F-box domain; LRRs: leucine-rich repeats). The location of the putative nucleolar localisation signal is indicated between the PHD and FBOX domains of KDM2B (NoLS: putative nucleolar localization signal).

(B) Multiple sequence alignment comparing the ZF-CxxC domain of mouse KDM2A/2B protein family with that of the Drosophila dKDM2 homologue (CG11033). All sequences were obtained from Uniprot protein database. The eight conserved cysteine residues (labelled 1 to 8) involved in zinc ion coordination are highlighted in yellow, the KFGG motif shown to be important for the DNA binding activity of ZF-CxxC domain containing proteins is shown in red and the conserved KQ motif mediating CG-specific recognition is highlighted in green. Given the highly conserved sequence of the DNA binding domain, it is tempting to speculate that the fly dKDM2 protein encodes a presumably functional ZF-CxxC domain.

An interesting observation came from alignment of the dKDM2 ZF-CxxC domain with that of its mammalian homologues KDM2A/2B (Figure 1.4, panel B), which showed significant sequence conservation across functional residues, thereby suggesting that dKDM2 retains a presumably functional ZF-CxxC DNA binding domain. As flies lack significant levels of genomic DNA methylation and associated DNA methyltransferase homologs (Goll and Bestor 2005; Lee et al. 2010), the observation that dKDM2 may recognize CpG dinucleotides is striking and its relevance remains to be investigated in the context of dRAF-mediated transcriptional regulation.

KDM2A and KDM2B share a highly similar domain architecture. Therefore, one possibility is that the two enzymes may play redundant biological roles. However, given its unique association with Polycomb group proteins, together with the observations that a mouse knockout model (in which only the long form of KDM2B was deleted) and a naturally occurring bovine mutant exhibit developmental defects (Fukuda et al. 2011; Testoni et al. 2012), suggest that KDM2B may play unique biological roles. Considering that KDM2B was reported to have both oncogenic and tumour suppressor properties in a series of conflicting reports (Suzuki et al. 2006; Pfau et al. 2008) and to contribute to leukemic development and maintenance in a mouse AML model (He et al. 2011), a clearer understanding of how this protein exerts its function and whether it recognizes non-methylated CpG island DNA will provide the framework for more detailed investigation into how the properties of KDM2B are coupled to transcriptional regulation.

1.6. Polycomb groups (PcG) proteins and their role in gene silencing

1.6.1. Discovery of Polycomb group proteins

The insect body is composed of a series of metameric segments, with individual segmental identity determined during early development by the selective expression of homeotic genes of the antennapedia and bithorax (ANT-C and BX-C) complexes (Lewis 1978; Ingham 1985; Lehmann and Nusslein-Volhard 1987; Moazed and O'Farrell 1992; Bienz and Muller 1995). A major subset of homeotic genes is represented by the *HOX* cluster, which encodes homeodomain-containing transcription factors that are expressed in a spatially restricted pattern throughout development and specify the anterior-posterior axial patterning in all animals with bilateral symmetry (Akam 1987; Garcia-Fernandez 2005; Mallo et al. 2010). Inappropriate expression of *Hox* genes results in dramatic homeotic transformations, whereby one body segment is transformed to resemble the identity of another, with striking examples including the *Antennapedia* fly mutant which has mesothoracic legs instead of antennae (Struhl 1981b) or four-winged flies in which disruption of the *Hox* gene *Ultrabithorax* (*Ubx*) leads to transformation of

haltere tissue into wing tissue (Lewis 1978; Kerridge and Morata 1982) . Although the boundaries of *Hox* genes transcription are established by segmentation proteins (encoded by the gap, pair-rule or segment polarity genes) early in embryogenesis, the patterns of *Hox* expression are maintained throughout the development of the fly, after the initiating transcription regulators have disappeared.

Genetic approaches led to the identification of the first two Polycomb group (PcG) proteins, Polycomb (PC) and Extra Sex Combs (ESC) (Lewis 1978; Struhl 1981a; Sato and Denell 1985; Struhl and Akam 1985). Homozygous fly mutants show embryonic lethality and display classic homeotic defects characterized by transformation of anterior segments to resemble more posterior body parts. In *esc*- embryos, the initial pattern of *Ubx* transcripts was identical to wild-type, but later showed abnormal ectopic *Ubx* expression indiscriminately in most or all of the segment primordia. Therefore, it was suggested that PcG proteins act as negative regulators of *Hox* gene expression by maintaining them in a silent state in those segments where they were not expressed during early development (Struhl and Akam 1985). Similarly, Trithorax group (TrxG) proteins were characterized as antagonists of PcG proteins and positive regulators of *Hox* genes, with mutations in TrxG subunits resulting in transformation of posterior to anterior structures, caused by reduced *Hox* gene expression (Ingham 1983; Schuettengruber et al. 2011; Lanzuolo and Orlando 2012). Although it was initially speculated that PcG proteins dictate *Hox* gene expression by forming a gradient in a manner analogous to the original segmentation genes, subsequent studies suggested instead that PcG and TrxG factors bind to specific genomic elements comprising both *Hox* and other homeotic genes and function more as “cellular memory” of gene activity in early development (Zink and Paro 1989; Chan et al. 1994; Ringrose and Paro 2004). Importantly, many more PcG proteins have subsequently been identified and homologues of fly PcG members were detected in vertebrates, worms and plants, indicative of an essential role in regulation of gene expression during development and differentiation of multicellular organisms.

1.6.2. Polycomb repressive complex 1 (PRC1)

1.6.2.1. Classical PRC1 in flies and mammals

The classical Drosophila core PRC1 complex consists of four PcG proteins, namely Polycomb (PC), Polyhomeotic (PH), Posterior Sex Combs (PSC) and Sex Combs Extra (SCE)/dRING (Figure 1.5, panel A) (Strutt and Paro 1997; Shao et al. 1999; Francis et al. 2001; Saurin et al. 2001). In addition to these stoichiometric components, studies using an epitope-tagging strategy to purify PRC1 from 0-12hr Drosophila embryos indicated the presence of sub-stoichiometric amounts of other factors such as Sex Combs on Midleg (SCM), the sequence-specific DNA binding factor ZESTE or the general transcription machinery dTAFII proteins, although the stability and relevance of these associations to polycomb function are unclear (Shao et al. 1999; Saurin et al. 2001). In contrast, mammalian genomes encode two SCE homologues RING1 (RING1A) and RNF2 (RING1B), three PH homologues (PHC1, 2, 3), at least five PC counterparts (CBX2, 4, 6, 7 and 8) and six PSC homologues (PCGF1, PCGF2 (MEL18), PCGF3, PCGF4 (BMI1), PCGF5 and PCGF6) (Figure 1.5, panel B) (Simon and Kingston 2009; Vandamme et al. 2011) .

Despite significant diversity among PRC1 complexes in mammals due to the existence of multiple paralogues of each PcG gene, Flag-tag based purification of PRC1 from HeLa cells yielded an essentially identical core complex to the fly counterpart, suggesting that flies and mammals are potentially using a similar mechanism to maintain homeotic gene expression patterns through cell divisions (Levine et al. 2002). This conservation of core PRC1 composition and function is further supported by the observation that mouse PC homologue M33/CBX2 can rescue fly *Pc*- mutants, albeit at less efficient levels (Muller et al. 1995). Although the core classical PRC1 complex generally retains the four-membered structure observed in flies, the diversity of mammalian homologues for each subunit complicates understanding of how the identity of individual PRC1 complexes is specified. Moreover, reports indicate that the composition of mammalian PcG complexes varies between different cell types and during differentiation, suggesting that the different homologues may not be entirely redundant (Jacobs and van Lohuizen 2002).

In agreement with non-overlapping biological roles of the different mammalian homologues, *Mel18/Pcgf2* and *Bmi1/Pcgf4* deficient mice are viable and display similar posterior transformations of the axial skeleton but also unique phenotypes such as neurological defects in *Pcgf4-/-* mice compared to defects in smooth muscle of the colon in *Pcgf2-/-* mice (Akasaka et al. 1996; van der Lugt et al. 1996; Akasaka et al. 2001). Furthermore, the CBX family proteins were proposed to define distinct PRC1 protein complexes (Vandamme et al. 2011), with CBX7 playing an essential role in dictating ESC pluripotency, while Cbx2 and Cbx4 are required for unique, non-overlapping aspects of early lineage specification (Morey et al. 2012; O'Loghlen et al. 2012). The mechanisms by which these distinct PRC1 complexes are regulated and the interplay between them remains however poorly defined and would require a systematic genetic ablation approach to dissect in greater detail.

The catalytic core of PRC1 is formed by RING1A/RING1B, subunits which encode an essential E3 ubiquitin ligase activity towards the H2AK119 residue (H2AK118 in flies) (Wang et al. 2004a; Cao et al. 2005). The enzymatic activity of RING1A/1B is stimulated through binding via RING-finger domains to PCGF proteins BMI1/PCGF4 (Li et al. 2006) and MEL18/PCGF2 (Elderkin et al. 2007). Importantly, H2AK119ub1 was found to be enriched on PRC1 targets, such as *Hox* genes or the inactive X-chromosome, suggesting that this histone modification may contribute to the mechanism by which PcG proteins regulate gene expression. In agreement, homozygous flies with a mutation abolishing the H2A ubiquitin ligase activity of dRING *in vitro* show misexpression of *Ubx* and *Abd-B* homeotic genes in imaginal discs (Fritsch et al. 2003), although later studies complicated interpretation of these findings through the discovery that the SCE^{R65C} protein encoded by this mutation is not stable and functions instead as a null disruption and not a site-specific amino acid exchange (Gutierrez et al. 2012). However, experiments expressing wild-type and catalytically inactive versions of RING1B in RING1A/1B double knock-out ES cells indicate that the E3 activity of RING1B is required for PcG-mediated repression of target genes and for the maintenance and differentiation of ESCs (Endoh et al. 2012).

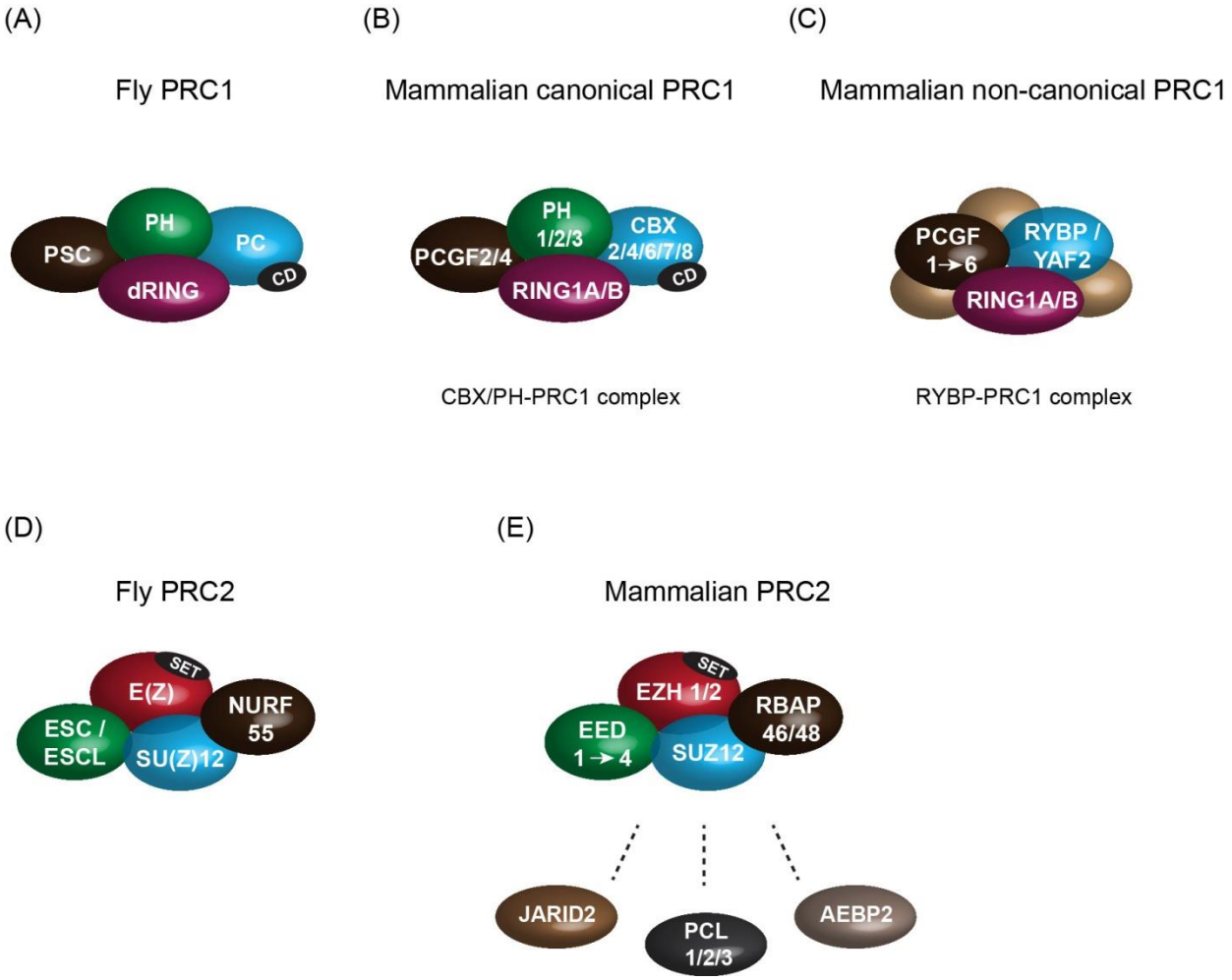


Figure 1.5. PRC1 and PRC2 subunit composition is conserved between flies and mammals.

(A) PRC1 in flies is composed of the four core subunits Polycomb (PC) which contains a chromodomain (CD) responsible for recognition of H3K27me3, Polyhomeotic (PH), Posterior Sex Combs (PSC) and Sex Combs Extra (SCE)/dRING. The dRING protein is responsible for the E3 ubiquitin ligase activity.

(B) Although the core subunit architecture is conserved in mammalian canonical PRC1, there is significant diversity due to the existence of multiple homologues for each PcG gene. Mammalian genomes encode two SCE/dRING homologues RING1 (RING1A) and RNF2 (RING1B), three PH homologues (PHC1, 2, 3), at least five PC counterparts (CBX2, 4, 6, 7 and 8) and six PSC homologues (PCGF1, PCGF2 (MEL18), PCGF3, PCGF4 (BMI1), PCGF5 and PCGF6).

(C) In non-canonical PRC1, the CBX subunit is replaced with RYBP/YAF2 and the PH component is absent. Proteomic studies suggest that each PCGF protein interacts with RING1A/1B in a mutually exclusive manner to define distinct variant PRC1 complexes characterized by a unique set of associated polypeptides (Gao et al. 2012).

(D) PRC2 in flies consists of the four core members Enhancer of Zeste E(Z), Extra Sex Combs ESC or the potentially redundant Extra Sex Combs-like ESCL, Suppressor of Zeste 12 homologue SU(Z)12 and the nucleosome remodelling factor p55/NURF-55. The E(z) subunit contains a SET-domain and is responsible for the H3K27 methyltransferase activity.

(E) In mammals, the PRC2 core subunit composition is conserved relative to flies, with much less subunit diversification compared to PRC1. However, there are two highly homologues E(Z) proteins, EZH1 and EZH2, as well as four translational EED mammalian isoforms and two closely related mammalian RB binding proteins , RbAp 48 and RbAp 46. In addition to the core subunits, a number of other proteins were reported to associate with

PRC2, among which JARID2, AEBP2 and one of the three fly PCL homologues (PCL1, 2 and 3). The importance of these factors to Polycomb function awaits further clarification.

The two other members of the classical core PRC1 complex, namely the PH and CBX subunits, also contain functional domains reported to be essential for PcG function. The PC protein and the corresponding mammalian CBX homologues are characterized by the presence of a conserved N-terminal chromodomain (CD) which was shown to be required for PC binding to polytene chromosomes (Messmer et al. 1992) and to specifically recognize H3K27 methylation catalysed by the PRC2 PcG complex (Cao et al. 2002). Finally, although *rae28-/-* (HPH1) homozygous mice showed perinatal lethality, posterior skeletal transformations and multiple developmental defects (Takihara et al. 1997), the contribution of the polyhomeotic PH component to PcG-mediated gene silencing remains incompletely understood. Interestingly, both the fly and its mammalian counterparts contain a Sterile Alpha Motif (SAM) domain known to be involved in protein-protein interactions and to participate in homo- and hetero-typic associations with other SAM-containing factors (Peterson et al. 1997; Kyba and Brock 1998; Kim et al. 2002). Two recent reports suggest that the ability of PH subunits to polymerize via their SAM domains is essential for PcG-mediated transcriptional repression in flies and mammals. According to these studies, overexpression of SAM domain mutants in Drosophila wing disc resulted in a derepression of the *Abd-B* homeotic gene (Robinson et al. 2012), while MEFs with knock-in of different PH2-SAM domain mutations indicated that the polymerization capacity of PH2-SAM is crucial for PRC1 clustering, stabilization of PRC1 and PRC2 binding, polycomb domain formation *in vivo* and robust gene silencing of developmental loci (Isono et al. 2013).

1.6.2.2. Non-canonical/variant PRC1 complexes in mammals

In Drosophila, the dRING-PSC E3 ubiquitin ligase catalytic core interacts with the Polycomb (PC) and Polyhomeotic (PH) subunits to form the canonical PRC1 complex. In mammals however, the discovery

from structural studies that RING1B can associate with CBX (mammalian PC orthologues) and RYBP/YAF2 proteins in a mutually exclusive manner suggested the existence of non-canonical PRC1 complexes (Figure 1.5, panel C) (Wang et al. 2010). Importantly, a recent comprehensive proteomic study in human 293T cells showed that, although RYBP/YAF2 co-purifies with all the six mammalian PSC homologues (PCGFs), only PCGF2 (MEL18) and PCGF4 (BMI1) have the capacity to interact with CBX/PH members and define canonical PRC1 complexes, although how this unique specificity is achieved remains poorly understood (Gao et al. 2012). Furthermore, the same study suggested that each PCGF associates with RING1B in an exclusive manner to define distinct PRC1 variant complexes, characterized by different associated polypeptides and potentially different chromatin-binding profiles depending on the subunit composition of the individual complexes (Gao et al. 2012).

In the case of canonical PRC1, the chromodomain of the CBX subunit is believed to mediate crosstalk with PRC2 through recognition of H3K27me3 and be responsible for the hierarchical recruitment of PRC1 to target loci. The factors which define the genomic localization of RYBP/YAF2-containing non-canonical PRC1s are currently unknown. However, it is becoming increasingly clear that the chromatin binding of variant PRC1 complexes is largely independent of PRC2-mediated H3K27 methylation. For example, ES cells lacking the PRC2 component EED and which are devoid of detectable H3K27me3 failed to recruit the PH2 polyhomeotic PRC1 subunit, but otherwise showed normal levels of H2AK119ub1 and recruitment of RING1B to the inactive X-chromosome (Schoeftner et al. 2006). Furthermore, genome-wide ChIP-Seq analysis suggested that in *Eed-/-* ES cells, low levels of RING1B were readily detectable at the majority of PcG target loci and this was sufficient to maintain normal global levels of H2AK119ub1 (Tavares et al. 2012). Importantly, CBX7 chromatin occupancy was drastically reduced in the absence of a functional PRC2, while enrichment of RYBP over TSS of RING1B targets was detected in both *Eed+/+* and *Eed−/−* mESCs, suggesting H3K27me3-independent recruitment of variant RYBP-containing PRC1 complexes.

It would be tempting to speculate that RYBP or the closely related protein YAF2 may function as targeting modules themselves, especially considering that RYBP interacts *in vitro* with YY1, the mammalian homologue of Drosophila sequence-specific DNA-binding PcG proteins PHO and PHO-L

(Garcia et al. 1999; Hisada et al. 2012). However, RYBP-deficient ESCs show unimpaired chromatin binding of PCGF2 and RING1B and unaffected ES cell self-renewal capacity. This suggests that the contribution of RYBP to silencing of PcG targets is modest and that RYBP/YAF2 are unlikely to play a major role in chromatin recruitment of variant PRC1 complexes (Hisada et al. 2012). Instead, they are most probably subservient to targeting by specific variant PRC1-associating polypeptides. Two such complexes have already been described, the one formed by PCGF1 containing KDM2B and BcoR (described in section 1.5.2.2), and a distinct complex formed by PCGF6. The PCGF6 subunit was reported to interact with the transcriptional repressor E2F6, the Malignant Brain Tumour (MBT) domain-containing protein L3MBTL2, HP1 proteins and histone deacetylases HDAC1/2, as well as the specific DNA-binding factors MAX/MGA (MAX-gene associated) to define a complex important in pluripotent stem cells and early development (Trojer et al. 2011; Qin et al. 2012).

The identification of distinct PRC1 complexes that show PRC2-dependent and independent chromatin binding profiles, together with an understanding of the unique composition of the different variant PRC1 complexes and their potential associated DNA-binding activities provides essential insight into the complexity of Polycomb-mediated gene regulation and remains an area of intense research.

1.6.3. Polycomb repressive complex 2 (PRC2)

1.6.3.1. Core PRC2 in flies and mammals

Drosophila and mammalian PRC2 complex consists of four core proteins, namely Enhancer of Zeste E(Z) (EZH1/EZH2 in mammals), Extra Sex Combs ESC or the potentially redundant Extra Sex Combs-like ESCL (Kurzhals et al. 2008) (corresponding to mammalian Embryonic Ectoderm Development EED1-4), Suppressor of Zeste 12 homologue SU(Z)12 (SUZ12 in mammals) and the nucleosome remodelling factor p55/NURF-55 (RbAp46/48, also known as RBBP7/4, in mammals) (Figure 1.5, panel D) (Cao et al. 2002; Czermin et al. 2002; Kuzmichev et al. 2002; Muller et al. 2002). The discovery that E(Z) contains a SET-histone methyltransferase domain capable of catalysing all three methylation states of H3K27, together

with observations indicating that mutations in the E(Z) SET-domain disrupt *Hox* gene repression in imaginal discs *in vivo* suggested that histone methylation plays an essential role in PcG-mediated gene regulation (Muller et al. 2002). In agreement with the importance of H3K27 methylation for Polycomb repression in flies, Drosophila larvae in which wild-type nucleosomes were replaced by $H3^{K27R}$ mutant nucleosomes show derepression of PcG target genes Ultrabithorax (*Ubx*), Abdominal-B (*Abd-B*), Sex combs reduced (*Scr*) and Engrailed (*En*) in imaginal wing disc cells, together with other segmental transformations due to misexpression of *Hox* genes, a phenotype which reproduces that observed in clones lacking the E(Z) PRC2 subunit (Pengelly et al. 2013).

Interestingly, EZH2 by itself does not demonstrate significant methyltransferase activity, but requires the presence of EED and SUZ12 members for catalytic activity (Cao and Zhang 2004b; Pasini et al. 2004). In addition to the C-terminal SET-domain, EZH1/EZH2 contains two SANT domains implicated in histone-tail binding (Boyer et al. 2004) and interaction with SUZ12, as well as a Pre-SET/CxC-motif which is believed to stabilize the catalytic domain (Ciferri et al. 2012). The two mammalian homologues EZH1 and EZH2 are not considered entirely redundant, especially as they seem to have different expression patterns, with EZH2 generally expressed in actively dividing cells, while EZH1 seems associated with both dividing and differentiated cells (Margueron et al. 2008). Although EZH1-PRC2 and EZH2-PRC2 were reported to co-occupy a similar set of target genes, the complex containing EZH1 was shown to possess less efficient H3K27 methyltransferase catalytic activity, further reinforcing the differences between the two related proteins. However, EZH2-null ESCs have a less severe phenotype than EED-null cells, potentially due to redundancy with EZH1, indicating that under certain conditions, EZH1 can partially complement EZH2 (Shen et al. 2008).

The four translational EED mammalian isoforms contain seven WD40 repeats which form a characteristic β-propeller structure, suggested to serve as a platform for protein-protein interactions, among which is EZH2 itself. Furthermore, EED was shown to interact with H3K27 trimethylated histone tails and promote allosteric activation of the PRC2 catalytic core towards unmodified H3K27 residues, thereby potentially acting as a mechanism of propagating this histone mark (Margueron et al. 2009). The SUZ12 component contains a zinc finger domain and the VRN2-EMF2-FIS2-SUZ12 (VEFS) domain,

which is required for interaction with EZH2 and was found to be essential for mediating inhibition of PRC2 by activating histone modifications such as H3K4me3 and H3K36me2/3 (Schmitges et al. 2011). In addition, SUZ12 was also reported to directly bind the long non-coding RNA ANRIL and recruit PRC2 to silence the *p15*^{INK4B} tumour suppressor gene (Kotake et al. 2011), although the significance of this association needs further investigation. Deletion of EZH2, SUZ12 or EED in mice results in lethality at early postimplantation stages (~ E7.5-8.5), underlining the importance of PRC2 core components to normal embryonic development (Faust et al. 1998; O'Carroll et al. 2001; Pasini et al. 2004; Leeb et al. 2010).

In contrast to the other three proteins which are unique to PRC2, NURF55 is also found in other chromatin-modifying complexes, including the ATP-dependent nucleosome-remodelling complexes NURF and NuRD (Suganuma et al. 2008). NURF55 also contains seven WD40 repeats and was shown to associate with SUZ12 and contribute to the nucleosome binding capacity of PRC2 (Nekrasov et al. 2005; Nowak et al. 2011). Moreover, its WD40 propeller structure was reported to bind unmodified H3 tails, an interaction which is abrogated by H3K4me3, further contributing to inhibition of PRC2 by active chromatin marks (Schmitges et al. 2011).

1.6.3.2. PRC2-associating proteins in mammals

In addition to core subunits, a number of other proteins were found associated with PRC2, such as AEBP2, JARID2 or PCLs (Figure 1.5, panel E), polypeptides which are especially interesting considering that they can either stimulate PRC2 catalytic activity or were reported to show potential DNA-binding specificity.

Adipocyte enhancer-binding protein 2 (AEBP2) was first isolated in phage screening assays based on its ability to bind to the AE-1 sequence in the proximal promoter region of the *aP2* gene, which encodes an adipose fatty acid-binding protein (He et al. 1999). Further characterization revealed that AEBP2 contains three tandemly repeated sequence motifs related to the Gli-Krüppel (Cys_2-His_2)-type zinc finger

and that it appears to function as a transcriptional repressor in transient transfection assays. Later purifications identified AEBP2 as a substoichiometric component of PRC2, which is important for optimal enzymatic activity, most probably as a consequence of its stabilizing effect on the PRC2 complex, given its extensive physical interactions with EZH2, EED, SUZ12 and RBAP48 (Cao et al. 2002; Cao and Zhang 2004b; Ciferri et al. 2012). In agreement with inclusion of AEBP2 in a PRC2 complex, AEBP2 and SUZ12 were shown to co-occupy a subset of PcG targets, although ChIP-Seq-type experiments have not been reported for this factor (Kim et al. 2009). Given its reported DNA-binding capacity, it remains tempting to speculate that AEBP2 may play a role in PRC2 recruitment to target loci, although a precise characterization of its DNA-binding specificity is currently lacking. Interestingly, the Drosophila homologue JING also partially functions as a transcriptional repressor and was reported to genetically interact with *Pc* , while mice homozygous for a truncation of *Aebp2* showed embryonic lethality (Culi et al. 2006; Kim et al. 2011), further underlining the need to understand the contribution of AEBP2 to PcG-mediated gene regulation.

Jumonji- and ARID-domain-containing protein 2 (JARID2) was also identified as a PRC2 component and was reported to either stimulate or repress its methyltransferase enzymatic activity (Peng et al. 2009; Shen et al. 2009; Landeira et al. 2010; Li et al. 2010; Pasini et al. 2010a). Importantly, although the founding member of the Jumonji C (JmjC) domain protein family of histone lysine demethylases, JARID2 lacks the essential residues for cofactor Fe(II) binding and is assumed to be catalytically inactive (Klose et al. 2006). In addition to the JmjC domain, JARID2 also contains a conserved ARID domain, which was reported to be important for its DNA-binding activity (Kim et al. 2003) and localization to target genes (Pasini et al. 2010a).

Furthermore, when JARID2 DNA-binding properties were investigated by the SELEX method, sequencing of the selected DNA suggested that JARID2 binds DNA without clear specificity, but shows a slight preference for GC-rich sequences (Li et al. 2010). Importantly, ChIP-Seq analysis indicated that JARID2 binds to more than 90% of PRC2-occupied targets and that artificial GAL4-JARID2 targeting is sufficient to recruit PRC2 and mediate H3K27me3-deposition at a heterologous promoter (Li et al. 2010; Pasini et al. 2010a). Depletion or deletion of JARID2 was shown to result in reduced enrichment of core PRC2

components to targets and subsequent loss of H3K27me3, although conflicting reports suggest unaffected histone modification levels in the absence of JARID2. Interestingly, *Jarid2* knockout mESCs show impaired differentiation towards neural or mesodermal fates (Landeira et al. 2010), underlining an essential contribution of this protein to normal development. Although the molecular mechanisms behind a potential JARID2-mediated recruitment of PRC2 remain undefined, a recent report suggested instead that JARID2 binds long-noncoding RNAs (lncRNAs) and that this interaction facilitates JARID2-PRC2 association and contributes to the recruitment and assembly of PRC2 on chromatin (Kaneko et al. 2014), further implicating JARID2 as a targeting module for PRC2.

Drosophila Polycomb-like (PCL) and the three mammalian homologues PCL1/PHF1, PCL2/MTF2 and PCL3/PHF19 were also reported to exist in a stable complex with PRC2 and to localize at PREs in Drosophila (Papp and Muller 2006) and at PcG-targets in mammals (Walker et al. 2010; Casanova et al. 2011; Hunkapiller et al. 2012). Removal of *Pcl* in fly embryos or in imaginal discs resulted in reduced but not abolished levels of Su(z)12 binding and H3K27me3, suggesting that in Drosophila, PCL is important to generate high levels of H3K27 trimethylation at PcG-occupied genes but cannot be a central PRC2 targeting mechanism (Nekrasov et al. 2007). In mammals, PCL2 knockdown impaired PRC2 recruitment to the inactive X in differentiating XX ES cells and also PRC2 occupancy at target genes in ESCs, but the levels of H3K27me3 appeared largely unaffected (Casanova et al. 2011). Although they lack direct DNA-binding activity, PCLs contain a Tudor domain, which was reported to specifically recognize H3K36me3 and in the case of PCL3, mediate recruitment of PRC2 to active loci and allow their transition to a Polycomb-repressed state, perhaps via concomitant recruitment of the NO66 H3K36me3 demethylase (Ballare et al. 2012; Brien et al. 2012; Cai et al. 2013). A more general role of PCLs in PRC2 targeting remains a matter of debate.

1.6.4. Additional PcG complexes in flies

1.6.4.1. Pho-RC

In addition to PRC1, PRC2 and dRAF (described in section 1.5.2.2. "KDM2B"), two other Polycomb complexes have been characterized in flies, Pho-RC and PR-DUB.

Pho-RC was first described by Klymenko et al. (2006) as a two-subunit complex consisting of the only sequence-specific DNA-binding PcG proteins Pleiohomeotic (PHO) or the homologous Pho-like (PHOL) in a mutually exclusive manner together with the Scm-related gene containing four MBT domains dSFMBT protein. dSFMBT contains four malignant brain tumor (MBT) repeats and a sterile α motif (SAM) domain and was shown to selectively recognize mono- or di-methylated H3K9 or H4K20, although the relevance of this binding is unclear. Importantly, although this complex does not encode chromatin modifying activities and is biochemically distinct from PRC1 and PRC2, artificial tethering of Gal4-dSFMBT fusion protein could repress transcription of a *UAS-Gal4-Ubx-LacZ* reporter gene, suggesting that dSFMBT functions as a repressor. However, its methyl-lysine binding ability is not a major chromatin recruitment mechanism, as dSFMBT requires PHO for DNA binding. Interestingly, PHO and PHOL binding sites are found in many different defined PREs, and *phol, pho* double mutants show derepression of homeotic genes *Ubx* and *Abd-B* in the wing disk (Brown et al. 2003). Nevertheless, occupancy of other PcG proteins such as PSC, PH, SCM and E(Z) to most targets appeared relatively unaffected in the absence of PHO and PHOL, questioning the contribution of these DNA-binding proteins to PRC recruitment, despite reports suggesting an association between Pho-RC and PRC1 or PRC2 subunits. The PHO/PHOL mammalian counterpart, Yin Yang 1 (YY1) was also reported to contribute to PcG recruitment, H3K27 trimethylation and transcriptional repression (Wilkinson et al. 2006). The importance of YY1 as a general mechanism of PcG targeting is not supported by the observation that in mouse ESCs, YY1 chromatin-binding profile shows almost no overlap with PRC2 or PRC1, but seems to exclusively co-localize with sites marked by K4me3 (Mendenhall et al. 2010). A recent report indicated that the fly Pho-RC complex also contains RPD3/HDAC1, HP1β, NAP1, and the orthologue of the mammalian Max gene-associated protein MGA, thus forming a complex resembling the one defined by mammalian PCGF6-PRC1 variant

complex (Alfieri et al. 2013), although how these other subunits contribute to Pho-RC mediated Polycomb repression in Drosophila and how this complex integrates with PRC1 and PRC2 remains to be defined.

1.6.4.2. PR-DUB

The most recent PcG complex characterized is PR-DUB (Polycomb Repressive Deubiquitinase), which was shown to consist of the H2A histone deubiquitinase CALYPSO (dBAP1) and the polycomb protein Additional sex combs (ASX) (Scheuermann et al. 2010). ASX was previously shown to partially co-localize with PH, PC and PCL on polytene chromosomes, but also occupy unique sites, suggesting that ASX could potentially bind chromatin in the absence of other PcG members (Sinclair et al. 1998). Intriguingly, *Asx*-homozygous embryos were reported to exhibit posterior transformations, but less severe than those observed for a classical PcG gene, *Pc* (Breen and Duncan 1986). Furthermore, one P-element-induced allele, Asx^{P1}, exhibits posterior and anterior transformations in adult flies (Sinclair et al. 1992), while homozygosity for Asx^{P1} strongly enhances both PcG and TrxG phenotypes (Milne et al. 1999), indicating that ASX could play a dual role in activation and repression of homeotic loci. Interestingly, PR-DUB is bound at the PREs of PcG target genes and Drosophila mutants lacking PR-DUB show a strong increase in the levels of monoubiquitylated H2A, suggesting that this complex specifically deubiquitinates the histone mark placed by PRC1. Although clearly further work is required to understand why PR-DUB is bound at polycomb targets, one possibility could be that a dynamic balance between H2A ubiquitylation and deubiquitylation is required for proper gene expression regulation. A similar complex was reported to exist in mammals, formed by BAP1 and one of the three ASX homologues ASXL1-3, but also containing other polypeptides such as the transcriptional regulator HCF-1, the forkhead transcription factors FOXK1 and FOXK2, YY1 or the O-linked N-acetylglucosamine transferase (OGT) (Katoh and Katoh 2004; Yu et al. 2010). Although the relevance of these associations in mammals remains to be determined, mice with reduced levels of ASX-like 2 (ASXL2) also show loss of bulk H3K27me3, as well as locus-specific depletion of this mark (Lai and Wang 2013), in agreement with a potential contribution to Polycomb-mediated repression.

1.6.5. Trithorax group proteins (TrxG) in flies and mammals

Trithorax group proteins were identified based on transformations that reproduce loss-of-function *Hox* mutations in Drosophila, but also in screens for extragenic suppressors of *Pc* mutations, where TrxG members were found to supress the Polycomb-mediated extra sex combs homeotic phenotype (Figure 1.6) (Kennison 2004; Allis et al. 2006). TrxG components are mainly associated either with histone-modifying activity or with chromatin-remodelling complexes, although other proteins have been classified as trxG members with less clear functional roles in maintaining active chromatin states (Ringrose and Paro 2007; Schuettengruber et al. 2011). The founding member of the group is TRX (Ingham 1998), which contains a C-terminal SET domain shown to catalyse H3K4 methylation (Shilatifard 2008). However, reports indicate that it is not the main H3K4 HKMTase in flies, with bulk H3K4me2/me3 mediated instead by COMPASS-member dSET1 (Ardehali et al. 2011). Interestingly, TRX was found in a complex with the H3K27 HAT dCBP, providing an elegant mechanism of counteracting PcG-based H3K27me deposition (Petruk et al. 2001).

Flies also contain a TRX-related protein TRR, which was reported to have a more specific role in associating with the ecdysone-receptor EcR and catalysing H3K4me3 at ecdysone-inducible promoters (Sedkov et al. 2003). However, more recent reports suggest that TRR associates with the demethylase dUTX to mediate removal of H3K27me3 at inactive promoters and prime them for H3K4me1 and H3K27ac (Herz et al. 2012). TRX and TRR are related to the mammalian Mixed Lineage Leukaemia methyltransferases MLL1/2 and MLL3/4, respectively (Eissenberg and Shilatifard 2010), which were discussed in greater detail in section 1.3.1.2.1. Another SET-domain containing TrxG member is Absent small and homeotic discs 1 (ASH1), protein shown to specifically dimethylate H3K36, mark reported previously to counteract PRC2-deposited H3K27me3 (as detailed in section 1.3.1.2.3) (Dorighi and Tamkun 2013).

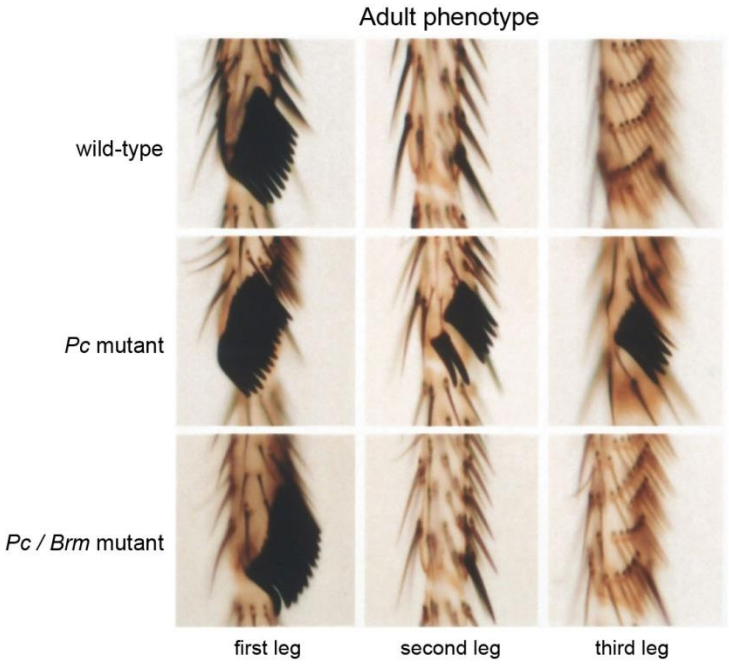


Figure 1.6. TrxG protein BRM suppresses Polycomb *Pc* mutant homeotic transformation phenotype.
In wild-type flies, sex comb teeth are unique to the first leg. In *Pc* mutant flies, the *Scr* gene becomes partially derepressed in the thoracic segments T2 and T3, leading to appearance of ectopic sex comb teeth on the second and third legs. This homeotic transformation phenotype is rescued by mutations in the Trithorax group protein BRAHMA, with flies showing normal presence of sex comb only on the first leg. Figure reproduced from (Allis et al. 2006) and (Kennison 2004).

Interestingly, the binding of both TRX and ASH1 appears to be facilitated by another TrxG component, KISMET, which is a member of the CHD subfamily of chromatin-remodelling factors (Srinivasan et al. 2008). In addition to KISMET, the TrxG proteins Brahma (BRM), OSA and moira (MOR) are also components of a SWI/SNF ATP-dependent chromatin-remodelling complex. These observations indicate that the contribution of TrxG to gene regulation may be two-fold, with chromatin-remodellers maintaining an accessible chromatin state at target genes potentially counteracting Polycomb-induced chromatin compaction, while the histone-modifying components antagonize PcG-mediated chromatin modifications. Importantly, multiple studies suggest that the function of TrxG may not be to drive transcription *per se*, but to act as anti-repressor to prevent inappropriate PcG-silencing as default state and render the gene available for activation (Poux et al. 2002; Klymenko and Muller 2004; Schwartz et al. 2010).

1.6.6. Polycomb complexes and their role in gene silencing

Polycomb group proteins are considered to function in maintaining transcriptional silencing, although the molecular mechanisms by which this is achieved remain unclear. Nevertheless, a multitude of potential pathways have been reported which can be broadly grouped into two classes, namely alteration of chromatin structure by compaction or interference with the transcriptional machinery, either directly or via "readers" of PcG-mediated histone modifications. Importantly, the silencing activity is primarily attributed to PRC1 (regardless of whether it is dependent or not on H2AK119ub1), with relatively few reports investigating a repressive role for H3K27me3 besides its ability to mediate canonical PRC1 recruitment. Admittedly, it is very difficult to distinguish between the activities of PRC1 and PRC2, considering their close connections. The discovery of PRC2-independent targeting of variant RYBP-containing PRC1s further complicates such a task, as all of these complexes co-localize at the majority of target genes and their interplay remains to be elucidated.

Core PRC1 was reported to compact polynucleosomal arrays *in vitro*, while other studies indicated that PRC1 can inhibit chromatin remodelling by the TrxG-related SWI/SNF complexes without altering nucleosomal organization (Francis et al. 2001; King et al. 2002; Francis et al. 2004; Grau et al. 2011). Furthermore, another report suggested that the ubiquitylating activity of RING1B is not required for chromatin compaction (Eskeland et al. 2010b), although in this particular case, the homologous RING1A protein was still present which could compensate for the loss of RING1B enzymatic capacity. Finally, Drosophila mutants lacking individual PRC1 subunits indicate that only a subset of targets become derepressed in the absence of dRING, while others are misexpressed only in animals which also lack the PSC-Su(z)2 and PH subunits (Gutierrez et al. 2012), suggesting that gene regulation by PRC1 may be a combination of its E3 ubiquitin ligase activity and chromatin compaction.

In mammals, Polycomb protein occupancy overlaps almost exclusively with a subset of CpG islands. Although CGIs are generally associated with an open chromatin conformation (see Section 1.4.4), PcG-occupied CGIs have a less accessible chromatin configuration than their non-PcG-associated

counterparts (Fenouil et al. 2012). The discovery that SAM-domain containing PH proteins can polymerize (Isono et al. 2013) most probably contributes to this phenomenon.

However, the question of how chromatin compaction actually interferes with transcription remains poorly understood. One answer could be that it prevents transcription factor access to its cognate binding site. However, studies suggest that PRC1 and general transcription factors can co-occupy target genes at the same time (Breiling et al. 2001) and that PcG subunits can bind to a PRE in both the ON and OFF transcriptional states (Langlais et al. 2012), indicating that Polycomb occupancy and transcription are not necessarily mutually exclusive.

Indeed, most bivalent genes characterized by presence of both H3K4me3 and H3K27me3 (see section 1.3.1.2.2) in mouse ESCs are transcribed at low levels (Guenther et al. 2007; Stock et al. 2007) and were shown to be occupied by poised RNAPII Ser-5 phosphorylated. Interestingly, deletion of RING1A/1B and subsequent loss of H2AK119ub1 was reported to result in otherwise poorly characterized changes in RNAPII complexes and a subsequent shift from poised RNAPII conformation to productive gene expression, supporting a role for PRC1-mediated H2AK119ub1 in promoting polymerase pausing. However, the reasons for this block to elongation or a potential connection to chromatin compaction were not investigated.

A role for PRCs in regulating transcription elongation is also supported by the observation that the TrxG protein KISMET acts downstream of positive transcription elongation factor (P-TEFb) recruitment and stimulates early transcriptional elongation (Srinivasan et al. 2008). Interestingly, there are also studies which indicate that, while TFIID and RNAPII are not prevented from binding by PcG occupancy, PcG components block transcription initiation (Dellino et al. 2004), a function which was suggested at least *in vitro* to be due to inhibition of Mediator assembly (Lehmann et al. 2012). Mediator is a multisubunit coactivator complex which is required to assemble the RNAPII preinitiation complex (PIC). Therefore, it seems that PcG components can in principle inhibit transcription initiation or the productive elongation step. Although these activities may not necessarily be mutually exclusive (perhaps depending on the gene target), the molecular mechanisms that underpin these processes are not clear. Perhaps a

systematic identification of H2AK119ub1-binding proteins could provide a link with the transcription machinery.

In addition to the mechanisms by which PcG contributes to transcriptional repression, another critical question in the field which was has not been unambiguously answered is whether polycomb occupancy is sufficient to determine the transcriptional status of a gene. Experiments in flies which place PREs upstream of a *Ubx-lacZ* reporter gene (Fritsch et al. 1999; Sengupta et al. 2004) or in front of the linked marker gene *mini-white* (DeVido et al. 2008) suggest that PcG recruitment is necessary and sufficient for transcriptional silencing. However, the observations that Polycomb-occupied genes can be activated and that, in the absence of PcG components, not all targets become derepressed (Endoh et al. 2008; Leeb et al. 2010; Schwartz et al. 2010), would indicate that PRC occupancy does not drive transcription repression *per se*. Instead, as suggested by early Drosophila studies, PcG appears to respond to the inactive transcriptional state of developmental regulatory genes and functions to maintain this state by creating a chromatin conformation that is inhibitory to transcription until a required activating signal is present to drive gene expression.

1.7. Polycomb recruitment mechanisms to target sites

1.7.1. Polycomb recruitment in flies

In Drosophila, it is accepted that PcG members are recruited to specific *cis*-regulatory genomic elements called PREs (Polycomb response elements) through the activity of different DNA binding proteins. However, on closer inspection, it becomes apparent that surprisingly little is understood about the architecture and sequence composition of PREs or the proteins responsible for PcG targeting.

Since their original identification in the homeotic gene clusters of the *Bithorax* (*BX-C*) and *Antennapedia* (*ANT-C*) complexes (Simon et al. 1993; Chan et al. 1994; Chiang et al. 1995; Gindhart and Kaufman 1995), many more PREs were defined based on their capacity to act as binding platforms for PcG proteins and mediate silencing of *Hox* reporter genes in a PcG-dependent manner in transgenic fly

embryos and larvae (Sengupta et al. 2004). However, bioinformatics analysis of the several hundred PREs identified based on genome-wide chromatin binding profiles of PcG subunits did not reveal a clear sequence consensus, although a number of DNA-binding factors have been reported to contribute to PRE function, such as PHO/PHOL (discussed previously in section 1.6.4.1), GAGA factor (GAF) (Horard et al. 2000; Mulholland et al. 2003), PIPSQUEAK (PSQ) which has an identical binding sequence-consensus and chromatin distribution as GAF (Huang et al. 2002), ZESTE (Z) (Dejardin and Cavalli 2004), Dorsal switch protein 1 (DSP1) (Dejardin et al. 2005), GRAINYHEAD (GRH) (Blastyak et al. 2006) or members of the SP1/Kruppel-like factor (KLF) family of zinc finger proteins (Brown et al. 2005). Among these factors, only PHO binding sites appear to be a consistent feature of PREs, with homozygous null mutants for many of the other DNA-binding proteins showing mostly unaffected *Hox* gene expression.

Importantly, PREs are not necessarily located immediately upstream of the gene they regulate but can be tens of kbs upstream or downstream of the target or within introns and were proposed to control promoters over long-distances *in-cis* but also *in-trans,* most probably via extensive looping and interactions between PREs into higher-order chromatin structures (Pirrotta 1997; Lanzuolo et al. 2007). Intriguingly, TrxG proteins seem to localize to the same genomic elements as PcG complexes (Orlando et al. 1998; Tillib et al. 1999; Schuettengruber et al. 2009), with PcG and TrxG members being bound at the *Ubx* PRE in both the ON and OFF transcriptional states, albeit binding of PRC1 and PRC2 was shown to be about twofold lower in Ubx^{ON} chromatin compared with Ubx^{OFF} chromatin (Papp and Muller 2006).

Another reported feature of PREs is that they appear to be more sensitive to DNaseI digestion, indicative of nucleosome free or unstable regions in the chromatin (Mohd-Sarip et al. 2006). These observations would therefore suggest that PREs and TREs most probably represent different states of the same genomic element (silent versus transcriptionally active, respectively), and that they act as binding platforms for both PcG and TrxG complexes due to their accessible chromatin configuration. In support of this, transcription through a PRE counteracts PcG occupancy and leads to TrxG acquisition and maintenance of the target gene in a transcriptionally active state (Schmitt et al. 2005). Interestingly, recent work by Sharif et al. (2013) studying GC content at the TSS of fly PcG targets with clear mammalian orthologues observed a peak in GC-richness at polycomb regulated genes that have

counterparts in mice, in contrast with the rest of the fly promoters which generally exhibit a drastic GC depletion. This result is in striking resemblance with mammalian PcG targets, which almost exclusively overlap with CpG island-associated genes. As CGIs are also characterized by an accessible chromatin configuration, it may be that PREs and mammalian polycomb targeting elements are defined by these two features: open chromatin structure and GC-richness, perhaps because of the specific requirements of the DNA-binding factor or factors which mediate PcG and TrxG recruitment.

The accepted view in flies is that DNA-binding proteins recruit PRC2 to PREs, which in turn brings in PRC1 via PC-dependent recognition of H3K27me3. This is referred to as the hierarchical recruitment mechanism. However, multiple studies indicate that while PcG proteins are mainly located at PREs, H3K27me3 forms broad domains sometimes encompassing the entire target gene and surrounding regions, but seems reduced at PREs most probably as a result of nucleosome depletion at those regions (Kahn et al. 2006; Papp and Muller 2006; Schuettengruber et al. 2009). Interestingly, only the PC chromatin-binding profile appears to be more spread than the other PcG proteins, potentially due to its association with trimethylated H3K27. Given the low H3K27me3 levels at PREs, these studies suggested the possibility that PRC1 proteins could be recruited to PREs independently of histone methylation (Kahn et al. 2006). Although H3K27-independent targeting of PRC1 has not been unambiguously demonstrated in flies, it has been tested and confirmed in mammals.

1.7.2. Polycomb recruitment in mammals

A schematic illustration of the potential mammalian Polycomb recruitment strategies referred to in this chapter is summarized in figure 1.7.

1.7.2.1. Mammalian PREs

Attempts to identify PRE-like elements in mammals have failed with only a couple of examples reported. Furthermore, the majority of fly DNA-binding factors reported to contribute to PRE function have no mammalian homologues, with the exception of PHO/PHOL. However, as mentioned previously, YY1 chromatin binding profile seems to not significantly overlap with that of PcG members, indicating that it cannot play a major contribution to genome-wide targeting. The mammalian PRE-L elements identified include PRE-kr, which regulates expression of the mouse *MafB/Kreisler* gene and was reported to contain a palindromic double PHO/YY1-binding site (Sing et al. 2009); a fragment from the human *HOXD* cluster – the D11.12 element, which again was suggested to contain a YY1-binding site whose disruption caused loss of PRC1 binding but did not affect SUZ12 recruitment (Woo et al. 2010); a couple of putative PREs from human T cells *SLC17A7* and *HoxA* genes which were shown to confer repressive activity to reporter gene expression in Drosophila (Cuddapah et al. 2012); and two novel prospective PcG-dependent regulatory elements within the human *HOXB* and *HOXC* clusters, whose activities were again reported to be dependent on YY1 (Woo et al. 2013). Although how these putative PREs function remains unclear, it is interesting to note that the genes they presumably regulate are all associated with CpG island elements.

1.7.2.2. Polycomb recruitment by sequence-specific transcription factors

Extrapolating from flies, different sequence-specific DNA-binding factors have been reported to mediate Polycomb recruitment in mammals via direct biochemical interactions. One example is the transcription factor REST (Neuron-Restrictive Silencing Factor or NRSF), which was originally identified as a repressor of neuronal genes in non-neuronal cells (Schoenherr and Anderson 1995). In mouse ES cells, as well as in human teratocarcinoma cells (NT2-D1), REST was shown to associate both with PRC1 components RING1B, PCGF1, PCGF4, multiple CBX proteins (particularly CBX7/8) and with PRC2 subunits EZH2 and

SUZ12 (Ren and Kerppola 2011; Dietrich et al. 2012). Importantly, ChIP-Seq analysis in wild-type and *Rest−/−* mES cells revealed a reduction in RING1B binding at a subset of polycomb targets normally co-occupied by REST, whereas the effect on SUZ12 seemed less dramatic (Dietrich et al. 2012). Furthermore, comparing RING1B and REST binding in WT control and *Eed−/−* mES cells, the authors concluded that, while the majority of RING1B peaks were lost in the absence of PRC2, RING1B was nevertheless maintained at REST-targets, suggesting that this transcription factor may contribute to PcG recruitment independently of PRC2 for at least a subset of sites.

Core binding transcription factors are heterodimeric complexes composed of CBF α (RUNX1/2/3) bound to a common CBFβ subunit. In megakaryocytic and T lymphocytic cells, RUNX1 was reported to directly interact with PCGF4/BMI1, while ChIP-Seq studies revealed an extensive overlap between RUNX1 and PRC1, with over 50% of the RING1B peaks also bound by RUNX1 (Yu et al. 2012). RUNX1-deficient thymocytes showed depletion of RING1B signal at the majority of co-occupied targets and ChIP-qPCR at a panel of targets in EZH2-deficient thymocytes showed no significant loss of RING1B, RUNX1 or CBFβ chromatin occupancy. These results supported the conclusion that RUNX1 can recruit PRC1 in the absence of PRC2, although it is not clear whether a similar RUNX1-dependent targeting function exists in pluripotent cells. Finally, the SNAIL family of transcription factors has also been reported to contribute to H3K27me3 deposition in neuronal progenitors, as assessed by a bioinformatics approach examining the correlation between acquisition of H3K27me3 during neuronal differentiation of murine stem cells and occurrence of computationally predicted TF binding sites (Arnold et al. 2013).

Although the contribution of these sequence-specific binding transcription factors cannot be disregarded, it remains unlikely that such interactions could account for all PcG binding events *in vivo,* especially as these TFs occupy only relatively small subsets of polycomb targets and cannot explain both PRC1 and PRC2 targeting.

1.7.2.3. Polycomb recruitment by complex-specific protein partners

Core PRC1 and PRC2 proteins do not show direct DNA-binding specificity. However, a number of polycomb-associated proteins were reported to possess such capacity either directly or via interaction partners. These include KDM2B, BCOR and MAX/MGA in the case of PRC1, while JARID2 and AEBP2 were suggested to contribute to PRC2 recruitment. Although the precise DNA-binding specificity for JARID2 and AEBP2 remains to be unambiguously determined (see Section 1.6.3.2) , it was shown that KDM2B can recognize non-methylated CpGs *in vitro* and localize to at least a few CGI-associated promoters, although its genome-wide chromatin binding profile has not been investigated. Furthermore, BCOR can specifically bind to BCL6 targets genome-wide (Hatzi et al. 2013). Interestingly however, only 36% of BCL6 targets were occupied by BCOR and many non-BCL6 BCOR bound regions seem instead coincident with CpG island promoters. MAX and its dimerization partner MGA were reported to bind DNA directly in a sequence-specific manner and regulate multiple physiological processes, including cell growth, survival, and differentiation (Hurlin et al. 1999; Grandori et al. 2000; Baudino and Cleveland 2001). The chromatin binding profile of MGA has not been investigated, but preliminary studies suggest that MAX, MGA and L3MBTL2 function as repressors of germ cell-related gene expression in mouse embryonic stem cells (Maeda et al. 2013), although whether their depletion affects polycomb occupancy at those sites was not addressed. Alternatively to DNA-binding activities, PCL proteins were suggested to contribute to PRC2 recruitment through recognition of a specific chromatin configuration, enabling intrusion of PRC2 into active chromatin regions (see Section 1.6.3.2). Clearly further work is required to understand whether any of these factors plays an essential role in Polycomb recruitment on a genome-wide scale.

1.7.2.4. Polycomb recruitment by long non-coding RNAs (lncRNAs)

Multiple non-coding RNAs (ncRNAs) have been reported to function *in cis* or *in trans* to achieve gene silencing through recruitment of polycomb repressive complexes. The classical example is the long ncRNA Xist, which is produced exclusively from the inactive X chromosome and spreads *in cis* to

completely coat the Xi. Although it was shown that *Xist* RNA is sufficient to initiate X inactivation, the mechanisms by which this is achieved remain less clear (Lyon 1961; Brockdorff et al. 1992; Herzing et al. 1997; Lee and Jaenisch 1997). One suggested pathway is that *Xist* directly interacts with and recruits silencing chromatin modifying activities to the inactive X chromosome. In agreement, PRC1 and PRC2 core components and their associated chromatin marks were found to be enriched on the Xi (Plath et al. 2003; Silva et al. 2003; de Napoles et al. 2004; Kohlmaier et al. 2004) , potentially as a result of direct interactions between PRC2 components EZH2 and SUZ12 and a stem-loop structure formed by the A-repeats of RepA ncRNA sequence within Xist intron 1 (Wutz et al. 2002; Zhao et al. 2008a).

Supporting an association between *Xist* and PRC2 is the observation that PcG recruitment is reversible and is dependent on continuous *Xist* expression, as demonstrated by experiments in which *Xist* transcription can be induced from a transgene integrated on an autosome (Kohlmaier et al. 2004). However, the same study showed that a RepA-deletion *Xist* mutant, which localizes to chromatin but not does lead to chromosome silencing, still results in H3K27me3 deposition, suggesting that methylation is established independently of silencing and is not sufficient to cause repression. Once established, PRC2 was suggested to result in PRC1 recruitment via H3K27me3. Work by Schoeftner et al. (2006) however showed that RING1B localizes to the Xi and mediates ubiquitylation of histone H2A in *Eed* deficient embryonic stem (ES) cells, suggesting that PRC1 recruitment can be independent of PRC2.

Despite intense research, it is not unambiguously proven that *Xist* directly interacts with and recruits PcG components (Brockdorff 2013), nor is it clear whether PRC1 and PRC2 play a role in initiating X chromosome inactivation or rather in the maintenance of a repressive chromatin state. This is further complicated by super-resolution immunofluorescence studies which indicate that *Xist* and PcG components do not co-localize but are instead in close proximity in the Xi territory (A. Cerase and N. Brockdorff, personal communication).

Other examples of lncRNAs recruiting PcG complexes include the human HOTAIR lncRNA from the *HoxC* locus which was reported to represses transcription *in trans* of the HOXD locus by association with PRC2 components (Rinn et al. 2007), the ANRIL lncRNA which was reported to bind to SUZ12 *in vivo* and recruit PRC2 to repress expression of *INK*^{4B} tumour suppressor gene (Kotake et al. 2011), as well as the

imprinted ncRNA Kcnqt1ot1 which was shown to interact with PRC2 components and the G9A HKMTase to establish lineage-specific transcriptional silencing of genes in the Kcnq1 domain (Pandey et al. 2008).

In addition to these specific examples of lncRNAs, short RNAs transcribed from polycomb repressed genes were reported to form stem-loop structures as observed in *Xist* and interact with SUZ12 to promote association of PRC2 with target genes (Kanhere et al. 2010). Importantly, these ~50-200nt long structures can be transcribed from promoters, introns and exons by RNAPII in a manner independent of polycomb activity, as they were still observed to exist in cells depleted of H3K27 methylation or H2AK119 ubiquitylation. When ES cells were induced to differentiate into precursor motor neurons, specific neuronal genes become derepressed concomitant with a decrease in short RNA transcripts and loss of Polycomb occupancy, potentially suggesting that short RNAs function to stabilize PRC2 on chromatin. However, it cannot be disregarded that Polycomb loss could be a direct result of gene activation and may be unrelated to short RNA transcription. Furthermore, these short transcripts were also produced from genes not associated with H3K27me3 and genes which also produce full-length mRNAs. As these short RNAs were reported to associate with SUZ12, it is not clear what prevents PRC2 binding and subsequent gene repression in such cases. Recently, it was shown that interactions between lncRNAs and the PRC2 associated factor JARID2 might promote JARID2-mediated recruitment of PRC2 to chromatin *in vivo*, although it remains to be established whether this is a genome-wide phenomenon (Kaneko et al. 2013a).

The mammalian genome was also reported to encode several thousand large intergenic ncRNAs (lincRNAs), and members of this species were also suggested to associate with various chromatin modifying complexes, including PRC2 subunits (Khalil et al. 2009; Guttman et al. 2011). siRNA –based depletion of a panel of these lincRNAs resulted in activation of some genes normally occupied by PRC2. However, it is not clear how PcG components specifically interact with lincRNAs. Interestingly, recent reports suggest that PRC2 binds RNAs promiscuously *in vitro and in vivo* and that this may represent a mechanism by which PRC2 scans both active and repressed loci to reinforce PcG-mediated repression at genes which escaped silencing and prevent PcG establishment at active genes (Davidovich et al. 2013; Kaneko et al. 2013a; Kaneko et al. 2013b). The striking conclusion from these experiments is that PRC2

may be broadly associated with gene promoters and not limited to classical PcG-targets that show high levels of PRC2 occupancy. However it remains to be determined whether this is a consequence of PRC2 recruitment by the underlying DNA sequence (i.e. GC-richness) or whether it is due to non-specific interaction and recruitment by RNAs (either nascent or non-coding), which seem to be a prominent characteristic of mammalian promoters.

1.7.2.5. H3K27me3-dependent and independent recruitment of PRC1

Based on co-occupancy and interdependence of PRC1 and PRC2 components at polycomb targets in flies, a hierarchical recruitment pathway was proposed according to which PRC2 is targeted to specific sites through a direct interaction with DNA-binding factors, where it catalyzes H3K27me3 deposition. This histone modification is then directly recognized by the chromodomain in the PC subunit, leading to PRC1 nucleation. Although this pathway is supported by some studies which show that depletion or deletion of E(Z) results in loss of PC binding at the *Ubx* PRE in both cultured S2 cells and fly wing imaginal discs (Cao et al. 2002; Cao and Zhang 2004a; Wang et al. 2004b), there are also reports which suggest a potential methylation-independent targeting of PRC1. For example, both Mohd-Sarip et al. (2002) and Kassis and Kennison (2010) indicated direct interactions between members of the PHO-RC complex and PRC1 subunits PH, PC and SCM, thereby potentially circumventing the need for PRC2 in binding of PRC1 to target loci. Moreover, staining of polytene chromosomes in *esc/escl*-double mutant larvae revealed that loss of H3K27 methylation does not immediately affect binding of the PSC PRC1 component, but the a gradual loss of PSC binding occurred. This would suggest that K27 methylation is not sufficient to recruit PRC1 but is essential for stabilization of PcG complex binding (Ohno et al. 2008). Finally, *Pcl* mutant embryos show a reduction in H3K27me3 levels, but PHO and PH X-ChIP levels are indistinguishable from their wild-type counterpart (Nekrasov et al. 2007), further reinforcing the need to better understand PcG targeting mechanisms in fly.

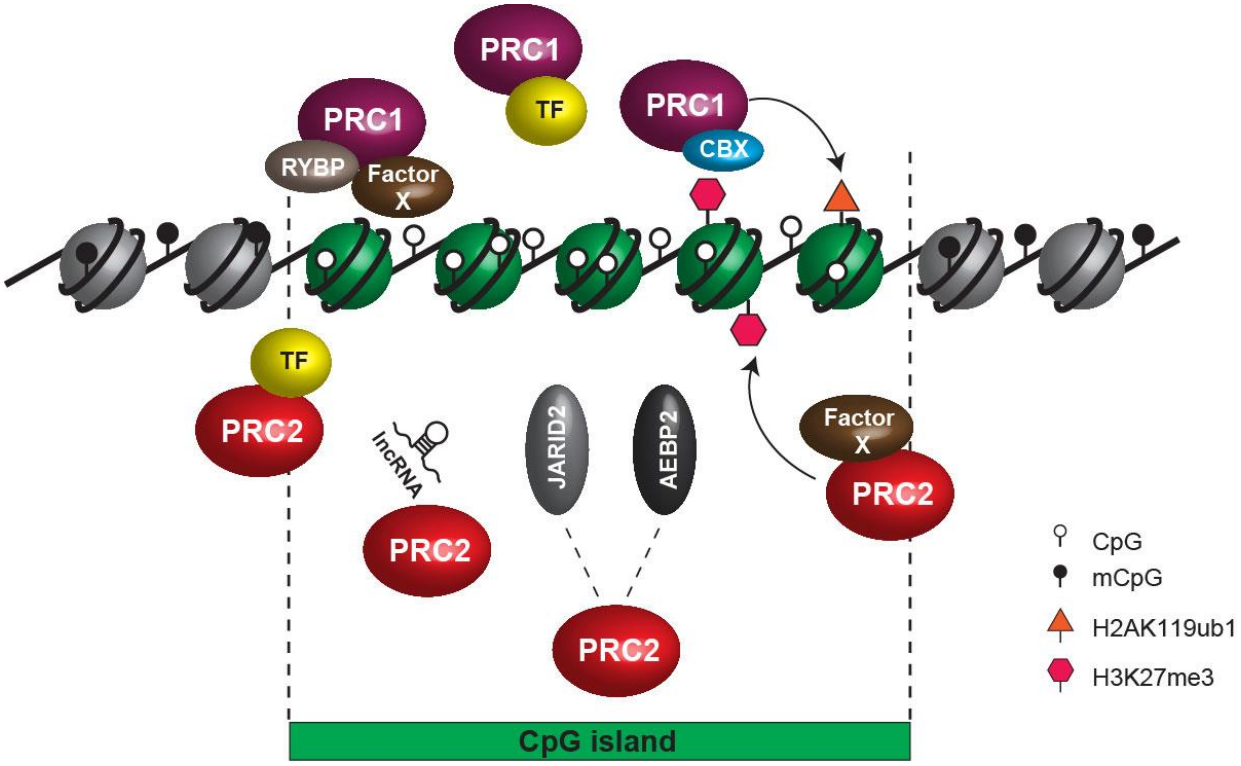


Figure 1.7. Potential recruitment mechanisms for Polycomb complexes to target genes in mammalian cells.

In flies, Polycomb complexes nucleate at specific genomic elements called PREs via the action of recruiter proteins. In mammals, very few PRE-like sequences were defined. Instead, PcG occupancy overlaps almost exclusively with CpG island elements, although the mechanisms behind this chromatin-binding profile remain poorly characterized. PRC1 and PRC2 were suggested to be recruited to target loci through direct biochemical interaction with various sequence-specific binding transcription factors (TFs), although the few examples currently identified are insufficient to account for all PcG-occupied sites *in vivo*.

Alternatively, PRC2 binding could be facilitated by its associated protein partners, JARID2 or AEBP2, which were reported to show DNA-binding specificity *in vitro*. Furthermore, non-coding RNAs (ncRNAs) were suggested to interact primarily with PRC2 and act as PcG recruiters either at selected target loci or more pervasively genome-wide. Factor X represents an unknown protein or pathway which can mediate PcG nucleation, potentially via recognition of the underlining non-methylated CpGs or the unique chromatin architecture at CpG islands.

Extrapolating from flies, PRC2 nucleation would result in the hierarchical recruitment of PRC1 through CBX-dependent recognition of H3K27me3. The observation that PRC1 binding is not entirely abrogated in the absence of H3K27me3, together with the discovery of non-canonical PRC1 complexes in which the CBX subunit is replaced by RYBP/YAF2, support the existence of H3K27me3-independent PRC1 targeting strategies, which are however currently poorly characterized. It is highly probable that PcG occupancy is the result of multiple diverse recruitment pathways and defining these represents a major challenge to understanding the role of Polycomb proteins in development and differentiation.

The picture seems equally complicated in mammals. Multiple studies indicate a causal link between H3K27me3 deposition and PRC1 recruitment (Agger et al. 2007; Lee et al. 2007b), with the most striking example represented by experiments in which a viral SET (vSET) protein catalysing H3K27 methylation is capable of recruiting PRC1 subunit CBX8 to mediate transcriptional repression of a reporter luciferase gene (Mujtaba et al. 2008). However, as discussed in greater detail in Section 1.6.2.2. (Non-canonical PRC1 complexes), an increasing number of reports indicate that PRC1 can be recruited to at least a subset of target genes in the absence of an intact PRC2 and independently of H3K27 methylation (Schoeftner et al. 2006; Pasini et al. 2007; Leeb et al. 2010), potentially mediated via variant RYBP-containing PRC1 complexes. Interestingly, ChIP-qPCR analysis at a panel of PcG target loci derepressed in RING1A/1B double KO ES cells indicated a concomitant decrease in EED binding (Endoh et al. 2008). Although this could be an indirect consequence of gene activation, it could also suggest a role of PRC1 in facilitating PRC2 binding to target loci, further complicating the issue of dissecting the complex interplay between PcG recruitment strategies.

1.7.2.6. Polycomb group proteins co-occupy a subset of CpG islands

Extensive whole-genome analysis in mouse and human embryonic stem cells revealed that Polycomb binding overlaps almost exclusively with a subset of CpG islands (~10%), with H3K27me3 marking the promoter of more than 2000 genes. Many of these PcG-occupied genes have essential developmental regulatory functions (Bernstein et al. 2006; Boyer et al. 2006; Lee et al. 2006; Mikkelsen et al. 2007; Ku et al. 2008). Importantly, H3K27me3 sites were frequently also enriched in H3K4me3 to create “bivalent domains” (see section 1.3.1.2.2 on “H3K27me3”). Interestingly, the majority of CpG island promoters were shown to be associated with significant H3K4me3 enrichment in ES cells regardless of their expression status, indicating that mammalian Trithorax-related proteins can specifically and pervasively interface these genomic elements most probably via the ZF-CxxC domain present in CFP1, MLL1 or MLL2. However, the molecular mechanisms by which Polycomb proteins nucleate at CGIs remain undefined.

Multiple studies show that GC-rich sequences devoid of activating transcription factor motifs are sufficient to recruit PcG complexes, suggesting that Polycomb proteins may have the capacity to recognize non-methylated CpG DNA *in vivo*. For example, Mendenhall et al. (2010) reported that removal of activating motifs from an otherwise active 'housekeeping' CpG island leads to PRC2 recruitment and subsequent H3K27me3 deposition. Moreover, integration into mouse ES cells of arbitrarily selected bacterial GC-rich sequences from the *E. coli* genome lacking eukaryotic activating signals resulted in both H3K4me3 and H3K27me3 enrichment, demonstrating a sufficiency of non-methylated GC-rich elements in establishing bivalent domains. This observation is in agreement with studies by Lynch et al. (2012), who use recombinase-mediated cassette exchange (RMCE) to insert fragments from the CGI-containing/PcG-occupied human α globin cluster into the orthologous non-CGI/non-PcG-associated region of the mouse locus and showed that insertion of these short DNA fragments is sufficient to confer novel bivalency to the mouse α locus. Interestingly, insertion of a constitutively active promoter upstream of the integrated fragments resulted in a complete clearance of PRC2 and H3K27me3 with a concomitant increase in H3K4me3 levels, reflective of a competition between PcG recruitment and transcriptional activation.

In addition to transcription, DNA methylation was also reported to be refractory to Polycomb nucleation, as indicated by experiments taking advantage of DNA methyltransferase-deficient cells, where a redistribution of PcG proteins from classic targets to novel non-methylated CpG-rich sequences was detected (Lynch et al. 2012; Reddington et al. 2013). Together with recent studies which indicate low but detectable levels of EZH2 at the majority of CGI-associated TSSs (Kaneko et al. 2013b), these observations point towards a model whereby, although both Polycomb and Trithorax-related proteins can interface non-methylated CpG island elements genome-wide, establishment and stabilization of PcG occupancy can only be achieved in the absence of activating signals, while TrxG binding and H3K4me3 levels are further reinforced by the process of active transcription. Thereby, according to this model, CGIs act as recruitment platforms allowing both the PcG and TrxG systems to assess the transcriptional state of the associated gene. Nevertheless, the mechanisms responsible for Polycomb recruitment to

CGI-associated target loci remain poorly characterized and this represents a major hindrance to our understanding of Polycomb-mediated gene regulation.

1.8. Aims of this thesis

A family of proteins was identified which can specifically recognize non-methylated CpGs via a ZF-CXXC DNA binding domain. One such example is KDM2A, which was reported to nucleate at CpG island elements genome-wide and specifically demethylate H3K36me1/me2. However, there are two highly related KDM2 proteins in mammals, with KDM2B also containing a highly conserved histone lysine demethylase catalytic JmjC domain and a ZF-CXXC DNA binding domain reported to be functional *in vitro* and required for KDM2B binding at specific loci *in vivo*. Therefore, the overarching aim of this thesis will be to investigate if KDM2B is selectively recruited to CpG islands genome-wide via its ZF-CXXC domain, and if so, to better characterize its contribution to CGI function and understand whether it acts redundantly with KDM2A or it has unique properties and functionality.

Chapter 3 aims to investigate the chromatin binding profile of KDM2B by using a newly generated antibody specifically recognizing this protein. Whole-genome sequencing results reveal that, while both KDM2A and KDM2B recognize CGIs in mouse ES cells, their chromatin binding profile is not entirely overlapping, with KDM2B preferentially enriched at Polycomb-occupied developmental regulatory gene-associated CGIs. Chapter 4 describes the design and use of an unbiased affinity-based purification strategy to investigate whether KDM2A and 2B form protein complexes, the conclusions of which further reinforce a connection between KDM2B and Polycomb proteins. Chapters 5 and 6 take advantage of KDM2B shRNA-based depletion or genetic ablation approaches to dissect whether KDM2B contributes to PRC1 binding to CGI-associated target loci. Results detailed in chapter 6 using a *de novo* targeting assay examine the effects of KDM2B tethering to an artificial CpG-free locus, further revealing an unexpected role of KDM2B in the formation of polycomb domains *in vivo*.

Together this work provides a novel mechanism by which KDM2B links recognition of non-methylated CpGs to nucleation of polycomb repressive complexes and formation of polycomb domains. In addition to offering at least a partial explanation for how CpG island elements may function as the mammalian counterpart to fly PREs, it also provides an exciting new avenue for exploring how polycomb activities are established and maintained during development and differentiation.

2. Chapter two - Materials and methods

2.1. DNA methods

2.1.1. DNA constructs used in this study

Inserts were PCR-amplified and generally inserted via ligation-independent cloning (LIC) into the expression vector of interest. All PCR generated constructs were verified by sequencing prior to use.

Table 2.1. Summary of main DNA constructs used in this study

DNA inserts		
Construct	Origin	Source
KDM2A	human	Full-length cDNA clone KIA1004 (kindly provided by T. Nagase)
KDM2B	human	Full-length cDNA clone (kindly provided by Dr. Ryo Koyama-Nasu)
RING1B	mouse	Full-length cDNA clone (Brockforff lab)
RYBP	mouse	IMAGE clone 8861345 (Source Bioscience)
YAF2	mouse	IMAGE clone 3488691 (Source Bioscience)
SKP1	mouse	IMAGE clone 3490159 (Source Bioscience)
PCGF1	mouse	Modified to full-length from IMAGE clone 1380201 (Source Bioscience)
PCGF2	mouse	IMAGE clone 4458285 (Brockdorff lab)
PCGF3	mouse	IMAGE clone 5371040 (Source Bioscience)
PCGF4	mouse	IMAGE clone 6854104 (Brockdorff lab)
PCGF5	mouse	Full-length (GeneArt synthesized, Invitrogen)
PCGF6	mouse	Full-length cDNA clone (kindly provided by Dr. Haruhiko Koseki)
cloning vectors		
Construct	lic-cloning adapted	Experimental use
pNIC28	yes	Prokaryotic expression vector
pLION2-Flag/2xStrepII	yes	Eukaryotic expression vector
pCAGIpuro-Flag/2xStrepII	yes (in this study)	Eukaryotic expression vector
pCAGIpuro-Flag/2xStrepII-TetR	yes	Eukaryotic expression vector used to generate N-terminal TetR fusions
pLKO.1 puro/blast	-	shRNA-expression vector

2.1.2. Polymerase chain reaction (PCR)

2.1.2.1. Taq-polymerase analytical PCR

Analytical PCR reactions were carried in 50 µl final volume. 200 ng of recombinant Taq DNA polymerase (purified by Dr. Jin Zhou), 0.5 µM of each forward and reverse primer, 5 nmol of dNTP mix (Fermentas) and 10 ng of plasmid template DNA were incubated together in the presence of Taq polymerase PCR buffer (20 mM Tris HCl pH 8.8, 10 mM (NH4)2SO4, 10 mM KCl, 2 mM MgSO4, 0.1% Triton X 100).

Table 2.2. General pipetting scheme for analytical PCR

Component	Volume
Plasmid DNA (10ng/µL)	1 µL
Forward primer (10 µM)	2.5 µL
Reverse primer (10 µM)	2.5 µL
dNTP mix (2 mM)	5 µL
10x Taq buffer	5 µL
Taq DNA polymerase (0.2 µg/µl)	0.5 µL
MQ H_2O	33.5µL

The solution was thoroughly mixed and briefly spun down to collect all liquid at the bottom of the PCR tube before placing in a BioRad DNA Engine Thermal cycler. Cycling parameters varied depending on DNA template and primer properties. The PCR products were purified using the QIAquick PCR Purification kit (QIAGEN) according to manufacturer instructions if required for cloning purposes. The DNA was eluted in 30 µl elution (EB) buffer.

Table 2.3. Thermal cycling conditions for analytical PCR

Cycling step	Temperature	Time
1. Initial denaturation	95°C	1 minute
2. Denaturation	95°C	30 seconds
3. Annealing	50-55°C	30 seconds
4. Extension	72°C	1kb/1minute
Repeat step 2 – 4 for 35x cycles		
5. Final extension	72°C	10 minutes
6. Hold	10°C	-

2.1.2.2. Phusion-polymerase high-fidelity PCR

For generation of mammalian expression constructs, all inserts were PCR-amplified using High-fidelity Phusion DNA polymerase (ThermoScientific/New Englans Biolabs). All reactions were performed in a 50 µL final volume. For problematic amplifications through GC-rich regions, 2% DMSO was added to the reactions. Alternatively, Phusion GC buffer was used instead of the recommended High-fidelity (HF) reaction buffer. Pipetting scheme and cycling conditions are summarized in Table 2.4 and Table 2.5, respectively.

Table 2.4. General pipetting scheme for high-fidelity Phusion PCR

Component	Volume
Plasmid DNA (10ng/µL)	1 µL
Forward primer (10 µM)	2 µL
Reverse primer (10 µM)	2 µL
dNTP mix (2 mM)	5 µL
5x HF buffer	10 µL
Phusion DNA polymerase	0.5 µL
MQ H_2O	29.5µL

The PCR reactions were gently mixed and all liquid was collected at the bottom of the tubes by a quick spin step before transferring the tubes to a BioRad DNA engine thermal cycler. Following Phusion polymerase-based PCR amplification, the products were generally purified via gel extraction for cloning purposes.

Table 2.5. Thermal cycling conditions for high-fidelity Phusion PCR

Cycling step	Temperature	Time
1. Initial denaturation	98°C	30 seconds
2. Denaturation	98°C	10 seconds
3. Annealing	50-55°C	30 seconds
4. Extension	72°C	1kb/30 seconds
Repeat step 2 – 4 for 35x cycles		
5. Final extension	72°C	10 minutes
6. Hold	10°C	-

2.1.3. Ligation-independent cloning (LIC)

2.1.3.1. Principle

Ligase independent cloning (LIC) is a simple and efficient method for cloning of expression constructs (Aslanidis and de Jong 1990). It takes advantage of the 3'-->5' exonuclease activity of T4 DNA polymerase to create specific 10-15 base single overhangs in a linearized expression vector. Inserts of interest are PCR-amplified with primers designed to contain complementary overhangs after treating them with T4 DNA polymerase. The annealing of insert and vector is therefore highly efficient and should in principle also work in the absence of DNA ligase. However, the efficiency of the process and stability of the ligated product is maximized in the presence of ligase. Importantly, once a vector is adapted for ligation-independent cloning, it easily allows insertion of multiple products amplified with identical overhang-containing primers. All LIC-adapted vectors in the Klose lab were generated to produce hybridisation-compatible single stranded overhangs after T4 polymerase treatment in the presence of dGTP. To create an insert with complementary overhangs to the LIC vectors, all PCR products were amplified with primers having the following overhangs:

Forward primer: 5' TACTTCCAATCCATG - specific sequence (~18-20bps) 3' (ATG start codon in red)

Reverse primer: 5' TATCCACCTTTACTGCTA - specific sequence (~18-20bps) 3' (CTA stop codon in red, if no C-terminal tag is used)

2.1.3.2. Generation of LIC-compatible vector pCAGIpuro-Flag/2xStrepxII

For subsequent ease of cloning of various inserts into a eukaryotic expression vector to generate stable ES cells lines expressing tagged versions of the proteins of interest, the pCAG-IRES-puro (pCAGIpuro) vector was first fully sequenced by primer walking. As no suitable restriction enzyme was found to adapt this vector for LIC-cloning, a LIC cassette containing the Flag/2xStrepII tag, a TEV protease cleavage site

and an optimal BaeI restriction site flanked by engineered 5’ and 3’ LIC overhangs was synthesized using GeneArt Gene synthesis service (Invitrogen). The cassette was flanked by XhoI/NotI restriction sites, which allowed cloning into pCAGIpuro multiple cloning site (Figure 2.1). BaeI digestion was used to linearize the vector, followed by T4 DNA polymerase treatment in the presence of dGTP. The 3'-->5' exonuclease activity of T4 DNA polymerase removes bases from both 3’- ends until the first guanine (G) residue within the LIC overhangs is reached, generating compatible single stranded ends for insert hybridisation.

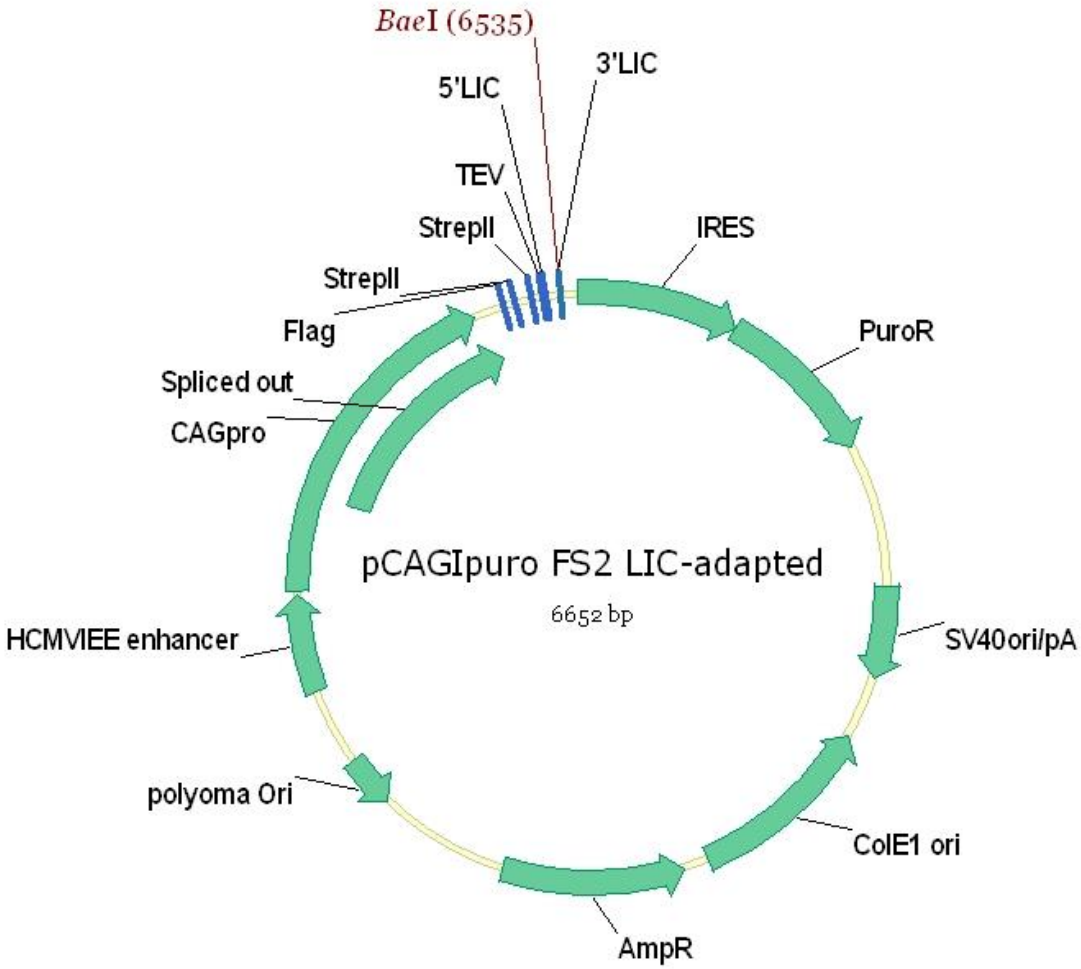


Figure 2.1. Schematic illustration of the LIC-adapted mammalian expression vector pCAG-IRES-puro Flag/2xStrepII.

(Construct map was generated using Vector NTI (version 10.3.1).

2.1.3.3. LIC-cloning protocol

5 µg of LIC-adapted vector is linearized by digestion with the appropriate enzyme (i.e. BsaI for pNIC28 and BaeI for pCAGIpuro-based vectors) in a 50µL reaction volume for at least 4 hours to overnight, according the individual enzyme requirements. The digestion products were resolved by agarose gel electrophoresis in 1X TAE, visualised by ethidium bromide staining and the linearized vector excised from the gel. Vector DNA was purified using the QIAquick Gel Extraction kit (QIAGEN), eluted in 30 µL elution buffer and quantified by nanodrop measurement. The purified vector was treated with T4 DNA polymerase (Fermentas) in the presence of dGTP to generate PCR-compatible single strand overhangs, as summarized in Table 2.6.

Table 2.6. Pipetting scheme and reaction conditions for T4 DNA polymerase treatment of LIC-adapted vector

T4 polymerase treatment pipetting scheme	
Component	Volume
Vector	1µg
5x T4 DNA polymerase buffer	10 µL
BSA (1mg/mL)	5 µL
dGTP (2mM)	2 µL
T4 DNA polymerase (3000units/mL)	2 µL
milliQ H_2O	to 50 µL
T4 polymerase treatment reaction conditions	
Temperature	Time
22°C (incubation)	30 minutes
75°C (heat inactivation)	20 minutes
10°C (hold)	-

Following T4 DNA polymerase treatment, DNA was purified using the clean-up protocol of the QIAquick Gel Extraction kit (QIAGEN) kit and eluted in 30 µL elution buffer. The DNA to be inserted via LIC-cloning was PCR-amplified with vector compatible overhang-containing primers using Phusion DNA polymerase. The PCR product was resolved by agarose gel electrophoresis, purified by gel extraction and eluted in 30 µL elution buffer. The PCR-amplified insert was treated with T4 DNA polymerase as described for the vector, with the exception that the treatment was performed in the presence of dCTP to generate

compatible single-stranded overhangs for the subsequent hybridisation reaction. The T4-treated insert was purified using the clean-up protocol of the QIAquick Gel Extraction kit (QIAGEN) kit and eluted in 30 µL elution buffer. The purified insert and vector DNA were mixed in different concentrations and ligated using the Rapid DNA ligation kit (Roche). A ligation reaction containing only vector was carried out in parallel to monitor self-ligation. The entire ligation mixture was transformed into chemically competent cells. Generally, 5-10 colonies are obtained following LIC-cloning, of which more than 90% are usually successful insertion events. Vector-only ligations do not yield any colonies, indicative of the high efficiency of the ligation-independent cloning approach.

2.1.4. DNA manipulations

2.1.4.1. DNA transformation

Heat-shock transformation of chemically competent XL-10 Gold *Escherichia coli* strain (prepared in-house) was performed as previously described (Sambrook and Russell 2001), with minor modifications. Briefly, the competent bacteria were gently thawed on ice and generally 1 µL plasmid DNA was added to the 100 µL thawed competent cells and gently mixed. The bugs were incubated on ice for 20 minutes, and heat-shocked in a 42°C water bath for 1 minute. The cells were subsequently incubated on ice for 2 minutes following the heat-shock. 1mL of Luria-Bertani (LB) media (10 g/L Tryptone, 5 g/L Yeast extract, 10 g/L NaCl, dissolved in distilled H_2O and pH adjusted to 7.0 prior to autoclaving) was immediately added to the bacteria. The cells were allowed to recover for 45 minutes at 37°C with shaking. Approximately 50-100 µL of the transformation mix was streaked on LB plates (with selective antibiotic, if required) and incubated at 37°C overnight to allow individual colony growth. If transformed DNA was the result of a ligation or LIC cloning reaction, generally the entire mix (~ 20 µL final reaction volume) was transformed into chemically competent cells. After recovery, the bacteria were collected by centrifugation at 4200rpm x 2 minutes, resuspended in approximately 100 µL LB media and spread on LB plates with appropriate antibiotic.

2.1.4.2. Isolation of plasmid DNA

For small-scale purification of plasmid DNA from bacterial cultures, individual colonies were picked using sterile pipette tips and inoculated into 5 mL LB media with appropriate antibiotic selection. The bacteria were grown overnight at 37°C with shaking at 200xrpm. Next day, the bugs were collected by centrifugation at 4600rpm x 10 minutes, and plasmid DNA purified using QIAGEN Spin Miniprep Kit according to the manufacturer's instructions. Isolated DNA was eluted in 30-50 µL of elution buffer and stored at -20°C until use.

For transfections into mammalian cells, a high-concentration of pure plasmid DNA was required. To this end, single transformed bacterial colonies were inoculated into 125 mL of LB media with required antibiotic selection and allowed to grow overnight at 37°C with shaking. The bacteria were harvested by centrifugation at 4600 rpm x 10 minutes, and high-quality plasmid DNA purified using the GenElute HP Maxiprep kit (Sigma-Aldrich) according to the manufacturer's instructions. After elution, the isolated plasmid DNA was further concentrated by precipitation following addition of 1/10 volume sodium acetate (3M solution, pH 5.2) and 7/10 volume isopropanol and extensively washed with 70% molecular ethanol solution. After air-drying, the DNA pellet was resuspended in elution buffer to a concentration of 1 mg/mL and stored at -20°C until use.

2.1.4.3. Analytical restriction digestion

Restriction digests were routinely performed in 30 µl reaction volume for conventional analytical purposes. All enzymes used purchased from New England Biolabs and the reactions were carried out at temperatures and under conditions indicated by the manufacturer (New England Biolabs, Roche). The results of the analytical digestions were assessed by running the digestion product on an appropriate concentration agarose gel (usually 1%) in 1x TAE (40 mM Tris, 20 mM glacial acetic acid, 1mM EDTA, pH

adjusted to 8.0) running buffer and visualized by ethidium bromide staining. The size of the products was determined by comparison with 1kb Plus DNA ladder (Invitrogen).

If required for cloning purposes, PCR or restriction digestion products were purified from agarose gels using the QIAquick Gel Extraction kit (QIAGEN) according to manufacturer's instructions and eluted in 30 µL elution (EB) buffer.

2.1.4.4. DNA mutagenesis

All site-directed mutagenesis reactions were performed using the QuickChange II XL site-directed mutagenesis kit (Stratagene) according to manufacturer instructions, with the mention that in-house generated chemically competent cells were routinely used for transformation of the mutagenesis reaction product. However, attempts to generate the KDM2B ZF-CxxC DNA binding domain mutation K643A in the pCAGIpuro FS2 vector background using the QuickChange kit were unsuccessful due to the high molecular weight of the plasmid (>10kb). Therefore for this experiment, a classic cloning approach was taken whereby the K643A mutation was first carried out using the QuickChange kit a smaller vector background (pEGFP hKDM2B) and then BstXI restriction sites flanking the ZF-CxxC DNA binding domain were used to excise the mutated fragment and ligate it into similarly digested KDM2B construct in the pCAGIpuro-FS2 eukaryotic expression vector used to generate the mammalian ES stable cell lines. All products generated from cloning steps were sequenced by Source BioScience LifeSciences (University of Oxford facility).

2.1.4.5. DNA ligation

All ligation reactions were carried out using the Rapid DNA Ligation Kit (Roche). Generally, 1 µL vector was mixed with different volumes of insert (1µL/3µL), together with 10 µL 2x T4 DNA ligation buffer and

1µL T4 DNA ligase. Final reaction volume was taken to 20µL by addition of 1X DNA Dilution Buffer. A vector-only control was used to monitor self-ligation. The samples were thoroughly mixed and ligation allowed to proceed for at least 1 hour at room temperature. The entire ligation mix was generally transformed into chemically competent bacterial cells.

2.1.5. Generation of knockdown constructs for KDM2B and PCGF1

2.1.5.1. Design of oligos containing 21-mer shRNA sequences against targets of interest

Selection of suitable 21-mer targets for the gene of interest was performed according to Addgene pLKO.1 protocol recommendations (http://www.addgene.org/tools/protocols/plko/#B) and Sigma-Alridch shRNA design guidelines. The sequences chosen were BLAST-searched using NCBI against mouse genome, to minimize degradation of potential off-target mRNAs. Multiple 21-mer target sequences were selected for KDM2B and PCGF1, as listed in Table 2.7.

Table 2.7. 21-mer shRNA sequences used for knockdown studies

Number	21-mer sequence (5' -> 3')	Notes
mKDM2B		
1	CGCTGTGGAAATATCTGTCAT	Targets FL+ SF KDM2B (used in (Farcas et al. 2012)
2	GCGGCTCATTATTCGCCATAT	Targets FL+ SF KDM2B
3	TGCGGCTCATTATTCGCCATA	Targets FL+ SF KDM2B
4	GCTGTGGAAATATCTGTCATA	Targets FL+ SF KDM2B
5	GCCTATTATTTCCCTTTGGAA	Targets 3'UTR of FL+SF KDM2B
6	GCTCCAACTCAGTTACTGTAA	Targets FL+ SF KDM2B (adapted from (He et al. 2011))
7	GTGGACTCACCTTACCGAATT	Targets FL KDM2B (adapted from (Fukuda et al. 2011))
8	CAGTTCTCACTTTAGTTGGTA	Targets FL+SF KDM2B
9	CAAGGATGGACTAGGGATCAA	Targets FL KDM2B

10	CATCTGCAACGAGATCATCCA	Targets FL+SF KDM2B
mPCGF1		
1	CCCAGATCACATGACAATGAA	Efficient PCGF1 KD (refer to Chapter 6)
2	ACCACTCTAAAGCCCACTATT	-
3	GCTTCTCCAATACAGTGTGAA	-
Scrambled shRNA control		
1	CCTAAGGTTAAGTCGCCCTCG	-

To design oligos compatible with pLKO.1 constitutive shRNA-expression vector, the 21-mer sequences were adapted to the following format:

Forward oligo: 5' CCGG—21bp sense—CTCGAG—21bp antisense—TTTTTG 3'

Reverse oligo: 5' AATTCAAAAA—21bp sense—CTCGAG—21bp antisense 3'

After annealing the forward and reverse oligos, the hairpin sequence generated comprises of a 21 base paired stem (sense and antisense strands), a 6 base loop which is also an XhoI restriction site (CTCGAG), 5' flanking region containing CCGG overhang for AgeI and 3' flanking sequence containing TTTTT termination for RNAPIII and AATT overhang for EcoRI.

2.1.5.2. Annealing oligos to clone into pLKO.1 puro/blast vector

The oligos from Section 2.1.5.1 were designed to contain the shRNA sequence flanked by sequences that are compatible with the sticky ends of EcoRI and AgeI restriction enzymes. Forward and reverse oligos are annealed and ligated into the pLKO.1 vector, producing a final plasmid that expresses the shRNA of interest. For annealing, the oligo stocks were diluted to 3mg/mL and then 1µL of forward and reverse oligo was incubated together with 48µL annealing buffer (100 mM NaCl, 50 mM HEPES ph7.9). The mixture was heated to 94°C for 4 minutes and then the heated block was taken out and allowed to cool to RT over at least 3 hours. The pLKO.1 cloning vector (encoding either puromycin (Addgene no

10878) or blasticidin (Addgene no 26655) resistance) was digested with AgeI and EcoRI for 2 hours at 37°C to release the stuffer. Digested DNA was run on 1% agarose gel and the vector gel extracted and eluted in 30µL elution buffer.

For ligation of shRNA oligos into the pLKO.1 vector, Rapid DNA ligation kit (Roche) was used in a final volume of 20µL, consisting of 2µL annealed oligo mix, 1µL digested and purified vector, 2µL T4 DNA ligase and 10µL 2x T4 DNA ligation buffer. The reaction was allowed to proceed for 2 hours at RT. All the ligation mix was transformed into chemically competent XL-10 Gold bacterial cells. The cells were plated on LB agar plates containing 100 µg/mL ampicillin and incubated overnight at 37°C. Successful clones were screened by NdeI and EcoRI-HF digestion, yielding 92 bp for vector only, while hairpin insertion yielded a product of 126 bp (unless hairpin oligos have these restriction sites in).The inserts were verified by sequencing with forward (5'-atacgatacaaggctgttagagagata-3') and reverse primers (5'-aggggaaagaatagtagaca-3'). The sequencing reactions were supplemented with 1M betaine and 2% DMSO to improve sequencing through the secondary structure of the shRNA constructs. Maxiprep-grade quality DNA was prepared from the positive clones of interest and used for lentivirus production.

2.2. Mammalian tissue culture methods

2.2.1. Tissue culture media preparation

Regular non-ES tissue culture media was composed of Dulbecco's modified Eagle's medium (DMEM) (with 4.5g/L glucose w/ L-glutamine, cat no BE12-604F, Lonza, Switzerland), supplemented with 10% Foetal Calf Serum (FCS) (Labtech, UK) and 100µg/mL penicillin/streptomycin (Lonza, Switzerland) per 500mL media bottle.

ES medium was composed of DMEM (high glucose, without sodium pyruvate, cat no 41965062, Gibco, Life Technologies Ltd, UK), supplemented with 15% FCS, 100µg/mL pen/strep, 2mM L-Glutamine (Gibco, UK), 1x Non-essential amino acids (MEM-NEAA) (Gibco, UK), 50µM beta-mercaptoethanol

(Gibco, UK) and 10ng/ml leukaemia inhibitory factor (LIF, prepared in house) per 500mL medium bottle. All new FCS serum batches were tested prior to use for lack of toxicity and embryonic stem cell growth promotion.

If required, antibiotic was added at the following concentrations for ES cells: puromycin (cat no P7255-100MG, Sigma-Aldrich, US) at 1-1.5µg/mL, blasticidin (Blasticidin S HCl, cat no R210-01, Gibco) at 10µg/mL, geneticin G418 sulfate (cat no 11811031, Gibco) at 400µg/mL. The optimum antibiotic concentrations were determined by kill curve experiments.

Medium was stored at 4°C and warmed up in a 33°C water bath for at least half an hour prior to use to avoid cold-shocking the cells.

2.2.2. Thawing, growing and freezing cells

Cryovials generally containing 1ml of frozen cells were thawed in a 33°C water bath and immediately transferred into a 15mL Falcon tube containing 9mL of fresh pre-warmed medium to dilute the DMSO and reduce its toxic effects on thawed cells. Cell were spun down at 1,000rpmx3 minutes at RT and the pellet resuspended in fresh medium and transferred to a tissue culture dish, according to the concentration at which the cells were frozen and the requirements of each cell line. The next day, medium was aspirated to remove floating dead cells and fresh medium was added.

All feeder-dependent ES cell lines used were grown on a monolayer of inactivated SNL fibroblasts. The SNL cell line is an immortalized subclone of the STO line manipulated to stably express the neomycin resistance and leukaemia inhibitory factor (LIF) genes (McMahon and Bradley 1990). ES cells were grown in dishes coated with 0.1 % gelatine to increase adherence. All the cell lines were cultured in dishes purchased from Greiner Bio-One GmbH (Germany) and incubated at 37°C in 5% CO_2 atmosphere. ES cells were examined and fed daily, while differentiated and transformed cells were media-changed every two-three days. All cultures were routinely tested for mycoplasma contamination according to manufacturer instructions (MycoAlert Mycoplasma detection kit, cat no LT07-318, Lonza).

Cells were passaged when 80-90 % confluent. To passage cells, medium was aspirated and the cells were briefly washed with phosphate buffered saline (PBS) without Ca^{2+}/Mg^{2+} or phenol red (cat no BE17-516F, Lonza, Switzerland). 0.05 % trypsin/0.53 mM ethylenediaminetetraacetic acid (EDTA) (cat no 25300-096, Gibco) was added to cover the cells. For passage of ESCs, the Trypsin-EDTA was supplemented with 2 % chicken serum (CS) which tends to be gentler on the cells. The cells were incubated for 2-5 minutes at 37°C or until they are uniformly dispersed into small clumps. Once the cells detached, the trypsin was inactivated by addition of fresh medium. The cells were gently resuspended and plated onto new dishes at the optimum split ratios. For feeder-dependent ESCs, SNLs were allowed to settle and adhere on the 0.1% gelatine-coated dishes for at least 2 hours before ES cells were plated.

For freezing, the cells were trypsinized as described above, resuspended in fresh medium and transferred into Falcon tubes followed by centrifugation at 1,000rpmx3 minutes at RT. The cell pellet was resuspended in freezing-medium consisting of FCS with 10% dimethyl sulfoxide (DMSO) (Sigma). 1mL of resuspended cells was aliquoted into 2mL cryovial tubes (cat no 122263, Greiner Bio-One) and immediately frozen down at -80°C overnight in special containers filled with isopropanol ("Mr Frosty", Nalgene). These containers help provide the critical, repeatable, 1°C/min cooling rate required for successful cryopreservation of cells. Frozen cryovials were stored long-term in liquid nitrogen vapour-phase canisters.

2.2.3. Mitomycin C inactivation of feeders

SNLP (puromycin-resistance) or SNLH (hygromycin-resistance) feeder cells were grown until confluency in DMEM media supplemented with 10 % FCS and 100 µg/ml penicillin/streptomycin. Old media was aspirated from the cells and replaced with media with 10µg/mL mitomycin C (Sigma). The 2mg mitomycin C stock was resuspended in 1xPBS to a concentration of 1mg/ml until the majority was in solution. The treated cells were placed at 37°C for 2 hours in a 5% CO_2 atmosphere incubator, followed by three gentle wash steps in 1x PBS before trypsinization. The inactivated feeders were collected,

counted using a hemocytometer and spun down at 1,000xrpm for 3 minutes. Aliquots of $4x10^6$ cells were frozen in standard freezing media (10% DMSO and 90% FCS) and stored in liquid N_2 until use. Approximately $2x10^6$ inactivated feeders were plated per 90mm Petri dish, with a culture area of 56.7 cm^2 and allowed to settle for at least 2 hours to overnight before addition of ES cells.

2.2.4. Cell lines used in this study

All cell lines used in this study and their properties are summarized in Table 2.8.

Table 2.8. Description of cell lines used in this work

Mouse embryonic stem cells (ESCs)			
Cells	**Feeder-dependent**	**Reference**	**Main experimental use**
E14	No	(Hooper et al. 1987)	Generation of stable cell lines / KDM2B knockdown
EFC-1	No	(Nichols et al. 1990)	Protein biochemistry
J1	No	(Li et al. 1992)	Immunofluorescence
V6.5	Yes (SNLs)	(Eggan et al. 2001)	ChIP-qPCR/ChIP-Seq
Ring1A–/–;Ring1Bfl/fl; Rosa26::CreERT2	Yes (SNLs)	(Endoh et al. 2008)	ChIP-qPCR/ChIP-Seq
Kdm2bfl/fl:Gt(Rosa26)Sor-ERT2Cre+/-	Yes (SNLs)	(Blackledge, Farcas, Kondo et al., *submitted*)	ChIP-qPCR/ChIP-Seq/ RNA-Seq
Non-ESC lines			
HeLa	-	(Scherer et al. 1953)	Transfections/Immunofluorescence
293T	-	(DuBridge et al. 1987)	Transfections/lentivirus production
IMR90	-	(Nichols et al. 1977)	Immunofluorescence
U2OS	-	(Ponten and Saksela 1967)	Immunofluorescence

2.2.5. Transfections

For transient over-expression studies or generation of stable cell lines, FuGene HD (Roche) or Lipofectamine 2000 (Invitrogen) reagents were used, with lipofectamine the method of choice for transfections into ES cells. The best efficiencies were obtained with DNA to transfection reagent ratio of 1:3 (i.e.: 1µg DNA + 3µL lipofectamine for 1 well of 6-well plate; 5µg DNA + 15µL lipofectamine for 10cm dish plate). Briefly, high-quality purified DNA and transfection reagent were diluted separately in DMEM media without any additions (no serum or penicillin/streptomycin), then mixed and incubated at room temperature for 20 minutes. The solution was then distributed dropwise onto the cells and the dish placed back in the tissue culture incubator. The transfection was allowed to proceed overnight. If lipofectamine was used, 1 hour before transfection, the cells were switched to media without any antibiotics to prevent cell death. Next day, the cells were fed with regular complete media and harvested 24-48 hours after transfection for transient over-expression experiments. For selection of stable integrants, the cells were passaged with suitable dilution and antibiotic selection began 48 hours after lipofectamine transfection. Puromycin (Sigma) was used at 1 – 1.5µg/mL concentration, geneticin / G418 sulphate (Gibco) at 400µg/ml and blasticidin (Gibco) at 10µg/mL. Stable cell lines were picked when fully formed, usually at 7 – 10 days after selection was started.

2.2.6. Isolation of ES clones

Colonies were generally picked from 10xcm or 15xcm plates. The cells were washed gently with 1x PBS, and then covered with sufficient 1xPBS to prevent the colonies from drying out. ES cell clones were picked under a Leica MZ75 microscope, by gently aspirating them using a 200µL pipette and placing them directly in a well of a round-bottom 96-well plate containing 30µL 0.0% 1x Trypsin-EDTA (Gibco) supplemented with 2% chicken serum (Gibco). After the colonies were picked, the plate was incubated at 37°C 5% CO_2 atmosphere for 30 minutes to allow the individual colonies to disperse. The cells were

then gently pipette up and down a few times and transferred onto a flat-bottom gelatin-coated 48-well plate (Lonza) with regular ES medium. Antibiotic selection was restarted 24 hours after the colonies were picked to minimize cell death during the procedure.

2.2.7. Lentivirus production and infection

For producing recombinant lentiviral particles, the shRNA constructs of interest were co-transfected with psPAX2 packaging plasmid and pMD2.G envelope helper plasmid into 293T cells using FuGene HD reagent (Roche). For transfection onto a 10cm plate, 6µg shRNA plasmid were mixed with 4µg psPAX2 packaging plasmid and 2µg pMD2.G envelope plasmid, at a total DNA : FuGene ratio of 1:3, thus using 36µL transfection reagent. After distributing to cells, transfection was allowed to proceed overnight. Next morning, the cells were fed with fresh complete media, and incubated for 4-6 hours. As the first virus produced in this time frame is sub-optimal, the cells were then fed again with 8mL fresh media. The first lentivirus batch was collected approximately 12 hours later (next morning). 8mL fresh media was again added to the transfected cells and the second lentivirus batch was collected 24 hours later, if a second collection was necessary. Residual 293T cells were removed by centrifuging the collected media for 5 minutes at 1,500xrpm, followed by a filtration step through a 0.45µm filter (a 0.2µm filter will shear the viral envelope) to remove any 293T cells carried over. The virus was aliquoted and frozen at -80°C until use. As the lentivirus was used to infect ES cells, ES - complete DMEM (i.e. containing LIF) was used to feed the 293T cells for lentivirus collection, so as to minimize side-effects of ES cells during infection due to dilution of LIF.

Lentiviral infection of feederless ES cells was performed overnight in the presence of 4 µg/mL polybrene. The following day, cells were diluted into fresh growth media and allowed to settle onto gelatine-coated dishes. Puromycin selection (1.5 µg/mL) was started 48 hr following transduction and stable lines were isolated, expanded and analyzed using qRT-PCR for efficient knockdown of the protein of interest.

2.2.8. Immunofluorescence (IF) studies

Staining was performed in HeLa, 293T, IMR90 and U2OS cells, which were cultured in DMEM (Lonza, Basel, Switzerland) supplemented with 10% FCS and penicillin/streptomycin. Transient transfection in HeLa cells was achieved using Fugene HD reagent (Roche, Basel, Switzerland) and 2 µg of Flag-tagged KDM2B expression plasmid. Briefly, cells seeded on cover slips were fixed with 4% paraformaldehyde for 20 min, permeabilized with 0.5% Triton X-100 for 10 min and blocked using 3% BSA for 30 min. Fixation was done approximately 24 hr after transfections. Cells were incubated with KDM2B rabbit polyclonal primary antibody for 2 hr or the Flag mouse monoclonal M2 antibody (Sigma, St Louis, MO) in case of exogenous expression, followed by incubation in secondary antibody conjugated with Rhodamine or FITC (Jackson ImmunoResearch Laboratories, West Grove, PA) for 1 hr. All intermediary washing steps and dilutions were performed in 1×PBS. After 4, 6-diamidino-2-phenylindole dihydrochloride (DAPI) nuclear staining, the cover slips were mounted on glass slides in fluorescent mounting medium (DAKO, Glostrup, Denmark) and allowed to dry overnight. Images were aquired using a Zeiss AxioSkop fluorescent microscope.

2.3. RNA methods

2.3.1. Routine RNA extraction using Trizol reagent

RNA was extracted from mammalian cell pellets using Trizol reagent (Invitrogen). Following a PBS wash step, 1mL Trizol was added (for pellet equivalent to material from one well of 6-well plate) and the pellet passed through a pipette several times to resuspend. The sample was incubated at room temperature for 5 minutes, after which 0.2mL chloroform was added, mixed vigorously and allowed to incubate for 2 minutes at room temperature. After a centrifugation step at 12,000xg for 15 minutes at 4°C, the upper colourless phase which contains the RNA was transferred to a fresh tube. A volume of 0.5mL isopropanol was added, mixed thoroughly and the samples incubated at room temperature for 10

minutes to allow the RNA to precipitate. Samples were centrifuged at 12,000xg for 10 minutes at 4°C, and the RNA pellet was rinsed twice with cold 70% molecular ethanol solution, followed by centrifugation at 7,600xg for 5 minutes. The supernatant was removed, the RNA pellet was dried at room temperature for 5-10 minutes, resuspended in a corresponding volume (usually 50 - 100µL) of DEPC-treated water (Invitrogen), quantified by Nanodrop and stored at -80°C until use.

2.3.2. RNA preparation for microarray and RNA-Seq gene expression analysis

Total RNA was extracted using the Qiagen RNeasy Mini kit. Approximately 10 µg nucleic acid were treated with Turbo DNase (Ambion, Carlsbad, CA) at 37°C for 30 min, according to the manufacturer's instructions. Genomic DNA-free RNA samples were further purified using the RNeasy kit RNA cleanup protocol. Samples were run on an 1% agarose gel to check quality of RNA preparation and integrity of 18S and 28S rRNA bands. For subsequent RT-PCR analysis, cDNA was synthesized with the ImProm-II Reverse Transcription System (Promega, Madison, WI), as described in the Reverse-transcription polymerase chain reaction (RT-PCR) section. Quantitative real-time PCR was performed in duplicate by using Quantace SYBR Green master mix, using *Gapdh* as housekeeping gene.

At least 1µg of total RNA in ultrapure water was submitted for RNA-Seq. polyA purification for mRNA-Seq directional library generation was done using the NEBNext UltraTM Directional RNA library Prep Kit for Illumina (NEB, Ipswich, MA), with minor modifications. Briefly, Wellcome Trust Centre for Human Genetics (WTCHG) in-house 8bp indexes were used (Lamble et al. 2013). No size selection was carried-out; instead a clean-up step was performed to remove adapter and primer dimers. Only 1/3 of the pre-PCR material was amplified in the PCR.

2.3.3. Reverse-transcription polymerase chain reaction (RT-PCR)

The RT-PCR reaction was performed using the ImProm-II™ Reverse Transcription System (Promega). The purified RNA was diluted to 100ng/µL in ultrapure DEPC-treated water (Invitrogen). 4µL of the diluted RNA (400ng in total) was mixed with 1µL random primers (random hexamers) stock (500ng/µL) and the tubes placed in a pre-heated 70°C block for 5 minutes. Afterwards, the tubes were immediately chilled on ice for at least 5 minutes and spun down for 10 seconds to collect any condensate and maintain the original volume. The reverse transcription reaction mix was prepared on ice and consisted of ImProm-II™ Reaction buffer (1x), 8mM $MgCl_2$, 1µL of 10mM dNTP mix, 0.5µL recombinant RNasin Ribonuclease Inhibitor (40U/ul), nuclease-free water and 1µL ImProm-II™ reverse transcriptase (RT) to a total volume of 15µL. The mix was added to the chilled 5µL RNA containing tube and mixed carefully. Samples were placed in a controlled-temperature heat block or PCR machine and incubated at 25°C for 5 minutes, followed by one hour step at 42°C. The reverse transcriptase was inactivated by heating the samples to 70°C for 15 minutes. The cDNA was diluted with an appropriate volume of DEPC-treated water (usually 300µL) and stored at -20°C until use.

2.3.4. Quantitative real-time polymerase chain reaction (qPCR)

Quantitative real-time PCR (qPCR) and reverse-transcription PCR (qRT-PCR) were performed on purified ChIP DNA or cDNA samples, respectively. Each 15µl reaction consisted of 5µL DNA sample, 7.5µL 2x SensiMix SYBR Green Hi-ROX mastermix (Quantace/Bioline Reagents, UK), 0.375µL each of forward (10µM) and reverse (10µM) primers and 1.75µL H_2O. The qPCR was performed on a Rotor-Gene 6000 (Corbett), with the following PCR conditions: 10 minutes hold at 95°C, followed by 45 cycles at 95°C for 10 seconds, 60°C for 15 seconds annealing step and 72°C for 15 seconds elongation step. A melting curve was generated after each experiment (from 65°C to 95°C, raising by 1°C each step, read every 5 seconds) to verify the specificity of the amplification. Detection was monitored by measuring the

increase in fluorescence throughout the cycle, with SYBR Green I having an excitation and emission maxima of 494 nm and 521 nm, respectively. Each qPCR reaction was run in duplicates. A sample with water instead of DNA was always used as control to rule out contamination or formation of primer dimers.

2.3.5. Primer design for SYBR Green qPCR and qRT-PCR

Primers were designed using the Primer 3 program (http://frodo.wi.mit.edu/) and the majority of the DNA sequences of interest were obtained from UCSC genome browser (http://genome.ucsc.edu). The product size range was limited to 80-120 bps, with primer size between 18-27 bps, melting temperatures of 59-61°C and the maximum Tm difference of 1.0°C. All primers were checked using the *in-silico* PCR function in the UCSC genome browser. If designing primers for real-time qRT-PCR, primers that span an intron (ideally quite large) were chosen, so that amplification product from cDNA is significantly different in size to amplification product from genomic DNA. All primers were ordered from Invitrogen, desalted and without any 5' or 3' modifications. Each primer pair was tested for efficiency by performing a serial 5-fold dilution curve on cDNA or genomic DNA. Only primers sets yielding a single amplification product, with a clean melting curve and good PCR efficiency were used. Due to the high number of ChIP-qPCR/qRT-PCR primer sets used in this study, their sequences are not included but are available upon request.

2.3.6. Microarray gene expression analysis

For microarray studies, RNA integrity was assessed on a BioAnalyzer. All samples had a RNA Integrity Number (RIN) ≥9.5 (Agilent Laboratories, Santa Clara, CA). Sense ssDNA was generated from 300 ng starting RNA with the Ambion WT Expression Kit according to the manufacturer's instructions. Sense ssDNA was fragmented and labeled using the GeneChip WT Terminal Labeling and Controls Kit. The

fragmented peak size was measured on the BioAnalyzer and was in the expected 40–70 nt range. The labeled ssDNA from four biological replicates was hybridized to Affymetrix Mouse Gene 1.0 ST Array (Affymetrix, Santa Clara, CA). Chips were processed on an Affymetrix GeneChip Fluidics Station 450 and Scanner 3000. Cel files were generated using Command Console (Affymetrix).

Affymetrix microarray probe intensities were normalised using RMA normalisation via the Bioconductor/R package oligo (version 1.18.1) (Carvalho and Irizarry, 2010) and differences tested using the Bioconductor/R package limma (version 3.10.3) (Smyth, 2004). In order for expression to have been considered changed, a majority of the probe sets targeting the gene needed a Benjamini and Hochberg corrected FDR of less than 0.05 and the most significant probe required a fold change of at least 1.5-fold between KDM2B knockdown and scrambled control samples. Enrichments in RING1B marked genes were tested by calculating the proportion of genes significantly changed compared to the fraction of genes on the array marked. Differences were tested using the hypergeometric test. Odds ratio for the RING1B marked genes to increase in expression vs decrease was tested using Fisher's exact test.

2.3.7. RNA-Seq gene expression analysis

The bioinformatics analysis was performed by Dr. Rob Klose and Mr. Hamish King. Briefly, for RNA-seq analysis paired end 51-bp reads were aligned to the mm9 reference genome using Tophat2 (v0.5) with default parameters. RPKM values were obtained from biological triplicates of the Kdm2bfl/fl and 72 hr tamoxifen treated cells using cuffdiff(v2.1.1) on a refGene set obtained from the UCSC table browser. For RNA-seq analysis each RING1B interval was annotated with the closest gene, and the fold change between mean RPKM values for treated and untreated KDM2b fl/fl cells was determined. Fold changes in RING1B, SUZ12 and RNA expression at RING1B peaks were plotted and tested for significance with the Welch Two Sample T-Test using R (version 3.0.1). Correlations between fold changes in RING1B and SUZ12, and RING1B and RNA levels, were tested using Pearson’s Product-Moment Correlation statistic in R. Lines of best fit were generated using a standard linear regression algorithm in R.

2.4. Protein methods

2.4.1. Total protein extraction (whole-cell extract, WCE)

Harvested mammalian cells were washed in 1xPBS and pelleted at 1000rpm x 5minutes.The cells were resuspend in an equal pellet volume of tissue culture lysis buffer, containing 1% NP-40, 150mM NaCl, 10% glycerol and 50mM TrisHCl pH 8.0, with 1mM DTT and 1x complete protease inhibitors (complete, EDTA-free, Roche, Basel, Switzerland) (both added fresh before use) and allowed to incubate on ice or at 4°C with gentle rotation for 30minutes. The lysed cells were spun at 13000rpm x 20 minutes, and the supernatant saved as whole cell extract.

2.4.2. Nuclear extract (NE) preparation

Cells were harvested by scraping, washed with 1× PBS and equilibrated in 10 pellet volumes of buffer A (10 mM Hepes pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF, and complete protease inhibitors (PIC) (Roche, Basel, Switzerland) added fresh before use). The cells were incubated on ice for 10 min and recovered by centrifugation at 1500×g for 5 min. The pellet was resuspended in 3 volumes of buffer A supplemented with 0.1% NP-40 and incubated on ice for 10 min with gentle agitation. Nuclei were recovered by centrifugation at 1500× rpm for 5 min. Recovered nuclei were then resuspended in 1× pellet volume of buffer B (5 mM Hepes [pH 7.9], 26% glycerol, 1.5 mM MgCl2, 0.2 mM EDTA, and complete protease inhibitors [Roche], 0.5 mM DTT) supplemented with 400 mM NaCl. The nuclei were first resuspended in a buffer with 250 mM salt. The volume of the buffer and the cells was taken into account, and the salt concentration was increased to 400 mM by gently adding concentrated 5 M NaCl dropwise while mixing. The extraction was allowed to proceed for 1 hr on ice with occasional agitation, and then the nuclei were pelleted at 13,000 rpm for 20 min at 4°C. The supernatant containing the nuclear cell extract was quantified using the BioRad Bradford protein assay (BioRad, Hemel Hempstead, UK) and stored at -80°C until use.

2.4.3. Protein concentration determination by Bradford assay

Bradford assay was carried out to determine protein concentrations. Briefly, Bradford dye (BioRad) was diluted 1:5 in water and mixed thoroughly. A series of volumes from the protein sample of interest (generally 1-5µL) was each added to 1mL of diluted Bradford reagent solution. OD at 595nm was measured using an UltroSpec 3300 spectrophotometer (GE Healthcare). In addition, a standard curve was generated with known concentrations of BSA (New England Biolabs), ranging from 1-10 ng/mL. Comparison of absorbance values with the standard curve allowed an approximate determination of the protein concentration for the sample of interest.

2.4.4. Acidic histone extraction

Cell pellets were washed in 1ml RSB buffer (10 mM Tris HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2), then lysed in RSB buffer supplemented with 0.5% NP-40 and incubated on ice for 10 minutes. After centrifugation at 500xg for 5 minutes, the nuclear pellet were resuspended in 0.5 ml 5 mM $MgCl_2$ and 0.5 ml 0.8 M HCl, and histone extraction allowed to proceed on ice for 20 minutes. The supernatant containing the extracted histones was collected by centrifugation at 13000xrpm for 20 minutes. The histones were subsequently precipitated by the addition of ice-cold TCA to a final concentration of 25% and incubation on ice for 30 minutes. After pelleting the histones by spinning (13000xrpm, 15 minutes), the histone pellet was washed twice with cold acetone without disturbing it. Following the last centrifugation step (13000xrpm, 15 minutes), the acetone was carefully aspirated and the histone pellet air-dried for 15-20 minutes at room temperature. The pellet was then resuspended in approximately 100µl 1x SDS-PAGE loading buffer and boiled at 95°C for 5 minutes before storage at -20°C until use.

2.4.5. SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated according to their molecular weights by gel electrophoresis, based on the principle described by Laemmli (Nature 227,680-685 (1970)). The stacking gel was composed of 5%

Acrylamide/Bis solution (BioRad, Hemel Hempstead, UK), 0.1% SDS, 0.1% ammonium persulphate (APS, Sigma), 0.1% *N*,*N*,*N*′,*N*′-Tetramethylethylenediamine (TEMED, Sigma) and 125mM Tris-HCl pH 6.8. The separating gel had an acrylamide concentration corresponding to the size of the protein of interest (i.e.: KDM2B resolves best on 6% separating gels, while PCGF1 is best resolved on 12% gels) ranging from 6-18% final concentration, with 0.1% SDS, 0.1% APS, 0.1% TEMED, Sigma and 400mM TrisHCl pH 8.8. All gels were cast and run using the Mini-Protean Tetra Cell system (BioRad). The protein samples were mixed with 1x SDS loading buffer(2% SDS, 100mM Tris pH 6.8, 100mM DTT, 10% glycerol and 0.1% bromophenol blue), and boiled at 95°C for 5 minutes. The gels were run in 1x SDS-PAGE running buffer (25mM Tris, 192mM glycine, 0.1% SDS) at 200V constant voltage until the bromophenol blue dye migrated off the gel. The gels were either probed by Coomassie blue protein stain, or used directly for Western blotting semi-dry protein transfer. Coomassie blue stain consisted of 50% methanol with 10% glacial acetic acid and 0.1% Brilliant blue R250 dye (Sigma). The gels were incubated with the stain for 15 minutes – 1 hour with gentle shaking, and destained until satisfactory resolution was achieved using 30% methanol, 10% acetic acid solution.

2.4.6. Western blot (WB) and enhanced chemiluminescence (ECL)-based detection

Following SDS-PAGE, the gels were transferred to 0.2μm (for histones) or 0.45μm (all other proteins) nitrocellulose membrane (BioRad). The nitrocellulose membrane was soaked in transfer buffer (39mM glycine, 48mM Tris, 0.037% SDS and 20% methanol) and placed on top of three layers of 0.3mm Whatman filter paper pre-soaked in the same buffer. After placing the gel on top of the membrane, three additional layers of pre-soaked filter paper were added on top. Semi-dry transfer was carried out for 1hr 30minutes at 200mA constant and 23V maximum in a Trans-Blot SD Semi-dry transfer cell (BioRad). The membrane was subsequently blocked using 5% milk in 1x PBS + 0.1% Tween 20 (Fisher) (PBST) solution for 30 minutes at room temperature. Incubation with primary antibody (generally used at 1:1000 dilution) was done in the same 5% milk/PBST solution overnight at 4°C. Following 3 x 5 minutes wash steps using PBST, HRP-conjugated secondary antibody (1:5000 dilution) in 1% BSA / PBST

solution was added for 1 hour at room temperature with gentle rocking. The following secondary antibodies were routinely used: ECL anti-mouse IgG HRP linked whole antibody from sheep (GE Healthcare), ECL anti-rabbit IgG HRP linked whole antibody from donkey (GE Healthcare) and ECL anti-sheep/goat IgG HRP linked from donkey (AbD Serotec). After the 3 x 5minutes wash steps in PBST, equal volumes of ECL solution I (2.5mM luminal (Sigma, made fresh in DMSO), 396µM p-coumaric acid (Sigma, made fresh in DMSO) and 100mM Tris HCl pH 8.5) and ECL solution II (5.6mM H_2O_2 and 100mM Tris HCl pH 8.5) were added to the membrane and allowed to react for 30 seconds – 1 minute. The membrane was exposed to Fuji Medical X-Ray film (Fisher) and developed using an Xograph Compact X4 system.

2.4.7. Superose 6 gel filtration chromatography

Size exclusion chromatography was performed using an FPLC system (GE Healthcare) under conditions recommended by the manufacturer (0.2ml/min flow rate and maximum 1.5MPa pressure). Briefly, a superose 6 10/300 GL gel filtration column (GE Healthcare) was pre-equilibrated with gel filtration standards thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75kDa), and ovalbumin (43 kDa). Approximately 2mg of nuclear extract prepared from EFC1 feederless mouse ES cells were loaded onto the column pre-equilibrated with buffer BC300 (50 mM HEPES pH 7.9, 300 mM KCl, 10% glycerol, 0.5 mM DTT added fresh). Fractions were collected (250µL), trichloroacetic acid (TCA)-precipitated, and subjected to Western blotting to identify KDM2A, KDM2B and RING1B – containing fractions.

2.4.8. Standard protein immunoprecipitation (IP)

For each immunoprecipitation, 0.5-1 mg nuclear extract was used in buffer BC150 (150 mM KCl, 10% glycerol, 50 mM HEPES pH 7.9, 0.5 mM EDTA, 0.5 mM DTT and 1x PIC [DTT and PIC added fresh before use]). Approximately 3 µg of antibody of interest was added and the reaction allowed to incubate overnight at 4°C with rotation. The next morning, 20 µL of packed protein A beads (RepliGen, Waltham, MA) were added, and the samples rotated for 1 hr at 4°C. If the antibody class and isotype required,

protein G sepharose beads (GE Healthcare) were occasionally used for optimal immunoprecipitation (i.e. Flag M2 mouse monoclonal antibody works best using protein G beads – based IP). Following three wash steps with BC300 (300 mM KCl, 10% glycerol, 50 mM HEPES [pH 7.9], 0.5 mM EDTA, 0.5 mM DTT [added fresh]), the beads were recovered by centrifugation at 1000xg for 1 minute, resuspended in a small volume of SDS-PAGE loading dye and boiled at 95°C for 5 min. After spinning at 1000×g for 5 min, the supernatant was used for western blotting, together with input and flow-through sample representing uncaptured material.

2.4.9. One-Step Strep protein complex purification approach

For purification of KDM2B-Flag/2xStrepII, the salt concentration of the nuclear extract was first reduced to 150 mM NaCl by dilution into nuclear extraction buffer B without salt (see Nuclear extraction preparation section). Between 10–15 mg of nuclear extract were used for each large-scale affinity purification. Avidin (IBA, Goettingen, Germany; 20 µg/1 mg extract) was added to remove biotinylated molecules which would bind unspecifically to the matrix. If benzonase treatment was performed, 75 U/mL nuclear extract benzonase nuclease (Novagen, Darmstadt, Germany) was added at this point. Extracts were pre-cleared for 30 min at 4°C, followed by a max speed centrifugation (13000× rpm) for 5 min. Cleared extract was then added to 50 µL packed StrepTactin superflow high-capacity resin (IBA) and allowed to rotate gently for 3 hr at 4°C. Eight wash steps were performed (20 mM Tris pH 8.0, 500 mM NaCl, 0.2% NP-40, 1 mM DTT (fresh) and 5% glycerol) in 1.5 mL low bind tubes (Eppendorf, Hamburg, Germany). Bound material was eluted in buffer containing 20 mM Tris pH 8.0, 150 mM NaCl, 0.2% NP-40, 1 mM DTT (fresh), 5% glycerol and 10 mM D-desthiobiotin (IBA). The same experimental setup was used to identify associated polypeptides for the other proteins used as baits.

Before submission for LC-MS/MS analysis, the efficiency of the purifications was assessed by running the elutions together with input nuclear extract and resin flow-through on gradient SDS-PAGE gels. For this purpose, either Mini-Protean TGX 4-20% (BioRad, Hemel Hempstead, UK) or Novex NuPAGE 4-12% Bis-

Tris (Invitrogen, Carlsbad, CA) gels were used. To confirm that proteins were pulled down in the purification, the gels were stained either by SyproRuby protein gel stain (BioRad, Hemel Hempstead, UK) or by using the SilverQuest Staining kit (Invitrogen, Carlsbad, CA). The gels were subsequently visualized with the Fuji FLA7000 phosphoimager. In addition, the elutions were probed by Western blotting against the 2xStrepII-tag, to confirm that at least the protein used as bait was pulled down during the purification.

2.4.10. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) and data analysis

The immunoprecipitated solution was desalted using chloroform-methanol extraction, followed by overnight in-solution tryptic digestion and subsequent peptide purification using Waters C18 Sep-Pak cartridges. The analysis of immunoprecipitated digested material was performed by LC–MS/MS using an Orbitrap Velos (Thermo mass spectrometer) coupled to a nano-UPLC system (NanoAcquity, Waters, Milford, MA) using a reversed phase 75 µm × 250 mm, 1.7µm particle size column, as described (Fischer et al., 2012). MS/MS spectra were searched against the UniProt SwissProt Mus database (v2011.11.18, 16,460 sequences) in Mascot v2.3.01, allowing one missed cleavage and 20 ppm/0.5 Da mass deviations in MS/MSMS, respectively. Carbamidomethylation of cysteine was a fixed modification. Oxidation of methionine, and deamidation of asparagine and glutamine were used as variable modifications. Protein assignment was based on at least two peptides identified. Proteins shown as significant hits have at least one unique peptide unambiguously assigned.

2.5. Antibody generation and purification

Three different rabbit polyclonal antibodies were generated during this study, and I will discuss them individually as each antibody has unique antigen design, expression and purification characteristics. All

rabbit immunizations were conducted at PTU/BS Scottish National Blood Transfusion Service, with roughly 20mL of serum delivered for each bleed number. Approximately 60mL of serum was collected for the terminal bleed. In addition to these three in-house produced antibodies, RING1B mouse monoclonal antibody was affinity purified from hybridoma cell tissue culture supernatant obtained from (Atsuta et al. 2001). This affinity purification of RING1B antibody was performed in order to minimize variability of antibody titers between different hybridoma supernatant batches. This was especially important for ChIP-qPCR/ChIP-Seq experiments where RING1B chromatin binding profiles were compared between different samples.

2.5.1. Generation and purification of KDM2B antibody

2.5.1.1. KDM2B Antigen design

A polyclonal antibody against KDM2B was generated by immunizing a rabbit with a Hisx6-tag fusion protein encoding amino acids 755–917 of human KDM2B protein, in a region showing most divergence from the highly similar protein KDM2A (see figure 3.1). The sequence of interest was PCR-amplified from KDM2B cDNA clone (a gift from Ryo Koyama-Nasu) and cloned into 6xHis-tagged/TEV-cleavable LIC adapted pNIC28 prokaryotic expression vector (a gift from U. Oppermann) to generate kanamycin-resistant construct pNIC28-KDM2B-2. All PCR generated constructs were verified by sequencing.

2.5.1.2. Expression and purification of 6xHis-tagged KDM2B antigen

To express and purify recombinant His-tagged KDM2B antigen fusion protein, 1µL (roughly 50-100ng/µL) sequenced construct miniprep was transformed into competent BL21 bacterial cells by heat shock method. BL21-based bugs are competent *E. coli* that are suitable for high level expression/production of heterologous proteins regulated by various expression vector systems. The bacteria also contain a

chloramphenicol-resistant pLysS plasmid which encodes the T7 lysozyme. T7 lysozyme lowers the background expression level of target genes under the control of the T7 promoter but does not interfere with the level of expression achieved following induction by IPTG. Generally, 25mL of overnight culture grown at 37°C of BL21-based bugs containing the KDM2B antigen-expression construct were added to 500mL of 2xTY media (16g/L tryptone, 10g/L yeast extract, 5g/L NaCl in distilled H_2O, pH adjusted to 7.0 and autoclaved), supplemented with antibiotics (final conc. = 50µg/mL of kanamycin & 34µg/mL of chloramphenicol, diluted from 1000X stock). The culture was grown at 37°C until OD_{600} reached approximately 0.5 and the cells were in exponential growth phase. Protein expression was induced with 1mM IPTG for 3 hours at 30°C. Small aliquots (5-10µL) of pre-induced and induced samples were collected for later analysis of efficiency of IPTG induction. The bacterial cells were harvested by spinning at 6,000xg for 20 minutes at 4°C in a Beckman centrifuge with JLA-8.1000 1L rotor. The cells were resuspended in lysis buffer (approximately 10mL buffer for 500mL culture), which consisted of 20mM Tris.HCl pH 8.0, 500mM NaCl, 0.1% NP 40 and 1X EDTA-free protease inhibitor cocktail (Roche). The resuspended cells were sonicated on ice using a VCX130PB microtip sonicator (Sonics & Materials) at 60% amplitude for 6 cycles (30seconds on/off) to release protein content. No more than 20mL of cells per tube were sonicated at once. The lysate was then cleared by centrifugation at 12,500rpmx20 minutes at 4°C.

The derived whole cell lysate was incubated with a corresponding amount of IMAC Sepharose 6 Fast Flow Ni-NTA Resin (GE Healthcare) pre-washed 2x times in the lysis buffer described above and allowed to rotate for 1 hour at 4°C. The amount of resin used depends on the amount of protein to purify, with an estimated binding capacity of 40mg/mL for proteins <30kDa. For KDM2B antigen, generally 1mL of Ni-NTA resin was used for lysate from 3L of induced bacterial culture. The lysate-resin mix was transferred to a disposable 10mL poly-prep chromatography column (BioRad) and the unbound material allowed to flow through. The resin was washed with 20-40 column volumes of Ni-NTA wash buffer (50mM NaH2PO4, 300mM NaCl, 20mM imidazole, pH adjusted to 8.0 with NaOH). Proteins were eluted 5-10 x times with 1 column volume of Ni-NTA elution buffer (50mM NaH2PO4, 300mM NaCl, 250mM imidazole, pH adjusted to 8.0 using NaOH). Generally, 5µL of each fraction was ran on SDS-PAGE and the

purity assessed by Coomassie blue protein staining. The most concentrated fractions were pooled together and dialysed overnight at 4°C using a sealed Spectra/Por dialysis tubing (Spectrum) with a molecular cut off range (MWCO) of 6-8 kDa in 2L of BC150 buffer (50mM Hepes pH 7.9, 150mM KCl, 10% Glycerol, 0.5mM DTT (add fresh)). Following dialysis, the pooled sample was collected and cleared of any precipitation by centrifugation at 13,000rpmx10 minutes at 4°C. Protein concentration was measured by Bradford assay and also by comparison with other proteins of known concentration by SDS-PAGE and Coomassie blue staining. Protein sample was stored at -80°C until use.

For immunization, at least 4x0.5mg antigen aliquots in less than 250µL volume were sent to PTU/BS.

2.5.1.3. Affigel-10 affinity purification of KDM2B antibody from serum

KDM2B antigen was coupled to AffiGel10 resin (BioRad, Hemel Hempstead, UK) and the antibody was affinity purified. Both Affi-Gel 10 and 15 supports are N-hydroxysuccinimide esters of a derivatized crosslinked agarose gel bead support, and both couple to ligands spontaneously in aqueous or non-aqueous solution. Affigel-10 was chosen over Affigel-15 as this support is recommended for coupling proteins best at a pH near or below their isoelectric point, in fitting with the pI of KDM2B antigen. Briefly, 500µL Affigel-10 was used for 10mg of antigenic peptide. The resin was washed 1x with 1mL MilliQ water, followed by two wash steps with 1mL PBS. 10mg of purified antigen was added to the washed affigel support and allowed to incubate at 4°C overnight or at RT for 3 hours gently rotating. The supernatant containing unbound material was collected, and the beads washed with 1mL of 0.2M ethanolamine solution. The beads were subsequently blocked with 1mL ethanolamine at 4°C for 2 hours, before loading onto a 10mL disposable BioRad column. The resin containing bound antigenic peptide was washed with 25mL of cold 1M NaCl and 25mL of cold PBS.

For antibody purification from serum, the beads were incubated with 10mL of serum from the immunized rabbit; the column was sealed with parafilm and allowed to rotate at RT for 3 hours. The affigel resin was washed with 25mL of 0.5M NaCl and 25mL of 1x PBS solution. The purified antibody

was eluted using 500µL 0.1 M glycine pH 3 into 50uL of 1M Tris pH 8.0 and the elutions ran on a 12% SDS-PAGE gel followed by Coomassie R-250 staining. The fractions were pooled together and concentrated using Vivaspin 20 columns (Sartorius Stedim Biotech) and brought to 100mM NaCl and 10% glycerol, aliquoted into 100µL samples and stored at -80°C until use. Importantly, the KDM2B antibody purification was highly efficient, yielding pure fractions of heavy and light chain in high concentrations. The effectiveness of the affinity purification procedure was assessed by checking the antibody response in Western blotting.

2.5.2. Generation and purification of Flag-2xStrepII (FS2) antibody

2.5.2.1. Antigen design

To generate an antibody against the Flag/2xStrepII tag, the following peptide covering the tags (underlined) and the linker regions between them was custom synthesized (GL Biochem, China), at 30-39mg and 80% min HPLC (see also supplementary figure S2):

MDYKDDDDKGSAASWSHPQFEKGGGSGGGSGGGSWSHPQFEKGARSC

The peptide was synthesized with a terminal cysteine residue. This has a sulfhydryl group that provides a highly specific conjugation site for crosslinking to the highly immunogenic keyhole limpet hemocyanin (KLH) carrier protein for immunization and antibody production. The synthesized peptide lyophilized powder was stored at -80°C until use.

2.5.2.2. Peptide coupling to KLH for antibody production

Keyhole limpet hemocyanin (KLH) is widely used as a carrier protein for conjugation to haptens and other antigens to make them more immunogenic for the purpose of antibody production. For coupling of the synthesized peptide, Imject Maleimide-Activated Mariculture KLH from Pierce (cat no 77606) was

used. 2mg of maleimide-activated KLH was gently dissolved in 200µL MilliQ water. Harsh pipetting or passing through a syringe will cause the KLH to precipitate out of solution and should be avoided. The peptide was dissolved in conjugation buffer (Pierce, cat no 77164) to a concentration of approximately 4mg/ml and 2mg of dissolved peptide allowed to incubate with the KLH solution for 2 hours at RT with gentle rotation. A 5mL desalting column (GE Healthcare) was equilibrated in Imject Purification Buffer Salts (Pierce, cat no 77159). The sample was applied by gentle injection to the desalting column and 5-6x 500µL fractions were collected after elution with Imject Purification buffer salts. This buffer-exchange of immunogens following conjugation is necessary to stabilize the product during freeze-thaw cycles. 3µL of each fraction was mixed with 200µL Bradford solution to determine the peak fractions. These were combined and the concentration was roughly estimated by Bradford assay. The KLH-conjugated immunogen sample was aliquoted into 0.5mg aliquots in less than 250µL per aliquot and stored at -80 until sent on dry ice to PTU/BS.

2.5.2.3. Sulfolink-based affinity purification of FS2-antibody from serum

Sulfolink coupling resin is routinely used to immobilize peptides having terminal cysteine residues to purify antibodies that were generated against peptide immunogens prepared by maleimide conjugation. 4mL of ~50% Sulfolink resin slurry (ThermoScientific/Pierce cat no 20401) was transferred into 15mL Falcon tube and spun at 2,000 rpm x 3 mins at 4°C. The resin was resuspended and equilibrated in 8mL of coupling buffer (50mM Tris, 5mM EDTA, pH adjusted to 8.5), followed by a centrifugation step at 2,000 rpm for 3 mins at 4°C. 4mL of synthesized peptide dissolved at a concentration of 0.5 mg/mL in coupling buffer was added to the Sulfolink resin. The peptide-Sulfolink resin sample was rotated at RT for 15 minutes and then allowed to sit for 30 minutes at RT. Following a short spin at 2,000rpm x 3minutes, the resin was washed twice with 6mL of coupling bugger spinning as above between the steps. Uncoupled sites on the Sulfolink resin were blocked by addition of 2mL of 50mM L-Cysteine HCl (made fresh before use) and incubation for 15 minutes with gentle rotation and then 30 minutes sit at RT. After centrifugation (2,000rpm x 3 mins), the supernatant was discarded and the Sulfolink resin-

coupled antigenic peptide was resuspended in 5mL of 1M NaCL and applied directly to a small 10mL BioRad disposable column. The column was washed twice with 3mL of 1M NaCl, and 3 times with 5mL of cold 1x PBS. The column can be used directly afterwards or can be stored at 4°C in 1xPBS until use.

For antibody purification, 10mL of serum was added to the sealed column and allowed to incubate at RT for 3 hours with gentle rotation. The flow through serum was collected and stored at -20°C in case the purification was not efficient. The column was washed with 25mL of 0.5M NaCl and 25mL of 1xPBS. The purified antibody was eluted using 1mL of L 0.1 M glycine pH 3 into 100uL of 1M Tris pH 8.0 and the elutions ran on a 12% SDS-PAGE gel followed by Coomassie R-250 staining. The fractions (generally between 10-16 elutions) were pooled together and concentrated using Vivaspin 20 columns (Sartorius Stedim Biotech) and brought to 100mM NaCl and 10% glycerol, aliquoted into 100µL samples and stored at -80°C until use.The column was equilibrated with 0.1M Tris pH 8.0, washed extensively with 1xPBS and stored at 4°C for re-use if within a year.

2.5.3. Generation and purification of PCGF1 antibody

For generation of an antibody specifically recognizing PCGF1, multiple attempts with different antigenic regions (either protein fusions or custom peptides) were made. The antigenic regions selected which show the highest divergence from the other 5 PCGF proteins are unstable at physiological pH and precipitate during expression or purification. Therefore, I will briefly describe the antigen design and purification scheme for the successful PCGF1 antibody generated and used extensively in Chapter 6.

2.5.3.1. Antigen design

For generation of a polyclonal antibody against PCGF1, a custom peptide was synthesized (GL Biochem, China, 30-39mg and 80% min HPLC) against the following sequence covering the N-terminus of human PCGF1:

MASPQGGQIAIAMRLRNQLQSVYKMDPLRNEEEVRC

As described above for FS2, the peptide was synthesized with a terminal cysteine residue to allow crosslinking to KLH-carrier protein. The synthesized peptide lyophilized powder was stored at -80°C until use.

2.5.3.2. Peptide coupling to KLH for antibody production

Coupling to Imject Maleimide-Activated Mariculture KLH was performed as described for FS2 antibody, with two important modifications. As the synthesized peptide was insoluble in Pierce conjugation buffer, DMSO was added to a final concentration of 20%. Due to the instability of the peptide even after KLH-conjugation, the desalting step was no longer performed, as the conjugate precipitated even if DMSO was added to the Purification Buffer salts during desalting. However, this meant that the buffer would not be exchanged to the stabilizing Imject Purification buffer salts. Therefore, custom conjugation buffer was prepared (0.083M sodium phosphate, 0.1M EDTA, 0.9M NaCl, 0.1M sorbitol, pH 7.2) which did not contain any sodium azide to allow direct injection into animals. The KLH-peptide conjugate was divided into 4x aliquots and shipped on dry ice to PTU/BS.

2.5.3.3. Sulfolink-based affinity purification of PCGF1-antibody from serum

PCGF1 peptide was covalently immobilized on SulfoLink resin (ThermoScientific/Pierce) as described previously for FS2 antibody. However, 20% final DMSO concentration was added at all steps during the Sulfolink resin-peptide column preparation. Approximately 10mL of serum was used for the purification. The purified antibody was eluted using 1mL of L 0.1 M glycine pH 3 into 100uL of 1M Tris pH 8.0 and the elutions ran on a 12% SDS-PAGE gel followed by Coomassie R-250 staining. The fractions (generally between 10-16 elutions) were pooled together and concentrated using Vivaspin 20 columns (Sartorius Stedim Biotech) and brought to 100mM NaCl and 10% glycerol, aliquoted into 100µL samples and stored at -80°C until use. The column was equilibrated with 0.1M Tris pH 8.0, washed extensively with 1xPBS and stored at 4°C in PBS.

2.5.4. Protein A purification of RING1B mouse monoclonal antibody from hybridoma tissue culture supernatant

2.5.4.1. Growth and maintenance of antibody expressing hybridoma cells

The hybridoma cells were generated and described by Atsuta et al. (2001) and kindly provided by Dr. Haruhiko Koseki. The cells were thawed according to standard procedure (see Tissue culture methods) into RPMI 1640 medium (Gibco) supplemented with 10% FCS, 100 µg/ml penicillin/streptomycin and 2 mM L-glutamine, with the note that these cells are adapted for growth in suspension. Although no significant attachment to plastic surfaces is expected, when it happens a single pipetting up and down is enough to detach the cells. The hybridoma cells can be passaged by directly transferring a certain volume of cells into new media (for example 1ml into 9ml fresh media per 10xcm dish), as the cells do not secrete toxic by-products. Alternatively, the cells can be centrifuged at 1000xrpm for 3minutes, the pellet resuspended and transferred to fresh plates. The cells can be frozen while still in serum-containing media according to standard protocol (10% DMSO + 90% FCS). When cell density reached 4million/ml, they were changed to AIM V serum-free media (Gibco), supplemented with 100 µg/ml penicillin/streptomycin and 2 mM L-glutamine. Media was collected every 36-48 hours at this density, or sooner (24 hours) if media became yellow. Media was collected and spun at 1000xrpm for 3 minutes. The supernatant containing the antibody was stored at 4°C until use and the cell pellet was returned to fresh media for a new collection. Three batches were usually collected in total before the cells looked less healthy (grainy appearance under the microscope). Although some protocols suggest filtering the supernatant using PVDF low protein binding filters 0.22-0.45µm, the supernatant was instead centrifuged again to remove any contaminating cells before use to avoid any antibody loss during filtration. The serum collected was combined (to a total of approx 700-800ml) and stored at 4°C for a maximum of 24-48 hours before the purification procedure.

2.5.4.2. Purification of mouse monoclonal RING1B antibody from hybridoma cells supernatant

RING1B mouse monoclonal antibody was purified from hybridoma cells culture supernatant using protein A beads (RepliGen), following the protocol described by Harlow and Lane (Using Antibodies, 1999). Protein A beads were chosen as the RING1B antibody class and specificity (namely IgG2b/k, as determined by Atsuta el al., 2001) allows for equally efficient purification using either protein A or protein G beads. Briefly, 1/10 volume of 1M Tris pH 8.0 was added to the collected serum to adjust the pH of the crude antibody preparation to 8.0. This was evident by a clear colour change from yellow-brown to dark pink when the pH changed. 1mL of packed protein A beads (RepliGen) were allowed to settle into large 20mL BioRad column. The beads were equilibrated with approximately 20 mL of 0.1M Tris pH 8.0 solution. The antibody solution was passed through the settled protein A beads column, flow-through was collected and applied again to the column for improved binding. Approximately 0.5mL input serum was retained for further testing.

The beads were washed with 10-20 volumes of 100mM Tris pH 8.0, followed by washing with 10-20 volumes of 10mM Tris pH 8.0 solution. The bound antibody was eluted using 1ml (1 column volume) of 0.1M glycine pH 3.0. Approximately 8 – 10 elutions were collected into tubed containing 100ul (1/10 column volume) of 1M Tris pH 8.0. Each elution was mixed gently to bring the pH back to neutral, otherwise the eluted antibody would be prone to denaturation. The immunoglobulin-containing fractions were identified by mixing 15ul of each elution into 200ul diluted Bradford reagent, and observing the colour change to blue. 5-10ul of each fraction was then run onto a 12% SDS-PAGE gel, together with input and flow-through material, followed by Coomassie R250 staining. Good purifications essentially yield pure heavy (~50kDa) and light (~25kDa) chain bands. The antibody containing fractions were combined, 5M NaCl was added to a final concentration of 50mM (for more sensitive antibodies, the salt concentration can be adjusted to 100-150mM) and glycerol to 10%. The combined antibody fractions were concentrated using Vivaspin 20 columns (MWCO of 5,000 Daltons, Sartorius), aliquoted and stored at -20°C. The relative antibody concentration was determined by running a 12% SDS-PAGE

gel comparing different volumes of the purified antibody to known concentrations of a standard protein / antibody preparation (i.e. BSA or Flag M2 antibody of 1µg/µL concentration). The purified antibody was tested in Western blot and ChIP experiments to determine its functionality and optimum working concentration for each application.

2.5.5. Other antibodies used extensively in this study

In addition to the rabbit polyclonal antibodies against KDM2B, Flag/2xStrepII and PCGF1 generated and characterized in this study, and the purification of RING1B mouse monoclonal antibody, other antibodies used extensively throughout this work are listed in Table 2.9, together with their experimental applications.

Table 2.9. List of main additional antibodies used in this study.

Antigen	Source of antibody	Application
EZH2	Cell Signaling 5246S	WB/ ChIP-qPCR
SUZ12	Cell Signaling 3737S	WB/ ChIP-qPCR/ ChIP-Seq
H3K27me3	Diagenode C15410069 (pAb-069)	WB/ ChIP-qPCR
H2AK119ub1	Described previously (Farcas et al. 2012)	WB/ ChIP-qPCR
HisH3	Described previously (Farcas et al. 2012)	WB/ ChIP-qPCR
KDM2A	Described previously (Blackledge et al. 2010)	WB/ ChIP-qPCR/ ChIP-Seq
LAMINB	Santa Cruz M20 (sc - 6217) and C20 (sc - 6216)	WB
STREPII	IBA (Fisher) IB21507001 (mAb)	WB
FLAG	Sigma Flag M2 F1804 (mAb)	WB/IF

(WB – Western blot, IF – immunofluorescence studies)

2.6. Chromatin Immunoprecipitation (ChIP)

For chromatin immunoprecipitation, feeder-dependent ES cells (i.e. V6.5, Ring1a$^{-/-}$; Ring1b $^{fl/fl}$ and Kdm2b $^{fl/fl}$ cells) were grown for at least one passage in feeder-free conditions on 0.1% gelatine-coated dishes. For tamoxifen-time course experiments, feeder-dependent conditional knockout cells were trypsinized, plated and allowed to settle overnight under feeder-free conditions. Early next day, of 4-hydroxytamoxifen (OHT) (Sigma) was added at a final concentration of 800 nM for the required time interval (i.e 24-72 hours). The cells were fed with fresh media supplemented with OHT on a daily basis. Feeder-independent ES cells were grown by standard method on 0.1% gelatine-coated 15xcm dishes in regular ES media supplemented with selection antibiotics, where required.

2.6.1. Chromatin preparation step

2.6.1.1. Single versus double cross-linked chromatin

Generally, single cross-linked chromatin was prepared when studying histone modifications, while double cross-linked chromatin was generated for chromatin-binding proteins. It is important to consider that not all antibodies behave in a similar manner, therefore it is essential to test both cross-linking methods for each new antibody used. For example, RING1B ChIPs work equally well with single and double cross-linked chromatin, while for KDM2A and KDM2B, double cross-linking significantly improves the ChIP enrichment levels obtained.

For KDM2A, KDM2B, PCGF1, RING1B, SUZ12 and EZH2 ChIPs, approximately 5x 10^7 cells were trypsinized and resuspended in 10 mL of 1x PBS pre-warmed in 33°C water bath. The cells were pelleted at 1000 rpm x 3 minutes, and the pellet resuspended in PBS as above. 23mg of Ethylene glycol bis[succinimidylsuccinate] (EGS) (ThermoScientific, cat no. 21565) were dissolved in 100 µL DMSO and 40 µL of this solution was added to 10 mL PBS/cells sample, with a final EGS concentration of 2 mM. The cells were fixed for 1 hour with EGS with gentle rotation at constant room temperature (23°C), followed

by 15 minutes crosslinking with 1% formaldehyde (Sigma, 270 µL of 37% FA solution were added to the 10 mL PBS/cells sample). The cells were allowed to incubate with formaldehyde for 15 minutes at set room temperature with gentle rotation. Formaldehyde was quenched by the addition of glycine to a final concentration of 125 µM (1.5 mL of 1M glycine stock was added to the 10 mL PBS/cells sample) and the solution left to incubate at room temperature for 3 minutes with rotation.

For histone modification ChIPs, generally ~1x10^7 cells were trypsinized, washed with 1x PBS as described above and fixed for 10 minutes with 1% formaldehyde, followed by glycine quenching. Importantly, H2AK119ub1 ChIPs require a higher amount of starting material to work efficiently, therefore ~5x10^7 cells were formaldehyde single cross-linked for this type of experiments.

2.6.1.2. Chromatin preparation by the SDS-method

Following formaldehyde quenching, the cross-linked cells were centrifuged at 500xg for 4 minutes at 4°C. Media was aspirated from cells and the pellet resuspended in 10 ml of ice cold 1xPBS. The PBS wash step was repeated, with the cells spun at 500xg for 4 minutes at 4°C in between. After the second PBS wash, the cell pellet was resuspended in fresh lysis buffer containing 1x protease inhibitor (PIC) (Roche) at a concentration of ~5x10^7 cells (chromatin binding proteins) or ~1x10^7 cells (histone modifications) per 1 mL of lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1). The resuspended cells were incubated in lysis buffer on ice for 30 minutes prior to sonication, with occasional gentle pipetting to avoid pelleting at the bottom of the tube.

2.6.1.3. Chromatin preparation by the Schmidt et al. (2009) method

Chromatin immunoprecipitation was performed as previously described (Schmidt et al. 2009), with minor modifications. Following formaldehyde quenching, the single or double cross-linked cells were

centrifuged at 2000 rpm for 4 minutes at 4°C. Media was aspirated from cells and the pellet resuspended in 10 mL of cold LB1 buffer (50 mM HEPES KOH, pH 7.9, 140 mM NaCl, 1mM EDTA, 10% Glycerol, 0.5% NP40, 0.25% Triton x 100, supplemented with 1x complete protease inhibitor cocktail). The cells were incubated with buffer LB1 for 10 minutes at 4°C with gentle rotation, to allow nuclei isolation. The successfully released nuclei were isolated by spinning the cells at 2000 rpm for 4 minutes at 4°C. The pellet was subsequently resuspended in 10 mL of ice cold buffer LB2 (10 mM Tris HCl, pH 8.0, 200 mM NaCl, 1mM EDTA, 0.5 mM EGTA, freshly supplemented with 1x complete protease inhibitor cocktail PIC). The resuspended pellet was gently rotated for 5 minutes at 4°C, and the nuclei pelleted at 2000 rpm for 4 minutes at 4°C.

The pellet was resuspended in an appropriate volume of buffer LB3 (10 mM Tris HCl, pH 8.0, 100 mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na Deoxycholate, 0.5% N-lauroylsarcosine, freshly supplemented before use with 1x protease inhibitor cocktail). 1 mL of LB3 buffer was used per ~$5x10^7$ of cells (chromatin-binding proteins) or for ~$1x10^7$ cells (in the case of histone modifications). If multiples of the required amount of cells was used, the pellet resuspended in the corresponding volume of LB3 buffer was divided into 1 mL aliquots in 15 mL sonication tubes and either used immediately for sonication or stored on ice with occasional mixing until sonicated. The Schmidt et al. method yields more reproducible chromatin preparations as it circumvents differential SDS-precipitation between samples and is therefore recommended.

2.6.1.4. Sonication

Sonication was performed in 1 mL aliquots in polystyrene 15 mL Falcon tubes. Chromatin was sheared using a BioRuptor sonicator (Diagenode) to produce fragments of approximately 500 bps (single cross-linked chromatin) or between 1-2 kb (double cross-linked chromatin). The sonicator water cooling system was set to 4°C and chromatin was sonicated for 30 minutes at full power with 30 second on/ 30 second off pulse, meaning that the actual sonication time is 15 minutes. If more than 1x1 mL aliquots

were sonicated, identical samples were pooled together. If chromatin was prepared by the (Schmidt et al. 2009) method, 1/10 volume of 10% Triton X-100 made up in LB3 was added to the sheared chromatin (i.e. 100 µl 10% Triton X-100 for 1 mL of chromatin) and the solution gently mixed. Chromatin samples were centrifuged in a 1.5 mL microcentrifuge tube at 13000 rpm for 10 minutes at 4°C to remove cellular debris. Chromatin samples were subsequently transferred to fresh tubes and either used immediately for immunoprecipitation or stored at -80°C until use.

2.6.1.5. Chromatin size verification

To test whether the sonication was efficient, the chromatin size is determined. Briefly, a 50 µL aliquot from 1 mL sheared chromatin is removed and 50 µL MQ water was added. To reverse crosslinking, 4 µL of 5 M NaCl solution was added, and the sample incubated at 65°C for at least 4 hours to overnight. Following addition of 4 µL DNase-free RNase (500 µg/mL), the sample was incubated for 1.5 hours at 42°C. 1 µL of Proteinase K (20mg/ml, Sigma) was added, and the solution was incubated at 45°C for 1 hour. DNA was recovered using either Invitrogen Purelink Micro PCR clean-up column or ChIP DNA Clean & concentration columns (Zymo) and eluted in 10 µL volume. The full volume was run on a 1% agarose gel with appropriate DNA marker for approximately 45 minutes at 100 V. Ethidium bromide staining allowed visualization and determination of chromatin size after sonication.

2.6.2. Immunoprecipitation step

2.6.2.1. ChIP setup using agarose beads

For chromatin-binding protein ChIPs, 100 µL of chromatin was used per IP, representing approximately $5x10^6$ cells/IP. For histone modification ChIPs, a lower concentration of cells (~$1x10^5$ cells) was used per individual immunoprecipitation reaction. For every IP to be performed, 100 µL chromatin was mixed

with 900 µL ChIP dilution buffer (1% triton X-100, 1 mM EDTA, 20 mM Tris-HCl (pH 8.0), 150 mM NaCl) supplemented with 1x complete protease inhibitor cocktail. If chromatin was prepared by SDS-method, it was centrifuged at 13 000rpm x 5 minutes at 4°C prior to use to remove any contaminating precipitates. If multiple IPs were performed, a chromatin + ChIP dilution buffer mastermix was prepared to minimize ChIP variability between samples. Approximately 40 µL blocked agarose A beads slurry (RepliGen) (equating to ~20 µL packed beads) was added per 1mL of chromatin/buffer mixture.

The beads were blocked for 1 hour at 4 °C with rotation in ChIP dilution buffer containing 1 mg/ml BSA and 1 mg/ml yeast tRNA and washed 3x times in ChIP dilution buffer prior to use. The remaining beads were stored at 4°C overnight in ChIP dilution buffer to the original slurry volume used.

Chromatin was pre-cleared by rotating the chromatin/beads mix at 4°C for at least 30 minutes, after which the beads were collected by centrifugation at 1000xg for 4 minutes at 4°C. For each IP to be performed, 1 mL aliquots from the pre-cleared chromatin master mix were transferred to fresh 1.5 mL LoBind tubes (Eppendorf). 100 µL Input sample was removed at this stage from remaining chromatin mastermix and stored at -20°C until use. The input therefore represents 10% of ChIP material.

Generally between 3-5 µg of antibody of interest was added to each sample. A non-specific antibody control (i.e. HA or Flag, if not epitope-tagged protein) were carried in parallel as control IPs. The antibody/chromatin mix was incubated overnight at 4°C with gentle rotation. Next day, 20 µL blocked and pre-washed agarose A beads was added per IP sample using cut-off tips and the mix rotated at 4°C for 1 hour. Beads were collected by centrifugation at 1000xg for 1 minute at 4°C. Washes were performed according to the Upstate protocol, with low salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 150 mM NaCl), high salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 500 mM NaCl), LiCl buffer (0.25 M LiCl, 1% NP40, 1% Deoxycholate, 1 mM EDTA, 10 mM Tris-HCl (pH 8.1)) and TE buffer (x2) (10 mM Tris-HCl (pH 8.0), 1 mM EDTA). All wash buffers were pre-cooled on ice prior to use. Each wash step was performed by resuspending the agarose beads in the appropriate wash buffer and the rotating at 4°C for 4 minutes. Samples were then centrifuged at 1000xg for 1 minute, supernatant removed and the beads resuspended in the next wash buffer. After the final

TE wash, beads were either resuspended in fresh elution buffer or directly mixed with Chelex-100 beads for ChIP DNA isolation (as detailed below).

2.6.2.2. ChIP setup using magnetic beads

The protocol is similar to the immunoprecipitation using agarose beads method, with a couple of essential modification. Firstly, the small, uniform size of the magnetic A beads and their reduced non-specific binding potential means that they does not require blocking and chromatin pre-clearing prior to use. Secondly, unlike the agarose beads-based IP protocol, the Dynabeads were first incubated with the antibody of interest, prior to addition of chromatin mastermix.

For each IP, 25 µ magnetic protein A Dynabeads (Invitrogen, cat no 10002D, 30mg/mL) were used. The beads were washed 3 times with 1 mL 0.5% BSA/1xPBS solution. After the last wash, the beads were resuspended in 1 mL 0.5% BSA/1xPBS and the required amount of antibody was added. The beads/antibody mix was incubated for at least 4 hours to overnight at 4°C with gentle rotation. The antibody-bound beads were collected using a DynaMag™-2 Magnet (Invitrogen) and washed 3 times with 0.5% BSA/1x PBS solution. 1 mL of chromatin/ChIP dilution buffer mix was added and incubated with the beads overnight at 4°C with gentle rotation. 100 µL from the remaining chromatin mastermix (or more if required) was collected as input and stored at -20°C until use. ChIP washes were carried out as described above (section 2.6.2.1).

2.6.2.3. ChIP setup for histone modifications versus chromatin-binding proteins

Although initial ChIP experiments were carried out using agarose beads for both chromatin-binding factors and histone modifications, Dynabeads were subsequently extensively used for protein factors as it generally yielded better enrichment levels and cleaner, more reproducible ChIPs with lower

background. It is especially recommended to use the Dynabeads-based ChIP approach for generating material for ChIP-Seq. Importantly, ChIPs for the study of histone modifications work best using agarose beads for the immunoprecipitation step.

2.6.2.4. Column-based ChIP DNA purification

After the final TE wash, the beads were resuspended in 100 µL fresh elution buffer (1 % SDS, 0.1 M NaHCO3) and vortexed at room temperature for 30 minutes. This elution approach was carried out only when ChIP samples were subsequently column-purified. The samples were centrifuged at 1000xg for 1 minute to collect beads and the eluate was transferred to fresh microcentrifuge tubes.

Following the elution step, 4 µL of 5M NaCl solution was added to the 100 µL Input and ChIP samples and mixed gently by repeated pipetting. The samples were incubated at 65° for at least 4 hours to overnight to allow cross-linking reversal. After at least 2.5 hours at 65°C, 2 µL RNase solution (Roche, 500 µg/mL) was added to each sample and the incubation continued for the remaining time. 1 µL Proteinase K (20mg/mL) was added per sample and allowed to react at 45°C for 1 hour. The samples were gently spun between the incubation steps to collect all condensate. The samples were purified using Invitrogen Purelink Micro PCR clean-up columns, following manufacturer instructions. Alternatively, for improved and more reproducible purification, Zymo ChIP DNA Clean & concentration columns were used. If samples were used for qPCR, ChIP DNA was eluted in 20 µL PCR grade H_2O and the final volume adjusted to 200 µL by addition of PCR grade H_2O. If used for ChIP-Seq, the samples were eluted in 10 µL sterile H_2O and if multiple identical ChIPs were carried out in parallel, they were pooled together in one tube following elution and mixed thoroughly prior to picoGreen DNA concentration quantification.

2.6.2.5. Chelex-100 resin-based ChIP DNA purification

ChIP washes were performed as usual, using either agarose or magnetic beads. If ChIP clean-up is carried out using Chelex-100 beads (BioRad), 40 µL of ChIP mastermix is removed as Input and diluted with 60µL MQ water prior to purification. During final ChIP wash with TE buffer, a 20% Chelex-100 bead slurry was prepared in PCR grade water (i.e. 200mg beads in 1mL H_2O) and 100µL of this 20% slurry was added to the Input sample. The remaining bead slurry was diluted 2-fold in MQ water to a final concentration of 10%. After the final TE wash, all supernatant was removed and 100µL of the 10% Chelex-100 slurry was added to each ChIP sample. The input and ChIP reactions were thoroughly mixed by vortexing for 10 seconds, and then the samples were boiled at 95°C for 10 minutes to reverse crosslinking. The samples were cooled to room temperature and briefly spun to pellet any condensate. 1.5 µL of Proteinase K was subsequently added and the samples were incubated at 55°C for 30 minutes with shaking. Proteinase K was heat-inactivated by boiling at 95°C for 10 minutes. The samples were centrifuged at 1000 rpm for 2 minutes to collect the beads (both Chelex-100 beads and the magnetic/agarose beads used during the ChIP procedure). The 100 µL supernatant was collected into fresh tubes. A second elution step was carried out by adding 100 µL of PCR grade H_2O to ChIP sample beads and 200µL PCR grade H_2O to Input sample beads. The tubes were mixed by vortexing briefly and centrifuged to collect beads. The supernatant containing ChIP DNA was transferred to the same tube as after the first collection, bringing the total volume for the ChIP sample to 200 µL and the input to 400µL. The DNA is ready for qPCR analysis, with input representing 2% of ChIP-ed DNA. Importantly, although the Chelex-100 ChIP DNA clean-up procedure reduces the time of cross-link reversal and DNA isolation, it does not yield pure DNA preparation. Therefore, this purification method can only be used for ChIP-qPCR –type experiments and not for ChIP-Seq, which requires column-based DNA purification.

ChIP qPCR data was expressed as either a percentage of input (y-axis labelled '% Input') as determined by the comparative C_T method (Livak and Schmittgen 2001) or enrichment relative to a control polycomb target site (i.e. *Bmp7* CGI) (y-axis labelled 'rel. enrichment').

2.6.3. ChIP-seq

2.6.3.1. ChIP setup for whole-genome sequencing

For ChIP-Seq analysis, generally approximately 10ng of purified DNA is required. To obtain this concentration, multiple ChIPs for the same factor of interest are pooled together after the column-based DNA isolation step. To generate the initial KDM2A and KDM2B ChIP-Seq profiles, 8x agarose beads IPs were pooled together. For ChIP-Seq studies in the KDM2B KD and KO experiments, magnetic Dynabeads were used for the immunoprecipitation step. 4x ChIP reactions were pooled together for KDM2B and SUZ12 ChIP-Seq, and 8x ChIPs were required to produce sufficient starting material for the RING1B ChIP-Seq investigation. Importantly, purified ChIP DNA material in the 50-350bp range is required for the ChIP-Seq library preparation stage. As double cross-linked chromatin is significantly larger than the required size, post-sonication of purified ChIP DNA was carried out. The ChIP sample was transferred to a 0.5 mL eppendorf tube and sonicated using a Bioruptor (Diagenode) for 1 hour at 30 second ON/ 30 second OFF pulse. In parallel, input DNA (diluted to a comparable concentration as ChIP DNA) was sonicated in the same manner. Input DNA can in principle be visualized on a 1% agarose gel to check size and efficiency of post-sonication, although it requires extensive ethidium bromide staining.

2.6.3.2. PicoGreen measurement of DNA concentration

Undiluted and a 1:10 diluted DNA sample purified after the ChIP procedure was quantified using PicoGreen® dsDNA Quantitation Kit (Molecular Probes). Briefly, 20x TE buffer (provided in the kit) was diluted to 1x using ultrapure MQ water. The concentration of input ChIP DNA was quantified by nanoDrop ND-100 spectrophotometer (Labtech) and used to generate a standard curve for picoGreen concentration measurements (generally the following input concentrations were measured: 1µg/mL, 200 ng/mL, 100 ng/mL, 40 ng/mL and 8ng/mL). 2µL of each input diluted were thoroughly mixed with

2µL diluted picoGreen solution and measured by nanoDrop 3300 fluorospectrometer (Labtech). Standard curve generated using input DNA of known concentrations was used to determine the concentration of purified DNA. At least 10ng total ChIP DNA in less than 50µL were sent for ChIP-Seq analysis. An exception to this was represented by RING1B ChIPs, for which the DNA concentration were generally in the range of maximum 1-2ng from 8x ChIP reactions pooled together.

2.6.3.3. ChIP-Seq analysis for KDM2B knockdown experiments

The bioinformatics analysis was performed by Dr. Rob Klose, in collaboration with Dr. Ian Sudberry and Dr. David Sims, from the group of Prof. Chris Ponting (CGAT, University of Oxford). Published datasets for CpG island (CGI) intervals were obtained from (Illingworth et al. 2010). Published RING1B ChIP-seq and input reads from mouse ESCs were obtained from GSM585229 (Tavares et al. 2012) and intervals were called as described below. EZH2 ChIP-seq from mouse ESCs was obtained from GSM480161 (Peng et al. 2009). TSS annotation and GO slim categories were obtained from Ensembl 66.

All sequences was mapped using Bowtie (Langmead et al. 2009) version 0.12.7 against the mouse genome (mm9). Up to two mismatches were allowed and only uniquely mapping reads were kept. Mapped reads were then de-duplicated using Picard to remove potential PCR-duplicates. Where appropriate, sets of mapped reads were normalized by random sampling of the larger set to the size of the smaller. Peak intervals were generated by MACS 1.4.2 (Zhang et al. 2008) using normalized chip and input samples. A p-value threshold of 1 × 10−5 and a fold enrichment over input threshold of 6 was used to filter the resulting peak intervals. Intervals closer than 200 bp were merged. Profile plots were generated using the custom script bam2geneprofile by counting reads mapping over each base of a pre-defined set of intervals and averaging the depth. In the case of Figure 5.5 (panel B), interval lengths were normalised. In all other cases a fixed window size about a point was used. Where two samples are shown on the same plot, the numbers of mapping reads has been normalised. Results are normalised over the number of intervals in the input set.

To calculate enrichment of KDM2B over KDM2A at CGIs overlapping TSSs, the coverage of each factor which overlapped a 1 kb window around any TSS was calculated using the coverageBed tool from BedTools (Quinlan and Hall 2010) on normalised read sets. For each CGI the ratio of KDM2B to KDM2A reads was calculated. The genes associated with CGIs that showed a greater than twofold enrichment of KDM2B over KDM2A were used to perform GO analysis. Enrichment of GO terms from the Biological Function category of the GO slim hierarchy was tested using the custom script GO.py. The significance of each category was tested using a hypergeometric test and FDRs calculated using the method of Benjamini and Hochberg (Benjamini and Hochberg, 1995). Terms with an FDR <0.05 and fold enrichment greater than 3.5 are shown.

To understand whether depletion of KDM2B affects RING1B occupancy, ChIP-Seq for KDM2B and RING1B in the control and KD cell line were performed in four biological replicates. However, the published work used only one replicate. ChIP-seq heatmaps were generated using the custom script bam2peakshape.py. Briefly, each CGI window was divided into 25 bp bins and the number of reads mapping to each bin counted. Reads were then normalised to the total number of reads mapping across all intervals. Intervals were then ranked by the number of reads in the RING1B set that map to that interval. Source code for custom scripts described is available in the mercurial repository at www.cgat.org/hg/cgat. The change-set used was 114addb46882. Sequencing and microarray data can be accessed via the geo accession GSE41267.

2.6.3.4. ChIP-Seq analysis for KDM2B$^{fl/fl}$ experiments

The bioinformatics analysis was conducted by Dr. Rob Klose and Mr. Hamish King (Klose lab). ChIP-Seq for KDM2B and RING1B was performed in biological triplicates, while the ChIP-Seq for SUZ12 was performed in biological duplicates. ChIP-seq sequencing reads for the antibody IPs or an input DNA sample were aligned to the mouse mm9 genome release using bowtie v1.1.2. The default alignment parameters were used in all cases with the exception that reads which could be aligned to more than

one site in the genome were supressed (-m 1). In samples comparing untreated and tamoxifen treated Kdm2b$^{fl/fl}$ cell lines the number of reads utilized for subsequent analysis in each replicate were down-sampled using Picard (v1.97) to the number of reads contained in the biological replicates that had the lowest total number of aligned reads. To identify a high confidence set of RING1B bound intervals, the independent RING1B ChIP-seq replicates from the untreated Kdm2b$^{fl/fl}$ sample were submitted independently to MACS (v1.4) with an input DNA sample and bound intervals were identified using the default parameters. The three independent RING1B interval sets were merged using Bedops (v2.2.0) and intervals not represented in all independent replicates were eliminated. When comparing ChIP-seq samples from the Kdm2b$^{fl/fl}$ and the 72hr tamoxifen treated cells, diffReps (v1.55.1) was applied using the default parameters and a negative binomial statistical test. diffReps utilizes the biological replicates to identify regions with significant changes in factor binding. Intervals identified by diffReps as changing significantly in RING1B or SUZ12 occupancy were intersected with the high-confidence RING1B and SUZ12 peaks respectively. This resulted in RING1B and SUZ12 interval sets which contain within them a region which significantly changed greater than 1.5 fold. Overlap analysis of RING1B and SUZ12 intervals that change was performed using the BedTools suite (Quinlan et al. 2010) and the VennDiagram package (version 1.0). Heat maps for BioCAP, KDM2B, RING1B and SUZ12 ChIP-seq data were generated for RING1B peaks (ranked by RING1B density) using HOMER and visualized with Java TreeView. Fold change heat maps were obtained by calculating the log2 fold difference between the normalised read counts within each RING1B peak interval in treated versus untreated KDM2b$^{fl/fl}$ ChIP-Seq data.

3. Chapter three - KDM2B is a nuclear protein which binds to CpG islands genome-wide

CpG islands (CGI) are short contiguous regions of the genome with a high frequency of non-methylated CpG dinucleotides. CGIs are usually associated with gene promoters and tend to remain non-methylated in somatic tissue regardless of the transcription status of the promoter they encompass. Although CGI methylation, either during development or aberrantly in cancer progression, results in gene silencing, it is still not clear how these elements are normally protected from *de novo* methylation or how they contribute to gene regulatory function. Interestingly, CGIs are characterized by an accessible chromatin conformation (Sabo et al. 2004; Blackledge and Klose 2011; Fenouil et al. 2012), which would make them an ideal binding platform for factors which specifically recognize the underlying DNA sequence or some aspects of its unique chromatin signature.

Chromatin immunoprecipitation (ChIP) studies have revealed that CGIs are generally depleted of the histone mark H3K36me2, which otherwise is found to coat intra- and intergenic regions (Kuo et al. 2011). This depletion correlates with nucleation of KDM2A, a Jumonji C (JmjC)-domain containing histone lysine demethylase which is selectively recruited to non-methylated CpG dinucleotides through its zinc finger-CxxC (ZF-CxxC) domain regardless of whether the gene associated with the CGI is active or repressed (Blackledge et al. 2010). Unexpectedly, knockdown of KDM2A followed by microarray-based gene expression analysis did not reveal dramatic changes in gene transcription profiles. This suggests that removal of H3K36me2 by KDM2A does not directly dictate transcriptional output but plays a more subtle role in mediating CGI function, most probably by delineating a unique architecture that differentiates CGI chromatin from the surrounding genome.

Importantly, there is a second KDM2 protein in mammals, KDM2B, which also contains a highly conserved catalytic JmjC domain and the ZF-CxxC DNA binding domain. Using *in vitro* formaldehyde-release histone demethylase assays, Tsukada et al. (2006) and He at al. (2008) showed that KDM2B is a catalytically active H3K36me2-demethylase. Furthermore, KDM2B was reported to localize to the *c-jun* CGI-associated promoter in a ZF-CxxC dependent manner (Koyama-Nasu et al. 2007), to potentially

tether a SIN3A repressor complex. However, through a putative nucleolar localization signal, KDM2B was also reported to be concentrated in the nucleolus and contribute to repression of ribosomal RNA genes by specifically demethylating H3K4me3 (Frescas et al. 2007).

A drawback of these previous studies investigating the chromatin distribution of KDM2B was that they relied on overexpressed, tagged versions of the protein to investigate its cellular localization or binding to only a few selected loci. This leaves open the question of how endogenous KDM2B functions at the genome scale and how it compares to KDM2A.

Therefore, to examine the cellular distribution and chromatin binding profile of KDM2B, an antibody recognizing endogenous KDM2B was raised and characterized. This essential tool allowed us to then define the subcellular localization of KDM2B, to understand whether KDM2B also associates with CGIs *in vivo* and to investigate potential chromatin binding profile differences between the two KDM2 proteins. A clearer understanding of how KDM2B exerts its function will provide the framework to better assess how the properties of ZF-CxxC containing proteins contribute to the unique chromatin architecture of CGIs and how this is coupled to transcriptional regulation.

3.1. Generation of an antibody specifically recognizing KDM2B

To examine the properties of endogenous KDM2B and to understand how it functions in comparison to KDM2A, a rabbit polyclonal antibody against KDM2B was raised. A 165 amino acid antigen was designed between the PHD and F-box domains of the human protein, in a region that was not conserved between the two otherwise highly similar KDM2 proteins (Figure 3.1, panel A). The antigen was expressed as a His-tag fusion protein in *E.coli* and affinity purified on Ni-NTA resin (Figure 3.1, panel B).

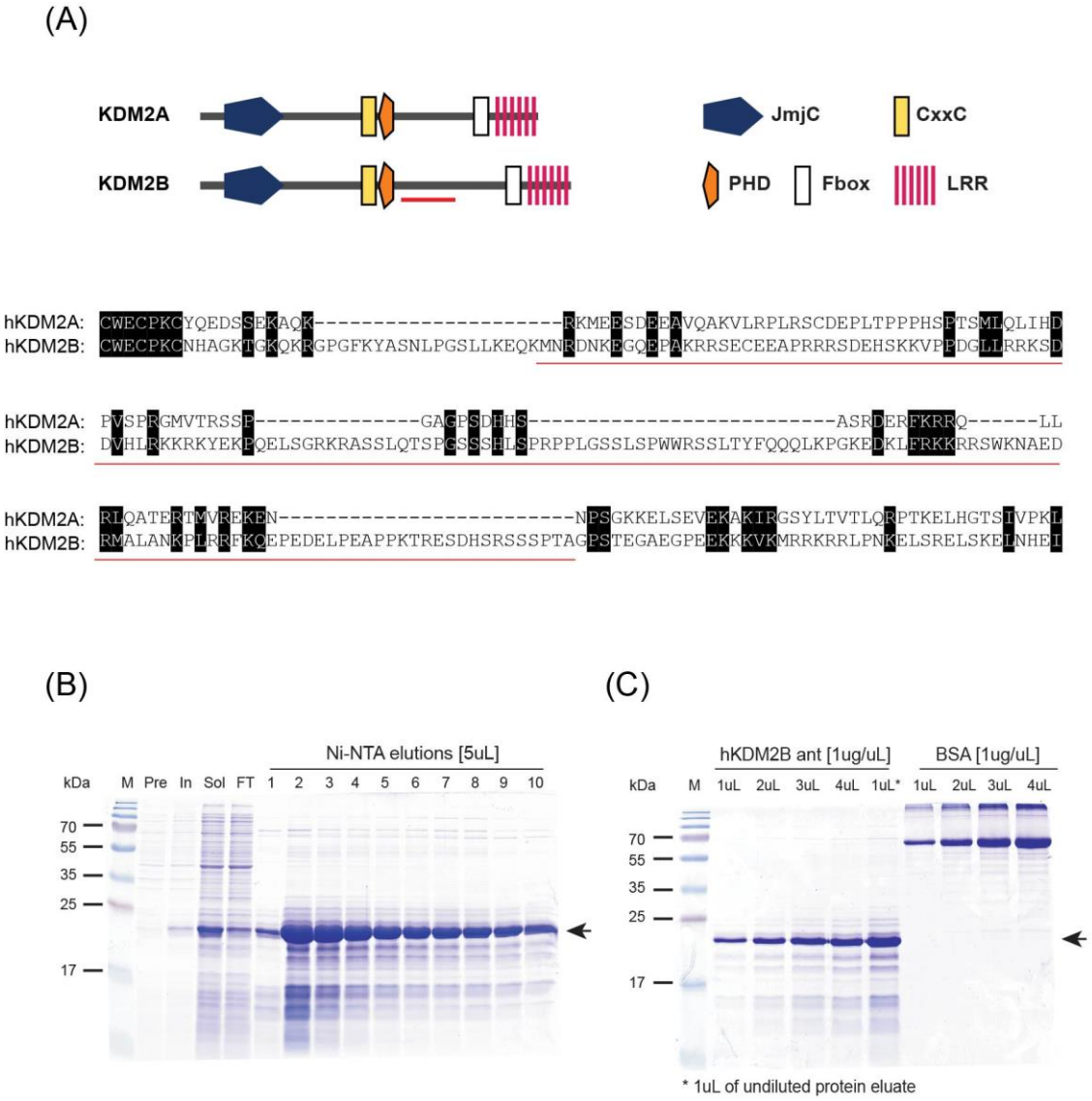


Figure 3.1. KDM2B antigen design and purification.

(A) Schematic representation of KDM2A and KDM2B illustrating the conserved domains. The red line indicates the antigen region chosen for generation of the KDM2B specific antibody. Partial protein sequence alignment of human KDM2A and KDM2B illustrating that the region selected for antigen production is not significantly conserved between the two proteins (antigen sequence underlined in red below the alignments). Full length sequences were obtained from the Uniprot database, with the accession numbers Q9Y2K7 (KDM2A) and Q8NHM5 (KDM2B).

(B) Coomassie stained SDS-PAGE following bacterial antigen expression and purification using Ni-NTA resin column. Pre: sample collected before IPTG-based protein induction; I: IPTG-induced sample; S: post-sonication soluble sample; FT: column flowthrough; 1-10: elutions collected from Ni-NTA column. The arrow indicates the band of the expected molecular weight for the 6xHis-tagged protein of interest.

(C) Determination of KDM2B antigen concentration after Ni-NTA affinity purification and subsequent dialysis compared to BSA samples of known concentration [1µg/µL].

The purified KDM2B antigen was subsequently used to immunize a rabbit and consecutive bleeds were tested for reactivity with endogenous and overexpressed KDM2B. With the exception of the pre-immune serum, all bleeds tested showed reactivity towards overexpressed KDM2B in Western blot analysis (Figure 3.2). It was unclear in nuclear extract from HeLa cells whether endogenous KDM2B was recognized due to multiple cross-reactive bands (Figure 3.2, panel A - NE lane). This is likely due to antibodies already present in the immunized rabbit that cross-react with human proteins as these are also present in the pre-immune serum. One of the purposes of this antibody is to immunprecipitate KDM2B. Therefore, bleed#3 was tested to determine whether it could carry out this function. As shown in Figure 3.2 (panel B), the crude antibody showed some capacity to immunoprecipitate overexpressed KDM2B.

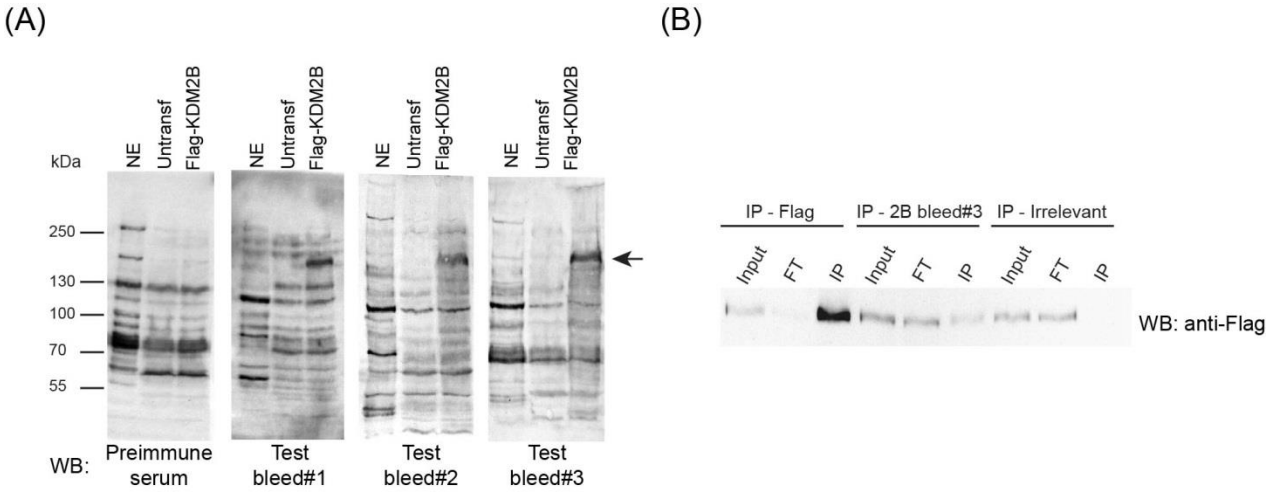


Figure 3.2. KDM2B antibody recognizes endogenous and overexpressed protein.

(A) Western blot analysis using pre-immune serum and 3 consecutive bleeds collected after immunization with KDM2B specific antigen. HeLa nuclear extract (NE) (30µg), untransfected HeLa whole cell extract (Untransf) and HeLa over-expressing Flag-tagged hKDM2B whole cell extract (Flag-KDM2B) were used to test the serum. Full length KDM2B migrates by SDS-PAGE at an expected size of 152 kDa (indicated by the arrow).

(B) KDM2B immunoprecipitated from HeLa cells expressing Flag-tagged KDM2B whole cell extract. Anti-Flag antibody or test bleed#3 serum (shown as representative) were used for immunoprecipitation. HA-specific antibody was used as a negative control (IP-Irrelevant). FT: flowthrough.

3.2. Validation of the KDM2B-specific antibody

To try and improve the specificity of the KDM2B antibody, the serum from bleed#3 was immunoaffinity purified using the immunization antigen covalently coupled to an AffiGel agarose solid support. The antibody was eluted with low pH and immediately neutralized to minimize irreversible unfolding of the antibody. Coomassie staining of the elutions from the immunoaffinity purification yielded characteristic IgG heavy and light chains (Figure 3.3, panels A and B). An additional contaminating band running at approximately 70kDa (indicated by arrow) is clearly visible, most probably corresponding to a highly-abundant serum protein (i.e. rabbit serum albumin). Samples corresponding to each step of the affinity purification were used to probe Western blots of nuclear extract and overexpressed KDM2B (Figure 3.3, panel C), with the affinity purified antibody showing the capacity to specifically recognize overexpressed KDM2B. It is unclear if endogenous KDM2B in HeLa cell nuclear extract is recognized due to presence of additional bands running at around the same size as KDM2B.

Importantly, a reduction in non-specific signal and improved efficiency of immunoprecipitation was observed following affinity purification as indicated by the detection of a distinct clear band of expected size in immunoprecipitation from mouse ES cell nuclear extract and depletion of this band in flow through (Figure 3.3, panel D). Detection of endogenous KDM2B in mouse ES cells nuclear extract appeared more efficient than that observed when using nuclear extract from HeLa cells (compare Figure 3.3, panels C and D). This could be due to KDM2B being more highly expressed in mouse ES cells, or the nuclear extraction procedure being more efficient at releasing KDM2B in ESCs. Additional immunoreactive protein species were detected by the affinity purified antibody, which may represent degradation products of KDM2B or additional isoforms due to the use of alternative promoters or splice sites. For example, a shorter form of KDM2B lacking the N-terminal region encoding the JmjC catalytic domain is known to be ubiquitously expressed during embryonic development and in various adult tissues (Fukuda et al. 2011). Although its molecular function remains unclear at this point, the existence of this shorter form and the fact that the KDM2B antibody recognizes both isoforms will be referred to later in this thesis (Figure 3.3, panel D, indicated by asterisk).

Another explanation for the detection of additional bands in Western blot analysis that are not detected in the immunoprecipitated (IP) sample could be that epitopes not accessible under native conditions are exposed during denaturation of proteins in polyacrylamide gel electrophoresis. Although the presence of non-specific immunoreactive bands was not completely eliminated by the affinity purification of the KDM2B antibody, the band corresponding to full length KDM2B was nevertheless a major species in the immunoprecipitation. Subsequent work using shRNAs targeted against KDM2B shows that the bands of interest (KDM2B full length and short form) are specifically depleted suggesting a reasonable specificity of the antibody (refer to Chapter 5). Antigen competition-type experiments were not conducted, but would have been useful as a further validation of the specificity of the KDM2B antibody.

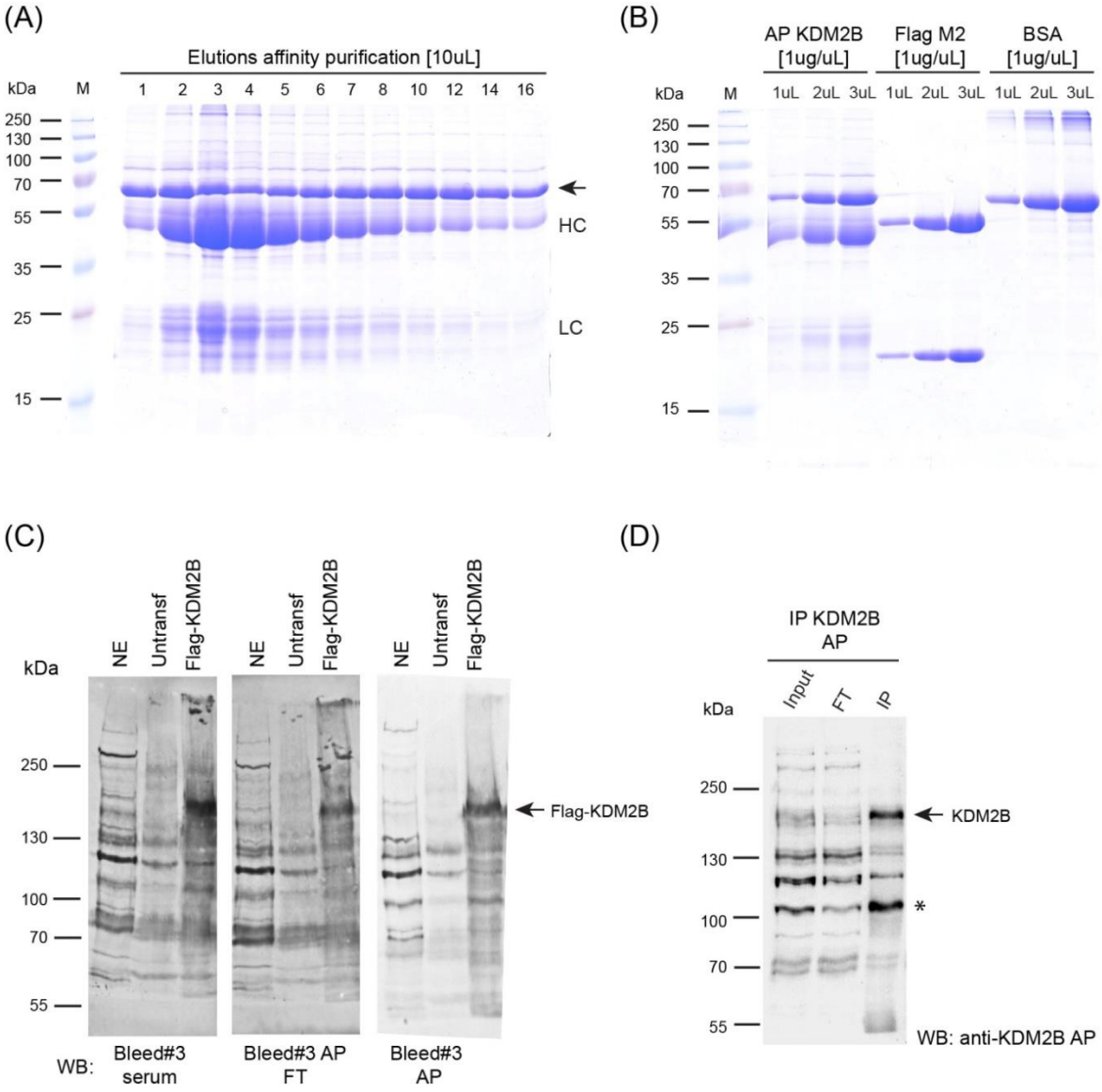


Figure 3.3. Affinity purification improves recognition and immunoprecipitation of KDM2B-specific antibody.

(A) Coomassie stained SDS-PAGE of fractions eluted from Affigel10 column-based immunoaffinity chromatography of KDM2B antibody from serum (elutions 1 to 16). The IgG heavy chain (HC) and light chain (LC) are indicated, while rabbit serum albumin is indicated by an arrow.

(B) Affinity purified KDM2B antibody concentration estimation at approximately 1mg/mL for the band of interest compared to BSA and Flag-M2 standards (as indicated).

(C) Western blot analysis using HeLa nuclear extract (NE – 30 µg), whole cell extract prepared from untransfected (Untransf) or Flag-tagged KDM2B expressing cells (Flag-KDM2B). The membranes were probed with bleed#3 serum, flowthrough sample from the affinity purification (FT) or with the final purified antibody (AP). A reduction in non-specific signal while retaining an efficient response to its target protein is observed.

(D) KDM2B immunoprecipitation from mouse ES cell nuclear extract followed by Western blot analysis. The immunoprecipitation and the WB were done using the affinity purified KDM2B antibody. The band at the expected molecular weight corresponding to full length KDM2B is indicated by an arrow, while the KDM2B short form is indicated by an asterisk.

Despite the fact that a divergent region of KDM2A and KDM2B was chosen to generate the KDM2B antibody, it was important to ensure that the KDM2B antibody does not cross react with KDM2A. To achieve this, tagged versions of KDM2A and KDM2B were transiently overexpressed and the response of the KDM2B antibody was tested by Western blotting. Previously, a KDM2A antibody was generated in the Klose lab. The specificity of this antibody was also tested. In both cases, the KDM2A and KDM2B antibodies specifically recognize their respective targets (Figure 3.4).

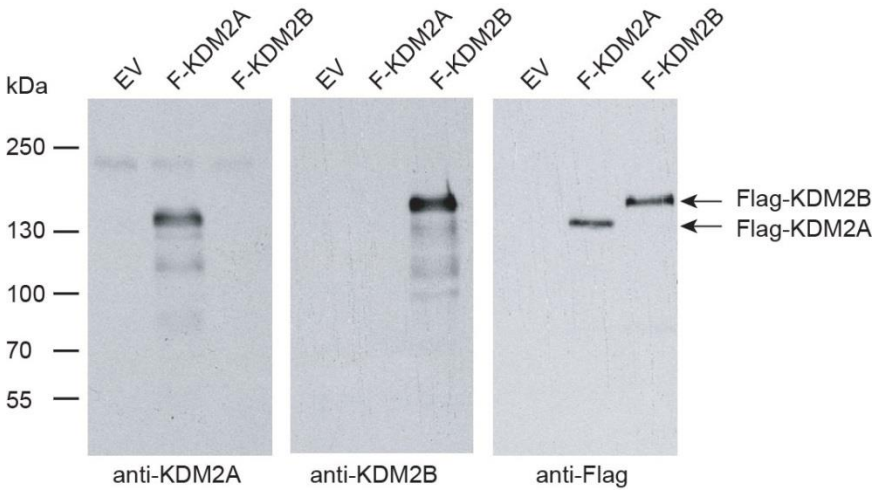


Figure 3.4. The KDM2A and KDM2B antibodies specifically recognize their respective antigens.
Empty expression vector (EV), Flag-KDM2A (F-KDM2A) and Flag-KDM2B (F-KDM2B) were transiently transfected into 293T cells. Whole cell extracts were prepared and probed with anti-KDM2A (left panel), anti-KDM2B (centre panel), or anti-Flag (right panel). In both cases, the KDM2A and KDM2B antibodies recognize their respective antigens and do not show significant cross-reactivity.

3.3. KDM2B is a nuclear protein with a broad nucleoplasmic staining pattern

Previous work had suggested that KDM2B localizes to nucleoli where it binds to the transcribed regions of ribosomal DNA and represses ribosomal RNA gene transcription by removing H3K4me3, potentially in a mitogen-regulated manner (Frescas et al. 2007). However, other reports have suggested non-nucleolar roles for KDM2B (Gearhart et al. 2006; Sanchez et al. 2007; He et al. 2008; Pfau et al. 2008). To address

these apparently contradictory observations, the subcellular localization of endogenous KDM2B was investigated using the newly generated affinity purified KDM2B-specific antibody. To achieve this, various mammalian cell lines were tested, including human transformed cell lines (HeLa, U2OS and 293T), non-immortalized and non-transformed fibroblasts (IMR90), as well as different mouse embryonic stem cell lines (J1 ES cells shown as representative). These immunofluorescence studies showed that KDM2B localizes to the nucleus, where it is broadly distributed throughout the nucleoplasm (Figure 3.5, panel A). In DAPI-stained nuclei, nucleoli appear as unstained, hollow, generally spherical structures, which are surrounded by DAPI positive chromatin (Boisvert et al. 2007). Interestingly, KDM2B seems to be largely excluded from these regions. To confirm that KDM2B is not concentrated in the nucleolus, immunofluorescence analysis was performed in HeLa cells overexpressing exogenous Flag-tagged KDM2B. This was carried out as these were the type of experiments in which KDM2B was previously reported to be nucleolar. Immunostaining of fibrillarin was included as a nucleolar marker (Figure 3.5, panel B). Fibrillarin is a small nuclear ribonucleoprotein (RNP) component which plays an important role in processing of the pre-rRNA and is found in the dense fibrillar component of the nucleolar structure (Bartova et al. 2010). The lack of co-localization between KDM2B and fibrillarin-rich regions was striking, indicating that, in contrast to previous reports, KDM2B is more broadly distributed throughout the nucleus. Importantly, the absence of a nucleolar staining pattern shows that KDM2B occupies a similar region of the cell as its paralogue KDM2A, thereby reinforcing the need to understand the functional properties of these two closely related enzymes.

3.4. KDM2B foci are distinct from Polycomb bodies, but appear to co-localize with 53BP1 foci

While performing the immunofluorescence studies, it was apparent that some cells exhibited punctate endogenous KDM2B foci (Figure 3.5, panel C). The percentage of cells having this distinct staining pattern differed between cell type, ranging from approximately 5% for ES cells to 10-15% in transformed lines like HeLa or U2OS. The number of foci was generally between 1 and 5 per cell, although statistically

relevant analysis was not performed. When HeLa cells were synchronized at the beginning of S phase by a double thymidine block and then released into the cell cycle by washing out the thymidine, the KDM2B foci were evident at 0, 4 and 8 hours post-release (corresponding to G1-S phase, S phase and G2/M phase of the cell cycle, respectively), but were not detectable at 12 hours post-release (late G2/M) when it was apparent that KDM2B does not remain associated to DNA during replication (data not shown). Although the significance of these foci is unclear, they do not correspond to nucleoli, as nucleoli appear as dark unstained regions in DAPI images.

Considering previously reported connections between KDM2B and polycomb E3 ubiquitin-ligase member RING1B (Gearhart et al. 2006; Sanchez et al. 2007), co-localization studies were conducted to understand whether the KDM2B foci overlap with Polycomb (PcG) bodies.

By using antibodies specifically recognizing three human PcG proteins (RING1, BMI1 and hPC2), Saurin et al. (1998) were able to show that the PcG complex forms a novel and unique nuclear structure that was termed “Polycomb bodies” (PcG bodies). PcG bodies are accumulation of Polycomb proteins in clustered regions of the nucleus, potentially tightly associated with large pericentric heterochromatin regions. Although the function of PcG bodies in mammalian cells remains to be determined, in *Drosophila* they seem to be involved in Polycomb protein-mediated gene pairing and silencing (Lanzuolo et al. 2007; Bantignies et al. 2011; Mao et al. 2011; Pirrotta and Li 2012). The numbers and sizes of these dynamic structures vary between cell types, with transformed lines like U2OS showing anywhere between 6 and 14 (per nucleus) large and distinct PcG bodies, mostly associated with the periphery of heterochromatin or centromeric satellites (Hernandez-Munoz et al. 2005; Smigova et al. 2011). However, when comparing the staining patterns of endogenous KDM2B and RING1B, it was obvious that KDM2B foci do not co-localize with PcG bodies, although there is an otherwise clear overlap between the two proteins at the pan-nuclear level. This lack of co-localization implies that KDM2B is most probably not a major component of PcG bodies and is not involved in the functional implications of this structure.

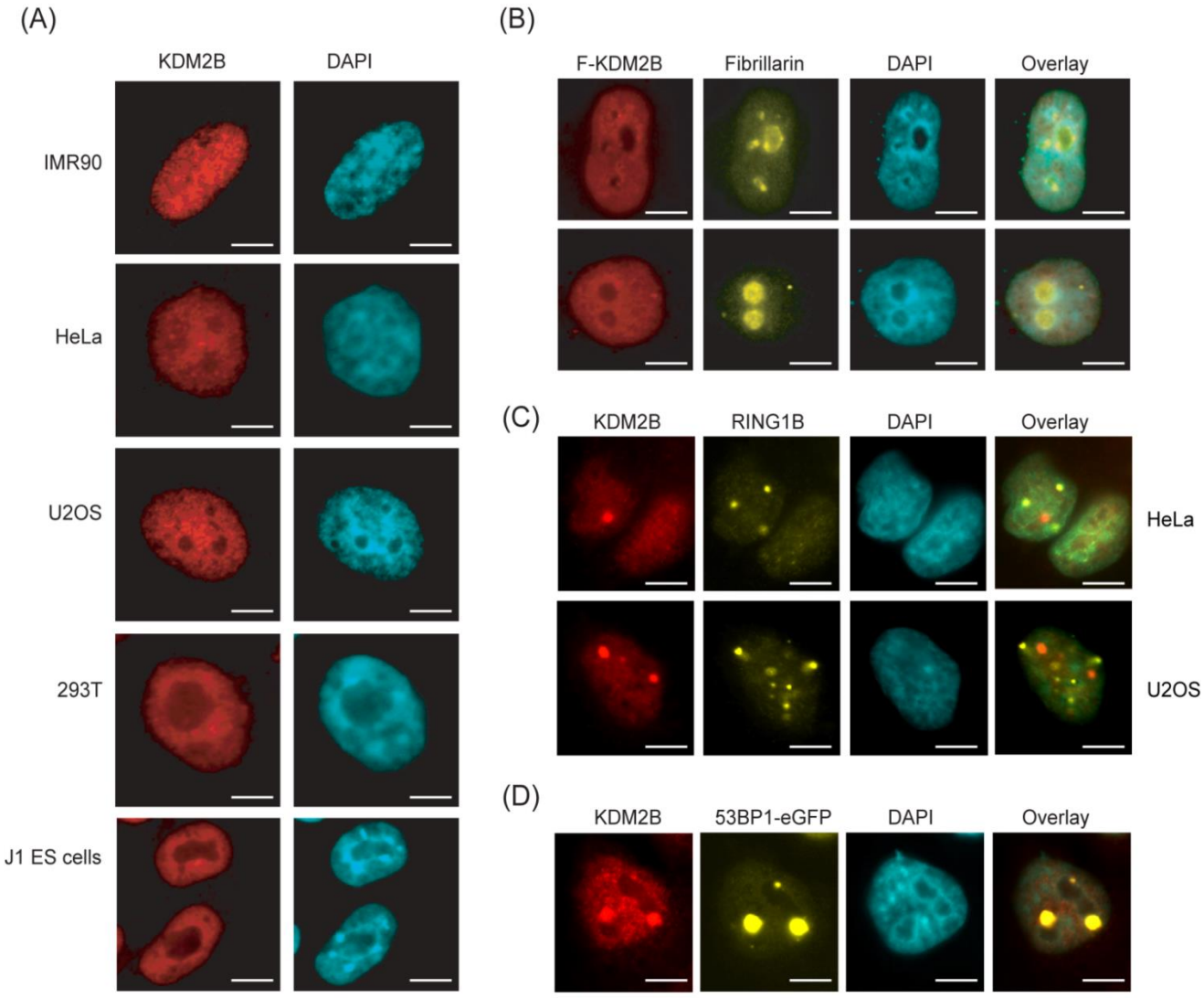


Figure 3.5. KDM2B is a nuclear protein, with a broad nucleoplasmic staining pattern.

(A) Endogenous KDM2B immunofluorescence in five different mammalian cell lines. KDM2B signal is in red (left) and DAPI signal in blue (right). In all cell types KDM2B is broadly distributed throughout the nucleus and not concentrated in the nucleolus as previously reported (Frescas et al., 2007).

(B) Immunofluorescence in HeLa cells for Flag-tagged KDM2B and endogenous fibrillarin, a nucleolar marker. KDM2B is largely excluded from nucleoli.

(C) In approximately 10-15% of cells, endogenous KDM2B forms distinct foci in the nucleus, whose identity and function remains to be clarified. Endogenous KDM2B and RING1B staining in HeLa and U2OS cells suggests that the KDM2B foci are distinct from nuclear PcG bodies.

(D) Overexpression of the DNA damage marker 53BP1 fused to eGFP and endogenous KDM2B staining in undamaged HeLa cells, showing co-localization between the two proteins. This observation is interesting, considering that KDM2B was detected in screens for proteins that become rapidly phosphorylated in response to DNA damage (Matsouka et al., 2007), but remains to be further investigated. Scale bar represents 5µm.

In an attempt to gain a better understanding of the nature of the KDM2B foci, a surprising observation was made when checking the staining pattern of the DNA damage marker 53BP1 in undamaged HeLa cells (Figure 3.5, panel D). Briefly, 53BP1 (also called TP53BP1) is a chromatin-associated factor that promotes DNA double-strand breaks (DSB) repair by the non-homologous end-joining (NHEJ) pathway (Schultz et al. 2000; Callen et al. 2013; Fradet-Turcotte et al. 2013). 53BP1 is recruited to chromatin by recognizing mononucleosomes marked by H4K20me2 and H2AK15Ub1, acting then as a recruitment platform for Rap1-interacting protein 1 (RIF1) to regulate DSB repair (Zimmermann et al. 2013). Interestingly, 53BP1 was reported to form distinct nuclear bodies in normally proliferating mammalian cells which correspond to common fragile-sites unresolved after replication stress (Lukas et al. 2011). As KDM2B was reported to be rapidly phosphorylated by the ATM/ATR DNA damage protein kinases in response to ionizing radiation (IR) DSBs (Matsuoka et al. 2007), the exciting possibility arose that KDM2B foci might be related to those formed by 53BP1. For this purpose, a construct expressing eGFP-tagged 53BP1 (kindly provided by Prof. Jiri Lukas) was transiently transfected into undamaged HeLa cells and the endogenous KDM2B staining pattern investigated. The co-localization between the two proteins was striking, potentially suggesting that KDM2B is recruited to common fragile sites in the genome to aid in the repair pathway.

Although the relevance of this staining pattern is unclear, the close and highly complex connections between histone modifications and DNA damage are beginning to be defined (Moore and Krebs 2004; Dinant et al. 2008; van Attikum and Gasser 2009). One important observation is that subjecting the cells to IR DNA damage leads to the complete dissolution of KDM2B foci, while 53BP1 relocates to multiple nuclear foci within minutes after exposure (data not shown), suggesting that if KDM2B is needed for repair of DNA breaks, its role is potentially different than that of factors which immediately accumulate at the site of DNA damage.

Based on these results, it is clear that KDM2B is a nuclear protein and not localized to nucleoli as previously suggested. Considering that KDM2B also encodes a ZF-CxxC domain, it remained possible that KDM2B could have a similar role to KDM2A in specifically binding non-methylated DNA and CGIs.

3.5. KDM2B binds non-methylated CGI DNA

Comparison of the ZF-CxxC domains of KDM2A and KDM2B shows that the architecture of the DNA binding domain is highly conserved between the two proteins (Figure 3.6), with a complete conservation of the eight cysteine residues involved in coordination of Zn^{2+} ions and preservation of the KFGG and KQ motifs within the linker region important for mediating specific DNA contacts (Allen et al. 2006; Cierpicki et al. 2010; Xu et al. 2011a).

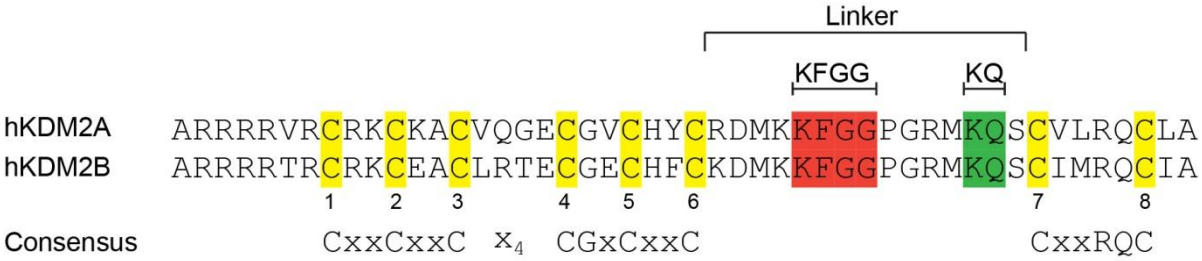


Figure 3.6. Primary sequence conservation of the ZF-CxxC DNA binding domain between KDM2A and KDM2B.

Shown here are the corresponding human protein sequences. The eight conserved cysteine residues involved in zinc ion coordination are highlighted in yellow, while the KFGG motif shown to be important for the DNA binding activity of ZF-CxxC domain containing proteins is shown in red. The conserved KQ motif (in green) makes specific contacts with the CpG-dinucleotide sequence in the DNA substrate. The highly conserved architecture of the ZF-CxxC DNA binding domain between the two mammalian KDM2 proteins suggests that, like KDM2A, KDM2B is able to specifically recognize non-methylated CpG dinucleotides.

Considering the highly conserved sequence of the DNA binding domain between the closely related KDM2 proteins, it was expected that the KDM2B ZF-CxxC domain could also function as a non-methylated DNA binder. Indeed, electrophoretic mobility shift assays (EMSA) have shown that recombinant KDM2B ZF-CxxC domain binds to a DNA probe containing non-methylated CpGs in a concentration-dependent manner, similar to KDM2A (see (Farcas et al. 2012)). This suggested that KDM2B could potentially also bind CGIs *in vivo*.

To address this possibility, chromatin immunoprecipitation (ChIP) was carried out using the affinity purified KDM2B-specific antibody. ChIP followed by quantitative PCR (ChIP-qPCR) showed that, in mouse embryonic stem cells (ESCs), KDM2B was bound to CGI-containing *Suv420h1*, *Ncoa2*, *Bdnf* and *Ink4a* promoters, but not to their gene bodies (Figure 3.7, panel A). Similar ChIP-qPCR analysis was carried out for several genes with non-CGI promoters (*Fgf7* and *Aim2* illustrated as representative) (Figure 3.7, panel B), where no KDM2B enrichment was detected. This result therefore suggests that KDM2B is preferentially targeted to non-methylated CGIs.

3.6. KDM2B binds CGIs genome-wide, with a chromatin binding profile resembling that of KDM2A

To determine whether the preferential association of KDM2B with CGIs is observed at the genome-scale, ChIP coupled to high-throughput parallel sequencing (ChIP-Seq) was carried out. The ChIP DNA was submitted to the Oxford Genomics Centre at the Wellcome Trust Centre for Human Genetics (WTCHG) for 51 base pair sequencing on the Illumina HiSeq2000. Two biological replicates were analysed for the KDM2B ChIP-Seq results. In addition, KDM2A ChIP-Seq reactions were set up in parallel, to directly compare the chromatin binding profiles of the two KDM2 proteins. The CGI track used throughout this study was obtained from UCSC mm9 table browser, while Bio-CAP-Seq profiles for mouse ESCs generated in the Klose lab are shown to indicate the location of non-methylated DNA (Blackledge et al. 2012; Long et al. 2013b).

A striking overlap was observed between KDM2B binding sites and algorithm-predicted CGIs. This overlap was evident when zooming in on regions of the genome containing both CGI and non-CGI-associated genes (Figure 3.7, panel C, representative region on chr3), with KDM2B enrichment specifically observed only at the CGI-containing promoters. To verify the specificity of these profiles, an input chromatin sample was analysed in a similar manner and showed no significant mappable reads at either promoter type. The results of the ChIP-Seq analysis show that KDM2B occupies over 90% of the approximately 23,000 CGIs (Illingworth et al. 2010).

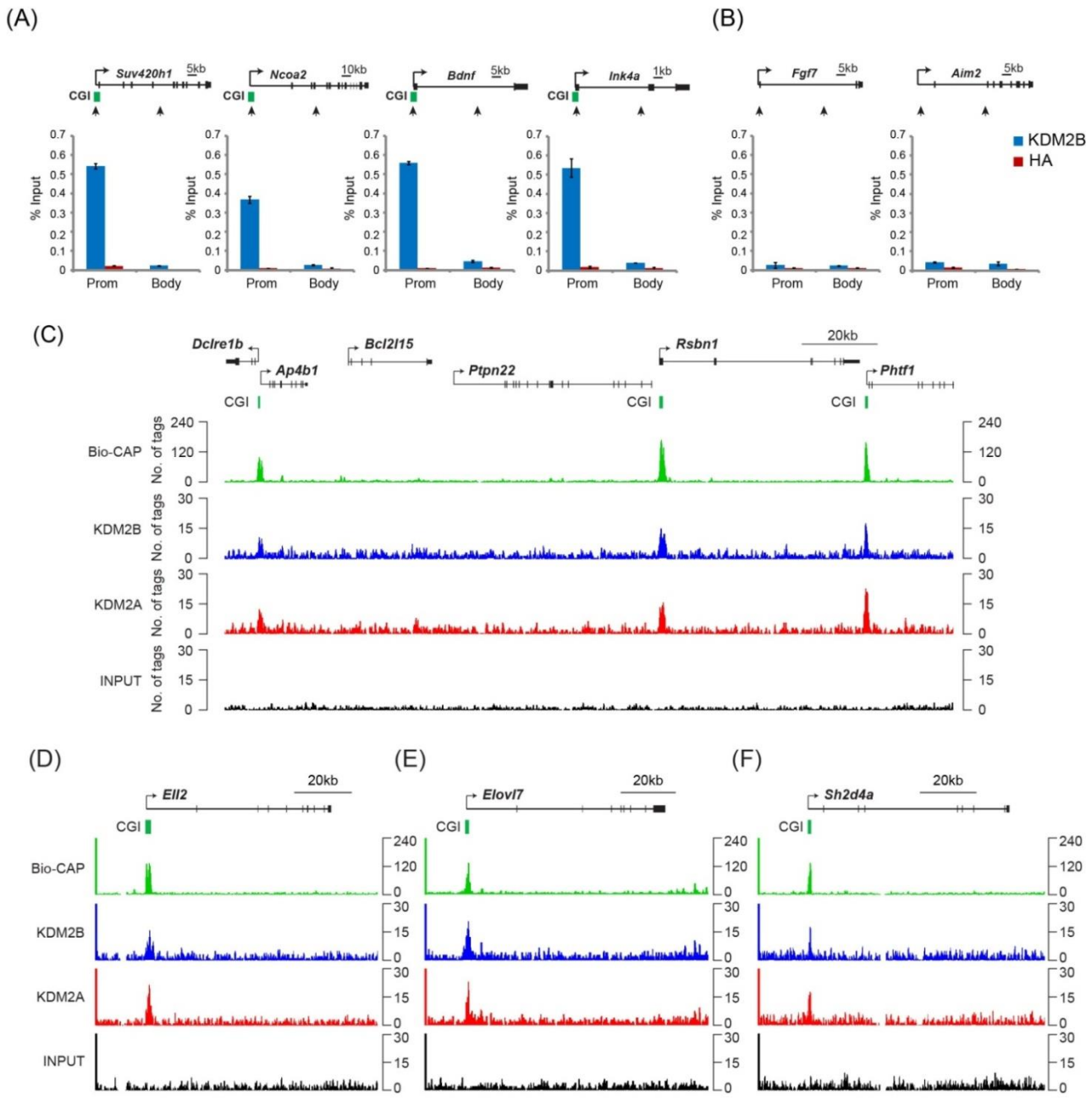


Figure 3.7. KDM2B binds to non-methylated CGIs genome-wide.

(A) ChIP analysis in mouse ES cells followed by quantitative PCR (ChIP-qPCR) for KDM2B (blue bar). An HA-specific antibody was used as a negative control (red bar). For each gene, a promoter (Prom) and body (Body) primer set was designed, as indicated. KDM2B shows enrichment at CGI- promoters compared to gene body regions and non-CGI- promoters (panel B). A schematic of each gene is shown above the ChIP-qPCR data, with exons indicated by verticals bars and CGIs (when present) by green horizontal bars. Error bars represent SEM of three biological replicates.

(C) Input and KDM2B ChIP-seq profiles over a region of the genome containing CGI and non-CGI associated genes. Bio-CAP-seq profiles are shown to indicate the location of non-methylated DNA (Blackledge et al., 2012). The ChIP-seq profile of the closely related CGI-binding protein KDM2A is shown as comparison (Blackledge et al., 2010). Above the sequencing traces individual genes are shown with the arrow indicating the transcription start site and

vertical black lines corresponding to exons. The locations of CGIs are indicated by green bars. KDM2B specifically associates with CGIs containing non-methylated DNA, similar to KDM2A.
(D) - (F) ChIP-seq profiles for KDM2A and KDM2B at a panel of CGI-containing genes.

Interestingly, the genome-wide chromatin binding profile for KDM2B appeared to be largely similar to that of KDM2A, with both proteins binding CGI elements and being excluded from non-CGI genes. The similar chromatin distribution of the two KDM2 proteins was apparent when focusing on a panel of CGI-containing genes (Figure 3.7, panels D-F), with KDM2A and KDM2B showing comparable enrichment levels. As both proteins are bound to CGIs genome-wide, this raised the question of what is the physiological relevance of having two closely homologous proteins with nearly identical chromatin distribution profile. They could function redundantly to remove H3K36me2 or they could be recruited to the same CGI but mediate non-redundant functions.

3.7. KDM2B, unlike KDM2A, is preferentially enriched at Polycomb-occupied CGIs

KDM2A and KDM2B show a largely similar chromatin distribution profile, with both KDM2s specifically binding to CGI elements genome-wide. When the tag density for the two proteins was analysed over all transcription start site (TSS)-associated CGIs and compared to non-CGI TSSs, similar enrichment for KDM2A and KDM2B was observed at the CGI-associated promoters and no signal was observed at non-CGI promoters (Figure 3.8, panel A). To understand whether there are any particular features of a CGI when it is promoter-associated that make it more susceptible to binding of the two enzymes, tag density for KDM2A and KDM2B was analysed over all CGIs, regardless of whether they were intergenic, found at the transcription end site or distinct from either class. The specificity of the enrichment held true, with both proteins showing binding to CGIs regardless of their promoter-association (Figure 3.8, panel B), in agreement with the capacity of KDM2A and KDM2B to recognize non-methylated CpG dinucleotides.

Despite the highly similar chromatin binding profile of the two KDM2 proteins, closer analysis of the ChIP-Seq data prompted an interesting observation. When the levels of KDM2A and KDM2B were

directly compared at individual CGIs genome-wide, it was apparent that not all CGI elements had similar enrichment levels for the two KDM2s. To investigate this observation further, a collaboration was initiated with bioinformaticians Drs. Ian Sudberry and David Sims, from the group of Prof. Chris Ponting. MA-plots allowed the genome-wide pairwise comparison of KDM2A and KDM2B at all gene-associated CGIs (Figure 3.8, panel C). The MA-plot indicates that, while the majority of CGIs show similar binding of the two KDM2s, a subset of CGIs shows preferential enrichment for KDM2B (Figure 3.8, panel C, coloured blue).

In order to understand whether there are any particular features shared among the CGIs with preferential enrichment of KDM2B, the genes associated with this CGI subset were extracted and subjected to gene ontology (GO) enrichment analysis (Figure 3.8, panel D). GO terms for these genes were investigated and the overrepresented classes identified using the Biological Function category of the GO slim hierarchy. Genes showing preferential enrichment of KDM2B were enriched for functions relating to anatomical structure development, cell morphogenesis, cell differentiation, developmental maturation and embryo development (Figure 3.8, panel D). An important feature shared by these developmental regulatory genes is that in embryonic stem cells they are often occupied by Polycomb group (PcG) protein complexes (Bracken et al. 2006; Leeb and Wutz 2007; Mikkelsen et al. 2007; Endoh et al. 2008; Simon and Kingston 2009).

PcG proteins reside in multimeric complexes, with the best studied ones being Polycomb Repressive Complex 1 (PRC1) and 2 (PRC2), and exert their function by modifying the chromatin structure at their target genes. PRC1 catalyses the mono-ubiquitylation of H2AK119 via its E3 ubiquitin ligase member RING1B, while PRC2 is responsible for the trimethylation of H3K27 by the histone lysine methyltransferase component EZH2 (Czermin et al. 2002; Kuzmichev et al. 2002; Muller et al. 2002; de Napoles et al. 2004).

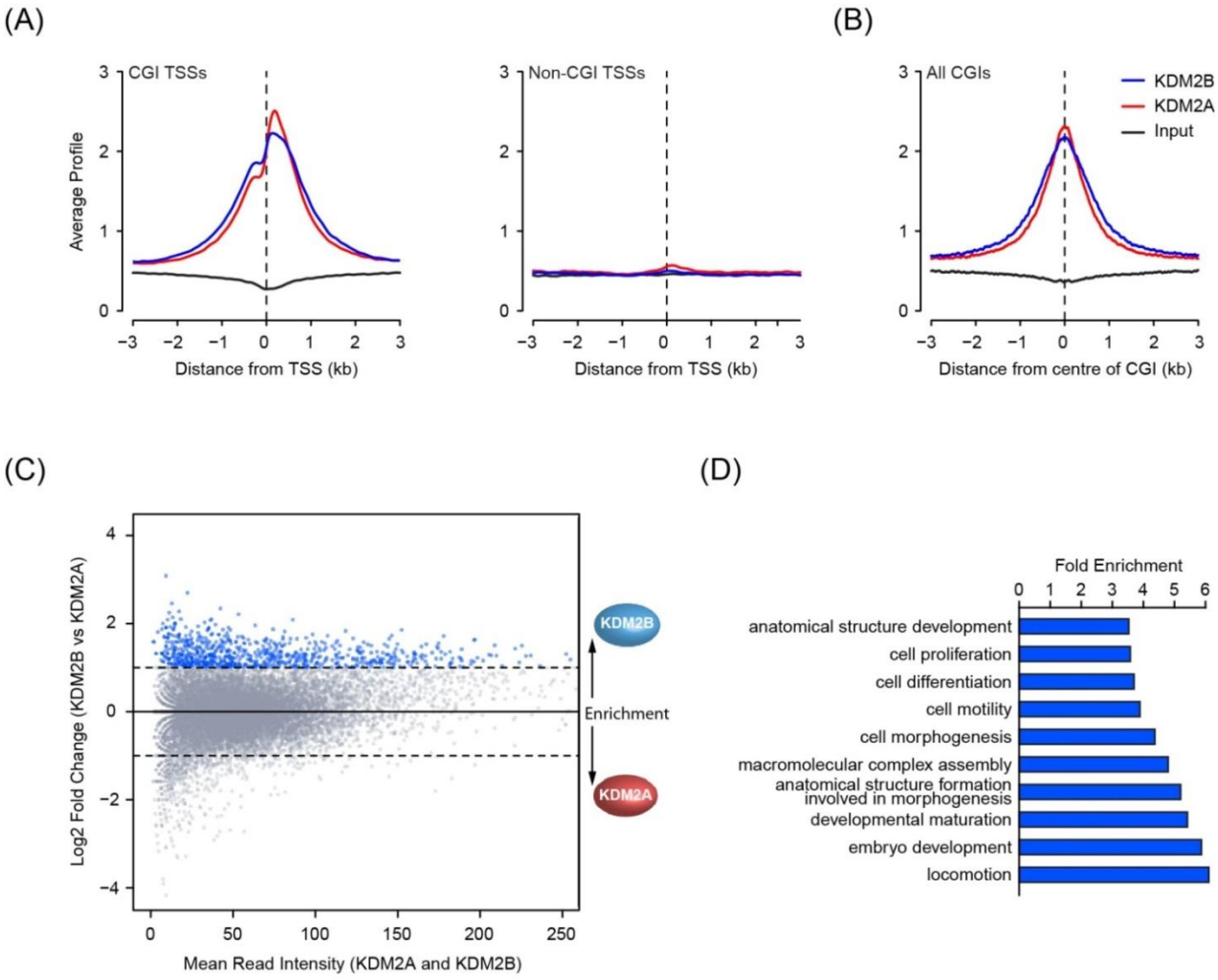


Figure 3.8. KDM2B is preferentially enriched at developmental regulatory genes.

(A) KDM2A and KDM2B ChIP-seq signal segregates specifically with CGI-associated gene promoters (left panel) and is excluded from gene promoters not associated with CGIs (right panel).

(B) KDM2A and KDM2B ChIP-seq signal is centred over CGIs, in agreement with their capacity to recognize non-methylated DNA at these sites.

(C) An MA-like plot depicting the relative enrichment of KDM2A and KDM2B at all gene associated CGIs in mouse ESCs. The log2 mean read intensity is displayed on x-axis and the log2 relative enrichment of KDM2B compared to KDM2A is displayed on the y-axis. KDM2B enriched CGIs are displayed above the MA-plot centred line (y=0 axis), while KDM2A enriched CGIs are shown below. The subset of CGIs highly enriched for KDM2B is coloured blue and represents approximately 5% of total KDM2B CGI peaks. **Note that there is no statistical significance attached to this.**

(D) A histogram displaying fold enrichment values for GO term analysis of the genes which are over twofold enriched for KDM2B (blue data points, part (C), approximately 580 genes at this defined threshold) at a FDR <5%. The bioinformatic analysis was performed by Dr. Ian Sudbery and Dr. David Sims from the group of Prof. Chris Ponting.

While in flies it is known that PcG proteins repress their target genes by binding to specific DNA regions called Polycomb Response Elements (PREs) via recruitment by dedicated transcription factors (Horard et al. 2000; Hodgson et al. 2001; Huang et al. 2002; Schwendemann and Lehmann 2002; Dejardin et al. 2005; Muller and Kassis 2006), it is not clear how PcG complexes bind their target genes in mammals. Interestingly, although no classical PREs were identified in mammals, ChIP-Seq analysis showed that PcG targets overlap almost exclusively with CGI elements (Ku et al. 2008). Furthermore, additional studies suggested that non-methylated CpG-rich sequences have the capacity to establish PRC2 recruitment in mammalian cells (Mendenhall et al. 2010; Lynch et al. 2012). The observation that KDM2B is preferentially enriched at PcG-occupied CGIs prompted the possibility that KDM2B could play a role in regulating polycomb-mediated repression of these developmental genes.

To address this possibility, the ChIP-Seq binding intensities of RING1B (Tavares et al. 2012) and EZH2 (Peng et al. 2009) were compared to those of KDM2A and 2B generated in this study at all CGI-associated TSSs (Figure 3.9). Visualizing the data as heat maps with signal intensities plotted at 5kb on either side of the TSS, it was apparent that KDM2A and KDM2B show similar binding profiles at the majority of CGIs. However, the subset of CGIs occupied by RING1B and EZH2 representing Polycomb targets show preferential enrichment for KDM2B and depletion of KDM2A. This striking relationship between PcG-bound CGI elements and preferential KDM2B enrichment held true when the ratio of KDM2B to 2A was plotted across all CGIs (Figure 3.9, scatter plot shown in red).

The enrichment of KDM2B at PcG-repressed CGIs is intriguing. Although the permissive chromatin architecture at CGIs renders them intrinsically accessible, PcG CGIs are believed to exist in a more compacted chromatin state, partly as a result of the histone modifications they impose (Eskeland et al. 2010a; Grau et al. 2011; Yuan et al. 2012). This repressive chromatin state is believed to be inhibitory to transcription, although an exact mechanism of how this is achieved has not been clearly demonstrated.

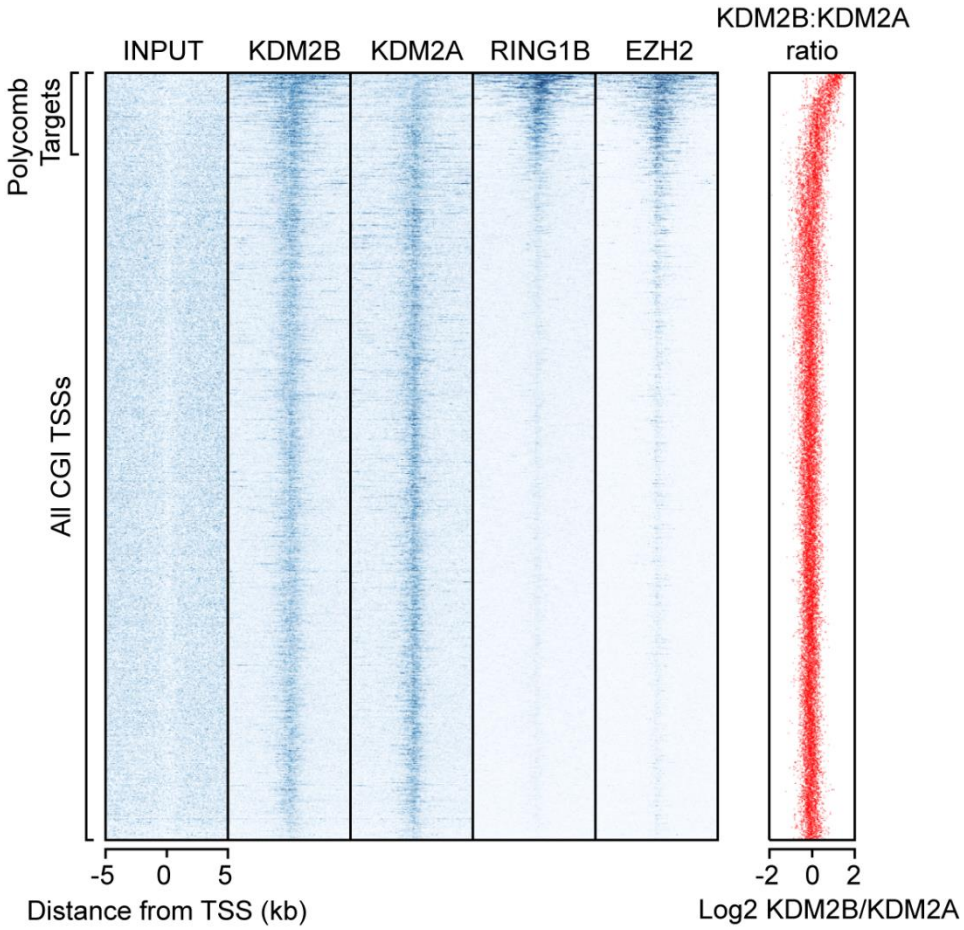


Figure 3.9. KDM2B is preferentially enriched at Polycomb-repressed CGIs.

Heat map illustrating ChIP-seq signal at all CGI-associated transcription start sites (TSSs) in mouse ESCs indicating depletion of KDM2A and enrichment of KDM2B at polycomb target CGIs marked by RING1B and EZH2. The scatter plot (far right in red) illustrates the log ratio of KDM2B to KDM2A enrichment at the same intervals depicted in the heat map as a scatter plot.

Interestingly, previous work by Thomson et al. (2010) suggested that PcG CGIs, potentially due to their compacted nature, are mostly refractory to binding of the ZF-CxxC domain containing protein CFP1. This observation appears to be valid for KDM2A as well (Figure 3.9). The reason for this is the nature of the ZF-CxxC domain itself. Although a small and compact DNA binding module, the ZF-CxxC domain requires access to both major and minor groove for efficient DNA binding (Xu et al. 2011a). This requirement implies that ZF-CxxC containing proteins cannot access DNA when occupied by nucleosomes, but can only interface chromatin in the accessible linker DNA regions , as was shown to be the case for KDM2A (Zhou et al. 2012). The compact chromatin state at Polycomb target loci would therefore limit the access of ZF-CxxC proteins to extra-nucleosomal linker DNA.

However, KDM2B does not appear to be depleted at PcG target CGIs, despite its ZF-CxxC binding domain potentially showing the same requirements in terms of accessing open linker DNA regions. While attempts to dissect this were carried out (as discussed in Chapter 5), it is still not clear how KDM2B preferentially binds Polycomb-occupied CGIs.

Another important observation resulting from the comparison of the KDM2s binding profile to those of RING1B and EZH2 (Figure 3.10) is the striking spatial relationship between PcG chromatin distribution and the underlying non-methylated DNA. RING1B and EZH2 do not map exclusively to gene-associated TSSs, but instead follow the chromatin binding profile of KDM2B and the corresponding non-methylated DNA signal at these regions. The exemplary panel of non-PcG CGIs (Figure 3.10, panels A-C) and developmental regulatory PcG-occupied genes (Figure 3.10, panels D-F) illustrates that PcG protein occupancy seems to closely mirror non-methylated CpG density.

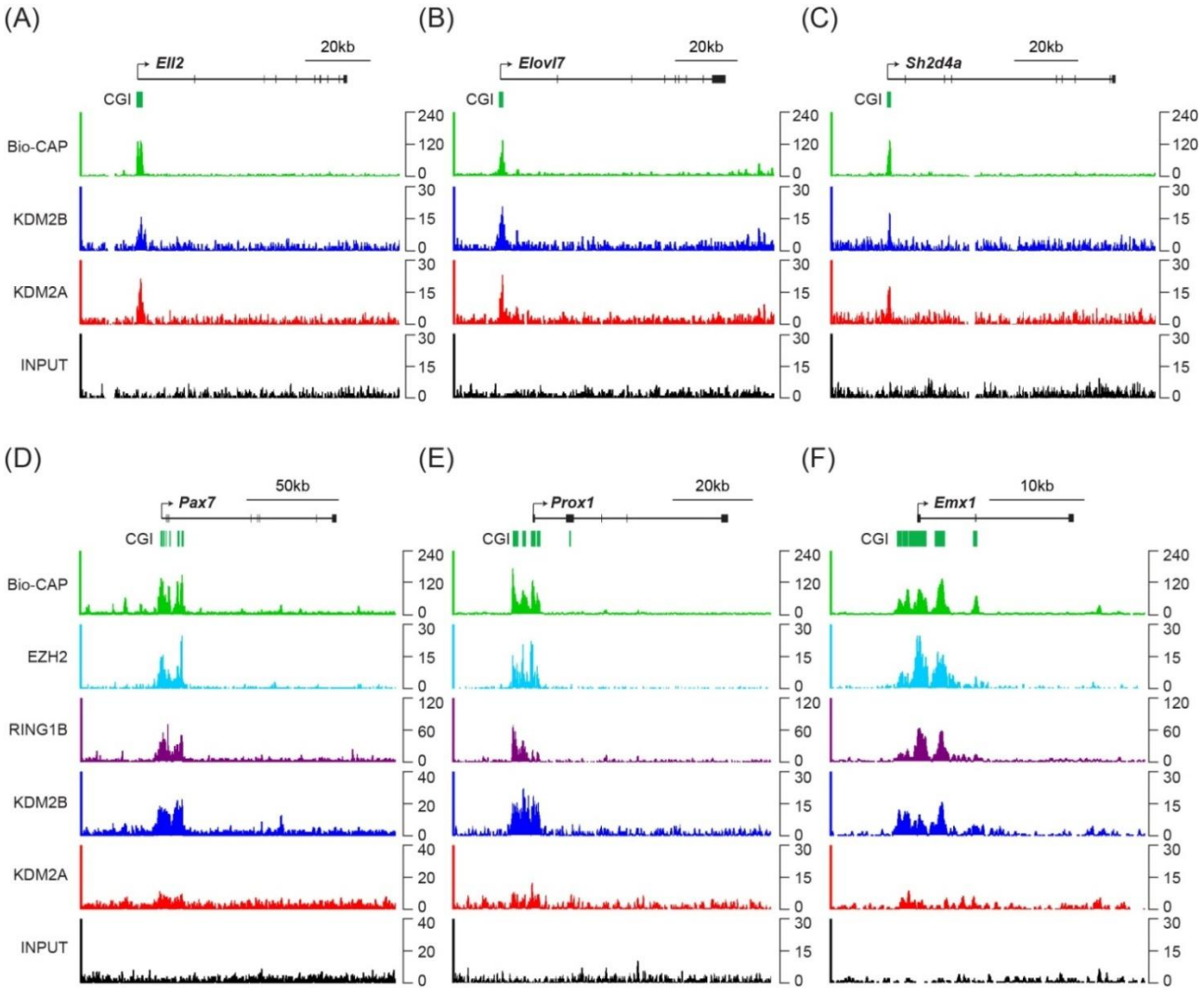


Figure 3.10. KDM2B is preferentially enriched at CGIs occupied by Polycomb group proteins.

(A) – (C) ChIP-seq profiles for KDM2A and KDM2B over a panel of non-polycomb associated CGIs. Similar enrichment of KDM2A and B is observed at these sites.

(D) – (F) ChIP-seq profiles for KDM2A, KDM2B, RING1B (PRC1), and EZH2 (PRC2) over a panel of polycomb associated CGIs. KDM2B is enriched and KDM2A depleted at these sites.

The preferential enrichment of KDM2B at Polycomb-associated CGIs, together with the striking overlap between PcG protein occupancy and non-methylated CpG density, opened the possibility that KDM2B could play a role in linking recognition of non-methylated DNA to PcG-mediated repression of developmental genes in mouse ESCs.

3.8. Summary and discussion

CGIs continue to remain enigmatic genomic elements with an unclear contribution to the regulation of gene expression. The discovery that ZF-CxxC domain-containing proteins are able to recognize non-methylated CpG dinucleotides and localize specifically to CGIs provided an exciting avenue to explore the functionality of these genomic elements. Previous work has shown that a histone lysine demethylase called KDM2A localizes to CGIs genome-wide via its ZF-CxxC DNA binding domain, where it catalyzes the removal of a specific histone mark, H3K36me2 (Blackledge et al. 2010). However, the biochemical properties and physiological role of the highly homologous protein KDM2B have been largely unexplored.

To better understand the contribution of KDM2B to CGI function, an antibody specifically recognizing this protein was first generated and characterized. Using this tool, work in this chapter has demonstrated that KDM2B is a nuclear protein, with a broad nucleoplasmic staining pattern. Furthermore, ChIP-Seq experiments using the newly generated antibody showed that KDM2B binds CGIs genome-wide in mouse ESCs. Interestingly, there are other ZF-CxxC domain containing proteins which were shown to localize to the same genomic elements. In addition, similar to KDM2A and 2B, these proteins are also often associated with specific chromatin modifying activities. For example, CFP1 is a component of the mammalian SET1 H3K4-methyltransferase complex (Lee and Skalnik 2005), MLL1/MLL2 mediate H3K4me3 at a subset of actively transcribed CGI-associated genes (Guenther et al. 2005), while TET1/3 were shown to be involved in active DNA demethylation by catalysing the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) (Ito et al. 2010).

The convergence of ZF-CxxC domain containing proteins at CGIs raises the important question of whether there is a relationship or interplay between these proteins. Although associated with distinct chromatin modifying activities, it will become essential to define whether the different ZF-CxxC proteins interface CGIs at the same time, with similar kinetics and whether they influence one another. Disruption of any of these chromatin-modifying complexes can have dramatic consequences, and many of these factors can promote genomic instability or are frequently mutated in various malignancies

(Kooistra and Helin 2012; Tan and Shi 2012). Therefore, it will become important to address how the different ZF-CxxC domain-containing proteins function, how their binding specificities are defined in the context of the nuclear environment and how they impact on the precisely coordinated and critical transcriptional programs during development.

The observation that two closely related KDM2 proteins have non-overlapping chromatin binding profiles, with KDM2B preferentially enriched at developmental regulatory gene-associated CGIs, further underlines the necessity of understanding the subtle differences between these two ZF-CxxC proteins. As KDM2A and 2B possess an almost identical DNA binding domain architecture, the distinct preferential enrichment of KDM2B at Polycomb-occupied CGIs is intriguing. Interestingly, Polycomb-group protein occupancy in mammals correlates almost exclusively with CpG islands, although the mechanisms responsible for this recruitment remain poorly defined (Ku et al. 2008). Therefore, the striking overlap between KDM2B and RING1B/EZH2 at Polycomb-occupied genes provides the possibility of an unexplored connection between KDM2B-mediated recognition of non-methylated DNA and Polycomb protein occupancy.

While this work was in progress, H3K36me2 was shown to inhibit PRC2-mediated H3K27me3 deposition (Yuan et al. 2011). This observation prompted the possibility that the demethylase activity of KDM2B might be required to remove K36me2, thereby allowing Polycomb-mediated repression of developmental regulatory genes. Although the co-localization of KDM2B and PcG-group proteins at a subset of CGIs could also be based on unrelated events, an alternative possibility is that, via its capacity to specifically recognize non-methylated DNA, KDM2B could contribute to the recruitment of Polycomb proteins. Further work is clearly needed to determine the specific DNA elements or protein interactions that mediate PcG recruitment in mammals. Nevertheless, the discovery that, unlike other ZF-CxxC proteins, KDM2B is preferentially enriched at PcG-occupied CGIs provides an unexpected link between Polycomb repressive proteins and non-methylated DNA.

4. Chapter four - KDM2B forms a variant PRC1 complex containing RING1B and PCGF1

ChIP-Seq chromatin binding profiles for KDM2A and KDM2B in mouse ESCs showed a preferential enrichment of KDM2B at developmentally regulated genes which are occupied by Polycomb group proteins. Interestingly, earlier studies using cancer cell lines suggested that the PRC1 E3 ubiquitin ligase RING1B is part of at least two distinct protein complexes, the classical PRC1 Polycomb complex and a novel complex containing the BCOR transcriptional corepressor and the histone demethylase KDM2B (Gearhart et al. 2006; Sanchez et al. 2007). Although not clear, this complex was believed to function as an alternative mechanism by which BCOR represses BCL6-target genes in an HDAC-independent manner. Importantly, both studies suggested that the RING1B-BCOR complex is catalytically active as an H2A E3 mono-ubiquitin ligase.

The link between KDM2B and RING1B seems to be a conserved feature, as Lagarou et al. (2008) showed that the Drosophila dRING protein forms a complex distinct from the classical PRC1 (Lagarou et al. 2008). Termed dRAF (dRING-associated factors), this complex comprises the single KDM2 homologue in flies, as well as the classical PcG member, Posterior Sex Combs (PSC). Interestingly, dKDM2 knockdown studies and *in vitro* reconstitution reactions suggested that dRAF rather than PRC1 was the main H2A ubiquitylating complex in flies. Furthermore, genetic interactions analysis revealed that dKDM2 enhances homeotic transformations in Polycomb (Pc) heterozygotes flies, underlining the contribution of dKDM2 to PcG-mediated gene repression.

Whether KDM2A and KDM2B form protein complexes in mouse embryonic stem cells (ESCs) has not been previously examined. Both KDM2s contain an F-box motif known in other proteins to interact with SKP1 and mediate formation of SCF (Skp1–Cullin–F-box containing)-type E3 ubiquitin ligase complexes involved in targeting proteins for proteasomal degradation (Jin et al. 2004). Furthermore, both proteins

have a number of leucine-rich repeat (LRR) motifs, which often act as a scaffold for protein-protein interactions (Kobe and Deisenhofer 1994; Enkhbayar et al. 2004). Despite their similarity, the two KDM2s do not have identical chromatin binding profiles, with KDM2B preferentially enriched at Polycomb-occupied CGIs. Therefore, to better understand how KDM2A and KDM2B might exert their unique functional properties, an affinity purification strategy was devised to identify potential protein interaction partners.

4.1. KDM2A and KDM2B have distinct gel filtration elution profiles

To understand the function of KDM2A and KDM2B, it was important to examine their biochemical properties and determine whether they are associated with protein partners. To begin addressing this question, gel filtration analysis was used, a technique which separates proteins based on their molecular size. Mouse embryonic stem cell (ESC) nuclear extract was fractionated using Superose 6 size-exclusion chromatography column and the fractions were probed by Western blot for KDM2A and KDM2B (Figure 4.1). Given that previous reports identified a RING1B-KDM2B multiprotein complex in transformed mammalian cells (Gearhart et al. 2006; Sanchez et al. 2007), the size exclusion chromatography fractions were also probed for RING1B. The majority of KDM2B fractionated with an apparent molecular size of 440 kDa to 2 MDa, and significantly overlapped with RING1B-containing fractions. In contrast, KDM2A eluted with an apparent molecular size of 200–500 kDa. Given that the molecular mass of KDM2B based on its amino acid sequence is approximately 150kDa, this elution profile suggests that KDM2B may reside in a multiprotein complex in mouse ESCs. It is important to consider that, although overlapping, the similar gel filtration profile for KDM2B and RING1B is not sufficient to indicate that the two proteins are part of the same complex. However, the distinct gel filtration elution profiles of KDM2A and KDM2B suggests that the two KDM2s could have potentially unique biochemical properties and likely do not reside in the same protein complexes.

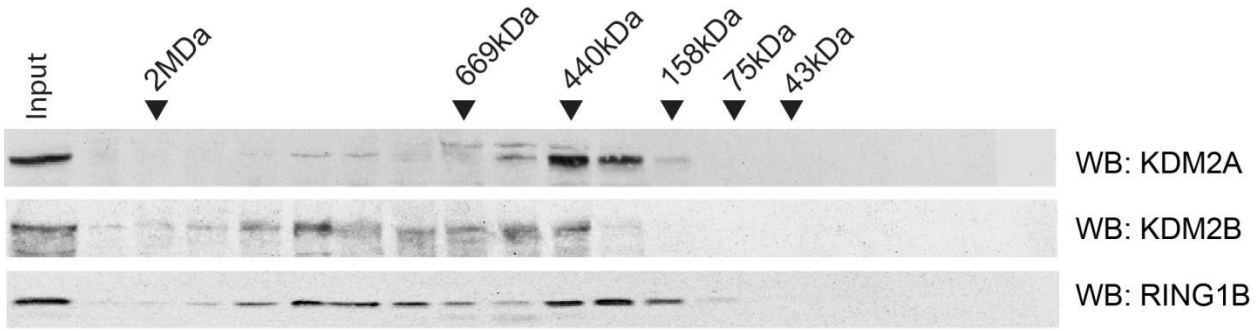


Figure 4.1. KDM2A and KDM2B have different gel filtration elution profiles.

Superose 6 size exclusion chromatography of nuclear extracts from mouse ES cells. Fractions were TCA-precipitated and probed by Western blot with antibodies against KDM2B, KDM2A and RING1B, as indicated. Input represents 40 μg of nuclear extract. The elution profile of the molecular weight standards is indicated above the blots. KDM2B resides in higher molecular weight fractions than KDM2A.

4.2. A Strep-tag based affinity purification approach for the identification of protein complexes

To understand whether KDM2A and KDM2B form protein complexes, an affinity-based purification strategy was devised which required the generation of cell lines stably expressing tagged versions of these proteins. For this purpose, a mammalian expression vector optimized for ease of cloning, stable constitutive expression and efficient and fast antibiotic selection was first constructed. The full length cDNA encoding the protein of interest was PCR-amplified and inserted via ligation independent cloning (LIC) into a pCAG-IRES-puro eukaryotic expression vector that has been modified to express a N-terminal Flag and 2x StrepII (FS2) tag followed by a TEV protease cleavage site. pCAG-IRES-puro is a bicistronic vector, with expression of insert cDNA and puromycin resistance genes driven by the strong synthetic CMV immediate enhancer/chicken β-actin (CAG) promoter.

The puromycin resistance gene follows the internal ribosomal entry site (IRES) element, allowing the selection of cell lines expressing the protein of interest. Mouse feeder-independent ES cells were transfected with the mammalian expression construct coding for Flag/2xStrepII (FS2)-tagged full-length KDM2A or KDM2B and clonal stable cell lines were isolated following 7-10 days of puromycin selection. Single ESC clones were expanded and screened by Western blotting of whole-cell extract for expression

of tagged protein (Figure 4.2, panel A). Because the gene insert and the antibiotic selection marker are translated from a single RNA, generally 90-100% of colonies that are resistant to puromycin also express the protein of interest.

Clones with stable expression of KDM2A and KDM2B were expanded under continued puromycin selection on approximately 100x15cm dishes, corresponding to ~5-8 x 10^9 cells. At confluence, the adherent cells were harvested by scraping and the pellet used for nuclear extract preparation by salt extraction method as described (Dignam et al. 1983), with minor modifications. 10-15mg of nuclear extract prepared from stable cell lines expressing epitope-tagged protein of interest was used to carry out affinity purification (Figure 4.2, panel B). The protein purification system employed takes advantage of the strong binding affinity of the Strep-II peptide tag for an engineered streptavidin resin called Strep-Tactin. Variants of this protocol were previously described and successfully used to characterize other protein complexes (Groth et al. 2005; Junttila et al. 2005; Schmidt and Skerra 2007). Nuclear extract was incubated with high-binding capacity Strep-Tactin resin, followed by stringent wash steps to remove non-specifically bound proteins. An additional benefit of the purification method is the highly efficient elution of bound material from the Strep-Tactin resin using biotin, or its reversibly binding and stable analogue, desthiobiotin, without the need for harsh or denaturing conditions. Following gentle elution with desthiobiotin, the fractions collected were subjected to in-solution tryptic digestion followed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. The mass spectrometry work was done in collaboration with Dr. Joanna McGouran and Dr. Benedikt Kessler at the Nuffield Department of Medicine, University of Oxford.

The acquired MS/MS spectra for individual peptides were then searched against a comprehensive and non-redundant protein database (Uniprot/Swissprot) using the Mascot search engine. A minimum of two matching peptides was required to assign each protein. Proteins without any unique peptide were excluded, as they could not be unambiguously assigned as present in the purified fraction. After subsequent optimisation, the fast and simple one-step Strep affinity purification method was then used to study whether KDM2A and KDM2B form protein complexes in mouse ESCs.

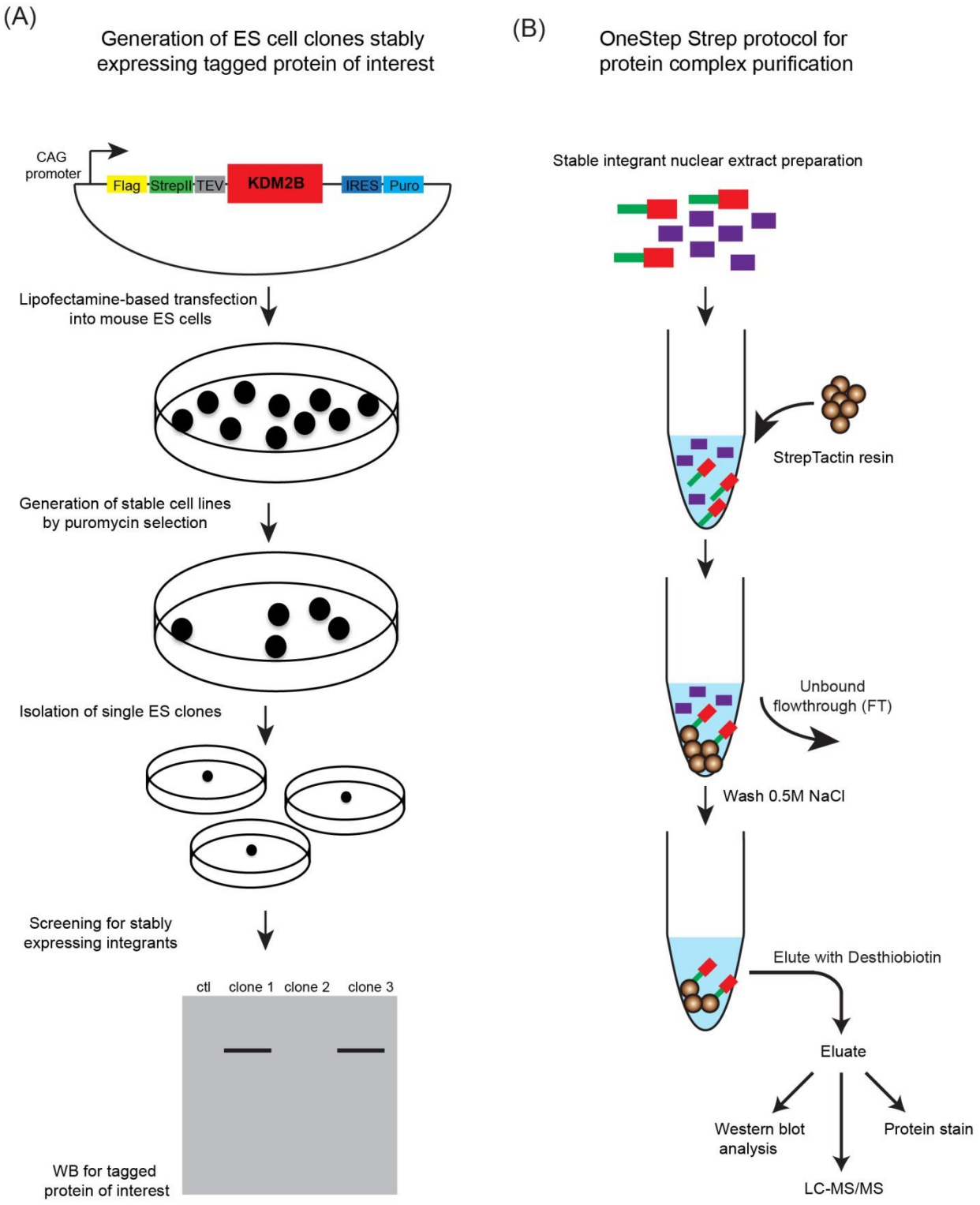


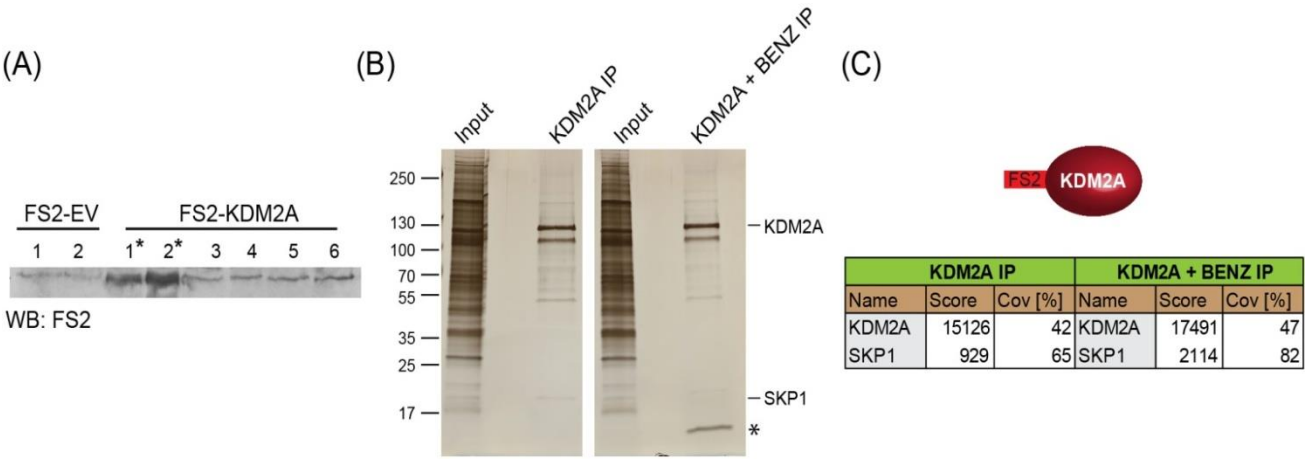
Figure 4.2. Generation of ES cell lines stably expressing tagged protein of interest and protein complex purification using the one-step Strep-tag approach.

(A) Schematic illustrating the construct used to generate stable ES cell lines expressing tagged full length protein of interest (KDM2B shown as example). Expression is driven by the CMV immediate enhancer/chicken β-actin (CAG) promoter. The puromycin resistance gene follows the IRES element allowing for efficient selection within 7-10 days in puromycin-containing media. Positive stably expressing clones were screened by Western blot of whole-cell extract with antibody against the StrepII tag.

(B) To purify potential protein complexes (in this case KDM2B and associated proteins), a mouse ESC line stably expressing Flag-2XStrepII-tagged KDM2B was generated. Nuclear extract was isolated from this cell line, and KDM2B affinity purified using StrepTactin resin with high binding capacity for the StrepII tag. 10-15mg of nuclear extract was generally used per affinity purification experiment. The efficiency of the purification was assessed by Western blot and visualization of the eluted fractions by protein staining (Silver or SyproRuby staining). The purified proteins were then analysed by tryptic digestion followed by peptide identification by mass spectrometry.

4.3. KDM2A associates with SKP1 in mouse ESCs

To gain insight into the molecular function of KDM2A, immunoprecipitation followed by mass spectrometry analysis was performed to identify potentially associated proteins. After transfection with a mammalian expression construct coding FS2-tagged full-length human KDM2A, stable cell lines were screened by Western blot (Figure 4.3, panel A). Interestingly, it was difficult to establish KDM2A stable cell lines, suggesting that overexpression of the protein is potentially toxic to the cells. Nuclear extract from stable ESC lines expressing epitope-tagged KDM2A was prepared and used for Strep-Tactin based affinity purification. To exclude the possibility that any identified interactions were indirect and mediated through DNA, parallel purifications in which the extract was pre-treated with benzonase to remove contaminating DNA were performed. In addition, purifications from a stable cell line containing only the empty tagging expression construct were done to exclude non-specific interactions with the affinity matrix. The eluted fractions were resolved by gradient SDS-PAGE and silver stained (Figure 4.3, panel B). Silver staining detects presence of full-length and a degradation product of KDM2A (double band between 100-130kDa), as confirmed by Western blotting (data not shown). To determine the identity of additional bands, the fractions were subjected to in-solution tryptic digestion followed by LC-MS/MS analysis. In the purifications from extracts expressing epitope-tagged KDM2A, only SKP1 was identified with high confidence as an interaction partner (Figure 4.3, panel C). A series of additional proteins, such as keratin, actin, tubulin, mitochondrial, ribosomal or hnRNP proteins, were also identified. However, they were also detected in analysis of empty vector cell line so were not considered KDM2A interaction partners.



KDM2A IP			KDM2A + BENZ IP		
Name	Score	Cov [%]	Name	Score	Cov [%]
KDM2A	15126	42	KDM2A	17491	47
SKP1	929	65	SKP1	2114	82

Figure 4.3. KDM2A associates with SKP1 in mouse ESCs.

(A) To purify KDM2A and any associated proteins, a mouse ESC cell line stably expressing Flag-2XStrepII-tagged (FS2) KDM2A was generated. Whole-cell extract screening allowed identification of FS2-tagged KDM2A cell lines (marked by asterisk), with the clones selected showing a reactive band at the expected size compared to empty vector lines.

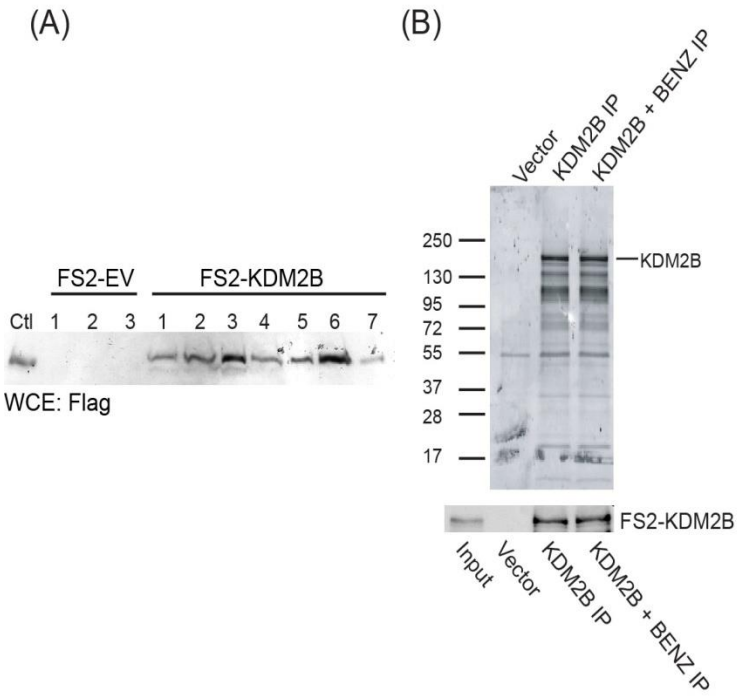
(B) Nuclear extract was isolated from tagged-KDM2A expressing cell line and KDM2A and associated proteins affinity purified using the One-step Strep purification method. Purified KDM2A fractions were resolved by gradient SDS-PAGE and visualized by silver staining. The purifications were performed in the absence and presence of benzonase to exclude DNA-mediated interactions and a cell line containing only the empty vector was used to control for non-specific binding to the affinity matrix. The band indicated by an asterisk most probably represents histone protein released from DNA by the benzonase treatment.

(C) Elutions from the KDM2A affinity purification were directly analysed by tryptic digestion followed by peptide identification by LC–MS/MS. The Mascot scores and peptide coverage are shown for the respective affinity purifications. KDM2A in ESCs stably associates with SKP1.

The F-box motif found in KDM2A is known in other proteins to interact directly with the SCF protein SKP1 and function as substrate adaptors to mediate proteasomal degradation of diverse targets (Skaar et al. 2013). However, SCF complexes are generally assembled around the essential structural scaffold protein CUL1 (cullin 1). The absence of cullin proteins in the KDM2A purifications suggests that KDM2A-SKP1 may not form a typical SCF complex, but could be involved in a non-canonical E3 ubiquitin ligase pathway, potentially mediating protein degradation of uncharacterized substrates.

4.4. KDM2B associates with Polycomb-group proteins in mouse ESCs

Given its preferential enrichment at PcG-occupied CGIs and previous reports that it associates with a variant PRC1 complex in transformed cells, it was important to identify potential KDM2B interaction partners in mouse ES cells. To this end, stable cell lines expressing FS2-tagged full-length human KDM2B were generated and screened by Western blot analysis (Figure 4.4, panel A). The majority of clones tested showed successful expression of tagged KDM2B. Stable ESC lines expressing FS2-tagged KDM2B were expanded, and nuclear extract generated for affinity purification using the one-step Strep approach (Section 4.2). The affinity purification was performed in the presence and absence of benzonase nuclease, to ensure interactions were not mediated through DNA. The eluted fractions were resolved by gradient SDS-PAGE and visualized by SyproRuby fluorescent protein stain (Figure 4.4, panel B). In addition to the band running at the expected size of KDM2B (approximately 150kDa), several other prominent bands not visible in the empty vector control were detectable. The elutions were probed by Western blot for tagged-KDM2B before analysis by LC-MS/MS, to verify that the purification efficiently pulled down the bait protein. Interestingly, the mass spectrometry analysis of the KDM2B purification indicates that in mouse ESCs, KDM2B associates with a series of proteins, including BCOR/BCORL1 and the PRC1 components RING1B, PCGF1 and RYBP/YAF2 (Figure 4.4, panel C). As in the case with KDM2A, SKP1 was identified as a strong interaction partner, presumably also mediated by the F-box motif present in KDM2B.



(C)

FS2 KDM2B

	KDM2B IP		KDM2B+BENZ IP	
Name	Score	Cov [%]	Score	Cov [%]
KDM2B	16279	78	16693	52
RING1B	330	58	320	45
BCoR	874	17	997	17
BCoRL1	332	11	487	12
PCGF1 (NSPC1)	277	39	273	39
RYBP	95	34	84	11
YAF2	23	13	48	30
SKP1	4321	88	5332	93

Figure 4.4. KDM2B associates with RING1B and PCGF1 to form a variant PRC1 complex in mouse ESCs.

(A) To purify KDM2B and associated proteins, a mouse ESC cell line stably expressing Flag-2XStrepII-tagged (FS2) KDM2B was generated. Whole-cell extract screening indicated that generation of stable cell lines using the pCAG-promoter system is highly efficient, with all clones screened showing a reactive band at the expected size of KDM2B compared to empty vector lines. Ctl indicates whole cell extract from HeLa cells transiently overexpressing FS2-tagged KDM2B.

(B) Nuclear extract was isolated from exogenous KDM2B expressing cell line and KDM2B and associated proteins affinity purified using the One-step Strep system. Purified KDM2B fractions were resolved by gradient SDS-PAGE and visualized by SyproRuby staining. The purifications were performed in the absence and presence of benzonase to exclude DNA-mediated interactions and a cell line containing only the empty vector was used to control for non-specific binding to the affinity matrix. The elutions were probed by western blot for KDM2B as indicated.

(C) Elutions from the KDM2B affinity purification were directly analysed by tryptic digestion followed by peptide identification by LC–MS/MS. The Mascot scores and peptide coverage are shown for the respective affinity purifications. KDM2B in ESCs associates with a variant PRC1 complex containing RING1B, BCOR/BCORL1, PCGF1, RYBP, YAF2 and SKP1.

Although the interaction between KDM2B and PRC1 components is in agreement with previous work in cancer cell lines (Gearhart et al. 2006; Sanchez et al. 2007), it is important to note that some of the other KDM2B interaction partners previously identified (such as subunits of casein kinase 2 or the HSP70 chaperone) were not detected in these purifications from mouse ESCs. This could be due to these proteins having lower expression levels in ESCs. Alternatively, they could represent weak or transient interaction partners of KDM2B, not captured or not preserved during the high salt stringent wash steps of the Strep-Tactin based purification procedure. Heat shock proteins are however common contaminants in mass spectrometry experiments, and HSP70-related proteins were detected in purifications from the empty vector control line.

The association of KDM2B with Polycomb-group proteins RING1B, PCGF1 and RYBP/YAF2 suggested that the relationship between KDM2B and PcG-occupied CGIs could be mediated through inclusion of KDM2B in a PRC1 complex. Interestingly, while this work was in progress, a comprehensive proteomic and genomic analysis identified six groups of PRC1 complexes, distinguished by the presence of a different member of the PCGF family (Gao et al. 2012). Importantly, this study also identified KDM2B as part of the PCGF1-containing PRC1 complex in transformed 293T cells.

The interaction between KDM2B and BCOR or BCOR-like 1 (BCORL1) proteins is intriguing. Both BCOR and BCORL1 genes are located on the X-chromosome and encode for large nuclear proteins (approximately 190kDa) associated with a number of diseases, such as OFCD (oculo-facio-cardio-dental syndrome) and AML (acute myeloid leukemia) (Huynh et al. 2000; Pagan et al. 2007; Tiacci et al. 2012). Furthermore, both proteins were reported to interact with specific DNA-binding repressive factors (BCOR with BCL6, while BCORL1 with CTBP1) and mediate transcriptional silencing of target genes in an HDAC-dependent manner. However, HDACs were not detected in the KDM2B affinity purification, suggesting that the KDM2B-BCOR/BCORL1 forms a distinct complex characterized by members of the Polycomb-group complex. The observation that the highly similar KDM2A protein does not interact with any Polycomb-group proteins suggests that differences in KDM2 protein interactions could potentially account for the distinct chromatin binding profiles of KDM2A and KDM2B and explain why KDM2B is preferentially enriched at PcG-targets.

4.5. KDM2B interacts with RING1B in the absence of the JmjC catalytic domain

The KDM2B gene has an alternative transcription start site downstream of the JmjC catalytic domain, giving rise to a short isoform of KDM2B (Figure 4.5, panel A). This shorter isofom was reported to be ubiquitously expressed during embryonic development and in various adult tissues (Fukuda et al. 2011). As the short form retains a presumably functional ZF-CxxC DNA binding domain, it is expected to recognize CGIs. However, the KDM2B antibody generated in this study was raised in a region shared between the two isoforms and cannot distinguish between them in ChIP experiments. Furthermore, as full-length KDM2B interacts with RING1B, it was important to test whether the short form of KDM2B also has the ability to form a variant PRC1 complex. To achieve this, a stable mouse ESC line expressing FS2-tagged KDM2B short form (SF) was generated and nuclear extract used to carry out affinity purification, as in the case of full-length KDM2B (FL). Mass spectrometry analysis revealed that this protein interacts with RING1B and forms an identical variant PRC1 complex as that previously identified for the long form of KDM2B (Figure 4.5, panel B).

Importantly, in the purification from extracts expressing epitope-tagged KDM2B SF, no peptides resulting from the MS fragmentation were found to cover the JmjC catalytic domain (data not shown). This suggests that KDM2B FL and SF do not associate in cells, but exist as independent complexes, both interacting with Polycomb-group components.

The role of KDM2B SF is currently poorly defined. However, the presence of the ZF-CxxC DNA binding domain and the ability to interact with RING1B and form a variant PRC1 complex suggest that the two KDM2B isoforms must be studied together to understand the contribution of KDM2B to CGI function. Furthermore, it will become important to address whether the complexes formed by the two KDM2Bs are interchangeable or if KDM2B FL and SF are differentially regulated in the cell.

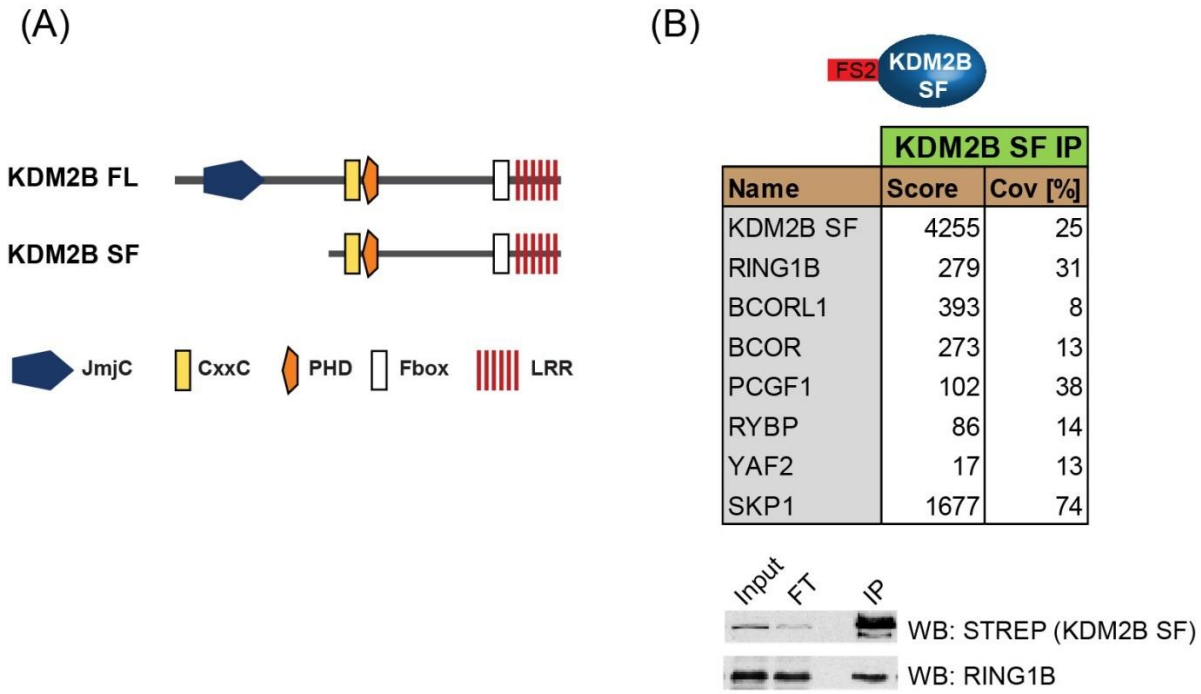


Figure 4.5. KDM2B interacts with RING1B in the absence of the JmjC catalytic domain.

(A) Schematic representation of KDM2B full-length (FL) and short-form (SF). Due to use of an alternative downstream promoter, KDM2B SF lacks the catalytic JmjC domain.

(B) Affinity purification followed by LC-MS/MS was carried out from mouse ESCs stably expressing epitope-tagged KDM2B SF protein. The eluted fractions were probed by Western blot for epitope-tagged KDM2B SF and RING1B. The Mascot scores and peptide coverage are shown for the respective protein hits. KDM2B SF retains the ability to interact with RING1B and form the variant KDM2B-PRC1 complex.

4.6. KDM2B forms a variant PRC1 complex characterized by the PCGF1 subunit

Affinity purification followed by mass spectrometry is a powerful technique to identify binding partners for a protein target of interest. However, this method will reveal all associating proteins, without the capacity to distinguish whether they reside in a single or in distinct complexes. Therefore, to better define the composition of the KDM2B core complex, reciprocal immunoaffinity purification experiments were performed for the interacting Polycomb components. Nuclear extracts prepared from cell lines expressing FS2-tagged RING1B, YAF2, RYBP or PCGF1 were used for affinity purification. SyproRuby fluorescent staining indicated the presence of multiple bands, the most prominent corresponding to the overexpressed epitope-tagged bait. This was confirmed by Western blot signal for the tagged proteins at

their expected molecular weight (Figure 4.6). Attempts to generate ESC lines expressing epitope-tagged BCOR and BCORL1 were unsuccessful, but should be pursued to better characterize their associated protein partners.

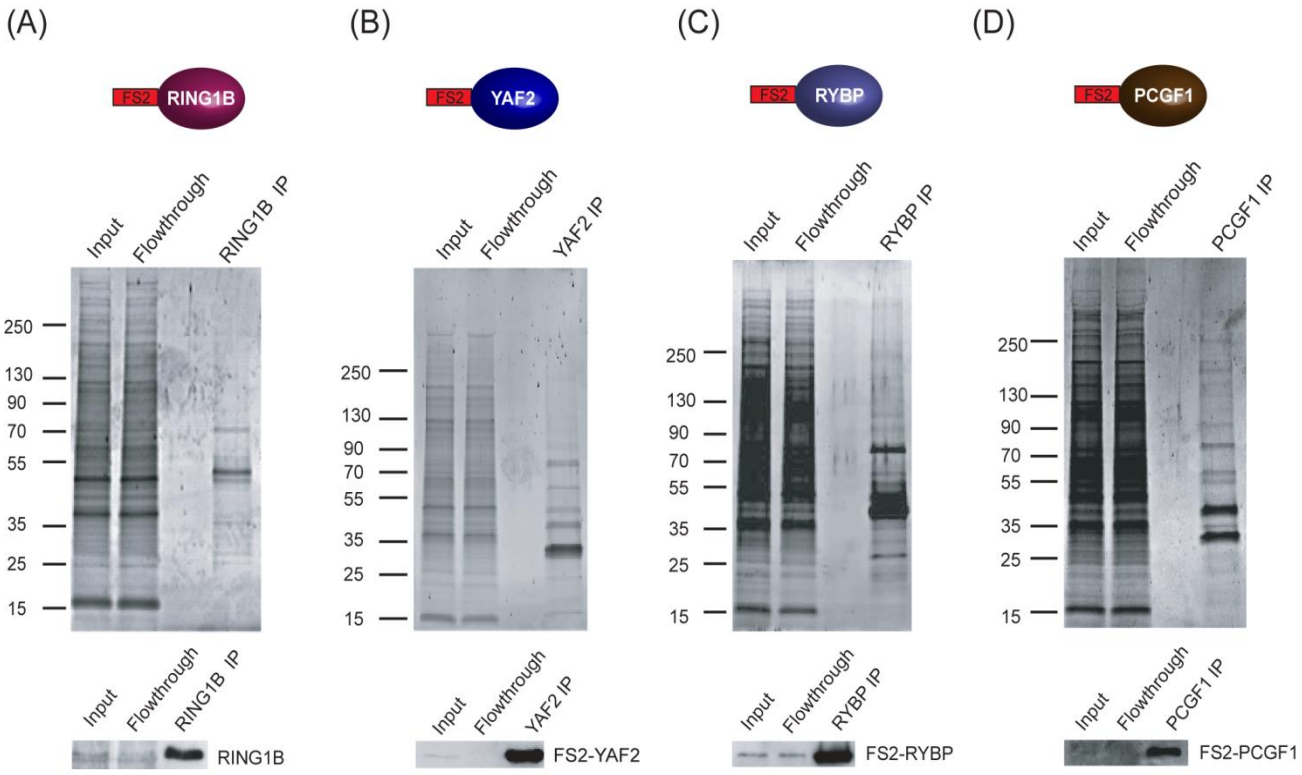


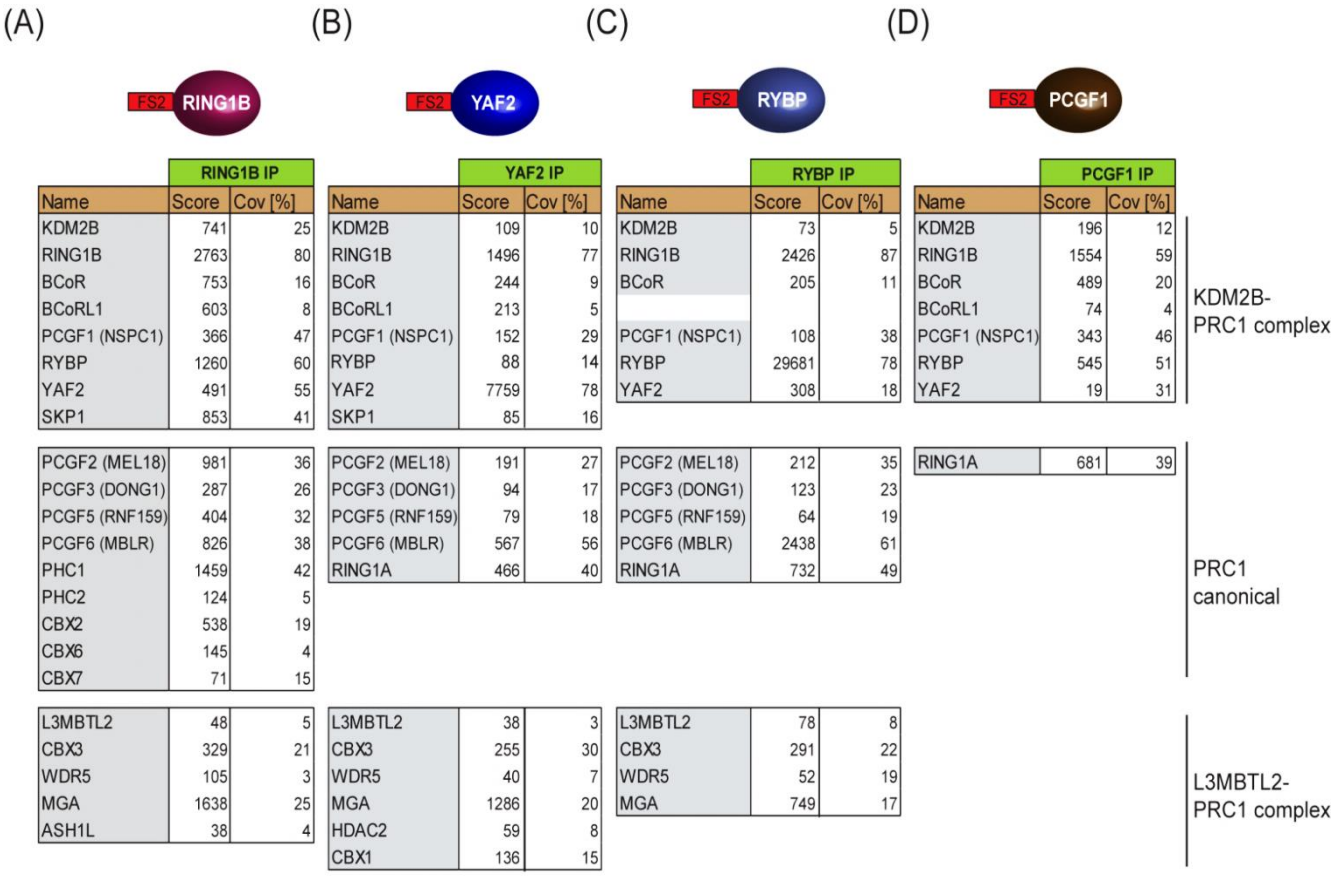
Figure 4.6. Reciprocal affinity purifications of KDM2B interaction partners.

(A)–(D) Stable cell lines expressing Flag-2XStrepII tagged RING1B (A), YAF2 (B), RYBP (C), and PCGF1 (D) were generated and each protein affinity purified from mouse ESC nuclear extract. Purified samples were resolved by gradient SDS-PAGE and visualized by SyproRuby staining. In each case the input, flowthrough, and elution is indicated above. The elutions were probed by western blot against the tagged protein as indicated below each SyproRuby stained gel.

The associated proteins for RING1B, YAF2, RYBP and PCGF1 were identified by mass spectrometry analysis. While the KDM2B-PRC1 variant complex and its associated components were identified in all the purifications, it was apparent that RING1B, YAF2 and RYBP also interact with additional proteins (Figure 4.7). As the E3 ubiquitin ligase catalytic core of the PRC1 complex, RING1B was shown to interact with Polycomb-group ring finger (PCGF) proteins via their N-terminal RING-finger domains (Cao et al.

2005; Buchwald et al. 2006; Li et al. 2006; Vandamme et al. 2011; Gao et al. 2012). Exclusive interaction with any of the six members of the PCGF family is believed to define functionally distinct PRC1 variant complexes, with different genomic localization profiles. In the RING1B purification, five out of the six PCGF proteins were identified (Figure 4.7, panel A). PCGF4 (also known as BMI1) was not unambiguously detected in any of the purifications, perhaps due to its reported low expression levels in ESCs (Ding et al. 2012). Importantly, polyhomeotic (PH) and chromobox-domain (CBX) proteins were also identified. In mammals, PRC1 complexes were shown to divide into canonical and non-canonical forms (Gao et al. 2012). Canonical PRC1 complexes are formed by the association of PCGF2 or PCGF4 with CBX and PH proteins. This is then believed to mediate crosstalk with PRC2 complexes through recognition of H3K27me3 mark by a chromodomain in the CBX subunit (Cao et al. 2002; Fischle et al. 2003). In non-canonical PRC1 complexes, the catalytic PCGF-RING1B core interacts with RYBP or the closely related protein YAF2, thereby excluding the CBX subunit (Wang et al. 2010). The PCGF protein then dictates the association of the non-canonical PRC1 complex with distinct additional subunits.

The RYBP and YAF2 affinity purifications yielded similar results to those obtained for RING1B (Figure 4.7, panel B and C), with the observation that RYBP or YAF2 complexes do not integrate CBX proteins, but are unique to non-canonical PRC1 complexes. Interestingly, a distinct set of proteins was identified in the RING1B, YAF2 and RYBP affinity purifications, which include L3MBTL2, MGA, WDR5 and CBX1 (HP1 β)/CBX3 (HP1 γ). These were previously reported to form a distinct variant PRC1 complex, characterized by the PCGF6 subunit (Trojer et al. 2011; Qin et al. 2012). The L3MBTL2-PRC1 complex is believed to regulate expression of a subset of target genes not occupied by canonical PRC1 and PRC2 (Qin et al. 2012), although how this is achieved remains poorly defined. To verify that the L3MBTL2 complex is distinct from the one formed by PCGF1, ESC cell lines stably expressing FS2-tagged PCGF6 were generated and nuclear extract from these lines used to carry out affinity purification followed by mass spectrometry analysis (supplementary figure S1). Importantly, no KDM2B or PCGF1 were detected in the PCGF6 affinity purification, suggesting that PCGF1 and PCGF6 define distinct non-canonical PRC1 complexes.



RING1B IP		
Name	Score	Cov [%]
KDM2B	741	25
RING1B	2763	80
BCoR	753	16
BCoRL1	603	8
PCGF1 (NSPC1)	366	47
RYBP	1260	60
YAF2	491	55
SKP1	853	41
PCGF2 (MEL18)	981	36
PCGF3 (DONG1)	287	26
PCGF5 (RNF159)	404	32
PCGF6 (MBLR)	826	38
PHC1	1459	42
PHC2	124	5
CBX2	538	19
CBX6	145	4
CBX7	71	15
L3MBTL2	48	5
CBX3	329	21
WDR5	105	3
MGA	1638	25
ASH1L	38	4

YAF2 IP		
Name	Score	Cov [%]
KDM2B	109	10
RING1B	1496	77
BCoR	244	9
BCoRL1	213	5
PCGF1 (NSPC1)	152	29
RYBP	88	14
YAF2	7759	78
SKP1	85	16
PCGF2 (MEL18)	191	27
PCGF3 (DONG1)	94	17
PCGF5 (RNF159)	79	18
PCGF6 (MBLR)	567	56
RING1A	466	40
L3MBTL2	38	3
CBX3	255	30
WDR5	40	7
MGA	1286	20
HDAC2	59	8
CBX1	136	15

RYBP IP		
Name	Score	Cov [%]
KDM2B	73	5
RING1B	2426	87
BCoR	205	11
PCGF1 (NSPC1)	108	38
RYBP	29681	78
YAF2	308	18
PCGF2 (MEL18)	212	35
PCGF3 (DONG1)	123	23
PCGF5 (RNF159)	64	19
PCGF6 (MBLR)	2438	61
RING1A	732	49
L3MBTL2	78	8
CBX3	291	22
WDR5	52	19
MGA	749	17

PCGF1 IP		
Name	Score	Cov [%]
KDM2B	196	12
RING1B	1554	59
BCoR	489	20
BCoRL1	74	4
PCGF1 (NSPC1)	343	46
RYBP	545	51
YAF2	19	31
RING1A	681	39

Figure 4.7. KDM2B forms a variant PRC1 complex characterized by PCGF1 subunit.

(A)- (D) Reciprocal affinity purifications and subsequent LC-MS/MS for RING1B, YAF2, RYBP, and PCGF1 confirm the interaction between KDM2B and these PRC1 components. PCGF1 is unique to the KDM2B-PRC1 complex. Protein identification scores and peptide coverage (Cov [%]) are indicated.

The proteins detected in the PCGF1 purification were almost identical to the KDM2B-PRC1 complex (Figure 4.7, panel D), suggesting that PCGF1 is unique to this variant complex. An additional protein identified in the PCGF1 but not in the KDM2B purification is RING1A, a functional H2AK119 ubiquitin ligase highly similar to RING1B. Although this difference could have a biological significance, the most likely explanation for this discrepancy is probably related to the low expression levels of RING1A in ESCs (Endoh et al. 2008). As RING1B was also detected with lower Mascot score in the KDM2B purifications compared to the PCGF1 mass spectrometry results, this suggests that KDM2B can associate with either RING1B or RING1A. From structural studies, PCGF1 interacts with RING1B via their RING-finger domains, while RING1B interacts with either RYBP or YAF2, this representing the Polycomb-core of the KDM2B complex. How the PCGF1 subunit then dictates specificity towards KDM2B or BCOR/BCORL1 is the focus

of ongoing work in the Klose lab. Interestingly, recent studies identified a protein interaction domain in BCOR which binds the RAWUL domain of PCGF1 but not of PCGF2 or PCGF4 (Junco et al. 2013). A direct interaction between KDM2B and BCOR was also proposed, but was difficult to characterize (Gearhart et al. 2006). Preliminary biochemical work in the Klose lab indicates that KDM2B can independently interact with both BCOR and PCGF1 (Dr. N.Rose, personal communication).

Based on the reciprocal affinity purifications, it can be concluded that KDM2B forms a non-canonical PRC1 complex consisting of RING1B, PCGF1, RYBP/YAF2 and BCOR/BCORL1. The association of KDM2B with Polycomb-group proteins in mouse ESCs suggests that the preferential enrichment of KDM2B at PcG-occupied CGIs could be mediated through its direct interaction with PRC1 components.

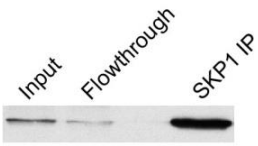
4.7. SKP1 associates with ZF-CxxC-domain containing F-box proteins

The SCF is the largest family of E3 ubiquitin ligases in mammals (Wei and Sun 2010) . SCF E3 ligases consist of the adaptor protein SKP1 (S-phase kinase associated protein 1), one F-box protein dictating the substrate specificity, one Cullin scaffold protein and one RING-family protein (RBX1 or RBX2) (Zhou et al. 2013). The human genome contains 69 F-box proteins, grouped into three classes depending on the types of substrate interaction domains identified in addition to the F-box motif (Jin et al. 2004). These classes were designated FBXW (WD40 interaction domain), FBXL (Leucine-rich repeats (LRRs) interaction domain) and FBXO (other or no recognizable interaction domains). Through promoting the degradation of many key regulatory proteins, SCF E3 ubiquitin ligases control numerous cellular processes, such as cell cycle progression, DNA replication, gene transcription or maintenance of genome integrity (Skaar and Pagano 2009; Silverman et al. 2012; Fukushima et al. 2013). In the affinity purifications for KDM2A and KDM2B, SKP1 was identified as a major interaction partner for both KDM2s. However, the scaffolding protein Cullin was not detected, suggesting that KDM2A and KDM2B do not form classical SCF complexes. One explanation could be that the inclusion of SKP1 in functional cullin-containing SCF complexes is mainly a cytoplasmic event.

Therefore, to better characterize the role of SKP1 in the nucleus, stable ESC lines expressing epitope-tagged SKP1 were established and affinity purification followed by LC-MS/MS was carried out to identify its associated nuclear proteins (Figure 4.8). Importantly, Polycomb-group proteins were identified, validating the involvement of SKP1 in the formation of a variant PRC1 complex containing KDM2B. In agreement with KDM2A affinity purification (Figure 4.3), KDM2A was also found to associate with SKP1 in mouse ESCs. Interestingly, SKP1 was found to associate with FBXL19, another ZF-CxxC domain containing protein highly similar to the two KDM2s, with the exception that it lacks the N-terminal JmjC catalytic domain. The presence of a conserved ZF-CxxC DNA binding domain suggests that FBXL19 can presumably recognize non-methylated CpGs and bind CGIs. Although otherwise poorly studied, FBXL19 was recently reported to interact with subunits of the Mediator complex, potentially establishing a link between the permissive chromatin state of CGIs and transcriptional regulation (Tan et al. 2013).

Interestingly, in the SKP1 purification from mouse ESC nuclear extract, two of the eight Cullin family members were detected (CUL1/CUL7), underlining the association of SKP1 with other F-box proteins in nuclear SCF complexes. The SCF complex formed by SKP1 and FBXL1 (SKP2) has been extensively studied and reported to be required for cancer development in multiple tumour-promoting conditions (Schulman et al. 2000; Chan et al. 2013). Unsurprisingly, proteins involved in cell-division, like CDK2, APC16 or CDC27, were also identified. The SCF and APC/C ubiquitin ligases are known important regulators of the cell cycle progression (Skaar and Pagano 2009; Fukushima et al. 2013).

The interaction of SKP1 with all three ZF-CxxC containing F-box proteins is intriguing. Although the biological significance of this association remains elusive, it is tempting to speculate that SKP1 is recruited to CGIs via its association with KDM2A and KDM2B and contributes to CGI function by an uncharacterized mechanism. One interesting possibility would be that SKP1 is recruited to CGIs to mediate ubiquitylation and subsequent degradation of other proteins binding to these genomic elements. One such example could be RING1B itself, whose activity and stability are known to be regulated by ubiquitylation (Ben-Saadon et al. 2006; de Bie et al. 2010).



SKP1 IP		
Name	Score	Cov [%]
SKP1	1747	76
KDM2B	874	45
BCOR	775	28
BCORL1	275	10
RING1B	539	49
PCGF1	196	54
RYBP	163	47
YAF2	30	26

SKP1 IP cont		
Name	Score	Cov [%]
KDM2A	1465	49
FBXL19	794	42
CULLIN1	924	48
CULLIN7	109	12
SKP2	368	31
FBXO38	1271	39
FBXO28	610	67
FBXO11	232	32
FBXO42	193	22
FBXW11	134	19
FBXO15	113	20
FBXO22	88	28
FBXW7	64	17
FBXO3	60	11
FBXL14	56	14
FBXO2	39	5
FBXL20	34	2.5
FBXO6	33	19
FBXL6	23	18
FBXO46	19	11
FBXL12	17	10

SKP1 IP cont		
Name	Score	Cov [%]
CKS1B	42	53
CDK2	40	23
APC16	21	12
CDC27	19	11
CDC16	18	2

Figure 4.8. SKP1 is an integral member of the KDM2B-PRC1 complex, in addition to functioning through other F-box proteins.

Following Western blot confirmation that the tagged protein bait was successfully pulled down, the immunoprecipitated SKP1 fraction was analysed by LC-MS/MS. The Mascot scores and peptide coverage are shown for the significant protein hits. SKP1 in mouse ES cells associates with the KDM2B-PRC1 variant complex (shown in grey), but also with additional F-box motif containing proteins and cullins to form SCF ubiquitin-ligase complexes. Interestingly, the two main SKP1 interaction partners (in addition to KDM2B) are the highly similar and CGI-binding proteins KDM2A and FBXL19, suggesting that SKP1 may potentially play an unexpected role in CGI function.

Although further work is clearly required to understand whether KDM2A and 2B form active SCF E3 ubiquitin ligase complexes and to define their potential substrates, these observations suggest that SKP1 may play a fundamental and pervasive role in regulation of CGI function. Furthermore, it also implies that the KDM2B-PRC1 variant complex could potentially have two ubiquitin ligase activities, one catalysing H2AK119 ubiquitylation via RING1B, and the other encoded by KDM2B-SKP1 towards an unexplored substrate. As KDM2B associates with SKP1 via the F-box domain, specifically disrupting this

interaction while leaving the rest of the KDM2B protein intact will help clarify the contribution of SKP1 to the PCGF1-PRC1 variant complex.

4.8. Summary and discussion

KDM2A and KDM2B recognize CGI elements via their ZF-CxxC DNA binding domain. However, in contrast to its paralogue, KDM2B is preferentially enriched at Polycomb-occupied CGIs in mouse ESCs. Work in this chapter has shown that KDM2B interacts with RING1B to form a variant PRC1 complex characterized by the PCGF1 subunit. The inclusion of KDM2B in a non-canonical PRC1 complex suggests that the distinct chromatin binding profile of KDM2B could be related to its association with Polycomb-proteins.

The KDM2B-PRC1 complex consists of RING1B, PCGF1, RYBP/YAF2 and BCOR/BCORL1. Mediated via its F-box motif, KDM2B also associates with SKP1, although this interaction is similar to the one formed by KDM2A. Importantly, the KDM2B-PRC1 complex does not include any CBX proteins. This distinguishes it from the canonical PRC1 complexes formed specifically by PCGF2 (MEL18) and PCGF4 (BMI1) (Figure 4.9). Thus, despite their extensive conservation at the amino acid level, the two KDM2s are associated with distinct proteins, with only KDM2B interacting with Polycomb-group components. Although the reasons behind this require further work, preliminary structural modelling studies in the Klose lab suggest that one exposed surface of the LRR domain is different between the two KDM2s, thus providing a unique binding platform for KDM2B to form the variant PRC1 complex (Dr. N.Rose, personal communication). Identifying the specific interaction domain between KDM2B and Polycomb-group components will be an invaluable tool to study the unique role of KDM2B at PcG-bound genes. The observation that the short form of KDM2B lacking the JmjC catalytic domain also associates with RING1B and PCGF1 suggests that the contribution of KDM2B to Polycomb-mediated repression may be unrelated to or extend beyond its histone demethylase activity.

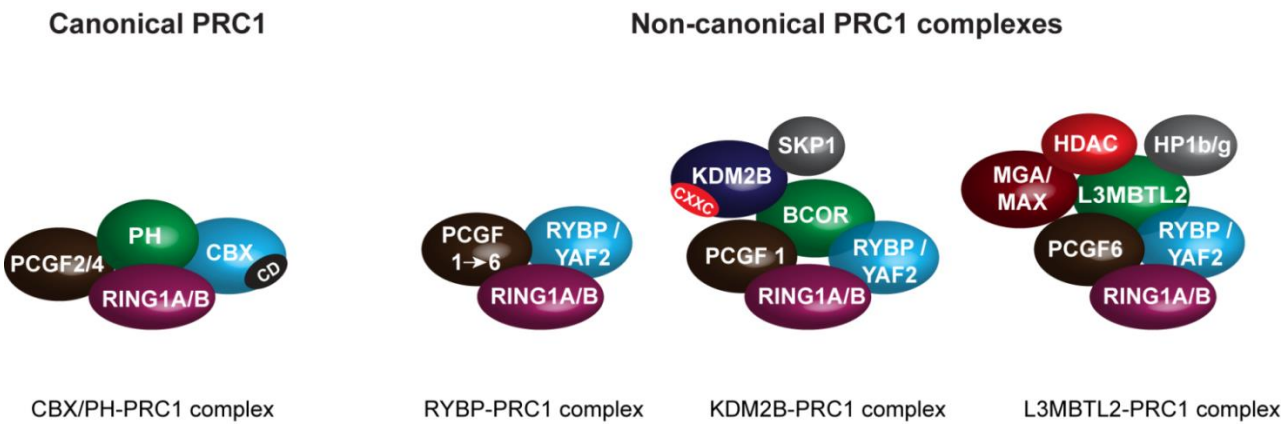


Figure 4.9. Canonical and non-canonical PRC1 complexes in mouse ESCs.

Canonical and non-canonical PRC1 complexes form around the catalytic subunit RING1A/1B, which mediates H2K119ub1, together with one of six PCGF subunits. Canonical PRC1 comprises the PCGF2 or PCGF4 protein, together with the chromodomain-contain (CD) CBX and polyhomeotic (PH) subunits. Non-canonical PRC1 complexes associate with RYBP or YAF2 instead of the CBX component. Two distinct variant PRC1 complexes are shown schematically, formed around the PCGF1 and PCGF6 subunits, respectively. While the mechanism by which the L3MBTL2-PRC1 complex binds chromatin remains elusive, the ZF-CxxC DNA binding domain of KDM2B may play a crucial role in recruiting RING1B to its target genes.

Canonical PRC1 recruitment to its targets was proposed to be mediated through interaction of the chromodomain in the CBX proteins with H3K27me3 placed by PRC2 (Cao et al. 2002; Fischle et al. 2003; Min et al. 2003; Lee et al. 2007b). Although the hierarchical PRC1 recruitment model is widely accepted, there is evidence suggesting that PRC1 targeting can also be achieved in a PRC2-independent manner, with RYBP-containing non-canonical PRC1 complexes mediating H2K119ub1 independently of H3K27me3 (Schoeftner et al. 2006; Tavares et al. 2012). However, nearly all PcG proteins lack inherent DNA binding activity, making the mechanisms by which PRC1 and PRC2 specifically localize to a subset of CGIs in mouse ESCs unclear. The discovery that KDM2B associates with RING1B and PCGF1 to form a non-canonical RYBP–containing complex prompts the question of whether the ZF-CxxC DNA binding domain of KDM2B could be important for PRC1 targeting to CGI elements. However, the PCGF1-PRC1 complex includes additional proteins with a reported capacity to specifically interpret the DNA sequence or its underlying chromatin status. Thus, BCOR and BCORL1 are associated with sequence-specific DNA-binding transcription factors, while RYBP was reported to be a ubiquitin-binding protein and interact with the transcription factor YY1 (Garcia et al. 1999; Arrigoni et al. 2006). It will therefore become

important to define the chromatin binding profile for BCOR and BCORL1, in order to determine whether they are restricted to binding only a specific set of target genes directed by their association with BCL6 or CTBP1, respectively. Alternatively, given their association with KDM2B, BCOR/BCORL1 would be expected to be recruited to CpG islands genome-wide. Interestingly, this possibility was supported by ChIP-Seq experiments in Diffuse-large B-cell lymphoma (DLBCL), which suggest that BCOR occupies only a minority of BCL6-targets, but is instead more widely associated with CpG island-promoters (Hatzi et al. 2013).

Although the contribution of these additional proteins from the PCGF1 variant complex to Polycomb function must be defined, the observation that KDM2B binds to CGIs genome-wide and colocalizes with RING1B on PcG-occupied CGIs suggests that the capacity of KDM2B to recognize non-methylated DNA may play a role in Polycomb-mediated repression. This is further supported by work in *Drosophila*, where the single KDM2B homologue dKDM2 retains a presumably functional ZF-CxxC DNA binding domain and its capacity to interact with dRING to form the major H2A ubiquitylating complex (Lagarou et al. 2008). Interestingly, alignment with its mammalian orthologues suggests that dKDM2 retains a potentially functional ZF-CxxC DNA binding domain. As flies lack significant levels of genomic DNA methylation and associated DNA methyltransferase homologs (Goll and Bestor 2005; Lee et al. 2010), the observation that dKDM2 could bind CpG dinucleotides is striking and remains to be investigated in the context of dRAF-mediated transcriptional regulation.

The KDM2B-PRC1 association represents one example from multiple non-canonical PRC1 complexes defined by distinct PCGF subunits. Although the interplay between these alternative Polycomb silencing activities is difficult to dissect, the DNA-specific binding activity of KDM2B could provide a simple and elegant link between recognition of non-methylated CpGs and nucleation of Polycomb-group proteins. At the same time, it suggests that, while the canonical PRC1 complexes are believed to be recruited to chromatin via the ability of CBX to recognize the H3K27me3 mark placed by PRC2, the non-canonical RYBP-containing PRC1 complexes evolved alternative targeting strategies. Defining the interplay between the diverse PRC1 complexes will be essential to understanding how Polycomb-mediated gene regulation is achieved.

5. Chapter five - KDM2B targets RING1B to CGIs genome-wide

ChIP-Seq chromatin binding profiles show that KDM2A and KDM2B bind CpG island elements genome-wide in mouse ESCs. However, despite their similar binding profile at the majority of CGI-associated genes, KDM2B shows preferential enrichment at CGIs occupied by the Polycomb repressive complexes. Furthermore, detailed affinity purification work in mouse ES cells indicated that KDM2B exists as part of a variant PRC1 complex including RING1B, PCGF1, RYBP/YAF2 and BCOR/BCORL1. The association of KDM2B with Polycomb-group components and its co-localization with RING1B at PcG-targets suggested that KDM2B could play a role in Polycomb-mediated repression.

Although in flies PcG proteins repress their target genes by binding to specific DNA regions called Polycomb Response Elements (PREs), it is still not clear how PcG complexes are specifically recruited to their target sites in mammals. Interestingly, genome-wide analysis indicated that PcG occupancy precisely correlates with the CpG islands of target genes (Ku et al. 2008). Together with work demonstrating that artificial GC-rich sequences devoid of transcriptionally activating signals have the inherent capacity to recruit PRC2 by default when introduced into mammalian ES cells, these observations suggest that CGIs may act as the mammalian counterpart to fly PREs (Mendenhall et al. 2010). However, a clear mechanistic model explaining the function of CpG islands in PcG complex recruitment is still undefined.

Extrapolating from flies, PRC2 is believed in mammals to be recruited to its target genes through a direct association with sequence-specific transcription factors. PRC2-mediated deposition of H3K27me3 is then recognized by the chromodomain in the PRC1 protein CBX, leading to hierarchical PRC1 recruitment. Recent discovery of non-CBX containing PRC1 variant complexes, together with studies showing PRC1/H2AK119ub1 targeting independently of H3K27me3 suggest that the hierarchical recruitment model cannot account for all PRC targeting and prompts a reevaluation of Polycomb recruitment

mechanisms in mammals (Schoeftner et al. 2006; Gao et al. 2012; Tavares et al. 2012). Binding of PRC1 to target loci is believed to be central in establishing transcriptional repression, although the mechanisms by which this is achieved remain unclear. Proposed mechanisms by which PRC1 mediates gene repression include H2AK119 ubiquitylation-dependent block to transcriptional initiation (Lehmann et al. 2012), inhibition of RNA polymerase II transcriptional elongation (Stock et al. 2007) and chromatin compaction (Francis et al. 2004; Grau et al. 2011). Interestingly, the compact chromatin nature of Polycomb-repressed regions has been suggested to render Polycomb-protein occupied chromatin refractory to binding of ZF-CxxC domain-containing proteins like KDM2A and CFP1 (Blackledge et al. 2010; Thomson et al. 2010).

However, unlike other ZF-CxxC containing proteins such as KDM2A and CFP1, KDM2B is preferentially enriched at Polycomb-occupied CGIs. This distinct chromatin binding profile suggests that KDM2B could be enriched at these regions in a ZF-CxxC independent manner, relying instead on its interaction with Polycomb-components (Figure 5.1, panel A). Interestingly, recent *in vitro* studies show that H3K36 methylation antagonizes PRC2-mediated H3K27me3 (Schmitges et al. 2011; Yuan et al. 2011). Therefore, it is tempting to speculate that the PCGF1-PRC1 complex could specifically recruit KDM2B to Polycomb-targets to remove H3K36me2 and allow H3K27 methylation.

Alternatively, KDM2B could function as a targeting module for PRC1, contributing to PcG-mediated repression in a manner dependent on its ZF-CxxC DNA binding domain (Figure 5.1, panel B). Despite the observation that PcG protein occupancy closely mirrors non-methylated CpG density, it is still unclear how Polycomb group proteins are recruited to their target loci. KDM2B could therefore provide an unexpected and direct link between recognition of non-methylated DNA and binding of PRC1 complex to its target genes in mouse ES cells. However, only a relatively small subset of CGIs are occupied by Polycomb proteins, while KDM2B is present at virtually all CGIs through its ZF-CxxC dependent recognition of non-methylated DNA, suggesting that PRC1 occupancy may not be determined by KDM2B alone.

Work in this chapter aims to understand the relationship between KDM2B and RING1B at PcG CGIs by specifically addressing the two models proposed (Figure 5.1). The resulting findings suggest that the

association of KDM2B with Polycomb-group components is essential for proper regulation of gene expression and PRC1 complex activity in mouse embryonic stem cells.

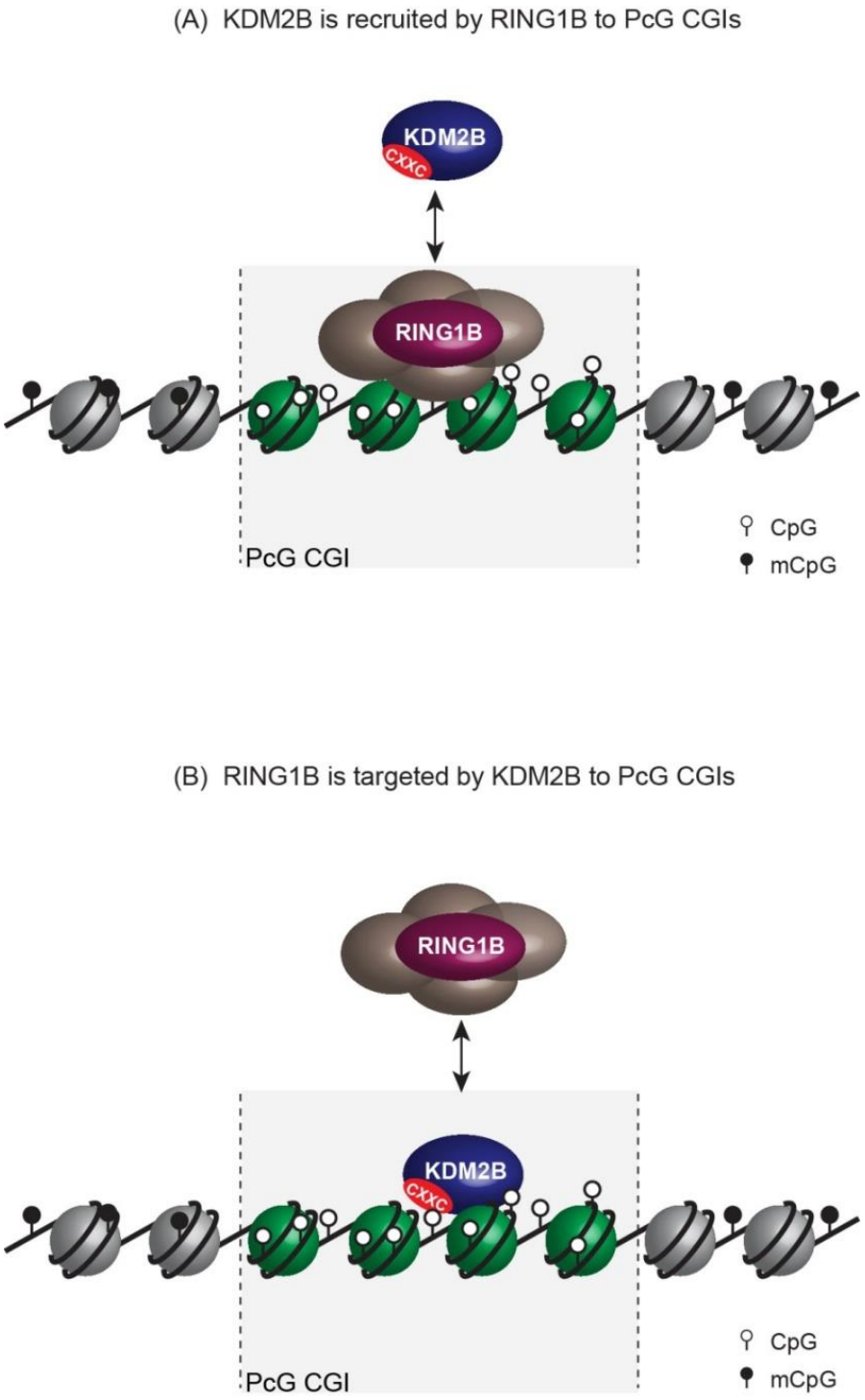


Figure 5.1. Alternative models for the relationship between KDM2B and RING1B at Polycomb-occupied CGIs.

(A) KDM2B is enriched at PcG-repressed CGIs independently of its ZF-CxxC DNA binding domain, but instead relies on RING1B.

(B) Alternatively, the capacity of KDM2B to recognize non-methylated CpGs is responsible for the recruitment of RING1B to PcG CGIs.

5.1. KDM2B relies on its ZF-CxxC DNA binding domain for localization to CGIs

5.1.1. KDM2B binds to PcG CGIs in the absence of an intact PRC1 complex

To begin addressing whether the preferential enrichment of KDM2B at PcG CGIs relies on its association with PRC1 components, Ring1A$^{-/-}$;Ring1B$^{fl/fl}$; Rosa26::CreERT2 mouse ESCs were obtained (Endoh et al. 2008). In these cells, RING1A is constitutively inactivated, while both alleles of RING1B have exon 2 encoding the ATG start codon flanked by loxP sites. For conditional deletion of RING1B, Ring1A$^{-/-}$;Ring1B$^{fl/fl}$ mice were crossed with Rosa26::CreERT2 transgenic mice, in which the Cre recombinase fused to a mutated ligand-binding domain for the human estrogen receptor (ERT) is inserted into the ubiquitously expressed Rosa locus. Ring1A$^{-/-}$;Ring1B$^{fl/fl}$; Rosa26::CreERT2 mouse ESCs were derived from blastocysts. Upon addition of 4-hydroxy - tamoxifen (OHT) (an estrogen receptor antagonist), the Cre-ERT construct migrates to the cell nucleus and catalyzes loxP-mediated recombination at Ring1b locus leading to loss of RING1B transcription.

RING1B protein levels are dramatically depleted 48 hours after tamoxifen (OHT) treatment, also leading to reduced levels of other PRC1 components like PCGF2, PH and CBX proteins (Endoh et al. 2008). Importantly, the RNA and protein levels of KDM2B are not affected following RING1A/RING1B deletion (data not shown). In the absence of RING1A/RING1B, the chromatin binding profile of KDM2B was investigated by ChIP-qPCR at a panel of non-CGI, non-Polycomb and Polycomb-target CGI associated genes. The chromatin binding profile of KDM2A was also investigated as a control for a ZF-CxxC DNA binding protein which does not associate with Polycomb-group components.

ChIP-qPCR analysis indicated that RING1B binding at Polycomb target CGIs is dramatically lost 2 days following tamoxifen treatment, in agreement with the reported depletion of RING1B (Figure 5.2, panel B). However, when binding of KDM2B was analysed, there was no significant change in KDM2B levels at Polycomb-associated CGIs following tamoxifen treatment (Figure 5.2, panel C).

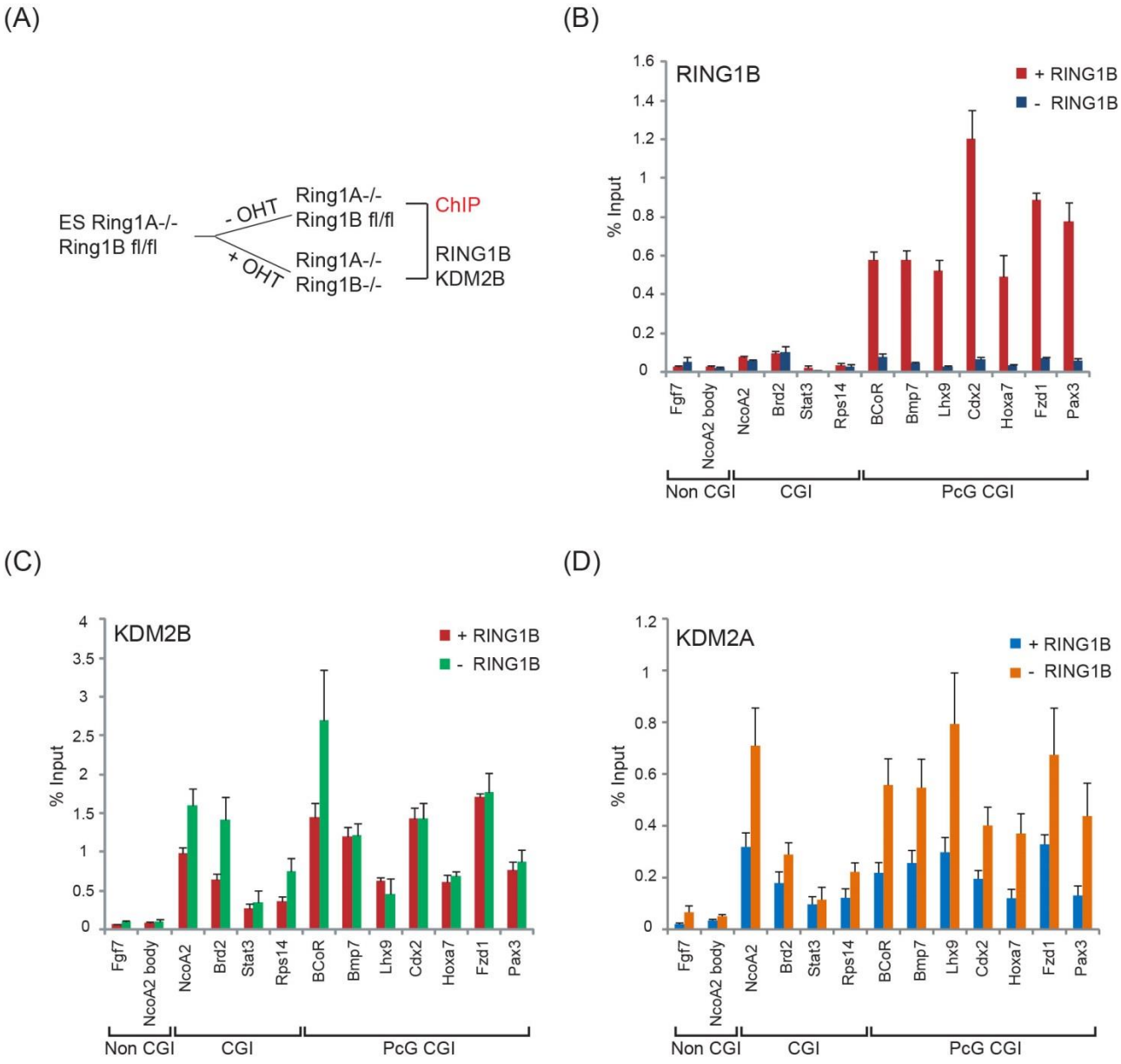


Figure 5.2. KDM2B binds to PcG CGIs in the absence of an intact PRC1 complex.

(A) Schematic representation showing removal of RING1B in ESC Ring1A−/− Ring1Bfl/fl cells by tamoxifen (OHT) treatment.

(B) ChIP analysis indicating that RING1B is enriched at polycomb target CGIs in untreated cells (red bars). After 48 hr of tamoxifen treatment RING1B binding is lost at polycomb targets (blue bars). Error bars represent SEM of three biological replicates.

(C) ChIP analysis demonstrating that removal of RING1B does not lead to loss of KDM2B binding at regular or polycomb associated target CGIs (compare red and green bars). This demonstrates that RING1B is not required to recruit KDM2B to polycomb repressed sites. Error bars represent SEM of three biological replicates.

(D) ChIP analysis showing an increase in KDM2A occupancy at regular and PcG CGIs after removal of RING1B (compare blue and orange bars). Error bars represent SEM of three biological replicates.

An exception to this was the BCOR PcG-occupied gene, which showed an increase in KDM2B binding upon PRC1 disruption for reasons which are not clear. Interestingly, ChIP-qPCR analysis showed an increase in KDM2A occupancy at PcG-target CGIs after removal of RING1B (Figure 5.2, panel D). This observation is in agreement with previous work which suggests that, potentially due to their compact chromatin architecture (Francis et al. 2004; Simon and Kingston 2009), PcG-occupied CGIs are refractory to binding of ZF-CxxC domain proteins. Therefore, disruption of Polycomb binding could result in a more accessible chromatin state at targets normally occupied by Polycomb proteins, potentially making binding of ZF-CxxC proteins more likely. Intriguingly, both KDM2s show increased chromatin occupancy upon RING1B removal at regular CGIs (Figure 5.2, compare panels C and D). This occurs in the absence of increased KDM2A and KDM2B protein levels (data not shown). Although this result requires further clarification, one possible explanation for the increased KDM2A and 2B binding at regular CGIs following RING1A/1B deletion will be discussed later in this chapter.

Nevertheless, the observation that removal of RING1A/RING1B does not disrupt the binding of KDM2B to Polycomb-bound CGIs suggests that KDM2B is not targeted to these sites by its association with the PRC1 complex.

5.1.2. Disruption of its ZF-CxxC domain impairs KDM2B binding to CGIs

As removal of RING1A/RING1B does not impair KDM2B binding to regular or Polycomb-associated CGIs, the next step was to test whether the ZF-CxxC domain is the principal DNA targeting activity for KDM2B. To achieve this, stable mouse ESCs were generated expressing Flag/2xStrep II (FS2)-tagged versions of wild-type KDM2B or a ZF-CxxC domain mutant KDM2B which lacks the ability to bind non-methylated DNA (Figure 5.3, panel A). The single amino acid point mutation (K643A) is part of the KQ-motif in the linker region of the ZF-CxxC DNA binding domain and is believed to be essential to forming specific side-chain and backbone interactions with the double-stranded CpG dinucleotide sequence (Cierpicki et al. 2010). A corresponding mutation was used previously to disrupt the DNA binding activity of KDM2A

(Blackledge et al. 2010). Although not conclusively demonstrated, this mutation is expected to disrupt the DNA binding capacity of KDM2B without affecting its native protein structure. An empty vector cell line was also generated and used as control for the ChIP-qPCR experiments. The enrichment levels of epitope-tagged WT or ZF-CxxCmut KDM2B were analysed at a panel of regular and Polycomb-repressed CGIs using a rabbit polyclonal antibody specifically recognizing the FS2 tag which was generated and characterized during this study (refer to Supplementary figure S2).

Importantly, ChIP-qPCR analysis shows that disrupting the ZF-CxxC DNA binding domain leads to loss of KDM2B binding to both regular and Polycomb-repressed CGIs (Figure 5.3, panel B). Therefore, the capacity of KDM2B to recognize non-methylated CpGs is essential for its chromatin binding.

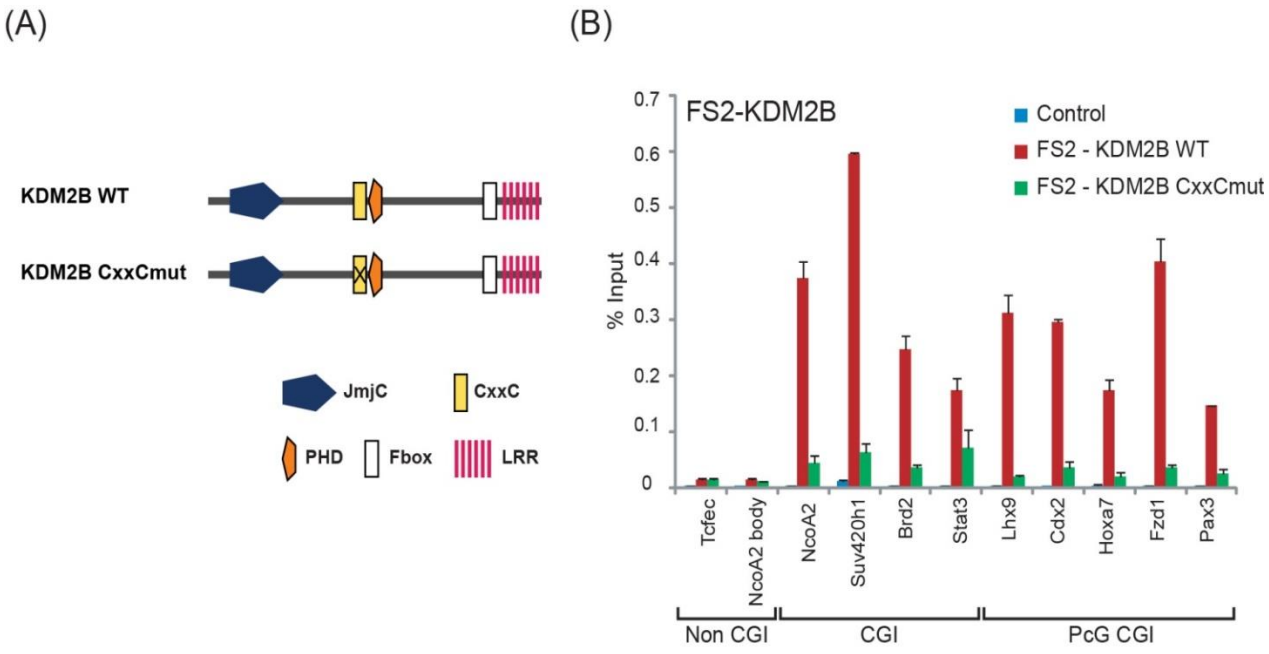


Figure 5.3. KDM2B requires a functional ZF-CxxC domain to bind CGIs.

(A) Schematic representation of the constructs used in this experiment, depicting wild-type KDM2B (WT) and KDM2B with a mutation in its ZF-CxxC DNA domain shown to abrogate its capacity to bind non-methylated CpGs (CxxCmut).

(B) KDM2B ChIP-qPCR analysis in ES cells stably expressing tagged wild-type (WT) KDM2B or a mutant KDM2B which disrupts its DNA-binding capacity (CxxCmut).The ZF-CxxC domain of KDM2B is required for KDM2B binding to CGIs regardless of their polycomb occupancy. Error bars represent SEM of two biological replicates.

Interestingly, KDM2B binding at some CGIs is higher than background after mutation of the ZF-CxxC DNA binding domain (i.e. *Suv420h1 or Stat3,* Figure 5.3, panel B). Although not clear, one possible explanation could be that the single amino acid mutation does not fully abolish the DNA binding activity of KDM2B. Alternatively, the presence of endogenous wild-type KDM2B could facilitate retention of the ZF-CxxC mutant KDM2B by an unknown mechanism.

Although this result suggests that there is no significant ZF-CxxC independent recruitment of KDM2B at PcG targets, an alternative possibility which cannot be excluded in this type of experiment is that RING1B could play a role in recruiting KDM2B to PcG CGIs, but that the ZF-CxxC domain of KDM2B is then essential for its retention at those loci. An additional possibility for lack of PRC1-dependent KDM2B targeting in the DNA binding mutant could be that an intact ZF-CxxC domain is actually required for its association with PRC1 components. However, this possibility was tested by checking the interaction partners in mouse ESCs of KDM2B WT and ZF-CxxCmut using a similar experimental setup as described in the previous chapter (refer to section 4.2). Briefly, LC-MS/MS analysis indicated that the DNA-binding domain mutant KDM2B retained the capacity to form the variant PRC1 complex (supplementary figure S3).

These results indicate that the ability of KDM2B to recognize and bind non-methylated CpGs is the main determinant required for chromatin binding, despite its association with Polycomb-group components. Therefore, if RING1B does not play a major role in recruiting KDM2B to PcG CGIs, the alternative possibility is that KDM2B functions as a targeting module for PRC1.

5.2. Depletion of KDM2B reduces RING1B occupancy at Polycomb-targets genome-wide

To investigate whether KDM2B contributes to Polycomb group protein occupancy on chromatin, the expression of KDM2B in mouse ESCs was first down-regulated using specific short hairpin RNAs (shRNAs). The constitutive knockdown approach was based on the pLKO.1 puro lentiviral expression vector (Addgene plasmid #10878 and (Moffat et al. 2006)). The vector carries the puromycin-resistance

gene and drives shRNA expression from a human U6 promoter. pLKO.1puro constructs expressing either KDM2B or scrambled control shRNA were co-transfected together with a packaging plasmid (psPAX2) (Addgene #12260) and an envelope plasmid (pMD2.G) (Addgene #12259) into 293T cells to allow production of lentiviral particles. Mouse ESCs were subsequently infected with lentivirus expressing the shRNA of interest.

Importantly, the KDM2B gene has an alternative transcription start site downstream of its JmjC catalytic domain, giving rise to a short form of the protein with a presumably functional ZF-CxxC DNA binding domain. Work in the previous chapter shows that this short isoform of KDM2B retains the capacity to associate with Polycomb proteins and form the variant PRC1 complex, similar to full length KDM2B (see section 4.5). Therefore, to study the contribution of KDM2B to Polycomb-mediated repression, shRNAs targeting both the full length and the short form of KDM2B were designed. Following lentiviral infection and puromycin selection, mouse ES clones stably expressing either the scrambled or KDM2B shRNAs were selected and expanded. The efficiency of KDM2B knockdown was tested by qRT-PCR and Western blot analysis.

In one KDM2B knockdown clone, approximately 80% knockdown of KDM2B was achieved at the RNA level as assessed by qRT-PCR (Figure 5.4, panel A), and clear depletion of the protein was detected when assessed by Western blot analysis (Figure 5.4, panel B). Importantly, KDM2A RNA and protein levels do not change, confirming the specificity of the KDM2B-targeting shRNA. Depletion of proteins involved in Polycomb-mediated repression can in some cases lead to destabilization of PRC1 or PRC2 core components (Pasini et al. 2007; Leeb et al. 2010; Tavares et al. 2012). However, knockdown of KDM2B does not affect the stability of RING1B or EZH2/EED proteins, permitting an examination of the effect that KDM2B depletion has on PRC1 function.

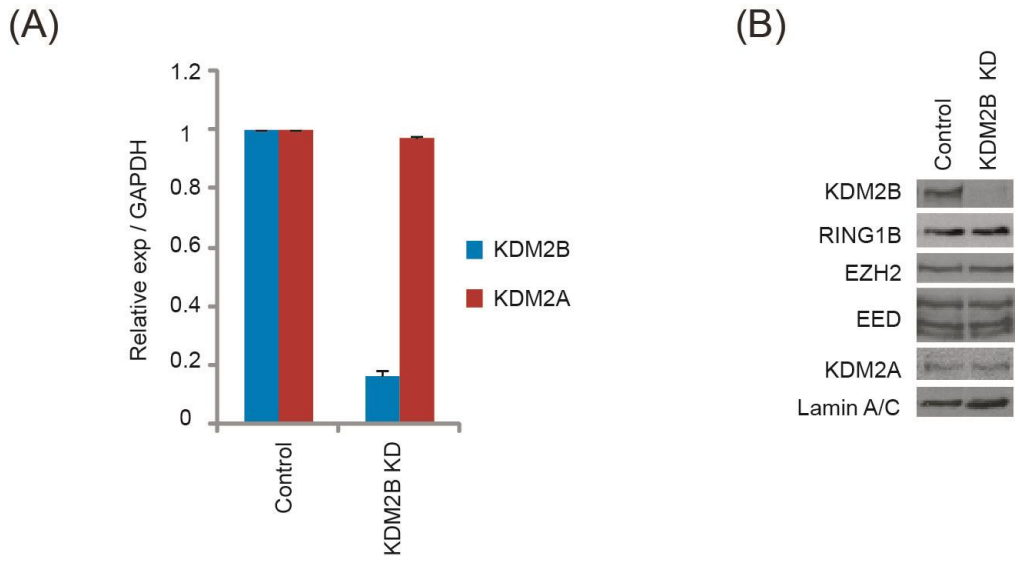


Figure 5.4. shRNA-based knockdown of KDM2B in mouse ESCs.

(**A**) qRT-PCR showing reduction of KDM2B at the mRNA level in the knockdown affecting both FL and SF KDM2B (KD) but not control scrambled cell line. The levels of the closely related KDM2A mRNA were not affected. Error bars represent SEM of three biological replicates.

(**B**) Western blot analysis showing that KDM2B knockdown depletes KDM2B but not the closely related KDM2A protein. Importantly depletion of KDM2B does not destabilize other polycomb group proteins including RING1B, EZH2 or EED. Lamin A/C indicates equal loading. The protein levels of the PRC1 and PRC2 proteins were analysed by a summer student, Mr. Thomas Sheahan, under my supervision.

To understand whether KDM2B plays a role in targeting RING1B to PcG CGIs via its capacity to recognized non-methylated DNA, ChIP-Seq experiments to assess KDM2B and RING1B occupancy were carried out in a scrambled control and KDM2B KD cell line. A matching control sample generated from input DNA was analysed in the same way to account for potential sequencing biases. In agreement with the clear reduction of KDM2B mRNA and protein levels (Figure 5.4), KDM2B occupancy was significantly reduced at CpG island elements genome-wide in the KDM2B KD compared to the control cell line (Figure 5.5, panel A). Interestingly, when plotting the RING1B ChIP-Seq signal over all PcG-associated CGIs, a clear and reproducible reduction in RING1B occupancy was observed across these sites after KDM2B depletion (Figure 5.5, panel B). The reduction in RING1B binding in the KDM2B KD cell line was evident when ChIP-Seq signals were plotted at individual PcG-associated genes, with *Nxph4* and *Bahcc1* PcG CGI-associated loci shown as examples (Figure 5.5, panels C and D).

Given the capacity of KDM2B to recognize non-methylated CGI DNA, its preferential enrichment at PcG-occupied CGIs and its association with RING1B and PCGF1 to form a variant PRC1 complex, it was hypothesized that KDM2B may contribute to PRC1 recruitment to PcG targets. The reduction in RING1B occupancy at these sites as a result of KDM2B shRNA-based depletion suggests that this is indeed the case. Furthermore, as depletion of KDM2B does not affect the protein levels of Polycomb-group proteins (Figure 5.4), the reduction in RING1B occupancy may directly result from a failure of KDM2B targeting it to PcG-occupied CGIs genome-wide.

The discovery that RING1B is retained at the majority of PcG targets in the absence of H3K27me3 (Pasini et al. 2007; Leeb et al. 2010; Tavares et al. 2012) suggested the existence of alternative recruitment strategies for the non-CBX containing variant PRC1 complexes. As PRCs do not contain any obvious inherent DNA-binding specificity, additional factors with which they associate, or which create a favourable chromatin environment, must be responsible for their specific chromatin recruitment. Multiple mechanisms have been suggested, such as recruitment of PRC1 directly to chromatin in a PRC2-independent manner by its association with the core binding transcription factor RUNX1/CBFβ (Yu et al. 2012), REST transcriptional repressor (Dietrich et al. 2012) or association with non-coding RNAs (Yap et al. 2010). However, the mechanisms proposed are valid only for locus-specific circumstances and cannot account for all PRC targeting. Therefore, the observation that depletion of KDM2B leads to reduction in RING1B occupancy across the majority of PcG-bound CGIs suggested that KDM2B may function as a PRC1 targeting factor that does not rely on recognition of H3K27me3 but instead interrogates the underlying non-methylated DNA content at PcG target genes.

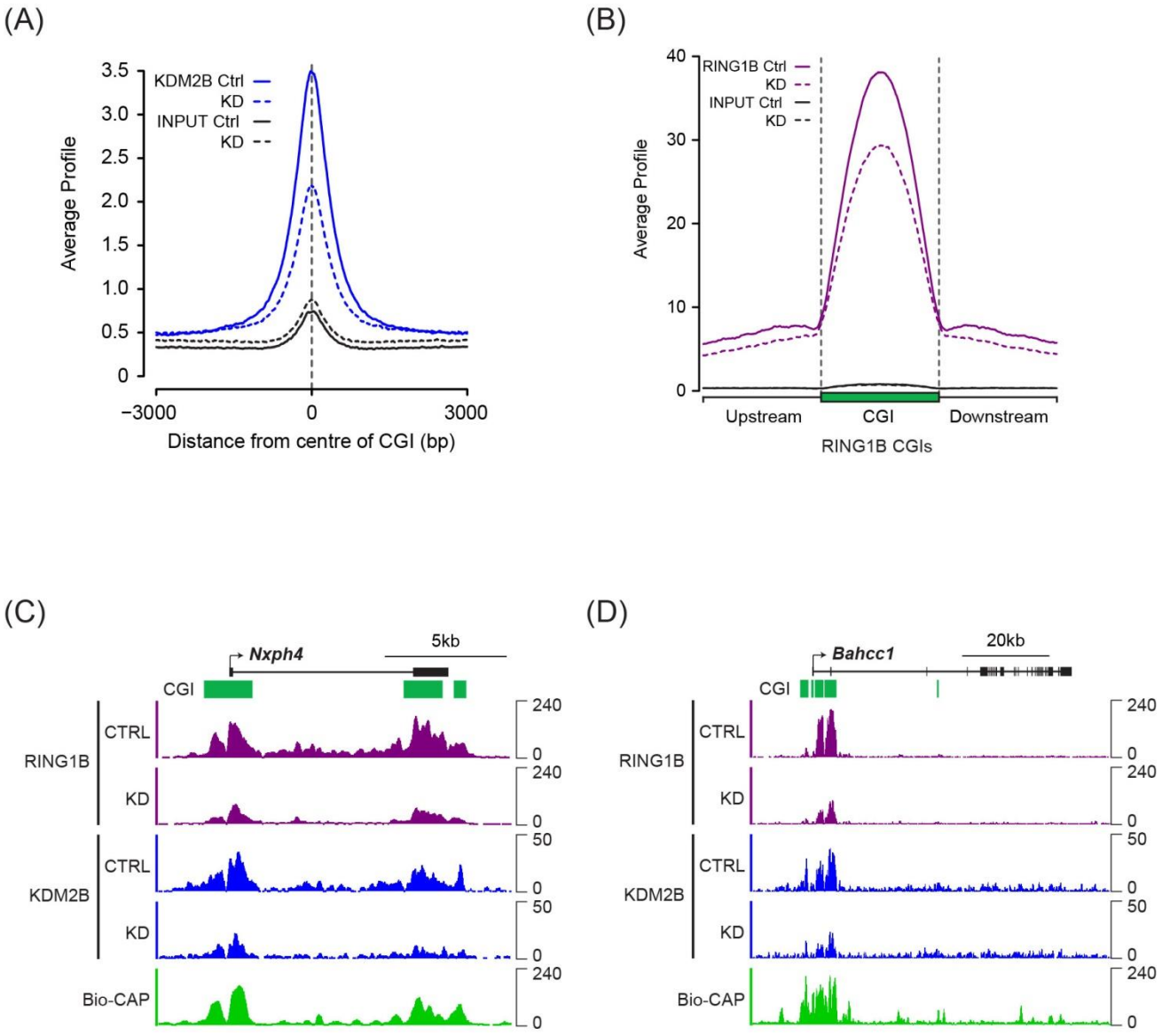


Figure 5.5. KDM2B depletion leads to reduction in RING1B occupancy at PcG CGIs.

(A) KDM2B ChIP-seq signal was plotted at CGIs in the scrambled control (Ctrl, solid blue line) and knockdown cell line (KD, dotted blue line). Sequencing signals in the input samples over the same regions are indicated as solid or dotted black lines. KDM2B ChIP-seq signal is specifically lost over CGIs genome-wide in the KDM2B knockdown cell line.

(B) RING1B ChIP-seq signal from the scrambled control (solid purple line) and KDM2B knockdown cell line (dotted purple line) were plotted over previously identified CGI associated RING1B binding sites. Input sequencing signal was plotted over the same regions. In the KDM2B knockdown line there is a specific reduction of RING1B binding over the CGI.

(C) - (D) Two examples of ChIP-seq profiles at PcG CGIs, showing reduction of RING1B following KDM2B knockdown. Bio-CAP-seq profiles are shown to indicate the location of non-methylated DNA (Long et al. 2013b).

5.3. KDM2B targets RING1B to CGIs genome-wide

Knockdown of KDM2B caused a reduction in RING1B binding at Polycomb targets (Figure 5.5). However, via its ZF-CxxC DNA binding domain, KDM2B binds to the majority of CpG island elements genome-wide (section 3.5). However, previous ChIP-Seq work has shown that PRC1 and PRC2 co-occupy only a subset of CGIs (Ku et al. 2008). Therefore, if KDM2B is indeed an essential targeting factor for PRC1 binding, one would expect KDM2B to recruit RING1B to all CGIs, not just Polycomb targets (Figure 5.6).

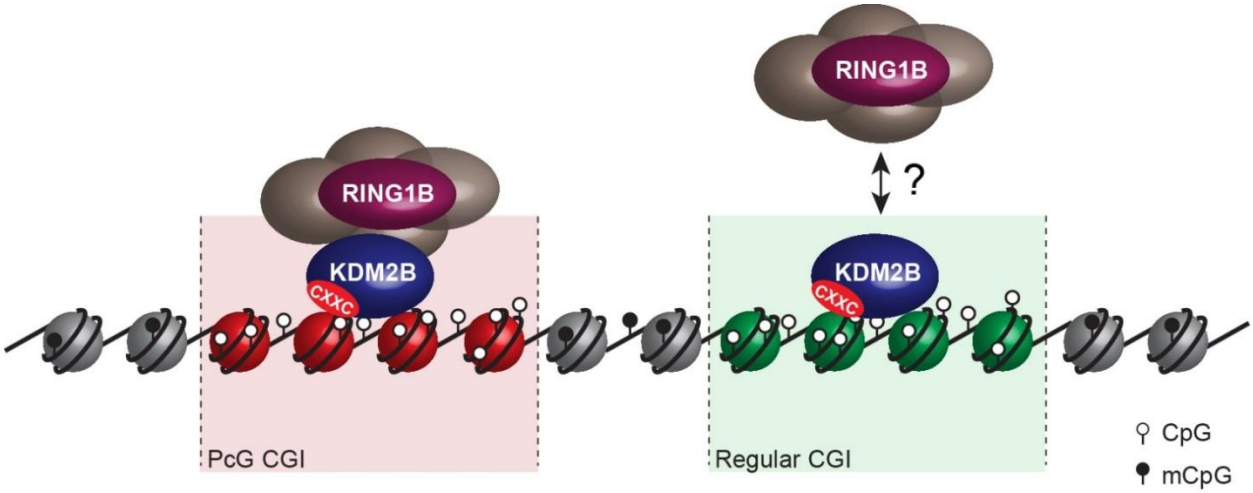


Figure 5.6. If PRC1 targeting is dependent on KDM2B, one would expect RING1B to localize to CGIs genome-wide.

Interestingly, a closer analysis of the RING1B ChIP-Seq profile in the scrambled control line revealed that the majority of regular non-Polycomb target CGIs in mouse ESCs show low magnitude but appreciable enrichment of RING1B. These novel low magnitude RING1B peaks are in addition to the classical high-magnitude peaks detected at established PcG targets, such as the *Atf3* gene shown as representative (Figure 5.7, panel A). The reasons for why these lower magnitude RING1B peaks at regular CGIs were previously unreported are unclear. The RING1B ChIP-Seq protocol used in this study was highly optimized, taking advantage of an affinity purified mouse monoclonal antibody against RING1B (described in (Atsuta et al. 2001)). Furthermore, the chromatin immunoprecipitation step was also

improved, using double crosslinking with the longer-arm protein-protein crosslinking agent EGS in addition to formaldehyde (as described in Materials and methods, section 2.6). Importantly, a closer analysis of available ChIP-Seq tracks for RING1B publicly available on the Gene Expression Omnibus (GEO) database suggests that small but appreciable levels of RING1B at non-PcG target CGIs are detectable (**GEO datasets** GSM656523 from R. Young lab; GSM850474 from A. Pombo lab and GSM585229 from N. Brockdorff lab, supplementary figure S4).

Importantly, together with the already characterized RING1B PcG CGIs, ChIP-Seq analysis indicated that clear RING1B peaks are detectable at approximately 70% of the 22,849 mouse CGIs genome-wide (CGI dataset used was from (Illingworth et al. 2010)). A minority of these novel RING1B occupied CGI sites showed detectable EZH2 (23%) (Peng et al. 2009) or H3K27me3 (20%) (Li et al. 2010), indicating that these sites are largely devoid of appreciable PRC2.

To understand the contribution of KDM2B to the binding of RING1B at these novel low magnitude sites, the RING1B ChIP-Seq signal over these low-magnitude regions was analysed in the scrambled control and KDM2B KD cell lines. Interestingly, this revealed a specific reduction in RING1B occupancy at these sites following KDM2B depletion (Figure 5.7, panel B). The reduction in RING1B binding in the KDM2B KD cell line was evident when ChIP-Seq signals were plotted at individual CGIs, with *Fah* and *Clip1* CGI-associated loci shown as examples of novel low-magnitude RING1B binding events (Figure 5.7, panels C and D). Importantly, the reduction in RING1B binding at these novel low-magnitude sites was even more dramatic than the loss of RING1B occupancy at PcG CGIs upon KDM2B knockdown (compare to Figure 5.5, panel B). This suggests that RING1B binding to these low-magnitude regions is even more dependent on the binding of KDM2B to these CGIs.

The unexpected discovery that there are detectable peaks of RING1B at nearly 70% of CGIs genome-wide suggested that KDM2B plays a pervasive role in PRC1 recruitment to CGIs. Nevertheless, it was essential to verify that these novel low-magnitude RING1B peaks are indeed real binding events and not a potential artefact of the ChIP-Seq technology or lack of antibody specificity.

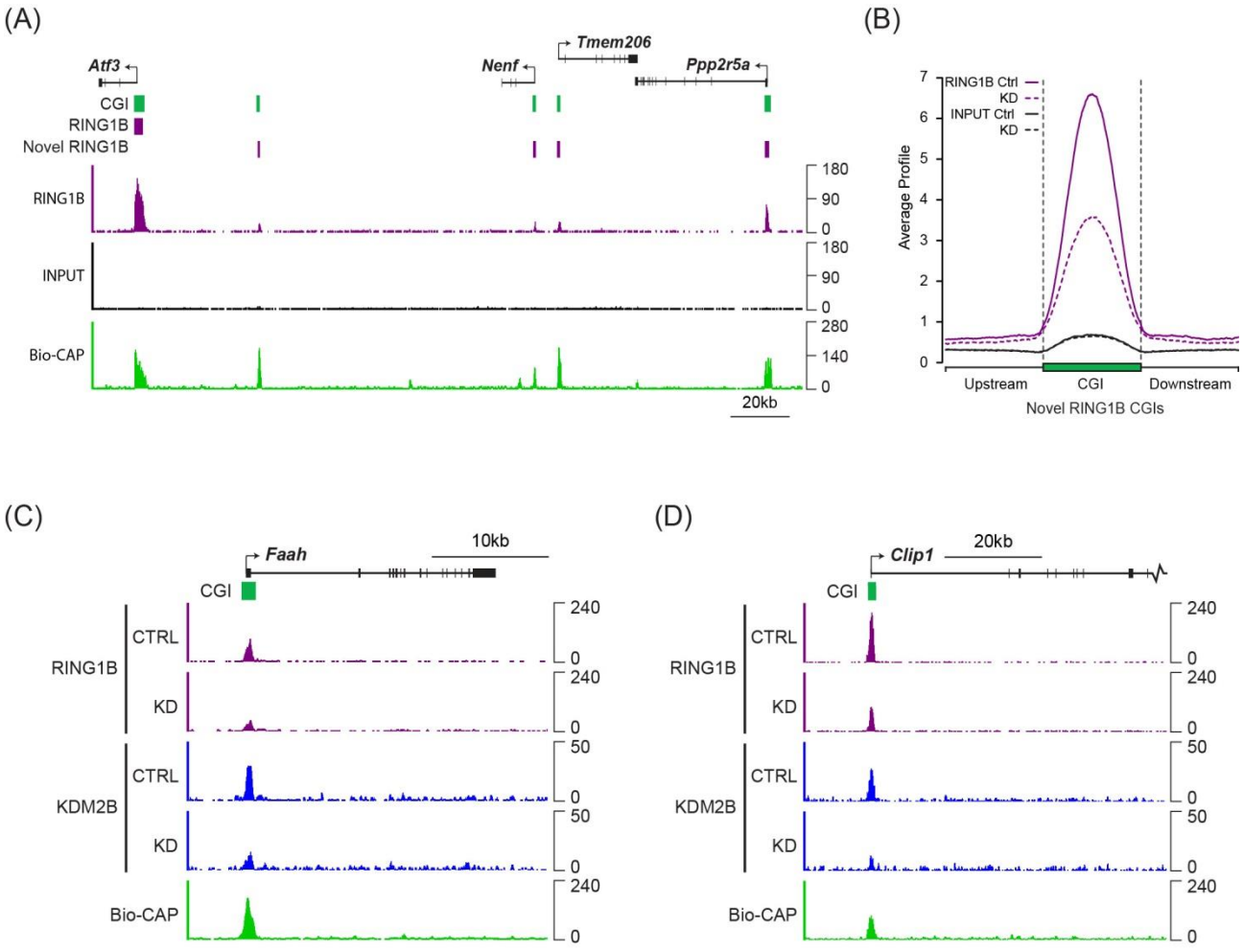


Figure 5.7. KDM2B targets RING1B to CGIs genome-wide.

(A) A representative region showing RING1B ChIP-seq signal in the scrambled control cell line illustrating previously identified high magnitude RING1B binding sites (*Atf3*) and novel low magnitude RING1B binding sites. Input sequencing traces over the same region are shown in black. Bio-CAP-seq signal indicates regions containing non-methylated DNA. Above the sequencing traces gene promoters are show by black arrows and exons by vertical black lines. CGIs are shown as green bars with previously identified RING1B peaks and novel RING1B peaks indicated with purple boxes.

(B) RING1B ChIP-Seq signal from the scrambled control and KDM2B KD cell line was plotted over the identified novel low magnitude binding sites. In the KDM2B knockdown cell line there is a clear loss of RING1B binding at these novel RING1B occupied sites.

(C) - (D) Two examples of ChIP-seq profiles at individual low magnitude RING1B binding sites showing clear reduction of RING1B following KDM2B knockdown. Bio-CAP-seq signal indicates regions containing non-methylated DNA.

To this end, ChIP-qPCR was carried out at a series of novel RING1B sites in cells where RING1A is constitutively inactivated and RING1B can be conditionally ablated by tamoxifen-inducible recombination (Endoh et al. 2008). Importantly, RING1B tamoxifen-mediated deletion resulted in a clear ChIP-qPCR loss of the RING1B signal at all eight genes tested (Figure 5.8), suggesting that these novel low-magnitude RING1B peaks detected in ChIP-Seq are real binding events.

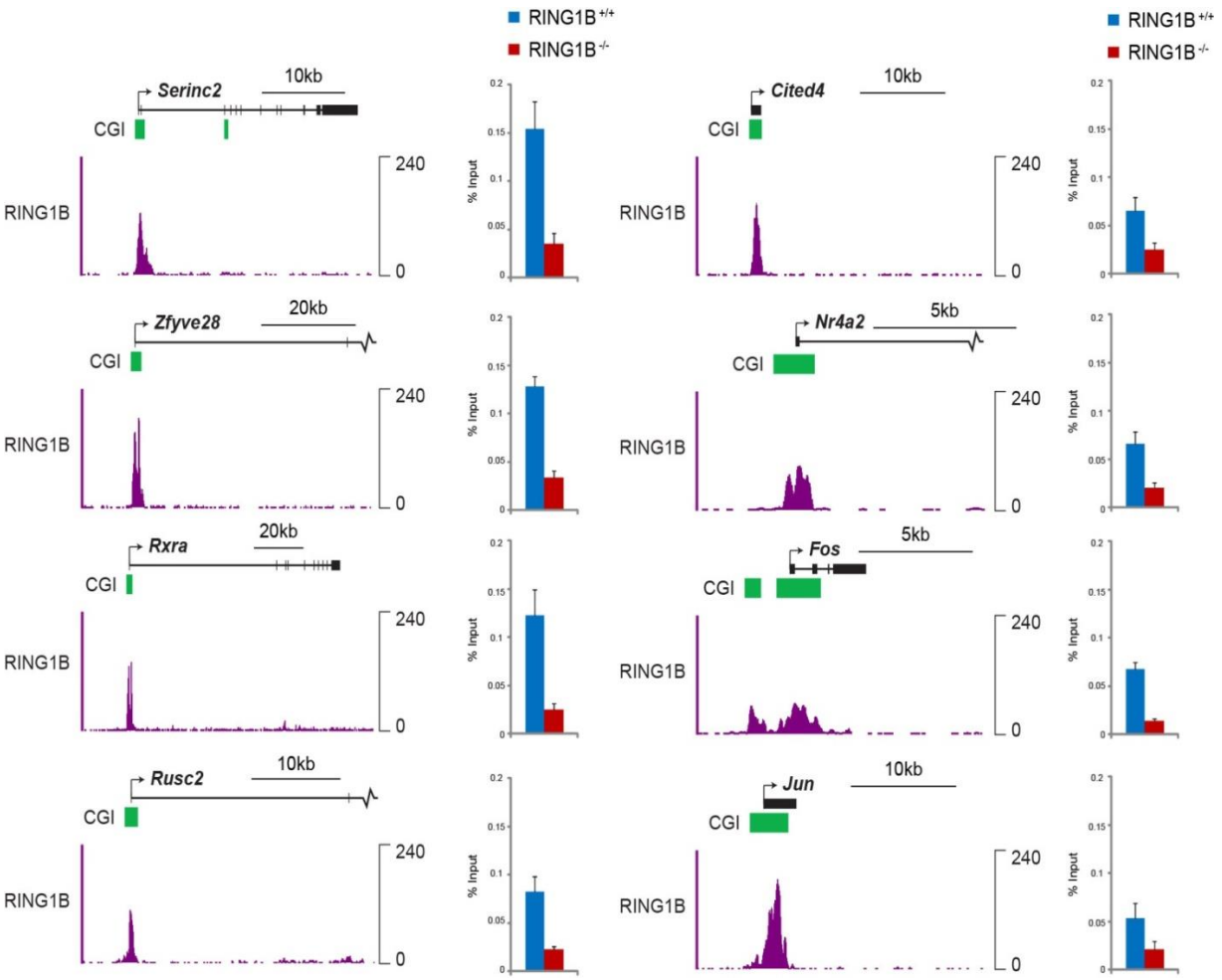


Figure 5.8. Validation of novel low-magnitude RING1B binding events in conditional RING1B knockout cells.

ChIP-qPCR analysis in RING1A$^{-/-,}$ RING1B$^{fl:fl}$ ES cells (Endoh et al. 2008) of RING1B occupancy at a panel of eight genes corresponding to novel RING1B peaks identified in ChIP-Seq in the scrambled control cell line. Upon tamoxifen-induced conditional RING1B deletion, there is a clear loss of RING1B binding levels, suggesting that these novel RING1B peaks identified correspond to real RING1B binding events. ChIP-seq traces detected in the RING1B ChIP-Seq in the scrambled control cell line are shown in each case for comparison. Error bars indicate the SEM for three biological replicates.

Furthermore, ChIP-Seq experiments in the RING1A$^{-/-,}$ RING1B$^{fl:fl}$ cells described above (Endoh et al. 2008) were carried out to confirm that RING1B binding events are lost following tamoxifen-mediated RING1B deletion genome-wide (Figure 5.9). In addition, RING1B genome-wide chromatin binding profile was analysed by ChIP-Seq using a commercial rabbit monoclonal antibody specific for RING1B (Cell Signalling, D22F2). RING1B ChIP-Seq chromatin binding profile generated using this antibody was indistinguishable from the one generated using the mouse monoclonal antibody described previously (Figure 5.9), suggesting that the widespread association of RING1B with CGIs is a reproducible feature.

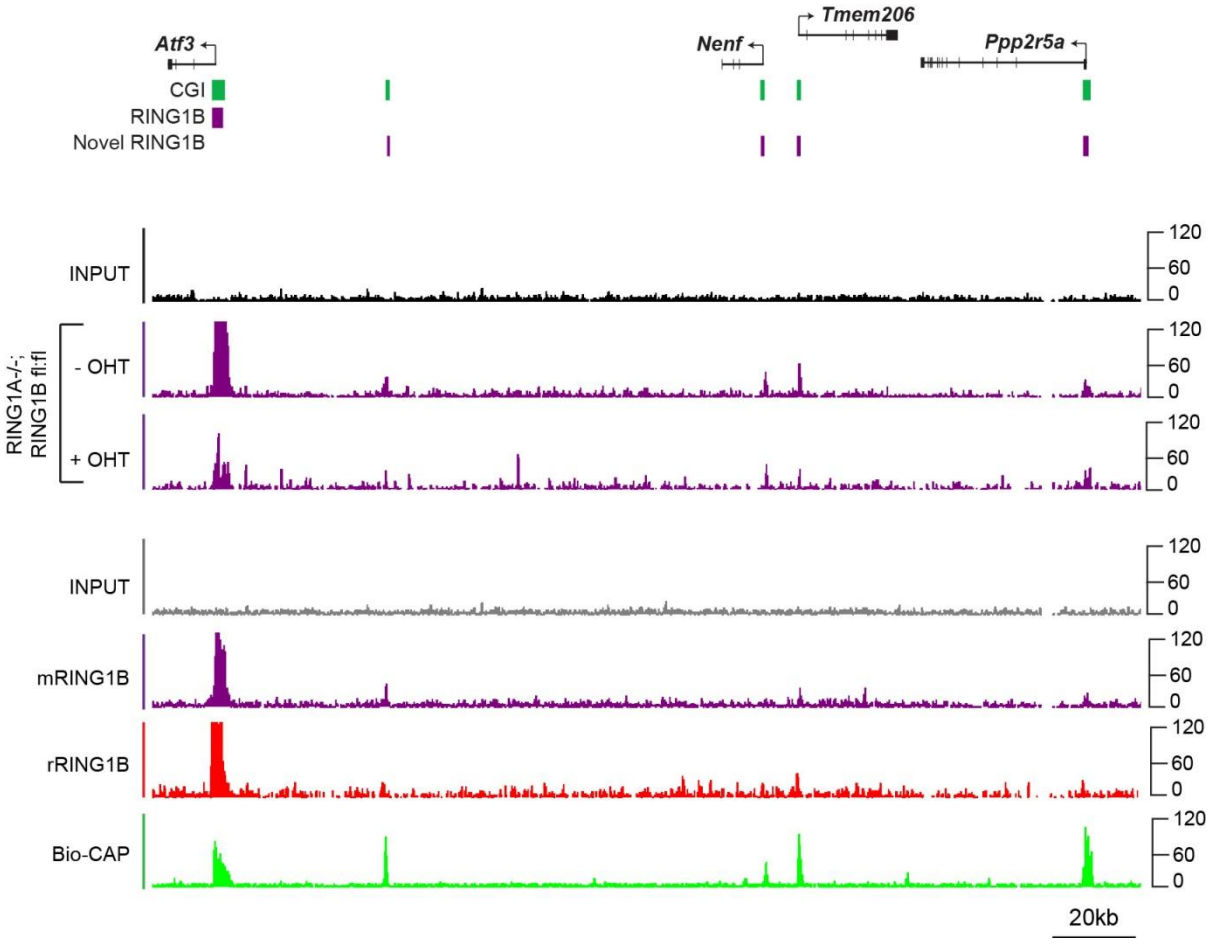


Figure 5.9.Validation of novel-low magnitude RING1B binding events in conditional RING1B knock-out cells and by using a different commercial RING1B antibody.

ChIP-Seq analysis in RING1A$^{-/-}$, RING1B$^{fl:fl}$ ES cells (Endoh et al. 2008) of RING1B occupancy at region shown previously as representative for the identification of novel low magnitude RING1B binding sites containing both a classical PcG target (i.e. *Atf3*) and regular CGIs. Upon tamoxifen-induced conditional RING1B deletion (+OHT), there is a clear loss in RING1B binding levels, suggesting that these novel RING1B peaks identified correspond to real RING1B binding events.

Furthermore, RING1B genome-wide chromatin binding profile was analysed by ChIP-Seq using a commercial rabbit monoclonal antibody specific for RING1B (rRING1B, shown in red). RING1B ChIP-Seq chromatin binding profile generated using this antibody was indistinguishable from the one generated using the mouse monoclonal antibody described previously (mRING1B). Sequencing signal in the corresponding input samples over the same regions are indicated. Bio-CAP-seq signal indicates regions containing non-methylated DNA. Above the sequencing traces gene promoters are show by black arrows and exons by vertical black lines. CGIs are shown as green bars with previously identified RING1B peaks and novel RING1B peaks indicated with purple boxes.

KDM2B is required for RING1B binding to PcG CGIs. Moreover, the discovery that there are detectable RING1B peaks at the majority of CGIs and that this association of RING1B with CpG island elements genome-wide is also dependent on KDM2B suggested that KDM2B acts as a targeting module for PRC1 recruitment. These findings therefore uncover a link between recognition of the underlying non-methylated DNA and PRC1 binding, providing a potential mechanism by which CpG islands could act as a mammalian counterpart to Drosophila PREs.

Interestingly, depletion of KDM2B in mouse ESCs resulted in a striking global loss of H2AK119ub1 ((see (Farcas et al. 2012)), indicating that KDM2B-dependent PRC1 targeting represents a significant contribution to H2AK119ub1 levels at the genome-scale. This striking global loss of H2AK119ub1 levels could be explained by the discovery that KDM2B recruits PRC1 at low-levels to CpG islands genome-wide. This could allow RING1B-mediated H2AK119ub1 deposition away from established Polycomb sites, which would be lost upon KDM2B depletion. The poor performance of H2AK119ub1 antibodies in ChIP-Seq experiments hinder genome-wide analysis of this histone mark in mouse ESCs, but would be essential to verify whether KDM2B-based targeting of PRC1 to CGIs genome-wide also translates into pervasive low level deposition of H2AK119ub1.

5.4. KDM2B targets PCGF1 to CGIs genome-wide

The widespread association between RING1B and CGIs genome-wide appears to be KDM2B dependent. If this binding is mediated by a KDM2B/PRC1 complex, one would expect that other members of the complex should also be broadly associated with CGIs. To test this possibility, an ESC line stably expressing Flag/2xStrep II (FS2)-tagged PCGF1 was generated and used for ChIP-Seq to determine the chromatin binding profile of epitope-tagged PCGF1. This exogenous approach was carried out due to lack of an antibody that specifically and efficiently recognizes endogenous PCGF1. To ensure that the ChIP-Seq profiles observed are not due to non-specific signal from the FS2 antibody, a cell line containing only the empty expression vector was used as control.

Importantly, ChIP-Seq analysis indicated that there are significantly more PCGF1 peaks than would be expected if PCGF1 were to be found only at classical PcG-target CGIs, with PCGF1 found at many CGI regions in the genome. The association between PCGF1 and CpG islands is evident when focusing on regions of the genome containing both classical PcG and non-PcG associated CGIs, while no such peaks were detected in the empty vector control ES line (Figure 5.10, panel A). On average between the two biological replicates, approximately 11,800 FS2-PCGF1 peaks were detected, significantly more than would be expected if PCGF1 was found only at the classical ~2,500 Polycomb-targets. The majority of PCGF1 peaks detected overlap with CpG island elements (Figure 5.10, panel B), as would be expected if the recruitment of PCGF1 is dependent on the ability of KDM2B to specifically recognize non-methylated DNA. Furthermore, PCGF1 peaks extensively overlap with RING1B-associated CGIs (Figure 5.10, panel C). Although fewer total FS2-PCGF1 peaks (~11,800) were observed compared to RING1B peaks (~16,400), this is most probably related to differences in the quality of the antibodies used for each ChIP-Seq experiment, with the RING1B monoclonal antibody functioning much more cleanly than the FS2 antibody in ChIP experiments.

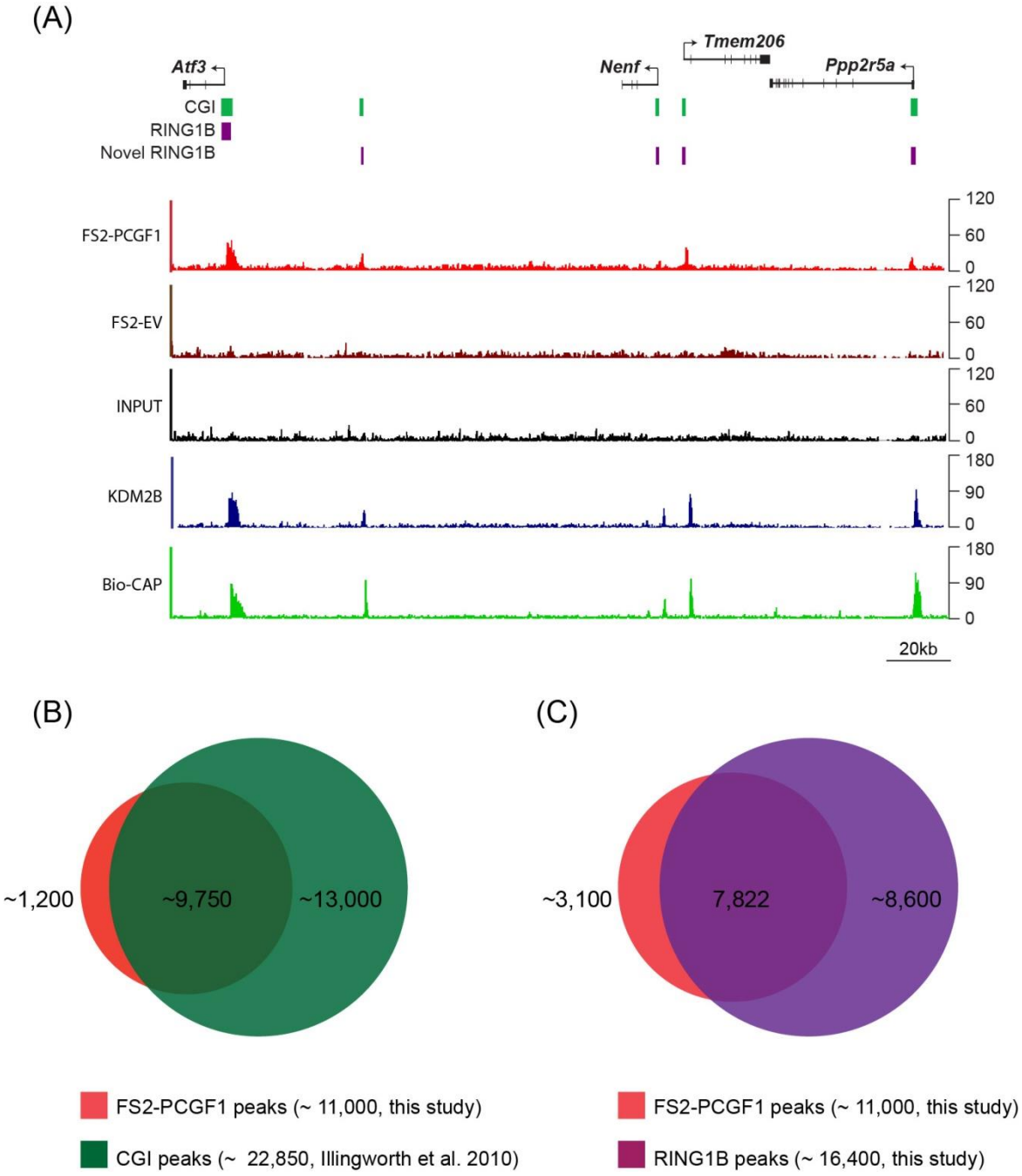


Figure 5.10. PCGF1 binds to CGIs genome-wide in mouse ESCs.

(A) A representative region showing epitope-tagged FS2-PCGF1 ChIP-seq signal in mouse ESCs comprised of both a classical PcG-target (*Atf3*) and non-Polycomb CGIs. The same region was used previously as representative for the identification of novel low magnitude RING1B binding sites (Figure 5.7, panel A). Input and empty expression vector control line sequencing traces over the same region are shown in black and brown, respectively. Bio-CAP-seq signal indicates regions containing non-methylated DNA. Above the sequencing traces gene promoters are show by black arrows and exons by vertical black lines. CGIs are shown as green bars with previously identified RING1B peaks and novel RING1B peaks indicated with purple boxes.

(B) – (C) Venn diagrams showing extensive overlap between epitope-tagged PCGF1 and CGIs (panel B) or RING1B (panel C) peaks. Published datasets for CpG island (CGI) intervals were obtained from (Illingworth et al. 2010).

Although ChIP-Seq using an epitope-tagged PCGF1 reveals a more widespread association between PCGF1 and CGIs, one drawback of this experiment that needs to be considered is that overexpressing PCGF1 might drive it to bind to sites where it would normally not do under physiological conditions. Therefore, ChIP-Seq analysis using an antibody against endogenous PCGF1 is required and will help clarify its chromatin binding profile.

A recent study by Gao et al. suggested that in addition to defining different PRC1 variant complexes, the six PCGF subunits also occupy distinct genomic loci reflecting functional differences (Gao et al. 2012). Importantly, Gao et al. also used an overexpression epitope-tagging system to analyse PCGF chromatin binding profile in a transformed human cell line. Despite some overlap in gene targets with other PCGF proteins, the Gao et al. study showed that PCGF1 target genes were enriched for gene ontology (GO) terms relating to cell junction organization and striated muscle cell differentiation, therefore suggesting that PCGF1 binding is restricted to specific and distinct regions of the genome. This is in contrast with the observations made in this study, which indicate that PCGF1 is instead associated with the majority of CGIs. A closer analysis of the Gao et al. (2012) HA-tagged PCGF1 chromatin binding profile in human transformed cells raised issues regarding the quality of their ChIP-Seq data and the significance of their identified PCGF1 binding sites (supplementary figure S5).

Results in this section suggest that in mouse ESCs, PCGF1 binds CpG island elements genome-wide and is not restricted to PcG-repressed CGIs. Although more work is clearly required to demonstrate the relationship between KDM2B binding and PCGF1 occupancy, it is tempting to speculate that the widespread binding of PCGF1 to CGIs is mediated via KDM2B. Therefore, it would be essential to check whether depletion of KDM2B affects the chromatin binding profile of PCGF1. However, the lack of an antibody that recognizes endogenous PCGF1 has prevented this experiment.

In conclusion, it appears that, through its ability to recognize non-methylated DNA and associate with Polycomb-group components, KDM2B has the capacity to recruit PRC1 to CGIs genome-wide. Together with the observation that depletion of KDM2B leads to reduction in RING1B occupancy at Polycomb and non-Polycomb CGIs, these results demonstrate that KDM2B provides an essential link between recognition of CGIs and function of Polycomb-group proteins.

5.5. Depletion of KDM2B leads to up-regulation of a subset of Polycomb targets

Polycomb proteins are believed to function as transcriptional repressors of developmental regulatory genes (Simon and Kingston 2009; Schwartz and Pirrotta 2013b). To understand whether KDM2B-mediated recruitment of PRC1 plays a role in Polycomb-mediated repression of target genes, Affymetrix microarray analysis was carried out on four biological replicates comparing the gene expression profile of scrambled and KDM2B knockdown (KD) cell lines. The microarray work was performed in collaboration with Dr. Sheena Lee, from the OXION facility at the Department of Physiology, Anatomy and Genetics, University of Oxford. The bioinformatics analysis of the microarray gene expression profiling was conducted in collaboration with Dr. Ian Sudberry (group of Prof. Chris Ponting) from the Computational Genomics Analysis and Training Programme (CGAT), University of Oxford.

KDM2B depletion affects RING1B binding to PcG CGIs genome-wide, suggesting that KDM2B may play a role in Polycomb-mediated repression. Therefore, it was essential to check whether Polycomb-targets showed an increase in gene expression following KDM2B knockdown. Gene expression analysis revealed that a total of 653 genes showed gene expression changes in response to KDM2B depletion by more than 1.5-fold, with approximately half of these being up-regulated. Interestingly, 24% of the 324 genes up-regulated at more than 1.5-fold after KDM2B depletion are characterized by RING1B occupancy in mouse ESCs. This is considerably higher than would be expected by random chance and it suggests that there is a propensity for RING1B-associated genes to show increased gene expression in response to KDM2B depletion. This is evident when visualizing the microarray gene expression profile in a volcano plot, with RING1B-occupied genes shown as red dots, while non-RING1B associated genes are shown in grey (Figure 5.11, panel A). Increase in gene expression of a panel of PcG-targets showing up-regulation in microarray profiling was further validated by qRT-PCR (Figure 5.11, panel B). Importantly, 11 out of the 12 genes selected (all except *Six1*) showed a striking increase in gene expression upon KDM2B depletion. ChIP-qPCR analysis further demonstrated that the same panel of PcG targets showing up-regulation of gene expression in response to KDM2B knockdown had reduced RING1B occupancy and locus-specific lower levels of H2AK119ub1 (see (Farcas et al. 2012)).

Closer analysis of the microarray expression profile (Figure 5.11, panel A) shows that there are also PcG-occupied genes which show down-regulation of gene expression in response to KDM2B depletion. Although this could be due to secondary effects of the shRNA-knockdown procedure, it also suggests that reduction in PRC1 occupancy at PcG CGIs will not necessarily result in transcriptional activation. Importantly, not all genes showing gene expression changes were associated with CGIs. This could be due to KDM2B having pleiotropic roles in regulation of gene expression, or it could be a consequence of secondary effects of the KDM2B knockdown. For example, depletion of KDM2B could affect the expression status of a specific transcription factor, which would then, indirectly of KDM2B KD, result in gene expression changes of its target genes.

Interestingly, microarray-based gene expression profiling suggested that a shRNA targeting only the full length KDM2B did not lead to any significant gene expression changes compared to scrambled shRNA control ES cell line, with only one gene (Ubiquitin-conjugating enzyme E2 variant 1, short = *Ube2v1*) being down-regulated at higher than 1.5 fold . Although the significance of this single change in gene expression remains to be investigated, this result supports the important observation that the short isoform of KDM2B could compensate for loss of the JmjC-containing full length KDM2B, suggesting that the importance of KDM2B is mediated via its ZF-CxxC DNA binding domain and its capacity to associate with the variant PRC1 complex.

Considering the essential and wide-spread role of KDM2B in PRC1 recruitment, one would expect that a higher percentage of the approximately ~ 2,500 classically repressed Polycomb-targets would show increase in gene expression following KDM2B knockdown. One possibility could be that the levels of KDM2B depletion are not sufficient to induce pervasive reactivation of Polycomb-targets. This suggests that the subset of PcG-bound genes showing increase in expression is the one more susceptible to reactivation upon reduction of RING1B occupancy and H2AK119ub1 levels below a certain threshold. However, the results obtained are in agreement with previous studies showing that only a subset of PcG-bound genes becomes reactivated after depletion of PRC1 component PCGF4 or deletion of RING1B (Bracken et al. 2006; Leeb et al. 2010). This suggests that PcG-occupancy *per se* does not determine the

transcriptional state of its bound targets, but plays a more subtle role perhaps by helping to maintain an overall transcriptionally repressive chromatin state of already silent genes.

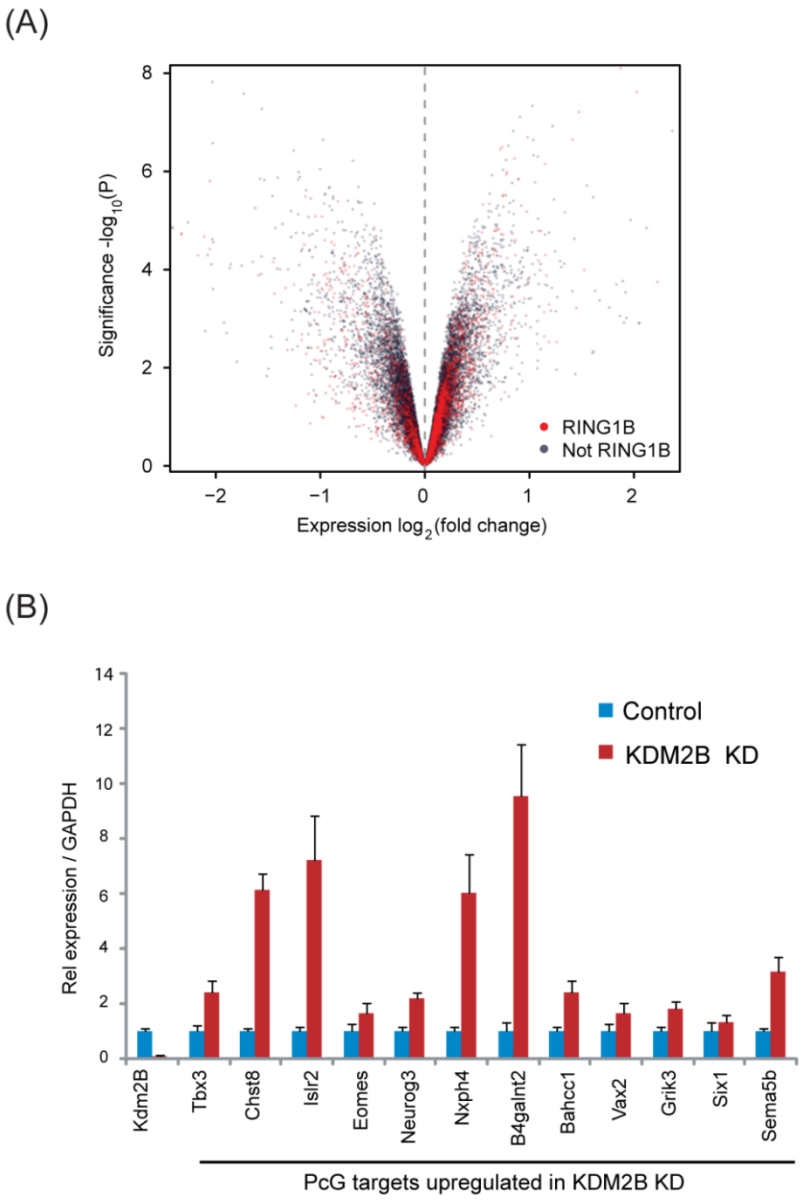


Figure 5.11. Depletion of KDM2B in mouse ESCs affects the gene expression profile and leads to up-regulation of a subset of Polycomb-repressed genes.

(**A**) Volcano plot illustrating the gene expression changes in the KDM2B depletion line compared to the control cell line. The x-axis corresponds to fold change and the y-axis to the significance level. RING1B associated genes are coloured red and show preferential enrichment in the genes up-regulated upon KDM2B knockdown.

(**B**) A panel of 12 polycomb target genes identified in microarray analysis that were up-regulated upon depletion of KDM2B were validated by qRT-PCR analysis. Error bars represent SEM of four biological replicates. Values are normalized to *Gapdh* and expression in the scrambled control line set to 1.

5.6. Summary and discussion

KDM2B binds virtually all CGIs through its ZF-CxxC domain-dependent recognition of non-methylated DNA. Furthermore, KDM2B associates with RING1B, PCGF1 and RYBP/YAF2 to form a variant PRC1 complex. The observation that, unlike other ZF-CxxC domain-containing proteins (i.e. KDM2A and CFP1), KDM2B is preferentially enriched at Polycomb-repressed targets suggested that KDM2B may be recruited to PcG CGIs in a ZF-CxxC independent manner, relying instead on PRC1 (Figure 5.1). In contrast to this idea, KDM2B binding to PcG CGIs was unaffected by loss of RING1A/RING1B (Figure 5.2). Interestingly, KDM2A shows increased occupancy at PcG-associated CGIs after RING1A/RING1B deletion, in agreement with the otherwise reported compact nature of Polycomb-repressed regions now becoming more accessible upon loss of intact PRC1 (King et al. 2002; Francis et al. 2004; Eskeland et al. 2010b).

Given the increase in KDM2A binding at PcG CGIs, one possibility that cannot be excluded is that removal of RING1B can reduce the occupancy of PRC1-associated KDM2B at PcG CGIs. However, the increased accessibility of these otherwise compact regions now allows binding of free KDM2B in a ZF-CxxC dependent manner, the overall ChIP-qPCR result being that the occupancy of KDM2B at PcG CGIs does not change. Nevertheless, the observation that disrupting the DNA binding domain of KDM2B impairs its ability to localize to both regular and PcG-repressed CGIs suggests that the ZF-CxxC domain is the main targeting activity for KDM2B (Figure 5.3). Importantly, upon removal of RING1A/RING1B, both KDM2s showed increased binding to regular CGIs (Figure 5.2, panels C and D). As this was not a consequence of increased KDM2A and 2B protein levels, the reasons for this observation remained elusive at the time. However, low-levels RING1B occupancy at the majority of CGIs could provide one possible explanation. From this perspective, low-magnitude RING1B occupancy of CGIs could be sufficient to impair via structural effect on chromatin (i.e. compaction) the binding of ZF-CxxC domain proteins. When intact PRC1 is removed via deletion of RING1B, the resulting increased accessibility could therefore allow less restrictive binding of ZF-CxxC proteins.

Knockdown of KDM2B using a shRNA-based approach resulted in a reduction of RING1B occupancy at Polycomb targets genome-wide (Figure 5.5), suggesting that KDM2B contributes to PRC1 recruitment to these sites. However, KDM2B binds CpG island elements genome-wide by virtue of its capacity to specifically recognize non-methylated DNA. Therefore, if KDM2B really plays an essential role in PRC1 recruitment, one would expect KDM2B to recruit PRC1 to all CGIs, not just classical Polycomb-targets (Figure 5.6). Indeed, our RING1B ChIP-Seq analysis showed that there are low-magnitude peaks of RING1B occupancy at the majority of CGIs (Figure 5.7). This suggests that targeting of RING1B relies on the recognition of non-methylated DNA by KDM2B. Interestingly, a comparison of the reduction in RING1B binding at regular versus PcG CGIs shows that the loss of RING1B occupancy at these novel low-magnitude CGI-associated sites is greater than its loss at established higher magnitude Polycomb-sites upon KDM2B depletion (compare Figure 5.5, panel B and Figure 5.7, panel B). This suggests that these low-magnitude RING1B peaks are more dependent on KDM2B, while PRC1 binding at classical Polycomb-targets is most probably stabilized by other mechanisms, such as positive reinforcement by PRC2 via the CBX-containing canonical PRC1 complexes.

Importantly, ChIP-Seq analysis of epitope-tagged PCGF1 revealed a more wide-spread association of PCGF1 with CGI elements genome-wide (Figure 5.10). Recent ChIP-Seq analysis of BCOR chromatin binding profiles in Diffuse-large B-cell lymphoma (DLBCL cells) showed that BCOR occupancy is not restricted to BCL6 targets, suggesting that it has BLC6-independent functions (Hatzi et al. 2013). Instead, BCOR was found more frequently associated with CpG island promoters, supporting the notion that BCOR is also recruited to CGIs via its association with KDM2B.

Therefore, it appears that KDM2B recruits PCGF1-PRC1 at low levels to CGIs genome-wide, allowing PRC1 targeting in a manner that does not rely on the recognition of PRC2-mediated H3K27me3. It should be noted that after the publication of this work (see (Farcas et al. 2012)), two other studies essentially showing the same KDM2B-dependent recruitment of PRC1 to Polycomb CGIs were published (He et al. 2013; Wu et al. 2013b). Interestingly, He et al. (2013) also report that more KDM2B-bound CGIs are co-occupied by PRC1 at low levels than just classical PcG targets.

The discovery that the majority of CGIs are occupied by low levels of PRC1 prompts the need to understand the relevance of these novel KDM2B-dependent RING1B sites. More specifically, if all CpG island elements have the capacity to engage PRC1, then this might suggest that something prevents establishment and accumulation of PRC1 and PRC2 group proteins at non-PcG target CGIs. One explanation could be that PRC2 components do not have the capacity to interrogate all CGIs and are restricted in their binding to Polycomb-targets, where they would get specifically recruited via dedicated transcription factors. This would therefore impede establishment of PRC accumulation outside PcG targets. However, the striking redistribution of H3K27me3 in severely DNA hypomethylated mouse somatic cells suggests that PRC2 can interpret non-methylated GC-rich DNA in a manner similar to PRC1 and is responsive to changes in DNA methylation (Reddington et al. 2013).

Polycomb occupancy has classically been considered a driver of transcriptional repression via its specific recruitment by dedicated transcription factors (King et al. 2002; Mohd-Sarip et al. 2002; Francis et al. 2004) . However, more and more studies indicate that Polycomb activity at target loci is responsive rather than instructive to the transcriptional state of the associated gene (Langlais et al. 2012; Caputo et al. 2013). Therefore, we propose that an alternative interpretation for this low-level PRC1 occupancy at CGIs could be that KDM2B acts as a sampling module for PRCs to interrogate the susceptibility of the CGI-associated gene for Polycomb repression (Klose et al. 2013). PRC1 can bind CGIs, but in the absence of the appropriate chromatin environment or signals, Polycomb establishment cannot take place. In terms of what the appropriate signals might mean, one can predict that the most valid candidate would likely be the process of transcription itself.

According to this model, the permissive chromatin architecture of CGIs enables KDM2B-PRC1 to scan these genomic elements regardless of their transcription status. In the absence of transcription, PRC1 and PRC2 can accumulate and establish a classical Polycomb-repressed domain. The stability of this association is further reinforced by feedback mechanisms between PRC1 and PRC2. For example, PRC1 occupancy and potentially H2AK119ub1-deposition were speculated to generate a more compact chromatin architecture favouring PRC2 enzymatic activity (Yuan et al. 2012), while H3K27me3 is recognised by the canonical CBX-containing PRC1 to mediate H2K119ub1 and amplify PRC occupancy.

However, although KDM2B-PRC1 and PRC2 can sample CGIs irrespective of the transcriptional state of the associated gene, productive transcription antagonizes the function of PcG complexes and blocks accumulation of PRCs (Figure 5.12). Although more work is required to understand which step or aspect of transcription is responsible for this antagonism, the observation that RNAPII is also present across silent regions of the genome (Guenther et al. 2007; Muse et al. 2007) suggests that the antagonism is most probably a result of productive transcriptional elongation and the accompanying chromatin modifications. For example, H3K4me3, H3K27ac and H3K36me2/3 which are associated with gene activation were shown to directly inhibit PcG activity (Schmitges et al. 2011; Reynolds et al. 2012), indicating a multilateral effort in counteracting PcG accumulation during the process of transcription.

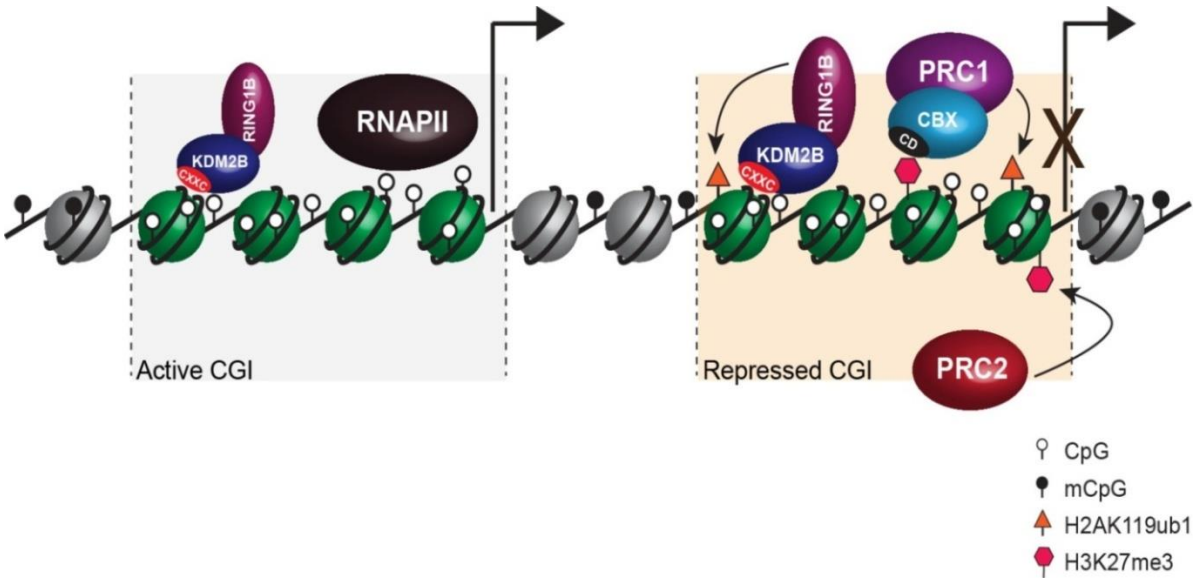


Figure 5.12. Polycomb occupancy is responsive to gene expression status.

By virtue of its ZF-CxxC DNA binding domain, the KDM2B-PRC1 complex samples CpG islands irrespective of the transcriptional state of the associated gene, suggesting that Polycomb-occupancy is the default state of CGIs. In the absence of transcription, PRC2-mediated H3K27me3 would allow binding of canonical CBX-containing PRC1 complexes, in addition to KDM2B-dependent PRC1 recruitment. This accumulation of PRCs would then reinforce the Polycomb-occupied established state and contribute to maintaining a transcriptionally inhibitory chromatin state at the respective gene.

In conclusion, work in this chapter has shown that, in mouse ESCs, KDM2B recruits PRC1 to PcG targets, but also at low levels to CGIs genome-wide, potentially as a sampling mechanism responsive to the transcriptional state of the CGI-associated gene. Although important, KDM2B-mediated recruitment of PRC1 is definitely not the sole targeting mechanism for Polycomb-group proteins. Further work is clearly required to understand how the KDM2B-PRC1 complex fits into the network of alternative PcG targeting mechanisms known to function in addition to the hierarchical recruitment model. Despite the complexity of the system, the discovery that PRC1 can interrogate non-methylated DNA via KDM2B provides an exciting avenue to explore the functionality of CGIs as mammalian PREs.

6. Chapter six - KDM2B targeting contributes to PRC1 and PRC2 occupancy at target sites *in vivo*

On the basis of its association with RING1B to form a variant PRC1 complex and its capacity to specifically recognize non-methylated CpG island DNA, the possibility that KDM2B may play a role in PRC1 recruitment to target sites was addressed in Chapter 5. Indeed, shRNA-based depletion of KDM2B resulted in a reduction of RING1B binding at Polycomb targets genome-wide, together with a reactivation of a subset of Polycomb-occupied genes. Interestingly, closer analysis of the RING1B ChIP-Seq profile revealed a more widespread low level association of RING1B with CGI elements genome-wide. This observation is further supported by epitope-tagged PCGF1 ChIP-Seq profile, which indicated that there are significantly more CGI-associated PCGF1 peaks than would be expected if PCGF1 were to be found only at classic Polycomb-targets. This low level KDM2B-dependent targeting of variant PRC1 to CGIs genome-wide was proposed to act as a sampling mechanism by which Polycomb proteins respond to the transcriptional state of the CGI-associated gene.

We and others have recently shown that KDM2B plays a role in targeting PRC1 to target sites (Farcas et al. 2012; He et al. 2013; Wu et al. 2013b). However, the conclusions drawn were based entirely on a shRNA-based knockdown of KDM2B. This system is suboptimal as it incompletely depletes the entire protein, thereby not permitting to dissect whether the contribution of KDM2B to PRC1 occupancy is directly dependent on its ability to recognize non-methylated DNA or whether it is non-ZF-CxxC related. In an attempt to study the *in vivo* biological significance of KDM2B, Fukuda et al. (2011) generated a constitutive homozygous mutation in the mouse *Kdm2b* gene. Mice deficient for KDM2B showed a failure of cranial neural tube closure, resulting in exencephaly at moderate frequency, a phenotype milder than would be expected for a protein with an essential role in Polycomb targeting. However, this strategy only deleted exons essential for the histone demethylase catalytic function, thereby affecting only the long form of KDM2B. As shown in previous chapters, the short form of KDM2B retains a presumably functional ZF-CxxC DNA binding domain and its capacity to associate with Polycomb

proteins to form a variant PRC1 complex. Therefore, the developmental defects observed do not take into account potential redundancy between the two isoforms of KDM2B regarding its contribution to Polycomb function.

The discovery that KDM2B contributes to PRC1 recruitment to target sites provides an essential link between recognition of non-methylated CGI DNA and nucleation of the Polycomb repressive system. However, the mechanisms by which PRC2 preferentially localizes to CpG islands remain poorly characterized (Ku et al. 2008), although multiple pathways have been suggested. For example, JARID2 was identified as a substoichiometric associating partner of PRC2 and depletion of JARID2 was shown to impair PRC2 recruitment to target genes and lead to a concomitant reduction in H3K27me3 levels at PcG-target promoters (Landeira et al. 2010; Li et al. 2010; Pasini et al. 2010a). Although JARID2 contains an ARID and a Zinc-finger DNA-binding domain, it shows no clear sequence specificity and only a slight bias towards GC-rich sequences, rendering its contribution to PRC2 targeting unclear. Another protein copurifying with PRC2 is AEBP2, which was shown to be required for optimal PRC2 enzymatic activity and whose Zinc-finger DNA-binding domain was also suggested to be essential for PRC2 recruitment, although this possibility was not tested at the genome-wide level (He et al. 1999; Cao and Zhang 2004b; Kim et al. 2009; Ciferri et al. 2012).

Dissecting the mechanisms by which CpG islands may act as the mammalian counterpart to fly PREs is further complicated by the multitude of factors functioning at these genomic elements, such as hierarchical recruitment mechanisms at natural PcG-targets or presence of non-coding RNAs. These multiple confounding factors functioning at CGIs complicate interpretation of how Polycomb complexes are targeted to chromatin sites. Furthermore, they highlight the need to develop improved experimental tools which would enable one to address specific steps in PcG nucleation at sites devoid of the intrinsic complexity of natural polycomb targets.

By taking advantage of a system to inducibly delete the ZF-CxxC DNA binding domain of KDM2B, together with a *de novo* targeting assay to examine the effects of KDM2B tethering to an artificial CpG-free locus, work in this chapter uncovers an unexpected role for KDM2B in mediating recruitment of both PRC1 and PRC2 and formation of polycomb domains in mouse ESCs.

6.1. Conditional deletion of KDM2B DNA-binding domain impairs both PRC1 and PRC2 occupancy at Polycomb targets genome-wide

6.1.1. Deletion of KDM2B ZF-CxxC DNA binding domain reduces RING1B, PCGF1 and SUZ12 occupancy at a panel of polycomb-target CGIs

To understand whether KDM2B links recognition of non-methylated CGI DNA to PRC1 recruitment and Polycomb function, homozygous mouse ES cells in which KDM2B exon 13 encoding the ZF-CxxC DNA binding domain is flanked by *loxP* sites were obtained on a collaborative basis from Dr. Takashi Kondo and Dr. Haruhiko Koseki at the RIKEN Research Center for Allergy and Immunology, Japan (Blackledge, Farcas, Kondo et al., *submitted*). These Kdm2b $^{fl/fl}$ cells stably express a tamoxifen-inducible Cre recombinase, allowing tamoxifen (OHT)-dependent excision of the ZF-CxxC DNA binding domain of both the long (LF) and short KDM2B (SF) isoforms (Figure 6.1, panel A). Importantly, Cre-mediated recombination at the *loxP* sites flanking exon 13 results in a 225 bps deletion (75 amino acids) of the ZF-CxxC domain, while leaving the rest of the KDM2B protein in frame. Tamoxifen-inducible Cre-recombination should therefore result in KDM2B isoforms which specifically lack the ability to bind non-methylated CGI DNA, but which retain the capacity to associate with Polycomb proteins and form the variant PRC1 complex (Figure 6.1, panel B and Supplementary figure S6).

A time-course experiment of tamoxifen-induced Cre/*loxP* recombination in the kdm2b$^{fl/fl}$ ES cells showed an increased mobility (by approximately 8kDa) for both KDM2B isoforms as early as 24 hours post-treatment, with near complete depletion of the ZF-CxxC DNA binding domain by 72 hours post-OHT addition (Figure 6.1, panel C). Interestingly, although equal concentrations of nuclear extract were loaded as indicated by the LAMIN B control Western blot, an apparent increase in KDM2B protein levels was observed following tamoxifen administration. As KDM2B RNA levels do not significantly change, this could potentially be due to an increase in protein stability. Alternatively, in the absence of an intact ZF-CxxC DNA binding domain, KDM2B is no longer tightly bound to DNA thereby allowing a more efficient

nuclear extraction of the ΔCxxC protein compared to the wild-type version. This would result in an apparent increase in KDM2B protein levels when analysed by Western blotting.

To verify that OHT-inducible deletion of the ZF-CxxC domain leads to efficient removal of KDM2B from chromatin, KDM2B ChIP-qPCR was performed in the Kdm2b$^{fl/fl}$ ES cells and at 48hr and 72hrs post-tamoxifen treatment (Figure 6.1, panel D). An almost complete loss of KDM2B occupancy at CGIs was observed following tamoxifen-inducible deletion of the ZF-CxxC DNA binding domain by 48 hrs. Despite near complete deletion of the DNA binding domain as determined by Western blot, the ChIP-qPCR levels of KDM2B after OHT treatment were slightly higher than background as detected for a control non-CGI body region. A potential biological significance of this observation is unclear, although it could suggest that KDM2B is retained at low-levels at CGIs in the absence of its DNA binding domain. However, a simpler explanation is that this result may be due to difficulties in experimentally establishing and interpreting background levels in ChIP-qPCR experiments, particularly at gene promoter regions which appear more sensitive to formaldehyde-mediated crosslinking (Teytelman et al. 2013).

Efficient removal of KDM2B from chromatin upon OHT treatment led to the conclusion that the KDM2B LF and SF ZF-CxxC DNA binding domain can be conditionally deleted in the Kdm2b$^{fl/fl}$ ES cell line by tamoxifen addition.

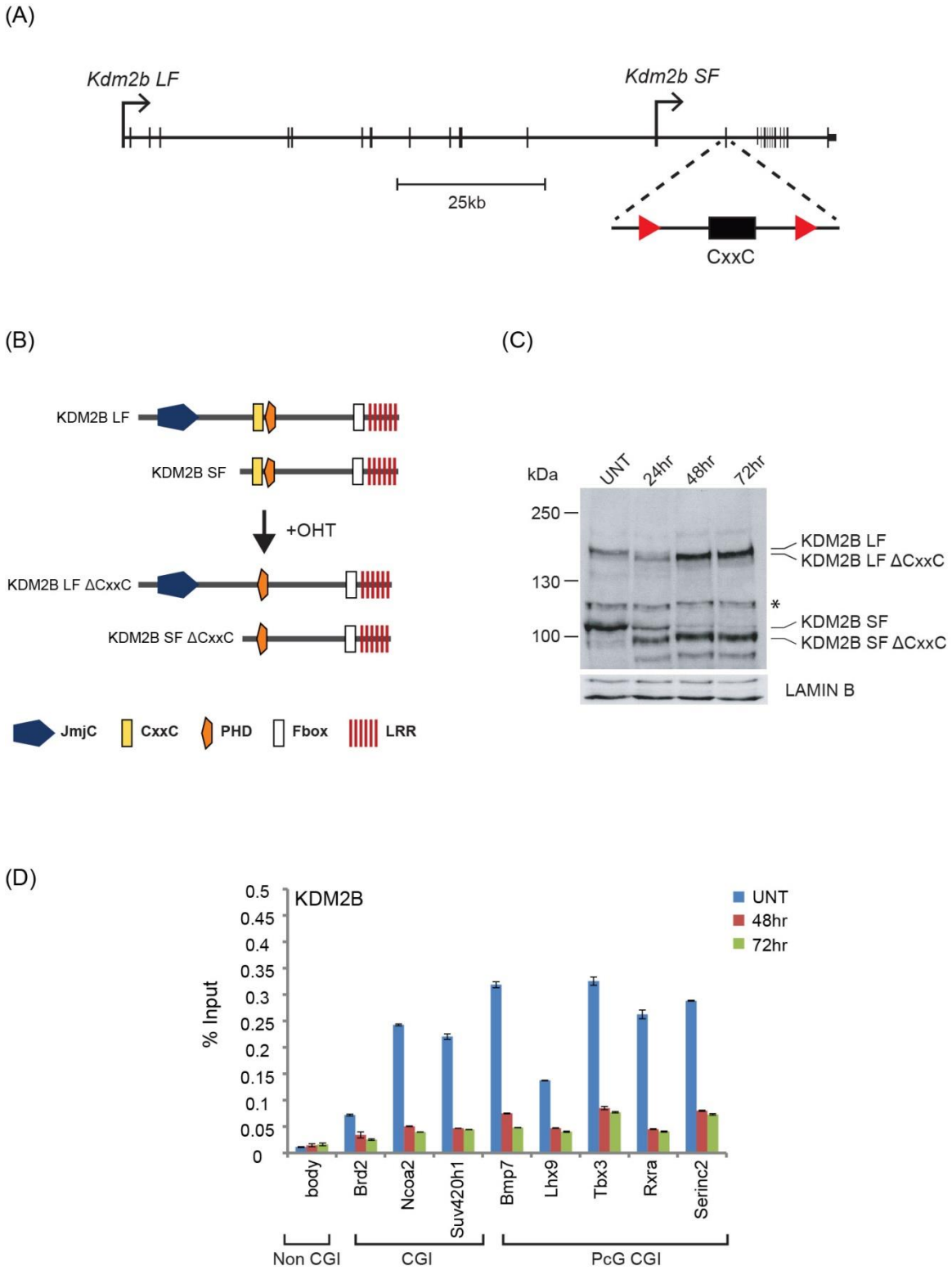


Figure 6.1. Conditional deletion of KDM2B ZF-CxxC DNA binding domain abrogates KDM2B binding to CGIs.

(A) Schematic representation of the *Kdm2b* gene showing that the ZF-CxxC DNA binding domain deletion affects both the long form (LF) and the short form (SF) of KDM2B. The positions of loxP sites flanking exon 13 encoding the ZF-CxxC domain are highlighted as red triangles.

(B) Schematic representation of KDM2B isoforms, illustrating that tamoxifen (OHT)-dependent removal of the ZF-CxxC domain specifically deletes the DNA binding domain of both LF and SF KDM2B, while leaving the rest of the protein intact (as supported by Supplementary figure S6).
(C) Western blot analysis of KDM2B protein levels in the untreated and following 24, 48, and 72 hours of tamoxifen (OHT) treatment in the Kdm2b$^{fl/fl}$ cell line. Near complete deletion of the ZF-CxxC domain is achieved by 72 hours of treatment as indicted by the appearance of a lower migrating band for both the long (LF) and short forms (SF) of the KDM2B protein. The (*) indicates a non-specific antibody cross-reactive band. LAMIN B was used as loading control. The apparent increase in band intensity is most probably a result of more efficient extraction of KDM2B from chromatin following deletion of its ZF-CxxC DNA binding domain.
(D) ChIP-qPCR analysis of KDM2B occupancy in mouse Kdm2b $^{fl/fl}$ ES cells at a panel of PcG and non-PcG target gene associated CGIs and a control gene body region (non-CGI). Tamoxifen-mediated deletion of the KDM2B ZF-CxxC DNA binding domain results in loss of KDM2B binding to CGIs at 48hr and 72hr post-treatment. ChIP experiments were performed in biological duplicates with error bars showing SEM.

Knockdown of KDM2B using a shRNA-based approach caused a reduction in RING1B binding at Polycomb targets genome-wide (chapter 5), suggesting that KDM2B acts as a targeting module for PRC1. Although not clearly demonstrated, it was hypothesized that the ability of KDM2B to specifically recognize non-methylated CpGs was critical for this function. It was therefore essential to verify whether conditional deletion of KDM2B ZF-CxxC DNA binding domain impacts on the chromatin occupancy of the PCGF1/PRC1 variant complex. To this end, ChIP-qPCR analysis was carried out comparing RING1B and PCGF1 chromatin binding at a panel of PcG and non-PcG target CGIs in the Kdm2b$^{fl/fl}$ and 48hr – 72hr tamoxifen-treated ES cells. Importantly, ablation of KDM2B ZF-CxxC domain did not result in destabilization of either PRC1 or PRC2 proteins (Figure 6.2, panel A), allowing an examination of the specific effect of this deletion on PRC1 chromatin occupancy. Interestingly, the same apparent increase in PCGF1 protein levels by Western blot was observed after tamoxifen addition, potentially also as a consequence of more efficient PCGF1 extraction from chromatin following loss of its KDM2B-dependent targeting mechanism. Considering the multitude of links between PRC1 and PRC2, the effect of KDM2B DNA-binding domain deletion was also investigated on the chromatin occupancy of PRC2 component SUZ12 (Figure 6.2, panel B).

As expected, tamoxifen treatment of the Kdm2b$^{fl/fl}$ cells lead to near complete loss of KDM2B binding at the two Polycomb-associated CGIs tested (i.e. *Bmp7* and *Serinc2*) (Figure 6.2, panel C). When RING1B binding levels were investigated, a specific reduction in RING1B signal at those sites was detected following KDM2B ZF-CxxC deletion, in fitting with a role for KDM2B in RING1B targeting (Figure 6.2, panel D). Interestingly, RING1B loss appears more dramatic at 72hrs versus 48hrs of tamoxifen treatment, potentially suggesting that RING1B occupancy can be stabilized at low-levels by alternative mechanisms, but prolonged culture in the absence of a functional targeting module will result in a dramatic reduction of RING1B occupancy.

PCGF1 was found to be more widely associated with CGIs than expected if it were restricted to Polycomb-occupied genes, presumably as a result of its association with KDM2B (Chapter 5). However, lack of an antibody recognizing endogenous PCGF1 prevented such a study in the KDM2B knockdown cell line. It was therefore important to generate such an antibody, to avoid relying on epitope-tagged versions of the protein. After multiple unsuccessful attempts, an antibody specifically recognizing PCGF1 and with the capacity to immunoprecipitate the protein under ChIP conditions was generated and characterized (Supplementary Figure S7). Using this newly generated antibody, it was possible to detect that deletion of KDM2B ZF-CxxC domain impairs PCGF1 binding to a panel of CGIs (Figure 6.2, panel E), in agreement with KDM2B-dependent recruitment of PCGF1/PRC1 variant complex to chromatin.

Interestingly, ChIP-qPCR analysis also revealed a clear reduction in SUZ12 occupancy at two Polycomb-associated CGIs following OHT-mediated ablation of KDM2B DNA binding domain (Figure 6.2, panel F). Importantly, the loss of SUZ12 binding as detected by ChIP-qPCR is not as dramatic as for RING1B. This observation suggests that impaired chromatin-bound levels of PRC2 are not due to loss of direct recruitment by KDM2B, but is most probably an indirect consequence, potentially mediated via reduced PRC1 occupancy in the absence of an intact KDM2B targeting pathway.

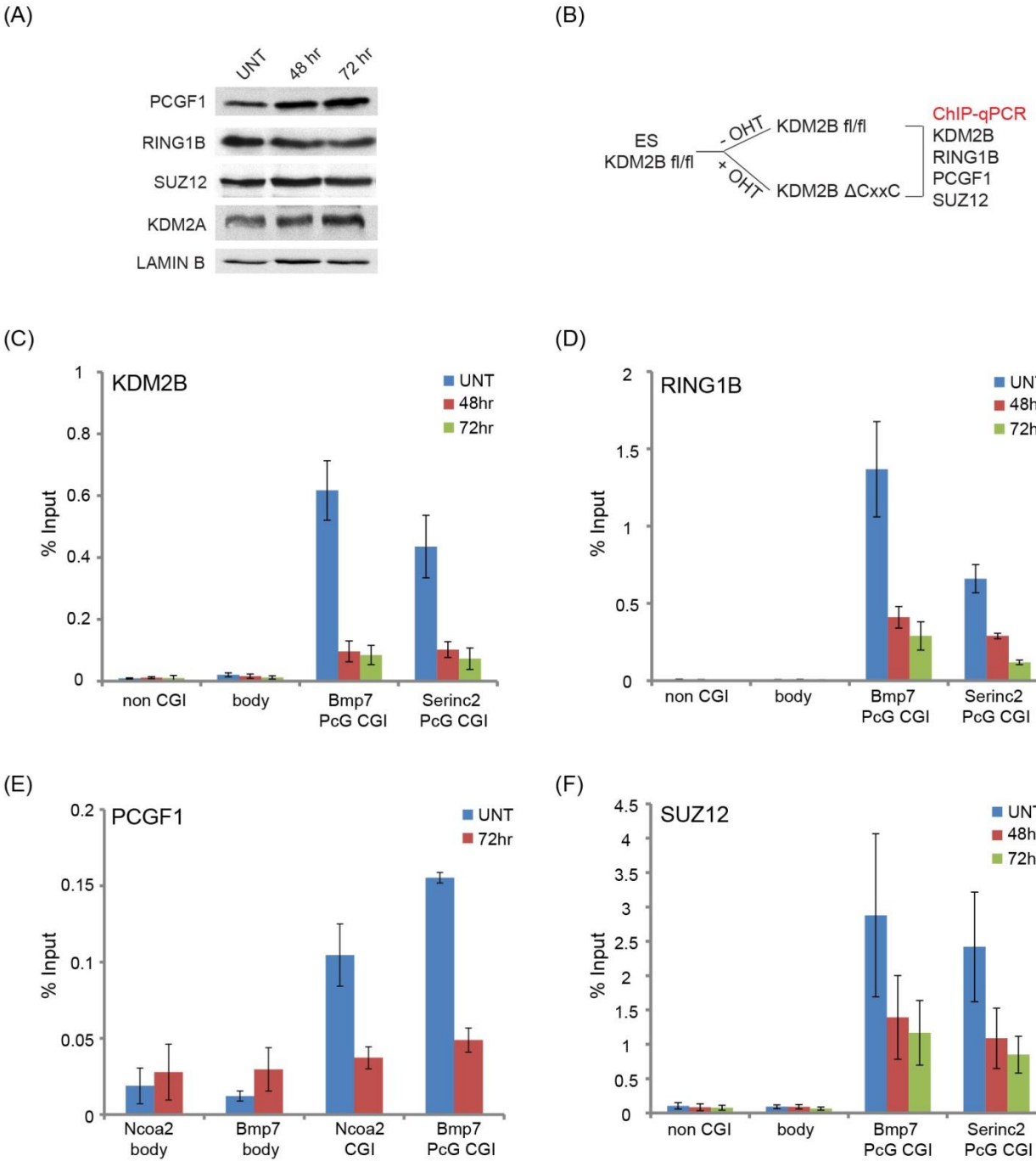


Figure 6.2. Disruption of KDM2B ZF-CxxC DNA binding domain leads to reduced RING1B, PCGF1 and SUZ12 occupancy at a panel of Polycomb CGIs.

(A) Western blot analysis showing that tamoxifen-inducible deletion of the KDM2B ZF-CxxC DNA binding domain does not affect the closely related protein KDM2A and does not destabilize other Polycomb group proteins such as RING1B, PCGF1 or SUZ12. LAMIN B indicates equal loading. This Western blot analysis was carried out by a summer student, Mr. Thomas Sheahan, under my joint supervision with Dr. Neil Blackledge.

(B) Schematic representation showing removal of KDM2B ZF-CxxC DNA binding domain in Kdm2b$^{fl/fl}$ ES cells following tamoxifen (OHT) treatment.

(C) ChIP-qPCR analysis indicating that KDM2B binding to two Polycomb target CGIs is lost after 48hr and 72hr of tamoxifen-mediated deletion of its ZF-CxxC domain. Non-CGI and body region are shown as control. Error bars represent SEM of three biological replicates.

(D) ChIP-qPCR analysis showing that RING1B occupancy at two PcG-target CGIs is reduced after tamoxifen-mediated deletion of KDM2B ZF-CxxC DNA binding domain. Error bars represent SEM of three biological replicates.
(E) ChIP-qPCR analysis showing that deletion of KDM2B ZF-CxxC domain abrogates PCGF1 occupancy at a regular and a PcG-target CGI. Two body regions are shown as control. Error bars represent SEM of two biological replicates.
(F) ChIP-qPCR analysis showing that SUZ12 occupancy at two PcG-target CGIs is reduced after tamoxifen-mediated deletion of KDM2B ZF-CxxC DNA binding domain, suggesting that KDM2B may play a role in PRC2 function. Error bars represent SEM of three biological replicates.

6.1.2. Deletion of KDM2B ZF-CxxC DNA binding domain impairs RING1B and SUZ12 occupancy at polycomb targets genome-wide

To understand whether reduced RING1B binding in the absence of KDM2B-mediated recognition of non-methylated CGI DNA is a genome-wide phenomenon, KDM2B and RING1B ChIP-Seq analysis was carried out in the untreated Kdm2b$^{fl/fl}$ cells and following 72 hrs of tamoxifen treatment. Given the clear reduction in SUZ12 binding at two Polycomb-target CGIs observed previously (Figure 6.2), SUZ12 genome-wide chromatin binding profile in the untreated and tamoxifen-treated Kdm2b$^{fl/fl}$ cells was also investigated, to address whether KDM2B could potentially play a more pervasive role in mediating PRC2 occupancy.

Approximately 8900 RING1B peaks were identified by ChIP-Seq in the untreated cells. Although less than determined in the previous scrambled shRNA-control experiment (Chapter 5), the number of RING1B peaks detected are still significantly higher than predicted if it were associated solely with classic Polycomb targets, suggestive of a more widespread association of RING1B with CGIs. One explanation for the lower number of RING1B peaks observed could be the high-stringency of the ChIP-Seq analysis. Briefly, ChIP-Seq experiments for RING1B in the Kdm2b$^{fl/fl}$ cells were performed in biological triplicates and the final RING1B peak set was called when overlapping RING1B peaks were detected in all three replicates.

When visualizing the ChIP-Seq profiles as heat maps covering a 10kb region around RING1B peaks arranged in decreasing order of RING1B binding levels, it was evident that tamoxifen-induced deletion of the ZF-CxxC DNA binding domain drastically abrogated KDM2B chromatin occupancy (Figure 6.3, panel A). In agreement with the results obtained after KDM2B knockdown-based depletion, ablation of KDM2B ZF-CxxC domain resulted in a greater than 1.5-fold reduction of RING1B binding at roughly 20% of the RING1B peaks identified, suggesting that the ability of KDM2B to recognize non-methylated CGI DNA is an essential contributor to PRC1 recruitment. When KDM2B was depleted by the shRNA-based approach, the most dramatic loss of RING1B occupancy was observed at low-magnitude RING1B peaks. In contrast, high-magnitude peaks corresponding to classic Polycomb targets showed the clearest reduction in RING1B binding after tamoxifen-mediated KDM2B ZF-CxxC domain deletion. Although the reasons behind this discrepancy remain unclear, it is important to consider that these are two different systems, with cells stably maintained in culture for a prolonged period under KDM2B shRNA-mediated depletion, while the effects of conditional deletion of ZF-CxxC DNA binding domain were investigated after 72hours of tamoxifen treatment. It is tempting to speculate that immediately following impaired KDM2B-dependent PRC1 recruitment by either knockdown or conditional ablation, RING1B occupancy at high-magnitude Polycomb targets is affected first. However, longer-term culture may allow PRC1 binding to be partially maintained by alternative targeting pathways and potential stabilizing feedback mechanisms between PRC1 and PRC2. In contrast, the dependence of low-magnitude RING1B peaks on KDM2B targeting becomes more apparent with increased culturing time in absence of a functional recruitment pathway.

Interestingly, when SUZ12 ChIP-seq was carried out in the Kdm2b$^{fl/fl}$ and 72 hrs tamoxifen treated cells, a striking and widespread reduction of SUZ12 occupancy was observed at polycomb targets genome-wide (Figure 6.3, panel A). Importantly, of the approximately 1900 SUZ12 peaks which showed a significant and greater than 1.5-fold loss upon ZF-CxxC domain deletion, more than 70% of these overlapped with sites also showing a greater than 1.5-fold reduction in RING1B occupancy, suggesting that impaired KDM2B-dependent recruitment of PRC1 to polycomb targets genome-wide also disrupts PRC2 binding (Figure 6.3, panel B).

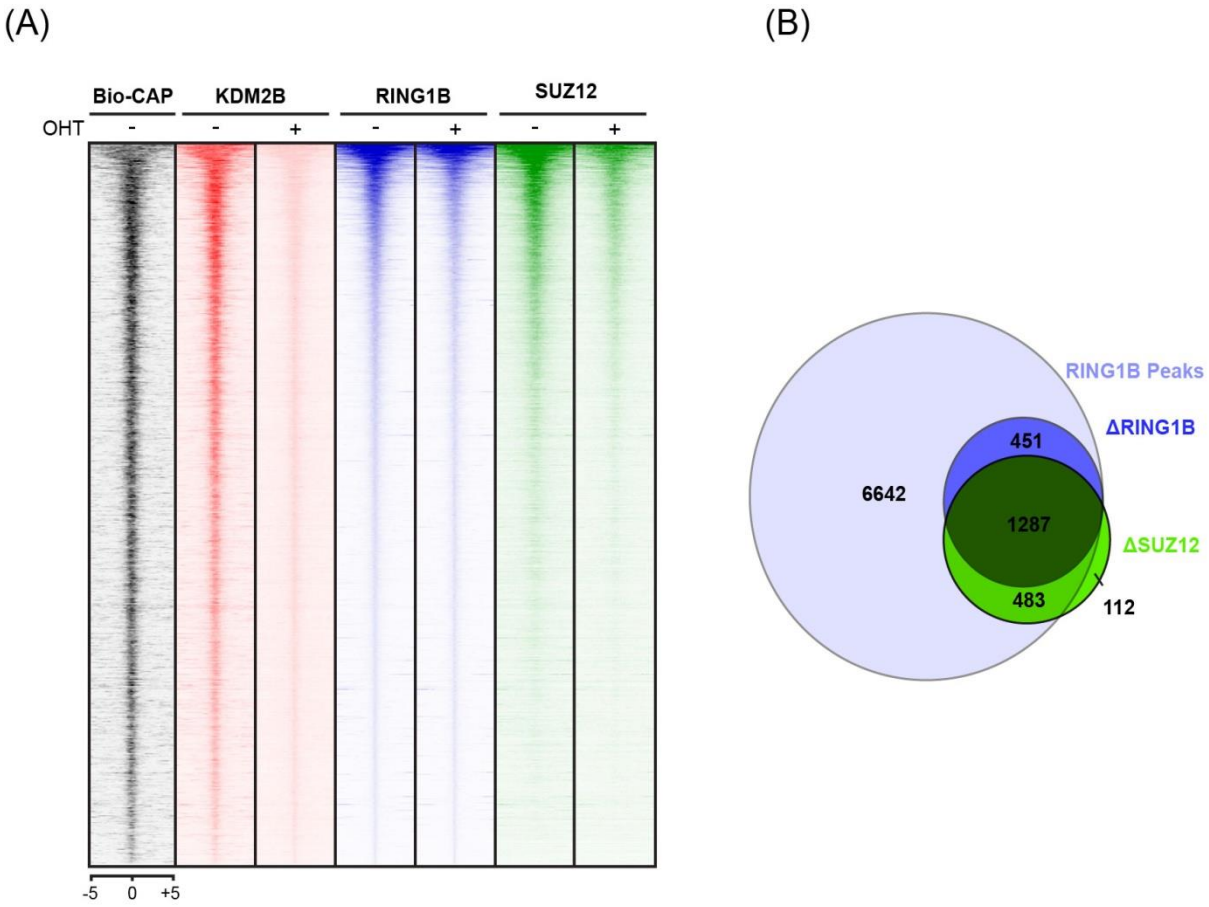


Figure 6.3. Deletion of KDM2B ZF-CxxC DNA binding domain results in loss of RING1B and SUZ12 binding at polycomb targets genome-wide.

(A) Heat map analysis of RING1B peaks (n=8961, as determined from three biological replicates) for KDM2B, RING1B and SUZ12 covering a 10 kilobase region centred over the RING1B peak prior to (-OHT) and following 72hrs (+OHT) tamoxifen treatment of kdm2b $^{fl/fl}$ ES cells. The peaks are arranged in decreasing order of RING1B binding magnitude. Bio-CAP is included to indicate non-methylated DNA signal at these sites.

(B) A Venn diagram showing that the majority of locations with significant changes in RING1B or SUZ12 occupancy following removal of the KDM2B ZF-CxxC domain overlap, and nearly all changes are restricted to RING1B peaks (total of n=8961).

The near complete loss of KDM2B chromatin occupancy and the striking reduction in RING1B and SUZ12 binding after tamoxifen-mediated ZF-CxxC DNA binding domain deletion was evident when ChIP-Seq signals were plotted at individual polycomb-associated CGIs, with *Klrg2, Epha7*, *Atf3* and *Npr1* loci shown as representative examples (Figure 6.4).

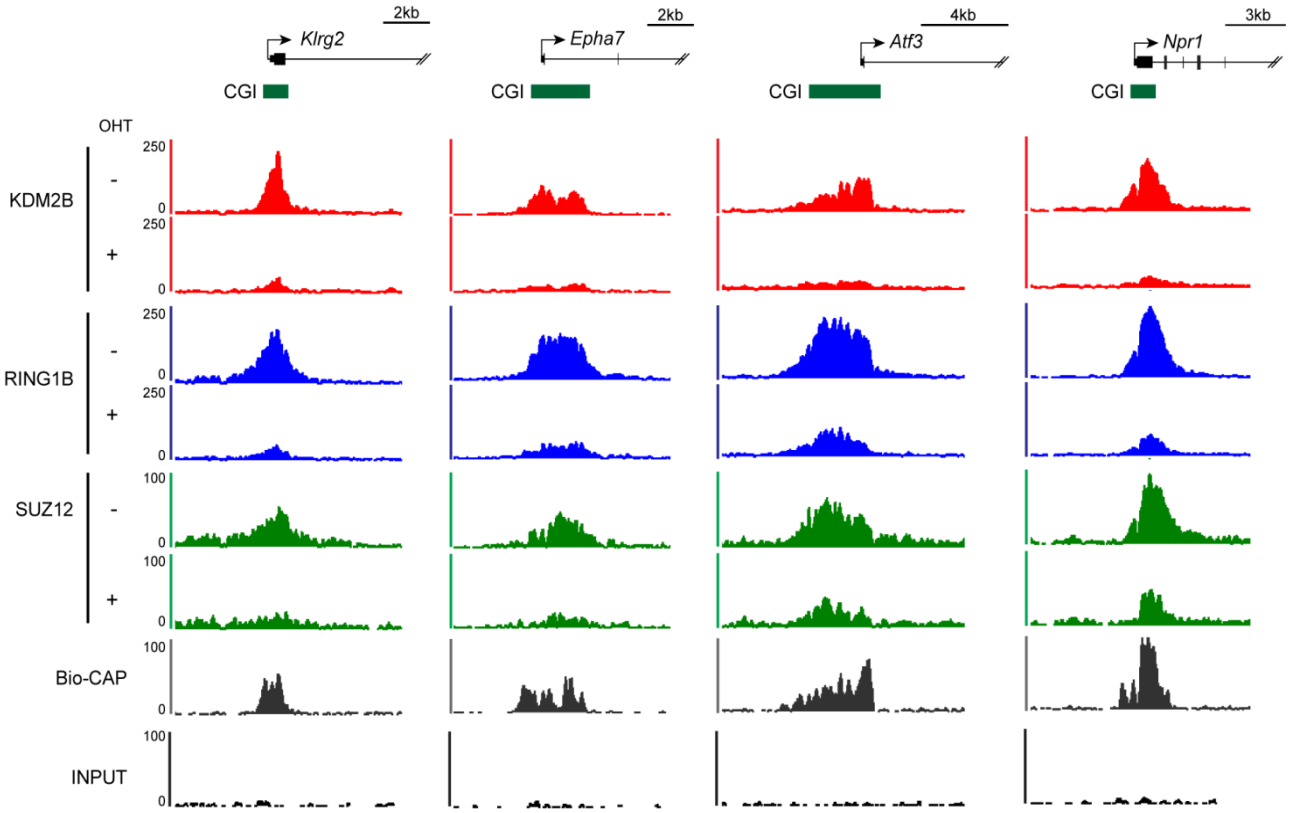


Figure 6.4. Deletion of KDM2B ZF-CxxC DNA binding domain results in reduction of RING1B and SUZ12 occupancy at classic polycomb targets.

ChIP-seq trace snapshots for KDM2B, RING1B, and SUZ12 in the Kdm2b$^{fl/fl}$ ES cells prior to (-OHT) and following 72hrs (+OHT) of tamoxifen treatment. Four classic Polycomb target CGI-associated genes (*Klrg2*, *Epha7*, *Atf3* and *Npr1*) are depicted illustrating the reduction in RING1B and SUZ12 occupancy after deletion of the KDM2B ZF-CxxC DNA binding domain. Corresponding input ChIP-Seq trace is shown as control for potential ChIP-Seq biases. In all cases, the Bio-CAP sequencing trace is shown in grey to indicate the location of non-methylated DNA (Long et al. 2013b). Above the sequencing traces individual genes are shown with the arrow indicating the transcription start site and vertical black lines corresponding to exons. CpG islands (CGI) are shown as green bars under the traces.

Using recombination-mediated cassette exchange, Lynch et al. (2012) showed that, although a high density of non-methylated CpG dinucleotides is sufficient for vertebrate Polycomb recruitment, there is a competition between Polycomb complex binding and transcriptional activation. To address whether loss of RING1B and SUZ12 is due to abrogation of KDM2B capacity to bind non-methylate CGI DNA or is an indirect consequence of increased gene expression following ZF-CxxC domain deletion, RNA-Seq analysis was carried out in the untreated Kdm2b$^{fl/fl}$ cells and after 72hrs of tamoxifen treatment. When significant fold changes in RNA expression of RING1B peak-associated genes were plotted together with

CHIP-Seq fold changes of KDM2B, RING1B and SUZ12 occupancy, it was apparent that only a modest increase in gene expression occurs at some of the polycomb targets showing reduced binding of both RING1B and SUZ12 (Figure 6.5, panel A). Furthermore, when RING1B binding fold changes were compared either with the level of SUZ12 alteration or changes in gene expression as determined by RNA-Seq, it was clear that, while a high correlation was observed between the magnitude of RING1B and SUZ12 change (Figure 6.5, panel B), no such correlation was evident between the level of gene expression change and the magnitude of RING1B loss (Figure 6.5, panel C). As this minor upregulation of gene expression is insufficient to explain the more widespread loss of PRC1 and PRC2 from target sites genome-wide, it was concluded that reduction in RING1B and SUZ12 occupancy was specifically a result of defective KDM2B-dependent targeting of PRC1 and not a consequence of massive transcriptional activation competing with binding of Polycomb complexes.

The essential contribution of KDM2B-dependent targeting to polycomb domain formation *in vivo* is further supported by the observation that attempts to generate mice homozygous for KDM2B ZF-CxxC DNA binding domain deletion have been unsuccessful and do not result in viable offsprings (Blackledge, Farcas, Kondo et al., *submitted*). Furthermore, the majority of Kdm2b$^{wt/\Delta ZF\text{-}CxxC}$ heterozygous mice die before 10 days postnatal. Interestingly, skeletal preparations from heterozygous mice indicates that they exhibit significant homeotic transformations, reminiscent of phenotypes observed for classic polycomb mutations in mice (Akasaka et al. 1996; van der Lugt et al. 1996; Akasaka et al. 2001; Isono et al. 2005). The striking reduction in PRC1 and PRC2 occupancy after deletion of KDM2B DNA binding domain, together with the strong homeotic transformations seen in Kdm2b$^{wt/\Delta ZF\text{-}CxxC}$ heterozygous mice suggests that the ability of KDM2B to target PRC1 has an essential contribution to Polycomb group protein function.

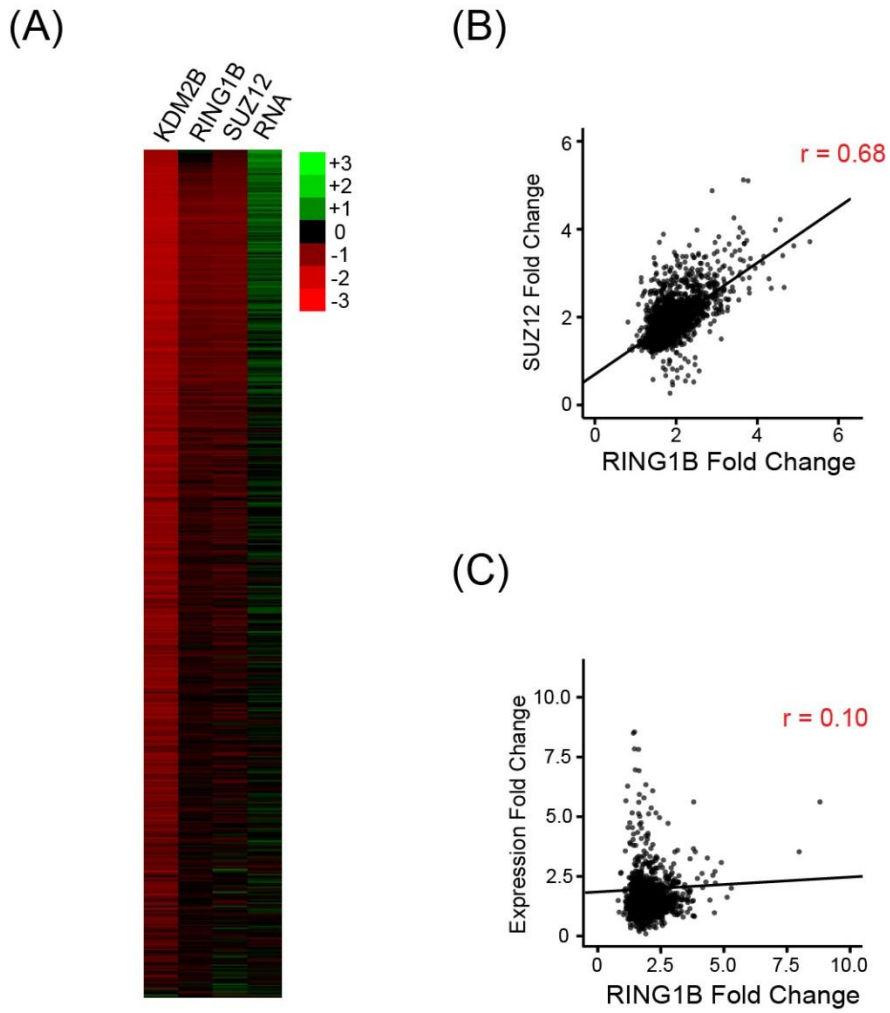


Figure 6.5. The reduction in RING1B and SUZ12 occupancy in the absence of KDM2B ZF-CxxC DNA binding domain is independent of large changes in target gene expression.

(A) Log2 fold changes in normalised read counts comparing the ChIP-seq and RNA-seq signal prior to and following 72hrs tamoxifen treatment of Kdm2b $^{fl/fl}$ ES cells. RNA-seq expression data is presented for the gene nearest each RING1B peak, as determined from three biological replicates. A widespread loss of KDM2B, RING1B, and SUZ12 is detected at these sites, with only a modest up-regulation of gene expression at genes showing the highest loss of PRC1 and PRC2.

(B) Scatter plot comparing the fold change of RING1B and SUZ12 at the RING1B sites that show significant changes after tamoxifen-mediated ablation of KDM2B ZF-CxxC domain. This indicates a high correlation (r=0.68) between the magnitude of RING1B and SUZ12 change suggesting that loss of PRC1 may be mechanistically linked with loss of PRC2.

(C) Scatter plot comparing the fold change of gene expression to the fold change in RING1B occupancy at sites that show significant RING1B alterations in response to tamoxifen-mediated ablation of KDM2B ZF-CxxC domain. There is no correlation (r=0.10) between the level of gene expression change as determined by RNA-Seq and the magnitude of RING1B change. The bioinformatics analysis used to generate this figure was performed by Mr. Hamish King (Klose lab) with help from Dr. Rob Klose.

6.2. A CpG-free *de novo* protein targeting assay

Disruption of the capacity of KDM2B to specifically bind CGIs results in a reduction of RING1B and SUZ12 occupancy at polycomb targets genome-wide, suggesting that KDM2B provides an essential link between recognition of non-methylated CpGs and polycomb recruitment. However, the association between KDM2B and RING1B represents just one example of a variant PRC1 complex, specifically characterized by the PCGF1 subunit. At natural CGI-associated polycomb targets, one can envisage that a multitude of Polycomb targeting mechanisms converge, such as the hierarchical PRC2-dependent recruitment of canonical PRC1, or targeting via the additional PRC1 variant complexes. As the connections between these alternative recruitment strategies are poorly defined, interpreting the results of disrupting one pathway may not necessarily be sufficient to specifically define its contribution to Polycomb function.

Therefore, to gain a more detailed understanding of PRC1 nucleation mechanisms *in vivo*, an artificial targeting strategy based on the high affinity and specificity of tetracycline repressor protein (TetR) for its cognate short DNA sequence binding site Tet operator (TetO), which was designed and built by Dr. Neil Blackledge, was used . Importantly, this tethering system allows one to study polycomb nucleation at sites devoid of the intrinsic complexity of natural polycomb targets.

To this end, a cassette containing 14 TetO sites interspersed with random CpG-free 23 bp linkers was first synthesized using GeneArt Gene Synthesis Service (Invitrogen) (Figure 6.6, panel A). To minimize position-effect influence, the TetO array cassette was amplified with homology arms and inserted by recombineering into a ~170 kilobase human gene desert bacterial artificial chromosome (BAC no RP11-419E6, Source Bioscience), with low GC-percentage, no CpG-islands and devoid of any identifiable genes or chromatin modifications. The BAC was modified to also contain flanking Tol2 sequences necessary for transposition (Kawakami 2007) and a Kanamycin/Neomycin(G418) selectable marker.

The TetO-containing BAC was transfected into mouse embryonic stem cells together with the helper plasmid capable of expressing the Tol2 transposase. Tol2-mediated transposition therefore allowed

single copy insertion of the TetO array cassette into mouse embryonic stem cells. Following G418 selection, single ES clones were isolated and screened for stable integration of the TetO array cassette. The precise genomic insertion site for the transposable element in the ES cloned used throughout this study was determined by splinkerette PCR (Potter and Luo 2010) and mapped to chromosome 8, resulting in a single stable integration of the TetO sequence array into chromatin surrounded on either side by greater than 75 Kb of inert human DNA (Figure 6.6, panel B).

By fusing a protein of interest to the bacterial TetR DNA binding domain, it can be directly targeted to the Tet-operator sites. For this purpose, the pCAG-Ires-puro mammalian expression vector was modified to contain the N-terminal Flag/2xStrepII tag, followed by 3x SV40 nuclear localization signals (NLS), the bacterial TetR DNA binding domain and LIC-compatible sites for efficient insertion of target proteins. The construct expressing the fusion between TetR and the protein of interest was transfected into mouse ES cells containing the stable integration of the TetO array cassette. Single ES clones were selected and screened for stable expression of the TetR-fusion protein in a TetO-array integration background.

Using the Flag/2xStrepII (FS2) antibody generated in this study (Chapter 5), ChIP-qPCR analysis allows the detection of fusion protein occupancy at the TetO and at various positions in the surrounding chromatin (Figure 6.6, panel C). Furthermore, as neither the TetO nor the surrounding inert DNA are subject to significant chromatin modifications, this TetR-dependent system enables one to directly assess the effects of occupancy of various fusion proteins.

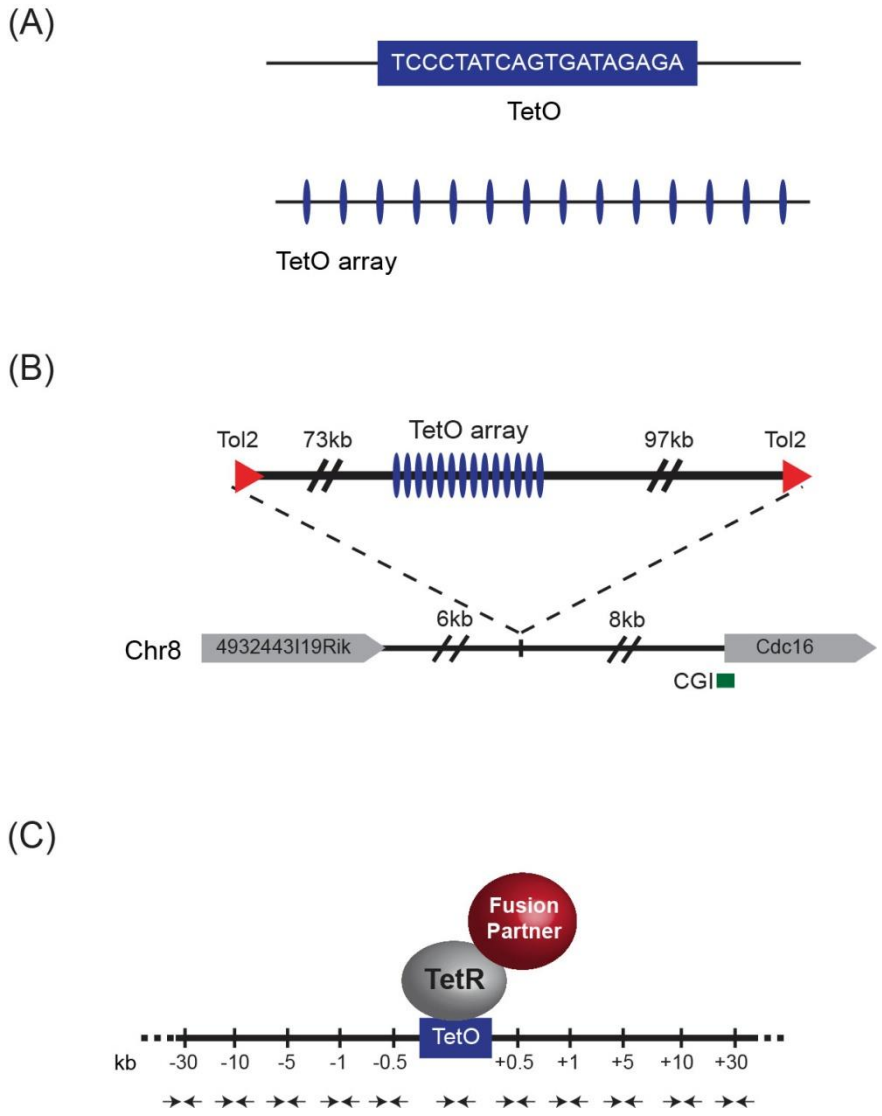


Figure 6.6. A *de novo* CpG-free tethering assay to study the effects of various proteins occupancy in mouse embryonic stem cells.

(A) Schematic illustrating the Tet-operator (TetO) CpG-free DNA sequence. 14 repeats of TetO sites (shown in blue) interspersed with random CpG-free 23bps linkers were commercially synthesized.

(B) Schematic depiction of the human gene desert BAC region containing the TetO array (blue) flanked by Tol2 integrase sites (shown as red triangles). Tol2-mediated integration enabled single copy insertion of the TetO array in mouse ES cells at a location on chromsosome 8. The TetO array is surrounded by more than ~170kbs of inert human DNA.

(C) The effects of TetR-dependent targeting of various fusion partners to the TetO array can be investigated by ChIP-qPCR. The primer locations of different positions in the TetO and surrounding chromatin (in kb) selected for analysis are indicated as forward and reverse black arrows.

6.3. Tethering of KDM2B is sufficient to recruit PRC1 and PRC2 and lead to polycomb domain formation *in vivo*

Taking advantage of the CpG-free TetR-dependent targeting system developed by Dr. Neil Blackledge, the first question addressed was whether KDM2B is sufficient to recruit RING1B to a genomic locus. This question was originally investigated by Dr. Neil Blackledge and later validated by my experiments. I will therefore only present the results obtained for my ChIP-qPCR analysis. One advantage of this *de novo* targeting system is that TetR and its associated fusion protein will only bind to the TetO array. For this purpose, the DNA-binding capacity of KDM2B was first disrupted by a specific point mutation in the ZF-CxxC DNA binding domain (K643A), as described in greater detail in the previous chapter (Chapter 5). This was carried out to prevent reduced TetO occupancy of the TetR-KDM2B fusion protein, due to KDM2B-mediated binding to CGIs genome-wide. A mouse ES clone with stable integration of the TetO array cassette and stably expressing TetR fused to full-length ZF-CxxC mutant human KDM2B was established. A cell line expressing only TetR was generated in parallel and used as control for subsequent ChIP-qPCR experiments. ChIP-qPCR results are presented as either percentage of input or as relative enrichment compared to a natural polycomb target site (i.e. *Bmp7* CGI-promoter).

ChIP with the FS2 antibody recognizing the tet repressor revealed that both TetR alone and TetR-KDM2B fusion protein occupy TetO and approximately 1kb around it, but diminish to background levels in the flanking regions (Figure 6.7, panel A). Importantly, only KDM2B targeting resulted in the recruitment of RING1B, despite lower levels of TetO occupancy for the TetR-KDM2B fusion compared to the TetR alone. This observation was not due to differences in ChIP efficiency levels between TetR and TetR-KDM2B fusion cell lines, which had comparable enrichment levels for all factors tested at the endogenous *Bmp7* locus (Supplementary Figure S8). Therefore, via its association with RING1B and PCGF1 in a variant PRC1 complex, KDM2B tethering is sufficient to recruit RING1B to a CpG-free genomic locus, albeit at lower levels than an endogenous polycomb target site. However, the levels of RING1B-mediated H2AK119ub1 placed in the KDM2B tethered cell line are significantly higher than those at the control *Bmp7* locus. This observation suggests that no linear relationship can be established between the levels of RING1B

occupancy and the histone modification it catalyzes. Alternatively, it could imply that at classic polycomb targets, mechanisms exist which regulate the levels of H2AK119ub1, such as histone deubiquitinases. The reduction in H2AK119ub1 levels detected at the TetO site is most probably due to nucleosome depletion at the TetO array, as shown later by analysing histone H3 enrichment levels (Figure 6.9).

ChIP-Seq experiments in the Kdm2b$^{fl/fl}$ and tamoxifen-treated cells showed that deletion of KDM2B ZF-CxxC DNA binding domain leads to reduced occupancy of both PRC1 and PRC2 at polycomb targets genome-wide. As the loss of SUZ12 was hypothesized to result from impaired KDM2B-dependent targeting of RING1B, the possibility that PRC1 could lead to PRC2 occupancy was tested in the TetR-KDM2B fusion cell line. When the levels of PRC2 component EZH2 were investigated by ChIP-qPCR, a striking enrichment was observed as a result of TetR-KDM2B targeting but not in the TetR control cell line (Figure 6.7, panel B). Surprisingly, very little EZH2 was detected at the Tet-operator but instead peaked at approximately 1 kb around TetO in the surrounding chromatin.

The reasons for this depletion at the TetO remain unclear, although one simple explanation could be that the TetO array is heavily occupied by TetR fusion protein and associated interaction partners, therefore impeding the binding of other factors. Alternatively, nucleosome depletion and the resulting low levels of H2AK119ub1 at the TetO site could be responsible for this unique EZH2 chromatin binding profile, as described in more detail later. Interestingly, as was the case for PRC1, EZH2 also appears to spread extensively, showing higher than background levels even at 10kb surrounding the TetO, although the mechanism for this spreading remains unclear. Importantly, KDM2B-dependent recruitment of PRC1 also resulted in deposition of EZH2-mediated H3K27me3, suggesting that KDM2B is sufficient to recruit both PRC1 and PRC2 and establish functional polycomb domains *in vivo*. The profile of H3K27me3 could suggest that RING1B and H2AK119ub1 spread beyond the otherwise very punctate peak of FS2-TetR/KDM2B potentially via a CBX-dependent hierarchical recruitment pathway.

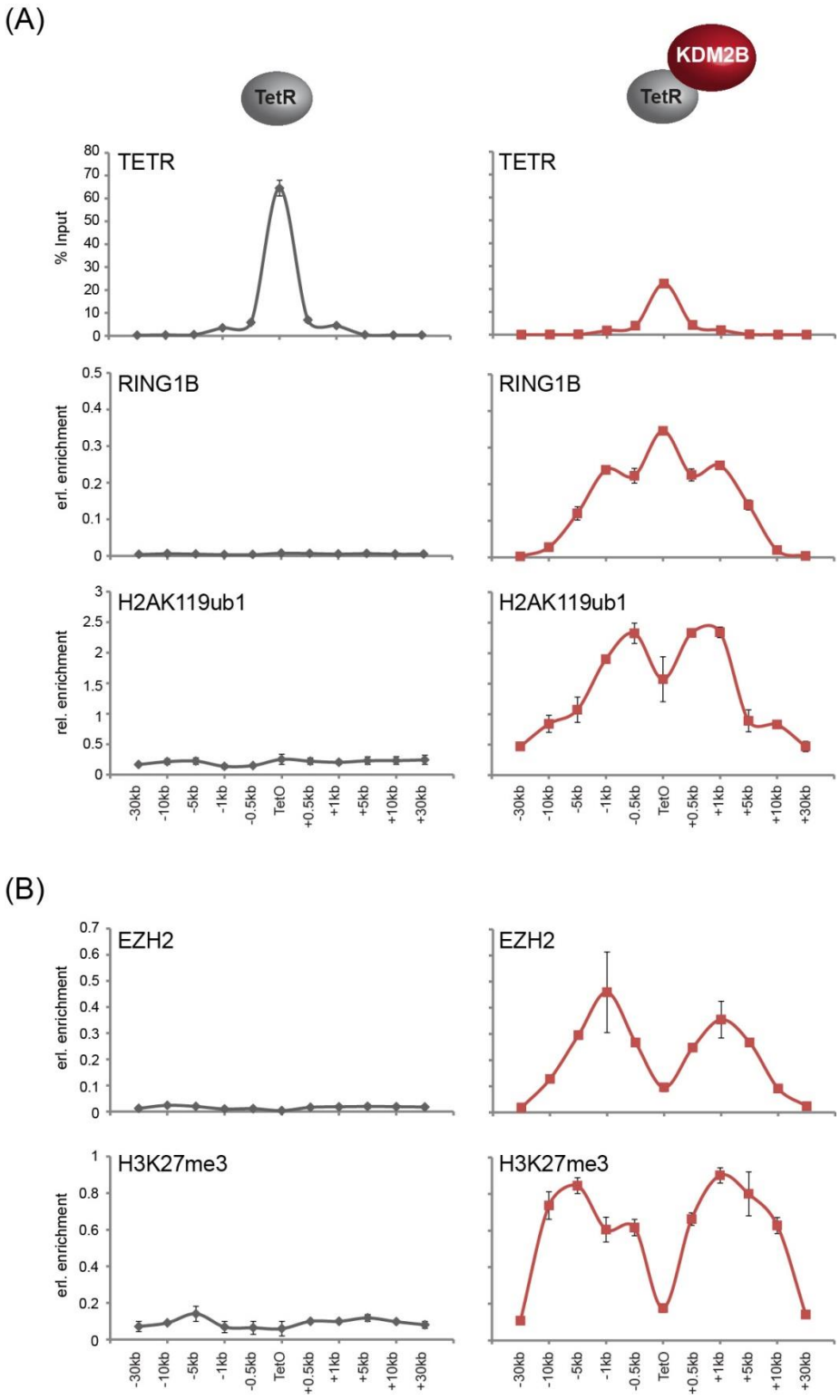


Figure 6.7. KDM2B-dependent targeting of PRC1 is sufficient to drive PRC2 recruitment and establish polycomb domains *in vivo*.

(A) ChIP-qPCR analysis of TetR, RING1B and H2AK119ub1 occupancy at the TetO and surrounding chromatin in the TetR only (shown in grey) and KDM2B fusion (shown in red) cell lines. Error bars represent SEM of two biological replicates. KDM2B-tethering is sufficient to recruit RING1B and mediate H2AK119ub1-deposition at this artificial chromosomal location. ChIP-qPCR results are presented as either percentage of input or as relative enrichment normalized to a natural polycomb target site (i.e. *Bmp7* CGI-promoter).

(B) ChIP-qPCR analysis EZH2 and H3K27me3 occupancy at the TetO and surrounding chromatin in the TetR only (shown in grey) and KDM2B fusion (shown in red) cell lines. Error bars represent SEM of two biological replicates. KDM2B-dependent targeting of RING1B mediates recruitment of PRC2 and H3K27me3 deposition.

6.4. KDM2B-mediated recruitment of PRC1 and PRC2 is PCGF1-dependent

The ability of KDM2B to recruit polycomb group proteins was hypothesized to be due to its association with PCGF1 to form a variant PRC1 complex. Even though PCGF1 was found by LC-MS/MS analysis to be unique to this complex (Chapter 4), the possibility that other undefined protein interactions or additional targeting mechanisms could be responsible for KDM2B-mediated PRC1 recruitment cannot be disregarded. To address whether PCGF1 is the bridging protein between KDM2B and RING1B targeting, I employed a shRNA-based approach to specifically deplete PCGF1 in the TetR-KDM2B cell line. A scrambled shRNA control and the shRNA targeting PCGF1 were cloned into pLKO.I/blasticidin constitutive expression vector. Single clones were isolated following lentiviral-based infection of TetR-KDM2B cell line and blasticidin selection, and the efficiency of PCGF1 knockdown was tested by qRT-PCR and Western blot analysis.

In one PCGF1 knockdown clone, approximately 80% knockdown was achieved at the RNA level as assessed by qRT-PCR (Figure 6.8, panel A) and clear depletion of the protein was detected when assessed by Western blot analysis using the PCGF1 antibody generated in this study (Figure 6.8, panel B). Importantly, PCGF1 depletion did not affect the TETR, RING1B and EZH2 protein levels, permitting an examination of the effect that PCGF1 knockdown has on KDM2B-mediated PRC1 recruitment (Figure 6.8, panel B). In agreement with the Western blot results, depletion of PCGF1 did not affect the occupancy of the TetR-KDM2B fusion protein at the TetO and immediate flanking region (Figure 6.8, panel C), but did result in a drastic reduction in PCGF1 targeting mediated by TetR-KDM2B fusion protein (Figure 6.8, panel D).

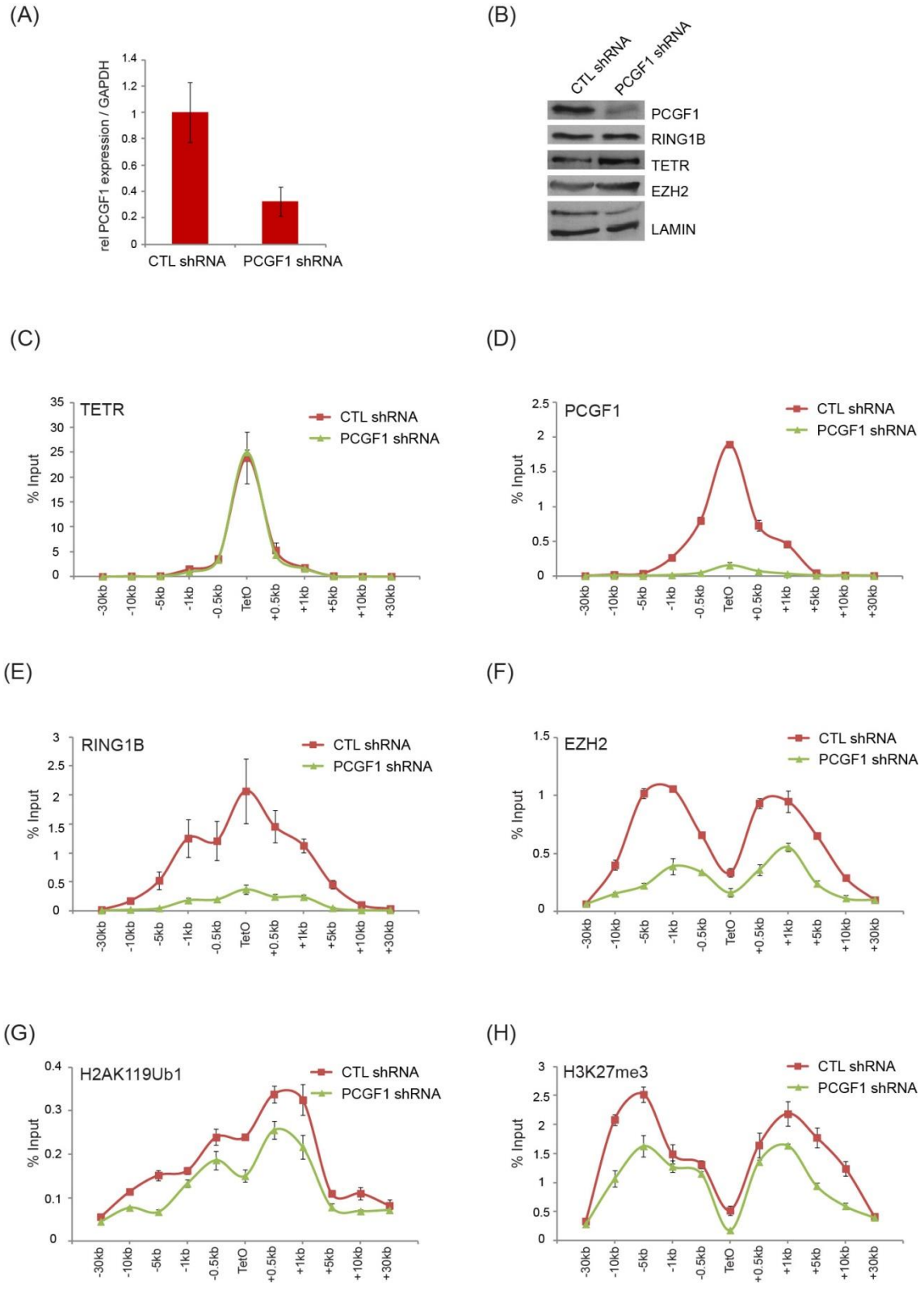


Figure 6.8. KDM2B-mediated recruitment of PRC1 and PRC2 is dependent on its association with PCGF1/PRC1 variant complex.

(A) qRT-PCR analysis showing reduction of PCGF1 at the mRNA level in the knockdown (PCGF1 shRNA) but not in the scrambled control TetR-KDM2B cell line (CTL shRNA). Error bars represent SEM of two biological replicates.

(B) Western blot analysis on nuclear extract showing that PCGF1 knockdown depletes PCGF1 but does not affect TetR, RING1B and EZH2 protein levels. Lamin B probing was used to indicate equal loading.

(C) – (H) ChIP-qPCR analysis of TetR, PCGF1, RING1B, EZH2, H2AK119ub1 and H3K27me3 levels in the TetR-KDM2B cell line expressing either PCGF1 shRNA (shown in green) or scrambled control (shown in red). PCGF1 depletion does not impair TetR-KDM2B protein occupancy, but clearly reduces the levels of PCGF1, RING1B and EZH2 binding at the TetO and surrounding chromatin. Error bars represent SEM of at least two biological replicates. This suggests that the ability of KDM2B to recruit PRC1 and PRC2 is dependent on its capacity to associate with PCGF1.

Interestingly, when PRC1 and PRC2 occupancy were assessed, a clear reduction in RING1B and EZH2 was observed at the TetO and surrounding chromatin in the TetR-KDM2B cell line expressing the PCGF1-specific shRNA compared to the scrambled control (Figure 6.8, panels E and F). Furthermore, PCGF1 depletion resulted in overall lower levels of H2AK119ub1 and H3K27me3, although this loss was not as dramatic as that observed for their respective polycomb chromatin-modifying enzymes (Figure 6.8, panels G and H). The striking loss of RING1B binding in response to PCGF1 knockdown suggests that the ability of KDM2B to drive polycomb domain formation at an artificial CpG-free genomic locus is mediated through KDM2B-dependent recruitment of the PCGF1/PRC1 complex and subsequent PRC2 nucleation.

ChIP-Seq analysis showed that deletion of KDM2B ZF-CxxC DNA binding domain led to loss of both RING1B and SUZ12 binding to polycomb targets genome-wide (Figure 6.3). However, one possibility in these types of experiments is that the concomitant loss of PRC2 in response to reduced PRC1 targeting is an indirect result due to a potential disruption of the chromatin state at natural polycomb targets. The use of a *de novo* tethering assay devoid of the complexities of polycomb sites suggested that targeting KDM2B is sufficient to drive PRC1 and PRC2 chromatin occupancy. Furthermore, loss of PRC1 recruitment at the TetO and surrounding chromatin by PCGF1 depletion also resulted in reduced EZH2 binding levels, indicating that PRC1 could play a role in mediating PRC2 nucleation. As no direct protein interaction was detected between KDM2B and PRC2 components and no known interactions exist between the two main polycomb complexes, the most plausible explanation is that PRC2 recruitment is responsive to a specific chromatin configuration determined by PRC1. Preliminary experiments in the Klose lab showing that deposition of H2AK119ub1 is sufficient to drive PRC2 recruitment suggest that

this is indeed the case (Blackledge, Farcas, Kondo et al., *submitted*). The mechanisms responsible for this PRC1-mediated PRC2 recruitment are however far from clear, although one interesting possibility is that H2AK119ub1 can be specifically recognized by a PRC2 component, in a manner similar to how CBX-dependent recognition of H3K27me3 drives binding of the canonical PRC1 complex.

6.5. All six PCGFs have the capacity to recruit RING1B, but only variant PRC1 complexes place significant levels of H2AK119ub1 and mediate PRC2 recruitment

KDM2B occupancy on DNA is sufficient to target PRC1 and mediate PRC2 recruitment, in a manner dependent on its ability to associate with PCGF1 and form a variant PRC1 complex. However, in mammals six major variant PRC1 complexes exist, each defined by a distinct PCGF subunit, a RING1A/RING1B ubiquitin ligase, RYBP or its homologue YAF2 and a unique set of associated polypeptides (Gao et al. 2012). Furthermore, PCGF2 and PCGF4 can form independent canonical PRC1 complexes, in which the catalytic subunits are associated with CBX and PH proteins. Therefore, it was important to understand whether the ability of KDM2B/PRC1 to recruit PRC2 is unique to this PCGF1-variant complex, or if the other PCGF-defined PRC1 complexes share this feature.

For this purpose, I fused all six mouse PCGF subunits (PCGF1 to PCGF6) to the TetR and stably expressed them in the mouse ES cell line with the TetO array cassette integration. A cell line expressing TetR alone was generated in parallel and used as control for the subsequent ChIP-qPCR studies. Western blot analysis showed relatively equal expression of the fusion proteins (Figure 6.9, panel A). In agreement with the inclusion of KDM2B in a PCGF1/PRC1 variant complex, only tethering of PCGF1 subunit resulted in clear recruitment of endogenous KDM2B to the TetO and immediate flanking regions (Figure 6.9, panel B), suggesting that the TetR-PCGF1 fusion is functional *in vivo*. ChIP-qPCR analysis shows that all six TetR-PCGF fusion proteins bind to the TetO array at comparable levels and diminish to background levels in the TetO surrounding chromatin (Figure 6.9, panel C). Furthermore, in agreement with the capacity of PCGFs to directly interact with RING1B via their RING-finger domains, all PCGF-TetR fusions

have the capacity to recruit RING1B. Interestingly, PCGF1 through PCGF5 targeted RING1B at significantly higher levels than regularly found at an endogenous polycomb target. This is in contrast with the RING1B occupancy detected as a result of TetR-KDM2B targeting, which was lower than that observed at a natural polycomb site (Figure 6.7). Although this could potentially indicate that KDM2B is a substoichiometric component of the PCGF1/PRC1 variant complex, it may also be due to KDM2B associating less tightly with RING1B than the direct interaction between PCGF1 and the PRC1 E3 ubiquitin ligase. TetR-PCGF6 occupancy resulted in relatively low RING1B enrichment. Although the reasons behind this are unclear, it should be mentioned that it was difficult to establish cell lines stably expressing the PCGF6 fusion protein, indicating that overexpression of TetR-PCGF6 may have deleterious effects to stably expressing cells.

Importantly, when H2AK119ub1 deposition was analysed, only the PCGF proteins which exclusively form non-canonical PRC1 complexes (PCGF1, 3, 5 and 6) were able to place significant levels of this histone mark , while canonical PCGF2 and PCGF4 did not lead to H2AK119ub1 deposition (Figure 6.9, panel C). This differential enzymatic activity between canonical and non-canonical PCGFs occurs despite comparable levels of RING1B nucleation at the TetO, as well as similar H2AK119ub1 enrichment at the endogenous *Bmp7* polycomb locus used for normalization (Supplementary Figure S9). The reduced occupancy of H2AK119ub1 at the TetO is due to nucleosome depletion at that site, as suggested by the similar reduction observed for histone H3.

If PRC1-dependent H2AK119ub1 deposition is responsible for mediating PRC2 recruitment, one would therefore expect the canonical PCGF2 and PCGF4 fusions should have reduced PRC2 occupancy and H3K27me3 levels compared to their non-canonical counterparts. In support of this hypothesis, ChIP-qPCR analysis showed that PCGF1,3,5 and 6 can efficiently recruit EZH2 and SUZ12 PRC2 subunits and H3K27me3, while the canonical PCGF subunits (PCGF2/PCGF4) showed low levels of PRC2 enrichment (Figure 6.9, panel D). Importantly, this is not due to experimental differences in ChIP efficiencies, as the TetR-PCGF fusions show comparable PRC2 and H3K27me3 enrichment levels at an endogenous polycomb target (Supplementary Figure S9). Therefore, it seems that despite their capacity to dimerize

with RING1B and form distinct PRC1 complexes, the PCGFs differ in their ability to place H2AK119ub1 and mediate PRC2 recruitment.

Contradictory evidence exists in literature regarding potential differences in the ability of canonical and non-canonical PRC1 complexes to catalyse H2AK119 mono-ubiquitylation. For examples, Tavares et al (2012) reported that RYBP-PCGF2/PRC1 and CBX-PCGF2/PRC1 complexes efficiently mono-ubiquitylate H2A at roughly equivalent levels *in vitro*. In contrast, Gao et al. (2012) showed that RYBP-containing PCGF4/PRC1 complexes exhibited stronger enzymatic activity than those containing PH2 and either CBX2 or CBX8, suggesting that *in vitro* RYBP stimulates the activity of RING1B on H2AK119. The results obtained here suggest that the canonical CBX-associating PCGF2 and 4 subunits are less capable of H2AK119ub1 deposition than the variant PCGF complexes. Furthermore, it shows that the ability of KDM2B to target PRC1 and mediate PRC2 recruitment is not a unique property of PCGF1, but that all variant PRC1 complexes can recruit PRC2 and drive polycomb domain formation *in vivo*. This important observation highlights the need to understand whether the other non-canonical PCGFs can also directly interpret and access natural polycomb targets and to define the mechanisms responsible for their target site binding specificity. Such an understanding could provide valuable insights into how polycomb domains are initiated *de novo* and potentially shed light on the identity of the targeting factors responsible for polycomb domain establishment during development.

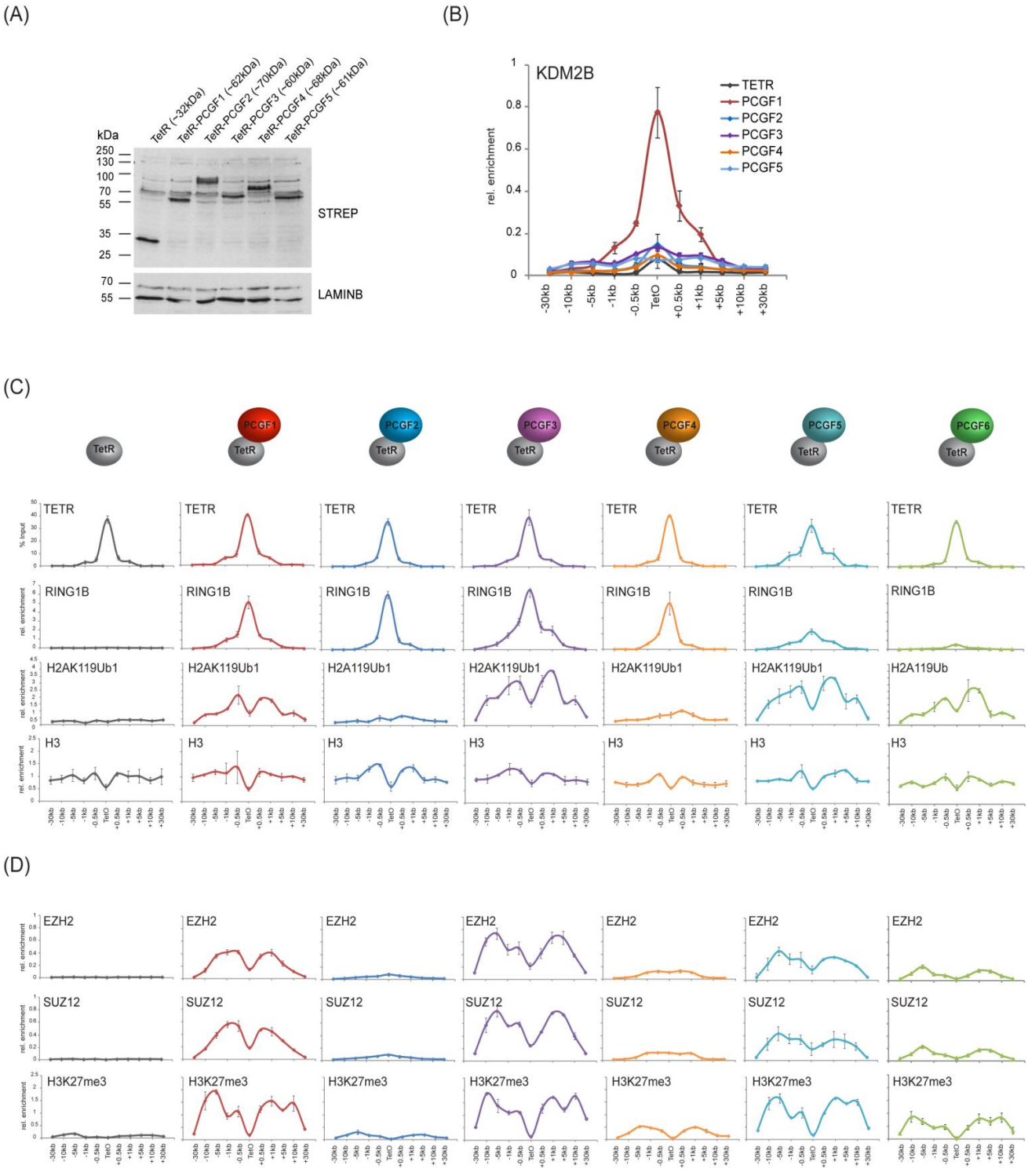


Figure 6.9. Non-canonical PRC1 complexes have the capacity to recruit PRC2 and mediate H3K27me3 deposition.

(A) Western blot analysis of the TetR-PCGF1 through PCGF5 fusion cell lines, indicating roughly equal expression levels. The Western blot was carried out on 25 µg of nuclear extract, using a Strep antibody recognizing TetR. Lamin B was used as loading control, and the predicted molecular weight of the different fusion proteins are indicated in brackets.

(B) ChIP-qPCR analysis of KDM2B enrichment across the TetO-containing locus in the TetR-PCGF fusion cell lines, indicating that only PCGF1 efficiently recruits KDM2B. The positions in the TetO surrounding chromatin are

indicated below the graph in kb. Error bars represent SEM of biological duplicates. Enrichment of KDM2B is shown normalized to an endogenous polycomb locus (i.e. *Bmp7*).

(C) ChIP-qPCR analysis for TetR, RING1B, H2AK119ub1 and histone H3 at the TetO and surrounding chromatin in the TetR-PCGF fusion cells lines. The fusion protein identity is indicated above each panel. Although all PCGFs recruit RING1B, only non-canonical PCGF1, 3, 5 and 6 are competent at H2AK119ub1 deposition. Error bars represent SEM of biological duplicates.

(D) ChIP-qPCR analysis of EZH2 and SUZ12 PRC2 components and H3K27me3 at TetO and flanking chromatin in the TetR-PCGF fusion lines. In agreement with the limited capacity of canonical PCGFs to place H2AK119ub1, only PCGF1, 3, 5, and 6 can efficiently recruit PRC2 and direct H3K27me3 placement. Error bars represent SEM of biological duplicates.

6.6. Summary and discussion

Tamoxifen-induced deletion of the ZF-CxxC DNA binding domain removes KDM2B from chromatin. In agreement with previous results obtained using a shRNA-based knockdown approach (Farcas et al. 2012; He et al. 2013; Wu et al. 2013b), loss of KDM2B binding leads to reduced RING1B occupancy at polycomb targets genome-wide. Interestingly, when PRC2 chromatin binding profile was investigated, a striking reduction in SUZ12 occupancy was detected primarily at sites which also lose PRC1. Furthermore, deletion of the KDM2B ZF-CxxC domain results in embryonic lethality in homozygous mice and strong homeotic transformations in heterozygotes (Blackledge, Farcas, Kondo et al., *submitted*), suggesting that the ability of KDM2B to recognize non-methylated CpG island DNA is essential for polycomb domain formation and normal embryonic development.

As no direct interaction was detected between KDM2B and PRC2 components, the striking reduction in SUZ12 occupancy in response to deletion of KDM2B DNA binding domain was hypothesized to be a consequence of impaired KDM2B-dependent recruitment of the PCGF1/PRC1 variant complex, which could potentially disturb via an uncharacterized pathway the feedback mechanisms reinforcing PRC2 occupancy at natural polycomb targets or potentially directly impair PRC2 recruitment. However, the intrinsic complexity and multitude of factors acting at natural polycomb sites *in vivo* make such an analysis difficult to interpret. The observation that tethering KDM2B to an artificial CpG-free genomic

locus is sufficient to recruit PRC1, lead to H2AK119ub1-deposition and mediate PRC2 occupancy enabled one to separate between the complexity imparted by the underlying non-methylated DNA and polycomb domain formation *in vivo*. The observation that PRC1-occupancy also leads to PRC2 nucleation is striking and represents the first clear example of PRC2 acting downstream of PRC1.

By fusing all the six known mouse PCGF subunits to the TetR, it was shown that the capacity of PRC1 to recruit PRC2 is not limited to PCGF1/PRC1, but that it is a feature shared amongst variant PCGF/PRC1 complexes. Importantly, this appears to correlate with the ability of PCGF1, 3, 5, and 6 to form enzymatically active PRC1 complexes *in vivo* that are more capable of placing high levels of H2AK119ub1 than canonical PCGF2/4-PRC1.

This result raises the question of what is different about canonical PRC1 so as to impede its capacity to place high levels of H2AK119ub1. One simple explanation could be that fusion of PCGF2 and PCGF4 to the TetR domain impairs their biological function, although it seems highly unlikely that only these two PCGF subunits should be specifically affected by this. Another possibility which cannot be completely disregarded is that at natural polycomb targets, an uncharacterized factor could function to stimulate PCGF2/4-PRC1 complex catalytic activity, a factor which would not be present at the artificial CpG-free TetO array or surrounding inert DNA chromatin. An alternative explanation to why lower levels of H2AK119ub1 were detected for PCGF2/4-complexes could be that canonical PRC1 can specifically associate with an H2AK119-specific histone deubiquitinase (DUBs), although no such interaction has been identified up to this point. One clear difference between PCGF2/4 and the other PCGF subunits is their ability to associate with CBX and PH proteins to form canonical PRC1 complexes, responsive to PRC2-mediated H3K27me3 deposition.

Further complicating the differential ability of PCGFs to drive polycomb domain formation is the observation that PCGF2 and PCGF4 can also associate with RYBP or its homologue YAF2 and form variant PRC1 complexes. It remains unclear how the association with either CBX or RYBP is regulated in this case. It would be interesting to test whether tethering PCGF2/4 versus variant PCGFs results in CBX recruitment at the TetO and surrounding chromatin versus RYBP/YAF2 enrichment. Furthermore, if association with either RYBP or CBX proteins plays a role in the different H2AK119ub1-deposition levels

observed between canonical and variant PCGFs, it is tempting to speculate that directly tethering RYBP or CBX should reproduce this differential competence in H2AK119ub1 placement and subsequent PRC2 recruitment.

H2AK119ub1-deposition by variant PCGFs led to efficient PRC2 nucleation. As no direct interaction was reported between PRC1 and PRC2, this result would suggest that H2AK119ub1 may be responsible for mediating PRC2 binding. Currently, two models come to mind by which H2AK119ub1 could affect PRC2 nucleation. Firstly, PRC1 occupancy has been intensely associated with chromatin compaction (Francis et al. 2004; Grau et al. 2011). Although chromatin accessibility-type experiments show that polycomb targets are generally characterized by a more compact chromatin architecture (Bell et al. 2010), this was however never unambiguously demonstrated to be a direct consequence of H2AK119ub1 placement. Nevertheless, the discovery that more dense nucleosomal arrays are better substrates for PRC2 enzymatic activity (Yuan et al. 2012) could allow one to speculate that H2AK119ub1 could induce chromatin compaction, thereby facilitating PRC2 recruitment and polycomb domain formation. In principle, DNase-type studies could be carried out to test whether variant PRC1 recruitment results in a more compacted chromatin state at the TetO and flanking regions.

The second possibility is however much more intriguing and virtually unexplored. According to this model, a PRC2 component or associating factor could specifically recognize H2AK119ub1 and drive PRC2 recruitment, in a manner analogous but reversed to the classic H3K27me3-dependent hierarchical recruitment pathway. Interestingly, a crosstalk was found to exist between H2B ubiquitination and H3K4me3 which plays essential roles in coordinating diverse cofactors during transcription activation (Kim et al. 2013; Wu et al. 2013a). This observation suggested the possibility that a similar pathway may connect H2AK119ub1 and H3K27me3 to establish and maintain polycomb domains. Interestingly, the crosstalk between H2Bub and H3K4me3 is mediated via the winged helix (WH) motif in the ASH2L subunit of the MLL1/2 trithorax mammalian counterpart complex, motif which is conserved in the polycomb protein Additional sex combs (ASX) (Aravind and Iyer 2012; Sanchez-Pulido et al. 2012).

Work in Drosophila identified an association between ASX and the histone deubiquitinase CALYPSO to form the Polycomb Repressive Deubiquitinase (PR–DUB) complex (Scheuermann et al. 2010).

Interestingly, Drosophila mutants lacking PR-DUB show a strong increase in the levels of monoubiquitylated H2A, suggesting that PR-DUB specifically deubiquitylates the histone mark placed by PRC1. Although the mechanisms by which ASX proteins contribute to gene regulation are not understood, mice with reduced levels of ASX-like 2 (ASXL2) protein homologue also show loss of bulk H3K27me3, as well as locus-specific depletion of this mark (Lai and Wang 2013). Clearly further work is required to understand why PR-DUB is bound at polycomb targets. However, one possibility could be that in addition to ensuring a dynamic balance between H2A ubiquitylation and deubiquitylation, it also acts as a link between PRC1-mediated H2AK119ub1 deposition and PRC2 recruitment.

Another candidate which could potentially play a role in linking H2AK119ub1 to PRC2 nucleation may be the PCL protein. Interestingly, H3K27me2 is broadly distributed and accounts for 50–60% of total nuclear histone H3 (Peters et al. 2003). As EZH1/EZH2 is the sole mammalian H3K27 histone methyltransferases, this implies that PRC2 can promiscuously dimethylate chromatin presumably by a "hit-and-run"mechanism. However, H3K27me3 is exclusively found at polycomb target promoters, suggesting that a specific PRC2-associated factor can recognize the chromatin state of natural polycomb targets and facilitate the conversion from H3K27 di- to tri-methylation. Work in flies identified a distinct form of PRC2 that contains the Polycomb group protein polycomb-like (PCL) protein, shown to be required for generation of high-levels of H3K27me3 but dispensable for H3K27me1/me2 (Nekrasov et al. 2007). It is therefore tempting to speculate that mammalian PCL homologues could potentially recognize a specific chromatin configuration at polycomb targets, leading to PRC2 recruitment and H3K27me3. Mammalian PCL homologues PCL1 (PHF1) and PCL3 (PHF19) were shown to bind H3K36me3 (Brien et al. 2012; Cai et al. 2013). However, their ability to potentially also engage with H2AK119ub1 as an alternative mechanism of linking H2AK119ub1 recognition to PRC2 recruitment was never investigated.

In conclusion, work in this chapter shows that the ability of KDM2B to recognize non-methylated DNA is responsible for PRC1 recruitment and subsequent PRC2 nucleation at polycomb targets genome-wide, reinforcing the central role of CGIs in the formation of polycomb domains. Importantly, KDM2B-PCGF1/PRC1 is one example from multiple variant PRC1 complexes. The preliminary results of the

tethering experiments suggest that all variant PRC1 complexes have in principle the capacity to initiate polycomb domains *de novo*. Therefore, it will be critical to define whether the other non-canonical PCGFs also have the ability to directly interrogate the underlying DNA sequence at natural polycomb sites, or whether KDM2B-dependent targeting of PCGF1/PRC1 is the principal mechanism responsible for the initiation of polycomb domains during development. The differences between canonical and variant PRC1 in mediating PRC2 recruitment is intriguing, and could potentially indicate that these two classes of PRC1 complexes have different roles in polycomb function. A model according to which variant PRC1s are responsible for the initiation of polycomb occupancy, while canonical CBX/PH-containing PRC1 may function primarily in the stabilization and reinforcement of established polycomb domains will be described in the final Conclusions chapter.

7. Chapter seven - Conclusions and implications for future work

Genome-wide analysis of PRC1 and PRC2 occupancy in embryonic stem cells revealed that polycomb complexes specifically associate with a subset of CGI-promoters encoding developmental regulators to help maintain a silent transcriptional state (Boyer et al. 2006; Ku et al. 2008). However, the reasons behind this specific chromatin binding profile and the causality of non-methylated GC-rich DNA in recruitment of polycomb group proteins remained poorly defined. Furthermore, increasing examples of PRC2-independent PRC1 targeting suggested that the hierarchical recruitment model was insufficient to explain the co-occupancy of PRC1 and PRC2 at natural polycomb targets (Schoeftner et al. 2006; Tavares et al. 2012; Simon and Kingston 2013). Work in this thesis was aimed at investigating the functional role of the ZF-CxxC domain containing histone lysine demethylase KDM2B. The results obtained reveal that this protein specifically recognizes CpG islands genome-wide and plays an essential role in polycomb domain formation through its inclusion in a PRC1 variant complex. The main findings of each chapter will be summarized and the questions they raise, as well as the models proposed to describe the implications of these findings, will be briefly discussed in the next few pages.

The generation of a KDM2B-specific antibody enabled an investigation of its chromatin binding profile. ChIP-seq results indicated that the ZF-CxxC domain of KDM2B is functional *in vivo*, with the protein localizing at CGIs genome-wide, in a manner analogous to its homologue KDM2A (Blackledge et al. 2010), but also other ZF-CxxC containing proteins such as CFP1 (Thomson et al. 2010) or MLL1 (at least at a subset of CGIs) (Guenther et al. 2005) (Figure 7.1). Therefore, it seems that CGIs act as binding platforms for multiple chromatin-modifying activities, presumably as a mechanism to create a unique and accessible chromatin architecture at these genomic elements which distinguishes them from the surrounding genome. However, the fact that these distinct proteins contain ZF-CxxC domains does not necessarily imply that they bind CGIs with similar affinities or that they co-occupy these sites. It would

be of great interest to understand whether other factors contribute to the localization of these proteins and to define the interplay between them, for example by studying the effects of deleting one ZF-CxxC domain-containing protein on the chromatin-binding distribution of another.

One major unanswered question in the field of chromatin biology is how CGIs are kept in a non-methylated state. It is clear that CGIs do have the potential to become methylated, as suggested by aberrant promoter methylation and silencing of tumour suppressor genes in cancer. Therefore, the hypothesis is that either certain factors prevent their *de novo* methylation during development or recently, that TET proteins actively remove methylation from these sites. It will be interesting to address whether occupancy by ZF-CxxC domain proteins has any protective function on maintaining CGIs free from DNA methylation, perhaps by competing with DNMTs. However, the observation that deletion of CFP1 in mouse embryonic stem cells did not lead to increased DNA methylation at CGI-promoters would argue against such a function for ZF-CxxC proteins (Clouaire et al. 2012).

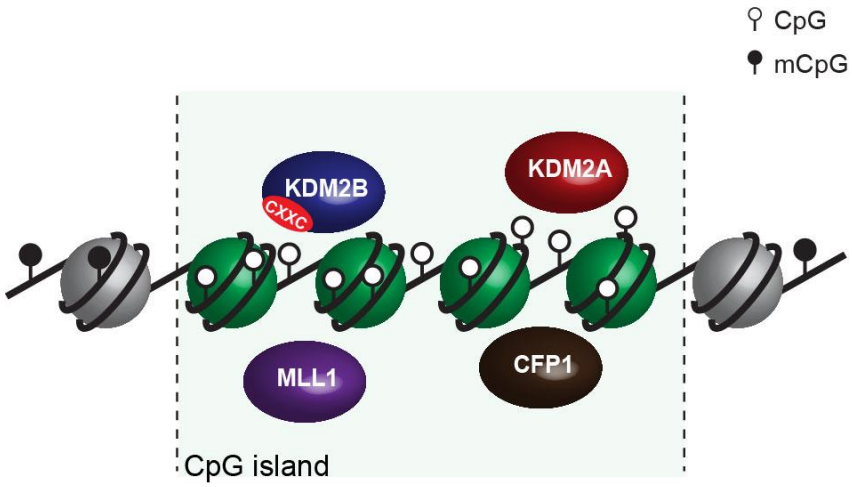


Figure 7.1. KDM2B binds CGIs genome-wide, in a manner analogous to other ZF-CxxC domain-containing proteins.

In ESCs, a subset of CGIs corresponding to genes with essential developmental regulatory roles was found to exist in a repressed chromatin state characterized by occupancy of polycomb repressive complexes 1 and 2 and their associated histone marks, H2AK119ub1 and H3K27me3, respectively.

Interestingly, ChIP-seq analysis performed in this study and others (Blackledge et al. 2010; Thomson et al. 2010) suggests that these Polycomb target CGIs are generally refractory to binding of ZF-CxxC proteins, potentially due to their reported more compacted chromatin state (Grau et al. 2011; Isono et al. 2013). However, results in chapters 3 and 4 indicate that KDM2B associates with PRC1 E3 ubiquitin ligase RING1B to form a variant PRC1 complex characterized by the PCGF1 subunit and absence of CBX or PH proteins. Furthermore, KDM2B was found to be preferentially enriched at PcG-occupied CGIs, in striking contrast to other ZF-CxxC proteins such as KDM2A or CFP1 (Figure 7.2).

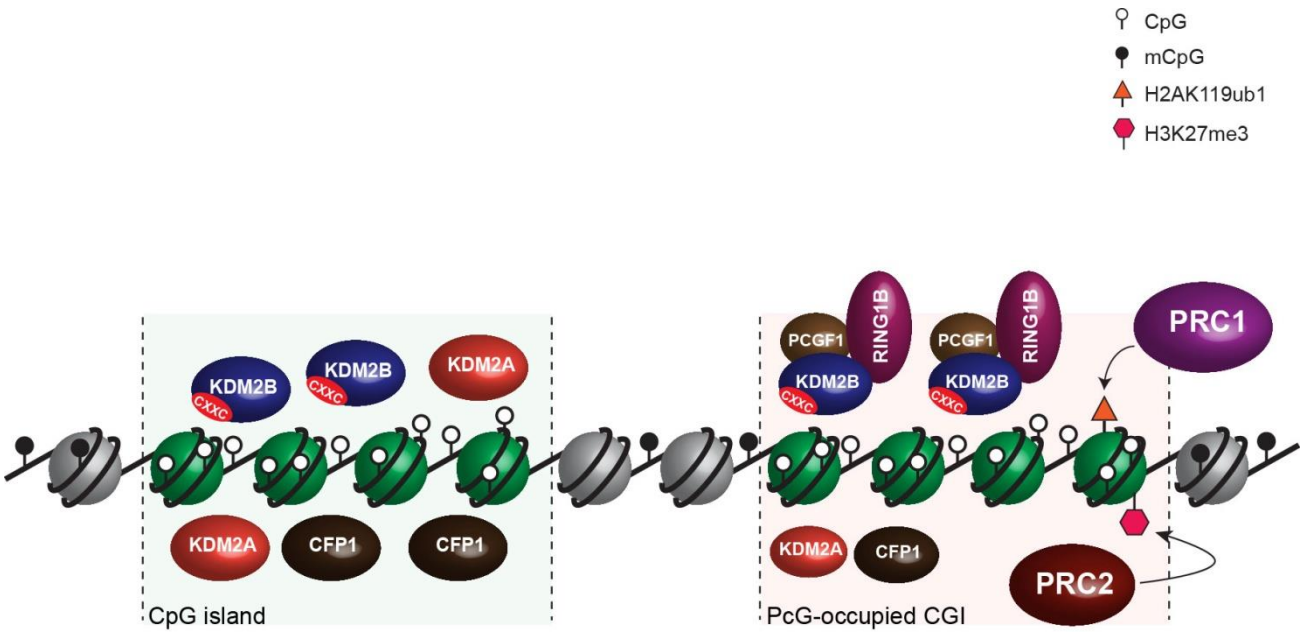


Figure 7.2. KDM2B forms a variant PRC1 complex and is preferentially enriched at Polycomb-target CGIs.

The reasons for which KDM2B is enriched at Polycomb targets remain undefined, especially considering that the ZF-CxxC DNA binding domain is highly conserved with that of KDM2A and therefore subjected to the same requirements in terms of non-methylated CpG dinucleotide accessibility. One possibility remains that, although the DNA binding domain is critical for KDM2B binding, its association with Polycomb proteins contributes to its stabilization at PcG-occupied CGIs. However, this suggestion awaits experimental testing, potentially by disruption of the residues or domains in KDM2B which mediate its

interaction with PcG subunits. One would then speculate that, in the absence of inclusion into a variant PRC1 complex, the chromatin binding profiles of KDM2A and 2B would be overlapping.

Given that KDM2B contains a ZF-CxxC DNA binding domain functional *in vitro* and *in vivo*, and considering its association with PRC1 members, one hypothesis which was tested was whether KDM2B plays a role in PRC1 recruitment to target loci. Indeed, as indicated by results in Chapter 5, depletion of KDM2B leads to reduced RING1B occupancy at PcG-associated CGIs. The most surprising observation however was the KDM2B-dependent widespread low-level association of RING1B with the majority of CpG island elements, in addition to the high-magnitude binding detected at classical Polycomb targets. In an attempt to define the functional relevance of these low-level magnitude RING1B peaks, the possibility that KDM2B enables PRC1 to interrogate all CGI-promoters for susceptibility to polycomb nucleation was proposed. This possibility is further supported by the wide-spread association of PCGF1 with CGIs detected in this study (see Chapter 5), but also by recent discoveries illustrating low-level occupancy of EZH2 at active promoters genome-wide (Kaneko et al. 2013b), therefore indicating that PRC2 also has the capacity to interrogate CGIs through an unknown mechanism.

It is interesting to mention that mammalian Trithorax group-related proteins CFP1 and MLL1/MLL2 also contain ZF-CxxC DNA binding domains, and H3K4me3 was detected at the majority of CGIs, regardless of the transcriptional status of the associated gene. Therefore, it appears that both PcG and TrxG proteins continuously “sample” CGIs. Establishment of either one or the other will most probably be directed by transcriptional output. According to this proposed model, active transcription reinforces TrxG binding and competes with PcG occupancy, although the molecular mechanisms of how this competition might occur remain to be clarified.

One possibility is that, through deposition of activating marks such as H3K4me3 or H3K36me3 shown to be inhibitory to PRC2-mediated histone methylation (Schmitges et al. 2011), polycomb binding is destabilized due to loss of H3K27me3-dependent reinforcement of PRC1 by PRC2. The transcription-directed reinforcement of TrxG is supported by ChIP-seq studies which show that, while H3K4me3 is found at the majority of CGI-promotes regardless of transcriptional status, it is nevertheless highly enriched at active genes (Guenther et al. 2007). Furthermore, studies using *Cfp1-/-* ES cells indicate that

transcription itself plays a key role in the targeting of CFP1 and H3K4me3 to active promoters, with only highly expressed genes showing reduction in H3K4me3 levels in response to Cfp1 deficiency (Clouaire et al. 2012).

As discussed in greater detail in Chapter 5, the absence of transcription or activating motifs could result in productive enrichment of polycomb binding and reinforcement of polycomb occupancy through positive feedback mechanisms between PRC1 and PRC2 favouring their mutual stabilization (Figure 7.3). It would be tempting to speculate that, given the competition between TrxG and PcG binding, disrupting either of the pathways responsible for H3K4me3 or H3K27me3 deposition should result in stabilisation of the opposing mark at CGI promoters. However, this would be an oversimplification of the model, as productive occupancy of either class of proteins is responsive to the transcriptional status of the associated gene. Therefore, to test the validity of this proposed model, one has to interfere directly with transcription and assess the effects of either activation or inhibition of this process on polycomb occupancy.

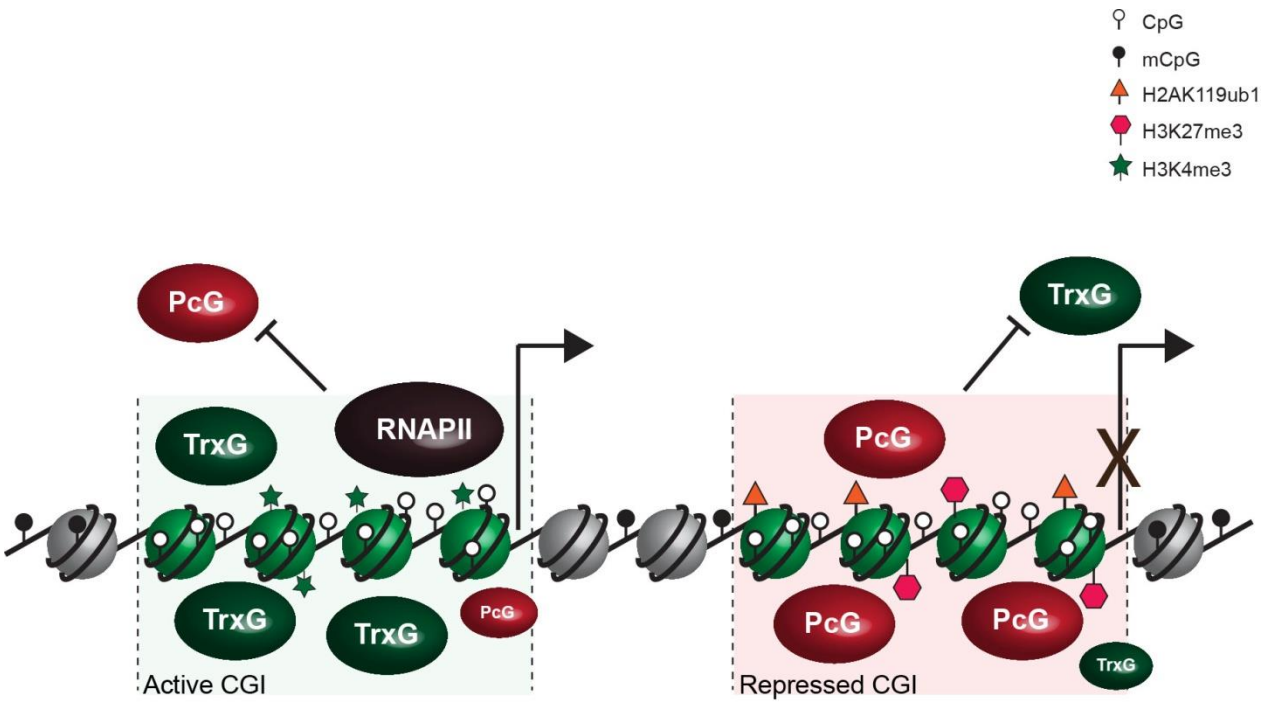


Figure 7.3. Polycomb and Trithorax group proteins sample CGIs and respond to the transcriptional status of the associated genes.

Although depletion of KDM2B resulted in decrease of RING1B binding at polycomb targets, the shRNA-based approach does not enable one to conclude that the ZF-CxxC domain of KDM2B is the principal mechanism by which this protein contributes to PRC1 occupancy. The observation that a specific genetic deletion which removes the DNA binding domain of KDM2B while leaving the rest of the protein intact also resulted in a significant loss of both RING1B and SUZ12 binding indicated that the capacity of KDM2B to recognize non-methylated CpGs plays an essential role in normal polycomb domain formation. The most unexpected finding in chapter 6 was the observation that all variant PRC1 complexes place high levels of H2AK119ub1, which leads to PRC2 recruitment and H3K27me3-deposition. In contrast, canonical PCGF2 and PCGF4 seem less competent to achieve this.

These results bring forth multiple questions which cannot be unambiguously answered at the moment. For example, what are the molecular mechanisms by which H2AK119ub1-deposition leads to PRC2 recruitment or what is the function of canonical CBX-containing PRC1 complexes if they appear less capable of placing high levels of H2AK119ub1. At the same time, it would be of great interest to understand whether any of the other variant PRC1 complexes contains DNA-binding or chromatin-recognition activity or whether KDM2B-PRC1 is the sole PRC1 with the capacity to specifically interrogate the underlying DNA of PcG targets. The overarching goal would be to understand how the interplay between these multiple complexes and recruitment strategies results in functional polycomb domain formation in the developing organism.

Although some of these aspects were discussed in Chapter 6, recent discoveries may help to provide insight into one potential function of canonical PRC1 complexes. In addition to the CBX subunit, canonical PCGF2/PCGF4-PRC1 is defined by incorporation of PH proteins. Multiple *in vitro* studies indicated that the PH subunit possesses head-to-tail self-polymerisation capacity through its C-terminal SAM domain (Peterson et al. 1997; Kim et al. 2002; Kim et al. 2005). Recently, Isono et al. (2013) demonstrated, with the use of SAM domain mutants, that the polymerisation of PH is required for PRC1 clustering, for the stabilisation of PRC1/PRC2 target binding and subsequent robust PcG-mediated repression of developmental loci.

Considering these observations, one could envisage a model according to which polycomb domain formation is initiated through recruitment of PCGF1-PRC1 via KDM2B-dependent recognition of non-methylated DNA, potentially in conjunction with targeting of other variant PRC1s by yet undefined mechanisms. As shown in chapter 6, tethering of variant PCGFs results in functional RING1B nucleation and high levels of H2AK119ub1 deposition. This in turn would lead to PRC2 recruitment and subsequent trimethylation of H3K27. This PRC2-mediated histone tail modification creates a binding site for the chromodomain of the CBX subunit in canonical PH-containing PRC1 complexes. Interactions via the SAM domain (i.e. polymerisation) of canonical PRC1s would then facilitate retention of PRC1/PRC2 at PcG-targets, potentially by imposing a condensed local chromatin structure which precludes RNAPII binding (in absence of activating transcription factors) and allows for high concentrations of Polycomb proteins to be achieved. According to this model, variant PRC1s would be responsible for initiation of polycomb domains, while canonical PRC1s would function in the stabilisation of PRC1/PRC2 binding and higher-order structure of these transcriptionally inhibitory domains (Figure 7.4).

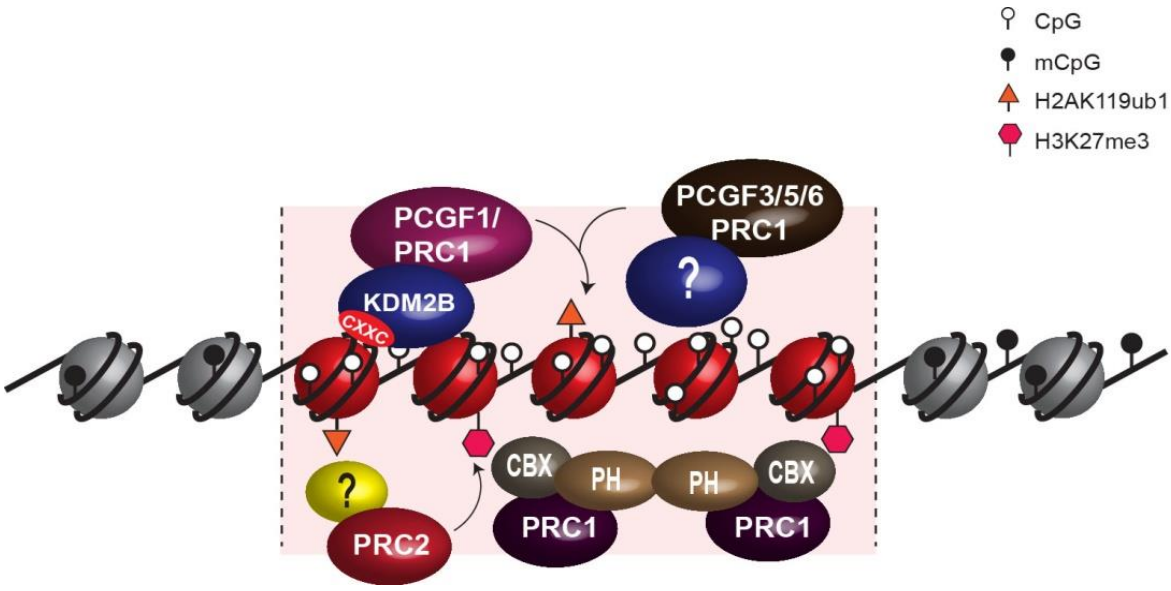


Figure 7.4. Variant PRC1 complexes may be responsible for initiation of polycomb domains, while canonical PRC1 functions in the stabilisation of Polycomb occupancy.

The model described is by no means sufficient to explain all Polycomb recruitment. Furthermore, it is clear that PRC2 can be targeted by either non-coding RNAs, sequence-specific DNA-binders or associated protein partners. Moreover, the differences in H2AK119ub1-deposition between variant and canonical PCGFs were observed at an artificial CpG-free locus. It could of course be that at natural CGI-associated PcG targets these differences are not valid, for example due to association of PCGF2/PCGF4-PRC1 with proteins which could stimulate their activity. The model described also does not exclude recruitment of PRC1 or PRC2 by a number of unrelated and uncharacterized pathways (Figure 7.4, Factor X). Nevertheless, it does provide a clear mechanistic link that explains a significant proportion of PcG protein occupancy at CpG islands.

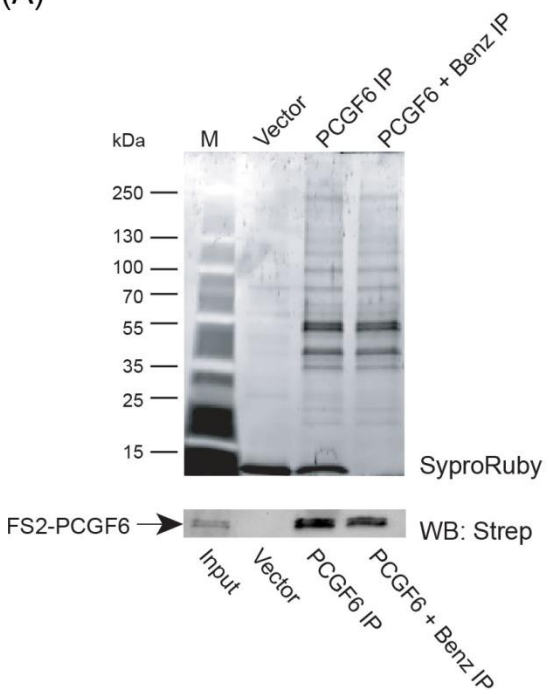
More work is clearly required to further understand the biological relevance of KDM2B. One aspect which was not address during this study is whether KDM2B has functions independently of its association with PcG proteins. Preliminary gel filtration studies presented in chapter 4, but also *in vitro* reconstitution experiments indicate that KDM2B appears to be a substoichiometric component of the variant PCGF1-PRC1 complex. Therefore, it will be important to characterize the role of free non-PcG associated KDM2B and define the interplay between these two states. One potential experiment to address this would be to rescue the KDM2B ZF-CxxC deletion cells with either wild-type or KDM2B mutants defective in their capacity to associate with PRC1.

Another aspect which was not fully addressed relates to the biological relevance of the H3K36me1/me2 demethylase activity of KDM2B. Experiments performed in this study show that the short form of KDM2B which lacks the JmjC catalytic domain retains the capacity to interact with PRC1 subunits. Therefore, it appears that the demethylase activity could in principle be dispensable for PcG recruitment. However, multiple lines of evidence suggest that H3K36me2/me3 is inhibitory to PRC2-catalysed H3K27 methylation (Schmitges et al. 2011; Yuan et al. 2011; Miyazaki et al. 2013). It would be interesting to assess whether rescue experiments in the KDM2B ZF-CxxC deletion mice with JmjC mutants completely rescue the strong homeotic transformations and embryonic lethality caused by loss of KDM2B and fully restore Polycomb occupancy in ES cells.

Nevertheless, the discovery that KDM2B can initiate *de novo* polycomb domains *in vivo,* capacity which appears to be shared with the other variant PRC1 complexes, provides a novel mechanism linking direct recognition of the underlying non-methylated DNA with recruitment of Polycomb proteins. In addition, although still at a hypothetical stage, it generated models to try and explain the observations made during this study. Importantly, the models discussed here are experimentally testable, and their results could potentially offer further insight into the complexities of PcG-mediated gene regulation. Despite joint intense research from multiple labs, the study of Polycomb function in various organisms remains an exciting field where numerous discoveries are waiting to be made. Finally, it brings us back to the importance of CpG islands, which, by acting as platforms for assembly of multiple essential chromatin modifying activities, play a central role in gene transcription regulation and integrity of cellular homeostasis.

8. Chapter eight - Supplementary figures

(A)



(B)

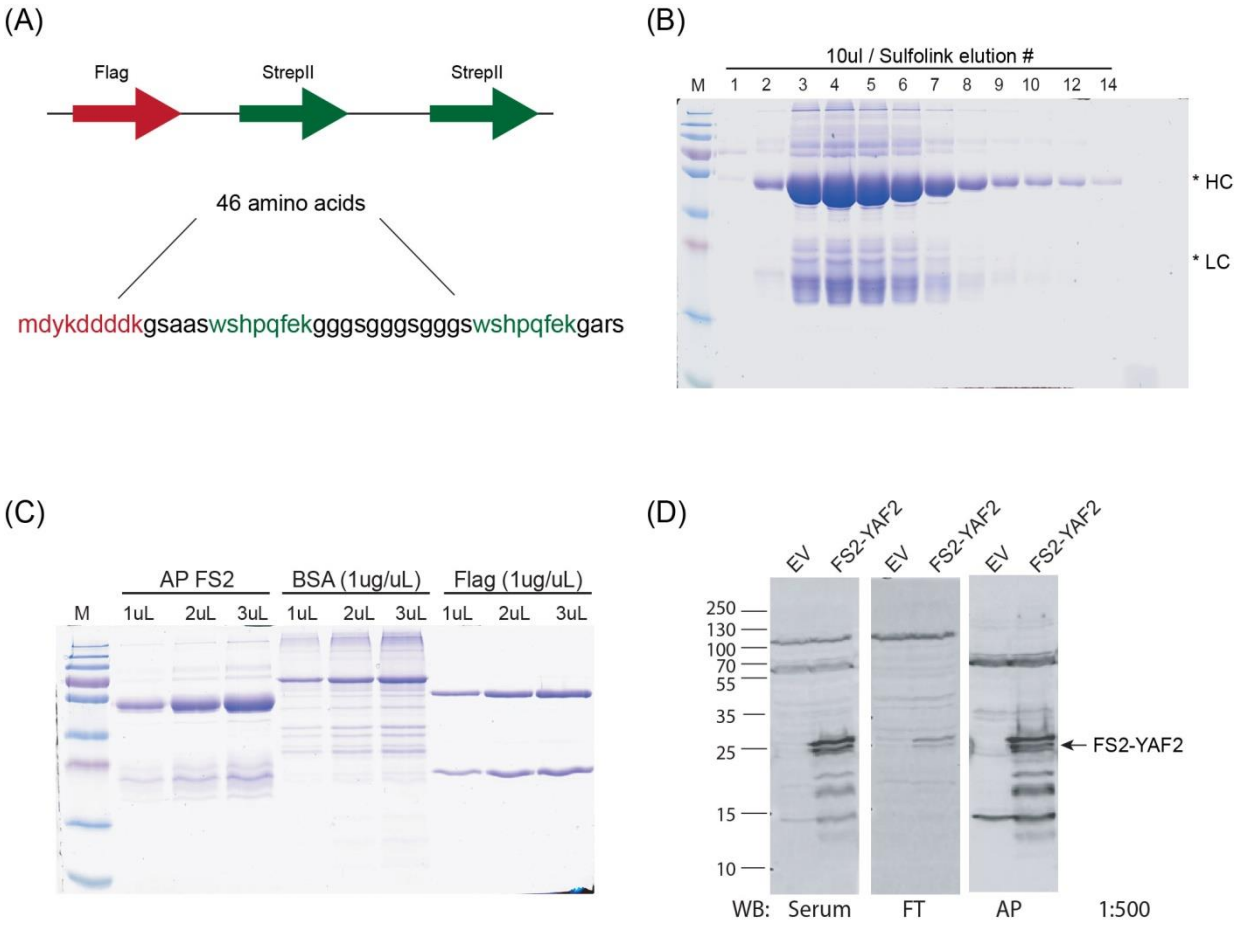
FS2 PCGF6

PCGF6 IP			PCGF6 + Benz IP		
Name	Score	Cov [%]	Name	Score	Cov [%]
MGA	3827	22	MGA	3572	22
PCGF6	3685	68	PCGF6	3350	56
RING1B	3046	45	RING1B	2356	52
RING1A	1609	44	RING1A	1295	40
CBX3/HP1g	1321	51	CBX3/HP1g	784	39
CBX1/HP1b	617	12	CBX1/HP1b	463	12
RYBP	1181	45	RYBP	1468	49
L3MBTL2	421	20	L3MBTL2	385	16
HDAC1	213	13	HDAC1	225	11
HDAC2	152	15	HDAC2	220	12
YAF2	182	25	YAF2	260	25
WDR5	70	10	WDR5	98	12

Supplementary figure S1. PCGF6 forms a distinct PRC1 variant complex, characterized by the MGA protein.

(A) Purified PCGF6 fractions were resolved by gradient SDS-PAGE and visualized by SyproRuby staining. The purifications were performed in the absence and presence of benzonase to exclude DNA-mediated interactions and a cell line containing only the empty vector was used to control for non-specific binding to the affinity resin. The elutions were probed by western blot for the tagged PCGF6 bait as indicated.

(B) Elutions from the PCGF6 affinity purification were directly analysed by tryptic digestion followed by peptide identification by LC–MS/MS. The Mascot scores and peptide coverage are shown for the respective affinity purifications. PCGF6 in mouse ES cells associates with a variant PRC1 complex containing the specific DNA-binding factor MGA, RING1B, HDACs and HP1b/g. Importantly, the PCGF6-PRC1 complex is distinct from the PCGF1-PRC1 complex and does not contain KDM2B.



Supplementary figure S2. Flag/2xStrep II (FS2) antibody design and purification.

(A) Schematic representation of Flag/2x StrepII tag and the amino acid sequence covering the tags and linker regions used to generate KLH-coupled antigen for rabbit immunizations. Flag tag is shown in red, StrepII tag in green, while linker sequences are illustrated in black.

(B) Coomassie stained SDS-PAGE gel of fractions eluted from Sulfolink column-based affinity chromatography of FS2 antibody from serum bleed#2 (elutions 1 to 14). 10ul of each fraction was loaded. The IgG heavy chain (HC) and light chain (LC) are indicated.

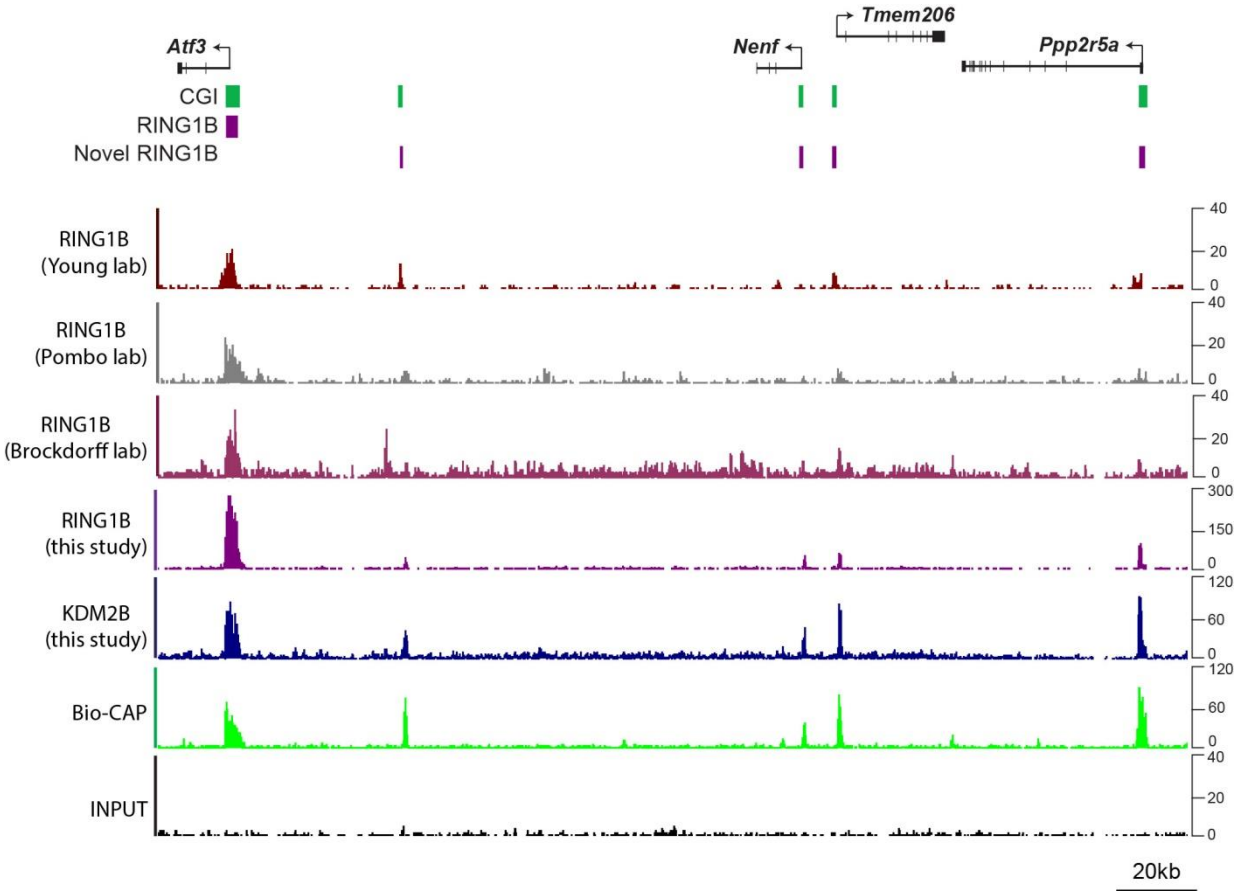
(C) Affinity purified (AP) FS2 antibody concentration estimation at approximately 2mg/mL for the band of interest compared to BSA and Flag-M2 standards (as indicated).

(D) Western blot analysis using whole cell extract prepared from empty vector (EV) and FS2-tagged YAF2 stably transfected ES cells. The membranes were probed with serum, flowthrough sample from the affinity purification (FT) or with the final purified FS2 antibody (AP). A reduction in non-specific signal while retaining an efficient response to its target FS2-tagged protein is observed (indicated by arrow).

KDM2B ZF-CxxCmut IP		
Name	Score	Cov [%]
KDM2B	5993	36
BCOR	224	6
PCGF1	113	15
RING1B	97	16
SKP1	331	58

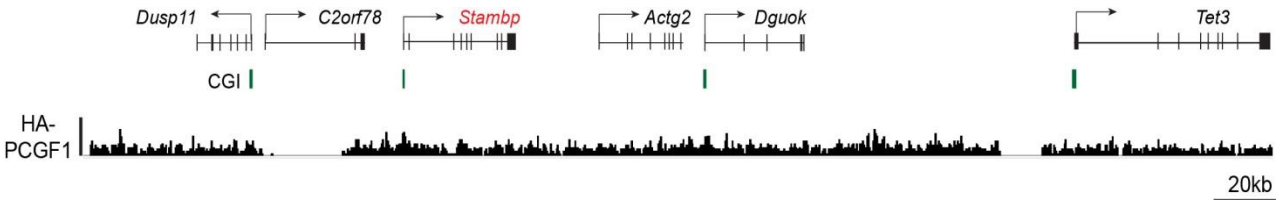
Supplementary figure S3. KDM2B ZF-CxxCmut retains the capacity to associate with Polycomb proteins.

To purify proteins associating with ZF-CxxC mutant KDM2B, nuclear extract from mouse ESC cell line stably expressing Flag-2XStrepII-tagged (FS2) KDM2B ZF-CxxCmut was prepared and used to carry out affinity purification followed by LC-MS/MS. The Mascot scores and peptide coverage are shown. Although the purification was less efficient, the results obtained suggest that KDM2B ZF-CxxC mutant associates with Polycomb proteins in ESCs, similar to wild-type KDM2B. The affinity purification was carried out by Mr. Thomas Sheahan, under my supervision.



Supplementary figure S4. Low-magnitude RING1B peaks are detectable at non-PcG target CGIs in other publicly available RING1B ChIP-Seq datasets.

A representative region comparison of RING1B ChIP-Seq profile determined in this study to other publicly available RING1B ChIP-Seq datasets in mouse ESCs. Input sequencing traces over the same region are shown in black. Bio-CAP-seq signal indicates regions containing non-methylated DNA. KDM2B ChIP-Seq profile as determined in this study is shown as indicative of CpG island regions. Above the sequencing traces gene promoters are show by black arrows and exons by vertical black lines. CGIs are shown as green bars with previously identified RING1B peaks and novel RING1B peaks indicated with purple boxes. The ChIP-Seq profiles were obtained from GEO with the following accession numbers: GSM656523 (R. Young lab); GSM850474 (A. Pombo lab) and GSM585229 (N. Brockdorff lab). Importantly, although the quality of the RING1B ChIP-Seq profile in this study is improved due to a highly optimized ChIP protocol, low-magnitude RING1B peaks are also detectable in the other datasets, suggesting that RING1B is more widely-associated with CGIs.



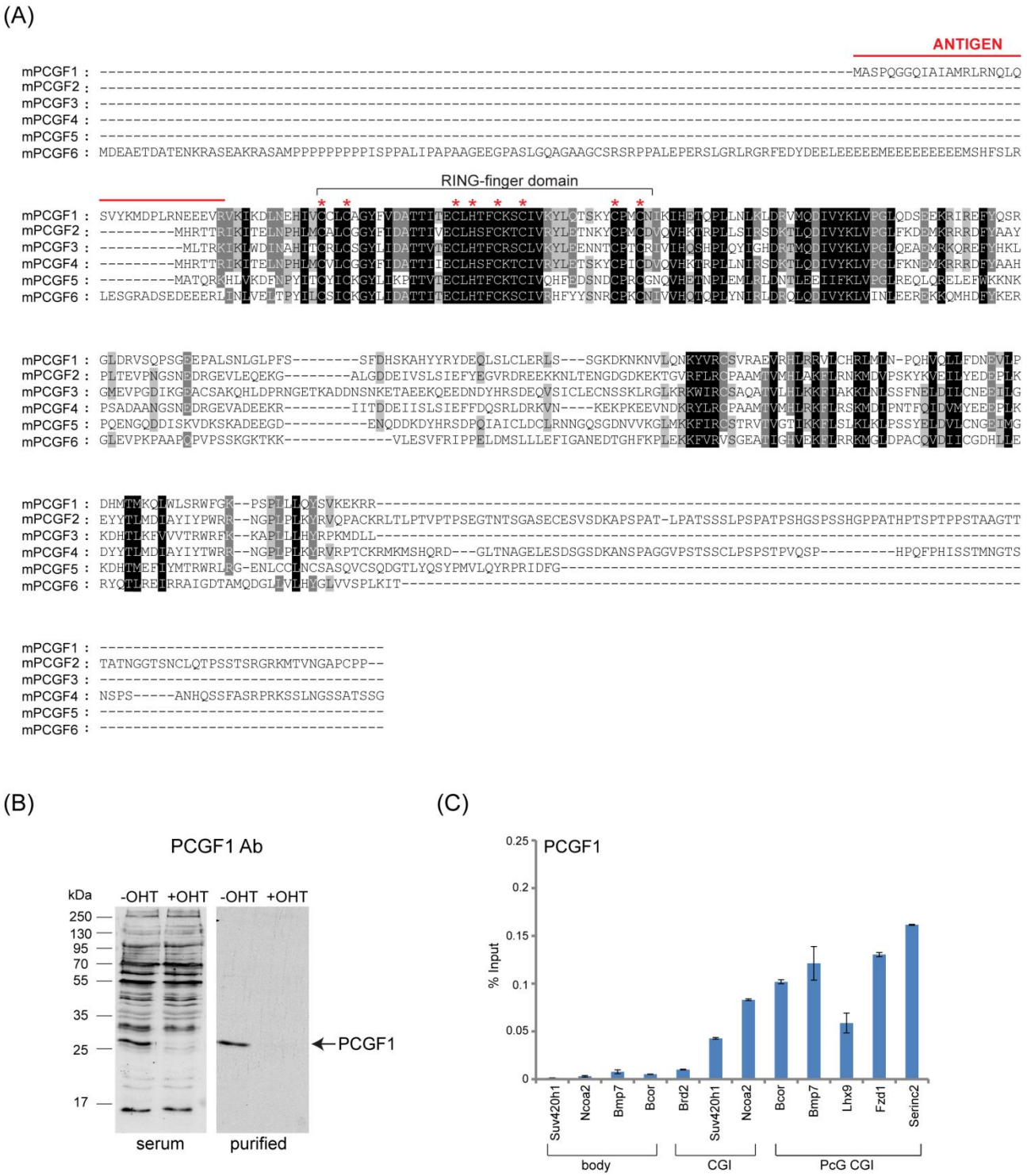
Supplementary figure S5. ChIP-Seq profile for epitope-tagged PCGF1, as obtained from Gao et al. (2012), does not have sufficient quality to unambiguously identify PCGF1 chromatin binding events.

Region on chromosome 2 showing epitope-tagged HA-PCGF1 ChIP-seq signal in 293T-Rex human transformed cell line showing both CGI and non-CGI associated promoters, as obtained by Gao et a. (2012). The ***Stambp*** gene (shown in red) was shown by Gao et. al (2012) to be a representative PCGF1 binding site. However, closer analysis of the ChIP-Seq profile for HA-tagged PCGF1 indicated that the quality of the dataset does not allow clearly identifying sites of preferential PCGF1 binding, as the tag density seems similar across the region. Therefore, it is impossible to conclude that PCGF1 has a unique chromatin binding profile, as compared to the other PCGF subunits. Above the sequencing traces gene promoters are show by black arrows and exons by vertical black lines, while CGIs are shown as green bars. The ChIP-Seq data set was aligned against the UCSC Human Feb. 2009 (GRCh37/hg19) Assembly.

	KDM2B WT		KDM2B ΔCxxC	
Name	Score	Cov (%)	Score	Cov (%)
KDM2B	36406	64	28907	66.3
SKP1	3188	73.6	8034	87.7
BCOR	413	12.2	5177	62.4
BCORL1	124	5.9	872	21.4
RING1B	259	33.9	1561	68.8
PCGF1	113	32.8	596	58.7
RYBP/YAF2	87	9.6	202	25.9

Supplementary figure S6. KDM2B retains the capacity to associate with the PCGF1-PRC1 variant complex in the absence of the ZF-CxxC DNA binding domain.

Affinity purification of full length Flag/2xStrepII (FS2)-epitope tagged wild-type (WT) KDM2B and KDM2B ZF-CxxC domain delete (KDM2BΔCxxC) followed by mass spectrometry-based analysis of associated proteins, as purified from mouse E14 embryonic stem cells. The mascot score and percentage coverage is indicated for the KDM2B-PCGF1/PRC1 variant complex components. As indicated, deletion of the ZF-CxxC domain encoding exon 13 generates a protein which lacks the ability to bind non-methylated DNA but retains the capacity to associate with Polycomb members. The stable cell lines expressing epitope tagged wild-type and mutant versions of KDM2B and the subsequent affinity purifications were performed and analysed by Mr. Lars Hanssen (Wellcome Trust rotation student) with my experimental advice.

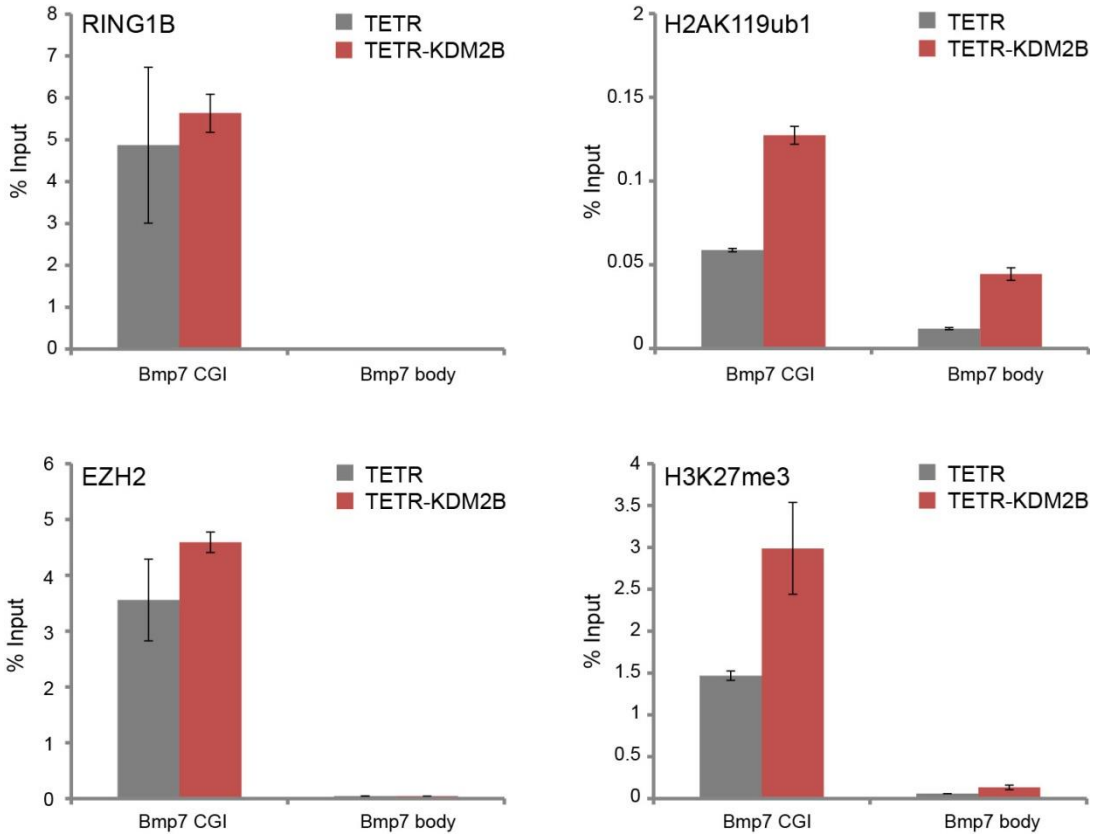


Supplementary figure S7. Generation and characterization of a PCGF1 specific rabbit polyclonal antibody.

(A) Protein sequence alignment of the six mouse PCGF proteins, illustrating that the region selected for PCGF1 antigen generation (shown in red above the alignment) is not conserved between the six polypeptides. Full-length sequences were obtained from Uniprot database, with the following accession numbers: Q8R023 (PCGF1), P23798 (PCGF2), Q8BTQ0 (PCGF3), P25916 (PCGF4), B7ZP24 (PCGF5) and Q99NA9 (PCGF6). The highly-conserved RING1B-finger interaction domain with the PRC1 E3 ubiquitin ligases RING1A/RING1B is also highlighted. Highly conserved regions between the six PCGFs are shown underlined in black, while amino acids partially conserved across the distinct proteins are coloured in grey.

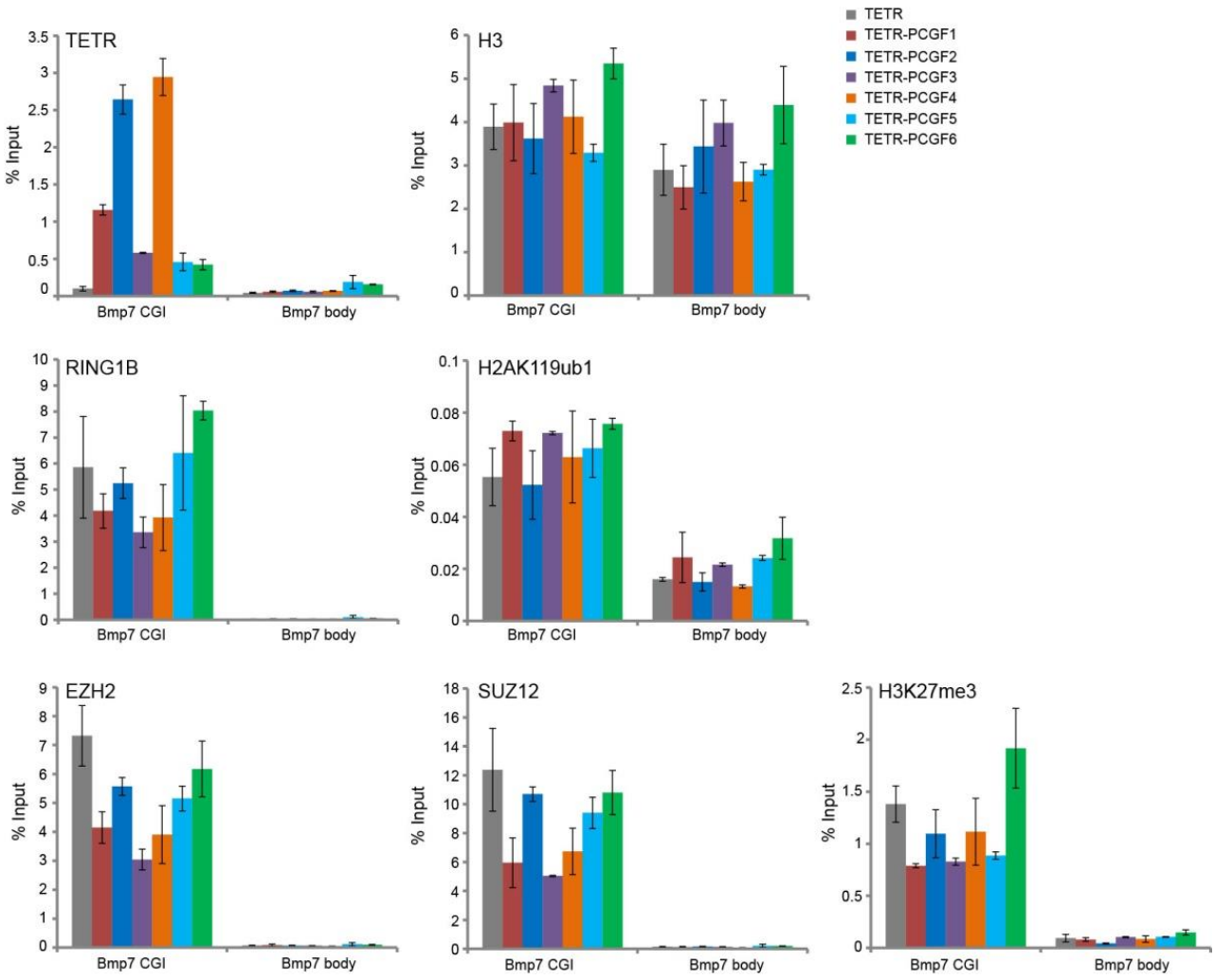
(B) Western blot analysis to test the response of PCGF1 antibody using nuclear extract from Pcgf1$^{fl/fl}$ ES cell line in which Pcgf1 can be conditionally-deleted by tamoxifen (OHT)-treatment (kindly provided by Dr. Haruhiko Koseki, unpublished). The membranes were probed with bleed#2 serum or with the final Sulfo-link affinity purified PCGF1 antibody. A single band of the expected PCGF1 molecular weight was detected by the affinity purified antibody, while no such band was detected after OHT-mediated Pcgf1 deletion. This suggests that the PCGF1 antibody generated and purified is highly specific for its cognate antigen.

(C) ChIP-qPCR analysis in mouse ESCs to test the response of the newly generated PCGF1 antibody in chromatin-immunoprecipitation type experiments. As can be observed, PCGF1 is enriched at regular and polycomb (PcG)-target CGIs, while no enrichment was detected at control non-CGI body regions. Importantly, the specific enrichment of PCGF1 at CGI versus non-CGI targets shows the suitability of the PCGF1 antibody for use in ChIP-experiments. Error bars represent SEM of two biological replicates.



Supplementary figure S8. Roughly comparable ChIP efficiencies at an endogenous polycomb target in the TetR and TetR-KDM2B cell lines.

ChIP-qPCR analysis at *Bmp7* CGI and corresponding body region in the TetR alone and TetR-KDM2B fusion cell lines, showing relatively similar enrichment levels of RING1B, H2AK119ub1, EZH2 and H3K27me3. This suggests that the recruitment of PRC1 and PRC2 detected in the KDM2B fusion line but not in the cell line expressing only TetR was a direct consequence of KDM2B targeting, and not due to differences in ChIP efficiencies between the two lines. Error bars represent SEM of two biological replicates.



Supplementary figure S9. Comparable ChIP efficiencies at an endogenous polycomb target between the different TetR-PCGF fusion lines.

ChIP-qPCR analysis at *Bmp7* CGI and corresponding body region in the TetR alone and TetR-PCGF1→6 fusion cell lines, showing relatively similar enrichment levels of RING1B, H2AK119ub1, EZH2, SUZ12, H3K27me3 and histone H3. This suggests that the differential ability of variant PRC1 complexes to catalyze H2AK119ub1 and mediate PRC2 recruitment is a direct consequence of tethering the respective individual PCGFs, and not due to differences in ChIP efficiencies between the TetR-PCGF fusion lines. Interestingly, all six TetR-PCGFs show some degree of above-background enrichment at the *Bmp7* promoter region (TETR panel). This suggests that the TetR-PCGFs are functional and can be recruited to endogenous natural polycomb targets irrespective of their fusion with TetR DNA binding domain. Interestingly, fusions of the canonical PCGF2 and PCGF4 seem to be especially efficient at nucleating at the *Bmp7* locus, although this does not result in higher levels of PRC1/PRC2 and their associated chromatin marks compared to tethering of variant PCGFs. Therefore, the validity of normalization to *Bmp7* CGI as a means of comparing between the different TetR-PCGF fusion lines is not affected. Error bars represent SEM of two biological replicates.

References

Agger K, Cloos PA, Christensen J, Pasini D, Rose S, Rappsilber J, Issaeva I, Canaani E, Salcini AE, Helin K. 2007. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**: 731-734.

Akam M. 1987. The molecular basis for metameric pattern in the Drosophila embryo. *Development* **101**: 1-22.

Akasaka T, Kanno M, Balling R, Mieza MA, Taniguchi M, Koseki H. 1996. A role for mel-18, a Polycomb group-related vertebrate gene, during theanteroposterior specification of the axial skeleton. *Development* **122**: 1513-1522.

Akasaka T, van Lohuizen M, van der Lugt N, Mizutani-Koseki Y, Kanno M, Taniguchi M, Vidal M, Alkema M, Berns A, Koseki H. 2001. Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development* **128**: 1587-1597.

Alfieri C, Gambetta MC, Matos R, Glatt S, Sehr P, Fraterman S, Wilm M, Muller J, Muller CW. 2013. Structural basis for targeting the chromatin repressor Sfmbt to Polycomb response elements. *Genes & development* **27**: 2367-2379.

Allen MD, Grummitt CG, Hilcenko C, Min SY, Tonkin LM, Johnson CM, Freund SM, Bycroft M, Warren AJ. 2006. Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *The EMBO journal* **25**: 4503-4512.

Allis CD, Jenuwein T, Reinberg D, Caparros ML. 2006. Epigenetics. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY*.

Andreu-Vieyra CV, Chen R, Agno JE, Glaser S, Anastassiadis K, Stewart AF, Matzuk MM. 2010. MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS biology* **8**.

Antequera F. 2003. Structure, function and evolution of CpG island promoters. *Cellular and molecular life sciences : CMLS* **60**: 1647-1658.

Aravind L, Iyer LM. 2012. The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell cycle* **11**: 119-131.

Ardehali MB, Mei A, Zobeck KL, Caron M, Lis JT, Kusch T. 2011. Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *The EMBO journal* **30**: 2817-2828.

Arnold P, Scholer A, Pachkov M, Balwierz PJ, Jorgensen H, Stadler MB, van Nimwegen E, Schubeler D. 2013. Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome research* **23**: 60-73.

Arrigoni R, Alam SL, Wamstad JA, Bardwell VJ, Sundquist WI, Schreiber-Agus N. 2006. The Polycomb-associated protein Rybp is a ubiquitin binding protein. *FEBS letters* **580**: 6233-6241.

Arthold S, Kurowski A, Wutz A. 2011. Mechanistic insights into chromosome-wide silencing in X inactivation. *Human genetics* **130**: 295-305.

Aslanidis C, de Jong PJ. 1990. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic acids research* **18**: 6069-6074.

Atsuta T, Fujimura S, Moriya H, Vidal M, Akasaka T, Koseki H. 2001. Production of monoclonal antibodies against mammalian Ring1B proteins. *Hybridoma* **20**: 43-46.

Ausio J. 1992. Structure and dynamics of transcriptionally active chromatin. *Journal of cell science* **102 (Pt 1)**: 1-5.

Ayoub N, Noma K, Isaac S, Kahan T, Grewal SI, Cohen A. 2003. A novel jmjC domain protein modulates heterochromatization in fission yeast. *Molecular and cellular biology* **23**: 4356-4370.

Ballare C, Lange M, Lapinaite A, Martin GM, Morey L, Pascual G, Liefke R, Simon B, Shi Y, Gozani O et al. 2012. Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nature structural & molecular biology* **19**: 1257-1265.

Ballestar E, Wolffe AP. 2001. Methyl-CpG-binding proteins. Targeting specific gene repression. *European journal of biochemistry / FEBS* **268**: 1-6.

Bantignies F, Roure V, Comet I, Leblanc B, Schuettengruber B, Bonnet J, Tixier V, Mas A, Cavalli G. 2011. Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell* **144**: 214-226.

Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. 2010. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**: 470-484.

Bartova E, Horakova AH, Uhlirova R, Raska I, Galiova G, Orlova D, Kozubek S. 2010. Structure and epigenetics of nucleoli in comparison with non-nucleolar compartments. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **58**: 391-403.

Baudino TA, Cleveland JL. 2001. The Max network gone mad. *Molecular and cellular biology* **21**: 691-702.

Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**: 482-485.

Bell O, Schwaiger M, Oakeley EJ, Lienert F, Beisel C, Stadler MB, Schubeler D. 2010. Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nature structural & molecular biology* **17**: 894-900.

Ben-Saadon R, Zaaroor D, Ziv T, Ciechanover A. 2006. The polycomb protein Ring1B generates self atypical mixed ubiquitin chains required for its in vitro histone H2A ligase activity. *Molecular cell* **24**: 701-711.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, 3rd, Gingeras TR et al. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**: 169-181.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-326.

Berry WL, Janknecht R. 2013. KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells. *Cancer research* **73**: 2936-2942.

Bestor TH, Ingram VM. 1983. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proceedings of the National Academy of Sciences of the United States of America* **80**: 5559-5563.

Bestor TH, Verdine GL. 1994. DNA methyltransferases. *Current opinion in cell biology* **6**: 380-389.

Bienz M, Muller J. 1995. Transcriptional silencing of homeotic genes in Drosophila. *BioEssays : news and reviews in molecular, cellular and developmental biology* **17**: 775-784.

Bird A. 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends in Genetics* **3, 342-347**.

-. 2002. DNA methylation patterns and epigenetic memory. *Genes & development* **16**: 6-21.

-. 2007. Perceptions of epigenetics. *Nature* **447**: 396-398.

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**: 91-99.

Bird AP, Wolffe AP. 1999. Methylation-induced repression--belts, braces, and chromatin. *Cell* **99**: 451-454.

Blackledge NP, Klose R. 2011. CpG island chromatin: a platform for gene regulation. *Epigenetics : official journal of the DNA Methylation Society* **6**: 147-152.

Blackledge NP, Long HK, Zhou JC, Kriaucionis S, Patient R, Klose RJ. 2012. Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA. *Nucleic acids research* **40**: e32.

Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, Klose RJ. 2010. CpG islands recruit a histone H3 lysine 36 demethylase. *Molecular cell* **38**: 179-190.

Blastyak A, Mishra RK, Karch F, Gyurkovics H. 2006. Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. *Molecular and cellular biology* **26**: 1434-1444.

Boa S, Coert C, Patterton HG. 2003. Saccharomyces cerevisiae Set1p is a methyltransferase specific for lysine 4 of histone H3 and is required for efficient gene expression. *Yeast* **20**: 827-835.

Boisvert FM, van Koningsbruggen S, Navascues J, Lamond AI. 2007. The multifunctional nucleolus. *Nature reviews Molecular cell biology* **8**: 574-585.

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. 2007. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**: 1760-1764.

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. 2001. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**: 2536-2539.

Boyer LA, Latek RR, Peterson CL. 2004. The SANT domain: a unique histone-tail-binding module? *Nature reviews Molecular cell biology* **5**: 158-163.

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**: 349-353.

Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & development* **20**: 1123-1136.

Breen TR, Duncan IM. 1986. Maternal expression of genes that regulate the bithorax complex of Drosophila melanogaster. *Developmental biology* **118**: 442-456.

Breiling A, Turner BM, Bianchi ME, Orlando V. 2001. General transcription factors bind promoters repressed by Polycomb group proteins. *Nature* **412**: 651-655.

Brien GL, Gambero G, O'Connell DJ, Jerman E, Turner SA, Egan CM, Dunne EJ, Jurgens MC, Wynne K, Piao L et al. 2012. Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nature structural & molecular biology* **19**: 1273-1281.

Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, Dent SY, Winston F, Allis CD. 2001. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes & development* **15**: 3286-3295.

Brockdorff N. 2013. Noncoding RNA and Polycomb recruitment. *Rna* **19**: 429-442.

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**: 515-526.

Brown JL, Fritsch C, Mueller J, Kassis JA. 2003. The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. *Development* **130**: 285-294.

Brown JL, Grau DJ, DeVido SK, Kassis JA. 2005. An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. *Nucleic acids research* **33**: 5181-5189.

Brownell JE, Allis CD. 1995. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 6364-6368.

Buchwald G, van der Stoop P, Weichenrieder O, Perrakis A, van Lohuizen M, Sixma TK. 2006. Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. *The EMBO journal* **25**: 2465-2474.

Byvoet P, Shepherd GR, Hardin JM, Noland BJ. 1972. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Archives of biochemistry and biophysics* **148**: 558-567.

Cai L, Rothbart SB, Lu R, Xu B, Chen WY, Tripathy A, Rockowitz S, Zheng D, Patel DJ, Allis CD et al. 2013. An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Molecular cell* **49**: 571-582.

Callen E, Di Virgilio M, Kruhlak MJ, Nieto-Soler M, Wong N, Chen HT, Faryabi RB, Polato F, Santos M, Starnes LM et al. 2013. 53BP1 mediates productive and mutagenic DNA repair through distinct phosphoprotein interactions. *Cell* **153**: 1266-1280.

Camporeale G, Shubert EE, Sarath G, Cerny R, Zempleni J. 2004. K8 and K12 are biotinylated in human histone H4. *European journal of biochemistry / FEBS* **271**: 2257-2263.

Cao R, Tsukada Y, Zhang Y. 2005. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Molecular cell* **20**: 845-854.

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

Cao R, Zhang Y. 2004a. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current opinion in genetics & development* **14**: 155-164.

-. 2004b. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Molecular cell* **15**: 57-67.

Caputo VS, Costa JR, Makarona K, Georgiou E, Layton DM, Roberts I, Karadimitris A. 2013. Mechanism of Polycomb recruitment to CpG islands revealed by inherited disease-associated mutation. *Human molecular genetics* **22**: 3187-3194.

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature genetics* **38**: 626-635.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**: 581-592.

Carvalho S, Raposo AC, Martins FB, Grosso AR, Sridhara SC, Rino J, Carmo-Fonseca M, de Almeida SF. 2013. Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic acids research* **41**: 2881-2893.

Casanova M, Preissner T, Cerase A, Poot R, Yamada D, Li X, Appanah R, Bezstarosti K, Demmers J, Koseki H 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. *Development* **138**: 1471-1482.

Cenciarelli C, Chiaur DS, Guardavaccaro D, Parks W, Vidal M, Pagano M. 1999. Identification of a family of human F-box proteins. *Current biology : CB* **9**: 1177-1179.

Chan CH, Morrow JK, Li CF, Gao Y, Jin G, Moten A, Stagg LJ, Ladbury JE, Cai Z, Xu D et al. 2013. Pharmacological inactivation of Skp2 SCF ubiquitin ligase restricts cancer stem cell traits and cancer progression. *Cell* **154**: 556-568.

Chan CS, Rastelli L, Pirrotta V. 1994. A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. *The EMBO journal* **13**: 2553-2564.

Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. 1998. DNA hypomethylation leads to elevated mutation rates. *Nature* **395**: 89-93.

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ. 1993. Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74**: 299-307.

Chiang A, O'Connor MB, Paro R, Simon J, Bender W. 1995. Discrete Polycomb-binding sites in each parasegmental domain of the bithorax complex. *Development* **121**: 1681-1689.

Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, Guszczynski T, Dressler GR, Copeland TD, Kalkum M et al. 2007. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *The Journal of biological chemistry* **282**: 20395-20406.

Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* **466**: 388-392.

Chow JC, Brown CJ. 2003. Forming facultative heterochromatin: silencing of an X chromosome in mammalian females. *Cellular and molecular life sciences : CMLS* **60**: 2586-2603.

Christensen J, Agger K, Cloos PA, Pasini D, Rose S, Sennels L, Rappsilber J, Hansen KH, Salcini AE, Helin K. 2007. RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. *Cell* **128**: 1063-1076.

Cierpicki T, Risner LE, Grembecka J, Lukasik SM, Popovic R, Omonkowska M, Shultis DD, Zeleznik-Le NJ, Bushweller JH. 2010. Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. *Nature structural & molecular biology* **17**: 62-68.

Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, Nogales E. 2012. Molecular architecture of human polycomb repressive complex 2. *eLife* **1**: e00005.

Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, Lee JH, Skalnik D, Bird A. 2012. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & development* **26**: 1714-1728.

Cooper DN, Taggart MH, Bird AP. 1983. Unmethylated domains in vertebrate DNA. *Nucleic acids research* **11**: 647-658.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845-1848.

Cosgrove MS, Wolberger C. 2005. How does the histone code work? *Biochemistry and cell biology = Biochimie et biologie cellulaire* **83**: 468-476.

Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. 1978. Molecular basis of base substitution hotspots in Escherichia coli. *Nature* **274**: 775-780.

Craig KL, Tyers M. 1999. The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. *Progress in biophysics and molecular biology* **72**: 299-328.

Cross SH, Meehan RR, Nan X, Bird A. 1997. A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nature genetics* **16**: 256-259.

Cuddapah S, Roh TY, Cui K, Jose CC, Fuller MT, Zhao K, Chen X. 2012. A novel human polycomb binding site acts as a functional polycomb response element in Drosophila. *PloS one* **7**: e36365.

Culi J, Aroca P, Modolell J, Mann RS. 2006. jing is required for wing development and to establish the proximo-distal axis of the leg in Drosophila melanogaster. *Genetics* **173**: 255-266.

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. 2002. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**: 185-196.

Daniel JM, Reynolds AB. 1999. The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. *Molecular and cellular biology* **19**: 3614-3623.

Davidovich C, Zheng L, Goodrich KJ, Cech TR. 2013. Promiscuous RNA binding by Polycomb repressive complex 2. *Nature structural & molecular biology* **20**: 1250-1257.

Dawlaty MM, Breiling A, Le T, Raddatz G, Barrasa MI, Cheng AW, Gao Q, Powell BE, Li Z, Xu M et al. 2013. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Developmental cell* **24**: 310-323.

de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, Levezinho H, Gut M, Eick D, Gut I et al. 2011. Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nature structural & molecular biology* **18**: 977-983.

de Bie P, Zaaroor-Regev D, Ciechanover A. 2010. Regulation of the Polycomb protein RING1B ubiquitination by USP7. *Biochemical and biophysical research communications* **400**: 389-395.

de Napoles M, Mermoud JE, Wakao R, Tang YA, Endoh M, Appanah R, Nesterova TB, Silva J, Otte AP, Vidal M et al. 2004. Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Developmental cell* **7**: 663-676.

De Santa F, Totaro MG, Prosperini E, Notarbartolo S, Testa G, Natoli G. 2007. The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell* **130**: 1083-1094.

Dejardin J, Cavalli G. 2004. Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. *The EMBO journal* **23**: 857-868.

Dejardin J, Rappailles A, Cuvier O, Grimaud C, Decoville M, Locker D, Cavalli G. 2005. Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. *Nature* **434**: 533-538.

Dellino GI, Schwartz YB, Farkas G, McCabe D, Elgin SC, Pirrotta V. 2004. Polycomb silencing blocks transcription initiation. *Molecular cell* **13**: 887-893.

Deshaies RJ, Joazeiro CA. 2009. RING domain E3 ubiquitin ligases. *Annual review of biochemistry* **78**: 399-434.

DeVido SK, Kwon D, Brown JL, Kassis JA. 2008. The role of Polycomb-group response elements in regulation of engrailed transcription in Drosophila. *Development* **135**: 669-676.

Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM. 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**: 491-496.

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. 2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *The Journal of biological chemistry* **285**: 26114-26120.

Dietrich N, Lerdrup M, Landt E, Agrawal-Singh S, Bak M, Tommerup N, Rappsilber J, Sodersten E, Hansen K. 2012. REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS genetics* **8**: e1002494.

Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic acids research* **11**: 1475-1489.

Dillon SC, Zhang X, Trievel RC, Cheng X. 2005. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome biology* **6**: 227.

Dinant C, Houtsmuller AB, Vermeulen W. 2008. Chromatin structure and DNA damage repair. *Epigenetics & chromatin* **1**: 9.

Ding X, Lin Q, Ensenat-Waser R, Rose-John S, Zenke M. 2012. Polycomb group protein Bmi1 promotes hematopoietic cell development from embryonic stem cells. *Stem cells and development* **21**: 121-132.

Dion MF, Altschuler SJ, Wu LF, Rando OJ. 2005. Genomic characterization reveals a simple histone H4 acetylation code. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 5501-5506.

Dorighi KM, Tamkun JW. 2013. The trithorax group proteins Kismet and ASH1 promote H3K36 dimethylation to counteract Polycomb group repression in Drosophila. *Development* **140**: 4182-4192.

DuBridge RB, Tang P, Hsia HC, Leong PM, Miller JH, Calos MP. 1987. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Molecular and cellular biology* **7**: 379-387.

Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* **287**: 560-561.

Dyson MH, Rose S, Mahadevan LC. 2001. Acetyllysine-binding and function of bromodomain-containing proteins in chromatin. *Frontiers in bioscience : a journal and virtual library* **6**: D853-865.

Edmunds JW, Mahadevan LC, Clayton AL. 2008. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The EMBO journal* **27**: 406-420.

Eggan K, Akutsu H, Loring J, Jackson-Grusby L, Klemm M, Rideout WM, 3rd, Yanagimachi R, Jaenisch R. 2001. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 6209-6214.

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C. 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic acids research* **10**: 2709-2721.

Eissenberg JC, Lee MG, Schneider J, Ilvarsonn A, Shiekhattar R, Shilatifard A. 2007. The trithorax-group gene in Drosophila little imaginal discs encodes a trimethylated histone H3 Lys4 demethylase. *Nature structural & molecular biology* **14**: 344-346.

Eissenberg JC, Shilatifard A. 2010. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental biology* **339**: 240-249.

Elderkin S, Maertens GN, Endoh M, Mallery DL, Morrice N, Koseki H, Peters G, Brockdorff N, Hiom K. 2007. A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. *Molecular cell* **28**: 107-120.

Endoh M, Endo TA, Endoh T, Fujimura Y, Ohara O, Toyoda T, Otte AP, Okano M, Brockdorff N, Vidal M et al. 2008. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**: 1513-1524.

Endoh M, Endo TA, Endoh T, Isono K, Sharif J, Ohara O, Toyoda T, Ito T, Eskeland R, Bickmore WA et al. 2012. Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. *PLoS genetics* **8**: e1002774.

Enkhbayar P, Kamiya M, Osaki M, Matsumoto T, Matsushima N. 2004. Structural principles of leucine-rich repeat (LRR) proteins. *Proteins* **54**: 394-403.

Eskeland R, Freyer E, Leeb M, Wutz A, Bickmore WA. 2010a. Histone acetylation and the maintenance of chromatin compaction by Polycomb repressive complexes. *Cold Spring Harbor symposia on quantitative biology* **75**: 71-78.

Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultchi AI, Wutz A et al. 2010b. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Molecular cell* **38**: 452-464.

Esteller M. 2002. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21**: 5427-5440.

Fang J, Hogan GJ, Liang G, Lieb JD, Zhang Y. 2007. The Saccharomyces cerevisiae histone demethylase Jhd1 fine-tunes the distribution of H3K36me2. *Molecular and cellular biology* **27**: 5055-5065.

Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW et al. 2012. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* **1**: e00205.

Faust C, Lawson KA, Schork NJ, Thiel B, Magnuson T. 1998. The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development* **125**: 4495-4506.

Fenouil R, Cauchy P, Koch F, Descostes N, Cabeza JZ, Innocenti C, Ferrier P, Spicuglia S, Gut M, Gut I et al. 2012. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome research* **22**: 2399-2408.

Ferrari KJ, Scelfo A, Jammula S, Cuomo A, Barozzi I, Stutzer A, Fischle W, Bonaldi T, Pasini D. 2013. Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Molecular cell*.

Fingerman IM, Wu CL, Wilson BD, Briggs SD. 2005. Global loss of Set1-mediated H3 Lys4 trimethylation is associated with silencing defects in Saccharomyces cerevisiae. *The Journal of biological chemistry* **280**: 28761-28765.

Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & development* **17**: 1870-1881.

Flemming W. 1879. Ueber das Verhalten des Kerns bei der Zellteilung und über die Bedeutung mehrkerniger Zellen. *Arch Pathol Anat* **77, 1–28**.

Fradet-Turcotte A, Canny MD, Escribano-Diaz C, Orthwein A, Leung CC, Huang H, Landry MC, Kitevski-LeBlanc J, Noordermeer SM, Sicheri F et al. 2013. 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* **499**: 50-54.

Francis NJ, Kingston RE. 2001. Mechanisms of transcriptional memory. *Nature reviews Molecular cell biology* **2**: 409-421.

Francis NJ, Kingston RE, Woodcock CL. 2004. Chromatin compaction by a polycomb group protein complex. *Science* **306**: 1574-1577.

Francis NJ, Saurin AJ, Shao Z, Kingston RE. 2001. Reconstitution of a functional core polycomb repressive complex. *Molecular cell* **8**: 545-556.

Freemont PS, Hanson IM, Trowsdale J. 1991. A novel cysteine-rich sequence motif. *Cell* **64**: 483-484.

Frescas D, Guardavaccaro D, Bassermann F, Koyama-Nasu R, Pagano M. 2007. JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. *Nature* **450**: 309-313.

Frescas D, Guardavaccaro D, Kuchay SM, Kato H, Poleshko A, Basrur V, Elenitoba-Johnson KS, Katz RA, Pagano M. 2008. KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. *Cell cycle* **7**: 3539-3547.

Fritsch C, Beuchle D, Muller J. 2003. Molecular and genetic analysis of the Polycomb group gene Sex combs extra/Ring in Drosophila. *Mechanisms of development* **120**: 949-954.

Fritsch C, Brown JL, Kassis JA, Muller J. 1999. The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. *Development* **126**: 3905-3913.

Fukuda T, Tokunaga A, Sakamoto R, Yoshida N. 2011. Fbxl10/Kdm2b deficiency accelerates neural progenitor cell death and leads to exencephaly. *Molecular and cellular neurosciences* **46**: 614-624.

Fukushima H, Ogura K, Wan L, Lu Y, Li V, Gao D, Liu P, Lau AW, Wu T, Kirschner MW et al. 2013. SCF-mediated Cdh1 degradation defines a negative feedback system that coordinates cell-cycle progression. *Cell reports* **4**: 803-816.

Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, Reinberg D. 2012. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Molecular cell* **45**: 344-356.

Garcia-Fernandez J. 2005. Hox, ParaHox, ProtoHox: facts and guesses. *Heredity* **94**: 145-152.

Garcia BA, Thomas CE, Kelleher NL, Mizzen CA. 2008. Tissue-specific expression and post-translational modification of histone H3 variants. *Journal of proteome research* **7**: 4225-4236.

Garcia E, Marcos-Gutierrez C, del Mar Lorente M, Moreno JC, Vidal M. 1999. RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. *The EMBO journal* **18**: 3404-3418.

Gardiner-Garden M, Frommer M. 1987. CpG islands in vertebrate genomes. *Journal of molecular biology* **196**: 261-282.

Gaydos LJ, Rechtsteiner A, Egelhofer TA, Carroll CR, Strome S. 2012. Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. *Cell reports* **2**: 1169-1177.

Gearhart MD, Corcoran CM, Wamstad JA, Bardwell VJ. 2006. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Molecular and cellular biology* **26**: 6880-6889.

Georgakopoulos T, Thireos G. 1992. Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. *The EMBO journal* **11**: 4145-4152.

Gindhart JG, Jr., Kaufman TC. 1995. Identification of Polycomb and trithorax group responsive elements in the regulatory region of the Drosophila homeotic gene Sex combs reduced. *Genetics* **139**: 797-814.

Glaser S, Lubitz S, Loveland KL, Ohbo K, Robb L, Schwenk F, Seibler J, Roellig D, Kranz A, Anastassiadis K et al. 2009. The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics & chromatin* **2**: 5.

Goldknopf IL, Taylor CW, Baum RM, Yeoman LC, Olson MO, Prestayko AW, Busch H. 1975. Isolation and characterization of protein A24, a "histone-like" non-histone chromosomal protein. *The Journal of biological chemistry* **250**: 7182-7187.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annual review of biochemistry* **74**: 481-514.

Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. 2005. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *The Journal of biological chemistry* **280**: 13341-13348.

Grandori C, Cowley SM, James LP, Eisenman RN. 2000. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annual review of cell and developmental biology* **16**: 653-699.

Grau DJ, Chapman BA, Garlick JD, Borowsky M, Francis NJ, Kingston RE. 2011. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes & development* **25**: 2210-2221.

Greer EL, Shi Y. 2012. Histone methylation: a dynamic mark in health, disease and inheritance. *Nature reviews Genetics* **13**: 343-357.

Groth A, Ray-Gallet D, Quivy JP, Lukas J, Bartek J, Almouzni G. 2005. Human Asf1 regulates the flow of S phase histones during replicational stress. *Molecular cell* **17**: 301-311.

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG et al. 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**: 606-610.

Guenther MG, Jenner RG, Chevalier B, Nakamura T, Croce CM, Canaani E, Young RA. 2005. Global and Hox-specific roles for the MLL1 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 8603-8608.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**: 77-88.

Gutierrez L, Oktaba K, Scheuermann JC, Gambetta MC, Ly-Hartig N, Muller J. 2012. The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development* **139**: 117-127.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**: 295-300.

Haig D. 2004. The (dual) origin of epigenetics. *Cold Spring Harbor symposia on quantitative biology* **69**: 67-70.

Hajkova P, Ancelin K, Waldmann T, Lacoste N, Lange UC, Cesari F, Lee C, Almouzni G, Schneider R, Surani MA. 2008. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**: 877-881.

Hallson G, Hollebakken RE, Li T, Syrzycka M, Kim I, Cotsworth S, Fitzpatrick KA, Sinclair DA, Honda BM. 2012. dSet1 is the main H3K4 di- and tri-methyltransferase throughout Drosophila development. *Genetics* **190**: 91-100.

Hampsey M, Reinberg D. 2003. Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* **113**: 429-432.

Handley PM, Mueckler M, Siegel NR, Ciechanover A, Schwartz AL. 1991. Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 258-262.

Hansen KH, Bracken AP, Pasini D, Dietrich N, Gehani SS, Monrad A, Rappsilber J, Lerdrup M, Helin K. 2008. A model for transmission of the H3K27me3 epigenetic mark. *Nature cell biology* **10**: 1291-1300.

Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, Carrozza MJ, Workman JL. 2002. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111**: 369-379.

Hata K, Okano M, Lei H, Li E. 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129**: 1983-1993.

Hatzi K, Jiang Y, Huang C, Garrett-Bakelman F, Gearhart MD, Giannopoulou EG, Zumbo P, Kirouac K, Bhaskara S, Polo JM et al. 2013. A hybrid mechanism of action for BCL6 in B cells defined by formation of functionally distinct complexes at enhancers and promoters. *Cell reports* **4**: 578-588.

He GP, Kim S, Ro HS. 1999. Cloning and characterization of a novel zinc finger transcriptional repressor. A direct role of the zinc finger motif in repression. *The Journal of biological chemistry* **274**: 14678-14684.

He J, Kallin EM, Tsukada Y, Zhang Y. 2008. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). *Nature structural & molecular biology* **15**: 1169-1175.

He J, Nguyen AT, Zhang Y. 2011. KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia. *Blood* **117**: 3869-3880.

He J, Shen L, Wan M, Taranova O, Wu H, Zhang Y. 2013. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nature cell biology* **15**: 373-384.

Heitz E. 1928. Das Heterochromatin der Moose. *1 Jahrb Wiss Bot* **69, 762–818.**

Hendrich B, Bird A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and cellular biology* **18**: 6538-6547.

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. 1999. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401**: 301-304.

Hernandez-Munoz I, Taghavi P, Kuijl C, Neefjes J, van Lohuizen M. 2005. Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. *Molecular and cellular biology* **25**: 11047-11058.

Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A. 2012. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & development* **26**: 2604-2620.

Herzing LB, Romer JT, Horn JM, Ashworth A. 1997. Xist has properties of the X-chromosome inactivation centre. *Nature* **386**: 272-275.

Hisada K, Sanchez C, Endo TA, Endoh M, Roman-Trufero M, Sharif J, Koseki H, Vidal M. 2012. RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. *Molecular and cellular biology* **32**: 1139-1149.

Hodgson JW, Argiropoulos B, Brock HW. 2001. Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Molecular and cellular biology* **21**: 4528-4543.

Hojfeldt JW, Agger K, Helin K. 2013. Histone lysine demethylases as targets for anticancer therapy. *Nature reviews Drug discovery* **12**: 917-930.

Holliday R. 1994. Epigenetics: an overview. *Developmental genetics* **15**: 453-457.

Holliday R, Pugh JE. 1975. DNA modification mechanisms and gene activity during development. *Science* **187**: 226-232.

Hooper M, Hardy K, Handyside A, Hunter S, Monk M. 1987. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**: 292-295.

Horard B, Tatout C, Poux S, Pirrotta V. 2000. Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. *Molecular and cellular biology* **20**: 3187-3197.

Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. 2013a. The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Molecular and cellular biology* **33**: 4745-4754.

Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, Shilatifard A. 2013b. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nature structural & molecular biology* **20**: 1093-1097.

Huang DH, Chang YL, Yang CC, Pan IC, King B. 2002. pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. *Molecular and cellular biology* **22**: 6261-6271.

Hunkapiller J, Shen Y, Diaz A, Cagney G, McCleary D, Ramalho-Santos M, Krogan N, Ren B, Song JS, Reiter JF. 2012. Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. *PLoS genetics* **8**: e1002576.

Hurlin PJ, Steingrimsson E, Copeland NG, Jenkins NA, Eisenman RN. 1999. Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *The EMBO journal* **18**: 7019-7028.

Huynh KD, Fischle W, Verdin E, Bardwell VJ. 2000. BCoR, a novel corepressor involved in BCL-6 repression. *Genes & development* **14**: 1810-1823.

Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP. 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS genetics* **6**: e1001134.

Ingham PW. 1983. Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. *Nature* **306: 591-593.**

-. 1985. A clonal analysis of the requirement for the trithorax gene in the diversification of segments in Drosophila. *Journal of embryology and experimental morphology* **89**: 349-365.

-. 1998. trithorax and the regulation of homeotic gene expression in Drosophila: a historical perspective. *The International journal of developmental biology* **42**: 423-429.

Inoue A, Zhang Y. 2011. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**: 194.

Isono K, Endo TA, Ku M, Yamada D, Suzuki R, Sharif J, Ishikura T, Toyoda T, Bernstein BE, Koseki H. 2013. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Developmental cell* **26**: 565-577.

Isono K, Fujimura Y, Shinga J, Yamaki M, J OW, Takihara Y, Murahashi Y, Takada Y, Mizutani-Koseki Y, Koseki H. 2005. Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of Hox genes. *Molecular and cellular biology* **25**: 6694-6706.

Issaeva I, Zonis Y, Rozovskaia T, Orlovsky K, Croce CM, Nakamura T, Mazo A, Eisenbach L, Canaani E. 2007. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Molecular and cellular biology* **27**: 1889-1903.

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**: 1129-1133.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**: 1300-1303.

Iyer LM, Abhiman S, Aravind L. 2011. Natural history of eukaryotic DNA methylation systems. *Progress in molecular biology and translational science* **101**: 25-104.

Izzo A, Kamieniarz K, Schneider R. 2008. The histone H1 family: specific members, specific functions? *Biological chemistry* **389**: 333-343.

Jablonka E, Lamb MJ. 2002. The changing concept of epigenetics. *Annals of the New York Academy of Sciences* **981**: 82-96.

Jacobs JJ, van Lohuizen M. 2002. Polycomb repression: from cellular memory to cellular proliferation and cancer. *Biochimica et biophysica acta* **1602**: 151-161.

Janzer A, Stamm K, Becker A, Zimmer A, Buettner R, Kirfel J. 2012. The H3K4me3 histone demethylase Fbxl10 is a regulator of chemokine expression, cellular morphology, and the metabolome of fibroblasts. *The Journal of biological chemistry* **287**: 30984-30992.

Jeltsch A. 2002. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem : a European journal of chemical biology* **3**: 274-293.

-. 2006. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics : official journal of the DNA Methylation Society* **1**: 63-66.

Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* **293**: 1074-1080.

Jin J, Cardozo T, Lovering RC, Elledge SJ, Pagano M, Harper JW. 2004. Systematic analysis and nomenclature of mammalian F-box proteins. *Genes & development* **18**: 2573-2580.

Jones PA, Baylin SB. 2007. The epigenomics of cancer. *Cell* **128**: 683-692.

Joo HY, Zhai L, Yang C, Nie S, Erdjument-Bromage H, Tempst P, Chang C, Wang H. 2007. Regulation of cell cycle progression and gene expression by H2A deubiquitination. *Nature* **449**: 1068-1072.

Joshi AA, Struhl K. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Molecular cell* **20**: 971-978.

Josse J, Kaiser AD, Kornberg A. 1961. Enzymatic synthesis of deoxyribonucleic acid. VIII. Frequencies of nearest neighbor base sequences in deoxyribonucleic acid. *The Journal of biological chemistry* **236**: 864-875.

Junco SE, Wang R, Gaipa JC, Taylor AB, Schirf V, Gearhart MD, Bardwell VJ, Demeler B, Hart PJ, Kim CA. 2013. Structure of the polycomb group protein PCGF1 in complex with BCOR reveals basis for binding selectivity of PCGF homologs. *Structure* **21**: 665-671.

Jung HR, Pasini D, Helin K, Jensen ON. 2010. Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Molecular & cellular proteomics : MCP* **9**: 838-850.

Junttila MR, Saarinen S, Schmidt T, Kast J, Westermarck J. 2005. Single-step Strep-tag purification for the isolation and identification of protein complexes from mammalian cells. *Proteomics* **5**: 1199-1203.

Kadosh D, Struhl K. 1998. Histone deacetylase activity of Rpd3 is important for transcriptional repression in vivo. *Genes & development* **12**: 797-805.

Kahn TG, Schwartz YB, Dellino GI, Pirrotta V. 2006. Polycomb complexes and the propagation of the methylation mark at the Drosophila ubx gene. *The Journal of biological chemistry* **281**: 29064-29075.

Kaneko S, Bonasio R, Saldana-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, D. R. 2014. Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin. *Molecular cell* **53, 1–11.**

Kaneko S, Bonasio R, Saldana-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D. 2013a. Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin. *Molecular cell*.

Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R. 2013b. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nature structural & molecular biology* **20**: 1258-1264.

Kanhere A, Viiri K, Araujo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW et al. 2010. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Molecular cell* **38**: 675-688.

Kass SU, Pruss D, Wolffe AP. 1997. How does DNA methylation repress transcription? *Trends in genetics : TIG* **13**: 444-449.

Kassis JA, Kennison JA. 2010. Recruitment of polycomb complexes: a role for SCM. *Molecular and cellular biology* **30**: 2581-2583.

Katan-Khaykovich Y, Struhl K. 2011. Splitting of H3-H4 tetramers at transcriptionally active genes undergoing dynamic histone exchange. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 1296-1301.

Katoh M, Katoh M. 2004. Identification and characterization of ASXL3 gene in silico. *International journal of oncology* **24**: 1617-1622.

Kawakami K. 2007. Tol2: a versatile gene transfer vector in vertebrates. *Genome biology* **8 Suppl 1**: S7.

Kennison JA. 2004. Introduction to Trx-G and Pc-G genes. *Methods in enzymology* **377**: 61-70.

Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ et al. 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**: 593-605.

Kerridge S, Morata G. 1982. Developmental effects of some newly induced Ultrabithorax alleles of Drosophila. *Journal of embryology and experimental morphology* **68**: 211-234.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 11667-11672.

Kim CA, Gingery M, Pilpa RM, Bowie JU. 2002. The SAM domain of polyhomeotic forms a helical polymer. *Nature structural biology* **9**: 453-457.

Kim CA, Sawaya MR, Cascio D, Kim W, Bowie JU. 2005. Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. *The Journal of biological chemistry* **280**: 27769-27775.

Kim H, Kang K, Ekram MB, Roh TY, Kim J. 2011. Aebp2 as an epigenetic regulator for neural crest cells. *PloS one* **6**: e25174.

Kim H, Kang K, Kim J. 2009. AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic acids research* **37**: 2940-2950.

Kim J, Kim JA, McGinty RK, Nguyen UT, Muir TW, Allis CD, Roeder RG. 2013. The n-SET domain of Set1 regulates H2B ubiquitylation-dependent H3K4 methylation. *Molecular cell* **49**: 1121-1133.

Kim T, Buratowski S. 2007. Two Saccharomyces cerevisiae JmjC domain proteins demethylate histone H3 Lys36 in transcribed regions to promote elongation. *The Journal of biological chemistry* **282**: 20827-20835.

Kim TG, Kraus JC, Chen J, Lee Y. 2003. JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. *The Journal of biological chemistry* **278**: 42247-42255.

King IF, Francis NJ, Kingston RE. 2002. Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. *Molecular and cellular biology* **22**: 7919-7928.

Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD. 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Molecular and cellular biology* **25**: 3305-3316.

Klimasauskas S, Szyperski T, Serva S, Wuthrich K. 1998. Dynamic modes of the flipped-out cytosine during HhaI methyltransferase-DNA interactions in solution. *The EMBO journal* **17**: 317-324.

Klose RJ, Bird AP. 2006. Genomic DNA methylation: the mark and its mediators. *Trends in biochemical sciences* **31**: 89-97.

Klose RJ, Cooper S, Farcas AM, Blackledge NP, Brockdorff N. 2013. Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. *PLoS genetics* **9**: e1003717.

Klose RJ, Gardner KE, Liang G, Erdjument-Bromage H, Tempst P, Zhang Y. 2007a. Demethylation of histone H3K36 and H3K9 by Rph1: a vestige of an H3K9 methylation system in Saccharomyces cerevisiae? *Molecular and cellular biology* **27**: 3951-3961.

Klose RJ, Kallin EM, Zhang Y. 2006. JmjC-domain-containing proteins and histone demethylation. *Nature reviews Genetics* **7**: 715-727.

Klose RJ, Yan Q, Tothova Z, Yamane K, Erdjument-Bromage H, Tempst P, Gilliland DG, Zhang Y, Kaelin WG, Jr. 2007b. The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* **128**: 889-900.

Klymenko T, Muller J. 2004. The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO reports* **5**: 373-377.

Klymenko T, Papp B, Fischle W, Kocher T, Schelder M, Fritsch C, Wild B, Wilm M, Muller J. 2006. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes & development* **20**: 1110-1122.

Knezetic JA, Luse DS. 1986. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell* **45**: 95-104.

Ko M, An J, Bandukwala HS, Chavez L, Aijo T, Pastor WA, Segal MF, Li H, Koh KP, Lahdesmaki H et al. 2013. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**: 122-126.

Kobayashi I, Nobusato A, Kobayashi-Takahashi N, Uchiyama I. 1999. Shaping the genome--restriction-modification systems as mobile genetic elements. *Current opinion in genetics & development* **9**: 649-656.

Kobe B, Deisenhofer J. 1994. The leucine-rich repeat: a versatile binding motif. *Trends in biochemical sciences* **19**: 415-421.

Kohlmaier A, Savarese F, Lachner M, Martens J, Jenuwein T, Wutz A. 2004. A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS biology* **2**: E171.

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature genetics* **41**: 376-381.

Kondo E, Gu Z, Horii A, Fukushige S. 2005. The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Molecular and cellular biology* **25**: 4388-4396.

Kooistra SM, Helin K. 2012. Molecular mechanisms and potential functions of histone demethylases. *Nature reviews Molecular cell biology* **13**: 297-311.

Kornberg RD. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**: 868-871.

Kornberg RD, Lorch Y. 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**: 285-294.

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y. 2011. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30**: 1956-1962.

Koyama-Nasu R, David G, Tanese N. 2007. The F-box protein Fbl10 is a novel transcriptional repressor of c-Jun. *Nature cell biology* **9**: 1074-1080.

Kriaucionis S, Heintz N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**: 929-930.

Krogan NJ, Dover J, Khorrami S, Greenblatt JF, Schneider J, Johnston M, Shilatifard A. 2002. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *The Journal of biological chemistry* **277**: 10753-10755.

Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M et al. 2003a. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Molecular cell* **11**: 721-729.

Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, Richards DP, Beattie BK, Emili A, Boone C et al. 2003b. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Molecular and cellular biology* **23**: 4207-4218.

Krokan HE, Drablos F, Slupphaug G. 2002. Uracil in DNA--occurrence, consequences and repair. *Oncogene* **21**: 8935-8948.

Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS et al. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS genetics* **4**: e1000242.

Kuo AJ, Cheung P, Chen K, Zee BM, Kioi M, Lauring J, Xi Y, Park BH, Shi X, Garcia BA et al. 2011. NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Molecular cell* **44**: 609-620.

Kurdistani SK, Grunstein M. 2003. Histone acetylation and deacetylation in yeast. *Nature reviews Molecular cell biology* **4**: 276-284.

Kurzhals RL, Tie F, Stratton CA, Harte PJ. 2008. Drosophila ESC-like can substitute for ESC and becomes required for Polycomb silencing if ESC is absent. *Developmental biology* **313**: 293-306.

Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & development* **16**: 2893-2905.

Kyba M, Brock HW. 1998. The SAM domain of polyhomeotic, RAE28, and scm mediates specific interactions through conserved residues. *Developmental genetics* **22**: 74-84.

Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**: 116-120.

Lagarou A, Mohd-Sarip A, Moshkin YM, Chalkley GE, Bezstarosti K, Demmers JA, Verrijzer CP. 2008. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes & development* **22**: 2799-2810.

Lai HL, Wang QT. 2013. Additional sex combs-like 2 is required for polycomb repressive complex 2 binding at select targets. *PloS one* **8**: e73983.

Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P. 2013. Improved workflows for high throughput library preparation using the transposome-based nextera system. *BMC biotechnology* **13**: 104.

Lan F, Bayliss PE, Rinn JL, Whetstine JR, Wang JK, Chen S, Iwase S, Alpatov R, Issaeva I, Canaani E et al. 2007. A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* **449**: 689-694.

Landeira D, Sauer S, Poot R, Dvorkina M, Mazzarella L, Jorgensen HF, Pereira CF, Leleu M, Piccolo FM, Spivakov M 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. *Nature cell biology* **12**: 618-624.

Langlais KK, Brown JL, Kassis JA. 2012. Polycomb group proteins bind an engrailed PRE in both the "ON" and "OFF" transcriptional states of engrailed. *PloS one* **7**: e48765.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: R25.

Lanzuolo C, Orlando V. 2012. Memories from the polycomb group proteins. *Annual review of genetics* **46**: 561-589.

Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V. 2007. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nature cell biology* **9**: 1167-1174.

Larsen F, Gundersen G, Lopez R, Prydz H. 1992. CpG islands as gene markers in the human genome. *Genomics* **13**: 1095-1107.

Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WA, Cohen A, Lasonder E, Stunnenberg HG. 2006. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Molecular and cellular biology* **26**: 843-851.

Lee JH, Skalnik DG. 2005. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *The Journal of biological chemistry* **280**: 41725-41731.

Lee JH, Voo KS, Skalnik DG. 2001. Identification and characterization of the DNA binding domain of CpG-binding protein. *The Journal of biological chemistry* **276**: 44669-44676.

Lee JS, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, Bhaumik SR, Shilatifard A. 2007a. Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* **131**: 1084-1096.

Lee JT, Jaenisch R. 1997. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature* **386**: 275-279.

Lee MG, Villa R, Trojer P, Norman J, Yan KP, Reinberg D, Di Croce L, Shiekhattar R. 2007b. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**: 447-450.

Lee N, Zhang J, Klose RJ, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. 2007c. The trithorax-group protein Lid is a histone H3 trimethyl-Lys4 demethylase. *Nature structural & molecular biology* **14**: 341-343.

Lee TF, Zhai J, Meyers BC. 2010. Conservation and divergence in eukaryotic DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 9027-9028.

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K et al. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**: 301-313.

Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A. 2010. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & development* **24**: 265-276.

Leeb M, Wutz A. 2007. Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *The Journal of cell biology* **178**: 219-229.

Lehmann L, Ferrari R, Vashisht AA, Wohlschlegel JA, Kurdistani SK, Carey M. 2012. Polycomb repressive complex 1 (PRC1) disassembles RNA polymerase II preinitiation complexes. *The Journal of biological chemistry* **287**: 35784-35794.

Lehmann R, Nusslein-Volhard C. 1987. hunchback, a gene required for segmentation of an anterior and posterior region of the Drosophila embryo. *Developmental biology* **119**: 402-417.

Leonhardt H, Page AW, Weier HU, Bestor TH. 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**: 865-873.

Levine SS, Weiss A, Erdjument-Bromage H, Shao Z, Tempst P, Kingston RE. 2002. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Molecular and cellular biology* **22**: 6070-6078.

Lewis EB. 1978. A gene complex controlling segmentation in Drosophila. *Nature* **276**: 565-570.

Li B, Howe L, Anderson S, Yates JR, 3rd, Workman JL. 2003. The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *The Journal of biological chemistry* **278**: 8897-8903.

Li B, Jackson J, Simon MD, Fleharty B, Gogol M, Seidel C, Workman JL, Shilatifard A. 2009a. Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *The Journal of biological chemistry* **284**: 7970-7976.

Li E, Beard C, Jaenisch R. 1993. Role for DNA methylation in genomic imprinting. *Nature* **366**: 362-365.

Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**: 915-926.

Li G, Margueron R, Ku M, Chambon P, Bernstein BE, Reinberg D. 2010. Jarid2 and PRC2, partners in regulating gene expression. *Genes & development* **24**: 368-380.

Li J, Wang J, Wang J, Nawaz Z, Liu JM, Qin J, Wong J. 2000. Both corepressor proteins SMRT and N-CoR exist in large protein complexes containing HDAC3. *The EMBO journal* **19**: 4342-4350.

Li Y, Trojer P, Xu CF, Cheung P, Kuo A, Drury WJ, 3rd, Qiao Q, Neubert TA, Xu RM, Gozani O et al. 2009b. The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *The Journal of biological chemistry* **284**: 34283-34295.

Li Z, Cao R, Wang M, Myers MP, Zhang Y, Xu RM. 2006. Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. *The Journal of biological chemistry* **281**: 20643-20649.

Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ. 2005. Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS biology* **3**: e328.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.

Long HK, Blackledge NP, Klose RJ. 2013a. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society transactions* **41**: 727-740.

Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grutzner F, Odom DT, Patient R, Ponting CP et al. 2013b. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *eLife* **2**: e00348.

Lorch Y, LaPointe JW, Kornberg RD. 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**: 203-210.

Lu T, Yang M, Huang DB, Wei H, Ozer GH, Ghosh G, Stark GR. 2013. Role of lysine methylation of NF-kappaB in differential gene regulation. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 13510-13515.

Lucio-Eterovic AK, Singh MM, Gardner JE, Veerappan CS, Rice JC, Carpenter PB. 2010. Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 16952-16957.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**: 251-260.

Lukas C, Savic V, Bekker-Jensen S, Doil C, Neumann B, Pedersen RS, Grofte M, Chan KL, Hickson ID, Bartek J et al. 2011. 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nature cell biology* **13**: 243-253.

Lynch MD, Smith AJ, De Gobbi M, Flenley M, Hughes JR, Vernimmen D, Ayyub H, Sharpe JA, Sloane-Stanley JA, Sutherland L et al. 2012. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *The EMBO journal* **31**: 317-329.

Lyon MF. 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* **190**: 372-373.

Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F et al. 2013. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nature neuroscience* **16**: 898-902.

Ma Q, Alder H, Nelson KK, Chatterjee D, Gu Y, Nakamura T, Canaani E, Croce CM, Siracusa LD, Buchberg AM. 1993. Analysis of the murine All-1 gene reveals conserved domains with human ALL-1 and identifies a motif shared with DNA methyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 6350-6354.

Maeda I, Okamura D, Tokitake Y, Ikeda M, Kawaguchi H, Mise N, Abe K, Noce T, Okuda A, Matsui Y. 2013. Max is a repressor of germ cell-related gene expression in mouse embryonic stem cells. *Nature communications* **4**: 1754.

Maeshima K, Hihara S, Eltsov M. 2010. Chromatin structure: does the 30-nm fibre exist in vivo? *Current opinion in cell biology* **22**: 291-297.

Mallo M, Wellik DM, Deschamps J. 2010. Hox genes and regional patterning of the vertebrate body plan. *Developmental biology* **344**: 7-15.

Mao YS, Zhang B, Spector DL. 2011. Biogenesis and function of nuclear bodies. *Trends in genetics : TIG* **27**: 295-306.

Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ, 3rd, Voigt P, Martin SR, Taylor WR, De Marco V et al. 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**: 762-767.

Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD, Reinberg D. 2008. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Molecular cell* **32**: 503-518.

Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, 3rd, Hurov KE, Luo J, Bakalarski CE, Zhao Z, Solimini N, Lerenthal Y et al. 2007. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**: 1160-1166.

Maunakea AK, Chepelev I, Cui K, Zhao K. 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell research* **23**: 1256-1269.

McMahon AP, Bradley A. 1990. The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62**: 1073-1085.

Meehan RR, Lewis JD, Bird AP. 1992. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic acids research* **20**: 5085-5092.

Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE. 2010. GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS genetics* **6**: e1001244.

Meselson M, Yuan R, Heywood J. 1972. Restriction and modification of DNA. *Annual review of biochemistry* **41**: 447-466.

Messmer S, Franke A, Paro R. 1992. Analysis of the functional role of the Polycomb chromo domain in Drosophila melanogaster. *Genes & development* **6**: 1241-1254.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553-560.

Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, Keightley PD, Bishop SM, Clarke AR, Bird A. 2002. Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297**: 403-405.

Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A. 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 12902-12907.

Milne TA, Dou Y, Martin ME, Brock HW, Roeder RG, Hess JL. 2005. MLL associates specifically with a subset of transcriptionally active target genes. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 14765-14770.

Milne TA, Sinclair DA, Brock HW. 1999. The Additional sex combs gene of Drosophila is required for activation and repression of homeotic loci, and interacts specifically with Polycomb and super sex combs. *Molecular & general genetics : MGG* **261**: 753-761.

Min J, Zhang Y, Xu RM. 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & development* **17**: 1823-1828.

Miyazaki H, Higashimoto K, Yada Y, Endo TA, Sharif J, Komori T, Matsuda M, Koseki Y, Nakayama M, Soejima H et al. 2013. Ash1l methylates Lys36 of histone H3 independently of transcriptional elongation to counteract polycomb silencing. *PLoS genetics* **9**: e1003897.

Moazed D, O'Farrell PH. 1992. Maintenance of the engrailed expression pattern by Polycomb group genes in Drosophila. *Development* **116**: 805-810.

Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, Piqani B, Eisenhaure TM, Luo B, Grenier JK et al. 2006. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**: 1283-1298.

Mohandas T, Sparkes RS, Shapiro LJ. 1981. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* **211**: 393-396.

Mohd-Sarip A, van der Knaap JA, Wyman C, Kanaar R, Schedl P, Verrijzer CP. 2006. Architecture of a polycomb nucleoprotein complex. *Molecular cell* **24**: 91-100.

Mohd-Sarip A, Venturini F, Chalkley GE, Verrijzer CP. 2002. Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. *Molecular and cellular biology* **22**: 7473-7483.

Moore JD, Krebs JE. 2004. Histone modifications and DNA double-strand break repair. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **82**: 446-452.

Morey L, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah SA, Di Croce L. 2012. Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell stem cell* **10**: 47-62.

Mujtaba S, Manzur KL, Gurnon JR, Kang M, Van Etten JL, Zhou MM. 2008. Epigenetic transcriptional repression of cellular genes by a viral SET protein. *Nature cell biology* **10**: 1114-1122.

Mulholland NM, King IF, Kingston RE. 2003. Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. *Genes & development* **17**: 2741-2746.

Muller J, Gaunt S, Lawrence PA. 1995. Function of the Polycomb protein is conserved in mice and flies. *Development* **121**: 2847-2852.

Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. 2002. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**: 197-208.

Muller J, Kassis JA. 2006. Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Current opinion in genetics & development* **16**: 476-484.

Murray K. 1964. The Occurrence of Epsilon-N-Methyl Lysine in Histones. *Biochemistry* **3**: 10-15.

Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K. 2007. RNA polymerase is poised for activation across the genome. *Nature genetics* **39**: 1507-1511.

Musselman CA, Lalonde ME, Cote J, Kutateladze TG. 2012. Perceiving the epigenetic landscape through histone readers. *Nature structural & molecular biology* **19**: 1218-1227.

Musselman CA, Mansfield RE, Garske AL, Davrazou F, Kwan AH, Oliver SS, O'Leary H, Denu JM, Mackay JP, Kutateladze TG. 2009. Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *The Biochemical journal* **423**: 179-187.

Nagy PL, Griesenbeck J, Kornberg RD, Cleary ML. 2002. A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 90-94.

Nakagawa T, Kajitani T, Togo S, Masuko N, Ohdan H, Hishikawa Y, Koji T, Matsuyama T, Ikura T, Muramatsu M et al. 2008. Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes & development* **22**: 37-49.

Nakanishi S, Lee JS, Gardner KE, Gardner JM, Takahashi YH, Chandrasekharan MB, Sun ZW, Osley MA, Strahl BD, Jaspersen SL et al. 2009. Histone H2BK123 monoubiquitination is the critical determinant for H3K4 and H3K79 trimethylation by COMPASS and Dot1. *The Journal of cell biology* **186**: 371-377.

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**: 386-389.

Neddermann P, Jiricny J. 1993. The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *The Journal of biological chemistry* **268**: 21218-21224.

Nekrasov M, Klymenko T, Fraterman S, Papp B, Oktaba K, Kocher T, Cohen A, Stunnenberg HG, Wilm M, Muller J. 2007. Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *The EMBO journal* **26**: 4078-4088.

Nekrasov M, Wild B, Muller J. 2005. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *EMBO reports* **6**: 348-353.

Ng HH, Robert F, Young RA, Struhl K. 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell* **11**: 709-719.

Nichols J, Evans EP, Smith AG. 1990. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110**: 1341-1348.

Nichols WW, Murphy DG, Cristofalo VJ, Toji LH, Greene AE, Dwight SA. 1977. Characterization of a new human diploid cell strain, IMR-90. *Science* **196**: 60-63.

Nishino Y, Eltsov M, Joti Y, Ito K, Takata H, Takahashi Y, Hihara S, Frangakis AS, Imamoto N, Ishikawa T et al. 2012. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO journal* **31**: 1644-1653.

Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, Shimamura S, Arita K, Kodama T, Ishikawa F et al. 2013. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* **502**: 249-253.

Nowak AJ, Alfieri C, Stirnimann CU, Rybin V, Baudin F, Ly-Hartig N, Lindner D, Muller CW. 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. *The Journal of biological chemistry* **286**: 23388-23396.

O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, Jenuwein T. 2001. The polycomb-group gene Ezh2 is required for early mouse development. *Molecular and cellular biology* **21**: 4330-4336.

O'Gara M, Horton JR, Roberts RJ, Cheng X. 1998. Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base. *Nature structural biology* **5**: 872-877.

O'Loghlen A, Munoz-Cabello AM, Gaspar-Maia A, Wu HA, Banito A, Kunowska N, Racek T, Pemberton HN, Beolchi P, Lavial F et al. 2012. MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell stem cell* **10**: 33-46.

Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. 2002. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296**: 1132-1136.

Ohno K, McCabe D, Czermin B, Imhof A, Pirrotta V. 2008. ESC, ESCL and their roles in Polycomb Group mechanisms. *Mechanisms of development* **125**: 527-541.

Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**: 247-257.

Okano M, Xie S, Li E. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature genetics* **19**: 219-220.

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD et al. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**: 714-717.

Orlando V, Jane EP, Chinwalla V, Harte PJ, Paro R. 1998. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. *The EMBO journal* **17**: 5141-5150.

Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, Dean W, Reik W, Walter J. 2000. Active demethylation of the paternal genome in the mouse zygote. *Current biology : CB* **10**: 475-478.

Pagan JK, Arnold J, Hanchard KJ, Kumar R, Bruno T, Jones MJ, Richard DJ, Forrest A, Spurdle A, Verdin E et al. 2007. A novel corepressor, BCoR-L1, represses transcription through an interaction with CtBP. *The Journal of biological chemistry* **282**: 15248-15257.

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C. 2008. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular cell* **32**: 232-246.

Panning B, Jaenisch R. 1996. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes & development* **10**: 1991-2002.

Papp B, Muller J. 2006. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes & development* **20**: 2041-2054.

Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K. 2007. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Molecular and cellular biology* **27**: 3769-3779.

Pasini D, Bracken AP, Jensen MR, Lazzerini Denchi E, Helin K. 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *The EMBO journal* **23**: 4061-4071.

Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappsilber J, Helin K. 2010a. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**: 306-310.

Pasini D, Hansen KH, Christensen J, Agger K, Cloos PA, Helin K. 2008. Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes & development* **22**: 1345-1355.

Pasini D, Malatesta M, Jung HR, Walfridsson J, Willer A, Olsson L, Skotte J, Wutz A, Porse B, Jensen ON 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. *Nucleic acids research* **38**: 4958-4969.

Passmore LA, Barford D. 2004. Getting into position: the catalytic mechanisms of protein ubiquitylation. *The Biochemical journal* **379**: 513-525.

Pastor WA, Aravind L, Rao A. 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature reviews Molecular cell biology* **14**: 341-356.

Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P et al. 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**: 394-397.

Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J. 2009. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**: 1290-1302.

Pengelly AR, Copur O, Jackle H, Herzig A, Muller J. 2013. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science* **339**: 698-699.

Peters AH, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AA, Perez-Burgos L, Kohlmaier A, Opravil S, Tachibana M, Shinkai Y et al. 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Molecular cell* **12**: 1577-1589.

Peterson AJ, Kyba M, Bornemann D, Morgan K, Brock HW, Simon J. 1997. A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. *Molecular and cellular biology* **17**: 6683-6692.

Petruk S, Sedkov Y, Johnston DM, Hodgson JW, Black KL, Kovermann SK, Beck S, Canaani E, Brock HW, Mazo A. 2012. TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. *Cell* **150**: 922-933.

Petruk S, Sedkov Y, Smith S, Tillib S, Kraevski V, Nakamura T, Canaani E, Croce CM, Mazo A. 2001. Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene. *Science* **294**: 1331-1334.

Pfau R, Tzatsos A, Kampranis SC, Serebrennikova OB, Bear SE, Tsichlis PN. 2008. Members of a family of JmjC domain-containing oncoproteins immortalize embryonic fibroblasts via a JmjC domain-dependent process. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 1907-1912.

Pirrotta V. 1997. PcG complexes and chromatin silencing. *Current opinion in genetics & development* **7**: 249-258.

Pirrotta V, Li HB. 2012. A view of nuclear Polycomb bodies. *Current opinion in genetics & development* **22**: 101-109.

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**: 131-135.

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**: 517-527.

Ponten J, Saksela E. 1967. Two established in vitro cell lines from human mesenchymal tumours. *International journal of cancer Journal international du cancer* **2**: 434-447.

Potter CJ, Luo L. 2010. Splinkerette PCR for mapping transposable elements in Drosophila. *PloS one* **5**: e10168.

Poux S, Horard B, Sigrist CJ, Pirrotta V. 2002. The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. *Development* **129**: 2483-2493.

Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA. 2012. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS genetics* **8**: e1002717.

Prokhortchouk A, Hendrich B, Jorgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E. 2001. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & development* **15**: 1613-1618.

Ptashne M. 2013. Epigenetics: core misconcept. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 7101-7103.

Qin J, Whyte WA, Anderssen E, Apostolou E, Chen HH, Akbarian S, Bronson RT, Hochedlinger K, Ramaswamy S, Young RA et al. 2012. The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. *Cell stem cell* **11**: 319-332.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox AV, Andrews TD, Howe KL, Otto T, Olek A et al. 2004. DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS biology* **2**: e405.

Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M, Smale ST. 2009. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**: 114-128.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 5237-5242.

Ramsey SA, Knijnenburg TA, Kennedy KA, Zak DE, Gilchrist M, Gold ES, Johnson CD, Lampano AE, Litvak V, Navarro G et al. 2010. Genome-wide histone acetylation data improve prediction of mammalian transcription factor binding sites. *Bioinformatics* **26**: 2071-2075.

Rando OJ, Winston F. 2012. Chromatin and transcription in yeast. *Genetics* **190**: 351-387.

Rao B, Shibata Y, Strahl BD, Lieb JD. 2005. Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Molecular and cellular biology* **25**: 9447-9459.

Ray-Gallet D, Almouzni G. 2010. Molecular biology. Mixing or not mixing. *Science* **328**: 56-57.

Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, Lerouge T, Hager GL, Chambon P, Losson R. 2003. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *The EMBO journal* **22**: 3153-3163.

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**: 593-599.

Reddington JP, Perricone SM, Nestor CE, Reichmann J, Youngson NA, Suzuki M, Reinhardt D, Dunican DS, Prendergast JG, Mjoseng H et al. 2013. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome biology* **14**: R25.

Ren X, Kerppola TK. 2011. REST interacts with Cbx proteins and regulates polycomb repressive complex 1 occupancy at RE1 elements. *Molecular and cellular biology* **31**: 2100-2110.

Reynolds N, Salmon-Divon M, Dvinge H, Hynes-Allen A, Balasooriya G, Leaford D, Behrens A, Bertone P, Hendrich B. 2012. NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *The EMBO journal* **31**: 593-605.

Richly H, Rocha-Viegas L, Ribeiro JD, Demajo S, Gundem G, Lopez-Bigas N, Nakagawa T, Rospert S, Ito T, Di Croce L. 2010. Transcriptional activation of polycomb-repressed genes by ZRF1. *Nature* **468**: 1124-1128.

Riggs AD. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenetics and cell genetics* **14**: 9-25.

Ringrose L, Paro R. 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annual review of genetics* **38**: 413-443.

-. 2007. Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* **134**: 223-232.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E et al. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**: 1311-1323.

Robertson KD, Wolffe AP. 2000. DNA methylation in health and disease. *Nature reviews Genetics* **1**: 11-19.

Robinson AK, Leal BZ, Chadwell LV, Wang R, Ilangovan U, Kaur Y, Junco SE, Schirf V, Osmulski PA, Gaczynska M et al. 2012. The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. *The Journal of biological chemistry* **287**: 8702-8713.

Roguev A, Schaft D, Shevchenko A, Pijnappel WW, Wilm M, Aasland R, Stewart AF. 2001. The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *The EMBO journal* **20**: 7137-7148.

Roh TY, Cuddapah S, Cui K, Zhao K. 2006. The genomic landscape of histone modifications in human T cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 15782-15787.

Russo VEA, Martienssen, R. A. & Riggs, A. D. (eds). 1996. Epigenetic Mechanisms of Gene Regulation *Cold Spring Harbor Laboratory Press, Woodbury*.

Ruthenburg AJ, Allis CD, Wysocka J. 2007. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell* **25**: 15-30.

Sabo PJ, Humbert R, Hawrylycz M, Wallace JC, Dorschner MO, McArthur M, Stamatoyannopoulos JA. 2004. Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 4537-4542.

Sambrook J, Russell DW. 2001. Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY*.

Sanchez-Pulido L, Kong L, Ponting CP. 2012. A common ancestry for BAP1 and Uch37 regulators. *Bioinformatics* **28**: 1953-1956.

Sanchez C, Sanchez I, Demmers JA, Rodriguez P, Strouboulis J, Vidal M. 2007. Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. *Molecular & cellular proteomics : MCP* **6**: 820-834.

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J, Kouzarides T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**: 407-411.

Sarraf SA, Stancheva I. 2004. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular cell* **15**: 595-605.

Sato T, Denell RE. 1985. Homoeosis in Drosophila: anterior and posterior transformations of Polycomb lethal embryos. *Developmental biology* **110**: 53-64.

Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE. 2001. A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature* **412**: 655-660.

Saurin AJ, Shiels C, Williamson J, Satijn DP, Otte AP, Sheer D, Freemont PS. 1998. The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *The Journal of cell biology* **142**: 887-898.

Saxonov S, Berg P, Brutlag DL. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 1412-1417.

Scherer WF, Syverton JT, Gey GO. 1953. Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *The Journal of experimental medicine* **97**: 695-710.

Schermelleh L, Haemmer A, Spada F, Rosing N, Meilinger D, Rothbauer U, Cardoso MC, Leonhardt H. 2007. Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic acids research* **35**: 4301-4312.

Scheuermann JC, de Ayala Alonso AG, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, Wilm M, Muir TW, Muller J. 2010. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**: 243-247.

Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**: 240-248.

Schmidt TG, Skerra A. 2007. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature protocols* **2**: 1528-1535.

Schmitges FW, Prusty AB, Faty M, Stutzer A, Lingaraju GM, Aiwazian J, Sack R, Hess D, Li L, Zhou S et al. 2011. Histone methylation by PRC2 is inhibited by active chromatin marks. *Molecular cell* **42**: 330-341.

Schmitt S, Prestel M, Paro R. 2005. Intergenic transcription through a polycomb group response element counteracts silencing. *Genes & development* **19**: 697-708.

Schoeftner S, Sengupta AK, Kubicek S, Mechtler K, Spahn L, Koseki H, Jenuwein T, Wutz A. 2006. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *The EMBO journal* **25**: 3110-3122.

Schoenherr CJ, Anderson DJ. 1995. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**: 1360-1363.

Schuettengruber B, Ganapathi M, Leblanc B, Portoso M, Jaschek R, Tolhuis B, van Lohuizen M, Tanay A, Cavalli G. 2009. Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS biology* **7**: e13.

Schuettengruber B, Martinez AM, Iovino N, Cavalli G. 2011. Trithorax group proteins: switching genes on and keeping them active. *Nature reviews Molecular cell biology* **12**: 799-814.

Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER, Finnin MS, Elledge SJ, Harper JW, Pagano M, Pavletich NP. 2000. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* **408**: 381-386.

Schultz LB, Chehab NH, Malikzay A, Halazonetis TD. 2000. p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *The Journal of cell biology* **151**: 1381-1390.

Schwartz YB, Kahn TG, Stenberg P, Ohno K, Bourgon R, Pirrotta V. 2010. Alternative epigenetic chromatin states of polycomb target genes. *PLoS genetics* **6**: e1000805.

Schwartz YB, Pirrotta V. 2013a. A new world of Polycombs: unexpected partnerships and emerging functions. *Nature reviews Genetics* **14**: 853-864.

-. 2013b. A new world of Polycombs: unexpected partnerships and emerging functions. *Nature reviews Genetics*.

Schwendemann A, Lehmann M. 2002. Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 12883-12888.

Sedkov Y, Cho E, Petruk S, Cherbas L, Smith ST, Jones RS, Cherbas P, Canaani E, Jaynes JB, Mazo A. 2003. Methylation at lysine 4 of histone H3 in ecdysone-dependent development of Drosophila. *Nature* **426**: 78-83.

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. 2008. Divergent transcription from active promoters. *Science* **322**: 1849-1851.

Seisenberger S, Peat JR, Reik W. 2013. Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells. *Current opinion in cell biology* **25**: 281-288.

Sengupta AK, Kuhrs A, Muller J. 2004. General transcriptional silencing by a Polycomb response element in Drosophila. *Development* **131**: 1959-1965.

Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE. 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**: 37-46.

Sharif J, Endo TA, Ito S, Ohara O, Koseki H. 2013. Embracing change to remain the same: conservation of polycomb functions despite divergence of binding motifs among species. *Current opinion in cell biology* **25**: 305-313.

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K et al. 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**: 908-912.

Shema-Yaacoby E, Nikolov M, Haj-Yahya M, Siman P, Allemand E, Yamaguchi Y, Muchardt C, Urlaub H, Brik A, Oren M et al. 2013. Systematic identification of proteins binding to chromatin-embedded ubiquitylated H2B reveals recruitment of SWI/SNF to regulate transcription. *Cell reports* **4**: 601-608.

Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan GC, Lee Y, Orkin SH. 2009. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* **139**: 1303-1314.

Shen X, Liu Y, Hsu YJ, Fujiwara Y, Kim J, Mao X, Yuan GC, Orkin SH. 2008. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Molecular cell* **32**: 491-502.

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**: 941-953.

Shiio Y, Eisenman RN. 2003. Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 13225-13230.

Shilatifard A. 2008. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Current opinion in cell biology* **20**: 341-348.

Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847.

Silva J, Mak W, Zvetkova I, Appanah R, Nesterova TB, Webster Z, Peters AH, Jenuwein T, Otte AP, Brockdorff N. 2003. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Developmental cell* **4**: 481-495.

Silverman JS, Skaar JR, Pagano M. 2012. SCF ubiquitin ligases in the maintenance of genome stability. *Trends in biochemical sciences* **37**: 66-73.

Simon J, Chiang A, Bender W, Shimell MJ, O'Connor M. 1993. Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. *Developmental biology* **158**: 131-144.

Simon JA, Kingston RE. 2009. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature reviews Molecular cell biology* **10**: 697-708.

-. 2013. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Molecular cell* **49**: 808-824.

Sinclair DA, Campbell RB, Nicholls F, Slade E, Brock HW. 1992. Genetic analysis of the additional sex combs locus of Drosophila melanogaster. *Genetics* **130**: 817-825.

Sinclair DA, Milne TA, Hodgson JW, Shellard J, Salinas CA, Kyba M, Randazzo F, Brock HW. 1998. The Additional sex combs gene of Drosophila encodes a chromatin protein that binds to shared and unique Polycomb group sites on polytene chromosomes. *Development* **125**: 1207-1216.

Sing A, Pannell D, Karaiskakis A, Sturgeon K, Djabali M, Ellis J, Lipshitz HD, Cordes SP. 2009. A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* **138**: 885-897.

Skaar JR, Pagan JK, Pagano M. 2013. Mechanisms and function of substrate recruitment by F-box proteins. *Nature reviews Molecular cell biology* **14**: 369-381.

Skaar JR, Pagano M. 2009. Control of cell growth by the SCF and APC/C ubiquitin ligases. *Current opinion in cell biology* **21**: 816-824.

Smigova J, Juda P, Cmarko D, Raska I. 2011. Fine structure of the "PcG body" in human U-2 OS cells established by correlative light-electron microscopy. *Nucleus* **2**: 219-228.

Song J, Rechkoblit O, Bestor TH, Patel DJ. 2011. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**: 1036-1040.

Srinivasan S, Dorighi KM, Tamkun JW. 2008. Drosophila Kismet regulates histone H3 lysine 27 methylation and early elongation by RNA polymerase II. *PLoS genetics* **4**: e1000217.

Sterner DE, Berger SL. 2000. Acetylation of histones and transcription-related factors. *Microbiology and molecular biology reviews : MMBR* **64**: 435-459.

Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, Brockdorff N, Fisher AG, Pombo A. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature cell biology* **9**: 1428-1435.

Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* **403**: 41-45.

Strahl BD, Grant PA, Briggs SD, Sun ZW, Bone JR, Caldwell JA, Mollah S, Cook RG, Shabanowitz J, Hunt DF et al. 2002. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Molecular and cellular biology* **22**: 1298-1306.

Strahl BD, Ohba R, Cook RG, Allis CD. 1999. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 14967-14972.

Struhl G. 1981a. A gene product required for correct initiation of segmental determination in Drosophila. *Nature* **293**: 36-41.

-. 1981b. A homoeotic mutation transforming leg to antenna in Drosophila. *Nature* **292**: 635-638.

Struhl G, Akam M. 1985. Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. *The EMBO journal* **4**: 3259-3264.

Strutt H, Paro R. 1997. The polycomb group protein complex of Drosophila melanogaster has different compositions at different target genes. *Molecular and cellular biology* **17**: 6773-6783.

Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. 2004. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *The Journal of biological chemistry* **279**: 27816-27823.

Suganuma T, Pattenden SG, Workman JL. 2008. Diverse functions of WD40 repeat proteins in histone recognition. *Genes & development* **22**: 1265-1268.

Surani MA. 1998. Imprinting and the initiation of gene silencing in the germ line. *Cell* **93**: 309-312.

Suzuki T, Minehata K, Akagi K, Jenkins NA, Copeland NG. 2006. Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice. *The EMBO journal* **25**: 3422-3431.

Szerlong HJ, Hansen JC. 2011. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **89**: 24-34.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L et al. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930-935.

Takai D, Jones PA. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 3740-3745.

Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, Tajima S. 2011. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proceedings of the National Academy of Sciences of the United States of America* **108**: 9055-9059.

Takihara Y, Tomotsune D, Shirai M, Katoh-Fukui Y, Nishii K, Motaleb MA, Nomura M, Tsuchiya R, Fujita Y, Shibata Y et al. 1997. Targeted disruption of the mouse homologue of the Drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. *Development* **124**: 3673-3682.

Tan L, Shi YG. 2012. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**: 1895-1902.

Tan MK, Lim HJ, Bennett EJ, Shi Y, Harper JW. 2013. Parallel SCF Adaptor Capture Proteomics Reveals a Role for SCF(FBXL17) in NRF2 Activation via BACH1 Repressor Turnover. *Molecular cell* **52**: 9-24.

Tanaka Y, Okamoto K, Teye K, Umata T, Yamagiwa N, Suto Y, Zhang Y, Tsuneoka M. 2010. JmjC enzyme KDM2A is a regulator of rRNA transcription in response to starvation. *The EMBO journal* **29**: 1510-1522.

Tavares L, Dimitrova E, Oxley D, Webster J, Poot R, Demmers J, Bezstarosti K, Taylor S, Ura H, Koide H et al. 2012. RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**: 664-678.

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature structural & molecular biology* **14**: 1025-1040.

Tazi J, Bird A. 1990. Alternative chromatin structure at CpG islands. *Cell* **60**: 909-920.

Testoni S, Bartolone E, Rossi M, Patrignani A, Bruggmann R, Lichtner P, Tetens J, Gentile A, Drogemuller C. 2012. KDM2B is implicated in bovine lethal multi-organic developmental dysplasia. *PloS one* **7**: e45634.

Teytelman L, Thurtle DM, Rine J, van Oudenaarden A. 2013. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 18602-18607.

Thompson JP, Granoff A, Willis DB. 1986. Trans-activation of a methylated adenovirus promoter by a frog virus 3 protein. *Proceedings of the National Academy of Sciences of the United States of America* **83**: 7688-7692.

-. 1988. Methylation of the promoter for an immediate-early frog virus 3 gene does not inhibit transcription. *Journal of virology* **62**: 4680-4685.

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD et al. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**: 1082-1086.

Tiacci E, Grossmann V, Martelli MP, Kohlmann A, Haferlach T, Falini B. 2012. The corepressors BCOR and BCORL1: two novel players in acute myeloid leukemia. *Haematologica* **97**: 3-5.

Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, Diaz MO, Scacheri PC, Harte PJ. 2009. CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* **136**: 3131-3141.

Tillib S, Petruk S, Sedkov Y, Kuzin A, Fujioka M, Goto T, Mazo A. 1999. Trithorax- and Polycomb-group response elements within an Ultrabithorax transcription maintenance unit consist of closely situated but separable sequences. *Molecular and cellular biology* **19**: 5189-5202.

Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM. 2004. An abundance of bidirectional promoters in the human genome. *Genome research* **14**: 62-66.

Trojer P, Cao AR, Gao Z, Li Y, Zhang J, Xu X, Li G, Losson R, Erdjument-Bromage H, Tempst P et al. 2011. L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Molecular cell* **42**: 438-450.

Trojer P, Reinberg D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? *Molecular cell* **28**: 1-13.

Tropberger P, Schneider R. 2010. Going global: novel histone modifications in the globular domain of H3. *Epigenetics : official journal of the DNA Methylation Society* **5**: 112-117.

Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y. 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**: 811-816.

Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR et al. 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to cells : devoted to molecular & cellular mechanisms* **11**: 805-814.

van Attikum H, Gasser SM. 2009. Crosstalk between histone modifications during the DNA damage response. *Trends in cell biology* **19**: 207-217.

van der Lugt NM, Alkema M, Berns A, Deschamps J. 1996. The Polycomb-group homolog Bmi-1 is a regulator of murine Hox gene expression. *Mechanisms of development* **58**: 153-164.

Vandamme J, Volkel P, Rosnoblet C, Le Faou P, Angrand PO. 2011. Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Molecular & cellular proteomics : MCP* **10**: M110 002642.

Venolia L, Gartler SM, Wassman ER, Yen P, Mohandas T, Shapiro LJ. 1982. Transformation with DNA from 5-azacytidine-reactivated X chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **79**: 2352-2354.

Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**: 58-69.

Vettese-Dadey M, Grant PA, Hebbes TR, Crane- Robinson C, Allis CD, Workman JL. 1996. Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *The EMBO journal* **15**: 2508-2518.

Voigt P, LeRoy G, Drury WJ, 3rd, Zee BM, Son J, Beck DB, Young NL, Garcia BA, Reinberg D. 2012. Asymmetrically modified nucleosomes. *Cell* **151**: 181-193.

Voigt P, Tee WW, Reinberg D. 2013. A double take on bivalent promoters. *Genes & development* **27**: 1318-1338.

Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG. 2000. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Molecular and cellular biology* **20**: 2108-2121.

Waddington CH. 1942. The epigenotype. *Endeavour* **1: 18–20.**

Wagner EJ, Carpenter PB. 2012. Understanding the language of Lys36 methylation at histone H3. *Nature reviews Molecular cell biology* **13**: 115-126.

Walker E, Chang WY, Hunkapiller J, Cagney G, Garcha K, Torchia J, Krogan NJ, Reiter JF, Stanford WL. 2010. Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell stem cell* **6**: 153-166.

Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y. 2004a. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**: 873-878.

Wang J, Tadeo X, Hou H, Tu PG, Thompson J, Yates JR, 3rd, Jia S. 2013. Epe1 recruits BET family bromodomain protein Bdf2 to establish heterochromatin boundaries. *Genes & development* **27**: 1886-1902.

Wang L, Brown JL, Cao R, Zhang Y, Kassis JA, Jones RS. 2004b. Hierarchical recruitment of polycomb group silencing complexes. *Molecular cell* **14**: 637-646.

Wang P, Lin C, Smith ER, Guo H, Sanderson BW, Wu M, Gogol M, Alexander T, Seidel C, Wiedemann LM et al. 2009. Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Molecular and cellular biology* **29**: 6074-6085.

Wang R, Taylor AB, Leal BZ, Chadwell LV, Ilangovan U, Robinson AK, Schirf V, Hart PJ, Lafer EM, Demeler B et al. 2010. Polycomb group targeting through different binding partners of RING1B C-terminal domain. *Structure* **18**: 966-975.

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics* **40**: 897-903.

Watt F, Molloy PL. 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & development* **2**: 1136-1143.

Weake VM, Workman JL. 2008. Histone ubiquitination: triggering gene activity. *Molecular cell* **29**: 653-663.

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics* **39**: 457-466.

Wei D, Sun Y. 2010. Small RING Finger Proteins RBX1 and RBX2 of SCF E3 Ubiquitin Ligases: The Role in Cancer and as Cancer Targets. *Genes & cancer* **1**: 700-707.

Wen YD, Perissi V, Staszewski LM, Yang WM, Krones A, Glass CK, Rosenfeld MG, Seto E. 2000. The histone deacetylase-3 complex contains nuclear receptor corepressors. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 7202-7207.

Wilkinson FH, Park K, Atchison ML. 2006. Polycomb recruitment to DNA in vivo by the YY1 REPO domain. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 19296-19301.

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K. 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**: 343-348.

Winkler GS, Kristjuhan A, Erdjument-Bromage H, Tempst P, Svejstrup JQ. 2002. Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 3517-3522.

Winston JT, Koepp DM, Zhu C, Elledge SJ, Harper JW. 1999. A family of mammalian F-box proteins. *Current biology : CB* **9**: 1180-1182.

Woo CJ, Kharchenko PV, Daheron L, Park PJ, Kingston RE. 2010. A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* **140**: 99-110.

-. 2013. Variable requirements for DNA-binding proteins at polycomb-dependent repressive regions in human HOX clusters. *Molecular and cellular biology* **33**: 3274-3285.

Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**: 389-393.

Wu L, Lee SY, Zhou B, Nguyen UT, Muir TW, Tan S, Dou Y. 2013a. ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes. *Molecular cell* **49**: 1108-1120.

Wu X, Johansen JV, Helin K. 2013b. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Molecular cell* **49**: 1134-1146.

Wutz A. 2011. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nature reviews Genetics* **12**: 542-553.

Wutz A, Rasmussen TP, Jaenisch R. 2002. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature genetics* **30**: 167-174.

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P et al. 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**: 86-90.

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

Xu C, Bian C, Lam R, Dong A, Min J. 2011a. The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. *Nature communications* **2**: 227.

Xu M, Long C, Chen X, Huang C, Chen S, Zhu B. 2010. Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. *Science* **328**: 94-98.

Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, Barbera AJ, Zheng L, Zhang H, Huang S et al. 2011b. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Molecular cell* **42**: 451-464.

Yamagishi T, Hirose S, Kondo T. 2008. Secondary DNA structure formation for Hoxb9 promoter and identification of its specific binding protein. *Nucleic acids research* **36**: 1965-1975.

Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM. 2010. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Molecular cell* **38**: 662-674.

Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, Weng Z, Rando OJ, Fazzio TG. 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* **147**: 1498-1510.

Yoh SM, Lucas JS, Jones KA. 2008. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes & development* **22**: 3422-3434.

Yoon HG, Chan DW, Reynolds AB, Qin J, Wong J. 2003. N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Molecular cell* **12**: 723-734.

Yu BD, Hess JL, Horning SE, Brown GA, Korsmeyer SJ. 1995. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**: 505-508.

Yu H, Mashtalir N, Daou S, Hammond-Martel I, Ross J, Sui G, Hart GW, Rauscher FJ, 3rd, Drobetsky E, Milot E et al. 2010. The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Molecular and cellular biology* **30**: 5071-5085.

Yu M, Mazor T, Huang H, Huang HT, Kathrein KL, Woo AJ, Chouinard CR, Labadorf A, Akie TE, Moran TB et al. 2012. Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Molecular cell* **45**: 330-343.

Yuan W, Wu T, Fu H, Dai C, Wu H, Liu N, Li X, Xu M, Zhang Z, Niu T et al. 2012. Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* **337**: 971-975.

Yuan W, Xu M, Huang C, Liu N, Chen S, Zhu B. 2011. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *The Journal of biological chemistry* **286**: 7983-7989.

Yun M, Wu J, Workman JL, Li B. 2011. Readers of histone modifications. *Cell research* **21**: 564-578.

Zhang S, Roche K, Nasheuer HP, Lowndes NF. 2011. Modification of histones by sugar beta-N-acetylglucosamine (GlcNAc) occurs on multiple residues, including histone H3 serine 10, and is cell cycle-regulated. *The Journal of biological chemistry* **286**: 37483-37495.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**: R137.

Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes & development* **13**: 1924-1935.

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008a. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**: 750-756.

Zhao Y, Lang G, Ito S, Bonnet J, Metzger E, Sawatsubashi S, Suzuki E, Le Guezennec X, Stunnenberg HG, Krasnov A et al. 2008b. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Molecular cell* **29**: 92-101.

Zhou JC, Blackledge NP, Farcas AM, Klose RJ. 2012. Recognition of CpG island chromatin by KDM2A requires direct and specific interaction with linker DNA. *Molecular and cellular biology* **32**: 479-489.

Zhou W, Wang X, Rosenfeld MG. 2009. Histone H2A ubiquitination in transcriptional regulation and DNA damage repair. *The international journal of biochemistry & cell biology* **41**: 12-15.

Zhou W, Wei W, Sun Y. 2013. Genetically engineered mouse models for functional studies of SKP1-CUL1-F-box-protein (SCF) E3 ubiquitin ligases. *Cell research* **23**: 599-619.

Zhou W, Zhu P, Wang J, Pascual G, Ohgi KA, Lozach J, Glass CK, Rosenfeld MG. 2008. Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Molecular cell* **29**: 69-80.

Zhou YB, Gerchman SE, Ramakrishnan V, Travers A, Muyldermans S. 1998. Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature* **395**: 402-405.

Zhu J, He F, Hu S, Yu J. 2008. On the nature of human housekeeping genes. *Trends in genetics : TIG* **24**: 481-484.

Zhu P, Zhou W, Wang J, Puc J, Ohgi KA, Erdjument-Bromage H, Tempst P, Glass CK, Rosenfeld MG. 2007. A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. *Molecular cell* **27**: 609-621.

Zimmermann M, Lottersberger F, Buonomo SB, Sfeir A, de Lange T. 2013. 53BP1 regulates DSB repair using Rif1 to control 5' end resection. *Science* **339**: 700-704.

Zink B, Paro R. 1989. In vivo binding pattern of a trans-regulator of homoeotic genes in Drosophila melanogaster. *Nature* **337**: 468-471.