KDM2A and KDM2B mediate widespread transcription repression in mouse embryonic stem cells

Anne Helen Turberfield

Merton College, University of Oxford

A thesis submitted in partial fulfilment of the
requirements for the degree of Doctor of Philosophy

Trinity Term, 2018

Abstract

KDM2A and KDM2B mediate widespread transcription repression in mouse embryonic stem cells

Anne Helen Turberfield, Merton College, Trinity Term 2018

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy of the University of Oxford

DNA methylation is widespread in vertebrate genomes, occurring on cytosine residues in the context of CpG dinucleotides. An important exception to this pervasive methylation are CpG islands (CGIs), short regions of CpG rich, non-methylated DNA that are associated with the majority of transcription start sites. CGIs are bound by proteins containing a zinc finger-CXXC (ZF-CXXC) domain, which specifically recognises non-methylated CpG dinucleotides. These proteins are found in a variety of chromatin modifying complexes, leading to the generation of a distinctive chromatin environment at CGI gene promoters. KDM2A and KDM2B (KDM2A/B) are recruited to CGI throughout the genome via their ZF-CXXC domain, and act as histone H3 lysine 36 (H3K36) demethylases. H3K36 methylation can inhibit transcription initiation, leading to the proposal that KDM2A and KDM2B may contribute to the generation of a transcriptionally permissive chromatin state at CGI-associated gene promoters through local removal of this repressive mark. However, KDM2A/B have been implicated in transcription repression in other contexts. Furthermore, both KDM2A and KDM2B also appear to possess demethylase-independent modes of transcription regulation, including via the short isoforms of these proteins which lack demethylase activity. Therefore, the genome-wide role played by KDM2A/B in regulating gene expression remains unclear.

Here I dissect the contribution of KDM2A/B to transcription regulation. I discover that the histone demethylase activity of KDM2A/B contributes little to the regulation of gene expression in mouse embryonic stem cells. In contrast, I find that loss of KDM2A/B from chromatin leads to a widespread increase in transcription, and show that KDM2B plays the predominant role in transcription repression. I demonstrate that the transcription upregulation following loss of KDM2A/B is not accompanied by changes in the accessibility of CGI gene promoters. Instead, I identify candidate substrates of the putative E3 ubiquitin ligase activity of KDM2A/B. Altered activity or stability of these substrates might result in transcription deregulation. Together, these observations reveal an unexpected and widespread role for KDM2B in transcriptional repression.

Acknowledgements

I thank Prof Rob Klose for his constant encouragement, guidance, enthusiasm and kindness. I have learned so much from being a member of your lab, and feel incredibly lucky to have had such a wonderful mentor.

Thank you to Emilia, Neil and Angelika for your guidance throughout my DPhil, for teaching me so many new techniques and for always being there to help me. I am very grateful to everyone in the Klose and Brockdorff labs, past and present, for their advice and for making the lab such a friendly environment. The work in this thesis would not have been possible without your support. Special thanks to Hamish and Nadya for teaching me so much about bioinformatics, to Deniz for your help with my mass spectrometry experiments, and to Tatyana and Emma for your tireless work to ensure that everything runs smoothly in the lab. I am very grateful to Prof Neil Brockdorff and Prof Jane Mellor for their advice during my thesis committee meetings.

Thank you to Prof Haruhiko Koseki and members of his lab for providing cell lines that were fundamental to this study. I would also like to thank the Wellcome Trust for funding my studentship, for their generosity in the face of health difficulties, and for offering me the opportunity to spend three months learning about science policy at the Academy of Medical Sciences.

I am hugely grateful to my close friends for their unwavering support. Thank you Hannah for being an amazing friend, and for your last minute proof-reading. Thank you to my sister Catherine for your love, and for travelling to Oxford to cheer me up this summer. Lorenzo, you mean the world to me, and I am so excited to plan our wedding for next year.

I would like to dedicate this thesis to my parents and grandparents. Your guidance and support have made this all possible.

Declaration of authorship

I declare that all work presented in this thesis is my own, unless otherwise acknowledged.

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.

Anne Helen Turberfield

4th October 2018

Contents

Abstract i

Acknowledgements ii

Declaration of authorship iii

Contents iv

List of figures viii

List of tables ix

List of common abbreviations x

1 Introduction 1
1.1 Chromatin regulates transcription in eukaryotes . . . 1
1.1.1 Eukaryotic DNA is packaged into chromatin . . . 1
1.1.2 Post-translational modification of histones . . . 2
1.1.3 Histone methylation . . . 4
1.1.4 H3K36 methyltransferases . . . 6
1.1.5 H3K36 demethylases . . . 10
1.1.6 Distribution of H3K36 methylation . . . 12
1.1.7 Regulation of gene expression by H3K36 methylation . . . 13
1.1.8 Antagonism between H3K36 methylation and H3K27 methylation . 16
1.1.9 H3K36 methylation regulates splicing and DNA methylation . . . 17
1.1.10 H3K36 methylation in development and disease . . . 18
1.2 CGI promoters as a platform for transcription regulation . . . 20
1.2.1 DNA methylation in vertebrate genomes . . . 20
1.2.2 Methyl-binding proteins recognise DNA methylation . . . 21
1.2.3 CGI are associated with the majority of vertebrate gene promoters . 22
1.2.4 ZF-CXXC proteins are recruited to CGI gene promoters . . . 24
1.3 The lysine demethylase 2 (KDM2) protein family . . . 26
1.3.1 KDM2A and KDM2B are H3K36me2 demethylases . . . 26
1.3.2 KDM2A/B have long and short isoforms . . . 29
1.3.3 Regulation of gene expression by the KDM2 family . . . 31
1.3.4 KDM2B targets PRC1.1 to CGI gene promoters . . . 35
1.3.5 Putative E3 ligase activity of KDM2 proteins . . . 38
1.3.6 Role of KDM2 proteins in development and disease . . . 40
1.4 Aims of this thesis . . . 42

2 Materials and Methods 44
2.1 Cell lines and culture . . . 44
2.1.1 Cell culture conditions . . . 44
2.1.2 Embryonic stem cell lines . . . 44
2.1.3 CRISPR-mediated genome editing . . . 45
2.1.4 Cell counting and cell size measurement . . . 45
2.2 Protein methods . . . 45
2.2.1 Acid extraction of histones . . . 45
2.2.2 Preparation of nuclear extracts . . . 46
2.2.3 Western blotting . . . 46

2.2.4 Antibodies . . . 47
2.2.5 Ubiquitylome mass spectrometry . . . 48
2.3 Chromatin immunoprecipitation . . . 50
2.3.1 Preparation of chromatin . . . 50
2.3.2 Immunoprecipitation . . . 51
2.4 RNA methods . . . 52
2.4.1 RNA extraction, reverse transcription and quantitative PCR . . . 52
2.4.2 Digital droplet PCR . . . 52
2.4.3 Isolation of nuclei . . . 53
2.4.4 Preparation of nuclear RNA for sequencing . . . 54
2.5 Chromatin accessibility . . . 54
2.5.1 Assay for transposase accessible chromatin (ATAC) . . . 54
2.5.2 ATAC-seq library preparation . . . 55
2.6 Massively parallel sequencing and data processing . . . 55
2.6.1 Massively parallel sequencing . . . 55
2.6.2 Read alignment . . . 55
2.6.3 Peak calling . . . 56
2.7 Normalisation, visualisation and analysis of sequencing data . . . 56
2.7.1 Calibration and differential analysis . . . 56
2.7.2 Histone ChIP-seq normalisation . . . 56
2.7.3 Data visualisation . . . 57
2.7.4 Functional annotation . . . 57
2.7.5 Datasets used in this study . . . 57
2.7.6 Correlation and significance testing . . . 58
2.7.7 Gene ontology analysis . . . 58

3 Examining the role of KDM2A/B histone demethylase activity 59
3.1 Introduction . . . 59
3.2 Results . . . 61
3.2.1 Conditional knockout of KDM2A/B-LF . . . 61
3.2.2 Regulation of bulk histone modification levels by KDM2A/B-LF . . . 62
3.2.3 Optimisation of H3K36me2 ChIP . . . 64
3.2.4 KDM2A/B knockout leads to a small increase in H3K36me2 at CGI gene promoters . . . 66
3.2.5 The gene bodies of highly transcribed genes are depleted of H3K36me2 . . . 68
3.2.6 Loss of KDM2A/B-LF results in minimal gene expression changes . . . 70
3.2.7 Loss of KDM2A/B-LF results in minimal changes to chromatin accessibility . . . 75
3.3 Summary and discussion . . . 77

4 KDM2A/B mediate widespread transcription repression 83
4.1 Introduction . . . 83
4.2 Results . . . 85
4.2.1 Identification of novel KDM2B isoforms . . . 85
4.2.2 Conditional deletion of the ZF-CXXC domain of KDM2A/B leads to their removal from chromatin . . . 88
4.2.3 KDM2A/B are required to sustain normal levels of transcription in mESCs . . . 90
4.2.4 Development of a calibrated nucRNA-seq protocol . . . 93

4.2.5 Removal of KDM2A/B leads to widespread transcription upregulation . . . 96
4.2.6 Transcription upregulation is not limited to polycomb target genes 102
4.2.7 Genes with low expression are more strongly upregulated following KDM2A/B removal . . . 107
4.3 Summary and discussion . . . 110

5 KDM2B plays the predominant role in transcription repression 115
5.1 Introduction . . . 115
5.2 Results . . . 116
5.2.1 Conditional removal of KDM2A from chromatin . . . 116
5.2.2 KDM2A chromatin occupancy is not required to maintain normal transcription . . . 117
5.2.3 Validation of the *Kdm2b-CXXC*$^{fl/fl}$ line . . . 118
5.2.4 Removal of KDM2B results in widespread transcription upregulation 120
5.2.5 KDM2A and KDM2B act as functionally redundant transcriptional repressors of a subset of genes . . . 126
5.2.6 Conditional deletion of FBXL19, KDM2A and KDM2B ZF-CXXC domains . . . 130
5.2.7 Removal of FBXL19 in addition to KDM2A/B largely recapitulates the transcription changes following removal of KDM2A/B alone . . . 131
5.2.8 A minority of genes are differentially regulated following removal of FBXL19 in addition to KDM2A/B . . . 133
5.3 Summary and discussion . . . 136

6 Investigating putative transcription repressive activities of KDM2A/B 142
6.1 Introduction . . . 142
6.2 Results . . . 144
6.2.1 Removal of KDM2A/B leads to minimal changes in chromatin accessibility . . . 144
6.2.2 SKP1 knockout leads to a dramatic accumulation of nucRNA . . . 144
6.2.3 Ubiquitylome-mass spectrometry identifies potential targets of KDM2A/B E3 ligase activity . . . 148
6.2.4 BRG1 TSS enrichment does not correlate with the gene expression changes following removal of KDM2A/B . . . 152
6.3 Summary and discussion . . . 153

7 Conclusions and future directions 161

A Appendices 167

Bibliography 177

List of Figures

1.1 Mammalian H3K36 methyltransferases . . . 7
1.2 The KDM2 family share a similar domain architecture . . . 28
1.3 KDM2A/B are proposed to contribute to the generation of a transcriptionally permissive chromatin state at CGI promoters . . . 31

3.1 Cre/loxP system for conditional knockout of KDM2A/B-LF . . . 63
3.2 Efficient KDM2A/B-LF knockout in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs . . . 64
3.3 KDM2A/B-LF knockout has a minor effect on global H3K36 methylation levels . . . 65
3.4 KDM2A/B-LF knockout has a minor effect on the levels of other CGI-associated histone modifications . . . 65
3.5 Optimisation of H3K36me2 ChIP . . . 67
3.6 KDM2A/B knockout leads to a modest increase in H3K36me2 at TSS . . . 69
3.7 H3K36me2 is depleted in the gene body of highly transcribed genes . . . 71
3.8 KDM2A/B-LF knockout leads to transcriptional changes of a small subset of genes . . . 72
3.9 A subset of polycomb target genes are deregulated following KDM2A/B-LF knockout . . . 74
3.10 Genes that significantly increase in expression are involved in developmental processes . . . 75
3.11 KDM2A/B-LF knockout leads to minimal changes in chromatin accessibility 76

4.1 Identification of novel *Kdm2b* isoforms . . . 87
4.2 Conditional deletion of the ZF-CXXC domain of KDM2A/B . . . 91
4.3 ZF-CXXC domain deletion leads to removal of KDM2A/B from chromatin 92
4.4 Removal of KDM2A/B leads to widespread deregulation of transcription . 93
4.5 Calibrated RNA-seq enables detection of a widespread reduction in transcription . . . 94
4.6 Calibrated RNA-seq reveals a global decrease in transcription following triptolide treatment . . . 95
4.7 nucRNA-seq calibration pipeline . . . 97
4.8 Removal of KDM2A/B leads to widespread transcription upregulation . . 98
4.9 Digital droplet PCR confirms the widespread increase in transcription . . . 101
4.10 Genes that significantly increase in expression are enriched for polycomb targets . . . 103
4.11 Widespread transcription upregulation is not explained by loss of polycomb-mediated gene repression . . . 106
4.12 Significantly increasing non-polycomb target genes are involved in a range of basic cellular processes . . . 108
4.13 Negative correlation between gene expression level and extent of transcription upregulation . . . 110

5.1 Conditional deletion of the ZF-CXXC domain of KDM2A . . . 117
5.2 Removal of KDM2A leads to minimal perturbation of transcription . . . 118
5.3 Conditional deletion of the ZF-CXXC domain of KDM2B . . . 119
5.4 Removal of KDM2B leads to widespread transcription upregulation . . . 120
5.5 Significantly upregulated genes are enriched for polycomb targets . . . 122
5.6 Transcription changes following removal of KDM2B correlate moderately with PCGF1 knockout and strongly with KDM2A/B knockout . . . 124

5.7 Calibration to *Drosophila* spike-in is required to reveal a widespread increase in transcription following KDM2B removal . . . 125
5.8 Significantly increasing non-polycomb target genes are involved in a range of basic cellular processes . . . 126
5.9 KDM2A and KDM2B act as functionally redundant transcriptional repressors of a subset of genes . . . 129
5.10 KDM2A is required to maintain normal transcription levels of a subset of genes . . . 130
5.11 Conditional deletion of the ZF-CXXC domains of KDM2A, KDM2B and FBXL19 . . . 131
5.12 Removal of FBXl19 in addition to KDM2A/B recapitulates the widespread increase in transcription . . . 132
5.13 Genes that increase in expression following removal of FBXL19 and KDM2A/B are involved in both developmental and basic cellular processes 134
5.14 Negative correlation between gene expression level and log2 fold change in transcription . . . 135
5.15 A minority of genes are differentially regulated following removal of FBXL19 in addition to KDM2A/B . . . 137

6.1 Removal of KDM2A/B leads to minimal changes in chromatin accessibility 145
6.2 SKP1 knockout leads to a dramatic increase in nucRNA levels . . . 147
6.3 Ubiquitylome mass spectrometry identifies potential KDM2A/B substrates 149
6.4 LC-MS/MS of peptides containing a ubiquitin remnant motif identifies potential KDM2A/B substrates . . . 150
6.5 Poor correlation between BRG1 occupancy and transcription changes following removal of KDM2A/B . . . 154

A.1 Tamoxifen treatment time-course for *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs . . . 167
A.2 A constitutive CRISPR knockout study confirms the expression of novel KDM2B isoform B in mESCs . . . 168
A.3 Removal of KDM2A/B leads to widespread upregulation of nascent transcripts . . . 169
A.4 Minimal enrichment of KDM2A, KDM2B and BioCAP signal at non-CGI TSS . . . 170
A.5 Conditional knockout of PCGF1 in *Pcgf1*$^{fl/fl}$ mESCs . . . 170
A.6 Conditional knockout of RING1B in *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ mESCs . . . 170
A.7 Tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs does not perturb RING1B or SUZ12 expression . . . 171
A.8 Non-CGI genes that significantly decrease in expression following KDM2A/B removal are relatively strongly downregulated . . . 171
A.9 Gene expression changes following removal of FBXL19 and KDM2A/B parallel those following loss of KDM2A/B . . . 172
A.10 KDM2A/B knockout is not accompanied by an increase in CGI methylation 174
A.11 Tamoxifen treatment of *Skp1*$^{fl/fl}$ mESCs leads to SKP1 knockout . . . 175
A.12 Tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs does not lead to a significant change in cell size . . . 175
A.13 Details of peptides that significantly decrease in ubiquitylation following removal of KDM2A/B . . . 176

List of Tables

2.1 Genotypes of conditional mESC lines 44
2.2 Antibodies used in this study 48
2.3 Published datasets used in this study 58

List of common abbreviations

ac	acetylation
ATAC	assay for transposase-accessible chromatin
ATAC-seq	ATAC followed by massively parallel sequencing
BioCAP	biotinylated CxxC affinity purification
CAGE	cap analysis of gene expression
CFP1	CXXC finger protein 1
CGI	CpG island
ChIP	chromatin immunoprecipitation
ChIP-seq	ChIP followed by massively parallel sequencing
Cre-ERT2	cre recombinase fused to the oestrogen receptor
ddPCR	digital droplet PCR
DNase	deoxyribonuclease
DNMT	DNA methyltransferase
DSB	double strand break
dsDNA	double-stranded DNA
E	embryonic day
FBXL	F-box and leucine-rich repeat
FCS	foetal calf serum
gDNA	genomic DNA
H3K36	histone H3 lysine 36
HDAC	histone deacetylase
HDR	homology dependent repair
HP1	heterochromatin protein 1
JmjC	jumonji C
KDM	lysine demethylase
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LRR	leucine-rich repeats
MBD	methyl binding domain
me1	monomethylation
me2	dimethylation
me3	trimethylation
mESC	mouse embryonic stem cell
MNase	micrococcal nuclease
MSC	mesenchymal stem cells
MSL	male-specific lethal
nucRNA	nuclear RNA
nucRNA-seq	nucRNA extraction followed by massively parallel sequencing
NuRD	nucleosome remodelling deacetylase
PCGF	polycomb group ring finger
PHD	plant homeodomain
PIC	protease inhibitor cocktail
PRC	polycomb repressive complex
PTM	post-translational modification
qPCR	quantitative real-time PCR
RNAPII	RNA polymerase II
RT-qPCR	reverse transcription followed by qPCR
SCF	SKP1-CUL1-Fbox
SDS-PAGE	SDS-polyacrylamide gel electrophoresis

SET	su(var)3-9-enhancer of zeste-trithorax
sgRNA	single guide RNA
SUMO	small ubiquitin-like modifier
TSS	transcription start site
ZF-CXXC	zinc finger-CXXC

1

Introduction

1.1 Chromatin regulates transcription in eukaryotes

Multicellular organisms are made up of a huge array of functionally diverse cell types. These contain the same genetic material, their DNA, which is wrapped around basic histone proteins to form chromatin (Luger et al. 1997; Arents et al. 1991). Given that individual cells contain the genetic blueprint required to make an entire organism, multicellular development relies on the establishment of cell-type-specific gene expression programmes. This requires the exquisite control of RNA polymerase II (RNAPII)-mediated transcription by a complex network of cellular pathways. Chromatin structure and post-translational modification (PTM) of histone proteins play an important role in shaping transcription. The dynamic regulation of chromatin is therefore crucial to the establishment and maintenance of distinct cell types during development.

1.1.1 Eukaryotic DNA is packaged into chromatin

Nucleosomes are the fundamental unit of chromatin. Each nucleosome comprises a symmetrical histone octamer, formed from two molecules each of histones H2A, H2B, H3 and H4, around which approximately 146 base pairs of DNA are wound. This nucleosomal packaging is a key mechanism by which genomic DNA is compacted and organised within the nucleus. Histone H1 binds linker DNA that bridges nucleosomes, and is thought to contribute to further compaction (Sato et al. 1999). Chromatin structure not only regulates DNA compaction, but also has profound effects on almost all processes that use DNA as a template, including transcription, replication, repair and recombination (Li et al. 2007). Importantly, packaging of genomic DNA into nucleosomes is inhibitory to transcription initiation (Knezetic and Luse 1986; Lorch et al.

1987), since nucleosomes act as a barrier to the binding of sequence-specific transcription factors and the basal transcription machinery (Cockerill 2011). The establishment of an accessible, nucleosome-depleted chromatin state at transcription start sites (TSS) is therefore required for productive transcription (Li et al. 2007; Struhl 1999).

1.1.2 Post-translational modification of histones

Through studying the role played by chromatin in regulating fundamental cellular processes such as transcription and DNA replication, it became clear that many of its effects are mediated through PTMs of histone proteins. The core histone proteins possess unstructured 20-35 residue N-terminal tails which protrude from the globular histone core of the nucleosome. These tails are subject to a variety of PTMs mediated by residue-specific histone modifying enzymes including, but not limited to, acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ribosylation, biotinylation and glycosylation (Kouzarides 2007; Shiio and Eisenman 2003; Zhang et al. 2011; Camporeale et al. 2004). In addition, recent studies have demonstrated that the structured fold domains of the histones can also be extensively modified (Tropberger and Schneider 2010). The combinatorial modification of different histone residues with different PTMs leads to a huge variation of possible PTM combinations. This led to the concept of a 'histone code', whereby a particular permutation of PTMs elicits a specific functional outcome (Turner 1993; Strahl and Allis 2000).

Histone acetylation can directly affect the physical properties of chromatin by disrupting DNA-histone contacts (Norton et al. 1989; Garcia-Ramirez et al. 1995; Anderson et al. 2001; Shogren-Knaak et al. 2006). This led to a simple charge-neutralisation model, proposing that histone acetylation contributes to transcription activation via a localised increase in chromatin accessibility (Wade et al. 1997). It is now clear that the predominant role of histone PTMs is to recruit 'reader' proteins, which recognize defined PTM states via specialised domains and possess chromatin-modifying or transcriptional co-activator activities. Thus histone tails provide a binding surface for the recruitment of effector proteins, including DNA/histone modifying enzymes, chromatin remodellers and histone chaperones, which coordinate fundamental processes

including transcription, replication, repair and compaction (Allis and Jenuwein 2016). Misregulation of histone PTMs is linked to cancer and developmental defects, underlining their important contribution to normal cellular function (Chi et al. 2010; Audia and Campbell 2016).

The recruitment of effector proteins to exert specific regulatory functions requires the localisation of particular histone modifications to, or their absence from, defined regions of the genome (ENCODE 2012). For example, promoters are characterised by high levels of histone H3 lysine 4 tri-methylation (H3K4me3) and low levels of H3K4me1, while distal regulatory elements are instead enriched for H3K4me1 and have low levels of H3K4me3 (Barski et al. 2007; Guenther et al. 2007). H3K36me3 is a hallmark of actively transcribed gene bodies (Krogan et al. 2003; Bannister et al. 2005) and histone acetylation is enriched at active promoters and distal regulatory elements (Hebbes et al. 1988; Pokholok et al. 2005; Karmodiya et al. 2012). Histone hypoacetylation is linked with transcription repression (Turner 2000) and H3K9me3 is enriched at inactive regions of heterochromatin (Nakayama et al. 2001; Lehnertz et al. 2003).

The enrichment of a variety of PTMs at a particular genomic location can coordinate the recruitment of a range of modifying activities to establish a specific chromatin environment. Modifications on adjacent residues can also positively or negatively regulate the binding of a given 'reader' protein, adding a further layer of complexity. However, the downstream molecular pathways associated with a particular set of histone PTMs depends on the genomic and cellular context in which they occur. Indeed, a specific PTM can be recognised by different 'reader' proteins. A particular PTM or set of PTMs can therefore lead to different or even opposing consequences depending on its context. This will be discussed in more detail with reference to histone lysine methylation in Section 1.1.3 below. Furthermore, it is unclear whether histone PTMs are truly instructive to transcription or instead are a reflection of transcription status in different contexts. PTMs therefore appear not to be interpreted as a rigid 'code', but instead contribute to fine-tuning DNA-dependent processes according to the cellular context (Peterson and Laniel 2004). A large body of research focused on enzymes that catalyse the removal of specific histone PTMs has revealed that most, if not all, of these PTMs are reversible. This allows the dynamic regulation of the chromatin environment

and hence fundamental DNA-dependent processes in response to internal molecular networks or external signals (Voss and Hager 2014).

1.1.3 Histone methylation

Histone methylation, and its potential involvement in transcription regulation, was first reported in the 1960s as a result of radiolabelling studies of cell extracts (Allfrey et al. 1964; Allfrey and Mirsky 1964; Murray 1964). However, it wasn't until 2000 that the first histone methyltransferase, SUV39H1, was identified (Rea et al. 2000). SUV39H1 catalyses H3K9 methylation via a conserved Su(var)3-9-Enhancer of Zeste-Trithorax (SET) domain, using S-adenosylmethioinine as a cofactor. The SET domain was first identified as a conserved region shared by the *Drosophila melanogaster* orthologue of this protein, Su(var)3-9, and two other *Drosophila* proteins that were also subsequently shown to act as histone lysine methyltransferases: E(z) (EZH2 in mammals) is a H3K27 methyltransferase (Czermin et al. 2002; Muller et al. 2002; Cao et al. 2002; Kuzmichev et al. 2002) while Trx is a H3K4 methyltransferase (Smith et al. 2004). The identification of Suv39H1 histone methyltransferase activity was rapidly followed by the identification of numerous other SET domain-containing histone lysine methyltransferases (Dillon et al. 2005). In contrast, the DOT1 family, which methylate K79 within the globular fold of histone H3, do not share this conserved SET domain (Feng et al. 2002), and a family of protein arginine methyltransferases instead catalyse the methylation of histone arginine residues. These also use SAM as a cofactor, and are classified according to whether they catalyse symmetric or asymmetric arginine dimethylation (Zhang and Reinberg 2001).

Methylation occurs on numerous histone residues. Furthermore, lysine methylation can occur in the mono- (me1), di- (me2) and tri- (me3) methylated state, while arginine methylation occurs in a number of symmetric and asymmetric mono- and di-methylated states (Kouzarides 2007). This variation in the methylation state of a given residue contributes to the complexity of potential histone PTM combinations. Histone methylation has a relatively small effect on electrostatic interactions as, unlike acetylation, it preserves the positive charge of the ε-amino nitrogen. Instead it is thought to positively or negatively regulate the recruitment of specific effector proteins, which

act to modify the local chromatin structure. Indeed, a wide variety of domains that specifically bind to methylated lysine residues have been identified, including plant homeodomain (PHD), WD40, chromo, Tudor, PWWP, MBT, CW and ankyrin repeat domains (Greer and Shi 2012).

Interestingly, the same modified histone lysine can be bound by different effector proteins. For example, recognition of H3K4me3 by the PHD domain of ING2 was shown to recruit the repressive mSin3a-Histone Deacetylase (HDAC) complex to the promoters of proliferation genes in response to DNA damage (Shi et al. 2006). H3K4me3 is also bound by the tandom chromodomains of the human ATP-dependent remodeller CHD1 (Sims et al. 2005), and by the PHD domain of BTPF (Li et al. 2006), a subunit of the ATP-dependent nucleosome remodelling factor (NURF) complex that is required for the induction of steroid responsive genes in *Drosophila* and correct Hox gene expression patterns in *Xenopus* embryos (Wysocka et al. 2006; Badenhorst et al. 2005). Furthermore, H3K4me3 is recognised by the tandem Tudor domains of the H3K9 and H3K36 demethylase KDM4A (Huang et al. 2006; Kim et al. 2006). H3K9 and H3K36 marks are implicated in transcription repression, therefore recruitment of KDM4A demethylase activity to gene promoters, which are enriched for the H3K4me3 mark to which KDM4A binds (Barski et al. 2007; Guenther et al. 2007), may promote transcription. H3K4me3 can also more directly influence transcription via its recognition by the PHD domain of the TAF3 subunit of general transcription factor TFIID (Vermeulen et al. 2007; Lauberth et al. 2013). Thus H3K4me3 is recognized by a variety of proteins which recruit diverse effector functions to chromatin, including histone deacetylase, chromatin remodelling and histone demethylase activities and the basal transcription machinery, which can act to promote or constrain transcription. This exemplifies the paradigm that a single histone modification can lead to a variety of downstream outcomes, depending on the cellular context in which it is interpreted.

Initially histone methylation was widely considered an irreversible mark that could only be removed via gradual dilution during DNA replication or by histone eviction, despite early biochemical studies that suggested that enzymatic histone demethylase activities exist in cells (Paik and Kim 1973; Paik and Kim 1974). This was refuted by the discovery of the histone demethylase KDM1A/LSD1, which demethylates H3K4

via an amine oxidase domain, using FAD as a cofactor (Shi et al. 2004). Soon after, KDM2A/JHDM1A/FBXL11 was shown to demethylate H3K36 via a Jumonj C (JmjC) domain, coordinating iron to mediate a 2-oxoglutarate-dependent demethylation reaction (Tsukada et al. 2006). These initial discoveries led to the identification of extended LSD and JmjC protein families, leading to our current view of histone methylation as a highly regulated and dynamic modification (Kooistra and Helin 2012). JMJD6 may function as an arginine-specific histone demethylase (Chang et al. 2007; Liu et al. 2013), however the exact nature of this reaction and its biological function are unclear (Mantri et al. 2010; Bottger et al. 2015; Liu et al. 2013). Instead, a recent study found that a number of JmjC histone lysine demethylases (KDMs) can also catalyse histone arginine demethylation *in vitro*, suggesting that these enzymes play a dual role (Walport et al. 2016). The importance of the robust regulation of specific histone methylation states is highlighted by the critical roles played by many histone methyltransferases and demethylases during development and in diseases, including in neurological disorders and many types of cancer (Nottke et al. 2009; Albert and Helin 2010; Pedersen and Helin 2010).

Here I will limit a more detailed discussion to the regulation of H3K36 methylation and its role in transcription regulation, which is central to this study. The H3K36 methylation system exemplifies a number of important paradigms that apply to the regulation of histone PTMs more broadly, which I will highlight throughout the discussion.

1.1.4 H3K36 methyltransferases

Set2 is the sole H3K36 methyltransferase in yeast, responsible for catalysing all H3K36 mono-, di- and tri-methylation (Strahl et al. 2002). In contrast, at least eight different H3K36 methyltransferases have been identified in mammals, all of which possess a catalytic SET domain (Figure 1.1). There is a division of labour between mono- and di-methyltransferases, including NSD1 and NSD2, and trimethylation by Set2 orthologue SETD2/HYPB/KMT3A (Bell et al. 2007; Edmunds et al. 2008). SET2 proteins from a range of species interact with the hyperphosphorylated C-terminal tail of RNA Polymerase II (RNAPII) (Schaft et al. 2003; Krogan et al. 2003; Xiao et al. 2003; Li

et al. 2003; Sun et al. 2005; Yoh et al. 2008). This requires the Set2 Rpb1 interacting (SRI) domain of SET2, which mediates its interaction with RPB1, the largest RNAPII subunit (Kizer et al. 2005). SET2 therefore migrates with elongating RNAPII through the body of actively transcribed genes (Li et al. 2002; Schaft et al. 2003). This leads to the characteristic enrichment of H3K36me3 throughout the transcribed region, peaking towards the 3' end, and a positive correlation between levels of H3K36me3 and transcription (Krogan et al. 2003; Bannister et al. 2005; Pokholok et al. 2005; Mikkelsen et al. 2007; Bell et al. 2007).

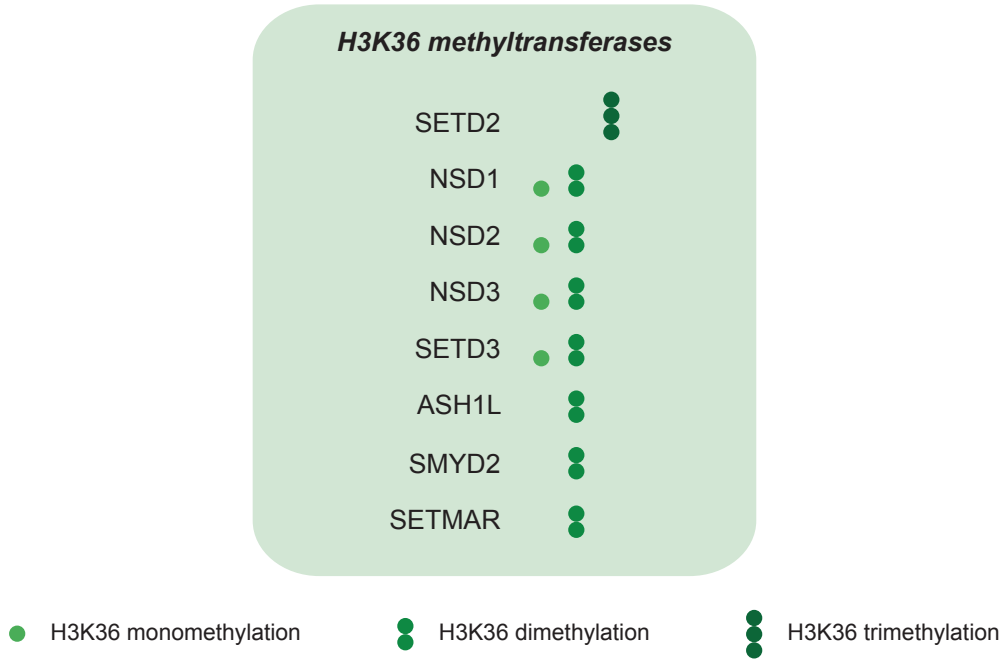


Figure 1.1: Mammalian H3K36 methyltransferases
Schematic of the eight H3K36 methyltransferases that have been identified in mammals. Their methylation state specificities are depicted, based on *in vivo* data where available.

SETD2 is thought to mediate virtually all H3K36 trimethylation in mammalian cells, as SETD2 knockdown leads to a dramatic loss of global H3K36me3 (Edmunds et al. 2008; Yuan et al. 2009; Hu et al. 2010; Ferrari et al. 2014). H3K36me1 and H3K36me2 levels are unaffected, in contrast to reports of SETD2 mono-, di- and tri-methyltransferase activity *in vitro*. Knockdown of the *Drosophila* SET2 homologue dHypb also leads to a decrease in H3K36me3 alone (Bell et al. 2007). These studies cast doubt on the ability of SETD2 to processively place all three methyl groups *in vivo*. SET2 catalytic activity appears to be modulated by a range of cofactors, including the histone chaperone Asf1 and the elongation factor Paf1 in yeast (Chu et al. 2007; Lin et al. 2010), and RNA

binding protein hnRNPL in mammalian cells (Yuan et al. 2009). SET2 activity may also be modulated by other histone modifications (Du et al. 2008; Nelson et al. 2006). Thus histone methyltransferase activity may be modulated at specific genomic regions via associated cofactors or the influence of the surrounding chromatin environment, enabling the fine-tuning of histone methylation patterns at different target regions in a context-dependent manner.

There are conflicting reports as to the residue targeted and the level of methylation (me1/me2/me3) catalysed by the other identified H3K36 methyltransferases: NSD1, NSD2, NSD3, SETMAR, SMYD2, ASH1L and SETD3. This is likely to be due to variation in the substrate (peptides, histones or nucleosomes), enzyme (SET domain or full length) and assay conditions (Western blotting, chromatin immunoprecipitation (ChIP) or mass spectrometry) used to assess histone residue specificity (Wagner and Carpenter 2012). Such discrepancies also plague research into histone methyltransferases that target other histone residues. Of the remaining H3K36 methyltransferases, NSD1 and NSD2 are best characterised.

NSD1 was shown to mono- and di-methylate H3K36, with conflicting reports as to whether it also has specificity for H4K20 (Rayasam et al. 2003; Berdasco et al. 2009; Lucio-Eterovic et al. 2010). This detection of alternative NSD1 activity may be explained by the use of histone octamers rather than nucleosomes to assess methyltransferase specificity (Li et al. 2009). The NSD1 crystal structure revealed an autoinhibitory loop, which appears to move to an active confirmation through allosteric activation by nucleosomal DNA (Qiao et al. 2011). Thus interaction with nucleosomes as opposed to histone octamers may impart specificity for H3K36 binding, as the H3K36 residue lies close to to the globular core of the nucleosome and therefore in proximity to nucleosomal DNA. Similarly, NSD2 was reported to not only dimethylate nucleosomal H3K36 (Kuo et al. 2011), but to methylate additional residues of recombinant histones including H3K27, H3K4 and H4K20 (Kim et al. 2006; Kim et al. 2008; Marango et al. 2008). However, addition of short lengths of DNA to histone octamers leads to preferential H3K36 dimethylation by NSD2, suggesting that DNA also acts as an allosteric effector of NSD2 to impart H3K36 specificity (Li et al. 2009). ASH1L also possesses specific H3K36 methyltransferase activity when incubated with nucleosomal substrates (Tanaka

et al. 2007; Yuan et al. 2011; Miyazaki et al. 2013), despite additional activities reported when using histone peptides. The structures of ASH1L and SETD2 also revealed an auto-inhibitory loop, suggesting that allosteric regulation of substrate specificity may be a conserved feature of H3K36 methyltransferases (An et al. 2011; Zheng et al. 2012; Yang et al. 2016).

NSD1-3 appear to localise to both promoter and gene body regions, where they catalyse H3K36 methylation (Kuo et al. 2011; Lucio-Eterovic et al. 2010; Rahman et al. 2011; Ram et al. 2011; Shen et al. 2015). The mechanisms that target the NSD family to chromatin remain incompletely characterised, however interactions with other proteins are suggested to play an important role. NSD1 was reported to interact with nuclear receptors (Huang et al. 1998) and localisation of NSD3 is regulated by its interaction with a number of proteins including BRD4 and LSD2, the latter of which appears to be recruited to elongating RNAPII (Rahman et al. 2011; Fang et al. 2010). Furthermore NSD1-3 possess a number of chromatin reader domains, including PWWP domains that recognise methylated H3K36. The PWWP domain of NSD2 preferentially binds H3K36me2, and this domain appears be required for stable binding of NSD2 to chromatin (Sankaran et al. 2016). This ability of NSD2 to bind the histone mark that it places was proposed to play a role in H3K36me2 spreading. The PHD2 domain of NSD2 may also contribute to its targeting (Huang et al. 2013). Overexpression of NSD2 leads to a broad increase in H3K36me2 (Kuo et al. 2011) while NSD1 knockdown leads to a broad decrease in H3K36me2 (Streubel et al. 2018), suggesting that both of these enzymes are targeted broadly to chromatin and play a widespread role in H3K36 dimethylation throughout the genome. Loss of function of the NSD1 orthologue in *Drosophila* or *C. elegans* leads to a global loss of H3K36me2, suggesting that NSD proteins play a conserved role in regulating bulk H3K36me2 levels (Bender et al. 2006; Bell et al. 2007).

ASH1L is reported to occupy the transcribed regions of active genes (Gregory et al. 2007), and to regulate Hox gene expression in both *Drosophila* and mammals (Miyazaki et al. 2013). The recruitment of *Drosophila* orthologue Ash1 is thought to be regulated through a variety of mechanisms, including via its chromatin binding domains and interaction with Mrg15, which recognises H3K36 methylation. While its recruitment to chromatin

is poorly understood in mammals, the conserved role in *Hox* gene regulation suggests that similar targeting mechanisms are at play.

1.1.5 H3K36 demethylases

The steady state of a given histone modification is governed by the balance between its addition and removal by specific enzymes. The localisation of histone demethylases to specific locations of the genome, and the modulation of this process in response to cellular cues, is therefore vital to the regulation of histone methylation levels. KDM2A/JHDM1A was the first JmjC-containing histone demethylase to be identified. Both KDM2A and its paralogue KDM2B/JHDM1B are H3K36 demethylases, with specificity for the mono- and di-methyl state (Tsukada et al. 2006; He et al. 2008). The crystal structure of KDM2A revealed that H3K36 specificity is mediated via threading of the H3 tail through a catalytic groove where it is positioned and oriented for H3K36 demethylation, with a number of interactions with specific residues contributing to specificity (Cheng et al. 2014). This study also provided insight into the specificity of KDM2A for the lower methylation states of H3K36, proposing that steric constraints following H3K36me3 binding could prevent α-ketoglutarate from undergoing a rotation within the active site that is necessary for the demethylation reaction. KDM2A and KDM2B both localise to regions of non-methylated DNA throughout the genome, called CpG islands (CGI), via their CXXC domains (Blackledge et al. 2010; Farcas et al. 2012; Wu et al. 2013; He et al. 2013). The roles played by KDM2A and KDM2B in transcription regulation will be discussed in detail in Section 1.3.

The KDM4/JHDM3 family were the first JmjC proteins reported to demethylate trimethylated substrates. There are four KDM4 family members in mammalian cells. While KDM4A, KDM4B and KDM4C are widely expressed, expression of KDM4D appears to be largely restricted to the testes (Iwamori et al. 2011). The KDM4 proteins catalyse demethylation of H3K36me2/3 and H3K9me2/3 (Cloos et al. 2006; Klose et al. 2006; Whetstine et al. 2006; Fodor et al. 2006). The crystal structure of KDM4A revealed that specificity for H3K36 and H3K9 is primarily determined by the local primary structure of the histone tail, while lack of demethylase activity towards monomethylated

substrates may be explained by the suboptimal positioning of the monomethyl group relative to the catalytic center (Chen et al. 2006). This demethylase selectivity for the degree of methylation leads to a greater control of lysine methylation patterns (Shi and Whetstine 2007). In addition to H3K36 and H3K9, KDM4 proteins were also reported to demethylate K26 of linker histone isoform H1.4, which is implicated in gene repression (Trojer et al. 2009).

The double Tudor domain of KDM4A can bind H3K4me3 and H3K20me2/3 *in vitro* (Mallette et al. 2012; Huang et al. 2006; Kim et al. 2006; Lee et al. 2008), while the double Tudor domain of KDM4B binds H3K20me2 and that of KDM4C binds H3K4me3 but not H3K20me3 (Mallette et al. 2012; Pedersen et al. 2014). KDM4A and KDM4C were detected at a number of gene promoters, where they were implicated in transcription activation, and were later shown to localise to H3K4me3-positive TSS by ChIP followed by sequencing (ChIP-seq) in several mammalian cell types (Pedersen et al. 2014; Pedersen et al. 2016). This suggests that the double Tudor domain is important for KDM4 localisation. Indeed, Tudor domain mutation or knockdown of the H3K4 methyltransferase subunit Dpy30 confirmed that recognition of H3K4me3 by the double Tudor domain of KDM4C is at least partially responsible for its chromatin binding pattern. Interestingly, KDM4B may also be linked to H3K4me3, through a reported interaction with the MML2 H3K4 methyltransferase subunit (Shi et al. 2011). However, 60% of KDM4B binding sites were reported to be distant from TSSs, suggesting that it may regulate histone methylation levels at different genomic regions (Das et al. 2014). Transcription factors may also contribute to KDM4 targeting, as reported for HIF1, E2F1 and ETS transcription factors (Luo et al. 2012; Wang et al. 2016; Kim et al. 2016; Li et al. 2018). In addition, KDM4 activity can also regulated by ubiquitylation and SUMOylation (Tan et al. 2011; Ipenberg et al. 2013; Yang et al. 2017).

The importance of the KDM4 double Tudor domain, and of the KDM2 CXXC domain, in protein localisation exemplifies the importance of chromatin- or DNA-binding domains in regulating demethylase targeting. Indeed, this may hold true even when the demethylase itself lacks a reader domain (Dimitrova et al. 2015). For example, the H3K36me2/3-specific demethylase NO66 appears to bind H3K36me3-enriched regions via its interaction with PHF19, whose Tudor domain binds H3K36me3

(Brien et al. 2012; Sinha et al. 2010). The recruitment of KDM4 H3K9/H3K36 demethylases to H3K4me3-enriched regions exemplifies the complex interplay between histone modifications. Simultaneous KDM4A and KDM4C knockout leads to increased H3K9me3 and H3K36me3 at TSS, with a correlation between increased H3K9me3 and transcription repression. This suggests that KDM4A and KDM4C continually act at TSS to prevent spurious accumulation of H3K9me3 and H3K36me3, which may otherwise perturb transcription (Pedersen et al. 2016).

Another JmjC protein, KDM8/JMJD5, was also reported to be a H3K36me2 demethylase involved in the repression of satellite repeat regions (Marcon et al. 2014). However, later studies failed to reproduce this demethylase activity and instead suggest that KDM8 instead acts as a protein hydroxylase (Del Rizzo et al. 2012; Youn et al. 2012; Hopkinson et al. 2013; Wilkins et al. 2018).

1.1.6 Distribution of H3K36 methylation

These H3K36 methyltransferases and demethylases form a regulatory network that dynamically controls H3K36 methylation levels in mammalian cells. The varied modes of regulation and targeting of these enzymes and their specificity for particular methylation states contribute to the complexity of this regulatory system. H3K36me3 is highly enriched in gene bodies, due to the evolutionarily conserved interaction of Set2 with elongating RNAPII. Early studies showed that H3K36me2 is also enriched over gene bodies in yeast (Bannister et al. 2005; Rao et al. 2005). This gene body enrichment and its initiation at a fixed distance from the TSS is consistent with H3K36me2 deposition by Set2 associated with elongating RNAPII (Rao et al. 2005; Pokholok et al. 2005). ChIP-chip analysis revealed that H3K36me2 is relatively depleted in promoter regions and transcriptionally silent chromatin, but is enriched in actively transcribed regions. However, unlike for H3K36me3, its enrichment doesn't correlate with transcription level. It is unclear how H3K36me2 and H3K36me3 deposition by Set2 are differentially regulated in yeast.

Whereas Set2 is the sole H3K36 methyltransferase in yeast, there is a division of labour between H3K36me1/2 and H3K36me3 methyltransferase activity in different

enzyme families in mammalian cells. Therefore, H3K36me2 deposition is not restricted to the activity of RNAPII-associated SETD2, but instead mediated by numerous methyltransferases including NSD1-3 and ASH1L. Mass spectrometry analysis revealed that H3K36me2 is an abundant modification in mammalian cells, present on between 40 - 70% of all histone H3, while approximately 10 - 20% of H3 is modified by H3K36me1 (Peters et al. 2003; Robin et al. 2007; Schotta et al. 2008; Jung et al. 2010). Kuo et al. reported broad coverage of intra- and intergenic regions with H3K36me2 in a myeloma cell line, and noted its enrichment in gene bodies where it positively correlates with the level of transcription (Kuo et al. 2011). H3K36me2 also broadly covers the genome in mouse embryonic stem cells (mESCs) (Streubel et al. 2018), however the level of H3K36me2 in gene bodies does not appear depend on transcription (Miyazaki et al. 2013). Miyazaki et al. revealed a pronounced depletion of H3K36me2 at TSS. This is in agreement with a study that demonstrated the specific depletion of H3K36me2 at CGI TSS in mESCs by ChIP-qPCR (Blackledge et al. 2010). This study noted the depletion of H3K36me2 throughout the gene body of highly transcribed genes, which may be due to its co-transcriptional conversion to H3K36me3 by SETD2.

1.1.7 Regulation of gene expression by H3K36 methylation

The best characterised function of H3K36 methylation is its suppression of cryptic transcription initiation within open reading frames. Transcription elongation is facilitated by co-transcriptional acetylation (Kristjuhan and Svejstrup 2004; Wittschieben et al. 1999), creating a hyperacetylated state which is potentially permissive for transcription initiation from cryptic TSS within coding regions. Co-transcriptional H3K36 trimethylation by Set2 is recognized by the Eaf3 subunit of the Rpd3s HDAC complex in yeast, which coordinates histone deacetylation in the wake of RNAPII to suppress cryptic initiation (Carrozza et al. 2005; Joshi and Struhl 2005; Keogh et al. 2005; Li et al. 2007). H3K36 dimethylation was shown to be sufficient for Eaf3 recruitment (Li et al. 2009; Youdell et al. 2008), suggesting that H3K36me2 and H3K36me3 may act redundantly in this pathway, and H3K36me2 was even reported to be the preferred Rpd3s substrate (Huh et al. 2012). In contrast, other studies showed that Rpd3s can directly interact with hyperphosphorylated RNAPII, independently of

H3K36 methylation (Drouin et al. 2010; Govind et al. 2010). Instead it appears that combinatorial recognition of H3K36 methylation by the chromodomain of Eaf3 and the PHD domain of Rco1, another Rpd3s subunit, is required to activate the Rpd3s complex (Li et al. 2007; Ruan et al. 2015; McDaniel et al. 2016).

Recent studies have shown that a number of additional mechanisms also underlie the suppression of cryptic initiation by Set-dependent H3K36 methylation in yeast (McDaniel and Strahl 2017). H3K36 methylation indirectly promotes a reduction of histone acetylation by suppressing the interaction of H3 with histone chaperones. This prevents histone exchange for new, acetylated histones within coding regions and is required to suppress cryptic initiation (Venkatesh et al. 2012; Venkatesh and Workman 2015). In addition, the Isw1b ATP-dependent chromatin remodelling complex is recruited to coding regions via the PWWP domain of its Ioc4 subunit, which binds H3K36me3 (Maltby et al. 2012). Here, Isw1b also acts to suppress histone exchange, and is required to suppress H4 acetylation and cryptic initiation (Smolle et al. 2012). This suppression of histone acetylation by the Isw1b chromatin remodelling complex may in part be mediated via indirect regulation of the Rpd3s deacetylase complex by chromatin remodelling, as the activity of the Rpd3s complex is modulated by nucleosomal spacing (Lee et al. 2013). Set2-mediated H3K36 methylation also appears to be required for robust transcription elongation itself, as yeast strains in which either Set2 is catalytically inactive, the SRI domain required for RNAPII interaction is mutated, or H3K36 is mutated are sensitive to the elongation inhibitor 6-azauracil (Kizer et al. 2005).

H3K36-dependent enforcement of a chromatin state that is inhibitory to transcription initiation may be conserved in mammals (Xie et al. 2011; Carvalho et al. 2013). However this appears to to be independent of histone acetylation, as SETD2 knockdown does not lead to changes in global histone acetylation levels in mouse cells, suggesting that H3K36me3-mediated inhibition of cryptic initiation occurs by a different mechanism (Edmunds et al. 2008). MRG15, an Eaf3 orthologue, was reported to form a mammalian Rpd3s-like complex containing HDAC1 and Sin3A and to bind H3K36me3 via its PWWP domain (Xie et al. 2011). This might suggest that the complex is recruited to H3K36me3-enriched coding regions to regulate transcription via histone deacetylation, as in yeast. However, Xie et al. instead found that the H3K4me2/3 demethylase activity of KDM5B,

which interacts with this complex, was critical for transcription regulation. Depletion of KDM5B or MRG15 led to increased intragenic H3K4me3 but not H4 acetylation, increased cryptic transcription initiation, and inhibited transcription elongation of target genes. Therefore, this complex appears to represses cryptic transcription through its removal of intragenic H3K4me3. In addition, SETD2-dependent recruitment of the FACT histone chaperone was shown to coordinate the reassembly of nucleosomes following their transcription-coupled displacement, thereby repressing cryptic initiation (Carvalho et al. 2013).

In addition to this role in the repression of cryptic initiation, H3K36 methylation has been implicated in both the activation and repression of transcription in a variety of contexts. Early work implicated promoter H3K36 methylation in transcription repression, as Set2 tethering to a heterologous promoter led to repression of reporter gene expression (Strahl et al. 2002) and Set2 H3K36 methylation represses basal expression of GAL4 (Landry et al. 2003). The worm NSD1 orthologue MES-4 is important for widespread transcriptional repression in late stage primordial germ cells, and is required to maintain H3K36me3 in the absence of transcription (Furuhashi et al. 2010). In contrast, H3K36 methylation is involved in the approximately two-fold upregulation of X-linked genes in males during dosage compensation in *Drosophila*, which is coordinated by the Male-specific lethal (MSL) complex. MSL is initially recruited to X-linked genes via GA-rich sequences, then spreads to local active genes and coordinates their activation via histone acetylation (Alekseyenko et al. 2008). Spreading of the MSL complex requires the chromodomain of MSL3, an Eaf3 orthologue that binds H3K36me3 *in vitro*, and Set2 (Sural et al. 2008; Larschan et al. 2007). This suggests that H3K36me3 in active gene bodies is required to recruit the MSL complex, leading to histone acetylation and hence promoting transcription activation. The activity of the histone methyltransferases NSD1-3 have been linked to both transcription activation and repression in different systems (Rahman et al. 2011; Huang et al. 1998; Berdasco et al. 2009; Kuo et al. 2011; Nimura et al. 2009). Therefore, H3K36 methylation appears to both promote and repress transcription in different contexts, with the transcriptional response regulated by both the chromatin context and the reader proteins that interpret the mark.

1.1.8 Antagonism between H3K36 methylation and H3K27 methylation

H3K36 methylation also appears to regulate gene expression via its inhibition of Polycomb Repressive Complex 2 (PRC2)-mediated H3K27 trimethylation, which is implicated in gene repression. H3K27me3 rarely co-exists with H3K36me2 or H3K36me3 in the same histone H3 molecule, and PRC2 methyltransferase activity is greatly inhibited by H3K36me2/3 *in vitro* (Yuan et al. 2011; Schmitges et al. 2011). Indeed, in *C. elegans* embryos H3K27me3 and H3K36me3 usually occur in mutually exclusive domains. In embryos lacking the NSD1 orthologue MES-4 there is a dramatic redistribution of H3K27me3 to germline genes, suggesting that H3K36me3 is required to repel H3K27me3 from these regions (Gaydos et al. 2012). In mESCs SET2-mediated H3K36me3 was reported to inhibit conversion of H3K27me1 to H3K27me2 by PRC2, leading to an accumulation of H3K27me1 in transcribed genes where it promotes transcription (Ferrari et al. 2014). While a more recent study confirmed the strong negative correlation between H3K36me3 and H3K27me3 in mESCs, it also reported a strong positive correlation between H3K27me2 and H3K36me2 (Streubel et al. 2018). In this study, NSD1 was shown to interact with PRC2, and NSD1 knockdown led to a broad increase in H3K27me3. This suggests that NSD1-mediated H3K36me2 is an important modulator of PRC2 activity, restricting the accumulation of H3K27me3 throughout the genome. The H3K36 methyltransferase ASH1L is a member of the trithorax group of proteins, which are implicated in gene activation, and also appears to play an evolutionarily conserved role in counteracting PRC2 activity. In *Drosophila*, H3K36me2 and H3K27me3 form mutually exclusive domains on salivary gland polytene chromosomes, and H3K27me3 increases in Ash1 mutants (Dorighi and Tamkun 2013). In mESCs, H3K36 methylation by ASH1L counteracts polycomb-mediated gene silencing and is required for robust induction of *Hox* gene expression (Miyazaki et al. 2013). Antagonism between H3K36 methylation and PRC2-mediated H3K27me3 therefore appears to be evolutionarily conserved, and shared between the variety of H3K36 methyltransferases present in higher eukaryotes.

H3K36me3 can also lead to the recruitment of PRC2 complexes during mammalian development, which is apparently at odds with this antagonistic relationship. The

Tudor domains of two substochiometric PRC2 subunits, PCL1/PHF19 and PCL3/PHF1, both recognise H3K36me3, leading to PRC2 recruitment to actively transcribed genes (Cai et al. 2013; Qin et al. 2013; Ballaré et al. 2012). PHF19 also recruits the H3K36me3 demethylase NO66 to these regions, coordinating PRC2-mediated H3K27me3, loss of H3K36me3 and silencing of expressed genes during differentiation (Brien et al. 2012). This coordination of H3K27 methylation via the PRC2 complex with H3K36 demethylation via NO66 is a further example of the opposing role of these modifications. In addition, H3K36 methyltransferases were shown to be inhibited by H2A ubiquitylation *in vitro* (Yuan et al. 2013). Given that PRC1 complexes, which work in concert with PRC2 to mediate gene repression, catalyse H2A K119 monoubiquitylation (ub1), antagonism between PRC1-mediated H2AK199ub1 and H3K36 methylation may reflect a broader interplay between H3K36 methylation and the histone modifications deposited by the polycomb system.

1.1.9 H3K36 methylation regulates splicing and DNA methylation

In addition to its role in transcription regulation, H3K36 methylation has also been implicated in regulating alternative splicing, another important step in the control of gene expression. For example, H3K36me3 has been shown to modulate splicing via its interaction with MRG15, which specifically recruits the negative splicing regulator PTB to H3K36me3-enriched exons (Luco et al. 2010). However, the crosstalk between H3K36 methylation and splicing is bidirectional (Almeida et al. 2011; Kim et al. 2011), therefore it remains a source of contention whether differential H3K36me3 methylation is a cause or effect of splicing.

Several studies have suggested a link between H3K36 methylation and DNA methylation, a key regulator of gene expression associated with repression of gene promoters (see Section 1.2.1). Both of the *de novo* DNA methyltransferases (DNMT) DNMT3A and DNMT3B contain a PWWP domain that binds H3K36me3 (Dhayalan et al. 2010; Rondelet et al. 2016). The interaction of the PWWP domain with H3K36me3 appears to be important for DNMT targeting and activity (Dhayalan et al. 2010). The PWWP domain of DNMT3B and SETD2-mediated H3K36 methylation are required

for the preferential enrichment of DNMT3 in the gene bodies of actively transcribed genes, which mirrors the pattern of H3K36me3 in peaking at the 3' end, and the hypermethylation of these regions (Baubec et al. 2015; Morselli et al. 2015; Jin et al. 2012). DNA methylome profiling revealed that gene body DNA methylation positively correlates with gene expression (Maunakea et al. 2010), in contrast to the association between promoter DNA methylation and gene silencing. A recent study showed that DNA methylation by DNMT3B represses cryptic initiation from within open reading frames, and that this depends on its recruitment by SETD2-mediated H3K36me3 (Neri et al. 2017). Therefore, it appears that elongating RNAPII coordinates the repression of cryptic initiation via a variety of molecular mechanisms that depend on H3K36 methylation, which not only include modifications of chromatin structure in gene body regions but also increased methylation of the genomic DNA itself. Gene body DNA methylation also appears to regulate splicing (Brown et al. 2012), which is also likely modulated by crosstalk with co-transcriptional H3K36 methylation.

1.1.10 H3K36 methylation in development and disease

While Set2 is non-essential in yeast, the H3K36 methyltransferases and demethylases present in higher eukaryotes play vital roles during embryogenesis in higher eukaryotes, indicating that precise regulation of H3K36 methylation is essential for normal development. SETD2 knockout in mice is embryonic lethal at around embryonic day (E)10.5, which is attributed to defects in angiogenesis (Hu et al. 2010). The methyltransferase NSD1 is also essential for early post-implantation development (Rayasam et al. 2003), while NSD2 knockout mice exhibit retarded growth and congenital heart defects and die soon after birth (Nimura et al. 2009). Whereas individual members of the KDM4 demethylase family are dispensable for embryogenesis, KDM4A and KDM4C play a redundant and essential role in early embryogenesis (Pedersen et al. 2016), and maternal expression of KDM4A is also critical for pre-implantation development (Sankar et al. 2017). KDM2 family members are also essential for embryonic development, which will be discussed in further detail in Section 1.3.6.

Members of this network of enzymes that regulate H3K36 methylation are also

implicated in a variety of diseases and cancers, reported to play both oncogenic and tumour supressor roles in different tumour types (Wagner and Carpenter 2012; Pedersen and Helin 2010). This indicates that control of H3K36 methylation is not only vital during embryogenesis, but must also be regulated throughout the lifetime of higher eukaryotes to maintain cell homoeostasis. Indeed, deregulation of H3K36 methylation has been mechanistically implicated in cancer progression in a number of studies. For example, fusion of NSD1 with the nucleoporin NUP98 occurs as a result of chromosomal translocation in a subset of human acute myeloid leukaemias (Jaju et al. 2001). This fusion protein localises to *Hox* genes, leading to increased H3K36 methylation and the suppression of PRC2-mediated H3K27me3, resulting in a methyltransferase-dependent increase in *Hox-A* gene activation and myeloid progenitor immortalization (Wang et al. 2007). NSD2 is overexpressed in multiple myeloma cells, leading to a global increase in H3K36 methylation accompanied by the expression of quiescent oncogenes, which depends on NSD2 catalytic activity (Kuo et al. 2011). This link between H3K36 methylation and tumour progression is further strengthened by the discovery of 'oncohistones', which contain a K36-to-methionine mutation in the histone H3.3 variant and drive over 90% of chondroblastomas (Fang et al. 2016; Lu et al. 2016). Oncohistone expression leads to a global loss of H3K36 methylation and a concomitant increase in H3K27 methylation, via a dominant-negative mechanism involving sequestration of H3K36 methyltransferases (Zhang et al. 2017).

It is important to consider that a number of the enzymes that regulate H3K36 methylation have also been shown to act on non-histone substrates. For example, NSD1 and KDM2A regulate NFκB via methylation and demethylation of its p65 subunit, respectively (Lu et al. 2010). In addition, KDM4C may demethylate several non-histone proteins, including Cbx4 (Yang et al. 2011; Ponnaluri et al. 2009). These proteins may also have roles beyond the regulation of protein methylation. Indeed, KDM4A was reported to act as a p53-specific E3 ubiquitin ligase, targeting this key cell cycle regulator for proteasomal degradation (Johmura et al. 2016). In addition, KDM2B plays an important role in recruiting the PRC1.1 complex to CGI gene promoters (Farcas et al. 2012, see Section 1.3.4). These additional roles may contribute to the phenotypes reported following deregulation of these proteins during development and

disease, making the direct impact of altered H3K36 methylation difficult to ascertain. Nevertheless, the implication of the many components of the H3K36 methylation regulatory network in critical developmental and disease pathways suggests that H3K36 methylation regulation itself plays a key role. Catalytic knockout or rescue experiments, for example in the NSD1 and NSD2 cancer studies described above, provide further evidence of the direct contribution of H3K36 regulatory activity to these processes.

1.2 CGI promoters as a platform for transcription regulation

1.2.1 DNA methylation in vertebrate genomes

While the histones that package DNA are subject to a range of modifications, genomic DNA is itself extensively methylated. The majority of CpG dinucleotides in vertebrate genomes are methylated on the cytosine residue at the C5 position (Lister et al. 2009), which is catalysed by a family of DNMTs. DNMT1 is required for the maintenance of DNA methylation (Jeltsch 2006). With the aid of PCNA and UHRF1, DNMT1 interacts with hemi-methylated DNA at replication forks generated by the semiconservative replication of symmetrically methylated CpG dinucleotides (Bostick et al. 2007; Chuang et al. 1997; Hashimoto et al. 2008). Here DNMT1 re-establishes symmetrical CpG methylation of the daughter strands following DNA replication (Bestor and Ingram 1983; Pradhan et al. 1999; Vertino et al. 2002). Importantly, this restoration of symmetrical CpG methylation following semi-conservative replication means that DNA methylation is a heritable epigenetic modification that can be stably maintained (Holliday and Pugh 1975; Riggs 1975). DNMT3A and DNMT3B are responsible for *de novo* DNA methylation (Okano et al. 1999), including the establishment of genome-wide patterns of DNA methylation following the epigenetic reprogramming that occurs during early embryogenesis and primoridal germ cell development (Hajkova 2011). The catalytic activity and chromatin targeting of DNMT3A/B is regulated by a related protein, DNMT3L, which itself lacks methyltransferase activity (Gowher et al. 2005; Chedin et al. 2002; Suetake et al. 2004; Ooi et al. 2007).

DNA methylation is closely linked with transcription repression, and plays an essential

role during mouse development (Bird and Wolffe 1999; Lei et al. 1996; Li et al. 1992; Okano et al. 1999). Early studies showed that DNA methylation is required to maintain the silencing of genes on the inactive X chromosome (Mohandas et al. 1981; Venolia et al. 1982) and of imprinted genes (Razin and Cedar 1994). They also uncovered an association between gene promoter DNA methylation and gene repression, and showed that loss of DNA methylation can result in reactivation of silenced genes (Antequera et al. 1990; Harris 1982; Holliday 1987; Jaenisch et al. 1985). It appears that DNA methylation is not the initiating event in gene silencing, but instead establishes stable repression (Payer and Lee 2008; Okamoto and Heard 2009). The role played by DNA methylation in transcriptional repression was also highlighted by experiments showing that methylated transgenes are transcriptionally inactive following transfer into recipient cells, in contrast to the expression of non-methylated controls (Vardimon et al. 1981; Waechter and Baserga 1982). Cancer studies have revealed that aberrant methylation of tumour supressor gene promoters results in their transcription inactivation (Jones and Baylin 2007), further underlining the importance of transcription regulation by DNA methylation.

1.2.2 Methyl-binding proteins recognise DNA methylation

How does DNA methylation at a gene promoter lead to transcription repression? One possibility is that DNA methylation prevents transcription factors from binding to their target genes and therefore directly interferes with transcription activation. Indeed, a number of transcription factors have been shown to be unable to bind to their target sequences when the sequences are methylated, including CTCF (Comb and Goodman 1990; Iguchi-Ariga and Schaffner 1989; Watt and Molloy 1988; Bell and Felsenfeld 2000). However, repression of methylated transgenes occurs only after chromatin assembly, and DNA methylation at some distance from the gene promoter can also lead to transcription repression, suggesting that another mechanism is also at play (Buschhausen et al. 1987; Kass et al. 1993). Indeed, DNA methylation is specifically recognised by proteins containing a methyl-CpG binding domain (MBD), which coordinate transcription repression via their association with a number of co-repressor complexes. MeCP2, the first MBD protein to be identified, plays an important

role in methylation-dependent gene silencing (Meehan et al. 1992; Lewis et al. 1992; Nan et al. 1997). MeCP2 recruits mSIN3A-containing HDAC complexes, resulting in chromatin deacetylation and thus transcription repression, and HDAC activity is required for MeCP2-mediated transcription repression *in vivo* (Nan et al. 1998; Jones et al. 1998). MeCP2 mutations lead to the severe neurological disorder Rett syndrome, highlighting the critical role played by this protein (Amir et al. 1999). Studies of *Mecp2* disease mutations revealed that MeCP2 also interacts with NCoR/SMRT co-repressor complexes (Lyst et al. 2013).

Homology searches led to the identification of additional proteins containing MBD binding domains, including MBD1, MBD2 and MBD4 (Cross et al. 1997; Hendrich and Bird 1998). MBD2 also links DNA methylation with hypo-acetylated chromatin, acting as a core subunit of the Mi2-Nucleosome Remodelling Deacetylase (NuRD) complex, and is required for efficient NuRD-mediated gene silencing (Ng et al. 1999; Ramirez et al. 2012). In addition to the MBD domain, a number of transcriptional repressors have been shown to bind methylated DNA via zinc finger domains. These include KAISO (Prokhortchouk et al. 2001), which was shown to repress methylated genes through its interaction with the NCoR deacetylase complex (Yoon et al. 2003). Thus a variety of DNA binding proteins recruit chromatin modifying activities to regions of methylated DNA, to coordinate histone hypoacetylation and gene repression. However, the overall role that methylated DNA binding proteins play in the regulation of gene expression remains unclear, particularly as the majority of CGI gene promoters remain free of DNA methylation irrespective of their transcriptional activity and therefore are not directly targeted by these proteins.

1.2.3 CGI are associated with the majority of vertebrate gene promoters

CpG dinucleotides are under-represented in vertebrate genomes, occurring at around one fifth of the expected frequency (Josse et al. 1961; Swartz et al. 1962). This is due to the spontaneous deamination of methylated cytosine residues to produce thymine (Coulondre et al. 1978; Duncan and Miller 1980). Incomplete repair of the resulting T:G mismatches leads to TpG conversion which, if it occurs in the germline,

results in the gradual depletion of CpG dinucleotides from the genome throughout evolution (Antequera 2003; Sved and Bird 1990). Despite this overall depletion of CpG dinucleotides, vertebrate genomes are punctuated by short contiguous regions of DNA, typically 1-2 kb in length, that have elevated GC content and CpG densities approaching the expected frequency (Bird et al. 1985; Gardiner-Garden and Frommer 1987). Critically, the majority of these CpG islands (CGI) remain free of DNA methylation, in contrast to the rest of vertebrate genomes. Indeed, CGI were first identified as genomic regions susceptible to cutting by DNA methylation-sensitive restriction endonucleases (Cooper et al. 1983; Bird et al. 1985). This hypomethylated state of CGI is likely to be key to their conservation through evolution, as it prevents CpG mutation via deamination (Bird et al. 1985).

Closer examination revealed that CGI are found at approximately 50-70% of vertebrate gene promoters (Tykocinski and Max 1984; Bird et al. 1985; Saxonov et al. 2006). The majority of these CGI gene promoters remain free of DNA methylation regardless of cell type or gene expression level (Meissner et al. 2008; Weber et al. 2007). The non-methylated state of CGI promoters appears to be required for normal transcription, as their increased methylation is associated with gene repression (Jones 2012; Bird 2002; Bird and Wolffe 1999), including during tumorigenesis and other pathologies (Baylin et al. 1998; Costello et al. 2000; Robertson and Wolffe 2000; Portela and Esteller 2010). A small percentage of CGI gene promoters acquire methylation during normal development, and this is associated with transcriptional repression (Schilling and Rehli 2007; Weber et al. 2007; Shen et al. 2007; Mohn et al. 2008; Illingworth et al. 2008; Illingworth and Bird 2009). The mechanisms that ensure that the majority of CGI remain free of DNA methylation remain incompletely characterised. They are likely to have evolved as a result of selective pressure to maintain normal levels of transcription of essential genes, given the strong link between CGI promoter methylation and gene repression.

The association of CGI with gene promoters and their refractory nature to DNA methylation suggests that they may play an important role in transcription regulation. Given that CGI gene promoters generally remain non-methylated irrespective of promoter activity, it appears that the presence of a CGI does not directly define the

transcriptional state of associated genes. Instead it has been proposed that CGI act as a platform for the recruitment of chromatin modifiers and transcription regulators to generate a transcriptionally permissive chromatin state (Blackledge and Klose 2011; Deaton and Bird 2011). Indeed, early studies provided evidence for this hypothesis, as CGI nucleosomes were found to be hyperacetylated and depleted of the linker histone H1, both signatures of actively transcribed regions (Tazi and Bird 1990). Furthermore, CGI were shown to be sensitive to nuclease digestion (Tazi and Bird 1990; Stalder et al. 1980; Bird et al. 1985; Antequera et al. 1989), and subsequent genome-wide micrococcal nuclease (MNase)-seq and deoxyribonuclease (DNase)-seq approaches demonstrated that chromatin accessibility and nucleosome depletion are defining characteristics of CGI rather than CpG poor promoters (Fenouil et al. 2012; Sabo et al. 2004; Teif et al. 2012; Li et al. 2011; Choi 2010). Elevated chromatin accessibility may facilitate binding of RNAPII, basal transcription factors and other transcription regulators to CGI, and has therefore been proposed to play an important role in their function as gene expression regulatory platforms (Blackledge and Klose 2011; Deaton and Bird 2011). Induction of CGI-associated immediate-early genes occurs rapidly and without a requirement for chromatin remodelling, in contrast to immediate-early genes that are not associated with CGI, adding further evidence that CGI may confer an accessible and transcriptionally permissive chromatin state (Ramirez-Carrozzi et al. 2009).

1.2.4 ZF-CXXC proteins are recruited to CGI gene promoters

The discovery that methylated CpG dinucleotides are recognized by MBD proteins prompted a search for a specialised protein domain that might instead recognise non-methylated CpG dinucleotides. Skalnik et al. performed a phage display screen for proteins that bind to non-methylated DNA, leading to the identification of a protein that binds to non-methylated CpG dinucleotides via a cysteine rich zinc finger(ZF)-CXXC domain (Voo et al. 2000; Lee et al. 2001). This protein was later named CFP1 (CXXC finger protein 1). The ZF-CXXC domain is characterised by two conserved cysteine-rich clusters that coordinate two Zn^{2+} ions, stabilising the domain in an extended crescent-shaped structure that interacts extensively with the major groove of DNA via a DNA-binding loop (Long et al. 2013). Structural studies revealed that the DNA-binding loop

is in close proximity to the CpG dinucleotide, and cytosine methylation leads to a severe steric clash which prevents the formation of hydrogen bonds that are essential for DNA binding (Cierpicki et al. 2010; Xu et al. 2011; Song et al. 2011). This underlies the specificity of the domain for non-methylated CpGs. Bioinformatic analyses led to the discovery of a family of 12 ZF-CXXC proteins in mice. These can be grouped into three subtypes, based on a divergent linker region that appears to define DNA sequence recognition properties.

Many ZF-CXXC proteins either possess histone-modifying activities or are part of chromatin modifying complexes. For example, CFP1 is a component of SETD1 complexes, one of the major H3K4me2/3 methyltransferases in eukaryotic cells (Lee and Skalnik 2005; Lee et al. 2007). CFP1 chromatin occupancy is determined by both its ZF-CXXC domain, which binds to non-methylated CpGs, and its PHD domain, which binds H3K4me3, leading to its preferential enrichment at actively transcribed CGI promoters that have elevated H3K4me3 (Lee et al. 2001; Thomson et al. 2010; Brown et al. 2017; Eberl et al. 2013). CFP1 is the central determinant of SETD1 promoter occupancy, and is required to maintain H3K4me3 levels at the promoters of actively transcribed genes (Brown et al. 2017; Clouaire et al. 2012; Clouaire et al. 2014). Two members of the MLL family of H3K4 methyltransferases also contain ZF-CXXC domains that bind to non-methylated DNA (Birke et al. 2002; Bach et al. 2009). While MLL1 localisation appears to be restricted to a subset of actively transcribed CGI (Guenther et al. 2005; Milne et al. 2005), MLL2 appears to bind more broadly to the majority of CGI (Denissov et al. 2014). This CGI localisation indicates that the ZF-CXXC domain of MLL1/2 is important for targeting. However, a number of studies suggest that other chromatin binding domains and interactions with other proteins also play a role (Zeleznik-Le et al. 1994; Chang et al. 2010; Wang et al. 2010; Milne et al. 2010; Park et al. 2010; Muntean et al. 2010).

CFP1, MLL1 and MLL2 therefore contribute to the enrichment of H3K4me3 at CGI promoters, which is a hallmark of these regions regardless of their transcriptional activity (Mikkelsen et al. 2007; Guenther et al. 2007; Thomson et al. 2010). H3K4me3 appears to promote transcription initiation via a number of mechanisms, including the recruitment of chromatin remodelling complexes and the TFIID general transcription factor (see Section 1.1.3). Therefore, enrichment of H3K4me3 may contribute to a

transcriptionally permissive chromatin state at CGI associated gene promoters (Deaton and Bird 2011; Blackledge and Klose 2011). CGI are also specifically depleted of H3K36me2 through the demethylase activity of the ZF-CXXC-containing proteins KDM2A and KDM2B, as will be discussed in detail in Section 1.3. Given that H3K36 methylation is inhibitory to transcription initiation (Section 1.1.7), its depletion at CGI promoters may also contribute to the generation of a distinctive chromatin state that is permissive to transcription.

1.3 The lysine demethylase 2 (KDM2) protein family

1.3.1 KDM2A and KDM2B are H3K36me2 demethylases

KDM2A/JHDM1A/FBXL11/NDY2 was the first JmjC histone demethylase to be identified. KDM2A demethylase activity was discovered using a screen for the release of ^{3}H-formaldehyde, one of the predicted products of the demethylase reaction, from radiolabelled nucleosomal histones (Tsukada et al. 2006). Tsukada et al. showed that the 2-oxoglutarate- and Fe^{2+}-dependent demethylase activity of KDM2A is conferred by its JmjC domain, and that KDM2A specifically demethylates H3K36me2, and to a lesser extent H3K36me1, *in vitro* and *in vivo*. In contrast, KDM2A is unable to demethylate H3K36me3. A crystal structure of KDM2A suggests that this specificity for H3K36me1/2 may be due to steric constraints following H3K36me3 binding, which could prevent 2-oxoglutarate from undergoing a rotation within the active site that is necessary for the demethylation reaction (Cheng et al. 2014). This study also revealed that a number of interactions between specific residues on the H3 tail and the catalytic groove of KDM2A contribute to its specificity for the H3K36 residue.

The KDM2A paralogue KDM2B/JHDM1B/FBXL10/NDY1 also acts as a H3K36 demethylase, with specificity for H3K36me1/2 *in vitro* and *in vivo* (Tsukada et al. 2006; He et al. 2008). KDM2B has also been reported to act as a H3K4me3 demethylase *in vitro* and at specific loci *in vivo* (Frescas et al. 2007; Janzer et al. 2012), however other studies suggest that H3K36me2 is the main substrate for KDM2B histone demethylase activity (He et al. 2008; Tsukada et al. 2006; Tzatsos et al. 2009). From here on I will refer to

KDM2A and KDM2B collectively as KDM2A/B.

KDM2A and KDM2B have a very similar domain architecture, including a Jumonji C (JmjC) demethylase domain, ZF-CXXC DNA binding domain, F box, PHD domain and leucine-rich repeats (LRR, Figure 1.2 A). While PHD domains in many proteins function as lysine or methyl-lysine binding modules (Taverna et al. 2007), the PHD domain of KDM2A does not appear to contribute to chromatin binding (Zhou et al. 2012). This is likely due to the absence of an aromatic cage that is required for methyl-lysine binding and the absence of conserved residues that contribute to lysine binding in other PHD domains. The ZF-CXXC domain of KDM2A binds to non-methylated DNA *in vitro* and targets KDM2A to CGI throughout the genome, regardless of the transcriptional activity of the CGI-associated genes (Blackledge et al. 2010). The acquisition of DNA methylation at a subset of CGI during differentiation is sufficient to inhibit KDM2A binding, adding further evidence that the localisation of KDM2A is restricted to regions enriched for non-methylated CpG dinucleotides. KDM2A specifically interacts with linker DNA and is unable to bind DNA that is wound around nucleosomes (Zhou et al. 2012). This is likely due to the requirement of the ZF-CXXC domain to access both the major and minor groove for efficient DNA binding, and therefore may be a more general property of ZF-CXXC proteins. Strikingly, loss of DNA methylation in *Dnmt3a*, *Dnmt3b*, *Dnmt1* triple knockout mESCs had no impact on the CGI binding profile of KDM2A (Zhou et al. 2012). This suggests that the occlusion of CpGs by nucleosomes and the low CpG dinucleotide frequency in the genome act as major barriers to KDM2A binding, and therefore contribute to CGI binding specificity.

The ZF-CXXC domain of KDM2B also specifically binds to non-methylated CpG dinucleotides (Farcas et al. 2012; Koyama-Nasu et al. 2007; Yamagishi et al. 2008), and KDM2B is recruited to CGI throughout the genome (Farcas et al. 2012; Wu et al. 2013; He et al. 2013). A subset of CGI associated with developmental genes are preferentially enriched for KDM2B and depleted of KDM2A in mESCs. Here KDM2B recruits the PRC1.1 complex to coordinate gene repression (Farcas et al. 2012), which will be discussed in detail in Section 1.3.4.

Members of the KDM2 family are found from budding yeast to human, and their JmjC

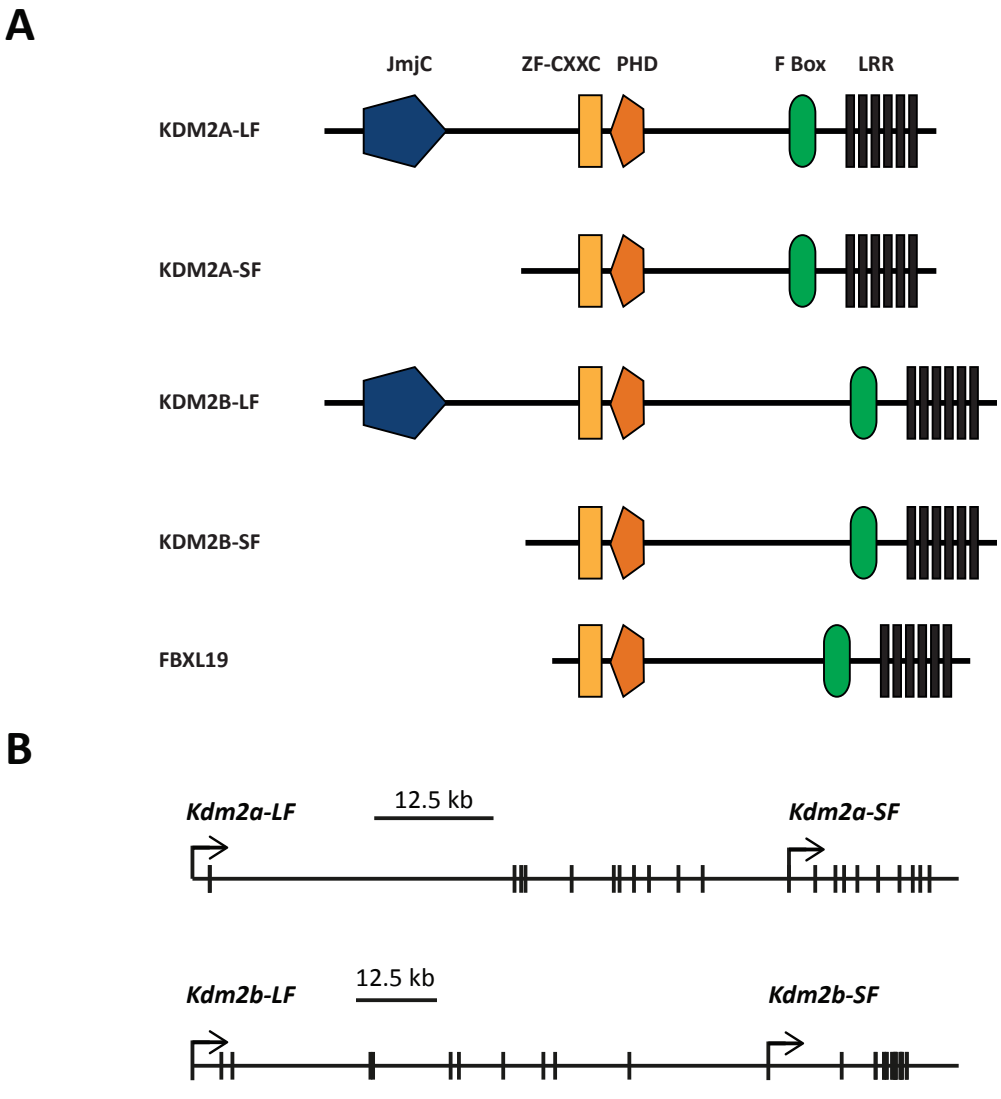


Figure 1.2: The KDM2 family share a similar domain architecture
(A) Schematic representation of the long (LF) and short (SF) isoforms of KDM2A/B and of FBXL19, showing their domain architecture. JmjC: jumonji C demethylase domain, ZF-CXXC: zinc finger CXXC domain, PHD: plant homeodomain, F box: F-box domain, LRR: leucine rich repeats.
(B) Schematic depicting the structure of the *Kdm2a* and *Kdm2b* genes. Transcription initiating from an internal promoter leads to the production of KDM2A/B-SF, which lack the JmjC demethylase domain.

domains are highly conserved (Tsukada et al. 2006; Klose et al. 2006). *Drosophila* dKDM2/CG11033 and *S. cerevisiae* JHD1 have been shown to catalyse H3K36me2 demethylation (Lagarou et al. 2008; Tsukada et al. 2006; Fang et al. 2007). Other KDM2 orthologues also contain highly conserved cofactor binding residues, suggesting that these also function as H3K36 demethylases (Klose et al. 2006). In contrast, a mutation in one of the Fe^{2+} coordinating residues renders the *S. pombe* KDM2 orthologue Epe1 catalytically inactive *in vitro* (Tsukada et al. 2006). Epe1 instead appears to limit heterochromatic domains in a JmjC-dependent manner (Ayoub et al. 2003; Wang et al. 2013). The KDM2 domain architecture is highly conserved in flies, mice and humans, whereas orthologues in lower eukaryotes lack several domains. *S. cerevisiae* JHD1 comprises the PHD and JmjC domains and functions with the H3K36me3 demethylase RPH1 to promote transcription by antagonising SET2-mediated H3K36 methylation (Kim and Buratowski 2007; Klose et al. 2007).

1.3.2 KDM2A/B have long and short isoforms

The *Kdm2a* and *Kdm2b* genes both have an internal TSS downstream of the exons encoding the JmjC domain (Tanaka et al. 2010; Long et al. 2013) (Figure 1.2 B). Transcription initiating from these alternative promoters gives rise to short isoforms of KDM2A and KDM2B, which lack the JmjC domain and therefore do not act as histone demethylases. However, they retain the other functional domains of these proteins, including the ZF-CXXC domain required for CGI targeting, and therefore are likely to share any demethylase-independent functions exhibited by the full-length isoforms (Figure 1.2 A). I will refer to these short isoforms as KDM2A/B-SF, and the full-length isoforms, which possess demethylase activity, as KDM2A/B-LF. The role played by KDM2A-SF remains poorly defined, while KDM2B-SF is sufficient to recruit the PRC1.1 complex to CGI (Blackledge et al. 2014), as will be discussed in Section 1.3.4.

The expression of these different KDM2A/B isoforms appears to be differentially regulated, which likely occurs through differential regulation of their independent gene promoters. KDM2B-LF protein is highly abundant in mESCs, embryonic germ cells and neural progenitors compared to more differentiated cell types, and KDM2B-

LF expression greatly decreases during embryoid body differentiation and embryonic development (He et al. 2013; Fukuda et al. 2011). In contrast, KDM2B-SF is ubiquitously expressed during embryonic development and in a variety of adult tissues (He et al. 2013; Fukuda et al. 2011). This suggests that KDM2B-LF and hence KDM2B demethylase activity may play a key role in the early embryo, while KDM2B-SF may predominate in adult tissues. *Kdm2a-LF* and *Kdm2a-SF* mRNA expression were detected in a number of mouse adult tissues and at a variety of embryonic developmental stages, with the exception of the brain front lobe (Lainović et al. 2017). In all of these tissues, *Kdm2a-LF* mRNA was more highly expressed than *Kdm2a-SF* mRNA, indicating that KDM2A-LF protein may be more abundant that KDM2A-SF in a range of cell types. KDM2A protein is also detected at a number of stages during mouse embryonic development, peaking at E10.5-14.5 where it is expressed throughout the whole body and is enriched in the forebrain and limbs (Kawakami et al. 2015). These studies provide evidence that the expression of the long and short isoforms of KDM2A/B is differentially regulated. This may potentially modulate the relative contribution of demethylase-dependent and -independent activities of KDM2A/B in different cell types and developmental stages.

The KDM2 family also has a third member in vertebrates, FBXL19 (Katoh and Katoh 2004). Unlike KDM2A and KDM2B, FBXL19 lacks a JmjC domain and, therefore, does not possess demethylase activity (Long et al. 2013). Instead FBXL19 is a paralogue of KDM2A/B-SF, and has the same domain architecture (Figure 1.2 A). FBXL19 was recently shown to be recruited to CGI via its ZF-CXXC domain (Lee et al. 2017; Dimitrova et al. 2018). Here FBXL19 is required for the activation of a subset of developmental genes during lineage commitment, through its recruitment of CDK8-mediator to prime the promoters of these genes (Dimitrova et al. 2018). FBXL19 has also been implicated in the recruitment of RNF20 to chromatin to mediate H2B mono-ubiquitylation (Lee et al. 2017). In addition to these chromatin-associated functions, FBXL19 has been shown to play a role in a variety of cytoplasmic processes, by acting as an E3 ubiquitin ligase which targets the GTPases RhoA, Rac1 and Rac3 and the interleukin receptor ST2L for proteasomal degradation (Dong et al. 2014; Wei et al. 2013; Zhao et al. 2012; Zhao et al. 2013).

1.3.3 Regulation of gene expression by the KDM2 family

KDM2A/B have been proposed to promote transcription initiation by removing H3K36me2 from CGI gene promoters. Blackledge et al. showed that a number of CGI promoters are specifically depleted of H3K36me2, using ChIP-qPCR. Loss of KDM2A from these CGI, following either KDM2A knockdown or CGI methylation during differentiation, leads to an increase in H3K36me2 at these regions, indicating that KDM2A plays an active role in H3K36me2 removal (Blackledge et al. 2010). There is a variety of evidence linking H3K36me2 with the inhibition of transcription initiation within gene body regions in yeast, and this function may be conserved in mammalian cells (Section 1.1.7). Therefore, local depletion of H3K36me2 at CGI by KDM2A/B may contribute to a transcriptionally permissive chromatin signature at CGI gene promoters (Figure 1.3, Blackledge et al. 2010; Blackledge and Klose 2011; Deaton and Bird 2011). Depletion of either KDM2A or KDM2B leads to deregulated gene expression (Blackledge et al. 2010; Farcas et al. 2012; Blackledge et al. 2014; He et al. 2013), however the contributions of demethylase-dependent or independent activities of KDM2A and KDM2B to these gene expression changes have not been characterised. In addition, these two paralogues may act as functionally redundant transcriptional regulators. Therefore the overall, genome-wide role played by KDM2A/B in transcription regulation remains unclear.

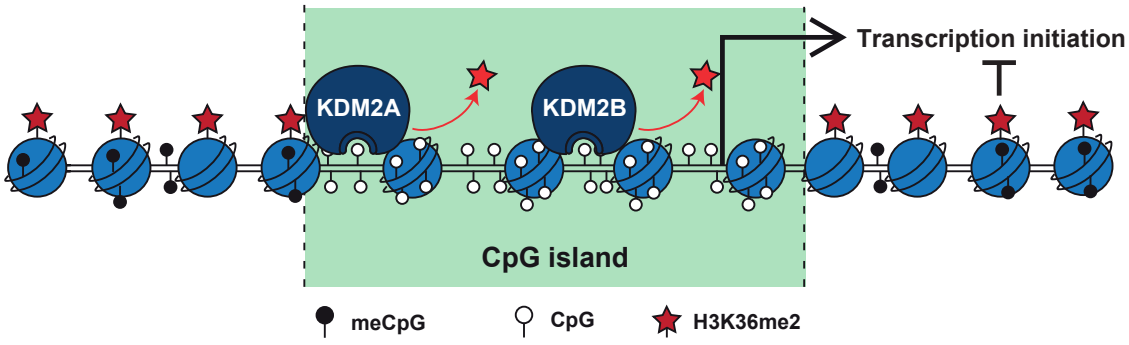


Figure 1.3: KDM2A/B are proposed to contribute to the generation of a transcriptionally permissive chromatin state at CGI promoters
Schematic depicting the proposed model for transcription regulation by KDM2A/B. KDM2A/B are recruited to CGI gene promoters throughout the genome via their ZF-CXXC domain. Here KDM2A/B may catalyse the demethylation of H3K36me2, a histone modification that is associated with the repression of transcription initiation. Local depletion of this mark at CGI through the activity of KDM2A/B may, therefore, contribute to the generation of a transcriptionally permissive chromatin state at CGI gene promoters.

KDM2A/B have also been implicated in transcription repression, including the repression of genes that play critical roles in cell proliferation, differentiation and senescence. For example, both KDM2A and KDM2B were reported to negatively regulate cell growth in response to starvation by repressing the transcription of ribosomal RNA genes, via H3K36me2 demethylation of their gene promoter region (Frescas et al. 2007; Tanaka et al. 2010). While these studies found that KDM2A/B localised to the nucleolus and reported H3K4me3 demethylase activity for KDM2B, multiple other studies failed to reproduce these observations (Blackledge et al. 2010; Farcas et al. 2012; He et al. 2008; Pfau et al. 2008; Han et al. 2016). The demethylase activity of KDM2A/B was also reported to promote, rather than constrain, cell proliferation in other contexts. For example, overexpression of either KDM2A or KDM2B leads to the immortalization of embryonic fibroblasts via a JmjC domain-dependent mechanism (Pfau et al. 2008). Transcriptional repression of the tumour suppressors $p15^{Ink4b}$ and $p16^{Ink4a}$ by KDM2B plays an important role in this process, and requires KDM2B demethylase activity (He et al. 2008; Tzatsos et al. 2009). In addition, KDM2B-LF is also implicated in the repression of $p19^{ARF}$, an inducer of apoptosis (Fukuda et al. 2011; Tzatsos et al. 2009). KDM2A was shown to regulate the proliferation and differentiation potential of mesenchymal stem cells (MSCs). KDM2A-mediated repression of $p15^{Ink4b}$/ $p27^{Kip1}$ and of Sox2/ Nanog expression is implicated in these processes. However, this appears to occur independently of the demethylase activity of KDM2A, as H3K36me2 levels did not change at the promoters of these genes (Gao et al. 2013; Dong et al. 2013). In contrast, repression of Ereg and Sfrp2 expression by KDM2A is associated with a reduction of H3K36me2 at their gene promoters, suggesting that the demethylase activity of KDM2A may play a role. This is implicated in the inhibition of MSC osteo/dentitogenic differentiation (Du et al. 2013; Yu et al. 2016).

KDM2A and KDM2B were both reported to promote somatic cell reprogramming via their demethylase activity (Wang et al. 2011; Liang et al. 2012). While Wang et al. found that repression of the *Ink4* locus by KDM2B demethylase activity played an important role in overcoming cell senescence during reprogramming (Wang et al. 2011), Liang et al. instead reported that the predominant role played by KDM2B demethylase activity is the facilitation of gene activation during early reprogramming (Liang et al. 2012).

KDM2A is also implicated in the transcriptional repression of centromeric satellite repeats, which appears to occur through both demethylase-dependent and -independent mechanisms. KDM2A knockdown leads to increased transcription of satellite repeats and an increase in H3K36me2 at these regions, and transcriptional repression depends on the KDM2A JmjC domain, implicating the demethylase activity of KDM2A in this process (Frescas et al. 2008). Interaction between KDM2A and Heterochromatin Protein 1 (HP1) is also reported to contribute to this transcriptional repression, by promoting the formation of pericentromeric heterochromatin. HP1 is known to interact with H3K9me3, a histone PTM associated with transcriptional repression, and to stimulate further H3K9me3 and DNA methylation via the recruitment of a H3K9 methyltransferase and DNMTs to generate transcriptionally repressive heterochromatin (Fanti and Pimpinelli 2008). Bartke et al. reported that KDM2A binds to H3K9me3-modified nucleosomes in the presence of HP1, and that this binding is abolished by DNA methylation (Bartke et al. 2010). In contrast, Borgel et. al found that KDM2A could be recruited to H3K9me3-modified nucleosomes containing methylated DNA in the presence of HP1α, and that HP1α could be recruited to nucleosomes lacking H3K9me3 in the presence of KDM2A. These studies led to a model in which KDM2A initiates silencing: KDM2A recruits HP1 to chromatin lacking DNA methylation and H3K9me3, where HP1 coordinates H3K9me3 and DNA methylation to form 'mature' heterochromatin (Borgel et al. 2017). KDM2A-SF was reported to interact with HP1 (Borgel et al. 2017; Bartke et al. 2010) and to form HP1α-dependent heterochromatin bodies (Lainović et al. 2017), suggesting that both isoforms may contribute to this demethylase-independent process. The yeast KDM2 orthologue Epe1 is also involved in heterochromatin maintenance, indicating that this may be a conserved role, although Epe1 has also been implicated in heterochromatin destabilisation (Ayoub et al. 2003; Zofall and Grewal 2006; Wang et al. 2013).

Early studies suggested that KDM2B is also involved in demethylase-independent transcriptional silencing (Koyama-Nasu et al. 2007; Yamagishi et al. 2008). This can be explained by its role in polycomb-mediated gene repression, which will be discussed in detail in Section 1.3.4. In addition, Boulard et al. proposed another demethylase-independent role for KDM2B, to protect polycomb-targeted genes from *de novo* promoter DNA methylation in order to sustain their transcription. Promoter methylation

otherwise coincides with transcriptional repression of these genes following constitutive KDM2B knockout (Boulard et al. 2015). KDM2B-SF expression was sufficient to rescue this hypermethylation phenotype, indicating that KDM2B demethylase activity is not required. Another potential demethylase-independent function of KDM2B is its reported E3 ubiquitin ligase activity (Janzer et al. 2012; Han et al. 2016), which will be the focus of Section 1.3.5. KDM2A has also recently been suggested to possess E3 ligase activity (Bueno et al. 2018). Furthermore, KDM2A may indirectly regulate protein ubiquitylation by recruiting the E3 ligase NEDD4, which leads to RNAPII ubiquitylation and enhancer deactivation in breast cancer cells (Tan et al. 2018).

The third KDM2 family member, FBXL19, was also recently shown to regulate transcription. Loss of FBXL19 binding to chromatin prevents the activation of a subset of developmental genes during cell lineage commitment. This is due to the loss of FBXL19-mediated recruitment of CDK8-mediator to the promoters of these genes, which is important for promoter priming (Dimitrova et al. 2018). FBXL19 lacks a JmjC demethylase domain and is a paralogue of KDM2A/B-SF. These short isoforms of KDM2A/B may also play an important role in transcription regulation, however their specific roles have not been studied in detail.

Finally, KDM2A/B may indirectly regulate gene expression through their demethylation of non-histone substrates. Both KDM2A and KDM2B were shown to demethylate the active, non-phosphorylated form of β-catenin in the nucleus, which promotes its proteasomal degradation. This prevents the over-accumulation of this transcriptional regulator following activation of the Wnt signalling pathway, limiting the extent of target gene upregulation (Lu et al. 2015). KDM2A was also reported to demethylate the p65 subunit of NF-κB, a critical transcriptional regulator of the innate and adaptive immune responses. KDM2A counteracts the activation of NF-κB, which is promoted by NSD1-mediated p65 methylation (Lu et al. 2010). Several other JmjC-containing demethylases have also been shown to demethylate non-histones substrates, which may complicate studies aiming to determine the direct role of their histone demethylase activities (Dimitrova et al. 2015).

A large number of studies therefore implicate demethylase-dependent or -independent

activities of KDM2A/B in transcriptional regulation. These studies were performed using a wide variety of cell types and experimental conditions, suggesting that the transcriptional response may depend on the target gene and the cellular context. The genome-wide role played by KDM2A/B in transcriptional regulation, and the contribution of their demethylase activity to this process, remains unclear.

1.3.4 KDM2B targets PRC1.1 to CGI gene promoters

Polycomb group proteins play a critical role in transcriptional repression in multicellular organisms, and are essential for normal embryonic development (Lewis 1978). These proteins are generally found in one of two chromatin-modifying complexes, Polycomb Repressive Complexes 1 and 2 (PRC1 and PRC2). PRC1 complexes monoubiquitylate H2AK119 via the E3 ligase activity of their RING1A/B subunit (Napoles et al. 2004; Wang et al. 2004), and are classified into subgroups based on their association with one of six polycomb group ring finger (PCGF) subunits (Gao et al. 2012). PCGF proteins act as bridging molecules within the PRC1 complex that confer unique subunit composition and are required for H2AK119ub1 catalysis (Cao et al. 2005; Li et al. 2006; Ben-Saadon et al. 2006; Buchwald et al. 2006). PRC2 complexes trimethylate H3K27 via the catalytic activity of their EZH1/2 subunit (Cao et al. 2002; Czermin et al. 2002; Kuzmichev et al. 2002; Muller et al. 2002). These chromatin modifications are thought to generate a compact and transcriptionally repressive chromatin state at PRC-occupied CGI (Di Croce and Helin 2013; Muller and Verrijzer 2009).

Target sites are generally co-occupied by PRC1 and PRC2 complexes, which together play an essential role in coordinating normal developmental gene regulation (Simon and Kingston 2013). In *Drosophila*, PRCs are targeted to Polycomb Response Elements through the activity of DNA-binding transcription factors (Steffen and Ringrose 2014). In contrast, while there is evidence that transcription factors and long non-coding RNAs contribute to PRC recruitment to specific loci in vertebrates, these targeting mechanisms do not appear to be widespread (Blackledge et al. 2015). Instead, PRC-occupied sites overlap almost exclusively with a subset of CGI (Ku et al. 2008). Together with the observation that DNA sequences with a high density of unmethylated CpG dinucleotides

are sufficient for PRC recruitment and H3K27me3 *in vivo* (Mendenhall et al. 2010; Lynch et al. 2012), this led to the proposal that CGI play a direct role in PRC recruitment.

Two groups identified KDM2B as part of a PRC1 complex in mammalian cancer cell lines, comprising KDM2B, SKP1, RING1A/B, PCGF1, BCOR and RYBP (Gearhart et al. 2006; Sanchez et al. 2007). Given that KDM2B possesses a ZF-CXXC domain, this finding raised the possibility that KDM2B might broadly target PRC1 to CGI gene promoters. dKDM2 was also shown to be part of a variant PRC1 complex in *Drosophila*, named dRAF. This complex includes dRING, the sole PRC1 E3 ligase, and PSC, a PCGF1 orthologue, suggesting that PRC1 regulation might be an evolutionarily conserved role of KDM2 proteins (Lagarou et al. 2008). *dKdm2* mutation increases the frequency of homeotic transformations in PC mutant flies, providing evidence that dRAF represents a *bona fide* variant PRC1 complex. KDM2B was subsequently shown to form a variant PRC1 complex in mESCs, characterised by the PCGF1 subunit and therefore referred to as PRC1.1 (Farcas et al. 2012; He et al. 2013; Wu et al. 2013). In contrast, SKP1 was the sole interaction partner identified for KDM2A (Klose lab, unpublished). Therefore KDM2B, but not KDM2A, is incorporated into PRC1 complexes. This requires the F box and LRR domains of KDM2B, and a structural study revealed that BCOR and PCGF1 cooperatively bind to the LRR (Wong et al. 2016; Blackledge et al. 2014; Inagaki et al. 2015). KDM2B plays an important role in PRC1 recruitment, as KDM2B depletion leads to a reduction in RING1B occupancy and H2AK119ub1 at target regions, and to the upregulation of a subset of polycomb target genes (Farcas et al. 2012; Wu et al. 2013; He et al. 2013; Blackledge et al. 2014). KDM2B-mediated targeting of the PRC1.1 complex is also required for normal PRC2 levels at a subset of target sites (Blackledge et al. 2014). Furthermore, KDM2B is sufficient to recruit PRC1.1 and establish a *de novo* polycomb domain containing PRC2, and this activity is shared by both KDM2B-LF and KDM2B-SF (Blackledge et al. 2014).

These studies revealed that KDM2B recruits PRC1.1 to CGI through direct binding to non-methylated, CpG rich DNA. This provided an alternative targeting mechanism to the prevailing hierarchical recruitment model, in which pre-existing PRC2 target sites recruit chromobox protein-containing canonical PRC1 complexes via their recognition of H3K27me3 (Cao et al. 2002; Min et al. 2003; Wang et al. 2004). This alternative

model explains the maintenance of H2AK119ub1 and gene repression at many PRC1 targets following loss of PRC2-mediated H3K27me3 (Schoeftner et al. 2006; Pasini et al. 2007; Leeb et al. 2010; Tavares et al. 2012). Loss of KDM2B leads to a reduction but not complete loss of RING1B at target sites and to the upregulation of only a subset of PRC1 target genes, indicating that the hierarchical recruitment system and direct targeting mechanisms function together in a cooperative and sometimes redundant fashion to coordinate PRC1 recruitment to target sites throughout the genome.

KDM2B is recruited broadly to CGI gene promoters, leading to the proposal that KDM2B may act as a sampling module that recruits low levels of PRC1 to most or all CGI. Indeed, CGI throughout the genome are occupied by low levels of PRC1, and this depends on KDM2B (Farcas et al. 2012; Blackledge et al. 2014). Local chromatin features associated with transcription may prevent polycomb accumulation at active gene promoters, leading to the establishment of polycomb domains at only the CGI promoters of inactive genes to enforce transcriptional repression (Blackledge et al. 2014; Long et al. 2013; Klose et al. 2013). This is consistent with the demonstration that polycomb domains are mainly formed after transcriptional silencing has been achieved (Hosogane et al. 2013; Yuan et al. 2012; Riising et al. 2014), suggesting that polycomb primarily functions to prevent inappropriate gene reactivation rather than to initiate silencing.

H3K36me2 is antagonistic to PRC2-mediated H3K27me3, and is deposited by a number of methyltransferases including the trithorax protein ASH1L (Section 1.1.8). KDM2B may therefore indirectly promote PRC2 activity at CGI by catalysing H3K36 demethylation (Wu et al. 2013). This may be part of a more general antagonism between the trithorax and polycomb systems acting at CGI gene promoters (Klose et al. 2013). For example, members of the trithorax family catalyse H3K4me3 and H3K27 acetylation, chromatin modifications associated with transcriptional activation which also inhibit the catalytic activity of PRC2 (Schmitges et al. 2011; Reynolds et al. 2012; Tie et al. 2009). Furthermore, H3K27me3 deposited by PRC2 may inhibit H3K4 methylation by SETD1 (Kim et al. 2013). In contrast to to this antagonistic activity, positive feedback mechanisms appear to promote the formation of either a polycomb or trithorax-dominant chromatin state. Canonical CBX-containing PRC1 complexes are recruited by PRC2-deposited H3K27me3, while PRC2 complexes can recognize, and are stimulated

by, PRC1-deposited H2AK119ub1 (Min et al. 2003; Wang et al. 2004; Cao et al. 2002; Blackledge et al. 2014; Cooper et al. 2014; Kalb et al. 2014). Positive feedback also drives the accumulation of H3K4 methyltransferases, and hence H3K4me3, at the promoters of highly transcribed genes, through the H3K4me3 binding domains of CFP1 and MLL1/2 (Chang et al. 2010; Milne et al. 2010) and through methyltransferase recruitment by RNAPII (Lee et al. 2008; Ng et al. 2003).

These antagonistic and positive feedback mechanisms have been proposed to underlie a bistable switch between trithorax and polycomb states (Klose et al. 2013). In this model the dominant state is robustly maintained by positive feedback mechanisms and can only be overcome following a concerted increase in the opposing system, conferring both epigenetic stability and the possibility to rapidly and robustly switch to the alternative state. KDM2B may contribute to this bistability by enabling PRC1.1 complexes to constantly sample the chromatin state of CGI gene promoters. Low level sampling may, if the appropriate conditions are found, nucleate polycomb domain formation and lead to the switch from a trithorax-dominant to a polycomb-dominant state. In addition, H3K36me2 demethylation by KDM2B may also contribute to this antagonism towards the trithorax system.

1.3.5 Putative E3 ligase activity of KDM2 proteins

Another potential mechanism by which KDM2 proteins may regulate transcription is through their putative E3 ligase activity. KDM2A, KDM2B and FBXL19 are F-box proteins, which are characterised by their ability to confer substrate specificity to SKP1-CUL1-F box (SCF) E3 ubiquitin ligase complexes via their protein-protein interaction domains (Ho et al. 2006). Ubiquitin is a highly conserved 76 amino acid protein that can be conjugated to specific lysine side chains of other proteins and to itself to create polyubiquitin chains. Ubiquitylation is classically associated with proteasomal degradation, but can also regulate the activity or interactions of the substrate protein (Komander 2009). SCF complexes transfer ubiquitin from an E2 enzyme to specific substrates. They contain a CUL1 protein which acts as a molecular bridge between the RING domain-containing E3 ligase Rbx1/2 and the adaptor protein SKP1, which binds

to one of a diverse family of F-box proteins (Cardozo and Pagano 2004). KDM2A/B and FBXL19 are classified into the FBXL F-box protein subfamily, characterised by leucine-rich repeat protein interaction domains (Jin et al. 2004, Figure 1.2 A). Furthermore, KDM2A/B and FBXL19 all interact with SKP1 (Anca Farcas and Emilia Dimitrova unpublished, Koyama-Nasu et al. 2007; Farcas et al. 2012; Tan et al. 2013; Gearhart et al. 2006; Han et al. 2016; Wong et al. 2016), and this interaction has been mapped to the F-box domain of KDM2B (Gearhart et al. 2006; Koyama-Nasu et al. 2007).

These properties suggest that the KDM2 family may be incorporated into SCF E3 ubiquitin ligase complexes. However, purifications of KDM2A/B and FBXL19 from cell extracts have generally failed to co-purify CUL1 (Anca Farcas, Emilia Dimitrova and Jessica Kelley unpublished, Tan et al. 2013; Farcas et al. 2012; Gearhart et al. 2006; Koyama-Nasu et al. 2007; Gao et al. 2012), suggesting that they do not form canonical SCF complexes. However, an interaction between KDM2B and CUL1 was reported in HEK293 cells (Han et al. 2016) and an FBXL19-CUL1 interaction was detected both *in vitro* and in cells overexpressing FBXL19 (Jessica Kelley, unpublished), suggesting that KDM2 proteins may interact with CUL1 under certain conditions. Interaction between KDM2 proteins and CUL1 may generally be sterically occluded by other interacting proteins, or may be regulated according to substrate availability (Reitsma et al. 2017).

KDM2B was reported to possess intrinsic E3 ligase activity via its PHD domain (Janzer et al. 2012) and to ubiquitylate c-Fos, targeting this transcription factor for proteasomal degradation (Han et al. 2016). In addition, the PHD domain of the KDM2A/B-SF paralogue FBXL19 also displays E3 ligase activity *in vitro* (Emilia Dimitrova and Jessica Kelley, unpublished) and FBXL19 has been shown to ubiquitylate the GTPases RhoA, Rac1 and Rac3 and the interleukin receptor ST2L, leading to their proteasomal degradation (Dong et al. 2014; Wei et al. 2013; Zhao et al. 2012; Zhao et al. 2013). KDM2A may also possess E3 ligase activity, as its overexpression stimulates 53BP1 ubiquitylation and leads to an increase in total ubiquitylation levels in whole cell extract (Bueno et al. 2018). Therefore, KDM2 proteins may possess both E3 ubiquitin ligase activity via their PHD domain and substrate selectivity via their LRR domains, activities usually exerted by distinct proteins within SCF complexes. This dual function of KDM2 proteins might abrogate the need for a bridging CUL1 subunit.

The proposal that PHD domains, including the PHD domain of KDM2B, may possess E3 ligase activity has sparked controversy. PHD domains are considered to function primarily as a chromatin binding module (Sanchez and Zhou 2011). However, E3 ligase activity has been reported for a number of PHD domains, including those of MEKK1 and MIR1/2 proteins (Lu et al. 2002; Coscoy et al. 2001; Mansouri et al. 2003; Yonashiro et al. 2006; Hewitt et al. 2002; Uchida et al. 2004; Wang et al. 2012). PHD and RING domains have a very similar conformation, in which the peptide chain folds back on itself to form a cross-brace structure with key cysteine residues coordinating two zinc ions (Pascual et al. 2000; Capili et al. 2001). Therefore, it is possible that the PHD domains shown to exert E3 ubiquitin ligase activity are instead misannotated RING domains (Aravind et al. 2003; Scheel and Hofmann 2003; Dul and Walworth 2007). Analysis of structural data suggests that KDM2A, KDM2B and FBXl19 contain *bona fide* PHD domains (Xu et al. 2018, Jessica Kelley unpublished), which could cast doubt on their E3 ubiquitin ligase activity. However, the structure of the first PHD finger of the autoimmune regulator AIRE, which possesses E3 ligase activity (Uchida et al. 2004), also confirmed its identity as a PHD domain (Bottomley et al. 2005). Therefore it is possible that PHD domains may themselves act as E3 ligases.

The E3 ligase activity of the KDM2 proteins remains to be validated, and the potential contribution of this activity to the regulation of chromatin structure and transcription is not clear. The H3K4me1/2 demethylase KDM1B/LSD2 also appears to function as an E3 ligase, targeting the polyubiquitylation of the O-GlcNAc transferase OGT, which leads to its proteasomal degradation (Yang et al. 2016). In addition, KDM4A was reported to act as a p53-specific E3 ubiquitin ligase, targeting this key cell cycle regulator for proteasomal degradation (Johmura et al. 2016). These studies raise the possibility that E3 ligase activity may be a more common property of histone demethylases than is currently appreciated.

1.3.6 Role of KDM2 proteins in development and disease

The KDM2 proteins play essential roles during development, as knockout of either KDM2A, KDM2B or FBXL19 leads to embryonic lethality. Knockout of either KDM2B-LF

or KDM2B-SF in isolation leads to increased perinatal lethality with neural tube defects (Boulard et al. 2015; Fukuda et al. 2011). Therefore, the individual isoforms of KDM2B are largely dispensable for embryonic development. Furthermore the demethylase activity of KDM2B, which is exhibited solely by KDM2B-LF, is evidently not fundamental to embryogenesis. Knockout of KDM2B-LF and KDM2B-SF together leads to a much more severe phenotype, with embryonic lethality occurring between E7 and E13 with complete penetrance (Blackledge et al. 2014; Boulard et al. 2015; Andricovich et al. 2016). The increased severity of complete KDM2B knockout reveals that KDM2B-LF and KDM2B-SF together play an essential role during embryonic development, revealing functional redundancy between the two isoforms. This suggests that a demethylase-independent activity of KDM2B is critical to development, consistent with the severe developmental defects observed in a bovine mutant carrying a missense mutation in the C-terminal region of KDM2B (Testoni et al. 2012).

In contrast, knockout of KDM2A-LF alone leads to impaired neural development and embryonic lethality between E10.5 and E12.5, indicating that KDM2A-LF in isolation is essential for embryonic development (Kawakami et al. 2015). The phenotype of KDM2A-SF knockout mice, and mice in which all KDM2A isoforms are knocked out, has not been reported to date. Deletion of the CXXC domain of FBXL19, which leads to its removal from chromatin, also leads to embryonic lethality at approximately E10.5, accompanied by clear growth retardation phenotypes including undeveloped facial mesenchyme and hypomorphic cardiac mesenchyme (Dimitrova et al. 2018). The essential role played by FBXL19 during development again highlights the importance of the demethylase-independent roles of the KDM2 family.

An early study reported that dKDM2 is essential for *Drosophila* development, suggesting that there is an evolutionarily conserved requirement for the KDM2 family during embryogenesis. However, the transposon insertion lines used were either hypomorphic or failed to affect dKDM2 expression, and also disrupted neighbouring genes (Zheng et al. 2018). Instead, the careful generation of dKDM2 null alleles using CRISPR-mediated genome editing revealed that *dKdm2* homozygous mutants are fully viable and fertile with no obvious developmental defects, but instead display circadian rhythm defects (Zheng et al. 2018). In contrast to this lack of an essential role for dKDM2

during fly development, KDM2A/B appear to play an important role during *Xenopus* embryogenesis, as they are required for specification of the body axis (Lu et al. 2015).

KDM2A/B are deregulated in a wide variety of cancer types, and have been reported to act both as oncogenes and tumour suppressors depending on the context. For example, overexpression of KDM2B is reported to promote tumour progression in acute myeloid leukemia, breast and pancreatic cancer (He et al. 2011; Tzatsos et al. 2013; Penzo et al. 2015), while overexpression of KDM2A promotes tumorigenesis in non small cell lung cancer and gastric and breast cancer (Dhar et al. 2014; Wagner et al. 2013; Liu et al. 2016; Huang et al. 2015). KDM2A-SF is overexpressed in a subset of aggressive breast cancers due to amplification of the C-terminal region of the *Kdm2a* gene. KDM2A-SF appears to as an oncogenic isoform in these cells, as knockdown of KDM2A-SF, but not KDM2A-LF, inhibits breast cancer cell growth and invasion (Liu et al. 2016). In contrast to these roles in promoting tumour progression, KDM2B is reported to act as a tumour suppressor in lymphoma and brain cancers (Suzuki et al. 2006; Frescas et al. 2007), while KDM2A is implicated as a tumour suppressor in lymphoma, prostate and brain cancer (Lu et al. 2010; Frescas et al. 2008). The implication of KDM2A/B in the progression of a broad range of cancers, and their apparent ability to act as oncogenes or tumour suppressors in different cancer types, highlights the context-dependent role that these proteins play in regulating cell proliferation and differentiation.

1.4 Aims of this thesis

KDM2A/B are recruited to CGI gene promoters throughout the genome via their ZF-CXXC domain. KDM2A/B catalyse the demethylation of H3K36me1/2 via their JmjC-catalytic domain, a histone modification that is associated with the repression of transcription initiation. This has led to the model that KDM2A/B contribute to the generation of a CGI-associated transcriptionally permissive chromatin state, through local removal of this repressive mark. However, KDM2A/B also appear to repress transcription in a number of contexts. Furthermore, KDM2A/B are also reported to regulate transcription in a demethylase-independent manner. In addition, both KDM2A and KDM2B have short isoforms that lack demethylase activity and are

relatively poorly characterised. Therefore, the genome-wide role played by KDM2A/B in regulating chromatin architecture and transcription, and the relative contributions of their demethylase-dependent and -independent activities, remains unclear. The central aim of this thesis is to dissect the contribution of KDM2A/B to transcription regulation.

To address this question, I combine genome engineering and conditional ablation strategies with genome-wide analyses in mESCs. I first examine the contribution of KDM2A/B histone demethylase activity to transcription regulation (Chapter 3). Given that KDM2A and KDM2B are highly homologous proteins, they may regulate transcription in a functionally redundant manner. Demethylase-independent activities of KDM2A/B may also play a key role in the regulation of gene expression, and these activities would likely be shared between the long and short isoforms of the proteins. I directly investigate these possibilities by studying the transcriptional changes following conditional removal of all KDM2A and KDM2B isoforms in mESCs (Chapter 4). I also investigate redundancy with the KDM2A/B-SF paralogue FBXL19 by studying the transcriptional changes following removal of KDM2A/B and FBXL19 in combination. I then dissect the individual contributions of KDM2A and KDM2B to transcription regulation (Chapter 5). Finally, I investigate the putative demethylase-independent mechanisms by which KDM2A/B may regulate transcription (Chapter 6). Together, this work provides insight into the genome-wide role played by KDM2A/B in regulating transcription and chromatin architecture, and the contribution of the demethylase-dependent and independent activities of KDM2A/B to these processes.

2

Materials and Methods

2.1 Cell lines and culture

2.1.1 Cell culture conditions

mESCs were grown at 37°C and 5% CO_2 in gelatin-coated dishes. Dulbecco's modified eagle medium (Gibco) was supplemented with 15% foetal calf serum (FCS), 10 ng/ml leukemia-inhibitory factor, penicillin/streptomycin, beta-mercaptoethanol, L-glutamine and non-essential amino-acids. Cells were passaged using trypsin-EDTA (0.25%, Gibco). Conditional mESC lines were treated with 800 nM tamoxifen for 96 h to induce Cre-mediated recombination.

Drosophila melanogaster SG4 cells were grown adhesively at 25°C in Schneider's *Drosophila* medium (Gibco) supplemented with penicillin/streptomycin and 10% heat-inactivated FCS.

2.1.2 Embryonic stem cell lines

The conditional mESC lines used in this study are summarised in Table 2.1. These were provided by Prof Haruhiko Koseki, RIKEN Center for Integrative Medical Sciences.

Name	Genotype
Kdm2a/b-JmjC$^{fl/fl}$	*Kdm2a-JmjC$^{fl/fl}$;Kdm2b-JmjC$^{fl/fl}$;Cre-ERT2*
Kdm2b-CXXC$^{fl/fl}$	*Kdm2b-CXXC$^{fl/fl}$;Cre-ERT2*
Skp1$^{fl/fl}$	*Skp1$^{fl/fl}$;Cre-ERT2*
Cre-ERT2	*Cre-ERT2*

Table 2.1: Genotypes of conditional mESC lines

2.1.3 CRISPR-mediated genome editing

Insertion of loxP sites was achieved using CRISPR-mediated genome editing. sgRNA-encoding oligos were cloned into the pSpCas9(BB)-2A-Puro plasmid, as previously described (Ran et al. 2013). Targeting constructs encoding the loxP sequence flanked by 150 bp homology arms were ordered from GeneArt (ThermoFisher). A 60-70% confluent 90 mm plate of mESCs was transfected with 3 µg of sgRNA plasmid and 10 µg of targeting construct, using Lipofectamine 2000 (30 µl) or Lipofectamine 3000 (30 µl + 26 µl p3000 reagent, Life Technologies). The following morning, cells were trypsinised and re-plated at a variety of densities, before selection with 1 µg/ml puromycin for 48 h. Individual colonies were later transferred to 96-well plates. Positive clones were identified using PCR screens of genomic DNA, and validated by Sanger sequencing of PCR products.

2.1.4 Cell counting and cell size measurement

A Countess II cell counter (Life Technologies) was used to count cells and to measure cell diameter. Cell diameter > 12 µm gating was used for mammalian cells, while cell diameter > 3 µm and cell elongation < 70 settings were used for *Drosophila* cells.

2.2 Protein methods

2.2.1 Acid extraction of histones

Throughout this study, centrifugation steps involving protein samples were performed at 4°C, unless otherwise indicated. For histone extraction, cell pellets were washed in 1ml RSB buffer (10 mM Tris HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2 and 20 mM NEM), then resuspended in RSB buffer supplemented with 0.5% NP-40 and incubated on ice for 10 min to allow cell lysis. Following centrifugation at 500xg for 5 min, the nuclear pellet was resuspended in 5 mM MgCl2 and 20 mM NEM. An equal volume of 0.8 M HCl was added, and the mixture was incubated on ice for 20 min to allow histone extraction. After centrifugation at 13000 rpm for 20 min, histones were precipitated from the supernatant

by adding TCA to 25% final concentration and incubating on ice for 30 min. Histones were pelleted by centrifugation at 13000 rpm for 15 min, and the pellet was washed twice in ice cold acetone. The histone pellet was resuspended in 1x SDS loading buffer (2% SDS, 100 mM Tris pH 6.8, 100 mM DTT, 10% glycerol and 0.1% bromophenol blue) by vortexing, and boiled at 95°C for 5 min.

2.2.2 Preparation of nuclear extracts

Cell pellets were resuspended 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 1x cOmplete protease inhibitor cocktail (PIC, Roche)) and incubated on ice for 10 min. Following centrifugation at 2000 rpm for 5 min, the cell pellet was resuspended in 3 volumes of Buffer A supplemented with 0.1% NP-40 and incubated on ice for 10 min. Nuclei were recovered by centrifugation at 4000 rpm for 5 min and resuspended in 1 pellet volume of Buffer B (250 mM NaCl, 5 mM Hepes pH 7.9, 26% glycerol, 1.5 mM MgCl2, 0.2 mM EDTA, 0.5 mM DTT and 1x PIC). The volume of nuclear suspension was measured and the NaCl concentration increased to 400 mM by dropwise addition of 5M NaCl, assuming a 250 mM NaCl starting concentration. After 1 h of rotation at 4°C, the suspension was pelleted at 13000 rpm for 20 min and the supernatant saved as nuclear extract. Extract concentration was determined by Bradford assay (BioRad), and Buffer B containing 400 mM NaCl was used to dilute extracts for western blotting.

2.2.3 Western blotting

Protein extracts were mixed with 1x SDS loading buffer and boiled at 95°C for 5 min. Proteins were separated according to their molecular weight using SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Gels were cast using the Mini-Protean Tetra Cell system (BioRad). Separating gel containing between 6% and 15% acrylamide/bis-acrylamide (BioRad) was used, depending on the size of the protein of interest. This also contained 0.1% SDS, 0.1% ammonium persulphate (Sigma), 0.1% TEMED (Sigma) and 400 mM Tris HCl pH 8.8. The stacking gel contained 5% acrylamide/bis-acrylamide and replaced Tris-HCl pH 8.8 with 125 mM Tris HCl pH 6.8. Gels were run at 200 V in 1x SDS-PAGE

running buffer (25mM Tris, 192mM glycine and 0.1% SDS).

Following SDS-PAGE, gels were transferred to nitrocellulose membrane by semi-dry transfer, using the Trans-Blot Turbo Transfer System (BioRad, 1.5 A for 7 min). The membrane was blocked in 5% milk in 1x TBS + 0.1% Tween 20 (Fisher, TBST) solution for 1 h at room temperature, then incubated with primary antibody in 5% milk in TBST (MTBST) overnight with gentle rocking. Following 3 quick and 3x 10 minute washes in TBST, the membrane was incubated in MTBST containing horse radish peroxidase- or infrared dye- conjugated secondary antibody (1:2500 or 1:15000 dilution, respectively) for 1 h at room temperature with gentle rocking. After 3 quick and 3x 10 minute washes in TBST, membranes either were soaked in equal volumes of ECL solution I (2.5mM luminal (Sigma), 396 M p-coumaric acid (Sigma) and 100 mM Tris HCl pH 8.5) and ECL solution II (5.6 mM H2O2 and 100 mM Tris HCl pH 8.5) and exposed to X-Ray film (GE Healthcare), or were imaged using an Odyssey Fc system (LI-COR, for infrared dye-conjugated antibodies).

2.2.4 Antibodies

Table 2.2 lists the antibodies used in this study. These include an antibody raised against the C-terminal region of KDM2A, which I affinity purified from rabbit serum. I expressed His-tagged KDM2A antigen in bacterial cells and performed Ni-NTA purification, then covalently coupled the antigen to Affi-Gel 10 (BioRad) and used this to immunoaffinity purify the KDM2A antibody. I also generated an antibody specific for H3K36me1. Histone H3 APATGGV(K)KPHRYRPGC peptide, with the indicated lysine mono-methylated, was conjugated to maleimide-activated keyhole limpet hemocyanin (Pierce) for antibody generation. The H3K36me1 peptide was immobilised on SulfoLink coupling resin (Pierce) and used to immunoaffinity purify the H3K36me1 antibody from rabbit serum. The specificity of the H3K36me1/2/3 antibodies was validated using Active Motif MODified histone peptide arrays (Millipore).

Antigen	Source of antibody	Dilution for western blotting
H3	Klose lab	1:3000
H4	Cell signalling L64C1	1:4000
H3K36me1	Klose lab	1:1000
H3K36me2	Klose lab	1:1000
H3K36me3	Klose lab	1:1000
H3K4me1	Klose lab	1:1000
H3K4me3	Klose lab	1:1000
H2AK199ub1	Cell signalling D27C4	1:2000
H3K9ac	Cell signalling C5B11	1:1000
H3K27ac	Cell signalling DSE4XP	1:1000
H3K27me3	Cell signalling C36B11	1:2000
FBXL19	Klose lab	1:2000
KDM2A	Klose lab	1:1000
KDM2B	Klose lab	1:1000
BRG1	Abcam EPNCIR111A	1:15000
CUL1	Bethyl A303-373A	1:1000
SKP1	Santa cruz sc-7163	1:750
TBP	Abcam Ab818	1:2000
anti-mouse secondary (IRDye800)	LI-COR 32210	1:15000
anti-rabbit secondary (IRDye680)	LI-COR 68071	1:15000
anti-mouse secondary (HRP)	GE Healthcare	1:2500
anti-rabbit secondary (HRP)	GE Healthcare	1:2500

Table 2.2: Antibodies used in this study

2.2.5 Ubiquitylome mass spectrometry

Protein LoBind tubes (Eppendorf) and RPT low retention tips (Starlab) were used throughout the preparation of peptide samples for mass spectrometry analysis. Approximately 20 x 145 mm dishes of confluent mESCs were treated with 10 µM MG-132 (Sigma) proteasome inhibitor for 4 h prior to harvesting. Plates were rinsed in PBS, then 3 ml of PBS containing 1xPIC and 10mM NEM was added per plate, and the cells harvested by scraping into falcon tubes. Dishes were rinsed with more of the PBS solution, then the falcon tubes were centrifuged at 800xg for 5 min. The supernatant was discarded, and the cell pellet was snap-frozen in liquid nitrogen. Cell pellets were thawed on ice, then nuclear extraction was performed using a modified version of the protocol above (Section 2.2.2). Buffer A was supplemented with 10mM NEM. Nuclei recovered after incubation in Buffer A + 0.1% NP40 were washed three times with 3 pellet volumes of Buffer A + 0.1% NP40. The washed nuclear pellet was resuspended

in 9 pellet volumes of freshly prepared 1.1x NUN buffer (1.65 M urea, 27.5 mM Hepes pH 7.6, 0.385 M NaCl, 1.1% NP40, 1.1 mM DTT, 1.1x PIC and 11 mM NEM), vortexed briefly and incubated on ice for 25 minutes before centrifugation at 13000 rpm for 10 min. Protein was quantified by Bradford assay (BioRad), and stored at -80°C in 10% glycerol.

Next, filter-aided sample preparation was used to remove NP40 and DTT from the nuclear extract, which otherwise dominate mass spectra. All centrifugation steps were performed at 14300xg for 30-40 min at room temperature. Vivacon 500 filters (Sartorius) were washed with 200 µl 0.1% TFA (Thermo Scientific) in 50% ACN by centrifugation (16 filters per sample). Nuclear extract was added (400 µg per filter), followed by centrifugation. The filters were washed 6 times using 200ul 8M urea in 100 mM TEAB (Thermo Scientific), then reduced with 10 mM TCEP (Thermo Scientific) in 200 µl 8M urea in 100 mM TEAB, and incubated at room temperature for 30 min with gentle shaking. Filters were alkylated by adding 50 mM C-IAA and incubated for 30 min at room temperature in the dark with gentle shaking. Following centrifugation, filters were washed twice with 1M urea in 50 mM TEAB, then the collection tube was changed. Trypsin was added to each filter (10 µg, Promega, diluted in 1 M urea and 50 mM TEAB) and incubated overnight at 37°C with gentle shaking. The following day, 10 ug trypsin diluted in 1 M urea and 50 mM TEAB was added, followed by incubation for 4 h at 37°C with gentle shaking. Peptides were eluted by centrifugation, and the filter washed with 150 µl 0.1% TFA then with 150 µl 50% ACN in 0.1% TFA into the same elution tube. The sample was dried using a speed vacuum concentrator at room temperature and stored -20°C.

Trypsin-digested peptides were purified using a 360 mg Sep-Pak c18 column (Waters), using gravity flow. Columns were pre-washed with 5 ml 100% ACN, then equlibrated with 2x 5 ml 0.1% TFA. Dried peptide was resuspended in 100 µl 0.1% TFA per tube and pooled before loading onto the column, and 0.1% TFA was used to wash the tubes and add residual peptide to the column. Columns were washed 2x with 4 ml 0.1% TFA. The collection tube was changed and peptides were eluted with 2 ml 50% ACN in 0.1% TFA, then with 2 ml 80% ACN in 0.1% TFA. An input sample (10 µg) was retained, and samples were dried using a speed vacuum concentrator at room temperature before

storage at -20°C.

PTMScan ubiquitin remnant motif immunoaffinity purification was performed according to the manufacturer's protocol (Cell signalling). Purified ubiquitylated peptides were dried using a speed vacuum concentrator and stored at -20°C. A further round of c-18 cleanup was performed prior to peptide analysis using liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 HPLC system (Thermo Scientific). An input sample was also analysed to control for changes in peptide abundance rather than ubquitylation, and MS data was analysed using the Perseus platform with a $p < 0.001$ threshold. LC-MS/MS and analysis were performed by Deniz Kaya, under the supervision of Shabaz Mohammed.

2.3 Chromatin immunoprecipitation

2.3.1 Preparation of chromatin

ChIP was performed as previously described (Blackledge et al. 2010; Schmidt et al. 2009), with minor modifications. For KDM2A/B ChIP, $5x10^7$ mESCs were fixed in 2 mM disuccinimidyl glutarate (Thermo Scientific) for 1 h at room temperature with gentle rotation, then in 1% formaldehyde for 12.5 min (methanol-free, Life Technologies). For histone ChIP, $1x10^7$ mESCs were fixed for 10 min in 1% formaldehyde alone. Reactions were quenched by adding glycine to a final concentration of 125 µM and incubating for 5 min, before centrifugation at 1000xg for 5 min and a PBS wash.

Different chromatin preparation methods were used for KDM2A/B ChIP and histone ChIP. For KDM2A/B ChIP, the cell pellet was resuspended in 10 ml LB1 buffer (50 mM HEPES-KOH pH 7.9, 140 mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X 100 and 1xPIC) and rotated for 10 min at 4°C. The released nuclei were pelleted by centrifugation at 1000xg for 5 min, then resuspended in 10 ml buffer LB2 (10 mM Tris HCl pH 8.0, 200 mM NaCl, 1mM EDTA, 0.5 mM EGTA and 1xPIC) and rotated for 5 minutes at 4°C. Following centrifugation at 1000xg for 5 min, the nuclear pellet was resuspended in 1 ml buffer LB3 (10 mM Tris HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5

mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine and 1xPIC). For histone ChIP, the cell pellet was resuspended in 1ml lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris HCl pH 8.1 and 1xPIC) and incubated on ice for 30 min.

Chromatin was sonicated with a BioRuptor sonicator (Diagenode) cooled to 4°C, using 25-30 cycles of 30 s on/ 30 s off at full power. After centrifugation at 20000×g for 20 min at 4°C, the supernatant was transferred to a fresh tube and a ten-fold dilution of the chromatin in 0.1 M NaOH was quantified by NanoDrop.

2.3.2 Immunoprecipitation

For KDM2A/B ChIP, 150 µg chromatin was used per IP. For histone ChIP, chromatin corresponding to $1x10^5$ cells was used for each IP. Chromatin was diluted ten-fold in ChIP dilution buffer (1% Triton-X100, 1 mM EDTA, 20 mM Tris HCl pH 8, 150 mM NaCl and protease inhibitors) and pre-cleared for 1-2 h with either protein A magnetic Dynabeads (Invitrogen, 10 µl, for KDM2A/B ChIP) or protein A agarose beads (Repligen, 20 µl packed beads, for histone ChIP) that had previously been blocked for 1 h with 500 µg/ml BSA and 50 µg/ml yeast tRNA. An input sample was saved, and immunoprecipitation was performed overnight at 4°C with approximately 3 µg of antibody. The following day, antibody-bound chromatin was isolated on blocked protein A magnetic/agarose beads (10 µl or 20 µl packed beads, respectively). ChIP washes were performed with 1 ml of buffer and rotation at 4°C for 4 min, followed by centrifugation at 1000 rpm for 1 min (agarose beads) or transfer to a magnet (Dynabeads): low salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris HCl pH 8.1, 150mM NaCl), high salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20mM Tris HCl pH 8.1, 500mM NaCl), LiCl buffer (0.25 M LiCl, 1% NP40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris HCl pH 8.1) and 2 washes with TE buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA). ChIP DNA was eluted by vortexing for 30 min in 100 µl elution buffer (1% SDS and 100 mM NaHCO3), Cross-links were reversed at 65°C overnight with 200 mM NaCl, combined with RNase treatment (1 µg, Roche). Proteinase K (20 µg) was added for 1 h at 45°C, before DNA purification using a ChIP DNA Clean Concentrator kit (Zymo). Histone ChIP-seq libraries were prepared using the NEBNext Ultra DNA library preparation kit

(NEB), from 10 ng ChIP material.

2.4 RNA methods

2.4.1 RNA extraction, reverse transcription and quantitative PCR

RNA was routinely extracted using TRIzol reagent (Invitrogen), according to the manufacturer's protocol. Purified RNA was diluted to 100 ng/µl in nuclease-free water (Life Technologies), and 4 µl was reverse transcribed using the imProm-II system (Promega) using random hexamer primers and RNasin ribonuclease inhibitor. A control lacking reverse transcriptase was included for each reverse transcription experiment.

The generated cDNA was typically diluted with 300 ul nuclease-free water and analysed by quantitative real-time PCR (qPCR). qPCR reactions consisted of 5 µl diluted (c)DNA sample, 7.5 µl 2x SensiMix SYBR Green No-ROX mastermix (Bioline), 0.375 µl each of forward and reverse primers (10 µM) and 1.75 µl water. qPCR was performed in a Rotor-Gene 6000 machine (Corbett): 10 min hold at 95°C followed by 45 cycles of 95°C for 10 s, 60°C for 15 s, 72°C for 15 s. Each qPCR reaction was run in duplicate, and a no-template control was included for each primer pair. A melting curve was generated at the end of each run to verify the amplification specificity.

qPCR primers were designed using the Primer 3 program, limiting the product size range to 80-120 bp, melting temperatures to 59-61°C and temperature difference to 1.0°C. cDNA primers were designed to span large introns, so that cDNA and gDNA products could be clearly distinguished. All primer pairs were checked for off-targets using UCSC primer blast, and their efficiency was tested using a serial dilution curve of (c)DNA.

2.4.2 Digital droplet PCR

For digital droplet PCR (ddPCR), total RNA was prepared from 10^6 mESCs using the RNeasy mini plus kit, including gDNA eliminator columns (QIAGEN). Purified RNA was eluted in 30 µl elution buffer, 7 µl was diluted with 8 µl water, and 4 µl of this dilution was used in the reverse transcription reaction (see Section 2.4.1). The generated

cDNA was diluted with 300 ul nuclease-free water.

ddPCR primers were designed using Primer 3 Plus with the following settings, as recommended by BioRad: 3.8 mM divalent cations, 0.8 mM dNTPs, 80-120 bp product range, 60-61°C melting temperature, using the SantaLucia 1998 salt correction formula and rodent mispriming repeat library. Their efficiency was tested using a serial dilution curve of cDNA by standard SYBR green qPCR.

RPT low retention tips (Starlab) were used throughout the ddPCR protocol, to increase pipetting accuracy. ddPCR reactions were prepared in 96-well PCR plates, and contained 12.5 µl 2x QX200 ddPCR EvaGreen Supermix (BioRad), 0.32 µl each of forward and reverse primers (10 µM), 7 µl diluted cDNA and 4.86 µl nuclease-free water. This 25 µl reaction was pipetted up and down to ensure even mixing, then 22 µl was transferred to a semi-skirted 96-well PCR plate (Eppendorf) and used for droplet generation with an AutoDG droplet generator (BioRad). Droplets were collected in a semi-skirted PCR plate, which was then sealed using a PX1 PCR plate sealer (BioRad). PCR was performed using a C1000 Touch thermal cycler (BioRad) with a 2°C/s ramp rate: 5 min at 95°C for enzyme activation, followed by 40 cycles of denaturation at 95°C for 30 s and annealing/extension at 60°C for 60 s, then signal stabilisation at 4°C for 5 min followed by 90°C for 5 min. Droplets were sorted into PCR-positive and PCR-negative fractions according to their fluorescence as a result of DNA-binding of the EvaGreen dye, using the QX200 droplet reader (BioRad). QuantaLife software (BioRad) was used to calculate the absolute concentration of template cDNA in the ddPCR reaction.

2.4.3 Isolation of nuclei

Nuclei were isolated from $10x10^6$ mESCs mixed with $2.5x10^6$ *Drosophila* SG4 cells, enabling calibrated nucRNA-seq and ATAC-seq analysis (see Section 2.7.1 below). Cells were lysed in HS lysis buffer (0.05% NP40, 50 mM KCl, 10 mM $MgSO_4.7H_20$, 5 mM HEPES, 1 mM PMSF, 3 mM DTT, 1xPIC), and nuclei pelleted by centrifugation at 1000xg for 5min. The pellet was washed twice in 1 ml ice-cold RSB buffer (10mM NaCl, 10mM Tris pH 7.4, 3mM MgCl2) then resuspended in RSB. A 10 µl aliquot was mixed with 10 µl trypan blue and transferred to a haemocytometer for counting. Nuclei were aliquoted

for preparation of nuclear RNA (nucRNA, $4x10^6$), genomic DNA ($1x10^6$) and assay for transposase accesible chromatin (ATAC, $5x10^5$), and pelleted by centrifugation at 1500xg for 5 min.

2.4.4 Preparation of nuclear RNA for sequencing

nucRNA was prepared from $4x10^6$ nuclei resuspended in 1 ml TRIzol reagent, according to the manufacturer's protocol (Invitrogen) and using Phase Lock tubes (Quantbio). Purified nucRNA was resuspended in 100 µl water and treated with the TURBO DNA-free kit (Thermo Scientific), using 6 U TURBO DNase. nucRNA (700 ng) was depleted of rRNA using the NEBNext rRNA depletion kit (NEB), then checked for efficient rRNA depletion on an RNA Pico Bioanlayser chip (Agilent) and quantified using the Qubit RNA HS assay kit (Life technologies). nucRNA-seq libraries were prepared from an equal amount of nucRNA using the NEBnext Ultra II directional RNA library prep (NEB), including 4.5 min fragmentation at 94°C. DNA LoBind tubes (Eppendorf) and RPT low retention tips (Starlab) were used during the preparation of all sequencing libraries.

2.5 Chromatin accessibility

2.5.1 Assay for transposase accessible chromatin (ATAC)

Chromatin accessibility was assayed by ATAC, using an adapted version of the published protocol (Buenrostro et al. 2013, adaptations made by Hamish King). Nuclei ($5x10^5$) were resuspended in Tn5 reaction buffer (10 mM TAPS, 5 mM MgCl2, 10% dimethylformamide and 2 µl Tn5 transposase (25 µM, purified by Hamish King)) and incubated for 30 min at 37°C. Tagmented DNA was purified using MinElute columns and eluted in 20 µl elution buffer (QIAGEN). Genomic DNA (50 ng) purified from the same preparation of nuclei was tagmented with Tn5 as above, to control for sequence bias of the Tn5 transposase and to determine the exact mouse/fly mixing ratio (see Section 2.7.1 below).

2.5.2 ATAC-seq library preparation

ATAC-seq libraries were prepared from 20 µl eluted ATAC sample or 10 µl input sample, by PCR amplification (8 cycles) using custom Illumina barcodes (Buenrostro et al. 2013). The NEBNext High-Fidelity 2X PCR master mix (NEB) was used, according to the manufacturer's protocol. Libraries were purified with two rounds of AMPure XP bead cleanup (Agencourt, 1.5x bead : sample ratio).

2.6 Massively parallel sequencing and data processing

2.6.1 Massively parallel sequencing

All ChIP-seq experiments were carried out in biological triplicate. Sequencing libraries were profiled on a DNA HS Bioanalyzer chip (Agilent) and quantified by qPCR alongside dsDNA standards (KAPA Biosystems). Libraries were sequenced using a NextSeq500 (Illumina) with 80 bp paired-end reads, with the help of Amanda Williams (Department of Zoology, University of Oxford).

2.6.2 Read alignment

The alignment pipelines used in this study were developed by Hamish King. For histone ChIP-seq and ATAC-seq (including sequencing of input gDNA), reads were aligned to concatenated mouse mm10 and *Drosophila melanogaster* dm6 genomes using Bowtie 2 (Langmead and Salzberg 2012) with the '- - no-mixed' and '- - no-discordant' options, before separation into mm10- and dm6-specific reads. Reads that were non-uniquely mapped were discarded. For ATAC-seq, reads that mapped to a custom 'blacklist' of genomic regions with artificially high counts, including mitochondrial DNA sequences, were also discarded. For nucRNA-seq, reads that aligned to the rRNA genomic sequence (GenBank: BK000964.3) were first identified using Bowtie 2 and discarded, before alignment to the concatenated mm10+dm6 genome using STAR (Dobin et al. 2013). To improve mapping of intronic sequences, reads that failed to map using STAR were aligned using Bowtie 2. PCR duplicates were removed using SAMtools (Li et al. 2009).

2.6.3 Peak calling

Peak calling was performed for ATAC-seq experiments, using the DANPOS2 dpeak function with matched inputs (Chen et al. 2013). A 1e-10 p-value cutoff was chosen for *Drosophila* samples, whereas 1e-30 was used for mouse. A total peak set, merging the peaks found in each separate replicate, was generated using BedTools (Quinlan and Hall 2010).

2.7 Normalisation, visualisation and analysis of sequencing data

2.7.1 Calibration and differential analysis

ATAC-seq and nucRNA-seq datasets were calibrated to the spike-in *Drosophila* genome using a custom pipeline. This enabled the detection of widespread, directional changes in accessibility/transcription (see Section 4.2.4 below for further details). A GFF file containing either a non-redundant mm10 gene set (King et al. 2018, for nucRNA-seq) or the total set of ATAC peaks (Section 2.6.3, for ATAC-seq) was annotated with read counts, using the raw mm10 BAM files generated by read alignment (Section 2.6.2). In contrast, individual dm6 BAM files were first pre-normalised to the relative mm10/dm6 ratio in the matched input gDNA library, to control for variation in the exact SG4/mESC cell mixing ratio between biological replicates. These normalised BAM files were then used to annotate a GFF file, containing a dm6 gene or ATAC peak set, with read counts. The annotated dm6 GFF file was used to calculate normalisation Size Factors using the DESeq2 package (Love et al. 2014). These dm6 Size Factors were then applied to DESeq2 analysis of the mm10 GFF file. Adjusted p-value <0.05 and fold change>1.4 thresholds were used to define significant changes.

2.7.2 Histone ChIP-seq normalisation

Histone ChIP-seq libraries were randomly downsampled to achieve the same total read number, using SAMtools. This method was also used for normalisation of nucRNA-seq

libraries to total library size. Bigwig tracks of H3K36me2 normalised to H3, and of the signal in tamoxifen-treated versus -untreated cells, were generated using the Deeptools bigwigCompare function with '- - operation ratio' setting (Ramirez et al. 2016).

2.7.3 Data visualisation

Quality controls including visual inspection of the data and genome-wide correlations were used to assess the reproducibility of biological replicates. High quality replicates were merged to create a representative bigwig track using DANPOS2 for ATAC-seq and the MACS2 pileup function for histone ChIP-seq and nucRNA-seq. Genome tracks were visualised using the UCSC genome browser (Kent et al. 2002; Casper et al. 2018). Heatmaps and metaplots were generated using the Deeptools plotHeatmap and plotProfile functions, respectively (Ramirez et al. 2016).

2.7.4 Functional annotation

Polycomb target genes were defined according to high levels of RING1B and SUZ12 binding at their TSS (analysis performed by Hamish King). For intron- and exon-specific nucRNA-seq analysis, paired reads were assigned to mm10 RefSeq gene introns, exon-exon junctions, or gene bodies by Angelika Feldmann, using a custom pipeline.

2.7.5 Datasets used in this study

A number of published mESC sequencing datasets were used in this study. These are listed in Table 2.3. Two unpublished calibrated nucRNA-seq datasets were also analysed. These were generated by Neil Blackledge and Nadya Fursova, from conditional mESC lines with tamoxifen-inducible RING1A/B or PCGF1 knockout respectively. The conditional *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ line was generated by Neil Blackledge (unpublished), and the conditional *Pcgf1*$^{fl/fl}$ line was generated in the lab of Prof Haruhiko Koseki.

Study	Accession code	Dataset(s)
Forrest et al. 2014	DRA000991	CAGE
Brown et al. 2017	GSE93538	H3K4me3 ChIP-seq
Brookes et al. 2012	GSE34520	H3K36me3 and RNAPII ChIP-seq
Long et al. 2013	GSE43512	BioCAP
King et al. 2018	GSE87822	BRG1 ChIP-seq, RNA-seq following BRG1 knockout
Farcas et al. 2012	GSE41267	KDM2A and KDM2B ChIP-seq
Rose et al. 2016	GSE83135	RING1B and SUZ12 ChIP-seq
Jonkers and Lis 2015	GSE48895	RNA-seq following triptolide treatment

Table 2.3: Published datasets used in this study

2.7.6 Correlation and significance testing

R was used to perform regression analyses using standard linear models, and for Spearman and Pearson correlation analyses. Empirical p-values were calculated using a custom script written by Angelika Feldmann. This method of significance testing was used to assess whether the distribution of a subset of genes into different gene classes was significantly different from the distribution of the total gene set into these classes. A random group of genes of the same size as the subset being tested was selected from the total gene set, and their distribution into the gene classes recorded. This was repeated 1000 times, allowing the calculation of the probability that the distribution of the test subset would be obtained by chance.

2.7.7 Gene ontology analysis

Gene Ontology analysis was performed using DAVID (Huang da et al. 2009b; Huang da et al. 2009b). The BP FAT setting and a FDR<0.1 cutoff were used, and the complete non-redundant mm10 gene set was used as a background.

3

Examining the role of KDM2A/B histone demethylase activity

3.1 Introduction

KDM2A/B are H3K36 demethylases, with specificity for the mono- and di-methyl state (Tsukada et al. 2006; He et al. 2008). They are targeted via their ZF-CXXC domain to CGI throughout the genome (Blackledge et al. 2010; Farcas et al. 2012; He et al. 2013; Wu et al. 2013), which are found at approximately 70% of gene promoters in mammals (Tykocinski and Max 1984; Bird et al. 1985; Saxonov et al. 2006). Here KDM2A/B have been proposed to catalyse demethylation of H3K36me2 via their JmjC catalytic domain (Blackledge et al. 2012).

Both H3K36me2 and H3K36me3 are enriched throughout the bodies of genes in yeast. This is due to the interaction between Set2, the sole H3K36 methyltransferase, and elongating RNAPII (Bannister et al. 2005; Rao et al. 2005; Pokholok et al. 2005). Indeed, H3K36me2 and H3K36me3 initiate at a fixed distance from the TSS, consistent with their deposition during transcriptional elongation. In contrast, the geographies of H3K36me2 and H3K36me3 differ markedly in vertebrates, due to an expanded system of H3K36 methyltransferases. There is a division of labour between a variety of mono- and dimethyltransferases and SETD2, which is thought to mediate virtually all H3K36 trimethylation (Bell et al. 2007; Edmunds et al. 2008). SETD2 plays an evolutionarily conserved role in vertebrates in depositing H3K36me3 throughout the coding region of actively transcribed genes, through its interaction with the hyperphosphorylated C-terminal tail of RNAPII (Schaft et al. 2003; Li et al. 2003). In contrast, deposition of H3K36me2 is not restricted to the activity of RNAPII-associated SETD2, but is instead mediated by numerous methyltransferases including NSD1-3 and ASH1L. H3K36me2

is an abundant histone modification in mammals, present on between 40 and 70% of all histone H3 (Peters et al. 2003; Robin et al. 2007; Schotta et al. 2008; Jung et al. 2010), and NSD1 and NSD2 appear to play a widespread role in depositing H3K36me2 throughout the genome (Kuo et al. 2011; Streubel et al. 2018).

H3K36me2 indirectly regulates transcription, through its recognition by a variety of specific reader domains such as the PWWP domain. These domains are found in a number of proteins that are part of chromatin modifying complexes, leading to their recruitment to regions of H3K36me2 throughout the genome. The chromatin modifying activities of these recruited complexes have been reported to regulate a number of stages of gene expression (see Section 1.1.7). The contribution of H3K36me2 to transcription regulation is best characterised in *S. cerevisiae*, where co-transcriptional H3K36 methylation by Set2 is required to suppress transcription initiation (Venkatesh et al. 2012). This link between H3K36 methylation, which is specifically deposited in gene bodies in yeast, and the regulation of transcription initiation, which instead occurs at gene promoter regions, may seem paradoxical. However, H3K36 methylation specifically suppresses transcription initiation from cryptic promoters found within gene body regions, where this mark is enriched. Both H3K36me2 and H3K36me3 are able to recruit the Eaf3 subunit of the histone deacetylase Rpd3s (Youdell et al. 2008; Li et al. 2009), which erases transcription elongation-associated acetylation and is required for suppression of cryptic transcription (Carrozza et al. 2005; Keogh et al. 2005). Recognition of H3K36 methylation by a number of other chromatin modifying complexes has also been linked to the suppression of cryptic initiation from within open reading frames (Smolle et al. 2012; Venkatesh et al. 2012; Venkatesh and Workman 2015).

A number of studies suggest that H3K36 methylation may play a conserved role in repressing cryptic initiation in mammalian cells. The Eaf3 orthologue MRG15 can form a mammalian Rpd3s-like HDAC complex, and is reported to bind H3K36 methylation via its PWWP domain. However, the repression of cryptic initiation by this complex may be mediated via H3K4 demethylation, rather than histone deacetylation as in yeast (Xie et al. 2011). In addition, H3K36 methylation was reported to repress cryptic initiation in mammalian cells through recruitment of the FACT histone chaperone, which coordinates the reassembly of nucleosomes following their transcription-coupled

displacement (Carvalho et al. 2013).

This link between H3K36 methylation and the suppression of transcription initiation led to the proposal that KDM2A/B may contribute to the generation of a promoter-associated transcriptionally permissive chromatin state, by locally removing H3K36me2 from CGI promoters (Blackledge et al. 2010; Blackledge and Klose 2011; Deaton and Bird 2011). However, removal of H3K36me2 by KDM2A/B has also been implicated in transcription repression in a number of studies, including the repression of genes that play critical roles in cell proliferation, differentiation and senescence (Frescas et al. 2007; He et al. 2008; Tzatsos et al. 2009; Fukuda et al. 2011; Du et al. 2013; Dong et al. 2013; Gao et al. 2013; Yu et al. 2016). Few of these studies provide direct evidence that the demethylase activity of KDM2A/B contributes to transcriptional repression, and these examples are limited to only a handful of genes, therefore the contribution of KDM2A/B demethylase activity to transcriptional repression is unknown.

The genome-wide role that KDM2A/B play in regulating H3K36me2 at CGI promoters and the impact that this has on gene expression therefore remain unclear. KDM2A and KDM2B are highly homologous proteins, both of which act as H3K36me2 demethylases and are targeted to CGI throughout the genome. Therefore, it is likely that there is functional redundancy between these paralogues. I aimed to directly examine the combined role of KDM2A and KDM2B in regulating H3K36me2 throughout the genome, and to determine the contribution of KDM2A/B demethylase activity to the regulation of CGI-associated transcription and chromatin accessibility.

3.2 Results

3.2.1 Conditional knockout of KDM2A/B-LF

In order to explore the role played by KDM2A/B demethylase activity I exploited a conditional mESC line developed by Prof Haruhiko Koseki, in which KDM2A/B-LF can be conditionally deleted while leaving the expression of the non-catalytic short isoforms intact. In this line, loxP recombination sites have been inserted within the *Kdm2a* and *Kdm2b* genes, surrounding exon(s) encoding a key portion of the JmjC demethylase

domain (*Kdm2a/b-JmjC*$^{fl/fl}$, Figure 3.1 A). The line also encodes Cre recombinase fused to the oestrogen receptor (Cre-ERT2). Addition of tamoxifen induces translocation of Cre-ERT2 into the nucleus, leading to Cre-mediated recombination of the loxP sites and exon excision. Deletion of the exon(s) leads to the generation of an mRNA that has a frameshift, resulting in a premature stop codon that effectively knocks out KDM2A/B-LF (Figure 3.1 B). Importantly, expression of KDM2A/B-SF should be unaffected, as they are expressed from a downstream promoter.

Reverse transcription of RNA followed by quantitative PCR (RT-qPCR) analysis confirmed that tamoxifen treatment of *Kdm2a/b-JmjC*$^{fl/fl}$ cells for 96 h led to efficient excision of the targeted exons (Figure 3.2 A). This resulted in loss of the KDM2A/B-LF proteins, while the KDM2A/B-SF proteins were unaffected (Figure 3.2 B). Therefore, this cell line enables the conditional knockout of KDM2A/B-LF in isolation, removing the forms of KDM2A/B that can demethylate H3K36.

3.2.2 Regulation of bulk histone modification levels by KDM2A/B-LF

KDM2A/B have specificity for H3K36me1/2. Therefore loss of KDM2A/B-LF would be expected to lead to an increase in these histone modifications at KDM2A/B target sites (Blackledge et al, 2010). However, while the demethylase activity of either KDM2A or KDM2B has been studied at specific target sites, their genome-wide role in regulating H3K36 levels has not been explored. In addition, the combined contribution of KDM2A and KDM2B to the regulation H3K36 methylation has not been investigated to date. To address these important considerations, I began by examining the contribution of KDM2A/B to the overall levels of H3K36 methylation. I prepared histone extracts from *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs before and after tamoxifen treatment to induce KDM2A/B-LF knockout, and performed western blotting for H3K36me1, H3K36me2 and H3K36me3. The levels of these modifications did not significantly change after KDM2A/B-LF knockout (Figure 3.3). There was a slight trend towards an increase in H3K36me2 and H3K36me3 following KDM2A/B-LF knockout, whereas no increase in H3K36me1 was evident. Given that KDM2A/B do not exhibit H3K36me3 demethylase activity (Tsukada et al. 2006; He et al. 2008), it is possible that the small increase in H3K36me3 resulted

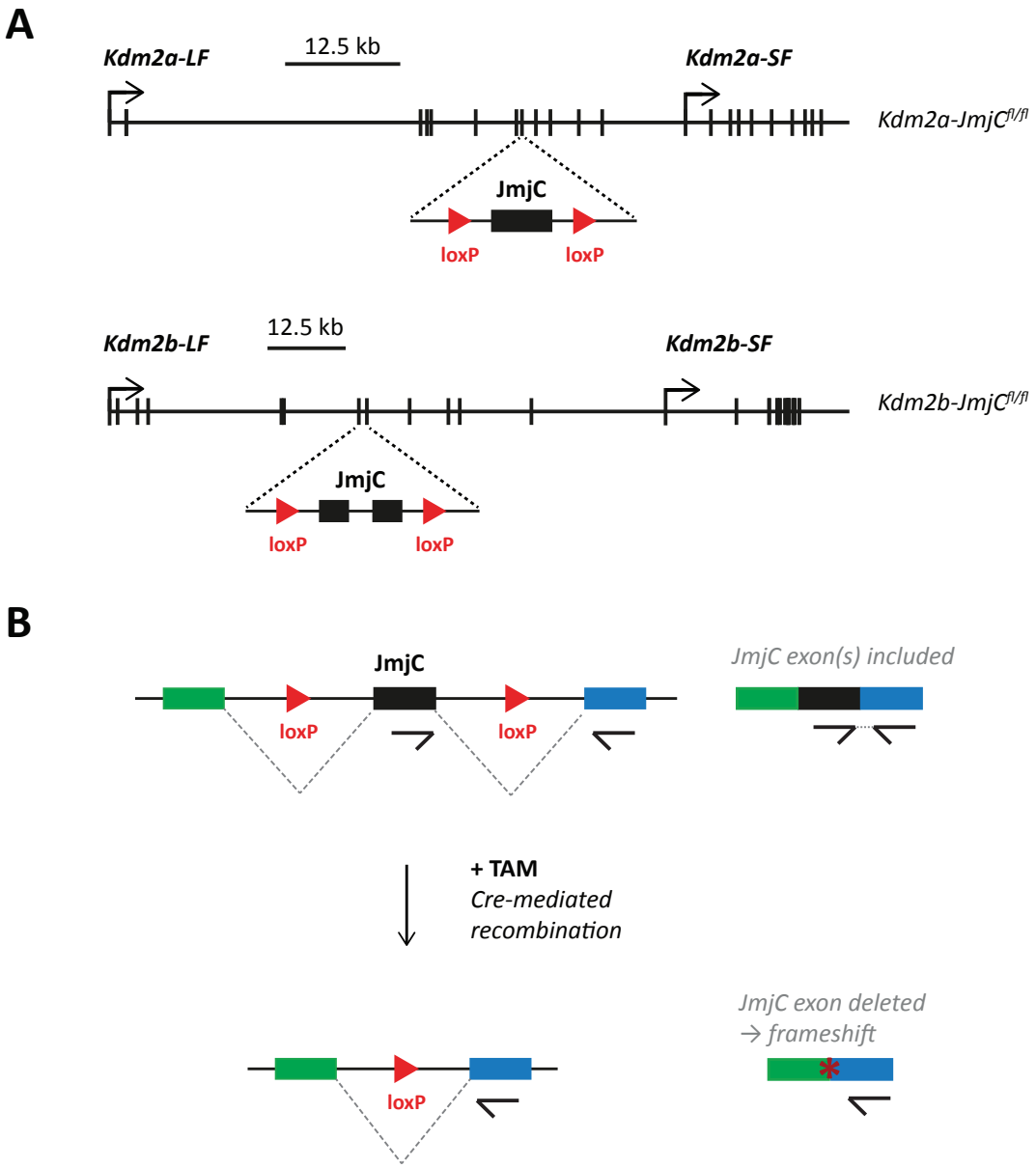


Figure 3.1: Cre/loxP system for conditional knockout of KDM2A/B-LF
(A) Schematic depicting the position of loxP sites flanking exon(s) encoding the JmjC demethylase domain for both KDM2A and KDM2B, enabling conditional knockout of KDM2A/B-LF. Expression of KDM2A/B-SF occurs from a downstream promoter and therefore should be unaffected.
(B) Addition of tamoxifen (TAM) leads to Cre-mediated recombination of the loxP sites, causing deletion of the JmjC-encoding exon(s). This leads to a frameshift mutation, resulting in loss of KDM2A/B-LF expression. RT-qPCR primers are depicted as arrows.

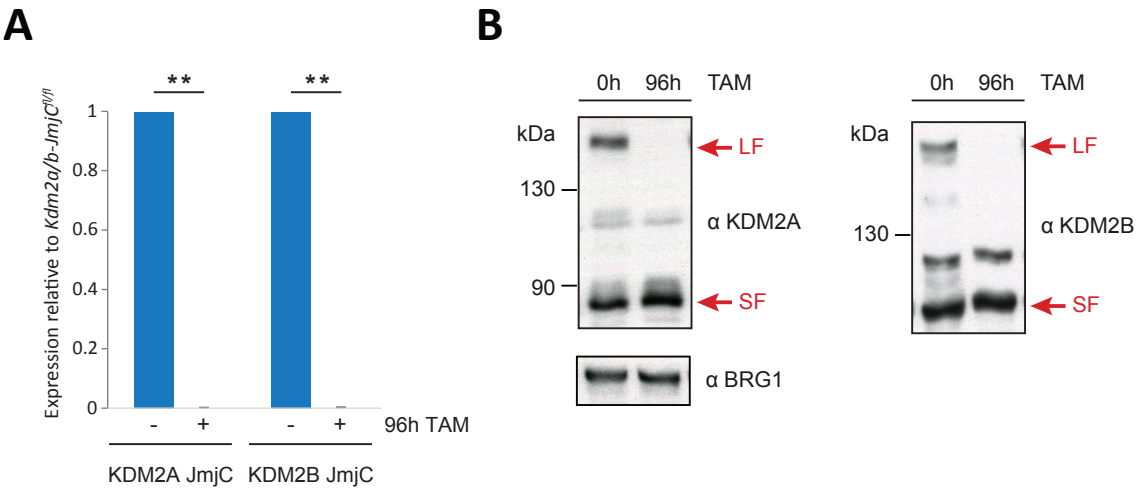


Figure 3.2: Efficient KDM2A/B-LF knockout in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs
(A) RT-qPCR analysis in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs before and after 96 hour tamoxifen (TAM) treatment to induce excision of exon(s) encoding the JmjC domain, using primers designed to detect splicing of the targeted exon (for primer positions see Figure 3.1). Expression was normalised to *b-actin* and *Hprt*, and is represented relative to *Kdm2a/b-JmjC*$^{fl/fl}$. Error bars for the tamoxifen treated sample show standard deviation of four biological replicates, asterisks represent significant changes (Student's T-test, ** represents $p<0.001$)
(B) Western blot showing loss of KDM2A/B-LF but not KDM2A/B-SF following 96h TAM treatment of *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs. BRG1 is a loading control for both KDM2A and KDM2B westerns.

from conversion of H3K36me2 to H3K36me3.

To further explore the effect of KDM2A/B-LF knockout, I examined the bulk levels of a number of other histone modifications that are associated with CGI gene promoters. There was also no significant change in the levels of these histone modifications following KDM2A/B-LF knockout (Figure 3.4). However, there was a slight trend towards increased H3K4me3, a hallmark of CGI promoters (Mikkelsen et al. 2007; Guenther et al. 2007; Thomson et al. 2010), and H3K27me3 and possibly H2AK119ub1, which are enriched at polycomb-repressed CGI promoters. In contrast, there appeared to be a subtle decrease in histone H3K9 acetylation (H3K9ac) and H3K27ac, which are associated with active promoters and enhancers, although this was not significant. Levels of H3K4me1, which is associated with enhancers, remained unchanged.

3.2.3 Optimisation of H3K36me2 ChIP

Having established that KDM2A/B-LF make a minimal contribution to the regulation of bulk H3K36me2 levels (Figure 3.3), I next wanted to explore the specific role that

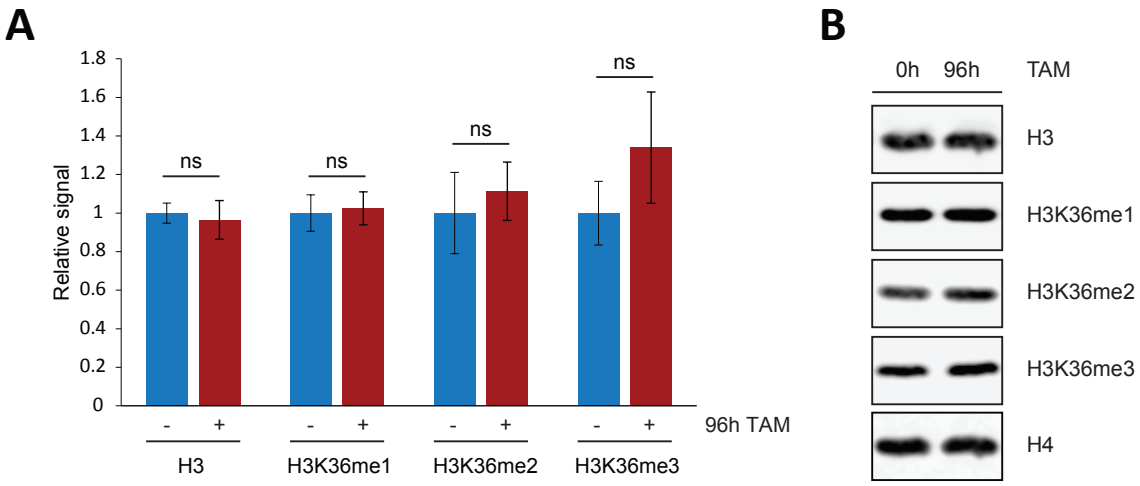


Figure 3.3: KDM2A/B-LF knockout has a minor effect on global H3K36 methylation levels
(A) Quantitation of western blots of histone extract from *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs, before and after 96 h tamoxifen (TAM) treatment to induce KDM2A/B-LF knockout. Signal is normalised to histone H4 and is represented relative to average *Kdm2a/b-JmjC*$^{fl/fl}$ signal. Error bars show standard deviation of three biological replicates. Significance was tested using a Student's T-test (non-significant (ns) if p>0.05).
(B) Representative western blots. Histone H4 is a loading control.

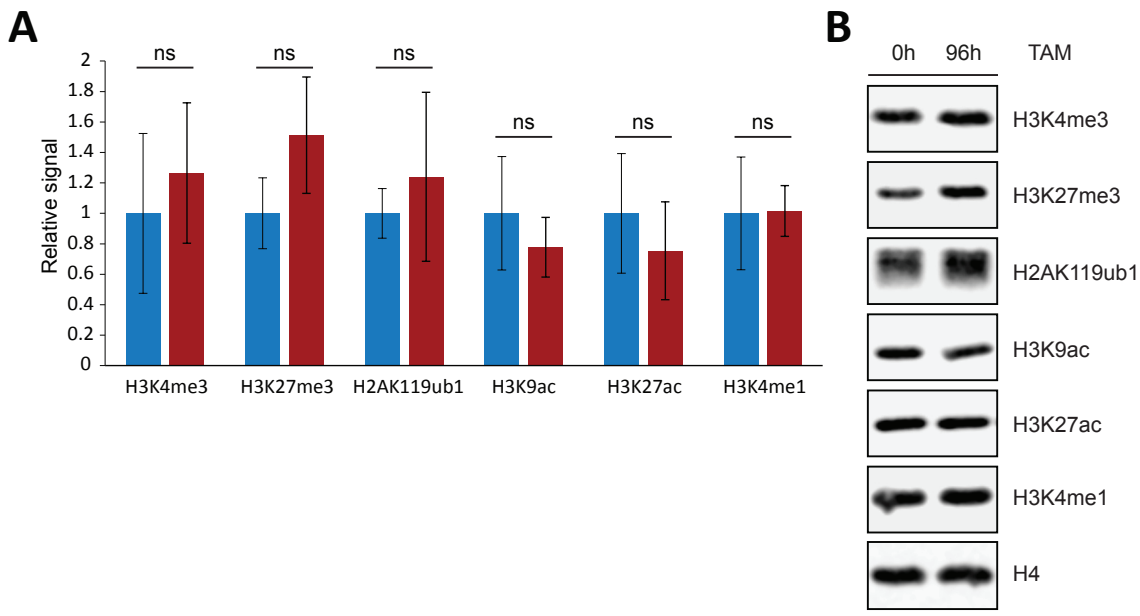


Figure 3.4: KDM2A/B-LF knockout has a minor effect on the levels of other CGI-associated histone modifications
(A) Quantitation of western blots of histone extract from *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs, before and after 96 h tamoxifen (TAM) treatment to induce KDM2A/B-LF knockout. Signal is normalised to histone H4 and is represented relative to *Kdm2a/b-JmjC*$^{fl/fl}$. Error bars show standard deviation of three biological replicates. Significance was tested using a Student's T-test (non-significant (ns) if p>0.05).
(B) Representative western blots. Histone H4 is a loading control.

KDM2A/B play in regulating H3K36me2 at CGI gene promoters. I therefore used ChIP to study the genome-wide distribution of H3K36me2 before and after KDM2A/B-LF knockout. Native ChIP can be used to study the distribution of histone modifications, as histones are naturally linked to the DNA they package and therefore crosslinking is not required to capture this interaction. Native ChIP benefits from high resolution, as MNase is used to digest the chromatin to predominantly mononucleosomal fragments, and high efficiency, as the epitope recognised by the antibody is not masked by crosslinks. However, unlike during fixed ChIP protocols, chromatin is isolated without the aid of detergents, and therefore certain epitopes can remain masked and inaccessible to antibody. A trial experiment demonstrated that native ChIP was able to capture an enrichment of H3K36me2 in the gene body relative to the gene promoter, however the recovery relative to the H3K4me3 control ChIP was extremely low (Figure 3.5 A, compare percentage input values). This suggests that the H3K36me2 epitope, unlike H3K4me3, remained masked under native conditions.

I tried using a semi-native protocol to improve the recovery of H3K36me2 ChIP. This combines low formaldehyde fixation (0.37%) with MNase digestion, and includes the use of detergent during chromatin isolation. However, my attempts gave inconsistent recovery of chromatin. Therefore, I instead used a previously described H3K36me2 ChIP protocol, which combines 1% formaldehyde fixation with sonication and again uses detergent to aid chromatin isolation (Blackledge et al. 2010). This gave recoveries that were now comparable to the H3K4me3 control ChIP, which works well under both conditions. This method again confirmed a depletion of H3K36me2 at gene promoters relative to the gene body (Figure 3.5 B).

3.2.4 KDM2A/B knockout leads to a small increase in H3K36me2 at CGI gene promoters

Having established a robust H3K36me2 ChIP protocol, I next wanted to use this method to investigate the contribution of KDM2A/B to H3K36me2 levels at CGI promoters throughout the genome. Therefore, I used ChIP followed by massively parallel sequencing (ChIP-seq) to profile H3K36me2 genome-wide in *Kdm2a/b-JmjC*$^{fl/fl}$

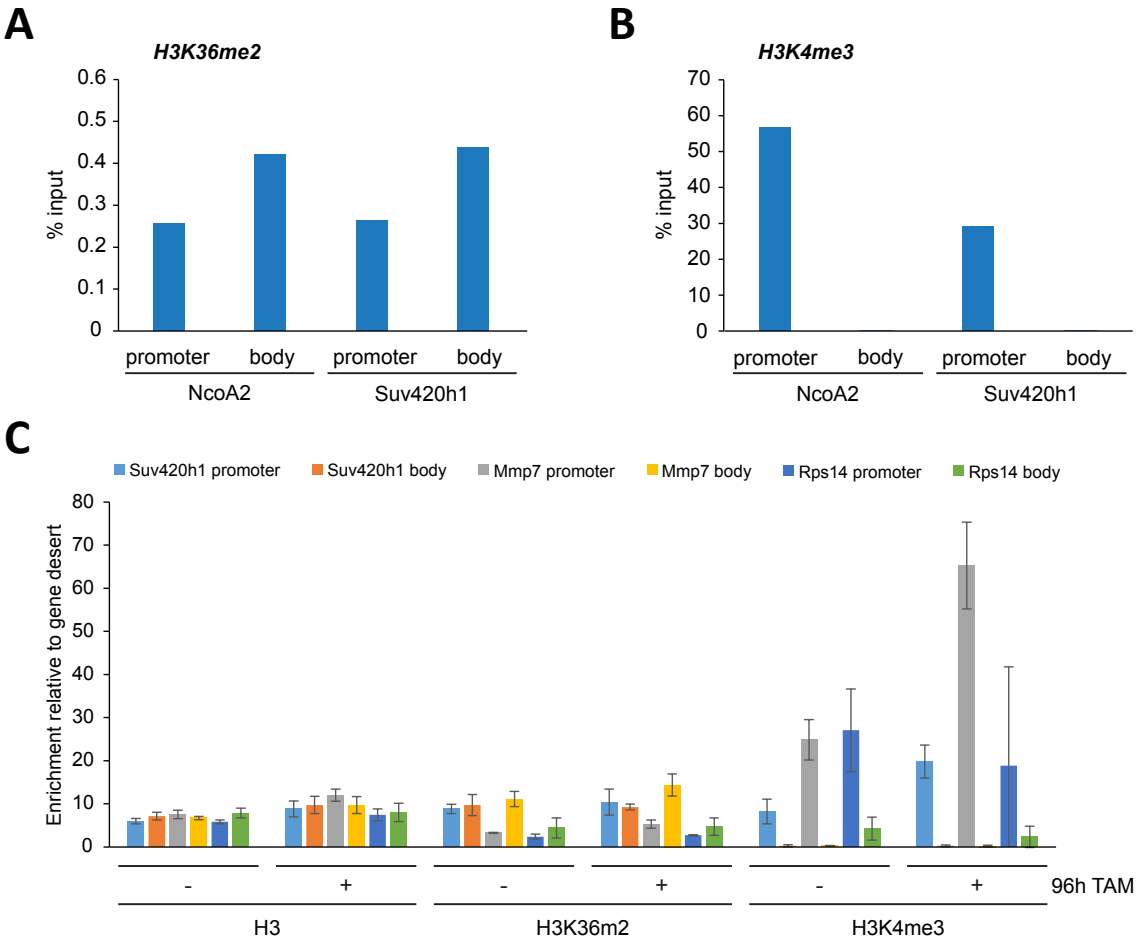


Figure 3.5: Optimisation of H3K36me2 ChIP
(A) Native ChIP-qPCR at CGI gene promoters and the corresponding gene body, showing low recovery of H3K36me2 ChIP.
(B) As (A) for H3K4me3 ChIP, showing high recovery at CGI gene promoters.
(C) Fixed ChIP-qPCR at a panel of CGI gene promoters and the corresponding gene body, normalised to a genomic DNA control region. H3, H3K36me2 and H3K4me3 ChIPs were performed in *KDM2A/B*$^{fl/fl}$ mESCs before and after 96 h tamoxifen (TAM) treatment to induce KDM2A/B-LF knockout. Experiments were performed in triplicate, error bars show standard deviation.

mESCs, before and after 96 h tamoxifen treatment to induce KDM2A/B-LF knockout. Importantly, I also performed H3 ChIP-seq and normalised the H3K36me2 signal to the local H3 signal as part of all analyses. This allowed me to control for decreases in histone occupancy, which would otherwise be indistinguishable from decreases in histone methylation. I examined H3K36me2 levels at gene promoters, which showed a pronounced dip in H3K36me2 levels at the TSS as previously described at individual CGI (Blackledge et al. 2010, Figure 3.6 A, B). Comparison of the ChIP-seq signal in tamoxifen treated and untreated cells revealed a small increase in H3K36me2 surrounding the TSS following KDM2A/B-LF knockout (Figure 3.6 A, B). This increase was more evident when represented as a ratio of the H3K36me2 ChIP-seq signal from treated and untreated cells (Figure 3.6 A - third heatmap, C). The increase in H3K36me2 levels at TSS is consistent with loss of KDM2A/B H3K36me2 demethylase activity at CGI gene promoters. Interestingly, KDM2A/B-LF knockout results in shrinking of the H3K36me2 depleted region surrounding the TSS, with H3K36me2 increasing at a more 5' position within the gene body and at a more 3' position upstream of the promoter (Figure 3.6 A). This observation likely reflects a failure to remove H3K36me2 from these regions following KDM2A/B knockout, leading to the encroachment of this modification from the surrounding methylated regions.

3.2.5 The gene bodies of highly transcribed genes are depleted of H3K36me2

Given that H3K36me2 was locally depleted around the TSS of genes, I investigated whether the H3K36me2 profile varied with gene transcription level. I separated expressed genes with CGI promoters into quartiles based on their transcription level in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs (see Section 3.2.6), with Q4 being the most highly transcribed, then plotted H3K36me2 across the gene body and the 10 kb up- and downstream. Strikingly, all four quartiles had a pronounced dip in H3K36me2 at the TSS (Figure 3.7 A), which was not apparent for non-CGI genes (Figure 3.7 B). In contrast, the surrounding chromatin was blanketed by H3K36me2, consistent with the high abundance of this histone modification. This confirms that local depletion of H3K36me2 is a specific hallmark of CGI gene promoters, as has been previously proposed (Blackledge et al. 2010; Blackledge and Klose 2011; Deaton and Bird 2011).

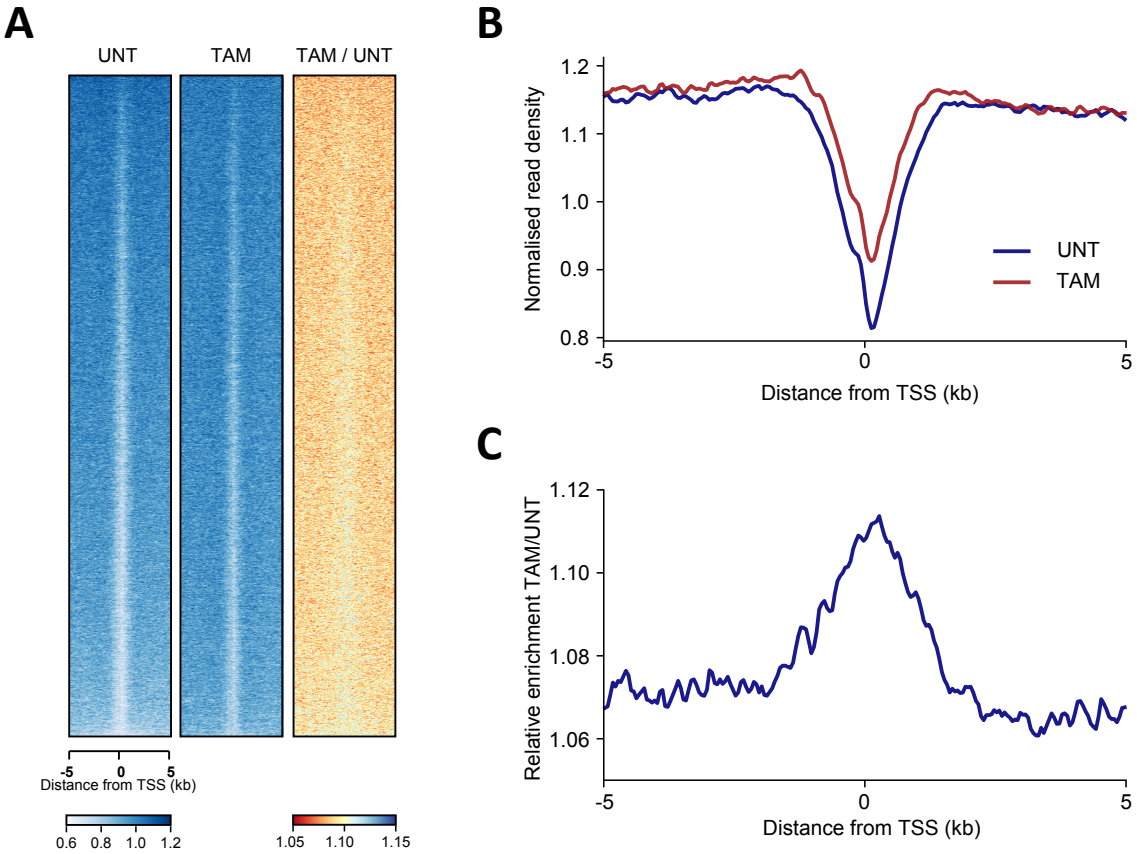


Figure 3.6: KDM2A/B knockout leads to a modest increase in H3K36me2 at TSS
(A) Heatmap of H3K36me2 surrounding the TSS of CGI gene promoters (n=20633), ranked by mean signal. Data for *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs before (UNT) or after tamoxifen (TAM) treatment to induce KDM2A/B-LF knockout is shown, alongside a ratio of the ChIP signal in TAM relative to UNT cells (TAM/UNT). H3K36me2 signal was normalised to H3 signal.
(B) Metaplot of normalised H3K36me2 ChIP-seq enrichment at TSS (n=20633), showing an increase in H3K36me2 following KDM2A/B-LF knockout.
(C) Metaplot of TAM/UNT H3K36me2 ChIP-seq enrichment at TSS (n=20633).

Highly transcribed genes also had relatively low H3K36me2 levels throughout the gene body (Figure 3.7 A, Q3 and Q4), as previously observed for isolated genes by ChIP-qPCR (Blackledge et al. 2010). This in unlikely to be due to KDM2A/B activity, as KDM2A/B do not localise to gene bodies (Blackledge et al. 2010; Farcas et al. 2012; Wu et al. 2013; He et al. 2013). Instead this depletion of H3K36me2 in the coding region of highly transcribed genes could be explained by high rates of conversion of H3K36me2 to H3K36me3, in agreement with previous observations that H3K36me3 is highly enriched within gene bodies and closely correlates with gene expression level (Krogan et al. 2003; Bannister et al. 2005; Pokholok et al. 2005; Mikkelsen et al. 2007; Bell et al. 2007). Indeed, H3K36me3 is highly enriched throughout the gene body of quartiles Q3 and Q4, and the characteristic peak in H3K36me3 towards the 3' end of the gene corresponded with a more pronounced dip in H3K36me2 at this position (Figure 3.7 C). This anti-correlation between H3K36me2 and H3K36me3 was not apparent for lowly transcribed quartiles Q1 and Q2, which are subject to relatively low levels of gene body H3K36me3.

Comparison of the change in TSS H3K36me2 signal following KDM2A/B-LF knockout for highly and lowly transcribed genes revealed that there was a similar increase in H3K36me2 irrespective of transcription level, in contrast to the minimal change that occurs at non-CGI gene promoters (Figure 3.7 D). This is consistent with the generic targeting of KDM2A/B to all CGI promoters via their ZF-CXXC domain, irrespective of the activity of the associated gene, and their widespread demethylation of these regions.

3.2.6 Loss of KDM2A/B-LF results in minimal gene expression changes

The increase in H3K36me2 levels at TSS following KDM2A/B-LF knockout suggested that the demethylase activity of KDM2A/B might play an important role in regulating gene expression. To investigate this possibility, I prepared nucRNA from *Kdm2a/b-JmjC*$^{fl/fl}$ cells before and after 96 h tamoxifen treatment to induce KDM2A/B-LF knockout, followed by library preparation and massively parallel sequencing (nucRNA-seq). Nuclear RNA is enriched for newly synthesised transcripts, and therefore represents a more sensitive measure of transcription changes than bulk mRNA levels. Visual examination of the nucRNA-seq signal revealed that the transcription of most

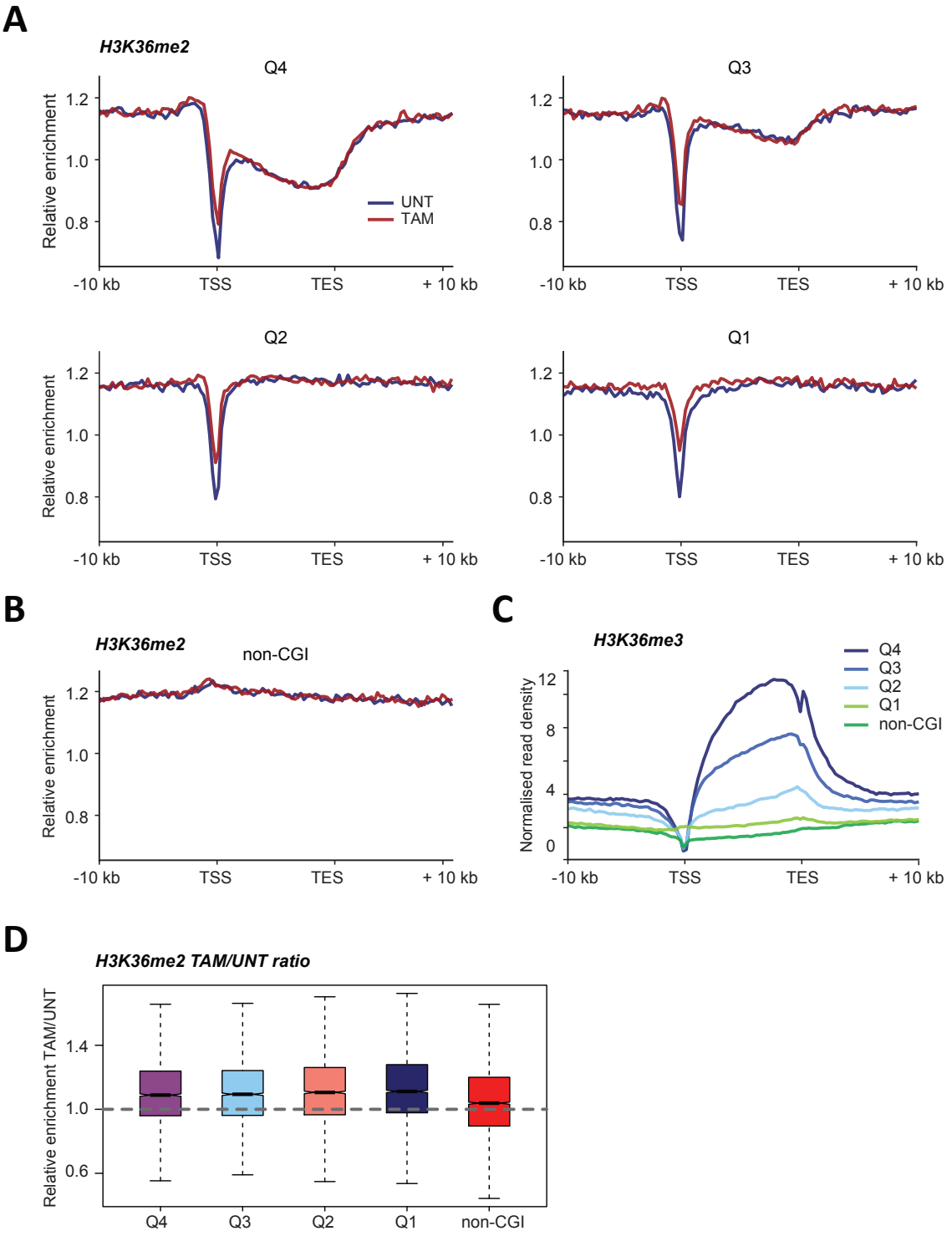


Figure 3.7: H3K336me2 is depleted in the gene body of highly transcribed genes
(A) Metaplots of H3K36me2 ChIP-seq enrichment throughout the gene body. Transcribed genes with CGI promoters were split into quartiles based on transcription level in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs (see Section 3.2.6), with Q4 being the most highly transcribed. H3K36me2 levels before (UNT) and after (TAM) 96 h tamoxifen treatment of *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs to induce KDM2A/B-LF knockout are shown.
(B) As (A), but for non-CGI genes (n=6527).
(C) Metaplot of H3K36me3 ChIP-seq enrichment throughout the gene body (Brookes et al. 2012), for CGI genes separated into transcription quartiles, as (A), and for non-CGI genes.
(D) Bar chart showing the ratio of TAM/UNT H3K36me2 ChIP-seq signal at the 500 bp region surrounding the TSS, for CGI gene promoters (split into quartiles based on expression level, as above) and non-CGI gene promoters.

genes did not change following KDM2A/B-LF knockout (*Mtpn*), while others either decreased (*Igsf8*) or increased (*Btg2*) in expression (Figure 3.8 A). A genome-wide analysis showed that the transcription of the vast majority of genes did not significantly change, while 114 genes significantly increased in transcription and 55 genes significantly decreased in transcription (p <0.05 and fold change>1.4, Figure 3.8 B).

I proceeded to explore the characteristics of these significantly changing genes. The

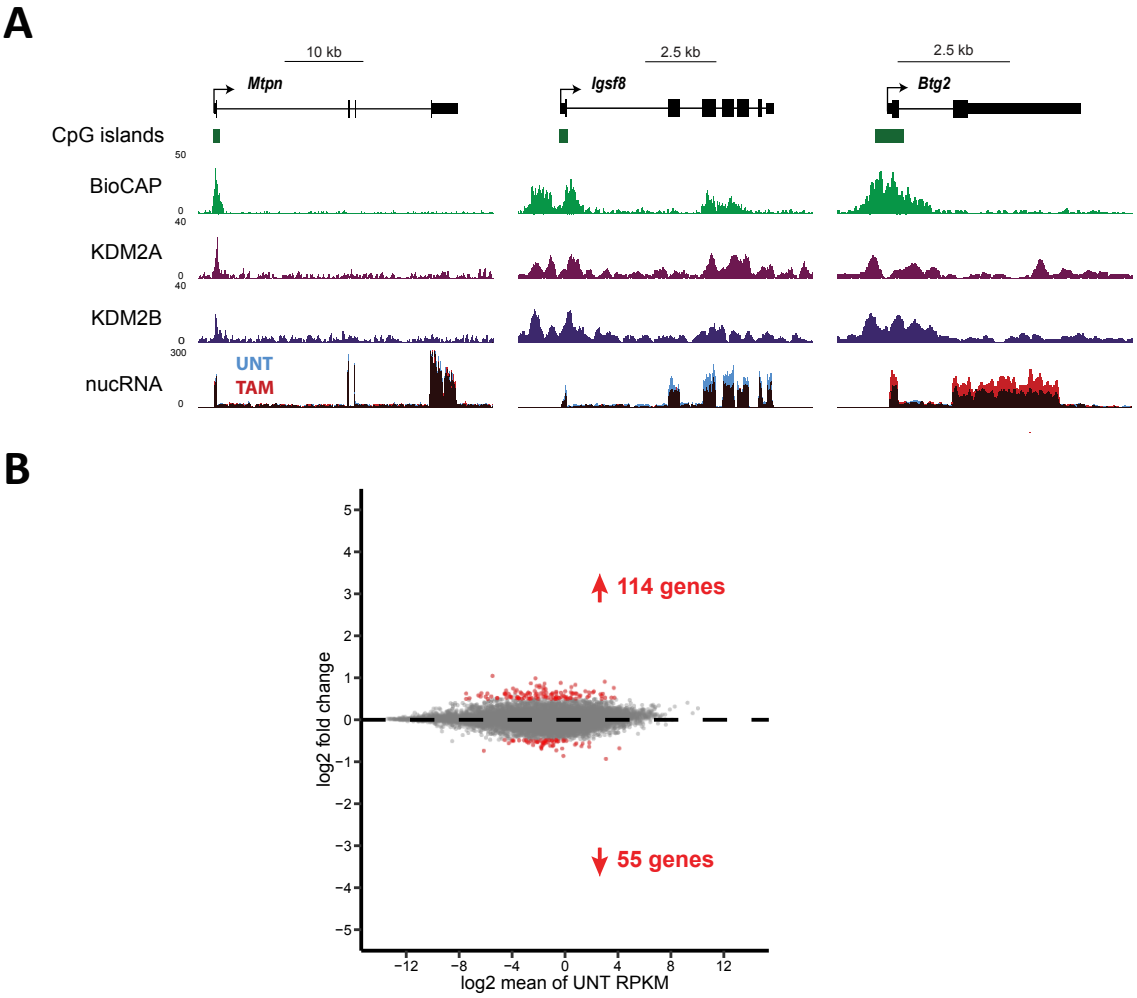


Figure 3.8: KDM2A/B-LF knockout leads to transcriptional changes of a small subset of genes
(A) Representative genome snapshots of genes with unchanged (*Mtpn*), decreased (*Igsf8*) or increased (*Btg2*) transcription following KDM2A/B-LF knockout. The CGI promoters of these genes are enriched for BioCAP signal, an experimental measure of non-methylated CpG-rich DNA, and for KDM2A and KDM2B (Datasets: Long et al. 2013; Farcas et al. 2012). Strand-specific nucRNA-seq signal in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs is depicted as an overlay plot, with the relevant strand shown. Signal in untreated cells is coloured blue and signal from tamoxifen treated cells is coloured red.
(B) Scatter plot of log2 mean nucRNA-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs for a non-redundant mm10 gene set (n=20,633, three biological replicates), plotted against log2 fold change in nucRNA-seq signal following addition of tamoxifen to induce KDM2A/B-LF knockout. Libraries were normalised to total library size. Significant changes were identified using DESeq2 and are coloured red (p<0.05 and fold change>1.4). The dotted line corresponds to no change in transcription.

subsets of genes that significantly increased or decreased in expression following KDM2A/B-LF knockout had relatively low average expression levels in wild-type cells, although this difference was subtle (Figure 3.9 A). The magnitude of the transcriptional changes following KDM2A/B-LF knockout were relatively small, with a median 1.48-fold increase and 0.69-fold decrease in transcription for the significantly increasing and decreasing subsets respectively. Therefore, KDM2A/B-LF knockout had a relatively small effect on the transcription of a small subset of genes. I compared the levels of a variety of chromatin features at the TSS of these gene sets, using datasets from wild-type cells. The average enrichment of Biotinylated CxxC affinity purification (BioCAP) signal, an experimental measure of non-methylated CpG-rich DNA, and of KDM2A ChIP-seq signal at the TSS of these subsets was similar to that of the entire gene set, whereas their small reduction in TSS H3K4me3 can be explained by their relatively low level of expression (Figure 3.9 B). Given that KDM2B is known to be part of the PRC1.1 complex, I next asked whether these deregulated genes were targeted by the polycomb system. The TSS of both significantly increasing and decreasing sets of genes were enriched for KDM2B and the polycomb proteins RING1B and SUZ12, characteristic of polycomb targets. Indeed, the significantly changing gene sets were significantly enriched for polycomb target genes defined by high RING1B and SUZ12 occupancy, and depleted of CGI-associated genes that are not polycomb targets (Figure 3.9 C).

In order to further explore the identity of these significantly changing genes, I performed gene ontology analysis (Huang da et al. 2009b; Huang da et al. 2009a). This revealed that the genes that significantly increased in expression following KDM2A/B-LF knockout were enriched for a variety of developmental terms typical of polycomb target genes, along with terms relating to cell migration and apoptosis (Figure 3.10). This enrichment of developmental genes, which are targeted by polycomb complexes in mESCs, is in keeping with the deregulation of a small subset of polycomb target genes as a result of KDM2A/B-LF knockout. These genes may be particularly sensitive to a reduction in KDM2B-mediated PRC1.1 targeting to CGI gene promoters following knockout of KDM2B-LF, while the majority of developmental genes remain appropriately repressed. In contrast, the significantly downregulated genes were not enriched for any ontology terms, which may be a result of the very small number of genes in this subset. It

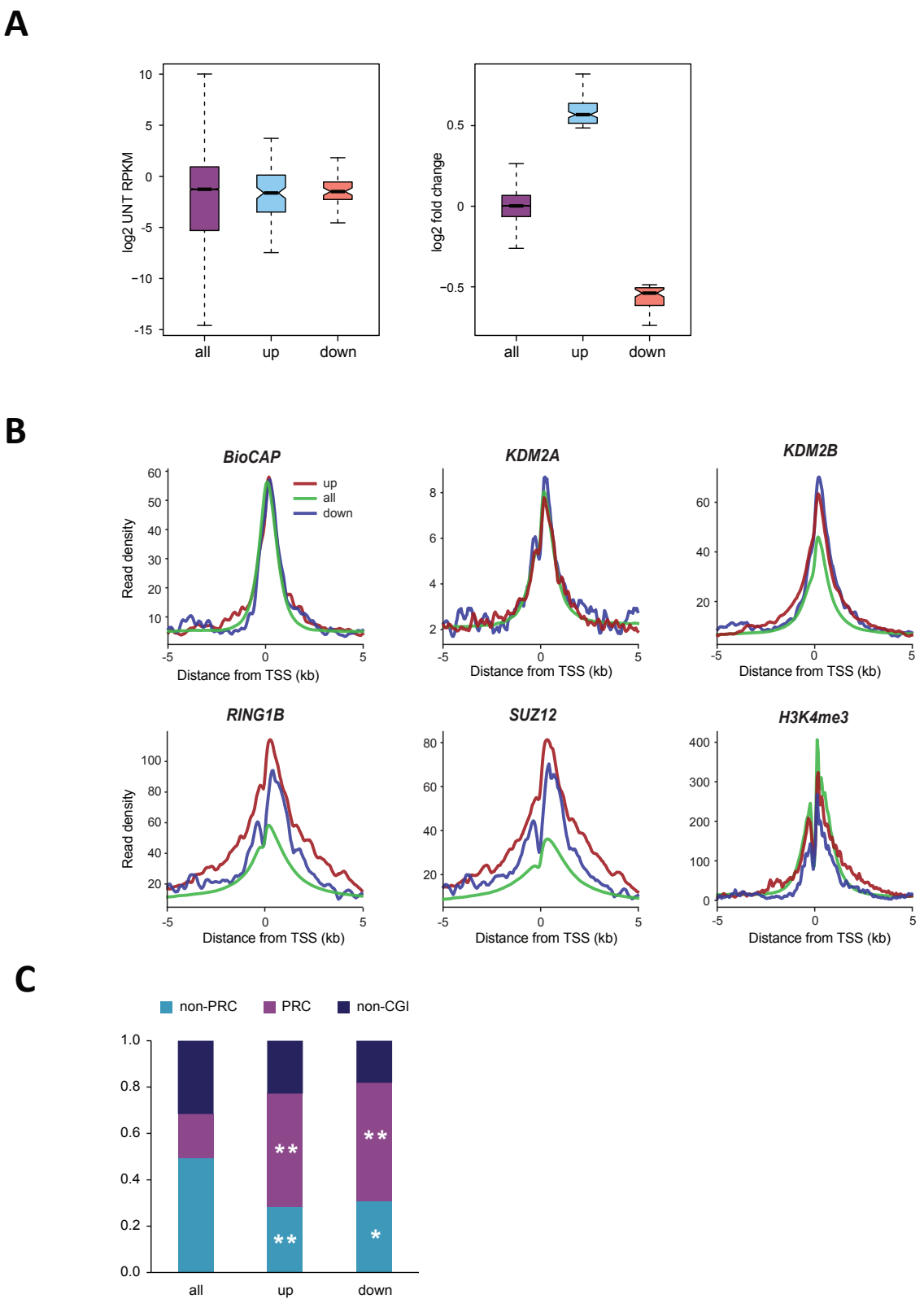


Figure 3.9: A subset of polycomb target genes are deregulated following KDM2A/B-LF knockout

(A) Box plots showing gene expression level in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs (log2 UNT RPKM) and log2 fold change in gene expression following 96 h tamoxifen treatment to induce KDM2A/B-LF knockout. Box plots are subdivided into all genes (n=20633, all) and genes that significantly increased (n=114, up) or decreased (n=55, down) in expression following KDM2A/B-LF knockout.
(B) Metaplots of BioCAP signal and KDM2A, KDM2B, RING1B, SUZ12 and H3K4me3 ChIP signal at the TSS, for all genes and for significantly increasing or decreasing subsets (Datasets: Long et al. 2013; Farcas et al. 2012; Rose et al. 2016; Brown et al. 2017).
(C) Bar chart showing the proportion of genes that belong to non-CGI, polycomb-target (PRC) CGI, and non-polycomb target (non-PRC) CGI classes, for all genes and the significantly increasing and decreasing subsets. PRC target genes were defined by RING1B and SUZ12 occupancy at the TSS (ChIP-seq datasets: Rose et al. 2016). Asterisks represent empirical p-value (* represents p<0.01, ** represents p<0.001), relative to all transcripts.

is intriguing that this set of genes, which are also enriched for polycomb targets, significantly decreased in expression: an increase in their expression might be expected following a reduction in their polycomb-mediated repression following KDM2B-LF knockout. This may be a secondary effect of KDM2A/B-LF knockout, or due to another function of KDM2A/B-LF.

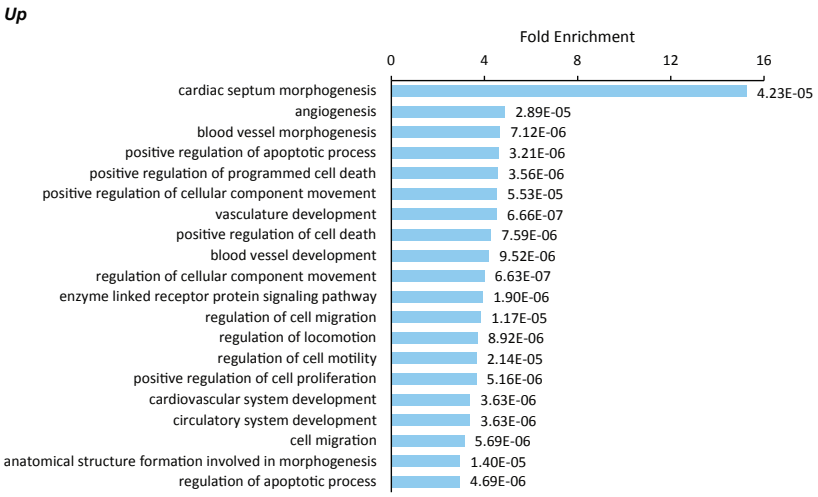


Figure 3.10: Genes that significantly increase in expression are involved in developmental processes
Gene ontology analysis of genes that significantly increased in expression following 96 h tamoxifen treatment of *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs to induce KDM2A/B-LF knockout (n=144). BP FAT analysis with FDR<0.1 cutoff was used, and p-values are annotated. The 20 most enriched terms are shown.

3.2.7 Loss of KDM2A/B-LF results in minimal changes to chromatin accessibility

CGIs, like other active regulatory elements, are characterised by a relatively accessible chromatin state (Fenouil et al. 2012; Sabo et al. 2004; Teif et al. 2012; Li et al. 2011;

Choi 2010). This has been proposed to be important for CGI to act as gene expression regulatory platforms (Blackledge and Klose 2011; Deaton and Bird 2011). Given that KDM2A/B-LF broadly regulate H3K36me2 at CGI promoters and are required for the normal transcription of a small number of genes, I wondered whether they are also regulators of CGI accessibility. To address this question, I performed ATAC followed by massively parallel sequencing (ATAC-seq) using nuclei prepared from *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs, before and after 96 h tamoxifen treatment to induce KDM2A/B-LF knockout. Visual examination of ATAC-seq signal at CGI promoters did not reveal any apparent differences between the tamoxifen-treated and wild type cells (*Asph*, *Fbxo8*, Figure 3.11 A). Genome-wide analysis confirmed that chromatin accessibility was unchanged, with only two sites significantly increasing in accessibility and no sites decreasing in accessibility following KDM2A/B knockout ($p<0.05$, fold change>1.4, Figure 3.11 B). Therefore, KDM2A/B-LF are not required to maintain the heightened accessibility of CGI throughout the genome.

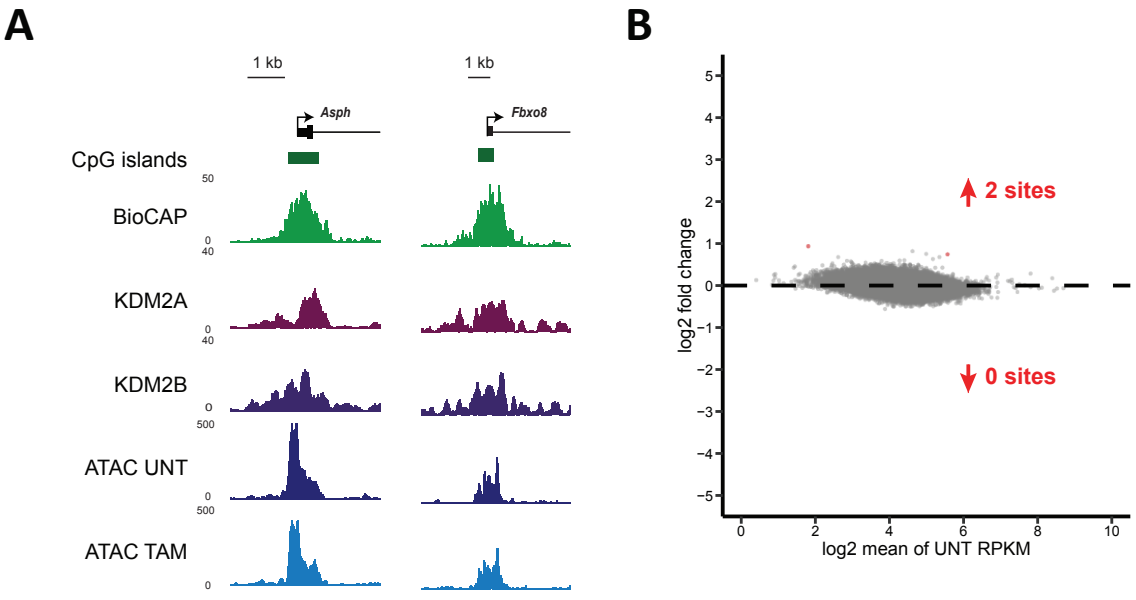


Figure 3.11: KDM2A/B-LF knockout leads to minimal changes in chromatin accessibility
(A) Representative genome snapshots of CGI TSS, which are enriched for BioCAP signal, KDM2A and KDM2B (Datasets: Long et al. 2013; Farcas et al. 2012). ATAC-seq signal in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs is shown before (UNT) and after 96 hours of tamoxifen (TAM) treatment to induce KDM2A/B-LF knockout. (B) Scatter plot of log2 mean ATAC-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs for ATAC peaks (n=74190, three biological replicates), plotted against log2 fold change in ATAC-seq signal following KDM2A/B-LF knockout. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change >1.4).

3.3 Summary and discussion

KDM2A/B are H3K36me2-specific demethylases, and are recruited to CGIs throughout the genome. Here I have explored the role of KDM2A/B demethylase activity in CGI regulation, using a conditional mESC line with tamoxifen-inducible KDM2A/B-LF knockout. Importantly expression of the non-catalytic short isoforms of KDM2A/B was unaffected following tamoxifen treatment, allowing the role of the demethylase-competent long isoforms to be studied in isolation.

KDM2A/B-LF made a minimal contribution to overall H3K36 methylation levels (Figure 3.3). Given that KDM2A/B specifically localise to CGI, which represent a minority of the total genome, this is perhaps unsurprising. Fang et al. reported that knockout of the yeast paralogue JHD1 also does not result in detectable changes to the overall levels of H3K36 methylation (Fang et al. 2007). In contrast, I revealed that KDM2A/B-LF contribute to the marked depletion of H3K36me2 at CGI TSS (Figure 3.6). Knockout of KDM2A/B-LF led to an increase in H3K36me2 levels surrounding the TSS and to shrinking of the H3K36me2 depleted region. This is consistent with the targeting of KDM2A/B-LF to CGI promoters throughout the genome, where their demethylase activity contributes to a H3K36me2-depleted state. In addition to this pronounced depletion of H3K36me2 surrounding the TSS, I showed that highly transcribed genes also had relatively low H3K36me2 levels throughout the gene body (Figure 3.7). This gene body depletion was unaffected by KDM2A/B knockout and was likely a result of conversion of H3K36me2 to H3K36me3, which is strongly enriched in the bodies of highly transcribed genes (Bannister et al. 2005; Mikkelsen et al. 2007). Indeed, H3K36me2 was more strongly depleted at the 3' end of genes, coinciding with the highest levels of H3K36me3.

A pronounced H3K36me2-depleted region persisted at CGI promoters in KDM2A/B-LF knockout cells. It is possible that longer term culture of mESCs in the absence of KDM2A/B-LF would lead to a more dramatic increase in H3K36me2 at CGI. This could be investigated by performing ChIP-seq at a later time-point following tamoxifen treatment of *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs. However, a large proportion of KDM2A/B-LF protein was lost after 24 h of tamoxifen treatment (Supplementary Figure A.1),

indicating that the absence of KDM2A/B-LF for 2-3 days had only a subtle effect on H3K36me2 CGI levels. This suggests that KDM2A/B demethylase activity is one of several mechanisms at play to ensure that TSS remain depleted of H3K36me2.

It is likely that there is functional redundancy between the KDM2 and KDM4 histone demethylase families at these regions. The KDM4 proteins catalyse H3K36me2/3 and H3K9me2/3 demethylation via their JmjC domain (Cloos et al. 2006; Klose et al. 2006; Whetstine et al. 2006; Fodor et al. 2006), and importantly KDM4A and KDM4C localise to gene promoters throughout the genome (Pedersen et al. 2014; Pedersen et al. 2016). The double Tudor domain of KDM4A and KDM4C appears to play a predominant role in their recruitment to regions enriched for H3K4me3 (Pedersen et al. 2014; Pedersen et al. 2016), which is a hallmark of CGI gene promoters, and their targeting may also be regulated by a number of sequence-specific transcription factors (Luo et al. 2012; Wang et al. 2016; Kim et al. 2016; Li et al. 2018). Pedersen et al. showed that simultaneous KDM4A and KDM4C knockout leads to increased H3K36me3 and H3K9me3 at TSS, including a subtle contraction of the H3K36me3 depleted region that parallels the shrinking of the H3K36me2 depleted region I observed following KDM2A/B knockout (Pedersen et al. 2016). The levels of H3K36me2 following KDM4A/C knockout were not reported in this study. However, given that KDM4 proteins can also demethylate H3K36me2, it is likely that KDM4A/C cooperate with KDM2A/B to remove H3K36me2 from gene promoters throughout the genome. The continued action of KDM4A/C may therefore explain the persistence of an H3K36me2 depleted region following KDM2A/B-LF knockout.

Are high levels of H3K36me2 targeted to CGI promoters, or do the KDM2 and KDM4 families play a relatively minor role in preventing the spurious accumulation of low levels of this modification? It is possible that CGI are inherently refractory to H3K36 methylation, and that H3K36me2 demethylase activity acts as a fail-safe mechanism to prevent encroachment of H3K36me2. NSD1-3 appear to play a widespread role in H3K36 dimethylation throughout the genome, as overexpression of NSD2 leads to a broad increase in H3K36me2 (Kuo et al. 2011) while NSD1 knockdown leads to a broad decrease in H3K36me2 (Streubel et al. 2018). Furthermore, these H3K36me1/2 methyltransferases appear to broadly localise to gene promoters as well as gene body

regions (Kuo et al. 2011; Lucio-Eterovic et al. 2010; Rahman et al. 2011; Ram et al. 2011; Shen et al. 2015). This suggests that H3K36me2 is actively targeted to CGI promoters, and that there are high fidelity systems in place to counteract the concerted action of H3K36 methyltransferases at these regions. I revealed that non-CGI gene promoters were not depleted of H3K36me2, indicating that this is not a more general property of all TSS (Figure 3.7). Instead, the depletion of H3K36me2 at CGI promoters is consistent with their demethylation as a result of the specific localisation of KDM2 and KDM4 proteins to these regions.

Combinatorial knockout of all H3K36me2 demethylases active in mESCs would be required to definitively establish their overall contribution to the H3K36me2-depleted state at CGI promoters, and to assess their combined contribution to transcription regulation and CGI accessibility. However, given the likelihood of functional redundancy between multiple demethylase proteins, this would prove immensely challenging. Alternatively, the transcriptional effect of the accumulation of H3K36 methylation at CGI promoters could be studied by targeting an H3K36me2 methyltransferase to these regions, for example by fusion to dead Cas9 or to a ZF-CXXC domain. The combined role of the KDM2 and KDM4 families could also be investigated by combining KDM2A/B-LF knockout with perturbation of H3K4me3 at CGI promoters to disrupt KDM4 targeting to these regions, for example by recruitment of dead Cas9 fused to an H3K4 demethylase. It is important to consider that the KDM2 and KDM4 families also appear to play additional roles beyond the demethylation of histones, which may contribute to transcriptional regulation. For example, KDM2B plays a well-characterised role in recruiting the PRC1.1 complex to CGI promoters (Farcas et al. 2012; He et al. 2013; Wu et al. 2013; Blackledge et al. 2014), while KDM4A was reported to act as an E3 ubiquitin ligase that targets the key cell cycle regulator p53 for degradation (Johmura et al. 2016), and KDM4B may interact with SWI/SNF chromatin remodelling complexes (Kawazu et al. 2011). Therefore, conditional catalytic point mutants would be required to truly study the contribution of the demethylase activity of these proteins to transcriptional regulation.

Here I investigated the transcriptional changes following loss of the demethylase-competent long isoforms of KDM2A/B. I showed that KDM2A/B-LF knockout led to the deregulated expression of only a small number of genes (Figure 3.8). This is

consistent with the lack of overt transcriptional changes following knockout of the KDM2 orthologue JHD1 in *S. cerevisiae* (Fang et al. 2007). The transcriptional changes following KDM2A/B-LF knockout may have resulted from loss of the demethylase-dependent or -independent activities of these isoforms. Indeed, transcriptional changes appear to have been largely restricted to a small subset of polycomb-targeted developmental genes (Figures 3.9 and 3.10), suggesting that a reduction in PRC1.1 targeting as a result of KDM2B-LF knockout was responsible for this deregulation of transcription, rather than a loss of demethylase activity. However the decrease in expression of a small number of genes cannot be explained by a reduction in polycomb-mediated repression. This may have resulted from other demethylase-dependent or independent activities KDM2A/B-LF, or may have been a secondary effect of KDM2A/B knockout. Conditional catalytic point mutation of KDM2A/B would be required to investigate the role of the demethylase activity of these proteins in isolation. Ultimately, the lack of pronounced changes in transcription or chromatin accessibility (Figure 3.11) following KDM2A/B-LF knockout demonstrated that KDM2A/B demethylase activity is not absolutely required for normal CGI function in mESCs.

This finding is in contrast to the prevailing model, in which the highly conserved KDM2A/B paralogues play an important role in CGI regulation. It is possible that KDM2A/B demethylase activity may play a more critical role in regulating transcription during cell fate changes, during which a new gene expression programme must be established and genes specifically induced or silenced. Strong H3K36me2 depletion may be required for a CGI promoter state that is highly permissive to transcription initiation and that facilitates gene induction. Therefore, it is possible that the demethylase activity of KDM2A/B may play a more important role during differentiation. In addition, KDM2A/B-LF may play a more critical role in different cell types where other mechanisms that contribute to the depletion of H3K36me2 at TSS may be less robust. It would be interesting to investigate transcription and chromatin accessibility changes following KDM2A/B-LF knockout during a cell fate transition, for example during retinoic acid-mediated differentiation of *Kdm2a/b-JmjC*$^{-/-}$ mESCs, or in a more differentiated cell type, and to perform H3K36me2 ChIP-seq with and without KDM2A/B-LF knockout in these differentiated cells.

Despite the high expression level of KDM2B-LF compared to KDM2B-SF in the early embryo, KDM2B-LF is not absolutely required for mouse development (Fukuda et al. 2011; He et al. 2013). This indicates that the demethylase activity of KDM2B does not play a critical developmental role. In contrast, KDM2A-LF plays an essential role in embryogenesis, as knockout of KDM2A-LF in isolation results in embryonic lethality by E12.5 (Kawakami et al. 2015). This might suggest that the demethylase activity of KDM2A plays a dominant and essential role during development. However, this phenotype may instead result from a reduction in putative demethylase-independent activities exerted by KDM2A-LF. While these are likely shared by KDM2A-SF, KDM2A-LF is more highly expressed than KDM2A-SF at a variety of embryonic stages (Lainović et al. 2017), suggesting that knockout of KDM2A-LF alone may lead to a pronounced reduction in both demethylase-dependent and independent activities. It would be interesting to investigate the specific role of KDM2A demethylase activity during development using a catalytic point mutant, and to explore potential functional redundancy between KDM2A-LF and KDM2B-LF by studying the development of a combinatorial catalytic mutant.

It would also be interesting to profile the changes in H3K36me1 levels following KDM2A/B-LF knockout using ChIP-seq. KDM2A/B-LF are able to demethylate H3K36me1, which is present on approximately 10 to 20% of all histone H3 (Peters et al. 2003; Robin et al. 2007; Schotta et al. 2008; Jung et al. 2010), but their contribution to H3K36me1 levels throughout the genome is unknown. However, the lack of a change in overall H3K36me1 levels following KDM2A/B-LF knockout suggests that any effect would be relatively modest or occur at a relatively small subset of H3K36me1 enriched sites (Figure 3.3).

In conclusion, KDM2A/B contribute modestly to H3K36me2 demethylation at CGI promoters throughout the genome. Loss of KDM2A/B demethylase activity resulted in an increase in H3K36me2 at CGI promoters and a contraction of the H3K36me2 depleted region. However, KDM2A/B-LF knockout led to only modest transcriptional changes of a small number of genes, and had minimal effect on chromatin accessibility. While KDM2A/B demethylase activity is therefore largely not required for normal CGI-associated transcription and chromatin accessibility in mESCs, it is possible that

demethylase-independent functions of the KDM2A/B proteins may play a key role at these sites.

4

KDM2A/B mediate widespread transcription repression

4.1 Introduction

The *Kdm2a* and *Kdm2b* genes contain an internal promoter downstream of the exons encoding the JmjC demethylase domain, which gives rise to short isoforms (SF) of the proteins (Tanaka et al. 2010; Long et al. 2013). While KDM2A/B-SF lack the JmjC domain present in KDM2A/B-LF, they retain the ZF-CXXC domain responsible for targeting KDM2A/B to CGI and the more C-terminal domains. The role played by KDM2A-SF is poorly defined, while KDM2B-SF is sufficient to support PCGF1/PRC1-dependent Polycomb domain formation, as it retains binding to the PCGF1/PRC1 complex and can recruit this complex to establish a polycomb domain (Blackledge et al. 2014).

KDM2A/B have both been proposed to play a variety of roles that are independent of their demethylase activity. For example, KDM2A has been implicated in the recruitment of HP1 to non-methylated DNA to initiate or maintain chromatin silencing (Borgel et al. 2017), and in the recruitment of the E3 ligase NEDD4 to enhancers, leading to RNA Polymerase II ubiquitylation and enhancer deactivation (Tan et al. 2018). In addition to its well-characterised role in targeting the PRC1.1 complex to CGI gene promoters, KDM2B may also possess intrinsic E3 ligase activity (Janzer et al. 2012), and is proposed to ubiquitylate the transcription factor c-Fos leading to its proteasomal degradation (Han et al. 2016). Such demethylase-independent activities are likely shared by the long and short isoforms of the proteins, as they share the same domain structure with the exception of the JmjC demethylase domain. This may lead to a degree of functional redundancy between the two isoforms. While KDM2A/B-LF, and therefore the histone

demethylase activity of KDM2A/B, are not required to maintain normal CGI-associated transcription in mESCs (Chapter 3), such demethylase-independent activities may play a more critical role.

Knockout of either KDM2A or KDM2B leads to embryonic lethality, demonstrating that both proteins play an essential role during development. Knockout of either KDM2B-LF or KDM2B-SF in isolation leads to increased perinatal lethality with neural tube defects, indicating that individually each isoform is not absolutely required for embryonic development (Boulard et al. 2015; Fukuda et al. 2011). In contrast, knockout of KDM2B-LF and KDM2B-SF together leads to embryonic lethality between E7 and E13 with complete penetrance (Blackledge et al. 2014; Boulard et al. 2015; Andricovich et al. 2016). This severe phenotype of the combined knockout reveals that KDM2B-LF and KDM2B-SF play an essential and redundant role during embryogenesis. This suggests that a demethylase-independent activity of KDM2B is critical to development, consistent with the severe developmental defects observed in a bovine mutant with a missense mutation in the C-terminal region of KDM2B (Testoni et al. 2012). In contrast, knockout of KDM2A-LF alone leads to impaired neural development and lethality by E12.5, indicating that KDM2A-LF itself plays an essential developmental role (Kawakami et al. 2015). The phenotype of KDM2A-SF knockout mice, and mice in which both KDM2A-LF and KDM2B-SF are knocked out in combination, remains to be examined. This would shed light on the importance of the demethylase-independent activities of KDM2A during embryonic development.

While the role played by KDM2A-SF during development has not been studied, KDM2A-SF has been shown to play an important role in the progression of a subset of aggressive breast cancers. KDM2A-SF is over-expressed in these cancer cells, due to high level amplification of the C-terminal region of the *Kdm2a* gene. Here KDM2A-SF acts as an oncogenic isoform, since knockdown of KDM2A-SF, but not KDM2A-LF, inhibits breast cancer cell growth and invasion (Liu et al. 2016). This critical role for KDM2A-SF in driving cancer progression demonstrates that demethylase-independent activity of KDM2A can play an important role in disease biology. This activity may also be required during embryonic development.

Here I aimed to explore the demethylase-independent role of KDM2A/B in transcription regulation. I generated a mESC line with tamoxifen-inducible deletion of the ZF-CXXC domains of KDM2A and KDM2B, leading to removal of both KDM2A/B-LF and KDM2A/B-SF from chromatin. Given that demethylase independent activities are likely shared between the long and short isoforms of KDM2A/B, the simultaneous removal of long and short isoforms allowed me to investigate the transcriptional effects following complete loss of these activities from chromatin. KDM2A and KDM2B are highly homologous proteins, so it is possible that there is functional redundancy between the demethylase-independent activities of these paralogues. Simultaneous removal of both KDM2A and KDM2B allowed me to investigate the total contribution of KDM2A/B to transcription regulation, while the conditional nature of this cell line ensured that the primary transcriptional effects of KDM2A/B removal were studied.

4.2 Results

4.2.1 Identification of novel KDM2B isoforms

In order to investigate the overall contribution of KDM2A/B to transcription regulation, I generated a mESC line in which all KDM2A/B protein could be removed from chromatin. This required a detailed understanding of the structure of the *Kdm2a/b* genes. Therefore, I began by examining the chromatin features found throughout the *Kdm2b* gene body. There was a prominent peak of BioCAP signal at the beginning of the *Kdm2b* gene, corresponding to the KDM2B-LF promoter, and another strong peak surrounding the exon upstream of ZF-CXXC-encoding exon, which corresponds to the KDM2B-SF promoter (Figure 4.1 A, promoters highlighted in blue). BioCAP marks non-methylated CpG rich regions throughout the genome, which are characteristic of mammalian promoters (Saxonov et al. 2006; Blackledge et al. 2012; Bird et al. 1985; Tykocinski and Max 1984). Both of these regions also contained elevated ATAC signal, indicative of active regulatory regions (Buenrostro et al. 2013; Buenrostro et al. 2015), and pronounced peaks of H3K4me3 and both total and Serine5-phosphorylated RNAPII, characteristic of gene promoter regions (Komarnitsky et al. 2000; Schroeder et al. 2000;

Bernstein et al. 2005; Schneider et al. 2004).

Strikingly, the *Kdm2b* gene contained two additional regions that were enriched for this collection of promoter-associated features, suggesting that these regions might function as alternative promoters (highlighted in yellow and orange, Figure 4.1 A). In order to investigate the possibility that transcription initiates from these regions, I examined the Ensembl database of *Kdm2b* transcripts. I identified four putative alternative exons that are located within these regions. Three putative exons (A1-3) coincided with the candidate promoter found upstream of the JmjC-encoding exons (highlighted in yellow, Figure 4.1 A), while a single putative exon (B) was present within the downstream candidate promoter region (highlighted in orange). The association of these four putative exons with a range of promoter-associated features suggested that they may serve as TSS for alternative KDM2B isoforms. This is reminiscent of the initiation of *Kdm2b-SF* transcripts from a unique first exon that is not incorporated into *Kdm2b-LF* transcripts. Putative exon B also had strong enrichment of Cap Analysis of Gene Expression (CAGE) signal. CAGE precisely maps the 5' end of cDNAs and therefore marks TSS (Kodzius et al. 2006), including the TSS of both *Kdm2a-LF* and *Kdm2b-SF*. This added further evidence that a novel *Kdm2b* transcript may initiate from the exon B region.

I designed an RT-PCR screen to investigate the possibility that these exons are incorporated into alternative *Kdm2b* isoforms. I generated cDNA from RNA isolated from mESCs, then performed PCR using a forward primer binding within the putative exon and a reverse primer binding within the canonical exon downstream (Figure 4.1 B). I obtained short PCR products of the expected size for putative exons A2, A3 and B (Figure 4.1 C), and validated all three by sequencing. This confirmed that these three alternative exons are indeed incorporated into *Kdm2b* transcripts in mESCs, revealing that multiple isoforms of KDM2B are expressed in mESCs. In contrast, examination of the *Kdm2a* gene did not reveal internal promoter regions in addition to the *Kdm2a-SF* promoter, indicating that KDM2A-LF and KDM2A-SF are the predominant KDM2A isoforms expressed in mESCs.

Translation of *Kdm2b* transcripts initiating from exons A2/3 was predicted to give rise to a novel 126 kDa isoform of KDM2B containing the JmjC demethylase domain (Figure

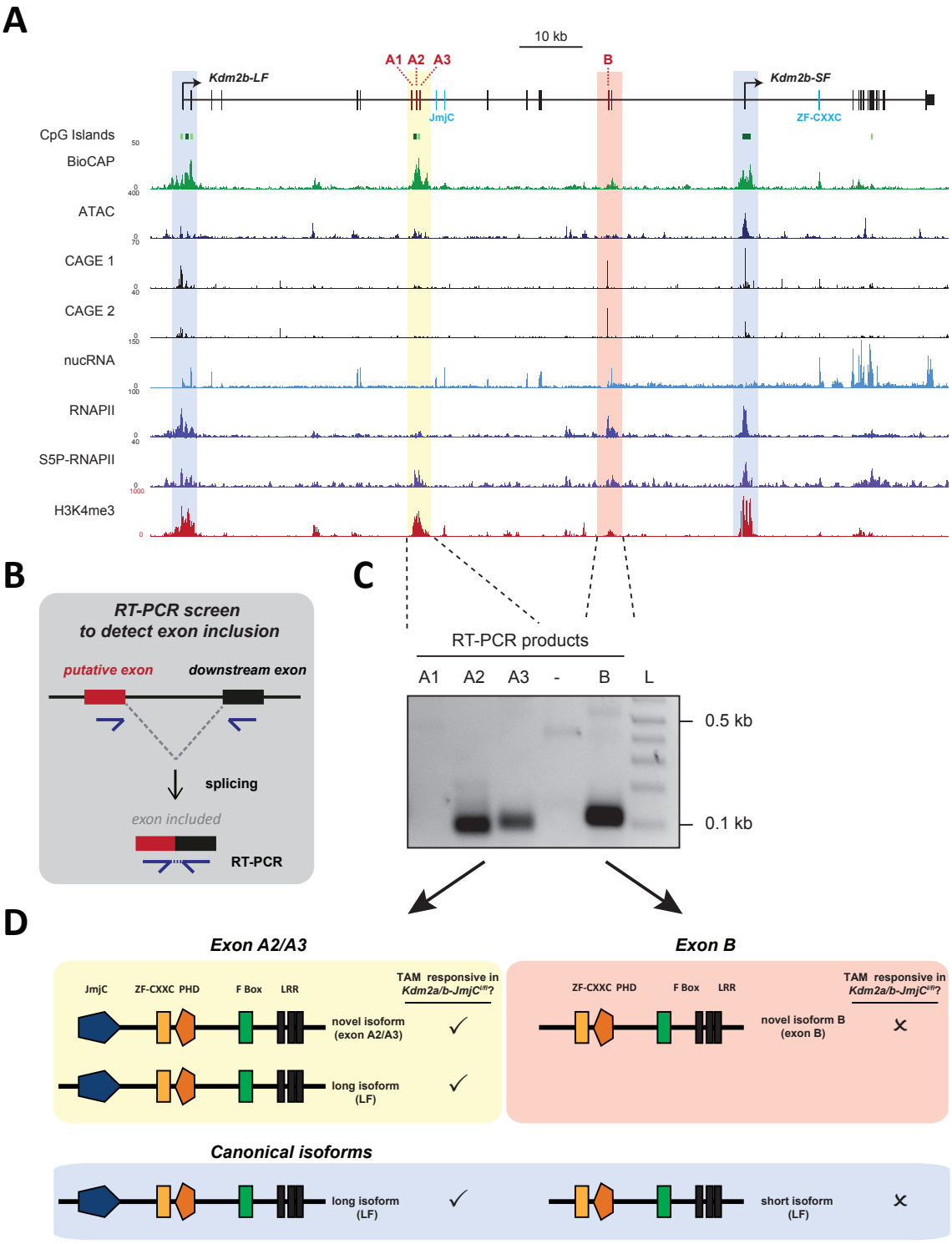


Figure 4.1: Identification of novel *Kdm2b* isoforms in mESCs

(A) Genomic snapshot of the *Kdm2b* gene. The *Kdm2b-LF* and *Kdm2b-SF* promoters are highlighted in lilac, while candidate promoter regions are highlighted in yellow and orange. Putative alternative first exons identified from the Ensembl database are coloured red, and critical exons encoding the JmjC and ZF-CXXC domains are coloured blue. Candidate promoter regions coincide with a number of promoter-associated features, including enrichment of BioCAP, H3K4me3, total and serine 5 phosphorylated RNAPII signal, and CAGE from two different mESC lines (Datasets: Long et al. 2013; Brown et al. 2017; Brookes et al. 2012; Forrest et al. 2014).
(B) Schematic of the RT-PCR screen used to detect inclusion of the putative exons into alternative *Kdm2b* transcripts. The forward primer binds within the putative exon, and the reverse primer binds within the canonical exon downstream. If the putative exon is included in the mature mRNA through splicing to the exon downstream, then a short PCR product will be detected following RT-PCR.
(C) Agarose gel showing RT-PCR products corresponding to exon A2/A3/B spliced onto the downstream exon. A RT-PCR product corresponding to splicing of putative exon A1 was not detected. Ladder (L) is shown for reference. (-) indicates negative control lane.
(D) Summary of putative KDM2B isoforms, showing their domain architecture. Left: exons A2/3 could represent the first exon of alternative *Kdm2b* isoforms, or could be incorporated into *Kdm2b-LF* transcripts through alternative splicing. In either case, these isoforms would likely be lost following tamoxifen (TAM)-induced deletion of the JmjC-encoding exons downstream in *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs. Right: expression of an 89 kDa isoform initiating from novel exon B downstream would be unaffected by TAM treatment in this line. Canonical KDM2B-LF and KDM2B-SF isoforms are shown below for reference.

4.1 D - left). However, an isoform of this size that was responsive to tamoxifen-induced deletion of the JmjC domain in the *Kdm2a/b-JmjC*$^{fl/fl}$ line studied in Chapter 3 was not evident by western blotting (Figure 3.2). Furthermore, the lack of a prominent peak of CAGE signal coinciding with exons A2/3 suggested that transcription does not commonly initiate from this region in mESCs. Instead, exons A2/3 are likely incorporated into *Kdm2b* transcripts as the result of alternative splicing to the exon upstream. In contrast, the coincidence of exon B with a prominent mESC CAGE peak provided strong evidence that transcription of an alternative KDM2B isoform initiates from this position. This would lead to the production of a 89 kDa protein, just 1 kDa larger than the canonical KDM2B-SF protein, which I will refer to as KDM2B isoform B (Figure 4.1 D - right). Indeed, a preliminary analysis using constitutive CRISPR-mediated knockout of KDM2B-SF and KDM2B isoform B confirmed the expression of this novel isoform in mESCs (Supplementary Figure A.2).

4.2.2 Conditional deletion of the ZF-CXXC domain of KDM2A/B leads to their removal from chromatin

I aimed to generate a mESC line in which I could remove all KDM2A/B isoforms, in order to study the total contribution of KDM2A/B to transcription regulation. However,

it would be challenging to individually target each of the multiple KDM2A/B isoforms with loxP sites to enable their tamoxifen-inducible knockout. In contrast, constitutive knockout of the KDM2A/B isoforms might lead to the gradual accumulation of gene expression defects with continued mESC culture and to secondary effects of KDM2A/B loss. Therefore, here I instead focus on the characterisation of a fully conditional mESC line with tamoxifen-inducible deletion of the KDM2A/B ZF-CXXC domains. Each of the identified KDM2A/B isoforms includes the ZF-CXXC DNA binding domain, which is required for KDM2A/B to bind to CGI throughout the genome (Koyama-Nasu et al. 2007; Yamagishi et al. 2008; Blackledge et al. 2010; Farcas et al. 2012). Therefore, tamoxifen treatment of this line should result in conditional removal of all KDM2A/B isoforms from chromatin, leading to loss of their chromatin-associated functions. The conditional nature of this line meant that I could examine the primary transcriptional effects of KDM2A/B removal.

I began with the previously reported *Kdm2b-CXXC*$^{fl/fl}$ mESC line, in which an exon encoding most of the ZF-CXXC domain of KDM2B is flanked by loxP sites (Figure 4.2 A, Blackledge et al. 2014). This line also expresses tamoxifen-inducible Cre recombinase, enabling conditional deletion of the ZF-CXXC domain. The downstream exon is in frame, therefore truncated versions of the KDM2B-LF and KDM2B-SF proteins lacking the ZF-CXXC domain are produced, which do not associate with chromatin but remain associated with the PRC1.1 complex (Blackledge et al. 2014). To achieve removal of KDM2A as well as KDM2B from chromatin, I sequentially inserted loxP sites flanking the exon encoding the majority of the KDM2A ZF-CXXC domain using CRISPR-mediated genome editing. This technique relies on the RNA-directed Cas9 nuclease, which creates a double strand break (DSB) in DNA sequences complementary to a 20 nucleotide region of an associated single guide RNA (sgRNA, Garneau et al. 2010; Gasiunas et al. 2012; Jinek et al. 2012). I supplied plasmids encoding two different sgRNAs, one complementary to a region upstream of the target exon and another complementary to a region downstream, and a HDR template, containing the loxP sequence and short 150 bp homology arms for repair of the sgRNA-targeted DSB (Figure 4.2 A). I identified successfully targeted clones using PCR screens of genomic DNA and cDNA, and confirmed that the exact loxP sequence had been inserted by Sanger sequencing.

I tested the response of the generated *Kdm2a/b-CXXC*$^{fl/fl}$ line to tamoxifen, using western blotting of nuclear extract. This revealed that 96 h tamoxifen treatment led to a decrease in molecular weight of both KDM2A/B-LF and KDM2A/B-SF, corresponding to efficient deletion of the ZF-CXXC domain (Figure 4.2 B). Interestingly, the abundance of KDM2B-LF appeared to increase following tamoxifen addition, whereas expression of KDM2B-SF and KDM2A-LF/SF appeared largely unchanged. This may have resulted from the increased efficiency of KDM2B-LF salt extraction when it lacked the ZF-CXXC domain and therefore was no longer bound to chromatin. The expression of the third KDM2 family member FBXL19, which is a paralogue of KDM2A/B-SF, remained unchanged (Figure 4.2 C). FBXL19 appeared as multiple bands by western blotting, which may indicate that the protein is heavily post-translationally modified.

Having verified that the *Kdm2a/b-CXXC*$^{fl/fl}$ line responded appropriately to tamoxifen, I next evaluated the chromatin binding profile of KDM2A/B following ZF-CXXC domain deletion, using ChIP-qPCR (Figure 4.3). This confirmed that 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs led to a near complete loss of both KDM2A and KDM2B from CGI promoters. Therefore, this line could be used to study the immediate transcriptional effects following removal of KDM2A/B from chromatin. Throughout this study I will refer to the loss of KDM2A/B binding to chromatin following tamoxifen treatment as KDM2A/B removal.

4.2.3 KDM2A/B are required to sustain normal levels of transcription in mESCs

Having confirmed that tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs led to deletion of the ZF-CXXC domain of KDM2A/B and to removal of KDM2A/B from CGI promoters, I was interested to investigate the transcription changes that occur in this cell line. Therefore, I performed nucRNA-seq analysis of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, before and after 96 h tamoxifen treatment to induce KDM2A/B removal. I initially used a conventional approach to RNA-seq analysis, normalising read counts to total library size. A genome-wide analysis revealed that the expression of more than 15% of all genes was significantly affected following removal of KDM2A/B using stringent thresholds ($p<0.05$

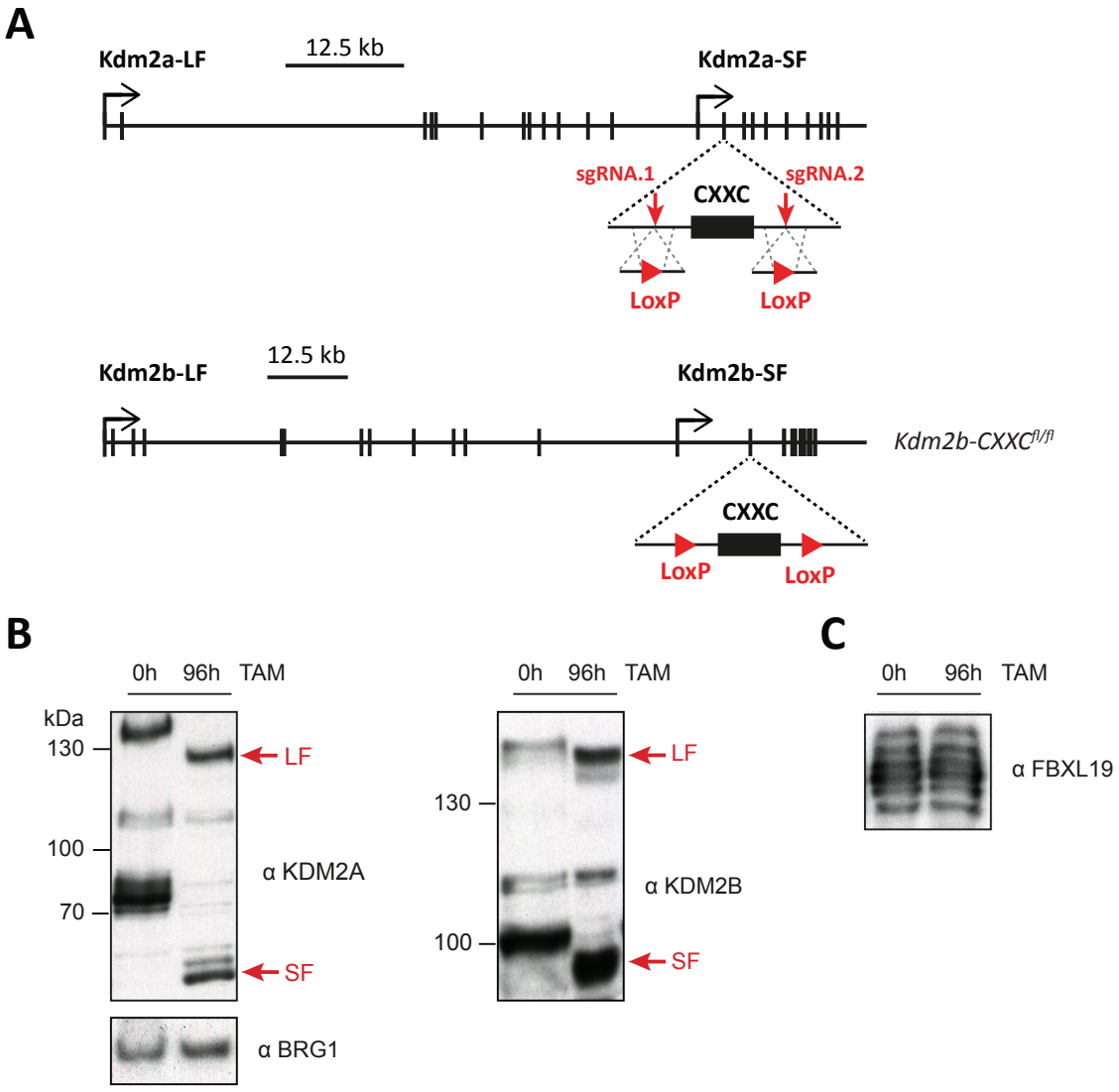


Figure 4.2: Conditional deletion of the ZF-CXXC domain of KDM2A/B
(A) Schematic depicting the position of loxP sites flanking an exon encoding the ZF-CXXC domain of KDM2B, and the position of sgRNAs used to target Cas9 to generate DSBs flanking an exon encoding the ZF-CXXC domain of KDM2A. LoxP sites were inserted via repair of the DSB using a supplied HDR template.
(B) Western blot showing deletion of KDM2A/B ZF-CXXC domains following 96 h tamoxifen (TAM) treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, leading to a size shift for both KDM2A/B-LF and KDM2A/B-SF. BRG1 is a loading control for both westerns.
(C) Western blot showing protein levels of FBXL19 before and after TAM treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs. Loading control: BRG1, see (B).

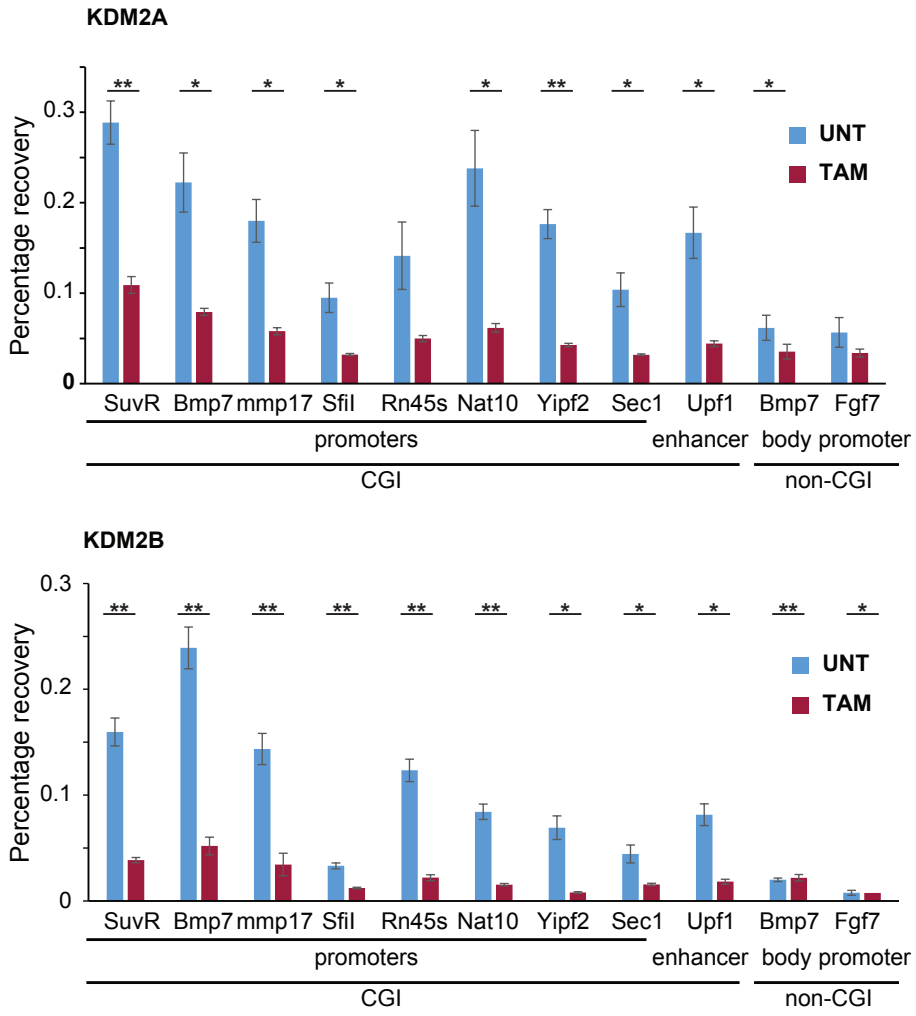


Figure 4.3: ZF-CXXC domain deletion leads to removal of KDM2A/B from chromatin
(A) ChIP-qPCR analysis of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs with or without 96 h tamoxifen (TAM) treatment to induce KDM2A/B ZF-CXXC domain deletion. KDM2A/B are enriched at a panel of CGI gene promoters and a CGI enhancer, but not at non-CGI promoter or gene body regions. Experiments were performed in triplicate and ChIP signal expressed relative to input. Error bars show standard error. Asterisks represent statistical significance calculated by Student T-test: *p<0.05, ** p<0.01

and fold change>1.4), with a large number of genes either increasing (1921 genes) or decreasing (1304 genes) in expression (Figure 4.4). This revealed that KDM2A/B are required to maintain normal transcript levels in mESCs. As KDM2A/B-LF make a minimal contribution to the regulation of gene expression (Section 3), this suggests that KDM2A/B-SF play a predominant role.

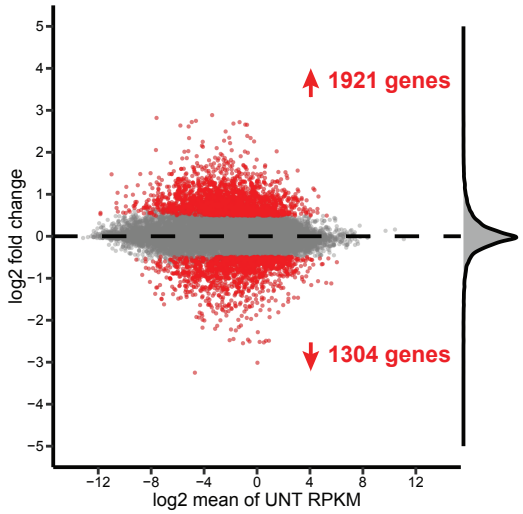


Figure 4.4: Removal of KDM2A/B leads to widespread deregulation of transcription
Scatter plot of log2 mean nucRNA-seq reads per kilobase per million mapped reads (RPKM) in untreated *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs for a non-redundant mm10 gene set (n=20633, three biological replicates), plotted against log2 fold change in nucRNA-seq signal following 96 h tamoxifen treatment to induce removal of KDM2A/B. Libraries were normalised to total read count. The dotted line corresponds to no change in transcription. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change>1.4)

4.2.4 Development of a calibrated nucRNA-seq protocol

Given that I observed widespread changes in transcription following KDM2A/B knockout (Figure 4.4), I considered whether the library normalisation method I used might have prevented the detection of a directional change in transcription. RNA-seq analysis pipelines generally rely on the assumption that only a minority of transcripts change in expression. This allows individual libraries to be normalised to total library size to account for differences in sequencing depth, with the resulting data typically expressed as Reads Per Million. However, if there were a global change in signal in the test samples compared to the untreated samples, for example due to a widespread decrease in transcription, then this effect would be not be detected when normalising by this method as the libraries are scaled according to their size (Figure 4.5). In contrast,

a spike-in approach would allow a widespread reduction or increase in transcription to be reliably detected. In this approach a constant amount of nucleic acid from another species is introduced into each sample, allowing the resulting libraries to instead be calibrated to the number of reads aligning to the spike-in genome. The spike-in genome therefore serves as an internal calibrator, enabling differences in the amount of recovered material between control and test samples to be conserved (Figure 4.5). Calibration of RNA-seq experiments was initially achieved through the introduction of synthetic calibrator RNAs (External 2005), while spike-in calibration through mixing with cells of another species at a defined ratio, in order to effectively spike-in a defined proportion of nucleic acid from this species, was first developed for ChIP-seq (Hu et al. 2015; Bonhoure et al. 2014; Orlando et al. 2014). KDM2A/B bind to the majority of gene promoters and appear to play a widespread role in transcription regulation, so I reasoned that normalisation to total library size might have prevented the detection of a widespread increase or decrease in transcription. Therefore, I decided to develop a spike-in calibration protocol to integrate into my nucRNA-seq pipeline.

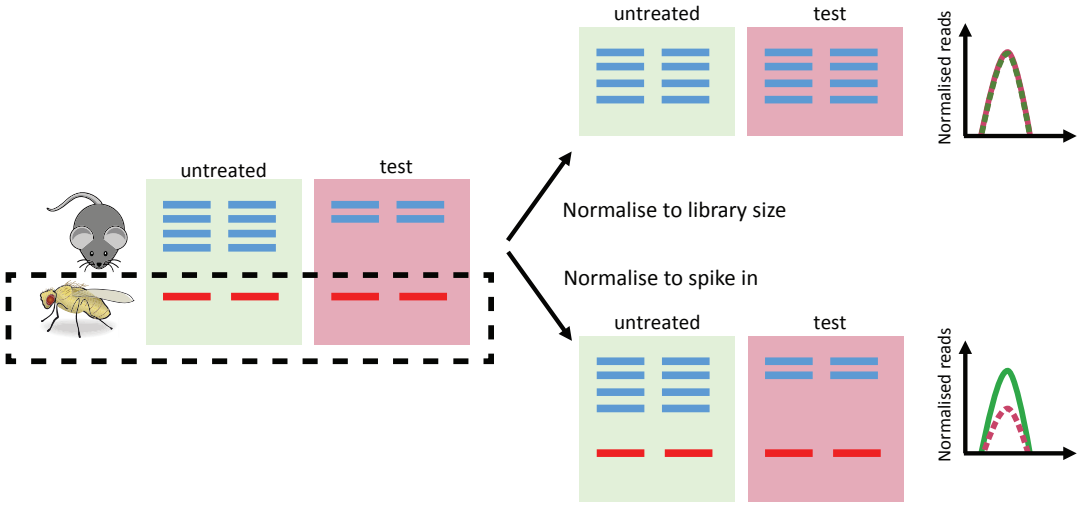


Figure 4.5: Calibrated RNA-seq enables detection of a widespread reduction in transcription
Schematic comparing conventional and spike-in library normalisation methods. A global decrease in transcription is not detectable when normalising to library size, whereas calibration to spike-in nucleic acid from a different species enables detection of this dramatic change.

For conventional RNA-seq analysis, I used the DESeq2 package to normalise libraries to their relative sequencing depths before performing differential gene expression analysis (Love et al. 2014). I modified this analysis pipeline to instead supply a normalisation

factor based on the number of reads aligning to a spike-in genome. I then verified that this modified pipeline enabled detection of a directional change in transcription, using a published mESC RNA-seq dataset with *Arabidopsis thaliana* spike-in. In this experiment the drug triptolide was used to inhibit transcription initiation in mESCs, leading to a dramatic global decrease in transcription by RNAPII (Jonkers and Lis 2015). Conventional analysis including normalisation to library size assigned 4388 genes as significantly increasing in expression and 3325 genes as significantly decreasing in expression ($p<0.05$, Figure 4.6 A). In stark contrast, calibration to the number of reads aligning to the spike-in *Arabidopsis thaliana* genome revealed a global decrease in transcription as a result of RNA polymerase II inhibition, with the vast majority of transcripts significantly decreasing in expression (17513 out of 20633 genes, $p<0.05$ - Figure 4.6 B). Therefore, this spike-in calibration pipeline was capable of revealing a global, directional change in gene expression.

In order to use this calibration pipeline to analyse nucRNA-seq data I developed a

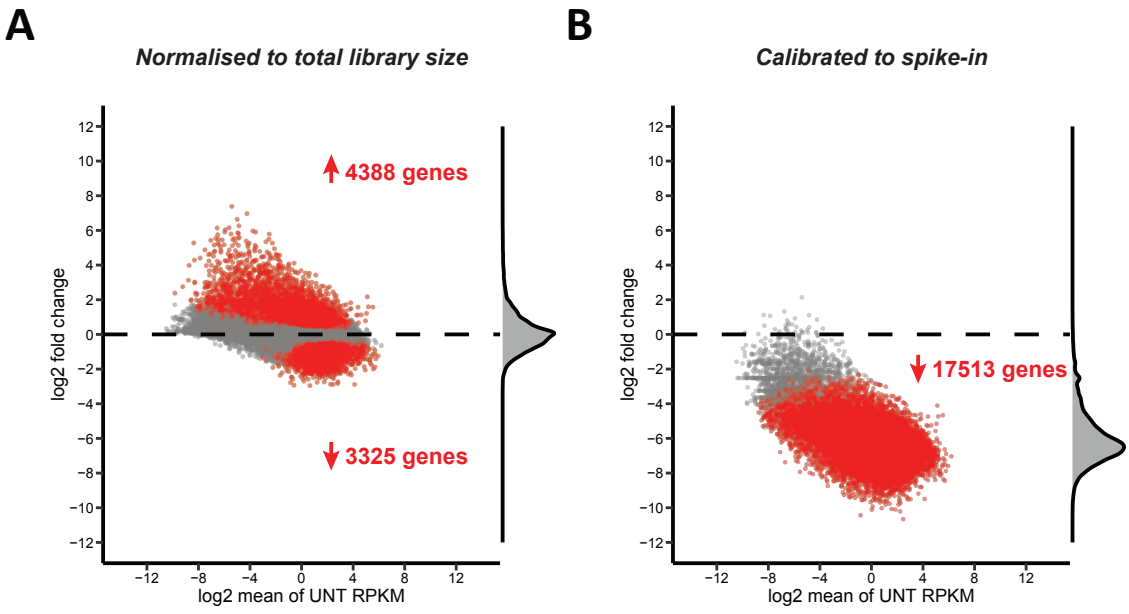


Figure 4.6: Calibrated RNA-seq reveals a global decrease in transcription following triptolide treatment
(A) Scatter plot of log2 mean RNA-seq reads per kilobase per million mapped reads (RPKM) in mESCs for a non-redundant mm10 gene set (n=20633, two biological replicates), plotted against log2 fold change in RNA-seq signal following treatment with triptolide for 50 min (RNA-seq data: Jonkers and Lis 2015). The dotted line corresponds to no change in transcription. Significant changes were identified using DESeq2, incorporating normalisation to total mm10 library size, and are coloured red ($p<0.05$). A density plot of the log2 fold change in RNA-seq signal is shown to the right.
(B) As (A), but instead incorporating DESeq2 calibration to total reads aligning to the spike-in *Arabidopsis thaliana* genome.

spike-in version of my nucRNA-seq protocol, in which 10x10^6 mESCs were mixed with 2.5x10^6 *Drosophila melanogaster* SG4 cells. nucRNA was subsequently prepared from this cell mixture. I chose to spike in SG4 cells as they have relatively simple cell culture requirements, there is a high quality sequence assembly of the *Drosophila* genome, and as there is minimal mapping of the *Drosophila* genome sequence to the mouse genome (Orlando et al. 2014). This enabled the assignment of sequencing reads to either the mouse or *Drosophila* genome with high confidence, ensuring accurate counting of *Drosophila* reads and therefore accurate spike-in calibration.

While the total number of reads is often used as a measure of sequencing depth, this total may be skewed by a few highly and differentially expressed genes. This renders the ratio of total read counts between two samples an inaccurate estimate of the ratio of read counts for the majority of genes (Anders and Huber 2010). Instead, the DESeq2 package incorporates normalisation of sequencing libraries to Size Factors, the median of the ratios of observed counts, which are a more robust measure of relative sequencing depths as they are not skewed by outliers. I reasoned that this Size Factor normalisation method could also be used as a more accurate measure of the sequencing depth of the spike-in genome, in order to ensure the robustness of spike-in normalisation. I therefore incorporated this approach into my spike-in analysis pipeline, by first calculating *Drosophila* Size Factors from a table of *Drosophila* transcript read counts, and then applying these Size Factors to the analysis of mouse transcripts. To further increase the accuracy of this calibrated analysis, I sequenced genomic DNA prepared from the input mixture of mouse and *Drosophila* nuclei, and pre-normalised the *Drosophila* samples to this ratio in order to account for variation in cell mixing ratios (Figure 4.7).

4.2.5 Removal of KDM2A/B leads to widespread transcription upregulation

Having established a custom spike-in calibration pipeline, I next wanted to use this approach to more accurately examine the effect of KDM2A/B removal on transcription. I therefore used this pipeline to investigate the transcription changes following removal of KDM2A/B in tamoxifen-treated *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs. Visual examination of the calibrated nucRNA-seq signal revealed that moderate upregulation of transcription

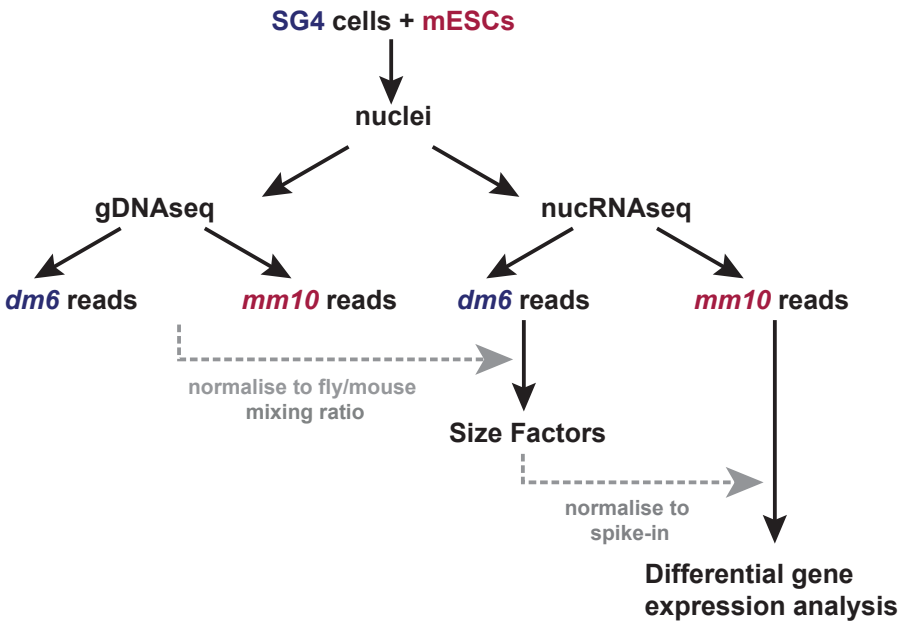


Figure 4.7: nucRNA-seq calibration pipeline
Schematic of the pipeline used to calibrate nucRNA-seq datasets using spiked-in *Drosophila* SG4 cells. SG4 cells were added to mESCs at a defined ratio, then nuclei were prepared from this mixture. Genomic DNA was prepared from a sample of nuclei and sequenced (gDNAseq), enabling normalisation to the exact fly/mouse mixing ratio. nucRNA was prepared from another sample of nuclei and sequenced (nucRNAseq). Size factors were calculated for a *Drosophila* gene set using the DESeq2 package, following pre-normalisation to the mixing ratio, and these were applied to the DESeq2 differential analysis of a mouse gene set.

was widespread (e.g. *Fbxl20*), while many genes were more dramatically upregulated (*Igf2*), and the expression of other genes was not dramatically affected (*Bzw2*, Figure 4.8 A). A genome-wide analysis confirmed that there was a widespread increase in gene expression, with 4528 significantly increasing and 531 significantly decreasing transcripts ($p<0.05$ and fold change>1.4, Figure 4.8 B). Furthermore, this calibrated analysis revealed that removal of KDM2A/B led to a more general increase in transcription, illustrated by the positive skew of log2 fold change density such that the majority of points lie above zero (Figure 4.8 B - density plot). This indicated that the nearly ten-fold greater number of significantly increasing compared to decreasing transcripts was representative of a more global trend.

The detection of directional transcription changes in this nucRNA-seq dataset relied on calibration to the *Drosophila* genome (Section 4.2.4). Therefore, I decided to validate the widespread increase in transcription revealed by this method using ddPCR, an alternative approach that does not rely on normalisation to an internal calibrator. There are two key differences between ddPCR and standard qPCR quantitation. The ddPCR

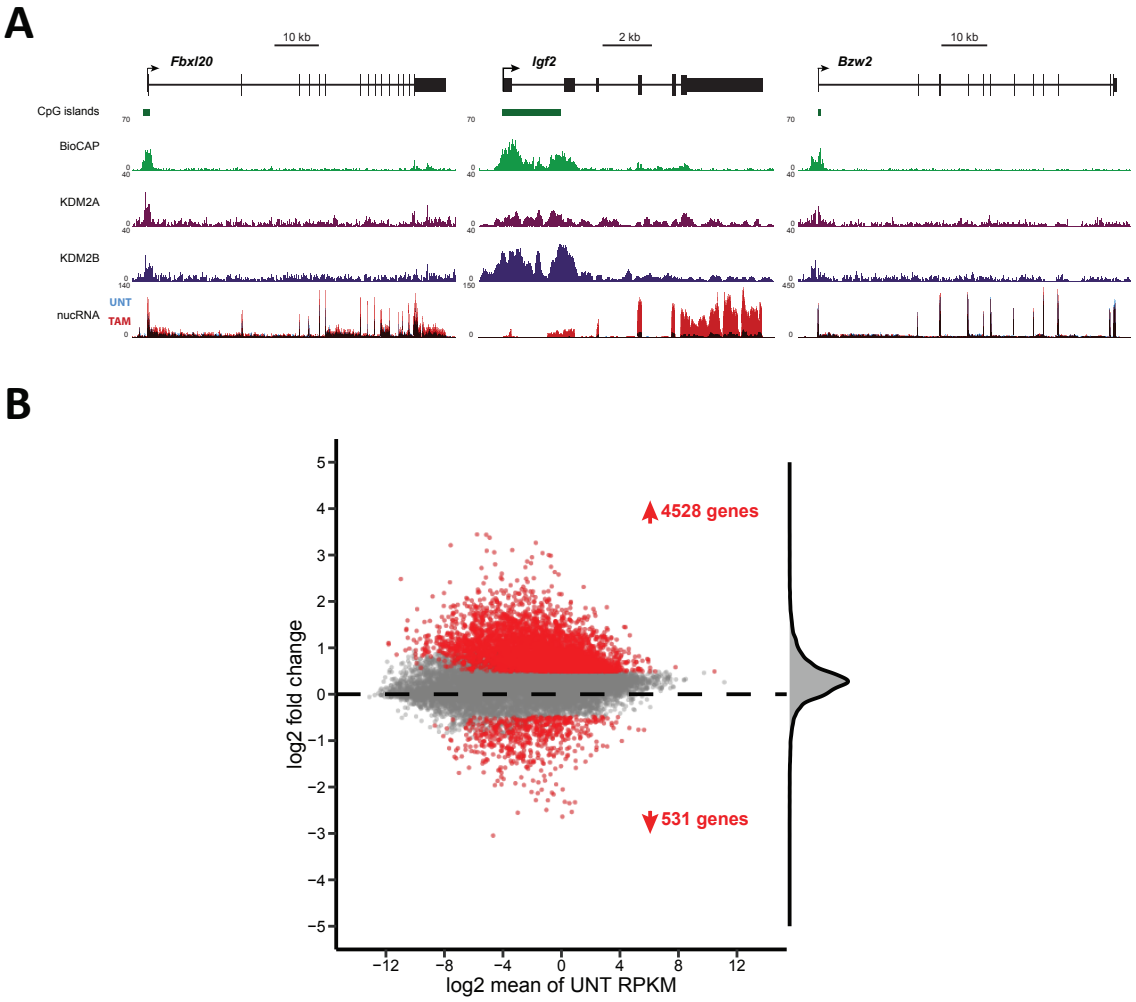


Figure 4.8: Removal of KDM2A/B leads to widespread transcription upregulation
(A) Genome snapshots of genes which moderately (*Fbxl20*) or dramatically (*Igf2*) increased in expression or were largely unaffected (*Bzw2*) following removal of KDM2A/B. The CGI promoters of these genes are enriched for BioCAP signal, KDM2A and KDM2B (Datasets: Long et al. 2013; Farcas et al. 2012). Strand-specific nucRNA-seq signal is depicted as an overlay plot, with the relevant strand shown. Signal in *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs before 96 h tamoxifen (TAM) treatment is coloured blue and signal after TAM treatment is coloured red.
(B) Scatter plot of log2 mean nucRNA-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs for a non-redundant mm10 gene set (n=20633, four biological replicates), plotted against log2 fold change in nucRNA-seq signal following 96 h tamoxifen treatment to induce removal of KDM2A/B. Libraries were calibrated to *Drosophila* spike-in. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change>1.4). A density plot of the log2 fold change in nucRNA-seq signal is shown to the right.

reaction is partitioned into thousands of droplets using a water-oil emulsion system prior to amplification (Figure 4.9 A), improving detection of small fold changes due to massively parallel sample partitioning. In addition data is collected at 'end-point', in contrast to the derivation of quantitative information during qPCR from a point on the analogue fluorescence curve, termed the cycle threshold (CT), where the signal increases above background (Hindson et al. 2011; Hindson et al. 2013). End point PCR, when performed using a limiting dilution of template molecules and combined with Poisson statistics, yields an absolute measure of DNA concentration (Vogelstein and Kinzler 1999; Sykes et al. 1992). Therefore, in contrast to qPCR, an internal calibrator is not required to precisely quantify the amount of target DNA in the ddPCR reaction. The ddPCR reaction mixture includes a double-stranded DNA(dsDNA)-intercalating fluorescent dye, enabling counting of fluorescent droplets which contained at least one copy of the target DNA, through droplet sorting (Figure 4.9 A).

I prepared cDNA from 10^6 *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, with or without 96 h tamoxifen treatment to induce removal of KDM2A/B, in triplicate. Whereas a defined concentration of DNA is typically added to a ddPCR reaction, which conceptually parallels normalisation of RNA-seq data to total library size, I instead added a defined volume of cDNA prepared from 10^6 mESCs, allowing me to conserve any difference in total RNA amount per cell following tamoxifen treatment. Normalisation to an internal calibrator is often used in ddPCR analysis of cDNA in order to account for variability in the many steps involved, including cell lysis, RNA extraction, reverse transcription and pipetting errors. I avoided the need for internal calibration by following a simplified and reproducible RNA extraction and cDNA preparation protocol, whereby a reproducible amount of RNA was recovered from a defined number of mESCs. The nucRNA isolation protocol involved several steps, including washing and resuspension of the nuclear pellet and phenol-chloroform RNA extraction, which might have led to variation in the number of recovered nuclei and the final amount of purified RNA. Therefore, I instead prepared total RNA by loading mESCs directly onto RNA extraction columns and eluting in a defined volume, in order to minimise variation between samples. I then prepared cDNA from this RNA in a defined reaction volume, and used a constant volume of this cDNA in the ddPCR reactions.

Analysis of the generated ddPCR droplets confirmed the moderate increase in transcription observed for many transcripts by nucRNA-seq analysis (Figure 4.9 B). ddPCR analysis (light blue bars) reproduced not only dramatic increases (*Igf2*, *Lamc2*) and decreases (*Klf8*) in transcription, but also more subtle increases that were either above (*Fbxl20*, *Cdx5rap3*, *Gmpr2*, *Mpc2*) or below (*Nsmce4a*, *Sos1*, *Top2a*) the 1.4 fold change threshold used to call significantly changing genes in the calibrated nucRNA-seq analysis (dark blue bars). These subtle increases were not captured when the nucRNA-seq libraries are instead normalised to total read count (green bars). This ddPCR analysis, which did not rely on calibration, therefore confirmed the moderate upregulation in transcription observed for thousands of transcripts by calibrated nucRNA-seq analysis, suggesting that this was an accurate representation of the gene expression changes following removal of KDM2A/B. Interestingly, the fold change calculated using ddPCR (light blue bars) was consistently higher than the value given by spike-in nucRNA-seq analysis (dark blue bars), with the exception of *Mpc2*, for both moderately increasing transcripts (*Fbxl20*, *Cdx5rap3*, *Gmpr2*) and for transcripts with no change in expression using spike-in analysis (*Map2k1*, *Dbt*, *Bzw2*). This may be due to the compression of log fold change values that occurs as part of rigorous DESeq2 analysis, suggesting that the calibrated nucRNA-seq analysis may in fact have underestimated the full extent of transcription upregulation that occurs.

I observed a strong enrichment of nucRNA-seq reads in the exons of genes compared to their introns (Figure 4.8 A). This indicated that a majority of the purified nucRNA was spliced rather than truly nascent. In order to investigate the effect of KDM2A/B removal on nascent transcription, I used a custom analysis pipeline to investigate the change in reads aligning to the introns, indicative of nascent unspliced RNA, or across exon-exon junctions, indicative of spliced mRNA, within this nucRNA-seq dataset. The widespread increase in transcription following removal of KDM2A/B was evident even when examining only nascent unspliced transcripts, albeit less dramatically than for spliced or all gene body transcripts (Supplementary Figure A.3). This analysis therefore suggests that removal of KDM2A/B from chromatin leads to an increase in transcription.

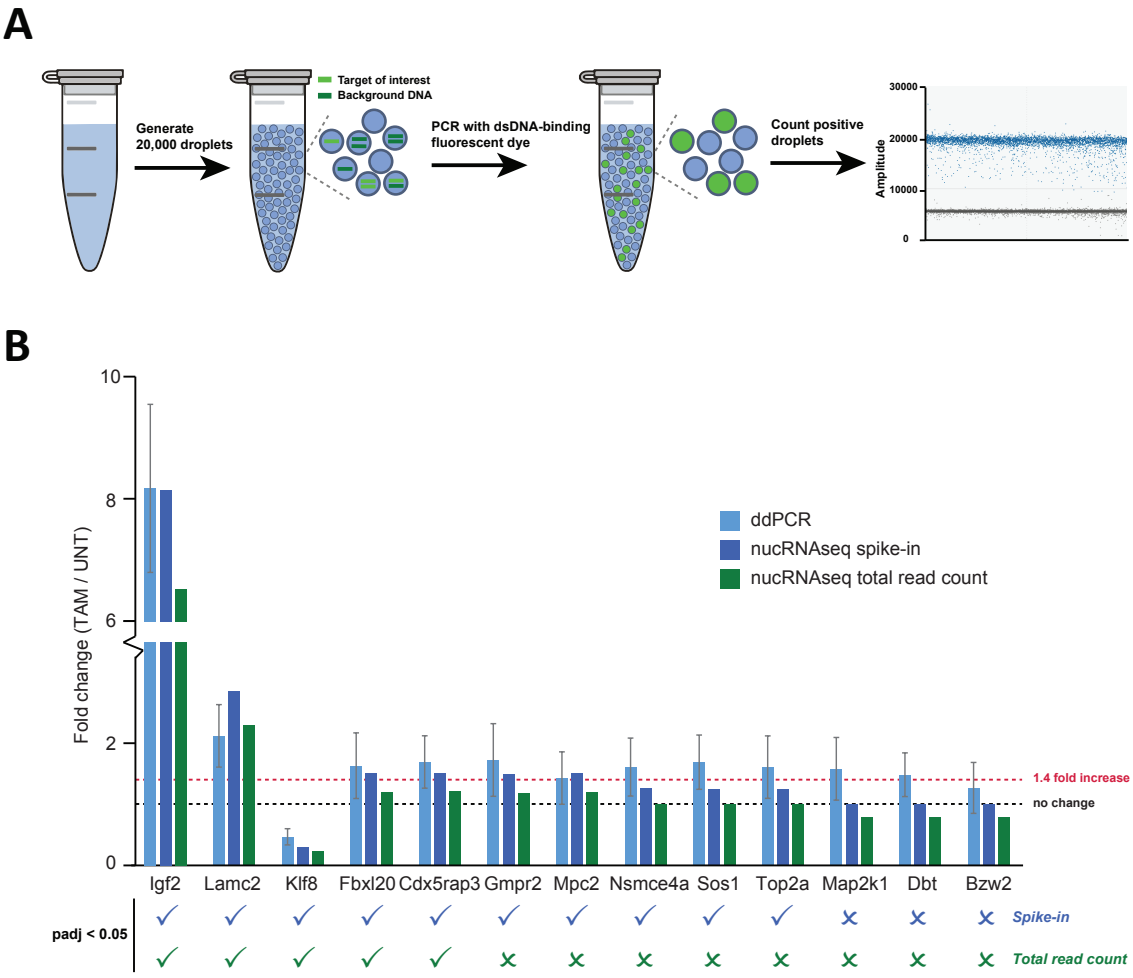


Figure 4.9: Digital droplet PCR confirms the widespread increase in transcription
(A) Schematic of the ddPCR method. The PCR reaction is partitioned into 20,000 droplets, a proportion of which contain the DNA target sequence. PCR amplification is performed within these droplets, in the presence of a dsDNA intercalating dye. Fluorescent droplets, which contain the DNA target molecule, are counted, and Poisson statistics applied to calculate the absolute concentration of target DNA in the reaction volume.
(B) ddPCR analysis of cDNA prepared from *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, with or without 96 h tamoxifen treatment to induce KDM2A/B removal. Bars represent fold change in target cDNA concentration following tamoxifen treatment (light blue), and the fold change values calculated from analysis of nucRNA-seq data using calibration to spike-in (dark blue) or normalisation to total library size (green) are shown for reference. The significance of these changes nucRNA-seq changes was calculated using DESeq2, and significant changes are indicated below (tick corresponds to padj<0.05). Horizontal lines represent no change in expression (black), and the 1.4 fold change threshold used for nucRNA-seq analysis (red). Experiments were performed in triplicate, and error bars show standard deviation.

4.2.6 Transcription upregulation is not limited to polycomb target genes

Having validated the widespread increase in transcription that occurs following removal of KDM2A/B from chromatin using ddPCR and intron-specific analysis, I next explored the characteristics of the genes that significantly changed in expression. Genes that significantly increased in expression following removal of KDM2A/B were more highly expressed in wild-type mESCs than genes that significantly decreased in expression, although this difference was subtle (Figure 4.10 A). I compared the levels of a variety of chromatin features at the TSS of these gene sets, using datasets from wild-type cells. This revealed that the promoters of significantly decreasing genes were moderately depleted, and the promoters of significantly increasing genes enriched, for BioCAP signal (Figure 4.10 B). This reflects the enrichment of non-CGI genes in the significantly decreasing set (Figure 4.10 C). The decrease in transcription of a subset of non-CGI genes was likely a secondary effect of KDM2A/B removal, as KDM2A/B are almost entirely depleted at non-CGI promoters in mESCs (Supplementary Figure A.4). Conversely, the enrichment of CGI genes in the significantly increasing set is consistent with transcription upregulation as a direct result of KDM2A/B removal from CGI gene promoters. Levels of H3K4me3 were also depleted at the TSS of the significantly decreasing gene set and enriched in the significantly increasing set, which reflects their relatively low and high starting transcription levels, respectively (Figure 4.10 A). The TSS enrichment of KDM2A paralleled BioCAP signal, as previously reported (Blackledge et al. 2012). In contrast, KDM2B was highly enriched at the TSS of genes that significantly increased in expression following tamoxifen treatment, along with the polycomb proteins RING1B and SUZ12 (Figure 4.10 B). This TSS enrichment of KDM2B, RING1B and SUZ12 is characteristic of polycomb targets. Indeed the significantly increasing set was enriched for polycomb target genes defined by high RING1B and SUZ12 occupancy (Figure 4.10 C).

I was intrigued by this potential link between the polycomb repressive system and the set of genes that signficantly increased in expression following removal of KDM2A/B: this gene set was also moderately enriched for H3K4me3 in wild-type mESCs and was relatively highly expressed, features which are not typical of polycomb repressed genes. To explore this link further I compared the profile of gene expression changes

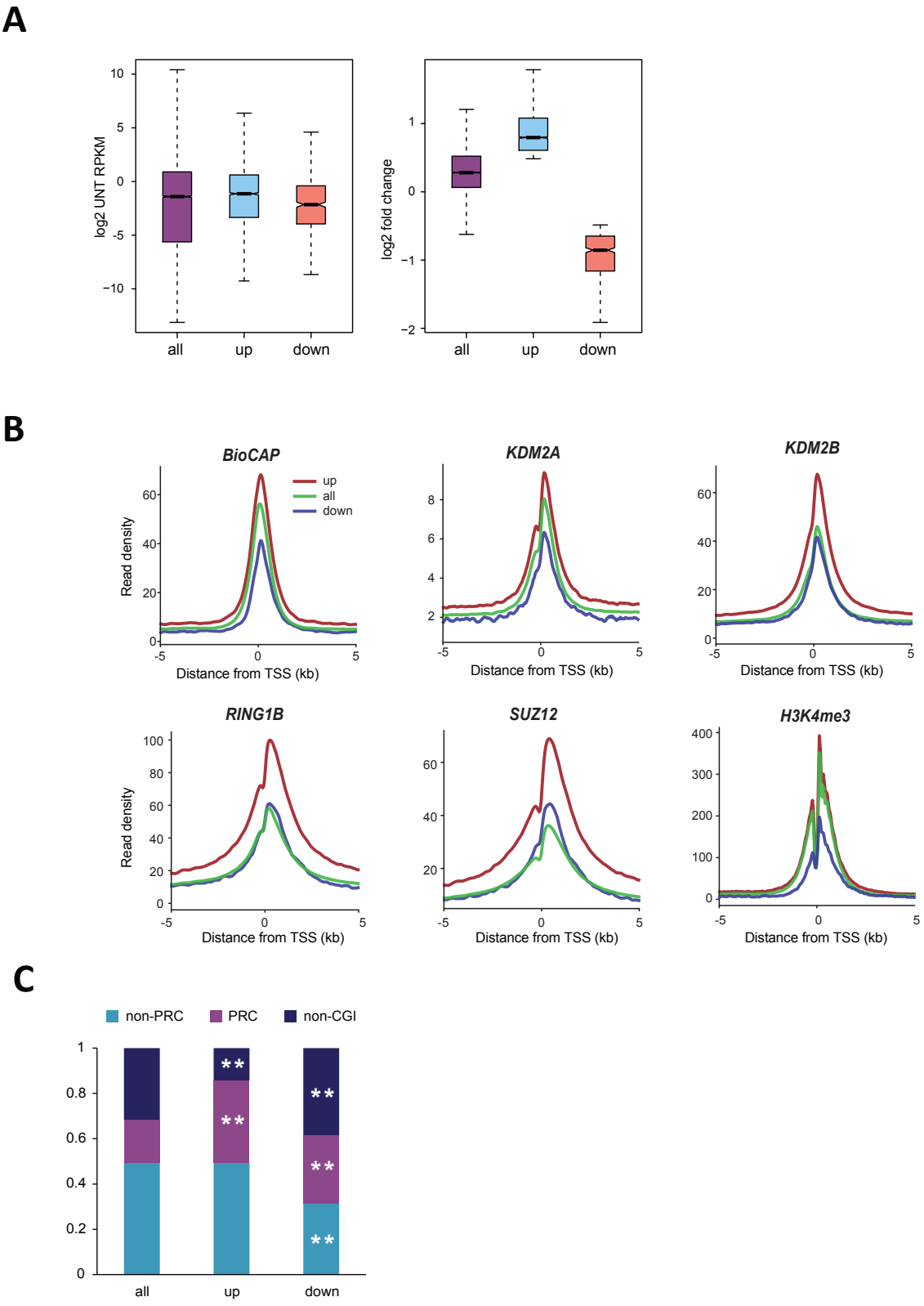


Figure 4.10: Genes that significantly increase in expression are enriched for polycomb targets

(A) Box plots showing starting expression level (log2 UNT RPKM) and log2 fold change in gene expression following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs to induce removal of KDM2A/B, for all genes (n=20633, all) and genes that significantly increased (n=4528, up) or decreased (n=531, down) in expression.
(B) Metaplots of KDM2A, KDM2B, RING1B, SUZ12 and H3K4me3 ChIP signal and BioCAP signal at the TSS for all genes, and for significantly increasing or decreasing subsets (Datasets: Long et al. 2013; Farcas et al. 2012; Rose et al. 2016; Brown et al. 2017).
(C) Bar chart showing the proportion of genes that belong to non-CGI and either polycomb-target (PRC) or non-polycomb target (non-PRC) CGI classes, for all genes and the significantly increasing and decreasing subsets. PRC target genes were defined by RING1B and SUZ12 occupancy at the TSS (ChIP-seq datasets: Rose et al. 2016). Asterisks represent empirical p-value (* represents p<0.01, ** represents p<0.001), relative to all transcripts.

following removal of KDM2A/B with the changes following tamoxifen-induced PCGF1 or RING1A/B knockout in *Pcgf1*$^{fl/fl}$ (Supplementary Figure A.5) and *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ (Supplementary Figure A.6) mESCs respectively (Figure 4.11 A). PCGF1 is a component of the PRC1.1 variant complex and is required for its ubiquitin ligase activity, while RING1A/B form the catalytic core of all PRC1 complexes. Calibrated nucRNA-seq analysis revealed that PCGF1 knockout led to the deregulation of nearly 1000 genes, while RING1A/B knockout led to the deregulation of almost 4000 genes (Figure 4.11 A). These significant changes in gene expression were biased towards transcriptional upregulation, as expected following loss of PRC1-mediated gene repression. The smaller number of genes that significantly increased in expression following PCGF1 knockout compared to RING1B knockout (783 compared to 3368), reflects the fact that the PRC1.1 complex is required to maintain the repressed state of only a subset of polycomb target genes (Blackledge et al. 2014). KDM2B is a component specifically of this PRC1.1 complex, and is required for its targeting to CGI gene promoters (Farcas et al. 2012; He et al. 2013; Wu et al. 2013; Blackledge et al. 2014). Therefore, if the gene expression changes following removal of KDM2A/B were solely due to loss of PRC1.1-mediated gene repression, then they should phenocopy the changes following PCGF1 knockout. However, more than five times as many genes significantly increased in expression following KDM2A/B removal compared to following PCGF1 knockout (Figure 4.8), suggesting that another polycomb-independent mechanism is also at play. Importantly, expression of the polycomb proteins SUZ12 and RING1B was unperturbed following removal of KDM2A/B from chromatin (Supplementary Figure A.7), as previously reported for the *Kdm2b-CXXC*$^{fl/fl}$ line (Blackledge et al. 2014), while

an increase in PCGF1 protein levels can be explained by a significant increase in PCGF1 transcription. Therefore, it is unlikely that removal of KDM2A/B indirectly perturbed polycomb-mediated gene repression more broadly by altering the abundance of PRC complexes. Indeed, the number of significantly increasing genes following removal of KDM2A/B (4528 genes) was greater than the number following loss of all PRC1 targeting in *Ring1a*$^{-/-}$*Ring1b*$^{-/-}$ mESCs (3368 genes), further evidence that the gene expression changes following KDM2A/B removal cannot be solely explained by a loss of polycomb-dependent gene repression.

I proceeded to compare the gene expression changes in more detail. A large proportion of the genes that significantly increased in expression following either PCGF1 or RING1B knockout were classified as polycomb target genes (Figure 4.11 B), as expected. This enrichment was far more pronounced than in the set of genes that significantly increased in expression following KDM2A/B removal, which again suggested that the transcriptional changes in this line could not be fully explained by a loss of polycomb-mediated gene repression. Less than half of the upregulated genes following RING1A/B knockout were also significantly upregulated following KDM2A/B removal (Figure 4.11 C). In contrast, nearly 80% of the significantly upregulated genes following PCGF1 knockout were also significantly upregulated following KDM2A/B removal. Extension of this comparison to all genes revealed a very poor correlation between the gene expression changes following RING1A/B knockout and following KDM2A/B removal, whereas the changes following PCGF1 knockout showed a moderate positive correlation (Figure 4.11 D). Collectively, these data confirmed that genes that require PRC1.1 to maintain their repressed state increased in expression following removal of KDM2A/B. However, the more global increase in transcription that occurred was a distinct phenomenon, as a pronounced increase in transcription of non-polycomb target CGI genes, as well as polycomb target genes, was observed only following removal of KDM2A/B (Figure 4.11 E). The relatively minor average change in expression of non-CGI genes was consistent with the majority of transcription changes being a direct result of KDM2A/B removal from CGI promoters.

Gene ontology analysis of the genes that significantly increased in expression following KDM2A/B removal revealed an enrichment for a variety of developmental terms typical

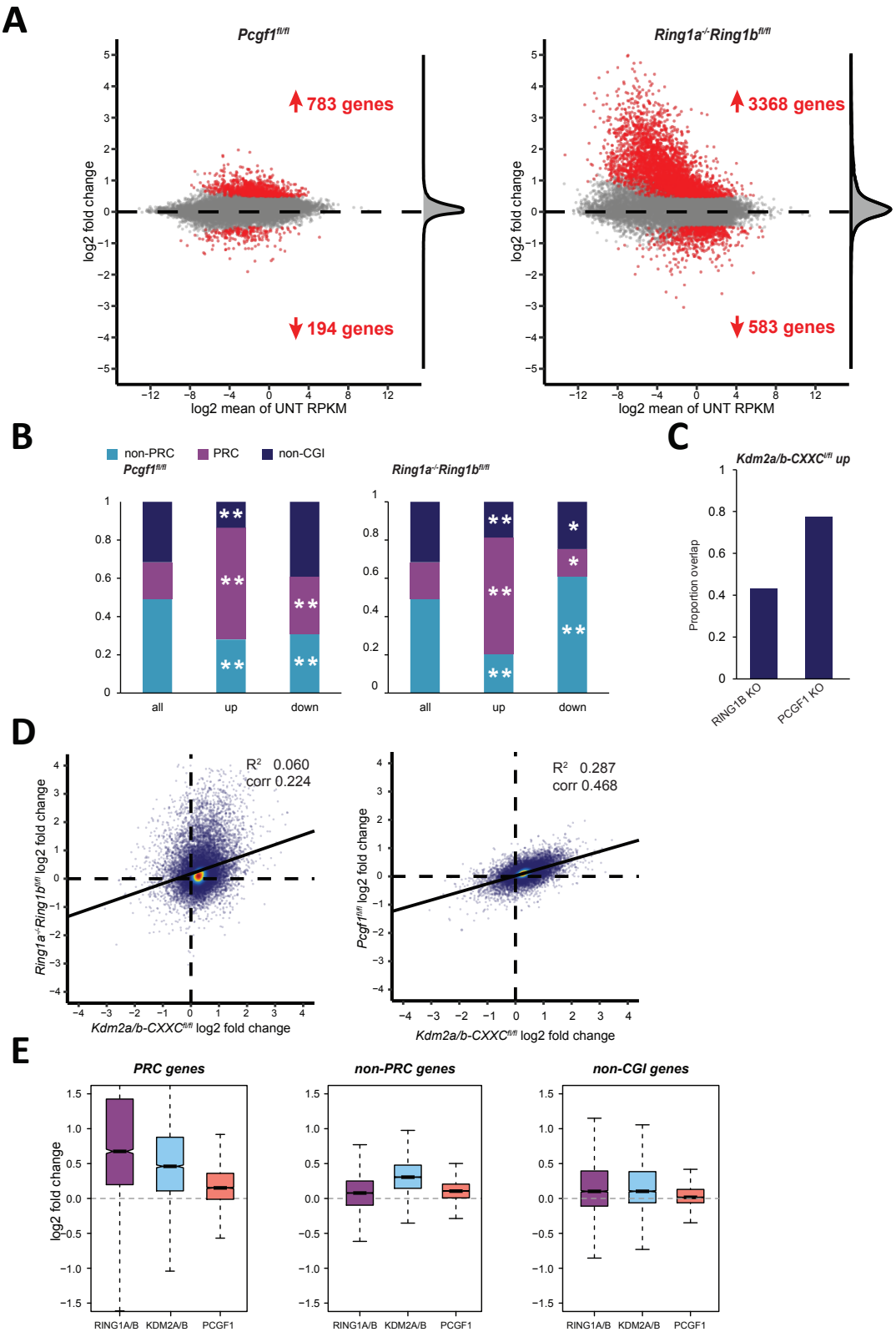


Figure 4.11: Widespread transcription upregulation is not explained by loss of polycomb-mediated gene repression

(A) Scatter plots of log2 mean nucRNA-seq reads per kilobase per million mapped reads (RPKM) in *Pcgf1*$^{fl/fl}$ (left) or *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ (right) mESCs for a non-redundant mm10 gene set (n=20633, three biological replicates), plotted against log2 fold change in nucRNA-seq signal following tamoxifen treatment to induce PCGF1 or RING1A/B knockout respectively. Libraries were calibrated to *Drosophila* spike-in. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change>1.4). A density plot of the log2 fold change in nucRNA-seq signal is shown to the right.
(B) Bar charts showing the proportion of genes that belong to non-CGI and either polycomb-target (PRC) or non-polycomb target (non-PRC) CGI classes, for all genes and the subsets that significantly increased or decreased in expression following PCGF1 or RING1A/B knockout. PRC target genes were defined by RING1B and SUZ12 occupancy at the TSS (ChIP-seq datasets: Rose et al. 2016). Asterisks represents empirical p-value (* represents $p<0.01$, ** represents $p<0.001$), relative to all transcripts.
(C) Bar chart showing the proportion of genes that significantly increased in expression following PCGF1 or RING1A/B knockout that also significantly increased in expression following removal of KDM2A/B.
(D) Scatter plots comparing the log2 fold change in nucRNA-seq signal following removal of KDM2A/B with the log2 fold change following RING1A/B or PCGF1 knockout. Solid lines represent linear regression, with the coefficient of determination (R^2) and Spearman correlation coefficient (corr) annotated.
(E) Box plots showing log2 fold change in gene expression following RING1A/B or PCGF1 knockout or KDM2A/B removal, for polycomb target (PRC), non-polycomb target (non-PRC) and non-CGI genes.

of polycomb targets (Figure 4.12 A), consistent with a subset of polycomb genes becoming derepressed. In contrast, the subset of these significantly increasing genes that have CGI promoters and were not classified as polycomb targets (2243 out of 4528 genes) was not enriched for developmental terms, but instead for a variety of basic cellular processes (Figure 4.12 B). This may reflect the generalised and widespread transcription upregulation that occurs following KDM2A/B removal, including of 'housekeeping' genes. The significantly decreasing genes were enriched for only two terms, relating to ion transport and response to external stimuli (Figure 4.12 C).

4.2.7 Genes with low expression are more strongly upregulated following KDM2A/B removal

Given that a reduction in polycomb-mediated targeting appeared insufficient to explain the widespread increase in transcription following KDM2A/B removal, I investigated the correlation of the log2 fold change in gene expression with the levels of a variety of promoter-associated chromatin features (Figure 4.13 A). This allowed a more detailed exploration of the relationship between these features and the extent of transcriptional change, extending the more qualitative conclusions that could be drawn by comparing TSS metaplots of these features (Figure 4.10 B). The extent of upregulation of significantly increasing genes showed a moderate positive correlation with polycomb-

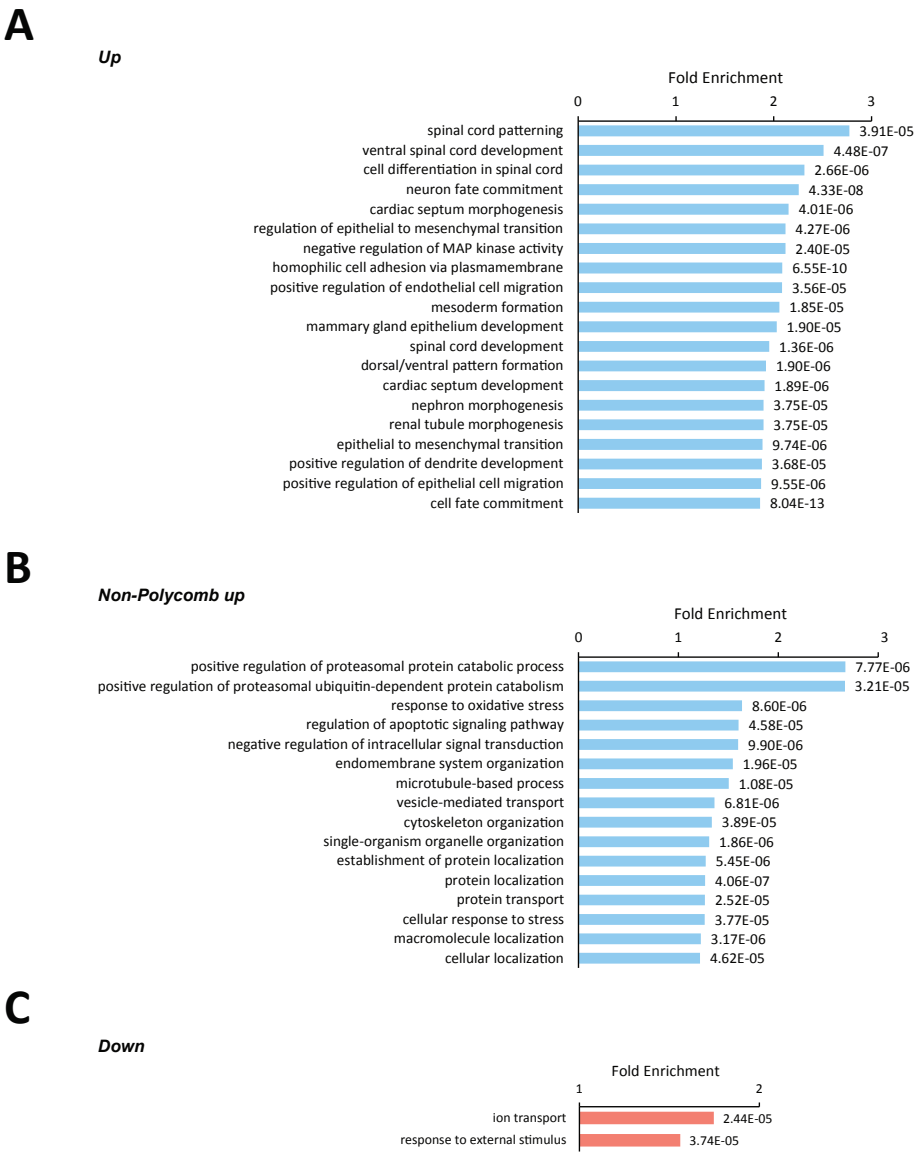


Figure 4.12: Significantly increasing non-polycomb target genes are involved in a range of basic cellular processes

(A) Gene ontology analysis of genes that significantly increased in expression following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs to induce removal of KDM2A/B (n=4528). BP FAT analysis with FDR<0.1 cutoff was used, and p-values are annotated. The 20 most enriched terms are shown.
(B) As A, but for significantly increasing non-polycomb target genes (n=2243).
(B) As A, but for significantly decreasing genes (n=531).

associated features (KDM2B, RING1B, SUZ12, H2AK119ub1 and H3K27me3 TSS enrichment), as previously shown by the enrichment of KDM2B, RING1B and SUZ12 in the TSS metaplots. There was also a moderate correlation between the log2 fold change in expression following removal of KDM2A/B and the log2 fold change in expression following RING1A/B knockout for this significantly increasing gene set, which can be explained by the upregulation of a subset of polycomb target genes. In contrast, the changes in gene expression correlated poorly with levels of BioCAP and KDM2A, which mark all CGI promoters, indicating that the transcriptional change was not simply proportional to the strength of the CGI and therefore the amount of KDM2A that was bound to the TSS. There was a weak negative correlation with features enriched at active promoters (H3K4me3, H3K27ac and H3K9ac), and a moderate negative correlation with starting transcription level, which both indicate that strong transcription upregulation was somewhat biased towards lowly transcribed genes.

Strong upregulation of lowly transcribed genes is characteristic of the loss of polycomb-mediated gene repression, however this correlation held true even for the subset of the significantly increasing genes that are not polycomb target genes, suggesting that lowly transcribed genes were also more sensitive to loss of the polycomb-independent mode of KDM2A/B mediated gene repression. This trend could be seen on a scatter plot of log2 fold change in gene expression against log2 expression level, where both polycomb targets (purple) and non-polycomb targets (light blue) showed a moderate degree of negative correlation (Figure 4.13 B). This comparison became clear when dividing non-polycomb targets into deciles based on their log2 fold change in expression (Figure 4.13 C).

In contrast, the extent of transcriptional repression of significantly decreasing genes did not correlate with polycomb-associated features, but instead showed a moderate correlation with BioCAP signal, KDM2A, KDM2B and H3K4me3 TSS enrichment and gene expression level. This likely reflected the more dramatic downregulation of non-CGI genes, which are depleted of these promoter-associated features, compared to CGI genes (Supplementary Figure A.8).

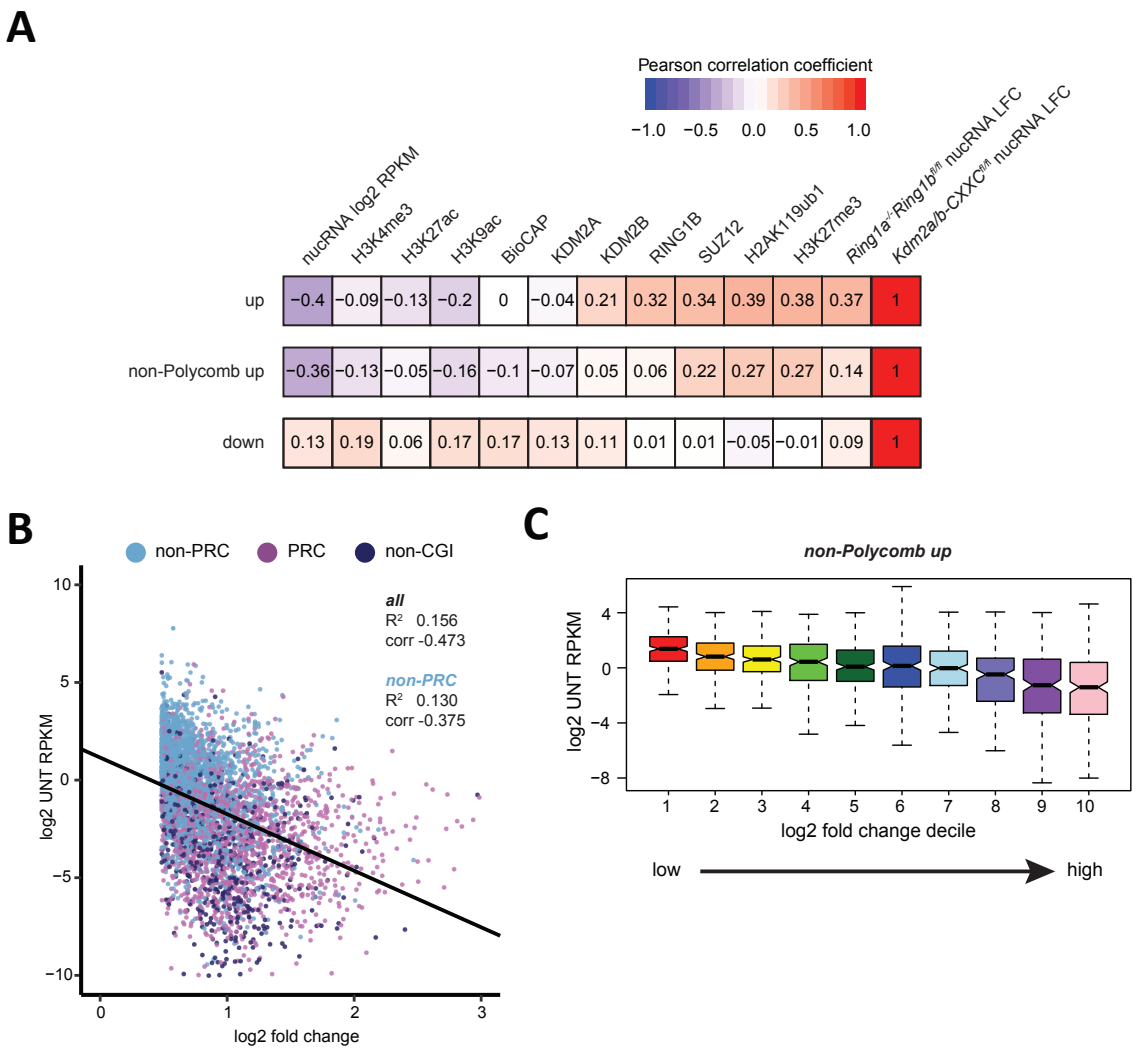


Figure 4.13: Negative correlation between gene expression level and extent of transcription upregulation
(A) Matrix of Pearson correlation coefficients, comparing the log2 fold change (LFC) in gene expression following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs with expression level (log2 UNT RPKM), the LFC in expression following tamoxifen treatment of *Ring1a*$^{-/-}$ *Ring1b*$^{fl/fl}$ mESCs, and the enrichment of a range of promoter-associated features at the 2kb region flanking the TSS. Correlations for all signficantly increasing genes (up, n=4528), significantly increasing CGI genes that are not polycomb targets (non-polycomb up, n=2243) and all significantly decreasing genes (down, n=531) are shown.
(B) Scatter plot of log2 fold change in nucRNA-seq signal following tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs plotted against log2 mean nucRNA-seq reads per kilobase per million mapped reads in untreated cells (log2 UNT RPKM). Points corresponding to polycomb target (PRC), non-polycomb target (non-PRC) and non-CGI transcripts are coloured purple, light blue and dark blue respectively.
(C) Box plot of the starting transcription level (log2 UNT RPKM) for non-polycomb target genes that significantly increased in expression following removal of KDM2A/B, divided into deciles based on the log2 fold change in gene expression.

4.3 Summary and discussion

Here I have examined the overall role played by KDM2A/B in the regulation of transcription. KDM2A/B are known to have both long isoforms, which catalyse

H3K36me2 demethylation via their JmjC catalytic domain, and short isoforms, which lack this activity but retain the ZF-CXXC domain required for CGI targeting and the more C-terminal domains (Tanaka et al. 2010; Long et al. 2013). I identified novel isoforms of KDM2B in mESCs, including an isoform that is expressed from an internal promoter and closely resembles KDM2B-SF (Figure 4.1 and Supplementary Figure A.2). In order to achieve removal of all KDM2A/B isoforms, I developed a *Kdm2a/b-CXXC*$^{fl/fl}$ mESC line, in which tamoxifen treatment led to deletion of the ZF-CXXC domains of KDM2A/B. Importantly, this domain is found in all identified KDM2A/B isoforms, and is required for KDM2A/B to bind to chromatin (Figure 4.3). Therefore, deletion of the ZF-CXXC domain led to loss of the chromatin-associated functions of KDM2A/B. The fully conditional nature of the *Kdm2a/b-CXXC*$^{fl/fl}$ line enabled the study of the initial transcriptional effects following KDM2A/B removal.

I generated a custom calibrated nucRNA-seq analysis pipeline (Figure 4.7), which revealed that removal of KDM2A/B led to an increase in transcription of the majority of genes and a significant increase in transcription of more than 20% of all transcripts (Figure 4.8). This suggests that KDM2A/B act broadly at CGI promoters to constrain transcription. Given that KDM2A/B-LF made a minimal contribution to the regulation of gene expression (Section 3), this indicates that KDM2A/B-SF, and therefore demethylase-independent activities of KDM2A/B, play the predominant role. I used ddPCR to validate the transcriptional upregulation observed in *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs. This confirmed that the calibrated nucRNA-seq pipeline provided an accurate representation of the transcriptional changes (Figure 4.9).

Restricting nucRNA-seq analysis to reads that mapped to intronic regions revealed that there was an increase in abundance of the majority of nascent unspliced transcripts, as well as spliced mRNAs, following removal of KDM2A/B (Supplementary Figure A.3). This added further evidence that removal of KDM2A/B had a direct effect on transcription. It is unclear why there was a more dramatic increase in spliced transcripts. This might have been a result of more sensitive analysis due to the greater proportion of reads mapping to exons, particularly as the DESeq2 package used tends to shrink the log2 fold change values for transcripts with low read counts. Alternatively this might have been a reflection of an over accumulation of spliced transcripts in the nucleus

following removal of KDM2A/B, which may have resulted from an increase in splicing or compromised mRNA nuclear export. The large proportion of spliced transcripts could also reflect contamination of the nucRNA sample with some cytoplasmic mRNA, which might have increased as bulk mRNA levels increased following KDM2A/B removal. The direct impact of KDM2A/B removal on nascent transcription could be studied using Global Run-On Sequencing (Core et al. 2008), which maps the binding profile of transcriptionally engaged RNAPII.

It would also be interesting to study the transcriptional response following a more acute loss of KDM2A/B protein. Study of the demethylase-independent isoforms of KDM2A/B in isolation could be achieved by inserting an AID or dTAG tag upstream of the unique first exons of KDM2A/B-SF and KDM2B isoform B, while C-terminal tagging would enable acute depletion of all KDM2A/B protein. These systems rely on ligand-induced ubiquitylation of the tagged protein by an exogenous or endogenous E3 ubiquitin ligase, respectively, leading to its proteasomal degradation (Nishimura et al. 2009; Nabet et al. 2018). This results in rapid depletion of the target protein within minutes or hours, in contrast to the 96 h tamoxifen treatment used to induce ZF-CXXC domain deletion in the *Kdm2a/b-CXXC*$^{fl/fl}$ cell line, offering the opportunity to study the very early transcriptional changes that follow a sudden depletion of the proteins.

Given the well-characterised role of KDM2B in targeting the PRC1.1 complex to chromatin, it was important to explore the extent to which the widespread increase in transcription following removal of KDM2A/B could be explained by loss of this polycomb targeting activity. Loss of PRC1.1 targeting, via PCGF1 knockout, also leads to the significantly increased expression of a large number of transcripts (Figure 4.11). However, more than five times as many genes significantly increased in expression following removal of KDM2A/B, and the increase in transcription of the majority of transcripts observed following KDM2A/B removal does not occur as a result of PCGF1 knockout. Therefore, removal of KDM2A/B leads to a more dramatic effect on transcription that does not phenocopy loss of the PRC1.1 complex.

One possible explanation is that KDM2B plays an important role in targeting other PRC1 complexes in addition to PRC1.1 to gene promoters. Removal of KDM2B might therefore

lead to impaired repression of a larger proportion of polycomb target genes. Indeed KDM2B was pulled down in a PCGF3 immunoprecipitation in 293T cells, although the signal was weaker than for PCGF1, suggesting that KDM2B might assemble in PRC1.3 complexes (Gao et al. 2012). However, this interaction between KDM2B and PCGF3 is questionable, as KDM2B immunoprecipitation identified only PCGF1, and not PCGF3, as an interacting partner in mESCs (Farcas et al. 2012), and an in vitro study showed that PCGF3 does not interact with KDM2B using the same binding mode as PCGF1 (Wong et al. 2016). If KDM2B were responsible for the majority of all PRC1 targeting to chromatin, then removal of KDM2B should reproduce the dramatic loss of H2AK119ub1 that results from RING1A/B knockout. However, only a relatively modest reduction in H2AK119ub1 levels was observed following deletion of the KDM2B ZF-CXXC domain compared to following RING1A/B knockout (Blackledge et al. 2014), suggesting that KDM2B is not responsible for the majority of PRC1 targeting.

To further explore the extent to which the transcription upregulation following KDM2A/B removal could be explained by loss of polycomb-mediated gene repression, I compared the gene expression changes to those following RING1A/B knockout. This revealed differences in the gene expression change profiles and poor correlation between the changes (Figure 4.11). Removal of KDM2A/B was unique in leading to widespread upregulation of both polycomb target and non-target genes. Overall, this analysis reveals an important role for KDM2A/B in transcription repression that is independent of the polycomb system.

The moderate increase in transcription that occurred for the majority of genes following removal of KDM2A/B indicates that KDM2A/B binding to CGI promoters does not function as a simple on/off switch for transcription. Indeed, given that KDM2A/B are enriched at all CGI gene promoters, cells would not survive if they exerted a dominant transcription repressive effect. Instead, it seems that the generic targeting of KDM2A/B to gene promoters serves to modestly constrain transcription. This activity might function to increase the threshold of activating signal from lineage-specific transcription factors that is required to achieve robust gene induction, which might prevent spurious gene activation as a result of more moderate fluctuations in transcription factor binding, contributing to the robustness of lineage-specific differentiation programmes. There was

a negative correlation between starting expression level and the extent of transcription upregulation following removal of KDM2A/B, indicating that genes with low expression are more susceptible to loss of the transcription constraining activity of KDM2A/B. This might result from a reduction in the threshold of low-level activator signals that are required to establish more robust transcription of these genes, leading to a relatively large fold change in expression. Loss of this constraining activity may have a relatively small effect on genes that are already strongly expressed and therefore already subject to strong transcription activation signals.

In conclusion, removal of KDM2A/B from chromatin led to a widespread increase in transcription, including of genes that are not targeted by the polycomb system. This suggests that KDM2A/B act broadly at CGI gene promoters to constrain transcription. This activity may be shared equally between KDM2A and KDM2B, or may be predominantly or entirely exerted by only one of these paralogue. In Chapter 5 I will therefore explore their relative roles, by investigating the transcription changes following conditional removal of KDM2A or KDM2B individually from chromatin. In Chapter 6 I will go on to explore the potential molecular mechanisms by which KDM2A/B may exert their transcription-limiting activity.

5

KDM2B plays the predominant role in transcription repression

5.1 Introduction

Conditional removal of KDM2A/B from chromatin revealed a widespread role for KDM2A/B in constraining transcription (Chapter 4). However, it is unclear from this analysis whether KDM2A and KDM2B act as functionally redundant transcriptional repressors, or whether one of these two homologues plays a predominant role. Although conditional loss of KDM2B binding from chromatin in *Kdm2b-CXXC*$^{fl/fl}$ mESCs was previously reported to lead to the modest upregulation of a subset of CGI genes (Blackledge et al. 2014), this analysis was not calibrated to a spike-in genome, therefore a more global increase in transcription may have been overlooked. Similarly, constitutive knockdown of KDM2A in human cervical carcinoma cells also led to subtle transcription changes for a subset of genes, however this analysis was also limited by the lack of spike-in calibration (Blackledge et al. 2010). In addition, the deregulated genes may reflect secondary effects of KDM2A depletion, as CGI and non-CGI genes were equally affected. This study noted that KDM2A was highly refractory to depletion in a variety of mESC and other primary cell lines, which suggests that KDM2A may play an essential role in these cell types. Therefore, the individual roles played by KDM2A and KDM2B in transcription regulation remain unclear.

FBXL19 is the third member of the KDM2 family, and is a paralogue KDM2A/B-SF (Katoh and Katoh 2004; Long et al. 2013). FBXL19 lacks a JmjC demethylase domain, but like KDM2A/B is recruited to CGI throughout the genome via its ZF-CXXC domain (Lee et al. 2017; Dimitrova et al. 2018). FBXL19 plays an essential role during development, as deletion of its ZF-CXXC domain in mice leads to embryonic

lethality at approximately E10.5 (Dimitrova et al. 2018). This highlights the importance of the demethylase-independent roles of the KDM2 family during embryogenesis. A recent study revealed that removal of FBXL19 from chromatin prevents the activation of a subset of developmental genes during cell lineage commitment, due to the loss of FBXL19-mediated recruitment of CDK8-mediator to these gene promoters which is important for promoter priming (Dimitrova et al. 2018). While conditional deletion of the ZF-CXXC domain of FBXL19 alone in mESCs does not lead to transcriptional deregulation (Dimitrova et al. 2018), it is possible that simultaneous loss of FBXL19, KDM2A and KDM2B might reveal a functionally redundant role for FBXL19 in transcription regulation that would otherwise be masked by continued targeting of KDM2A/B to chromatin.

Here I aimed to directly examine the individual contributions of KDM2A and KDM2B to the widespread increase in transcription that occurs following their simultaneous removal from chromatin. I performed calibrated nucRNA-seq analysis using cell lines in which the ZF-CXXC domain of either KDM2A or KDM2B was conditionally deleted, allowing the effect of KDM2A removal from chromatin to be studied in a system where KDM2B binding remains intact, and vice versa. In addition, I aimed to investigate the potential functional redundancy of FBXL19 with KDM2A/B. I generated a mESC line in which the ZF-CXXC domains of FBXL19, KDM2A and KDM2B could be simultaneously deleted, allowing the effect of simultaneous removal of all three paralogues be studied for the first time.

5.2 Results

5.2.1 Conditional removal of KDM2A from chromatin

Having uncovered a widespread role for KDM2A/B in transcriptional repression, I next wanted to investigate the individual contribution made by KDM2A. In order to study the role played by KDM2A in isolation, I generated a conditional cell line in which I could remove KDM2A from chromatin by inducing deletion of the KDM2A ZF-CXXC domain. I inserted loxP sites flanking the exon encoding the majority of this domain, using the

same CRISPR targeting strategy used to generate the *Kdm2a/b-CXXC*$^{fl/fl}$ line (Figure 4.2 A) and targeting an mESC line in which tamoxifen-inducible Cre recombinase is expressed from the *Rosa26* locus (supplied by Prof Haruhiko Koseki). Having identified putative homozygous clones for upstream and downstream loxP sites, I used western blotting to validate that 96 h tamoxifen treatment of the generated *Kdm2a-CXXC*$^{fl/fl}$ line led to a decrease in the molecular weight of both KDM2A-LF and KDM2A-SF proteins, corresponding to efficient deletion of the KDM2A ZF-CXXC domain (Figure 5.1 A). In contrast, protein levels of KDM2B-LF, KDM2B-SF and FBXL19 were unperturbed following tamoxifen treatment (Figure 5.1 B). Therefore, this *Kdm2a-CXXC*$^{fl/fl}$ line could be used to study the effect of removal of KDM2A from CGI promoters in isolation.

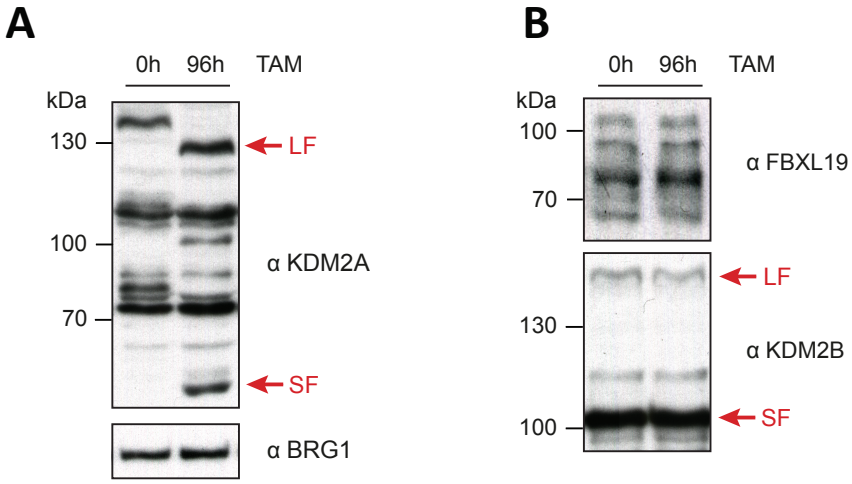


Figure 5.1: Conditional deletion of the ZF-CXXC domain of KDM2A
(A) Western blot showing deletion of the KDM2A ZF-CXXC domain following 96 h tamoxifen (TAM) treatment of *Kdm2a-CXXC*$^{fl/fl}$ mESCs, leading to a size shift for both KDM2A-LF and KDM2A-SF. BRG1 is a loading control.
(B) Western blot showing unperturbed expression of FBXL19 and KDM2B following TAM treatment of *Kdm2a-CXXC*$^{fl/fl}$ mESCs. Loading control: BRG1, shown in (A).

5.2.2 KDM2A chromatin occupancy is not required to maintain normal transcription

Having generated a mESC line in which I could conditionally remove KDM2A from chromatin, I proceeded to use this line to investigate the role played by KDM2A in transcription regulation. I performed calibrated nucRNA-seq using *Kdm2a-CXXC*$^{fl/fl}$ mESCs, with and without 96 h tamoxifen treatment to induce KDM2A removal. A genome-wide analysis revealed that there were virtually no changes in transcription

following tamoxifen treatment, with only 15 significantly upregulated and 3 significantly downregulated genes ($p<0.05$ and fold change>1.4, Figure 5.2 A). In addition, there was no skew in the distribution of the log2 fold change density (red, Figure 5.2 B), in stark contrast to the widespread upregulation observed following simultaneous loss of KDM2A/B (blue). Therefore, removal of KDM2A alone did not affect transcription, and did not recapitulate the widespread increase in transcription that occurred following simultaneous removal of KDM2A/B.

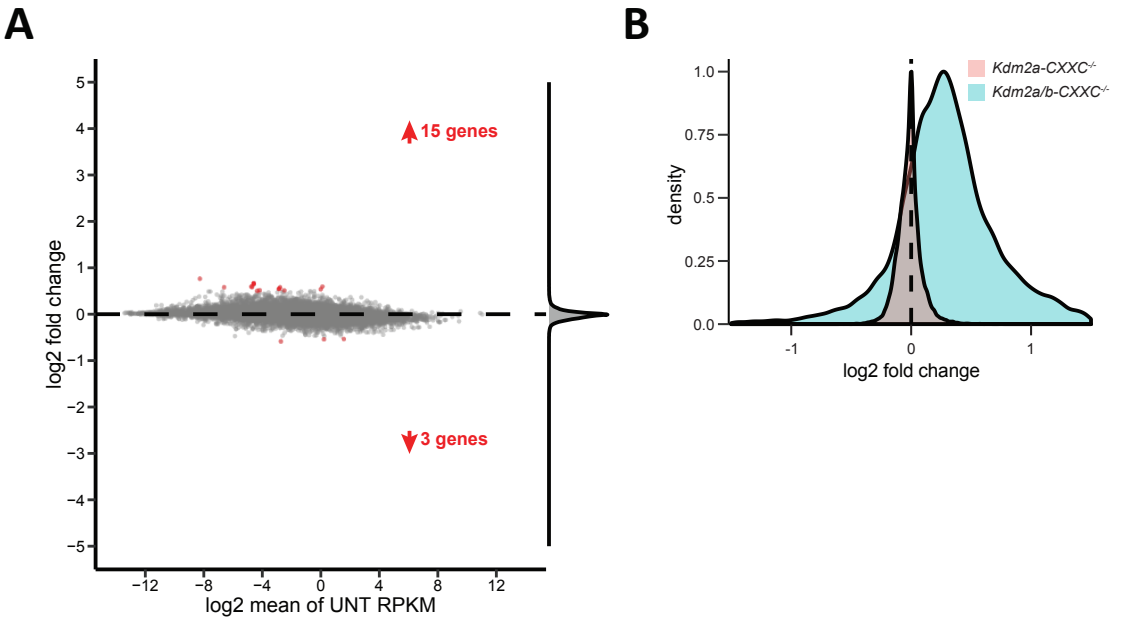


Figure 5.2: Removal of KDM2A leads to minimal perturbation of transcription
(A) Scatter plot of log2 mean nucRNA-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2a-CXXC*$^{fl/fl}$ mESCs for a non-redundant mm10 gene set (n=20633, three biological replicates), plotted against log2 fold change in nucRNA-seq signal following 96 h tamoxifen treatment to induce removal of KDM2A. Libraries were calibrated to *Drosophila* spike-in. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change>1.4). A density plot of the log2 fold change in nucRNA-seq signal is shown to the right.
(B) Density plot showing the log 2 fold change distribution of nucRNA-seq signal for all transcripts, following tamoxifen treatment of *Kdm2a-CXXC*$^{fl/fl}$ (red) or *Kdm2a/b-CXXC*$^{fl/fl}$ (blue) mESCs. The dashed line indicates no change in transcription.

5.2.3 Validation of the *Kdm2b-CXXC*$^{fl/fl}$ line

This nucRNA-seq analysis revealed that chromatin-bound KDM2A is not required for normal levels of transcription in mESCs. Given that KDM2A/B knockout led to a widespread increase transcription (Figure 4.8), I wondered whether KDM2B might instead play the predominant role in transcription regulation, or alternatively whether KDM2A/B act as functionally redundant transcription repressors. To address these

possibilities, I required a mESC line in which KDM2B binding to chromatin could be conditionally ablated. Therefore, I began by validating the previously reported *Kdm2b-CXXC*$^{fl/fl}$ mESC line (Blackledge et al. 2014), which I had used to generate the *Kdm2a/b-CXXC*$^{fl/fl}$ line (Figure 4.2 A). In this line an exon encoding most of the KDM2B ZF-CXXC domain is flanked by loxP sites, and the line also encodes ERT2-Cre recombinase, enabling tamoxifen-inducible deletion of this exon. Western blotting showed that 96 h tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs led to a decrease in the molecular weight of both KDM2B-LF and KDM2B-SF, corresponding to efficient deletion of the ZF-CXXC domain (Figure 5.3 A). Expression of KDM2B-LF appeared to increase following tamoxifen treatment, as observed in the *Kdm2a/b-CXXC*$^{fl/fl}$ line (Figure 4.2 B). This may be due to increased efficiency of salt extraction of KDM2B-LF following its removal from chromatin. Importantly, expression of KDM2A-LF, KDM2A-SF and the KDM2 family member FBXL19 was largely unperturbed by tamoxifen addition (Figure 5.3 B), indicating that tamoxifen treatment should not affect KDM2A or FBXL19 occupancy at CGI promoters. Therefore, this line could be used to study the effect of loss of CGI-targeting of KDM2B in isolation.

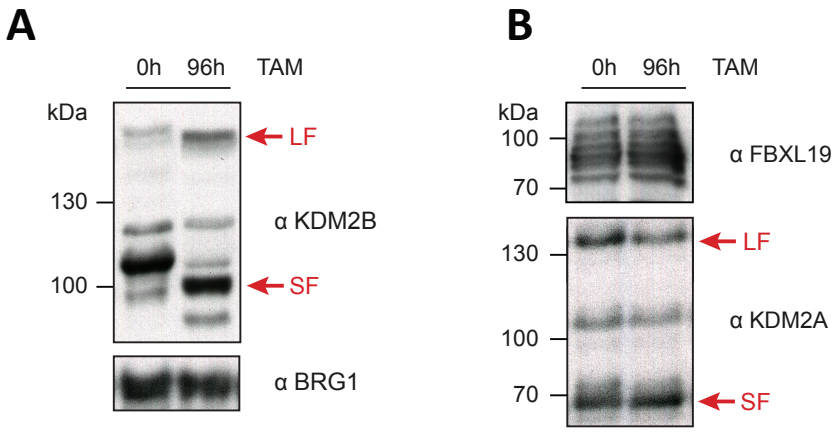


Figure 5.3: Conditional deletion of the ZF-CXXC domain of KDM2B
(A) Western blot showing deletion of the KDM2B ZF-CXXC domain following 96 h tamoxifen (TAM) treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs, leading to a size shift for both KDM2B-LF and KDM2B-SF. BRG1 is a loading control.
(B) Western blot showing largely unperturbed expression of FBXL19 and KDM2A following TAM treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs. Loading control: BRG1, shown in (A).

5.2.4 Removal of KDM2B results in widespread transcription upregulation

I proceeded to investigate the transcriptional changes following removal of KDM2B, by performing calibrated nucRNA-seq for *Kdm2b-CXXC*$^{fl/fl}$ mESCs with and without 96 h tamoxifen treatment to induce deletion of the KDM2B ZF-CXXC domain. A genome-wide analysis revealed that the expression of thousands of genes was perturbed, with 3024 significantly increasing and 220 significantly decreasing genes ($p<0.05$ and fold change>1.4, Figure 5.4 A). This was reminiscent of the transcriptional changes following simultaneous KDM2A/B ZF-CXXC deletion, with the majority of significant changes representing increases in transcription. Furthermore, when I compared the log2 fold change density profile between the *Kdm2b-CXXC*$^{fl/fl}$ and *Kdm2a/b-CXXC*$^{fl/fl}$ nucRNA-seq datasets, it became clear that the global upregulation of transcription was largely recapitulated following removal of KDM2B alone (Figure 5.4 B).

Next, I investigated the characteristics of the significantly changing genes. The

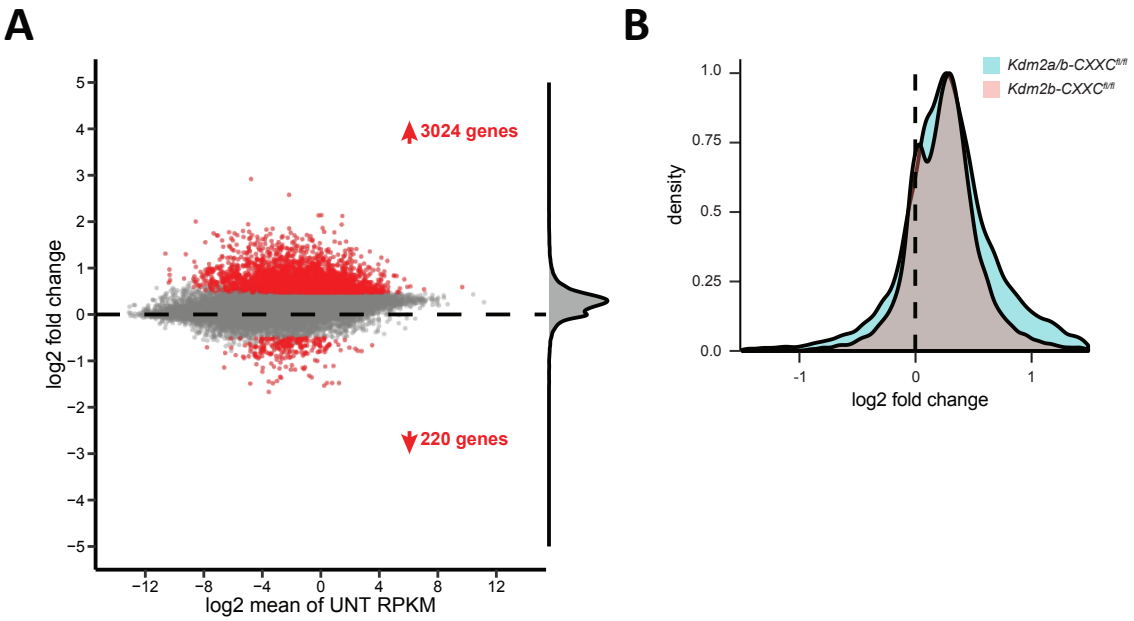


Figure 5.4: Removal of KDM2B leads to widespread transcription upregulation
(A) Scatter plot of log2 mean nucRNA-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2b-CXXC*$^{fl/fl}$ mESCs for a non-redundant mm10 gene set (n=20633, three biological replicates), plotted against log2 fold change in nucRNA-seq signal following 96 h tamoxifen treatment to induce removal of KDM2B. Libraries were calibrated to *Drosophila* spike-in. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change>1.4). A density plot of the log2 fold change in nucRNA-seq signal is shown to the right.
(B) Density plot showing the log 2 fold change distribution of nucRNA-seq signal for all transcripts, following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ (red) or *Kdm2a/b-CXXC*$^{fl/fl}$ (blue) mESCs. The dashed line corresponds to no change in transcription.

significantly upregulated genes were on average more highly expressed than the significantly downregulated genes (Figure 5.5 A). I compared the levels of a variety of chromatin features at the TSS of these gene sets, using datasets from wild-type cells. This revealed that BioCAP signal and KDM2A binding was moderately enriched at the TSS of significantly upregulated genes compared to all genes (Figure 5.5 B). This likely resulted from the relative depletion of non-CGI genes in this gene set (Figure 5.5 C), which are depleted of these CGI signatures. Levels of H3K4me3 were depleted in the downregulated gene set and enriched in the upregulated set, reflecting their relatively low and high starting transcription levels respectively. There was a more striking enrichment of KDM2B, RING1B and SUZ12 at the TSS of the set of significantly upregulated genes, indicating that this set was enriched for polycomb target genes. This was confirmed following categorization of a relatively large proportion of these genes as polycomb targets, defined by high RING1B and SUZ12 occupancy (Figure 5.5 C). Overall, the characteristics of the upregulated and downregulated gene sets were very similar to the corresponding sets following simultaneous removal of KDM2A/B, suggesting that similar groups of genes became deregulated in both experiments. Importantly, I observed widespread upregulation of both polycomb target and non-polycomb target CGI genes following loss of KDM2B only (Figure 5.5 C), suggesting that loss of KDM2B alone at least partially recapitulated the transcription upregulation following loss of the polycomb-independent mode of KDM2A/B mediated gene repression.

I was interested to compare the transcription changes that occurred following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ and *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs in more detail. More than 75% of the transcripts that were significantly upregulated following KDM2B removal were also significantly upregulated following removal of KDM2A/B together (2356 out of 3024 genes), and 75% of the significantly decreasing transcripts were also shared with the KDM2A/B significantly downregulated gene set (164 out of 220 genes, Figure 5.6 A). This indicated that the transcription changes following removal of KDM2B alone largely recapitulated the changes following removal of KDM2A/B together. Indeed, extension of this comparison to all genes revealed a good correlation between the gene expression changes in these two datasets (Figure 5.6 B, left).

Next I wanted to compare the gene expression changes following removal of KDM2B

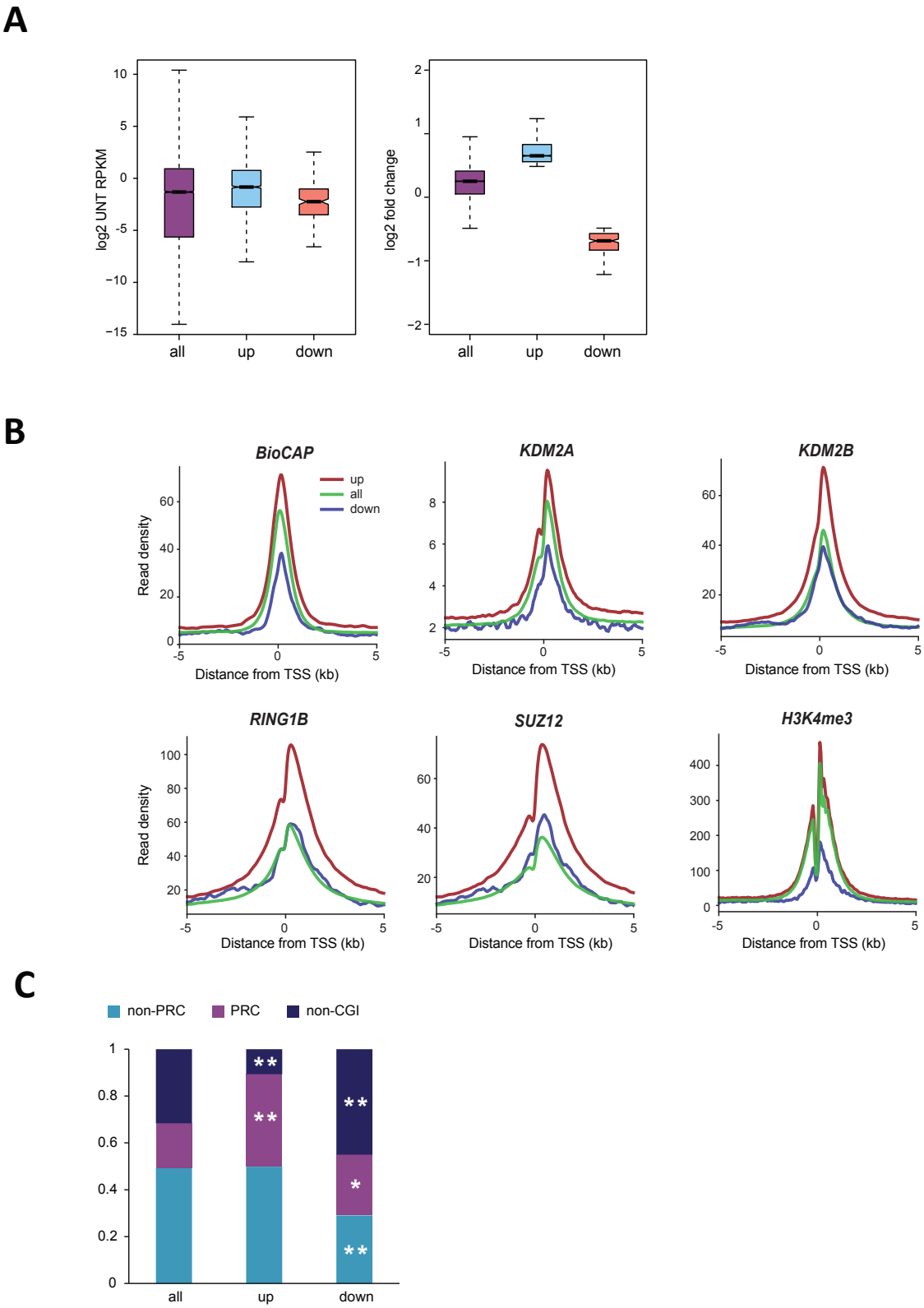


Figure 5.5: Significantly upregulated genes are enriched for polycomb targets

(A) Box plots showing starting expression level (log2 UNT RPKM) and log2 fold change in gene expression following 96 h tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs to induce removal of KDM2B, for all genes (n=20633, all) and genes that significantly increased (n=3024, up) or decreased (n=220, down) in expression.
(B) Metaplots of KDM2A, KDM2B, RING1B, SUZ12 and H3K4me3 ChIP signal and BioCAP signal at the TSS for all genes, and for significantly upregulated or downregulated subsets (Datasets: Long et al. 2013; Farcas et al. 2012; Rose et al. 2016; Brown et al. 2017).
(C) Bar chart showing the proportion of genes that belong to non-CGI and either polycomb-target (PRC) or non-polycomb target (non-PRC) CGI classes, for all genes and the significantly upregulated and downregulated subsets. PRC target genes were defined by RING1B and SUZ12 occupancy at the TSS (ChIP-seq datasets: Rose et al. 2016). Asterisks represents empirical p-value (* represents p<0.01, ** represents p<0.001), relative to all transcripts.

with those following loss of either PRC1.1-mediated gene repression, in *Pcgf1*$^{fl/fl}$ mESCs, or all PRC1-mediated gene repression, in *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ mESCs. The gene expression changes following KDM2B removal correlated very poorly with the changes following RING1A/B knockout, whereas they correlated moderately with the changes following PCGF1 knockout (Figure 5.6 B). This resembled the correlation of these datasets with the *Kdm2a/b-CXXC*$^{fl/fl}$ nucRNA-seq data (Figure 4.11). Fewer than 35% of the significantly upregulated genes following RING1A/B knockout were also significantly upregulated following KDM2B removal, whereas nearly 70% of the significantly upregulated genes following PCGF1 knockout were shared (Figure 5.6 C). This suggested that genes that require PRC1.1 for repression were upregulated following KDM2B removal. However, as observed following removal of KDM2A/B in combination, these only represented a subset the significantly increasing genes, indicating that KDM2B also limited the transcription of genes that are not targeted by the polycomb system.

I was interested to compare the gene expression changes observed in this calibrated analysis of nucRNA generated from *Kdm2b-CXXC*$^{fl/fl}$ mESCs to the gene expression changes revealed following mRNA-seq analysis without calibration to a spike-in genome (Blackledge et al. 2014). Genome-wide analysis revealed that over a thousand genes became deregulated in this published dataset, with more significantly upregulated than downregulated genes (Figure 5.7 A). However, there were a number of notable differences between these significant changes and the changes I identified using calibrated nucRNA-seq analysis of the same cell line. Most importantly, no global upregulation of transcription was detected using the dataset without spike-in calibration, illustrated by the centering of the log2 fold change distribution around zero (Figure 5.7 B), despite the moderate overall correlation of gene expression changes between the two experiments

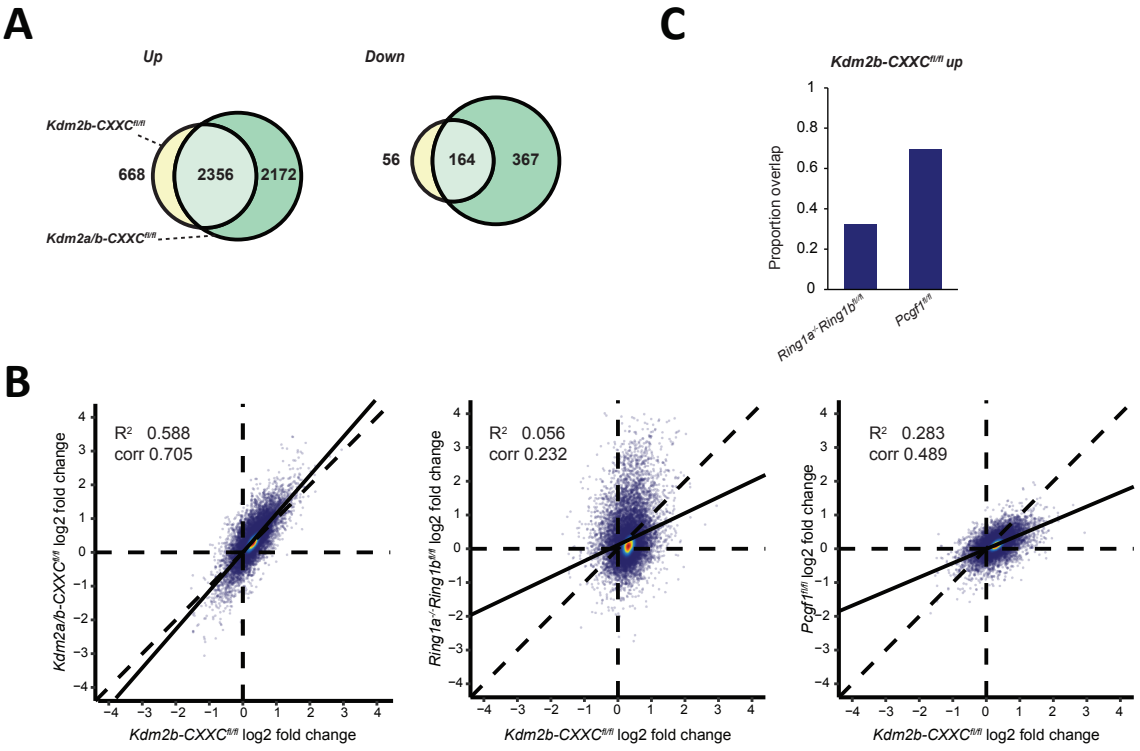


Figure 5.6: Transcription changes following removal of KDM2B correlate moderately with PCGF1 knockout and strongly with KDM2A/B knockout

(A) Venn diagrams comparing genes that significantly increased (Up) or decreased (Down) in expression following 96 h tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ (yellow) or *Kdm2a/b-CXXC*$^{fl/fl}$ (green) mESCs to induce removal of KDM2B alone or KDM2A/B together, respectively.

(B) Scatter plots comparing the log2 fold change in nucRNA-seq signal following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs with the log2 fold change following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ (left), *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ (middle), or *Pcgf1*$^{fl/fl}$ (right) mESCs. Solid lines represent linear regression, with the coefficient of determination (R^2) and Spearman correlation coefficient (corr) annotated.

(C) Bar chart showing the proportion of the set of genes that significantly increased in expression following tamoxifen treatment of either *Pcgf1*$^{fl/fl}$ or *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ mESCs that also significantly increased in expression following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs.

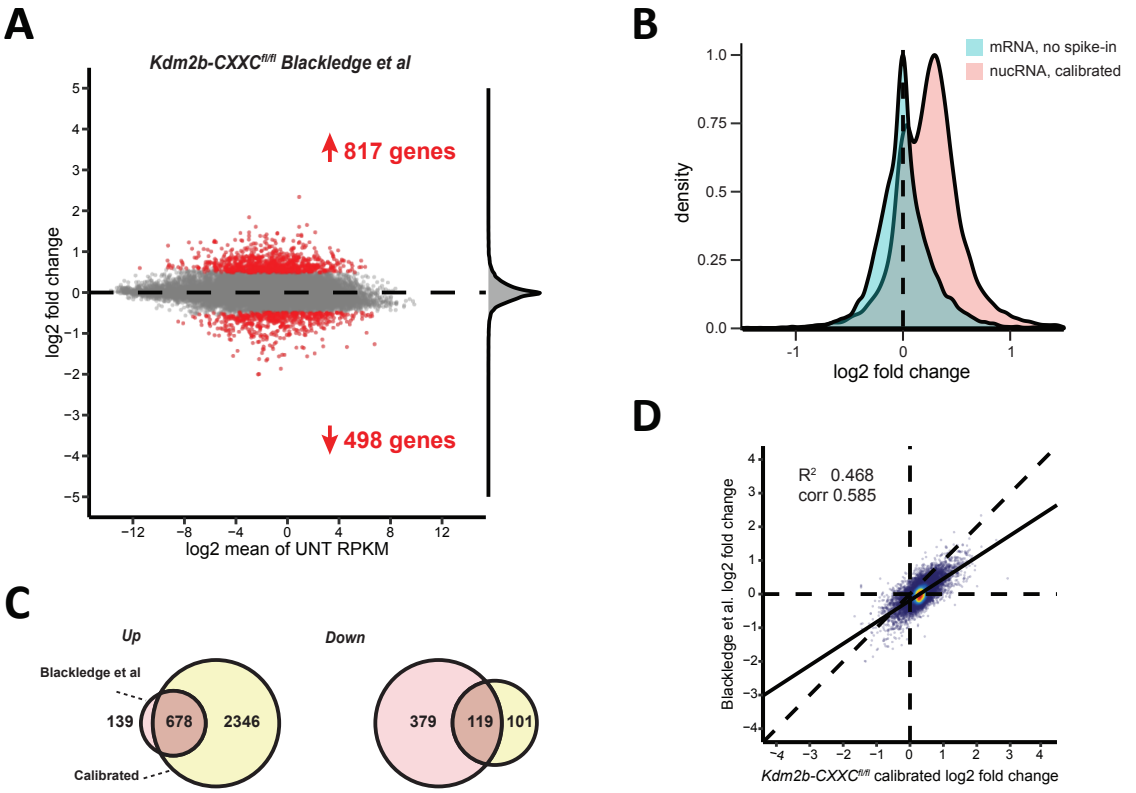


Figure 5.7: Calibration to *Drosophila* spike-in is required to reveal a widespread increase in transcription following KDM2B removal
(A) Scatter plot of log2 mean mRNA-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2b-CXXC*$^{fl/fl}$ mESCs (Blackledge et al. 2014) for a non-redundant mm10 gene set (n=20633, three biological replicates), plotted against log2 fold change in mRNA-seq signal following 96 h tamoxifen treatment to induce KDM2B removal. Libraries were normalised to total library size. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change>1.4). A density plot of the log2 fold change in nucRNA-seq signal is shown to the right.
(B) Density plot showing the log2 fold change distribution of RNA-seq signal for all transcripts following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs, either normalised to total library size (blue) or calibrated to *Drosophila* spike-in (red). The dashed line indicates no change in transcription.
(C) Venn diagrams comparing the genes that significantly increased (Up) or decreased (Down) in expression following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$, for published mRNA (red, Blackledge et al. 2014) and calibrated nucRNA (yellow) datasets.
(D) Scatter plot comparing the log2 fold change in RNA-seq signal following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs for published mRNA (Blackledge et al. 2014) and calibrated nucRNA datasets. Solid line represents linear regression, with the coefficient of determination (R^2) and Spearman correlation coefficient (corr) annotated.

(Figure 5.7 C, D). This demonstrated that spike-in calibration is required to detect widespread directional changes in transcription, as discussed in Section 4.2.4.

Gene ontology analysis revealed that the genes that significantly increased in expression following KDM2B removal were enriched for a variety of developmental terms (Figure 5.8 A). In contrast, the subset of these genes that that have CGI promoters and were not classified as polycomb targets (1506 out of 3024 genes) were instead enriched for terms relating to basic cellular processes, including microtubule-mediated localization (Figure

5.8 B). The downregulated gene set were not significantly enriched for any gene ontology terms.

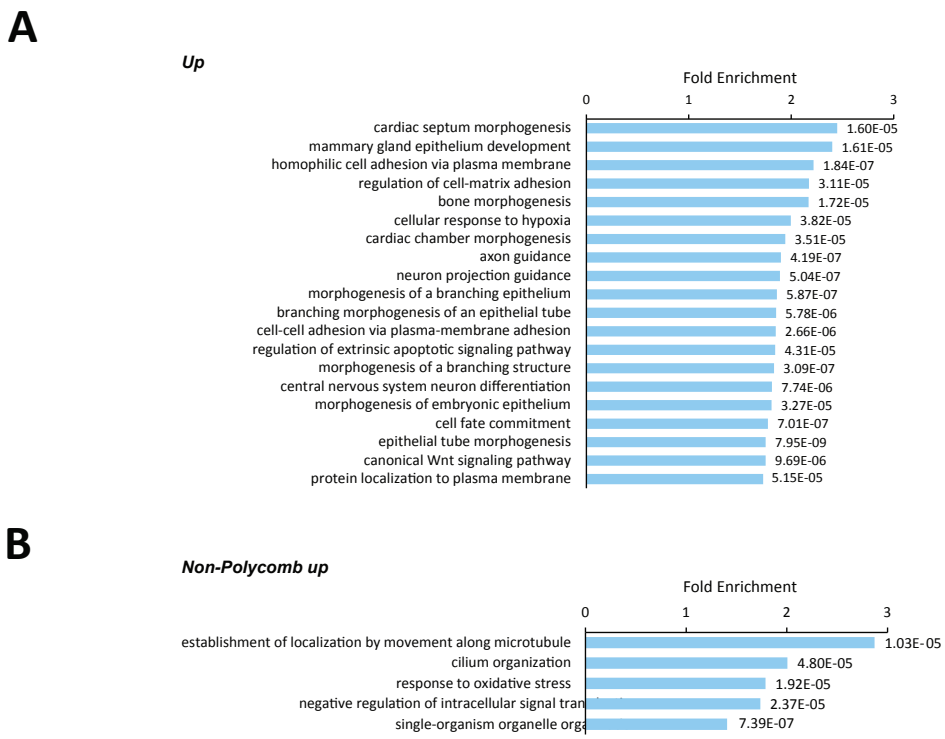


Figure 5.8: Significantly increasing non-polycomb target genes are involved in a range of basic cellular processes
(A) Gene ontology analysis of genes that significantly increased in expression following KDM2B removal (n=3024). BP FAT analysis with FDR<0.1 cutoff was used, and p-values are annotated. The 20 most enriched terms are shown.
(B) As (A), but for the subset of significantly increasing genes that are not polycomb targets (n=1506).

5.2.5 KDM2A and KDM2B act as functionally redundant transcriptional repressors of a subset of genes

There was a strong positive correlation between the gene expression changes following removal of KDM2B alone and those following removal of KDM2A/B together. However, the gradient of the linear regression line was greater than one. While this deviation was small, it indicated that there may have been a stronger upregulation of a subset of genes following removal of KDM2A in addition to KDM2B (Figure 5.6 B - left, compare to the dashed line of equality). I investigated this possibility by plotting the ratio of the fold change in gene expression following KDM2B removal compared to following KDM2A/B removal (Figure 5.9 A). The change in gene expression was similar between these two cell lines for the majority of transcripts. This is illustrated by the heatmap intensity

centred at zero, which corresponds to no difference between the gene expression changes. However, over 1500 genes were differentially regulated, with more than 1.5 fold higher or lower expression following removal of KDM2B alone compared to KDM2A/B together. Comparison of the log2 fold change in gene expression of these differentially regulated classes revealed that the lower expression of a subset of genes following removal of KDM2B alone was due to a weaker increase in their transcription (Figure 5.9 B - right). In contrast, the higher expression of a subset of genes following removal of KDM2B alone was due to a less dramatic decrease in their expression (Figure 5.9 B - middle). Therefore, both sets represented genes that were less dramatically deregulated following removal of KDM2B alone compared to simultaneous removal of KDM2A/B. This implied that KDM2A plays a functionally redundant role in sustaining the normal transcription level of these genes, and is able to compensate for the removal of KDM2B from their promoters.

The subset of genes with relatively low expression following removal of KDM2B alone compared to KDM2A/B together contained nearly three times as many genes as the subset with relatively high expression (1274 compared to 438 genes). This indicated that the majority of the differential gene expression changes between the two cell lines could be explained by the continued transcriptional repression of a subset of genes by KDM2A following removal of KDM2B in isolation. Closer inspection revealed that only 20% of this subset were significantly upregulated following removal of KDM2B alone, whereas nearly 75% were significantly upregulated following removal of KDM2A/B (Figure 5.9 C). This was consistent with stronger upregulation of these genes following removal of KDM2A in addition to KDM2B.

I was interested to investigate whether the genes that significantly increased in expression in these cell lines were generally more strongly upregulated following removal of KDM2A/B together compared to KDM2B alone. On average there was no difference in the gene expression changes between the two cell lines (Figure 5.9 D, orange). The set of genes that significantly increased in expression following KDM2B removal were subtly less strongly upregulated following removal of KDM2B alone compared to KDM2A/B together (3024 genes, light blue). In contrast, the set of genes that significantly increased in expression following simultaneous removal of KDM2A/B

were much less strongly upregulated following removal of KDM2B alone (4528 genes, purple - illustrated by the pronounced reduction in the ratio of the log2 fold change in gene expression). The majority of the significantly increasing genes were shared between the two cell lines, and this shared set were less strongly upregulated following removal of KDM2B alone compared to KDM2A/B together (2356 genes, dark blue). This indicated that the trend of stronger gene upregulation following removal of KDM2A/B together compared to KDM2B alone was a more general feature of genes that significantly increased in transcription.

Further investigation of this trend revealed a negative correlation between the fold change in expression following KDM2A/B removal and the fold change difference comparing removal of KDM2B alone and KDM2A/B together (Figure Figure 5.9 E). Thus genes that moderately increased in expression following removal of KDM2A/B showed a similar increase following removal of KDM2B alone, whereas genes that strongly increased in expression following KDM2A/B removal were much less strongly upregulated following removal of KDM2B in isolation. This suggested that continued targeting of KDM2A to chromatin constrained the expression of a subset of genes in the absence of KDM2B. Removal of KDM2A in addition to KDM2B then led to a relatively large increase in the expression of these genes.

The set of genes that appeared to rely on KDM2A for repression in the absence of KDM2B were enriched for gene ontology terms relating to a variety of developmental processes (Figure 5.10 A). In contrast, the set that failed to undergo strong downregulation in the absence of KDM2B were enriched for genes involved in ion transport (Figure 5.10 B). These are similar to the ontology terms of the set of genes that were significantly deregulated following removal of KDM2A/B together (Figure 4.12), suggesting that genes that rely on KDM2A to maintain relatively normal levels of transcription do not represent a specific subset, at least based on their ontology.

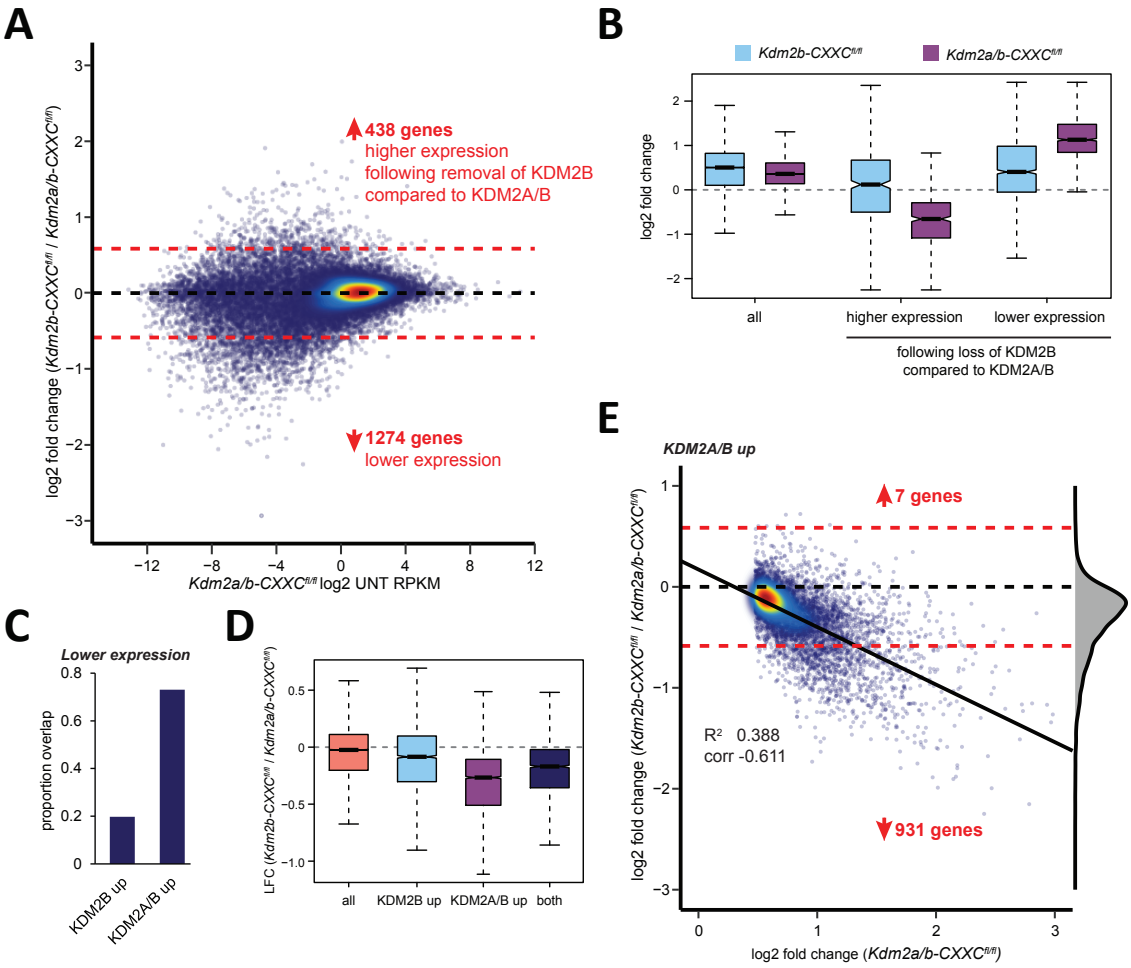


Figure 5.9: KDM2A and KDM2B act as functionally redundant transcriptional repressors of a subset of genes

(A) Scatter plot of log2 mean nucRNA-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs for a non-redundant mm10 gene set (n=20633), plotted against the ratio of the log2 fold change in gene expression following removal of KDM2B alone versus KDM2A/B together, in tamoxifen treated *Kdm2b-CXXC*$^{fl/fl}$ and *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs respectively. A 1.5 fold threshold (red dotted lines) was used to define genes that were differentially regulated between the two cell lines.

(B) Box plot of the log2 fold change in gene expression following 96 h tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ (blue) or *Kdm2a/b-CXXC*$^{fl/fl}$ (purple) mESCs, for all genes (n=20633) or genes with more than 1.5 fold higher (n=438) or lower (n=1274) expression following loss of removal of KDM2B alone compared to removal of KDM2A/B together.

(C) Bar chart for the set of genes with more than 1.5 fold lower expression following loss of KDM2B compared to KDM2A/B (n=1274), showing proportion overlap with the set of genes that significantly increased in expression following removal of KDM2B alone (n=3024, KDM2B up) or KDM2A/B together (n=4528, KDM2A/B up).

(D) Box plot of the ratio of the log2 fold change in gene expression following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ compared to *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, for all genes, KDM2B up and KDM2A/B up gene sets, and genes that are shared between these two sets (n=2356, both).

(E) Scatter plot of the log2 fold change in gene expression following tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, plotted against the ratio of the log2 fold change following tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ versus *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, for the KDM2A/B up gene set. A 1.5 fold threshold (red dotted lines) was used to define genes which were differentially regulated between the two datasets. Solid line represents linear regression, with the coefficient of determination (R^2) and Spearman correlation coefficient (corr) annotated. A density plot of the log2 fold change in nucRNA-seq signal is shown to the right.

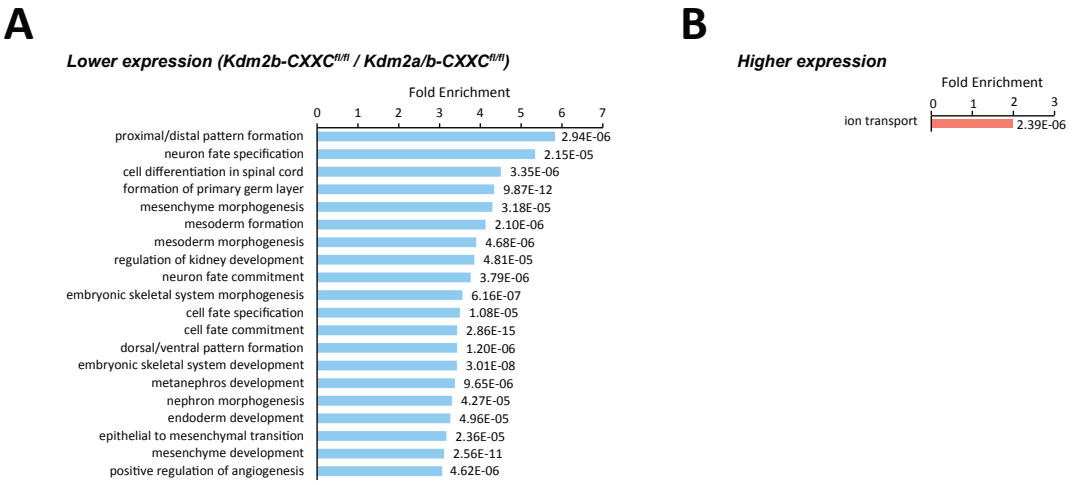


Figure 5.10: KDM2A is required to maintain normal transcription levels of a subset of genes
(A) Gene ontology analysis of genes that were more that 1.5 fold more highly expressed following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ compared to *Kdm2b-CXXC*$^{fl/fl}$ mESCs (n=438). BP FAT analysis with FDR<0.1 cutoff was used, and p-values are annotated. The 20 most enriched terms are shown.
(B) As (A), but for genes that were more than 1.5 fold less highly expressed following removal of KDM2A/B together compared to KDM2B alone (n=1274).

5.2.6 Conditional deletion of FBXL19, KDM2A and KDM2B ZF-CXXC domains

While conditional deletion of the ZF-CXXC domain of KDM2A did not perturb normal gene expression (Section 5.2.2), comparison of the transcriptional changes that resulted from KDM2B removal with those following removal of KDM2A/B together revealed that KDM2A appears to act as a functionally redundant transcriptional repressor of a subset of genes (Section 5.2.5). This led me to consider the possibility that the third KDM2 family member, FBXL19, might also contribute to the regulation of gene expression in mESCs. While conditional removal of FBXL19 alone does not lead to transcription deregulation in mESCs (Dimitrova et al. 2018), simultaneous loss of KDM2A, KDM2B and FBXL19 from chromatin might reveal a functionally redundant role for FBXL19 as part of the transcription repression coordinated by the KDM2 family. To test this hypothesis, I generated a cell line with simultaneous tamoxifen-inducible deletion of the ZF-CXXC domains of KDM2A, KDM2B and FBX19, which I will refer to as *Kdm2a/b/f19-CXXC*$^{fl/fl}$. I used CRISPR-mediated genome editing to sequentially insert loxP sites in the *Fbxl19* gene flanking the exon encoding the majority of the ZF-CXXC domain, using *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs (Figure 5.11 A). I identified successfully targeted clones using

PCR screens of genomic DNA and cDNA, and validated that the exact loxP sequence had been inserted using Sanger sequencing. I used western blotting to confirm that treatment of *Kdm2a/b/F19-CXXC*$^{fl/fl}$ mESCs with tamoxifen for 96 h led to a decrease in the molecular weight of FBXL19, corresponding to deletion of its ZF-CXXC domain (Figure 5.11 B). Tamoxifen treatment also induced a size shift of the KDM2A/B-LF and KDM2A/B-SF proteins, indicating that I had achieved simultaneous deletion of the ZF-CXXC domains of KDM2A, KDM2B and FBXL19.

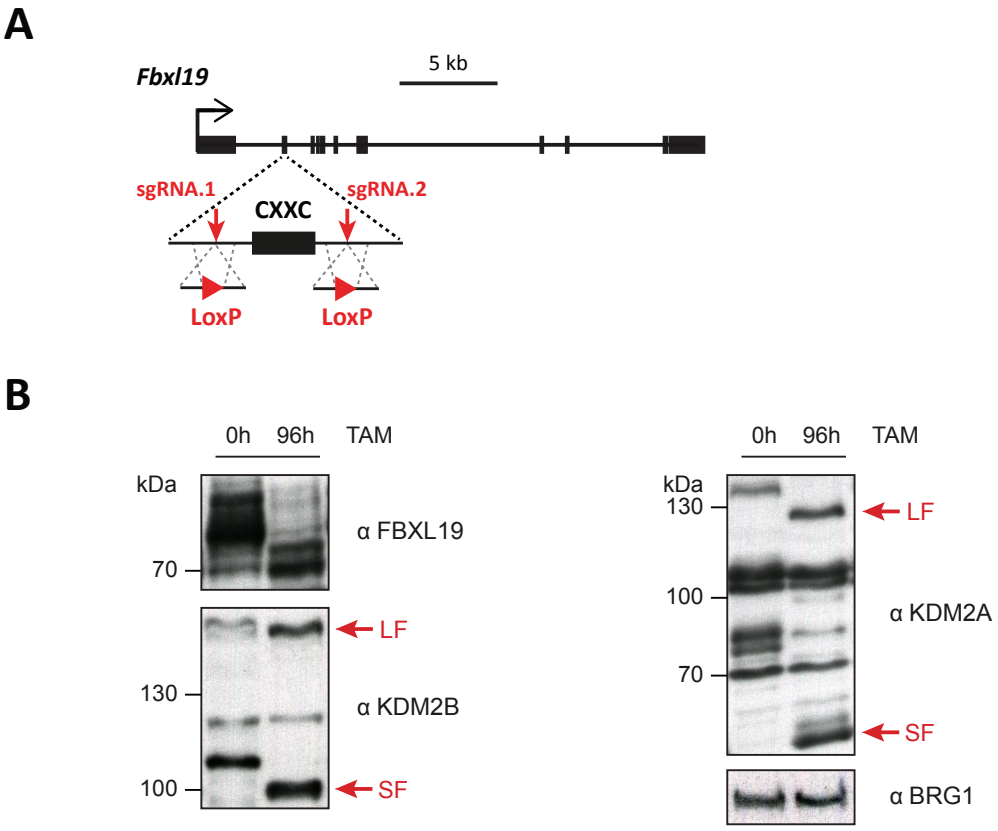


Figure 5.11: Conditional deletion of the ZF-CXXC domains of KDM2A, KDM2B and FBXL19
(A) Schematic depicting the position of sgRNAs used to target Cas9 to generate DSBs flanking an exon encoding the ZF-CXXC domain of FBXL19. LoxP sites were inserted via repair of the DSB using a supplied HDR template with 150 bp homology arms.
(B) Western blot showing a size shift of KDM2A/B-LF, KDM2A/B-SF and FBXL19 proteins following 96 h tamoxifen (TAM) treatment of *Kdm2a/b/F19-CXXC*$^{fl/fl}$ mESCs, corresponding to deletion of their ZF-CXXC domain. BRG1 is a loading control.

5.2.7 Removal of FBXL19 in addition to KDM2A/B largely recapitulates the transcription changes following removal of KDM2A/B alone

I was interested to investigate whether simultaneous loss of KDM2A, KDM2B and FBXL19 from chromatin would reveal functional redundancy between these KDM2

family members. To explore this possibility I performed calibrated nucRNA-seq for *Kdm2a/b/f19-CXXC*$^{fl/fl}$ mESCs, with and without 96 h tamoxifen treatment to induce removal of KDM2A/B and FBXL19 from chromatin. A genome-wide analysis revealed that that the widespread increase in transcription that resulted from KDM2A/B removal was recapitulated following additional removal of FBXL19, with 4277 significantly increasing and 243 significantly decreasing transcripts ($p<0.05$ and fold change>1.4, Figure 5.12 A). Comparison of the log2 fold change density revealed a very similar distribution for the *Kdm2a/b-CXXC*$^{fl/fl}$ and *Kdm2a/b/f19-CXXC*$^{fl/fl}$ cell lines, indicating that the gene expression changes were very similar overall.

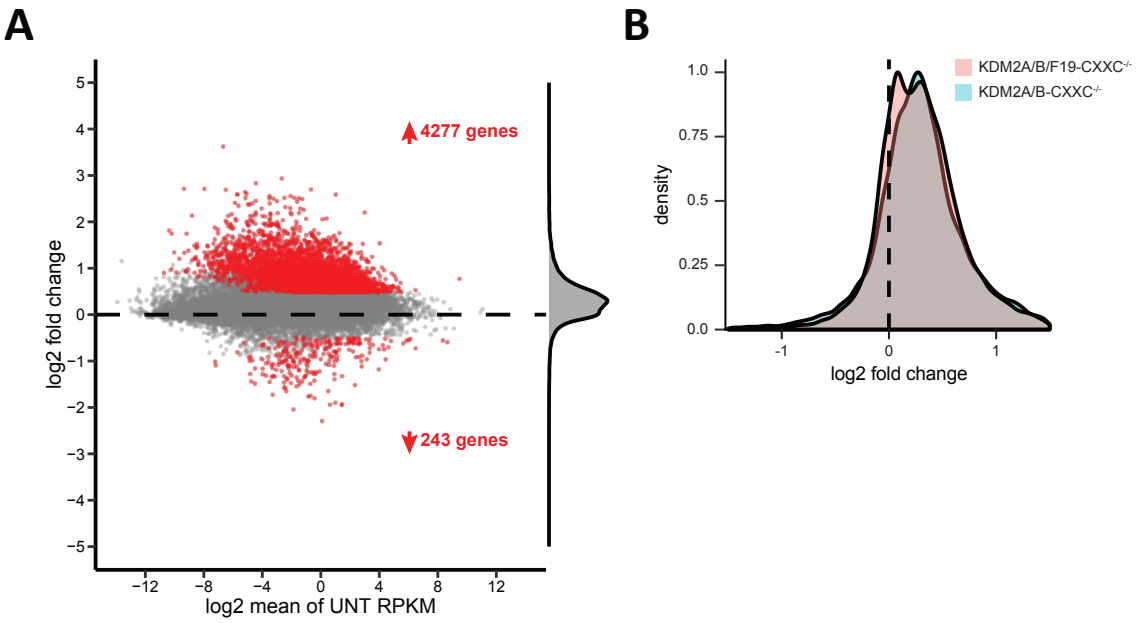


Figure 5.12: Removal of FBXl19 in addition to KDM2A/B recapitulates the widespread increase in transcription
(A) Scatter plot of log2 mean nucRNA-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2a/b/f19-CXXC*$^{fl/fl}$ mESCs for a non-redundant mm10 gene set (n=20,633, three biological replicates), plotted against log2 fold change in nucRNA-seq signal following 96 h tamoxifen treatment to induce removal of KDM2A/B and FBXL19. Libraries were calibrated to *Drosophila* spike-in. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change>1.4). A density plot of the log2 fold change in nucRNA-seq signal is shown to the right.
(B) Density plot showing the log2 fold change distribution of nucRNA-seq signal for all transcripts, following tamoxifen treatment of *Kdm2a/b/f19-CXXC*$^{fl/fl}$ (red) or *Kdm2a/b-CXXC*$^{fl/fl}$ (blue) mESCs. The dashed line corresponds to no change in transcription.

I proceeded to investigate the properties of the genes that significantly changed in expression. This revealed that genes that are sensitive to loss of PRC1.1 targeting were a subset of the upregulated genes (Supplementary Figure A.9), as I found for the genes that were upregulated following removal of KDM2A/B only. In contrast, genes that significantly decreased in expression were enriched for non-CGI genes, suggesting that

their deregulation may have been a secondary effect of the removal of KDM2A/B and FBXL19 from CGI promoters. Gene ontology analysis confirmed that the set of genes that significantly increased in transcription following loss of FBXL19 and KDM2A/B from chromatin contained both developmental genes and genes involved in a variety of basic cellular process (Figure 5.13 A and B), reflecting the widespread transcription upregulation of both polycomb target and non-polycomb target genes. In contrast, the genes that significantly decreased in expression were not significantly enriched for any terms.

5.2.8 A minority of genes are differentially regulated following removal of FBXL19 in addition to KDM2A/B

Next, I investigated the gene expression changes that occurred following removal of FBXL19 and KDM2A/B in more detail using correlation analysis. The gene expression changes following simultaneous removal of FBXL19 and KDM2A/B strongly correlated with the gene expression changes following removal of KDM2B alone or KDM2A/B together for the significantly upregulated gene sets, reflecting the widespread upregulation of similar sets of genes in these three cell lines (Figure 5.14). The gene expression changes in these cell lines also correlated moderately with the enrichment of a variety of polycomb-associated features (KDM2B, RING1B, SUZ12, H2AK119ub1 and H3K27me3 TSS enrichment), reflecting the upregulation of a subset of polycomb target genes in these cell lines. There was a moderate negative correlation between the gene expression changes and starting transcription level, indicating that lowly transcribed genes tended to be more strongly upregulated, although this correlation was weaker following removal of KDM2B alone. In contrast, there was poor correlation with the gene expression changes that occurred following RING1A/B knockout (bottom row), as only a subset of polycomb target genes became deregulated following loss of KDM2A/B and FBXL19 from chromatin.

I extended this comparison of the three cell lines to compare the gene expression changes for all genes, using scatter plots to compare the gene expression changes following simultaneous removal of KDM2A/B and FBXL19 with those following removal

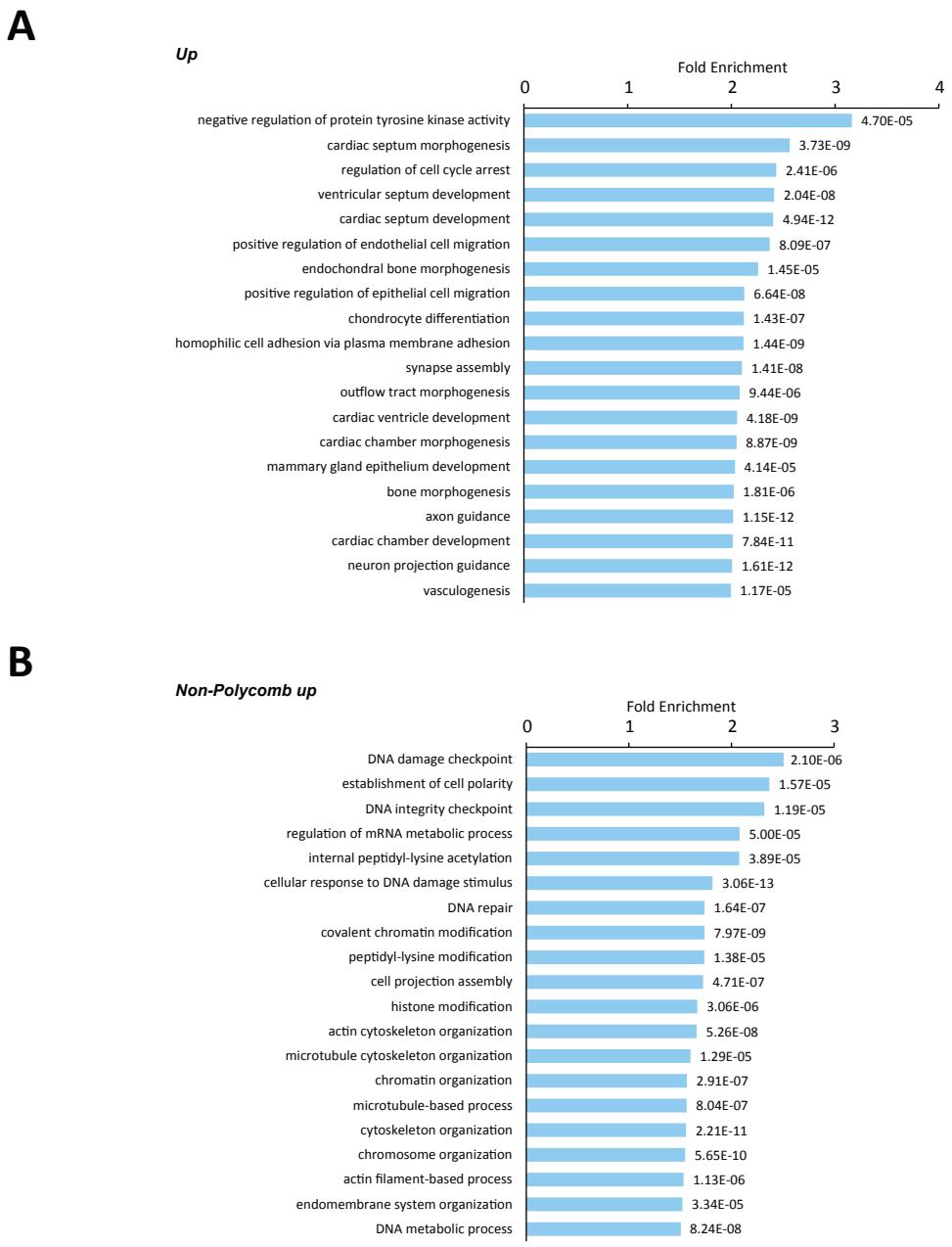


Figure 5.13: Genes that increase in expression following removal of FBXL19 and KDM2A/B are involved in both developmental and basic cellular processes
(A) Gene ontology analysis of genes that significantly increased in expression following 96 h tamoxifen treatment of *Kdm2a/b/f19-CXXC*$^{fl/fl}$ mESCs to induce removal of KDM2A, KDM2B and FBXL19. BP FAT analysis with FDR<0.1 cutoff was used, and p-values are annotated. The 20 most enriched terms are shown. (B) As (A), but for non-polycomb target genes that significantly increased in expression.

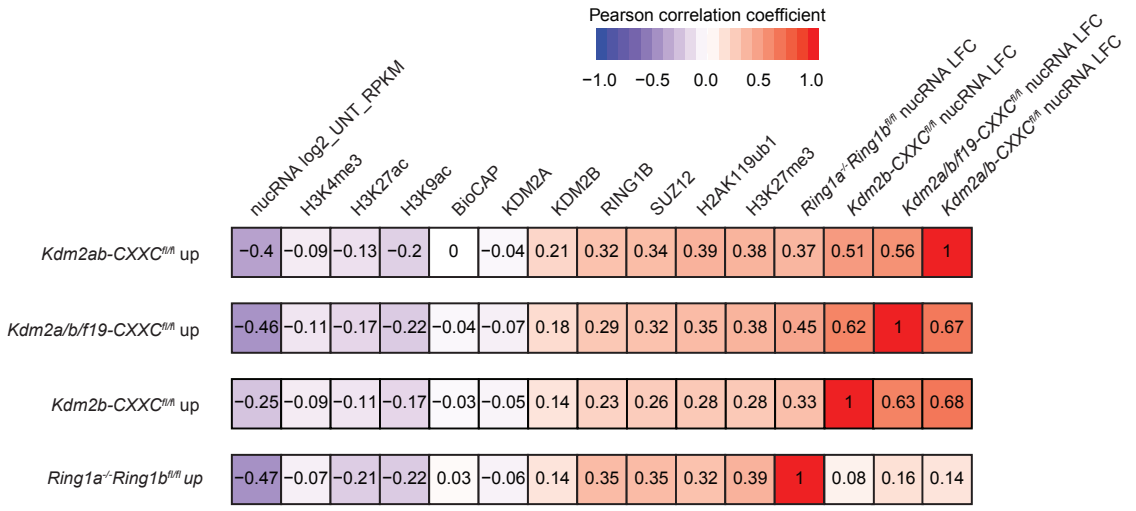


Figure 5.14: Negative correlation between gene expression level and log2 fold change in transcription
Matrix of Pearson correlation coefficients, comparing the log2 fold change (LFC) in gene expression following 96 h tamoxifen treatment of *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$, *Kdm2b-CXXC*$^{fl/fl}$, *Kdm2a/b/f19-CXXC*$^{fl/fl}$ and *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs with expression level (log2 UNT RPKM) and the enrichment of a range of promoter-associated features at the 2 kb region flanking the TSS. Correlations for the subset of genes that significantly increased in expression following tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ (top row), *Kdm2a/b/f19-CXXC*$^{fl/fl}$ (second row), *Kdm2b-CXXC*$^{fl/fl}$ (third row) or *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ mESCs (bottom row) are shown.

of KDM2B alone (Figure 5.15 A) or KDM2A/B together (5.15 B). This revealed a relatively strong overall correlation between the gene expression changes. There was a slightly weaker correlation with the transcription changes following removal of KDM2B alone, which may reflect the derepression of genes that are otherwise constrained by continued KDM2A chromatin targeting (Figure 5.9).

A comparison of the fold change in gene expression following simultaneous removal of KDM2A/B and FBXL19 with that following removal of KDM2A/B only confirmed that the gene expression changes were similar for the majority of transcripts (Figure 5.15 C). However, this analysis also revealed the differential regulation of a subset of genes. There were 859 genes that were greater than 1.5 fold more highly expressed following removal of FBXL19 in addition to KDM2A/B due to a more dramatic increase in their transcription (Figure 5.15 D - middle), while 963 genes were more than 1.5 fold less highly expressed due to their failure to become strongly upregulated (Figure 5.15 D - right).

I was interested to investigate whether genes were generally more strongly upregulated following loss of FBXL19 in addition to KDM2A/B. On average there was no difference

between the gene expression changes following removal of KDM2A/B alone and following removal of KDM2A/B together with FBXL19 (Figure 5.15 E, orange). The set of genes that significantly increased in expression following simultaneous removal of KDM2A/B and FBXL19 were on average slightly more highly expressed following removal of FBXL19 in addition to KDM2A/B (Figure 5.15 E, 4277 genes, light blue), whereas the set of genes that significantly increased in expression following removal of KDM2A/B alone were on average less highly expressed following additional removal of FBXL19 (4528 genes, purple). Importantly, there was no difference in the gene expression changes for the subset of genes that significantly increased in expression in both cell lines (2539 genes, dark blue). Therefore, while removal of FBXL19 in addition to KDM2A/B led to differences in the gene expression changes of a subset of genes, FBXL19 did not appear to play a widespread role in either restricting or enhancing the gene expression changes following removal of KDM2A/B, and did not seem to feed into the broad transcription repressive pathway coordinated by KDM2A/B.

5.3 Summary and discussion

Here I investigated the individual roles played by KDM2A and KDM2B in transcription regulation, by studying the effect of conditional removal of either KDM2A or KDM2B in isolation from chromatin. I also explored the possibility that FBXL19 plays a functionally redundant role with KDM2A and KDM2B in transcription repression, by exploring the gene expression changes following simultaneous removal of FBXL19, KDM2A and KDM2B.

Tamoxifen treatment of *Kdm2a-CXXC*$^{fl/fl}$ mESCs revealed that KDM2A is not required to maintain normal transcription in mESCs (Figure 5.2). It would be interesting to use this cell line to investigate whether KDM2A plays a more important role in gene expression regulation in a more differentiated cell type, or during a cell fate transition such as retinoic acid mediated differentiation. In addition, generation of *Kdm2a-CXXC*$^{fl/fl}$ mice from these mESCs would enable the developmental role of KDM2A chromatin binding to be studied for the first time. While KDM2A-LF knockout leads to embryonic lethality between E10.5 and E12.5 (Kawakami et al. 2015), it is possible that the simultaneous loss

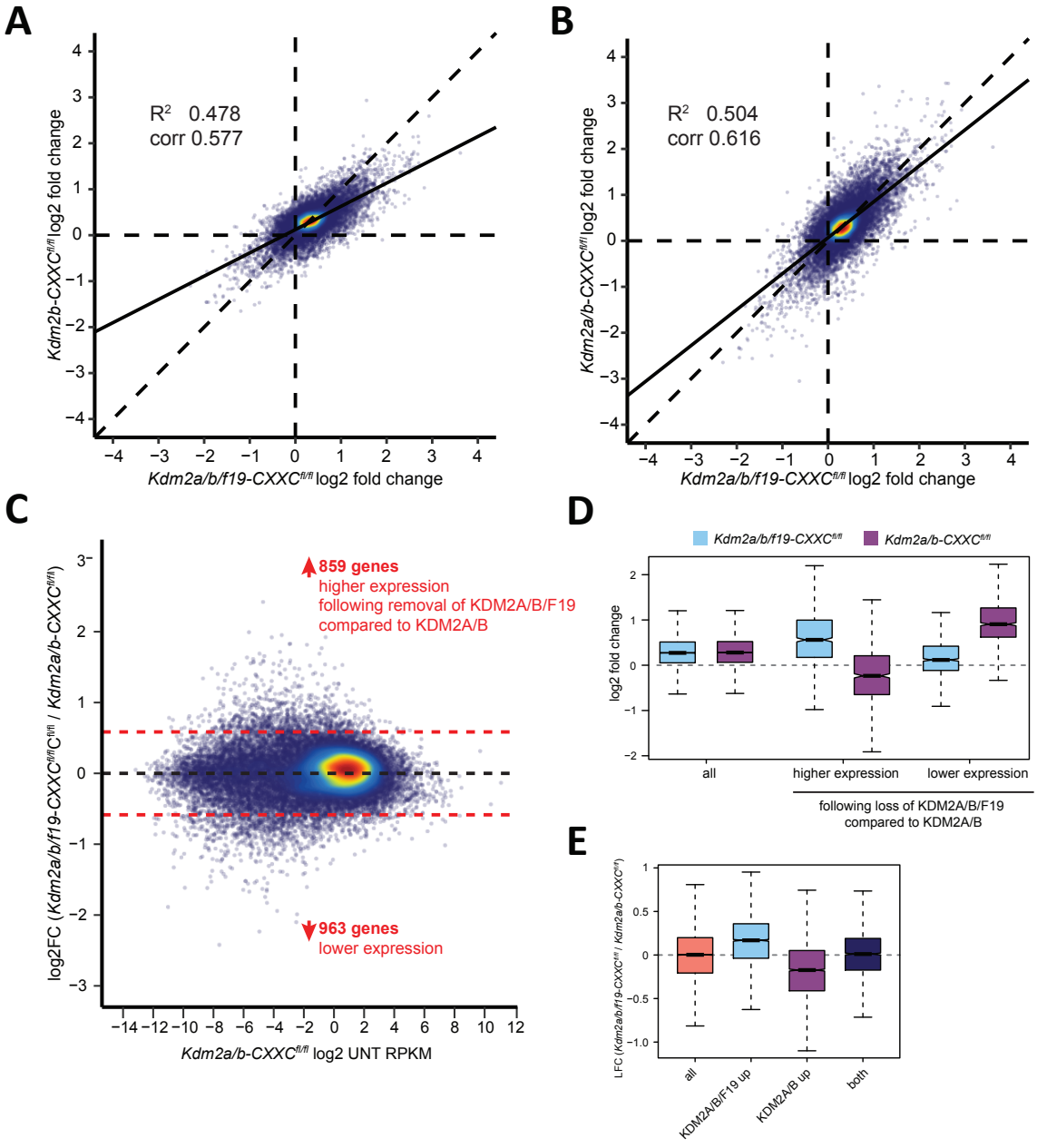


Figure 5.15: A minority of genes are differentially regulated following removal of FBXL19 in addition to KDM2A/B

(A) Scatter plot comparing the log2 fold change in nucRNA-seq signal following 96 h tamoxifen treatment of *Kdm2a/b/f19-CXXC*$^{fl/fl}$ mESCs with the log2 fold change following 96 h tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs. Solid line represents linear regression, with the coefficient of determination (R^2) and Spearman correlation coefficient (corr) annotated.
(B) As (A), but comparing to the log2 fold change following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs.
(C) Scatter plot of log2 mean nucRNA-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs for a non-redundant mm10 gene set (n=20633), plotted against the ratio of the log2 fold change (log2FC) following tamoxifen treatment of *Kdm2a/b/f19-CXXC*$^{fl/fl}$ versus *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs. A 1.5 fold threshold (red dotted lines) was used to define genes which are differentially regulated between the two datasets.
(D) Box plot of the log2FC in gene expression following tamoxifen treatment of *Kdm2a/b/f19-CXXC*$^{fl/fl}$ (blue) or *Kdm2a/b-CXXC*$^{fl/fl}$ (purple) mESCs, for all genes (n=20633) or genes with more than 1.5 fold higher or lower change in expression following removal of FBXL19 in addition to KDM2A/B.
(E) Box plot of the ratio of the log2FC in gene expression following tamoxifen treatment of *Kdm2a/b/f19-CXXC*$^{fl/fl}$ compared to *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, for all genes (n=20633, all) or for the subset of genes that significantly increased in expression following tamoxifen treatment of *Kdm2a/b/f19-CXXC*$^{fl/fl}$ (n=4277, KDM2A/B/F19 up) or *Kdm2a/b-CXXC*$^{fl/fl}$ (n=4528, KDM2A/B up) mESCs, and genes that were shared between these two sets (n=2539, both).

of KDM2A-LF and KDM2A-SF from chromatin in this line would yield a more dramatic developmental phenotype.

In contrast, tamoxifen treatment of *Kdm2b-CXXC*$^{fl/fl}$ mESCs largely recapitulated the widespread increase in transcription that occurred in *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs (Figure 5.4), with good overall correlation between the gene expression changes (Figure 5.6). Therefore, within the KDM2 family KDM2B plays the predominant role in constraining transcription. The nucRNA levels of *Kdm2a* were higher than those of *Kdm2b*, suggesting that the relatively minor role played by KDM2A in transcription repression is unlikely to be explained by a relatively low abundance of KDM2A protein compared to KDM2B. However, a direct comparison of KDM2A and KDM2B protein levels at CGI promoters would require endogenous tagging of *Kdm2a* and *Kdm2b* with the same epitope tag, followed by ChIP-seq. Comparison of the gene expression changes following removal of KDM2B to those following PCGF1 or RING1A/B knockout confirmed that genes that are sensitive to loss of PRC1.1 targeting were a subset of the upregulated genes (Figure 5.6), while there was poor overall correlation with the changes following RING1A/B knockout, as observed for the *Kdm2a/b-CXXC*$^{fl/fl}$ dataset. Loss of KDM2B from chromatin recapitulated the widespread upregulation of both polycomb target and non-target genes observed following loss of KDM2A/B, demonstrating that KDM2B exerts both polycomb-dependent and -independent modes of transcription repression.

A previous study proposed a demethylase-independent role for KDM2B in the protection of polycomb target gene promoters from DNA methylation, as constitutive KDM2B knockout led to methylation of these promoters and transcriptional repression of the associated genes (Boulard et al. 2015). However, in a series of unpublished experiments I failed to observe an increase in DNA methylation of CGI gene promoters following KDM2A/B knockout (Supplementary Figure A.10). This analysis included targets that Boulard et al. found to increase in DNA methylation following constitutive KDM2B knockout. It appears that long term culture of KDM2B knockout cells may have led to the accumulation of DNA methylation in the cell line used in the published study. The transcriptional changes in this line are likely unrelated to the widespread increase in transcription I observed following conditional removal of KDM2B in *Kdm2b-CXXC*$^{fl/fl}$ mESCs.

Despite the predominant role played by KDM2B in transcriptional repression, loss of KDM2A/B together from chromatin led to a more dramatic increase in the expression of a subset of genes than loss of KDM2B alone, revealing that KDM2A and KDM2B act as functionally redundant transcriptional repressors of these genes (Figure 5.9). Validation of this variation in the extent of transcription upregulation using ddPCR would add weight to this evidence that KDM2A is able to compensate for the loss of KDM2B promoter binding at a subset of genes.

Importantly, while a subset of genes experienced a more dramatic increase in transcription following removal of KDM2A/B together compared to KDM2B alone, these genes still moderately increased in transcription following loss of KDM2B alone, to an extent similar to the average increase in transcription of all genes (Figure 5.9 B - 'lower expression'). Therefore, KDM2A was insufficient to prevent an increase in the transcription of these genes following loss of KDM2B, but counteracted a more dramatic upregulation. Thus KDM2A and KDM2B are not truly functionally redundant at these sites, but perhaps play additive roles to restrict transcription. Indeed, extending this analysis to the entire group of genes that significantly increased in transcription following removal of KDM2A/B revealed that the genes that most strongly increased in transcription following loss of KDM2A/B were most strongly constrained by the continued presence of KDM2A following loss of KDM2B only ((Figure 5.9 E). This

suggests that KDM2A plays a widespread role in constraining the transcription of genes that would otherwise undergo a relatively large increase in transcription following simultaneous loss of KDM2A/B from chromatin. It would be interesting to explore whether the relative contributions of KDM2A and KDM2B to transcription repression vary in a more differentiated cell type. I noted a negative correlation between starting expression level and the extent of transcription upregulation following removal of KDM2A/B from chromatin. This correlation was less strong following removal of KDM2B only (Figure 5.14). This suggests that the transcription constraining role of KDM2A predominantly acts at lowly expressed genes, leading to a reduction in the correlation between expression level and transcription upregulation following removal of KDM2B alone.

Simultaneous removal of FBXL19, KDM2A and KDM2B in *Kdm2a/b/f19-CXXC*$^{fl/fl}$ mESCs largely recapitulated the widespread increase in transcription observed following loss of KDM2A/B (Figure 5.12). Comparison of the gene expression changes revealed that a minority of genes are differentially regulated following additional loss of FBXL19 from chromatin. There were an equal number of genes that were more or less strongly upregulated following additional loss of FBXL19, indicating that FBXL19 does not play a general role in either enhancing or counteracting the transcription repressive activity of KDM2A/B. This is also reflected in the similar profile of log2 fold change density for both cell lines (Figure 5.12 B).

It would be interesting to study the combined developmental role of KDM2A/B, with or without FBXL19, in *Kdm2a/b-CXXC*$^{-/-}$ and *Kdm2a/b/f19-CXXC*$^{-/-}$ mice. Despite the relatively minor contribution of KDM2A to transcription repression in the context of *Kdm2b-CXXC*$^{-/-}$ mESCs, it is possible that this might lead to an earlier embryonic lethality phenotype, while an early requirement for the combined activity of FBXL19, KDM2A and KDM2B might be apparent in *Kdm2a/b/f19-CXXC*$^{-/-}$ mice.

In conclusion, loss of KDM2A or KDM2B individually from chromatin revealed that KDM2B plays the predominant role in transcription repression, while KDM2A constrains the degree of transcription upregulation of a subset of genes. The molecular mechanism that underlies this widespread repression of transcription by KDM2B and more subtle

modulation of transcription by KDM2A may be shared or distinct between the two paralogues, and will be the focus of Chapter 6.

6

Investigating putative transcription repressive activities of KDM2A/B

6.1 Introduction

Conditional removal of KDM2B leads to a widespread increase in nucRNA levels, revealing a broad role for KDM2B in constraining transcription (Chapter 5). This is independent of KDM2B demethylase activity, as conditional knockout of the JmjC domain-containing KDM2A/B long isoforms leads to minimal transcription changes (Chapter 3). Polycomb target genes that are sensitive to loss of PRC1.1-mediated gene repression represent a subset of the genes that increased in expression following removal of KDM2B, while a large number of genes that are not polycomb targets also significantly increased in expression. This indicates that KDM2B has both polycomb-dependent and polycomb-independent modes of transcription repression. KDM2A on its own is not required to sustain normal transcription. However, KDM2A appears to contribute to the same repressive pathway as KDM2B, as removal of KDM2A in addition to KDM2B leads to the more dramatic upregulation of a subset of genes. There are a variety of potential mechanisms by which KDM2A/B may exert transcription repression at CGI-associated gene promoters. These may be distinct for KDM2A and KDM2B or shared by both paralogues.

Early studies revealed that active gene promoters and CGIs are sensitive to nuclease digestion (Tazi and Bird 1990; Stalder et al. 1980; Bird et al. 1985; Antequera et al. 1989), and subsequent genome-wide MNase-seq and DNase-seq approaches demonstrated that chromatin accessibility and nucleosome depletion are defining characteristics of CGI, in contrast to CpG-poor, promoters (Fenouil et al. 2012; Sabo et al. 2004; Teif et al. 2012; Li et al. 2011; Choi 2010). Elevated chromatin accessibility may facilitate binding

of RNA polymerase, basal transcription factors and other transcription regulators to CGIs, and has therefore been proposed to play an important role in their function as gene expression regulatory platforms (Blackledge and Klose 2011; Deaton and Bird 2011). KDM2A/B bind broadly to CGI gene promoters, therefore one possibility is that the widespread transcription repression coordinated by KDM2A/B is mediated through their action to limit chromatin accessibility at these regions, resulting in reduced binding of transcription-promoting factors.

Another potential mechanism by which KDM2A/B might act to broadly repress transcription is their putative E3 ligase activity. KDM2B was reported to possess intrinsic E3 ligase activity via its PHD domain (Janzer et al. 2012) and to ubiquitylate the transcription factor c-Fos, leading to its degradation by the proteasome (Han et al. 2016). In addition, the PHD domain of the KDM2A/B-SF paralogue FBXL19 also displays E3 ligase activity *in vitro* (Emilia Dimitrova, unpublished) and FBXl19 has been shown to ubiquitylate the GTPases RhoA, Rac1 and Rac3 and the interleukin receptor ST2L, leading to their proteasomal degradation (Dong et al. 2014; Wei et al. 2013; Zhao et al. 2012; Zhao et al. 2013). KDM2A has also been suggested to possess E3 ligase activity, as overexpression of KDM2A was reported to stimulate 53BP1 ubiquitylation and lead to an increase in total ubiquitylation levels in whole cell extract (Bueno et al. 2018).

KDM2A, KDM2B and FBXL19 are members of the F-box family of proteins, which are characterised by their ability to confer substrate specificity to Skp1-Cul1-Fbox (SCF) E3 ubiquitin ligase complexes. F-box proteins bind to the core SCF component SKP1 and bind to ubiquitylation targets via their additional protein-protein interaction domains. KDM2A, KDM2B and FBXL19 are classified into the FBXL subfamily, characterised by leucine-rich repeat protein interaction domains (Jin et al. 2004), and all interact with SKP1 (Koyama-Nasu et al. 2007; Farcas et al. 2012; Tan et al. 2013; Gearhart et al. 2006; Han et al. 2016), which is also localised to CGI gene promoters (Emilia Dimitrova, unpublished). Therefore it is possible that KDM2A/B and FBXL19 act as E3 ubiquitin ligases at CGI promoters to regulate the turnover of specific target proteins, which they may interact with via their LRR domains. A role in modulating the level of transcription regulators at CGI could explain how KDM2A/B broadly regulate transcription. However, a role for KDM2A/B in protein ubiquitylation remains to be tested in mESCs.

I aimed to explore the potential mechanisms by which KDM2A/B regulate transcript levels. I investigated the effect of KDM2A/B removal on chromatin accessibility and the nuclear ubiquitylome of mESCs.

6.2 Results

6.2.1 Removal of KDM2A/B leads to minimal changes in chromatin accessibility

To investigate whether KDM2A/B regulate accessibility at CGI-associated gene promoters, I performed calibrated ATAC-seq in *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, with and without 96 h tamoxifen treatment to induce removal of KDM2A/B. A genome-wide analysis of all accessible regions, including both promoters and enhancers, revealed that KDM2A/B were not required to maintain the elevated accessibility of these regions. Removal of KDM2A/B led to virtually no change, with less than 0.1% of all accessible regions significantly changing in accessibility (Figure 6.1 A). A minimal effect on accessibility was also evident when KDM2A or KDM2B were removed individually (Figure 6.1 B and C). To investigate potential functional redundancy with the third KDM2 family member FBXL19, I also performed calibrated ATAC-seq in *Kdm2a/b/f19-CXXC*$^{fl/fl}$ mESCs, where tamoxifen treatment induces simultaneous removal of KDM2A, KDM2B and FBXL19. Again, there were virtually no changes in chromatin accessibility in this line (Figure 6.1 D). Therefore, KDM2A, KDM2B and FBXL19 are not required to maintain the elevated chromatin accessibility that characterises active regulatory regions throughout the genome. Evidently the widespread transcription repression coordinated by KDM2A/B is not a consequence of, and does not immediately result in, an alteration in the accessibility of CGI gene promoters.

6.2.2 SKP1 knockout leads to a dramatic accumulation of nucRNA

Both KDM2A and KDM2B have been reported to possess intrinsic E3 ligase activity (Janzer et al. 2012; Han et al. 2016; Bueno et al. 2018). Furthermore, KDM2A/B both interact with SKP1, a core adaptor protein found in SCF E3 ligase complexes, and SKP1

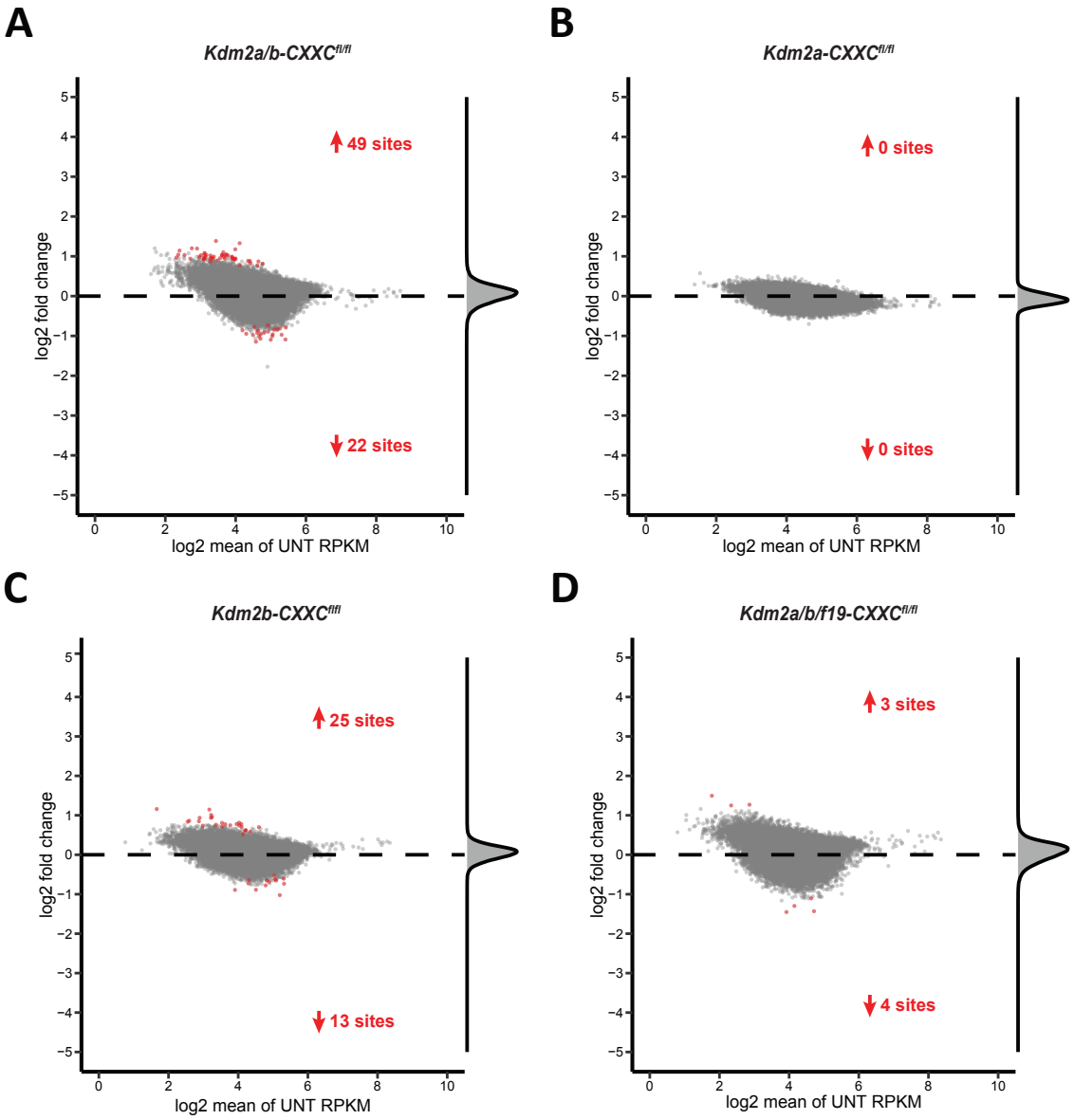


Figure 6.1: Removal of KDM2A/B leads to minimal changes in chromatin accessibility
(A) Scatter plot of log2 mean ATAC-seq reads per kilobase per million mapped reads (RPKM) in *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs for ATAC peaks (n=69063, three biological replicates) plotted against log2 fold change in ATAC-seq signal following 96 h tamoxifen treatment to induce KDM2A/B removal. Libraries were calibrated to *Drosophila* spike-in. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change>1.4). A density plot of the log2 fold change in ATAC-seq signal is shown to the right.
(B) As (A), but for *Kdm2a-CXXC*$^{fl/fl}$ mESCs (n=73136, three biological replicates).
(C) As (A), but for *Kdm2b-CXXC*$^{fl/fl}$ mESCs (n=82135, three biological replicates).
(D) As (A), but for *Kdm2a/b/f19-CXXC*$^{fl/fl}$ mESCs (n=85561, three biological replicates).

colocalises with KDM2A/B at CGIs (Emilia Dimitrova, unpublished). Therefore, SKP1 may be required for the putative E3 ligase activity of KDM2A/B at CGI promoters. In this case, SKP1 knockout might be expected to recapitulate the widespread increase in transcription that occurred following removal of KDM2A/B. To investigate this possibility, I extracted nucRNA from *Skp1*$^{fl/fl}$ mESCs, before and after 72 h tamoxifen treatment to induce SKP1 knockout (Supplementary Figure A.11, cells provided by Prof Haruhiko Koseki). I then investigated the gene expression changes by RT-qPCR. This revealed that SKP1 knockout led to a dramatic increase in transcription of all genes tested (green bars), including genes that were shown by nucRNA-seq analysis to significantly decrease in expression following removal of KDM2A/B (blue bars, *Tdrd12*, *Pcsk9*) and genes that were unaffected or only subtly increased in expression (*Ctsc*, *Usp10*, *Fpgs*, *Lrrc45*, *Wdr92*, Figure 6.2).

The transcriptional response following SKP1 knockout was markedly different to that following removal of KDM2A/B. This may be explained by the important role played by SKP1 in regulating cell cycle progression. SKP1 was initially identified through its interaction with cyclin A/CDK complexes, and therefore named S-phase kinase-associated protein 1 (Zhang et al. 1995). SKP1 is required for ubiquitin-mediated proteolysis of cyclins Cln2 and Clb5 and the cyclin-dependent kinase inhibitor Sic1, and is therefore required for both G1/S and G2/M cell cycle transitions (Bai et al. 1996; Connelly and Hieter 1996). I attempted to measure the size of SKP1 knockout cells, but this fell outside of the range that could be accurately measured using the standard technique used in the lab. However, visual inspection of SKP1 knockout cells and the nuclei I prepared from them revealed that they were much larger than the cells prior to tamoxifen treatment. It is likely that these cells continued to grow in the absence of SKP1 but were unable to enter mitosis due to cell cycle inhibition, leading to increased transcription to maintain the concentration of RNA in the cytoplasm (Marguerat and Bahler 2012; Kempe et al. 2015; Padovan-Merhar et al. 2015). This would account for the dramatic increase in nucRNA levels I observed by RT-qPCR, and would mean that any effects of SKP1 knockout on transcription regulation itself are difficult to resolve from this dramatic accumulation of RNA that may result from cell cycle arrest in this line. This therefore limits the usefulness of comparing the gene expression changes in this

line to those following removal of KDM2A/B in the context of potential CGI-focused E3 ligase activity.

The observation that SKP1 knockout cells increased in size prompted me to investigate whether removal of KDM2A/B also led to cell size changes, given the broad, if less dramatic, increase in transcription that also occurred in this condition. I detected no significant change in cell size following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, while a comparison to cells known to have a smaller (SG4) or larger (HeLa) diameter confirmed the sensitivity of this approach (Supplementary Figure A.12). Therefore, the widespread changes in transcription following removal of KDM2A/B appear to have occurred in the absence of an alteration in cell size.

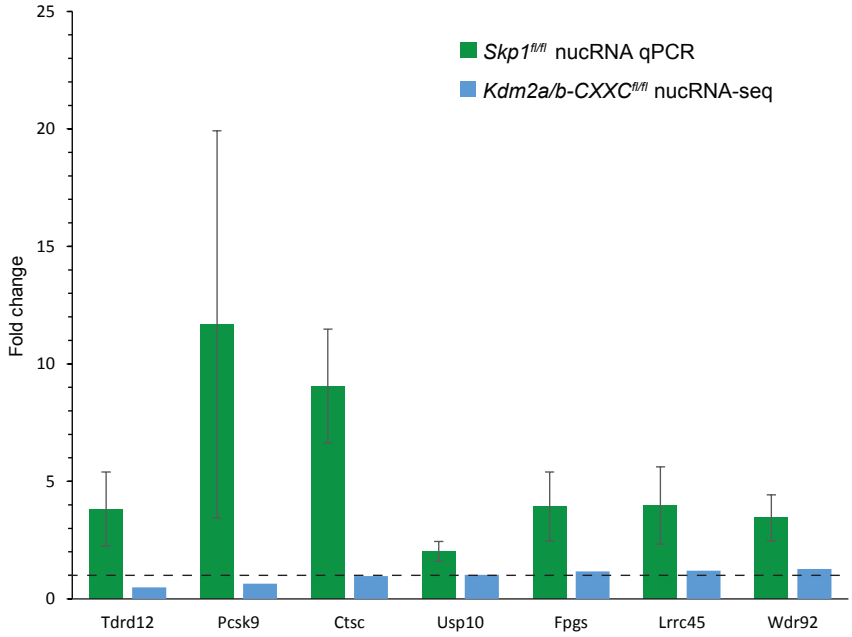


Figure 6.2: SKP1 knockout leads to a dramatic increase in nucRNA levels
RT-qPCR analysis of a panel of genes, showing the fold change in nucRNA levels following 72 h tamoxifen treatment of *Skp1*$^{fl/fl}$ mESCs to induce SKP1 knockout (green). The fold change in nucRNA-seq signal following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs is shown in blue for reference. The dashed line represents no change in expression. Expression was normalised to *Drosophila Rpl32* and *a-tubulin* genes and calibrated to the precise spike-in ratio, calculated using mouse and Drosophila qPCR analysis of genomic DNA prepared from the same nuclei preparations. Error bars show standard deviation of three biological replicates.

6.2.3 Ubiquitylome-mass spectrometry identifies potential targets of KDM2A/B E3 ligase activity

In order to more directly examine the potential E3 ligase activity of KDM2A/B, I adopted a ubiquitylome-mass spectrometry approach. This unbiased method involved identifying all ubiquitylated peptides present in nuclear extract, and comparing the profiles of these ubiquitylated peptides before and after 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs to remove KDM2A/B. Any ubiquitylated peptides that were under-represented in the tamoxifen treated sample represent potential targets of KDM2A/B E3 ligase activity (Figure 6.3). I decided to use the *Kdm2a/b-CXXC*$^{fl/fl}$ line for this experiment, as the most dramatic changes in ubiquitylation would likely occur following simultaneous removal of KDM2A and KDM2B.

To generate material for nuclear ubiquitylome analysis I treated cells with MG132 proteasome inhibitor prior to harvesting, to ensure that ubiquitylated proteins that are normally rapidly targeted for proteasomal degradation would instead accumulate. I prepared nuclear extract in the presence of protease inhibitors, N-ethylmaleimide and tricholoroacetic acid to inactivate deubiquitinating enzymes, and subsequently denatured proteins in 8M urea to prevent the degradation or turnover of ubiquitylated proteins during sample preparation. I then enriched the ubiquitylated peptides by immunoaffinity purification using an antibody specific for the di-glycine remnant left following trypsin digestion of ubiquitylated proteins, and used LC-MS/MS to identify the proteins from which these ubiquitylated peptides originated. I also used LC-MS/MS to analyse an input sample prior to immunoaffinity purification, which enabled the identification of peptides that changed in abundance following removal of KDM2A/B. I normalised for these changes in the subsequent analysis of ubquitylated peptides, as they would otherwise be indistinguishable from changes in protein ubiquitylation.

I identified peptides from 28 different proteins that significantly decreased in ubiquitylation following removal of KDM2A/B ($p<0.001$, Figure 6.4). In the tamoxifen treated *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs used for this study KDM2A/B are removed from chromatin but are still expressed at levels comparable to wild-type (Figure 4.2). Therefore, the proteins identified by this mass spectrometry screen are putative

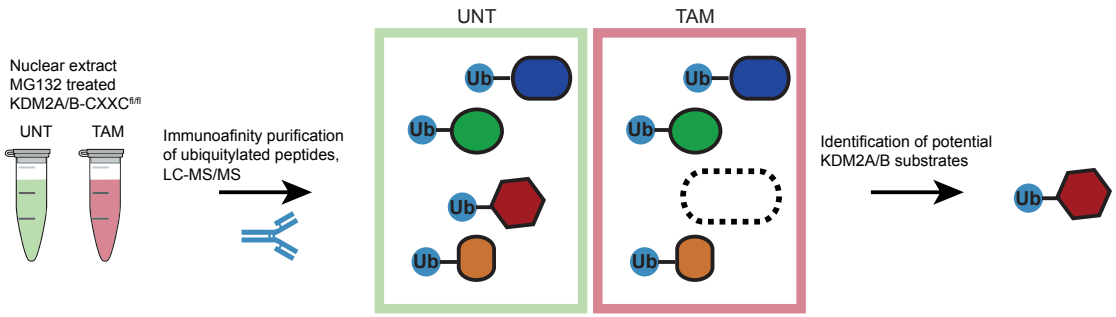


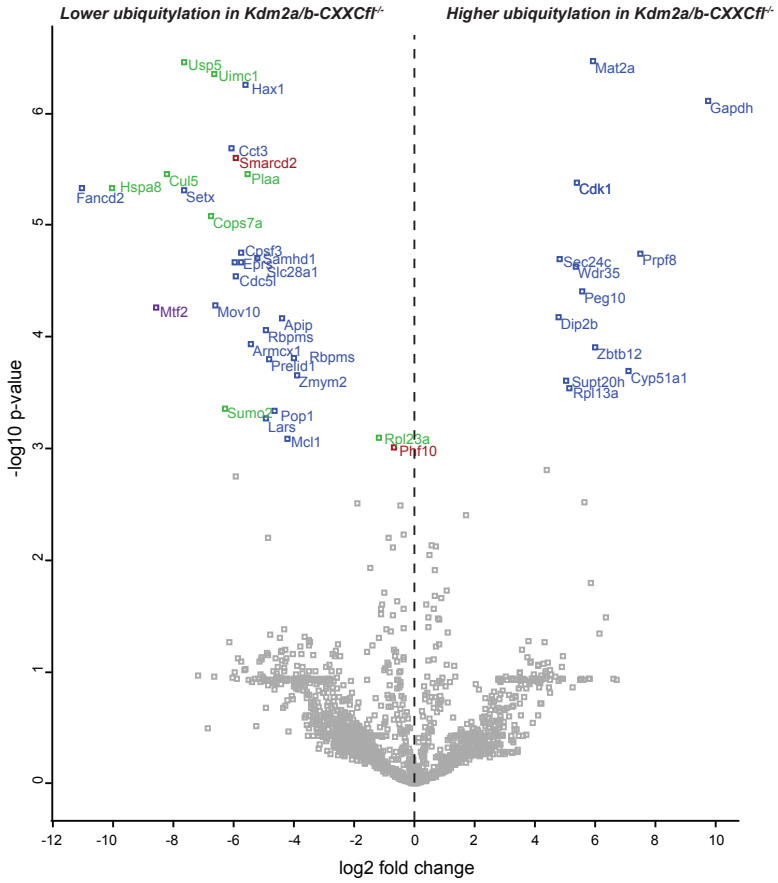
Figure 6.3: Ubiquitylome mass spectrometry identifies potential KDM2A/B substrates
Schematic of the ubiquitylome-mass spectrometry approach used to identify proteins that are differentially ubiquitylated following removal of KDM2A/B. mESCs are treated with the proteasome inhibitor MG132 prior to nuclear extraction and Filter Aided Sample Preparation including trypsin digestion. Ubiquitylated peptides are purified using an antibody specific for the di-glycine remnant left following trypsin digestion of ubiquitylated proteins (PTMScan® Ubiquitin Remnant Motif Kit, Cell Signalling) and identified using LC-MS/MS. Ubiquitylated peptides that are under-represented in the tamoxifen treated sample are potential targets of KDM2A/B E3 ligase activity.

substrates of the E3 ligase activity of KDM2A/B while bound to chromatin, as their ubiquitylation appears to be dependent on KDM2A/B chromatin occupancy. A smaller number (12 peptides) instead increased in ubiquitylation, and therefore do not represent potential KDM2A/B substrates. The increase in ubiquitylation of these peptides may be a secondary effect of KDM2A/B removal and the gene expression changes that result.

Interestingly, 8 of the 28 identified putative KDM2A/B substrates are themselves involved in protein ubiquitylation pathways (Figure 6.4 A and B - green, also see Supplementary Figure A.13 for a more detailed description), including Cullin-5, the Cops7a subunit of the COP9 signalosome complex responsible for Cullin deneddylation, and ubiquitin hydrolase Usp5. This may reflect cross-talk between KDM2A/B and other regulators of protein ubiquitylation. Ubiquitylation of an MTF2 peptide also significantly decreased following removal of KDM2A/B (Figure 6.4 - purple). MTF2 was recently shown to be important for recruitment of the PRC2 complex to its target sites throughout the genome (Perino et al. 2018), raising the interesting possibility that KDM2B, which itself recruits the PRC1.1 complex to CGI promoters, also modulates MTF2-dependent PRC2 recruitment to these target regions.

Two components of SWI/SNF complexes were also identified as potential substrates of KDM2A/B ubiquitylation: BAF45a, encoded by the *Phf10* gene, and BAF60b, encoded by the *Smarcd2* gene (Figure 6.4 - red). SWI/SNF are an evolutionarily conserved family of multisubunit ATP-dependent chromatin remodelling complexes, which can be grouped

A



B

-log10 p-value	log2 fold change	Gene name
6.46	-7.65867	Usp5
6.35	-6.64011	Uimc1
6.26	-5.61044	Hax1
5.69	-6.06478	Cct3
5.60	-5.92587	Smarcd2
5.46	-8.23717	Cul5
5.45	-5.52919	Plaa
5.33	-10.049	Hspa8
5.31	-7.64591	Setx
5.08	-6.76523	Cops7a
4.75	-5.77043	Cpsf3
4.70	-5.22314	Samhd1
4.67	-5.95934	Slc28a1
4.67	-5.77384	Eprs

-log10 p-value	log2 fold change	Gene name
4.54	-5.92823	Cdc5l
4.28	-6.63303	Mov10
4.24	-8.64773	Mtf2
4.17	-4.40491	Apip
4.06	-4.92474	Rbpms
3.93	-5.45111	Armcx1
3.79	-4.82983	Prelid1
3.66	-3.91579	Zmym2
3.36	-6.30551	Sumo2
3.34	-4.66413	Pop1
3.27	-4.9404	Lars
3.09	-1.20041	Rpl23a
3.09	-4.22189	Mcl1
3.01	-0.69662	Phf10

Figure 6.4: LC-MS/MS of peptides containing a ubiquitin remnant motif identifies potential KDM2A/B substrates

(A) Volcano plot illustrating the log2 fold change in abundance of ubiquitylated peptides following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs to induce removal of KDM2A/B, and the corresponding -log10 p-value for these changes. Peptides that change in abundance in nuclear extract prior to ubiquitin-remnant motif immunoaffinity purification were excluded. Significant changes are highlighted and labelled with the gene name ($p<0.001$, three biological replicates). Proteins involved in ubiquitylation pathways are coloured green, components of SWI/SNF complexes are coloured red, and MTF2 is highlighted in purple. LC-MS/MS and Perseus analysis were performed by Deniz Kaya.

(B) Table providing details of the peptides that significantly decreased in ubiquitylation following removal of KDM2A/B, coloured as above.

into two subfamilies, named BAF and PBAF in higher eukaryotes. These contain a set of highly conserved core subunits, and a number variant subunits that are thought to contribute to the targeting, regulation and lineage-specific functions of the complexes (extensively reviewed by Mohrmann and Verrijzer 2005). A core catalytic ATPase subunit (BRM or BRG1 in BAF complexes, BRG1 in PBAF complexes) acts as a DNA translocating motor that disrupts histone-DNA interactions, which can either positively or negatively regulate gene expression depending on the context (Simone 2006).

The BAF60b and BAF45a SWI/SNF subunits I identified as potential targets of KDM2A/B E3 ligase activity were both previously shown to be regulated by ubiquitin-mediated proteasomal degradation (Lores et al. 2010; Tatarskiy et al. 2017). A BAF60 subunit is found in most BAF and PBAF complexes, with BAF60a, BAF60b or BAF60c predominating in different cell types due to tissue-specific regulation (Goljanek-Whysall et al. 2014). BAF60 proteins have been shown to interact with a number of transcription factors, and this interaction is thought to mediate recruitment of SWI/SNF to target genes (Ito et al. 2001; Hsiao et al. 2003; Debril et al. 2004; Lickert et al. 2004; Flajollet et al. 2007). In contrast, the BAF45a subunit is a signature of PBAF complexes. BAF45a appears to be critical for stability of the PBAF complex: its depletion in human cells leads to PBAF destabilisation (Brechalov et al. 2014), and depletion of its homologue SAYP in *Drosophila* also destabilises the complex and dramatically decreases its binding to gene promoters (Vorobyeva et al. 2009). The reduced ubiquitylation of these subunits following KDM2A/B knockout, both of which are implicated in SWI/SNF complex regulation, raises the intriguing possibility that KDM2A/B regulate the level or activity of SWI/SNF complexes at CGI promoters.

The ubiquitylation of a number of other proteins involved in transcription regulation also decreased following removal of KDM2A/B. These included SETX, which is required for efficient transcription termination (Skourti-Stathaki et al. 2011), and RBPMS, which acts as a transcription coactivator for Smad proteins (Sun et al. 2006). Therefore, this experiment revealed a number of putative substrates whose regulation by KDM2A/B-mediated ubiquitylation could potentially lead to the transcriptional changes observed following removal of KDM2A/B.

6.2.4 BRG1 TSS enrichment does not correlate with the gene expression changes following removal of KDM2A/B

I was interested to investigate the potential regulation of SWI/SNF remodelling complexes by KDM2A/B. mESC ChIP-seq datasets are not available for the BAF45a and BAF60b subunits which I identified as potential substrates of KDM2A/B E3 ligase activity, however the chromatin binding profile of the core ATPase subunit BRG1 has been determined. BRG1 is the exclusive ATPase subunit of both BAF and PBAF complexes in mESCs (Ho et al. 2009), and is not found outside of SWI/SNF complexes (Lessard et al. 2007). Therefore, the BRG1 ChIP-seq profile should be representative of the chromatin occupancy of SWI/SNF complexes in mESCs. BRG1 is enriched at both distal regulatory regions and TSS, where its enrichment correlates with gene expression level (King et al. 2018; Ho et al. 2009; Kidder et al. 2009; Dieuleveult et al. 2016). The enrichment of both KDM2A/B and the core SWI-SNF component BRG1 at TSS is compatible with the possibility that KDM2A/B directly target SWI/SNF complexes while bound to chromatin. If the abundance of SWI/SNF complexes at TSS is indeed regulated by KDM2A/B, and if loss of this regulation leads to the gene expression changes following removal of KDM2A/B, then the change in SWI/SNF TSS occupancy following loss of KDM2A/B might be expected to correlate with the extent of transcription upregulation of the corresponding gene. I was unable to profile SWI/SNF binding in these conditions due to time constraints, but instead examined the correlation between BRG1 TSS enrichment in wildtype mESCs and the gene expression changes following removal of KDM2A/B.

There was a slight enrichment of BRG1 at the TSS of the genes that significantly increased in expression following removal of KDM2A/B compared to the average over all TSS (Figure 6.5 A). I investigated this enrichment in further detail, and detected a very slight depletion of BRG1 at the TSS of genes that strongly increased in expression following KDM2A/B removal (Q5 - Figure 6.5 B). Although this is a subtle difference, it might indicate the KDM2A/B play a role in limiting SWI/SNF accumulation at these TSS. Loss of KDM2A/B would then be expected to lead to a relatively large increase in SWI/SNF occupancy at these regions, which could potentially underlie the relatively large increase

in expression of the associated genes. If targeting of SWI/SNF complexes for degradation underlies the widespread transcriptional repression mediated by KDM2A/B, then I would expect a negative correlation between the gene expression changes following BRG1 knockout and those following KDM2A/B removal. However, there was poor correlation between these datasets (Figure 6.5 C). The alterations in SWI/SNF occupancy or activity that might result from removal of KDM2A/B may not be proportional to the TSS enrichment of BRG1 in wild-type cells. Therefore the lack of correlation in this preliminary analysis does not necessarily imply that KDM2A/B do not regulate SWI/SNF complexes. A direct investigation of the changes in SWI/SNF chromatin occupancy or activity following removal of KDM2A/B is required to address this possibility.

6.3 Summary and discussion

Here I have investigated the potential mechanisms by which KDM2A/B may broadly repress transcription. I profiled chromatin accessibility following removal of KDM2A/B, and used a ubiquitylome-mass spectrometry approach to identify putative targets of KDM2A/B E3 ligase activity.

ATAC-seq profiling revealed that conditional removal of KDM2A and/or KDM2B from chromatin led to virtually no change in the accessibility of regulatory regions throughout the genome, including both promoters and enhancers (Figure 6.1). Therefore the widespread transcription repression coordinated by KDM2A/B appears not to be a result of, or lead to, altered CGI accessibility. Simultaneous loss of both PRC1 and PRC2 transcription repressive complexes does not lead to changes in chromatin accessibility (King et al. 2018), and conditional knockout of ZF-CXXC family proteins CFP1 or FBXL19 also leads to minimal changes in accessibility (Hamish King, unpublished). Furthermore, the dramatic loss of transcription following inhibition of RNAPII in mammalian cells does not affect the nucleosome depleted state of CGI promoters, as measured by MNase-seq (Fenouil et al. 2012), or their elevated accessibility, as measured by ATAC-seq (King et al. 2018). Therefore, it appears that the accessibility of CGI promoters does not depend on their occupancy by a variety of transcription regulators, including KDM2A, KDM2B, FBXL19 and other ZF-CXXC family proteins and

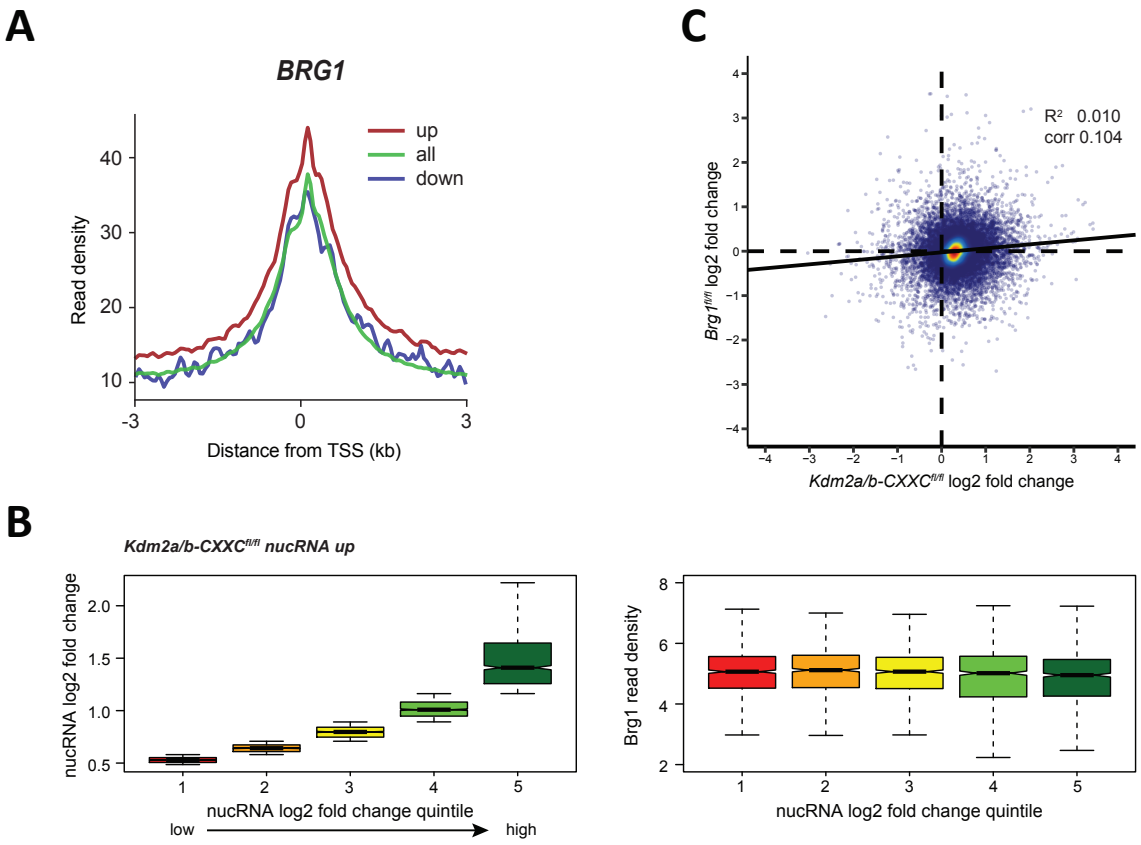


Figure 6.5: Poor correlation between BRG1 occupancy and transcription changes following removal of KDM2A/B

(A) Metaplot of BRG1 ChIP signal in the 6 kb region surrounding the TSS for all genes (n=20633, all) and genes that significantly increase (n=4528, up) or decrease (n=531, down) in expression following removal of KDM2A/B (ChIP-seq dataset: King et al. 2018).

(B) Left: bar chart demonstrating the division of genes into quintiles based on the log2 fold change in nucRNA-seq signal following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs to induce removal of KDM2A/B (left), for genes that significantly increased in expression (n=4528). Right: bar chart showing the BRG1 read density at the TSS for these quintiles.

(C) Scatter plot comparing the log2 fold change in nucRNA-seq signal following removal of KDM2A/B with that following BRG1 knockout (nucRNA-seq following BRG1 knockout: King et al. 2018). Solid line represents linear regression, with the coefficient of determination (R^2) and Spearman correlation coefficient (corr) annotated.

the polycomb repressive complexes, or on their transcribed state.

I proceeded to explore the potential E3 ligase activity of KDM2A/B. I began by investigating the transcriptional changes following knockout of the core SCF complex component SKP1. SKP1 knockout led to a dramatic accumulation of nuclear RNA (Figure 6.2). However, given the well-characterised role for SKP1 in cell cycle progression and the greatly enlarged cell size I observed in the knockout line, it seems likely that this accumulation of nucRNA resulted from cell cycle arrest but continued cell growth. Therefore I am unable to draw conclusions as to the direct role of SKP1 in transcription regulation, which might be linked to the E3 ubiquitin ligase activity of SKP1 interactors KDM2A/B at CGI promoters. Reassuringly, I did not observe a change in cell size following removal of KDM2A/B, suggesting that the widespread transcription upregulation in this line is unlikely to be a secondary effect of cell size.

Finally, I adopted a ubiquitylome-mass spectrometry approach to investigate the potential E3 ligase activity of KDM2A/B. I identified peptides from 28 different proteins that significantly decrease in ubiquitylation following removal of KDM2A/B, which are therefore putative substrates of KDM2A/B (Figure 6.4). These include a variety of proteins known to play chromatin-associated roles in transcription regulation, consistent with them being direct targets of the E3 ligase activity of CGI-bound KDM2A/B.

Interestingly, 8 of the 28 identified putative KDM2A/B substrates are themselves involved in protein ubiquitylation pathways. This is not a general observation for ubiquitylome-mass spectrometry profiles generated from MG132 treated mESCs (Emilia Dimitrova, unpublished), suggesting that there may be cross-talk between KDM2A/B and other components of the E3 ligase system. For example, Cops7a is part of the COP9 signalosome complex that mediates the deneddylation of the cullin subunits of cullin-RING E3 ligases (Wei and Deng 1999; Lyapina et al. 2001), leading to decreased ubiquitin ligase activity (Saha and Deshaies 2008), and also regulates cullin-RING ligases by a variety of other mechanisms (Cope and Deshaies 2003). Ubiquitylation of Cops7a by KDM2A/B could therefore modulate this regulatory activity. KDM2A/B do not appear to interact with a cullin subunit (Farcas et al. 2012; Gearhart et al. 2006; Koyama-Nasu et al. 2007; Gao et al. 2012), which generally acts as a molecular bridge

between the E3 ligase and substrate selector protein. However, given that KDM2A/B may possess both E3 ligase activity via their PHD domain and substrate selectivity via their LRR domains, this might bypass the need for a bridging Cullin subunit for their E3 ligase function. KDM2A/B-mediated Cops7a ubiquitylation might instead modulate the activity of other cullin-containing E3 ligase complexes. Indeed, Cul5 is also under-represented following removal of KDM2A/B, suggesting that KDM2A/B might regulate other E3 ligase complexes through subunit ubiquitylation. Alternatively, KDM2A/B-mediated Cops7a ubiquitylation might regulate COP9-mediated recruitment of the deubiquitinating enzyme UBP12 to chromatin (Zhou et al. 2003).

Ubiquitylation of Small Ubiquitin-like Modifier 2 (SUMO2) also decreased following removal of KDM2A/B (Figure 6.4). The SUMO modification is found on thousands of substrates, transferred via a three step SUMOylation cascade reminiscent of the ubiquitylation pathway (Hay 2005; Johnson 2004), often leading to modulation of protein-protein interactions. SUMO2, like ubiquitin, can form poly-SUMO2 chains (Tatham et al. 2001; Matic et al. 2008), and it has become clear that there is extensive crosstalk between the ubiquitylation and SUMOylation pathways (Ulrich 2005; Schimmel et al. 2008). While SUMO and ubiquitin were initially thought to represent distinct homotypic signals, subsequent studies revealed the formation of SUMO2-ubiquitin heterologous chains via the activity of SUMO-targeted ubiquitin ligases (Guzzo and Matunis 2013), which may promote more efficient proteasomal degradation (Aillet et al. 2012). These hybrid chains are specifically recognised by proteins containing tandem SUMO- and ubiquitin-interacting motifs (Guzzo et al. 2012), and may also be recognised by proteins with a SUMO-interacting motif in close proximity to a ubiquitin binding domain, leading to an extensive list of putative SUMO-ubiquitin chain binding proteins (Guzzo and Matunis 2013). The antibody I used in this study to enrich for ubiquitylated species does not cross-react with SUMO2 modified peptides, as these do not contain a di-gylcine remnant motif following trypsin digestion. Instead, the under-representation of SUMO2 ubiquitylation following removal of KDM2A/B raises the intriguing possibility that KDM2A/B can ubiquitylate SUMO2 chains. This could be tested using *in vitro* assays.

Another protein that decreased in ubiquitylation following removal of KDM2A/B is also

implicated in the regulation of SUMO-ubiquitin hybrid chains. *Uimc1* encodes RAP80, a subunit of the BRCA1-A complex, which was the first SUMO-ubiquitin chain receptor to be identified. The tandem SUMO- and ubiquitin-interacting motifs of RAP80 lead to its 80-fold higher affinity for hybrid SUMO-ubiquitin chains relative to homotypic chains, and mediate BRCA1 recruitment to DNA damage sites that become modified with SUMO-ubiquitin chains by RNF4 (Guzzo et al. 2012). Therefore, RAP80 ubiquitylation by KDM2A/B may also modulate this SUMO/ubiquitin pathway.

I identified two SWI/SNF subunits as potential targets of KDM2A/B E3 ligase activity, BAF45a and BAF60b, raising the possibility that KDM2A/B exerts its transcription regulatory role through the regulation of SWI/SNF chromatin remodelling complexes. Indeed, a number of studies have revealed an important role for ubiquitylation and proteasomal degradation in regulating SWI/SNF subunit levels (Chen and Archer 2005; Sohn et al. 2007; Lores et al. 2010; Tatarskiy et al. 2017), including both BAF45a and BAF60b (Lores et al. 2010; Tatarskiy et al. 2017). If KDM2A/B regulate transcription via their ubiquitylation of SWI/SNF subunits, then loss of this regulation might be expected to lead to alterations in chromatin accessibility. However, I detected virtually no change in chromatin accessibility following removal of KDM2A/B (Figure 6.1). Interestingly, Trizzino et al. recently showed that depletion of ARID1A, a core subunit of BAF SWI/SNF complexes, leads to widespread transcription inhibition in the absence of pronounced changes in promoter accessibility in ovarian clear cell carcinoma cells (Trizzino et al. 2018). Instead, a severe impairment of RNAPII pausing appears to explain the transcription changes. This raises the possibility that modulation of SWI/SNF complexes by KDM2A/B could directly regulate the transcription cycle, independently of SWI/SNF dependent regulation of chromatin accessibility.

RNAPII pausing has emerged as a key step in transcription regulation in higher eukaryotes, allowing integration of multiple regulatory signals to fine-tune gene expression (Adelman and Lis 2012). Knockout of the pause-inducing factor NELF in mESCs led to the deregulated expression of genes involved in cellular growth and proliferation, with an enrichment of genes involved in metabolic and biosynthetic processes (Williams et al. 2015). I observed an enrichment of similar gene ontology terms in the set of non-polycomb target genes that significantly increased in expression

following removal of KDM2A/B (Figure 4.12). It is tempting, therefore, to speculate that KDM2A/B broadly modulate transcription by regulating the pausing of RNAPII initiating from CGI gene promoters. This may be mediated either via their regulation of SWI/SNF subunits or via an independent mechanism. This could be investigated using GRO-seq to map the distribution of engaged RNAPII throughout the gene body and to determine whether the pausing index is significantly affected following removal of KDM2A/B (Core et al. 2008). This could be combined with profiling of total RNAPII, Serine 5-phosphorylated paused RNAPII and Serine 2-phosphorylated elongating RNAPII by ChIP-seq.

Intriguingly, the BAF45a and BAF60b SWI/SNF subunits I identified in the ubiquitylome mass spectrometry screen represent minor components of ESC-specific BAF complexes, although they are both expressed in mESCs. BAF45d is the most abundant BAF45 subunit in SWI/SNF complexes mESCs (Ho et al. 2009), whereas BAF45a predominates in neural stem cells and is required for neural progenitor proliferation (Lessard et al. 2007). Similarly BAF60a predominates in mESC BAF complexes, whereas BAF60b is the dominant BAF60 subunit in embryonic fibroblasts (Ho et al. 2009) and is required for granulopoiesis (Michel and Kadoch 2017; Witzel et al. 2017). Therefore, KDM2A/B may play a role in regulating the incorporation of tissue-specific subunits into BAF complexes, which may have important functional consequences. Given the cell type-specific modulation of SWI/SNF subunit incorporation, the potential role played by KDM2A/B in regulating SWI/SNF complexes might lead to distinct functional consequences at different developmental stages in different tissues.

Further work is required to validate these putative substrates of KDM2A/B E3 ligase activity, and to determine whether KDM2A or KDM2B or both are required for their ubiquitylation. This could include investigation of whether KDM2A/B interact with their candidate substrates, for example by expression of HA-tagged candidate and FLAG-tagged KDM2A or KDM2B followed by co-immunoprecipitation experiments. However, interaction between E3 ligases and their substrates are often transient, and therefore may not be detected using this approach. In this case, ligase trapping could be used to strengthen the interaction between KDM2A/B and their potential substrates. In this affinity purification approach the E3 ligase is expressed fused to a FLAG tag and a

polyubiquitin-binding domain, which stabilises the interaction of the E3 ligase with its ubiquitylation substrates. Hexahistidine-tagged ubiquitin is also expressed, to enable robust identification of ubiquitylated substrates: FLAG immunoprecipitation is first used to enrich for substrates of the E3 ligase, which remain bound via the ligase trap, then affinity purification under denaturing conditions is used to capture proteins covalently bound to hexahistidine ubiquitin, which represent ubiquitylation substrates of the E3 ligase (Mark et al. 2016). These substrates can subsequently be identified by western blotting. KDM2A/B E3 ligase activity should be investigated using *in vitro* assays, in which recombinant KDM2A or KDM2B is combined with E1, E2, ATP and FLAG- or Cy5- tagged ubiquitin. Auto-ubiquitylation can then be detected via western blotting or in-gel fluorescence, respectively. *In vitro* assays of KDM2B E3 ligase activity have yielded conflicting results to date (Emilia Dimitrova and Jess Kelley, unpublished), while KDM2A E3 ligase activity has not been examined *in vitro* to my knowledge.

If E3 ligase activity is required for the regulation of transcription by KDM2A/B, then expression of KDM2A/B lacking a region required for substrate recognition (LRR, F-box domains) or E3 ligase activity (PHD domain) should recapitulate the widespread transcription upregulation observed following removal of KDM2A/B. However, this may prove difficult as attempts to overexpress the KDM2 proteins lacking either their F-box or PHD domain revealed that these domains are important for protein stability (Klose lab, unpublished). Furthermore, the LRR domain of KDM2B is required for its binding to the PRC1.1 complex via PCGF1 and BCORL1 (Wong et al. 2016), therefore deletion of this region would lead to loss of both substrate ubiquitylation and PRC1 interaction and not allow these two activities to be uncoupled. An alternative approach would be to focus on characterisation of the ubiquitylation substrate. Removal of KDM2A/B might lead to an accumulation of this substrate at CGI promoters, due to loss of KDM2A/B-mediated ubiquitylation and subsequent proteasomal degradation. If this accumulation is responsible for the widespread increase transcription following removal of KDM2A/B, then knockout of this substrate would be expected to counteract this increase in transcription. Furthermore, if the residue that is ubiquitylated by KDM2A/B is identified, then mutation of this residue alone should be sufficient to counteract this increase in transcription. Combination of the ligase trapping approach, to confirm that

the candidate is ubiquitylated by KDM2A/B, and knockout or mutation of this candidate, potentially preventing transcription deregulation following loss of KDM2A/B, has the potential to provide more convincing evidence of transcription regulation via KDM2A/B E3 ligase activity.

In conclusion, I have shown that removal of KDM2A/B from chromatin does not lead to immediate changes in chromatin accessibility. Furthermore, I have identified putative substrates of KDM2A/B E3 ligase activity, including a number of proteins involved in ubiquitylation pathways and components of SWI/SNF remodelling complexes. Further work is required to validate these potential substrates, and to explore whether their regulation by KDM2A/B-mediated ubiquitylation is an important mechanism by which KDM2A/B broadly regulate transcription. Additional mechanisms by which KDM2A/B may regulate transcription should also be explored, such as the regulation of RNAPII pausing.

7

Conclusions and future directions

KDM2A/B are recruited to CGI gene promoters throughout the genome via their ZF-CXXC domain (Blackledge et al. 2010; Farcas et al. 2012; Wu et al. 2013; He et al. 2013), and catalyse the demethylation of H3K36me2 via their JmjC catalytic domain (Tsukada et al. 2006; He et al. 2008). Studies of individual genes have shown that localisation of KDM2A/B to CGI promoters is correlated with a local depletion in H3K36me2, however the genome-wide role played by KDM2A/B in regulating H3K36 methylation has remained unclear. Here I showed that CGI gene promoters throughout the genome are depleted of H3K36me2 in mESCs, and that the histone demethylase activity of KDM2A/B contributes modestly to this depletion (Chapter 3). Thus KDM2A/B demethylase activity is not absolutely required the generation of a H3K36me2-depleted state, but instead may function as a fail-safe mechanism to protect CGI promoters from spurious H3K36 methylation. It is likely that functional redundancy between KDM2A/B and the KDM4 family, which also appear to limit the accumulation of H3K36 methylation at gene promoters (Pedersen et al. 2014; Pedersen et al. 2016), may contribute to the persistence of a H3K36me2-depleted CGI state in the absence of KDM2A/B.

H3K36 methylation is inhibitory to transcription initiation in yeast, and may play an evolutionarily conserved role in mammalian cells (Section 1.1.7). Therefore, it has been proposed that KDM2A/B contribute to the generation of a transcriptionally permissive chromatin state at CGI gene promoters by locally removing this repressive mark (Blackledge et al. 2010; Blackledge and Klose 2011; Deaton and Bird 2011). In contrast to this prevailing model, I showed that the long isoforms of KDM2A/B contribute little to transcription regulation. As demethylase activity is unique to KDM2A/B-LF, this reveals that H3K36me2 demethylation by KDM2A/B is generally not required for normal transcription in mESCs. Given the likelihood that KDM2 and KDM4 proteins cooperate to generate the H3K36me2-depleted state that characterises CGI promoters, simultaneous conditional catalytic point mutation of these demethylase

proteins would be required to determine the overall contribution of H3K36me2 to transcription regulation at CGI promoters. Alternatively, the transcriptional effect of the accumulation of H3K36 methylation at CGI promoters could be studied by targeting an H3K36me2 methyltransferase to these regions, for example by fusion to dead Cas9 or to a ZF-CXXC domain. Robust depletion of H3K36me2 at CGI promoters may be more crucial during cell fate transitions, in which a new gene expression programme must be established, compared to during maintenance of steady-state levels of transcription in mESCs. Strong H3K36me2 depletion may be required for a CGI promoter state that is highly permissive to transcription initiation and facilitates gene induction. Therefore, the demethylase activity of KDM2A/B may play a more important role during differentiation. Whether this activity would predominantly promote or constrain transcription remains unclear.

The majority of transcripts increased in expression following removal of KDM2A/B from chromatin, revealing an unexpected and widespread role for KDM2A/B in transcriptional repression (Chapter 4). Given that loss of demethylase activity as a result of KDM2A/B-LF knockout had minimal effect on transcription, this instead suggests that demethylase-independent role(s) of KDM2A/B are important for transcription regulation. I dissected the relative contributions of KDM2A and KDM2B to transcriptional repression, demonstrating that KDM2B plays the predominant role in constraining transcription (Chapter 5). Importantly, this repression is not limited to polycomb target genes, which are known to be targeted by the KDM2B-PRC1.1 repressive complex. Instead, KDM2B appears to act broadly to dampen the transcription of the majority of genes in mESCs. KDM2A is required to restrict the transcription of a small subset of genes in the absence of KDM2B. KDM2A also appears to more broadly cooperate with KDM2B and play a minor role in restricting transcription, particularly the transcription of lowly expressed genes. In contrast, the KDM2A/B-SF paralogue FBXL19 does not appear to cooperate with KDM2A/B to function as a general transcription repressor.

I investigated potential demethylase-independent mechanisms by which KDM2A/B may regulate transcription (Chapter 6). The transcriptional repression coordinated by KDM2A/B does not appear to be a result of, or to result in, changes to the accessibility

of gene promoters. Instead I provided preliminary evidence that KDM2A/B may act as E3 ubiquitin ligases, by identifying a number of candidate substrates. These include a number of proteins involved in ubiquitylation pathways and subunits of SWI/SNF chromatin remodelling complexes. This raises the intriguing possibility that KDM2A/B may ubiquitylate one or more of these candidates while they are associated with CGI promoters, which may regulate their stability or activity. If this substrate acts to broadly modulate transcription, then its regulation by ubiquitylation could potentially underlie the transcriptional repression coordinated by KDM2A/B. SWI/SNF complexes were recently implicated in broadly promoting transcription by regulating RNAPII pausing (Trizzino et al. 2018). It is possible that the transcription repression exerted by KDM2A/B is mediated through the regulation of RNAPII pausing, which could occur via SWI/SNF ubiquitylation or through an independent mechanism.

The function of KDM2A has remained enigmatic, as KDM2A is generally not thought to be incorporated into protein complexes and appears to function redundantly with KDM2B to remove H3K36me2. Here I propose that KDM2A plays an, albeit minor, role in transcriptional repression, and this is likely to be independent of its demethylase activity. It will be interesting to determine the mechanism by which KDM2A dampens transcription, which may involve E3 ubiquitin ligase activity and may be shared with or distinct from that of KDM2B, and to explore whether KDM2A plays a more dominant role in transcription regulation in other cell types or developmental stages.

I have uncovered a novel and widespread role for KDM2B in constraining transcription. This appears to be one of multiple functions exerted by KDM2B at CGI gene promoters that may promote transcriptional repression. Importantly, CGI promoters generally remain free of DNA methylation irrespective of the transcription level of the associated gene. Therefore, they are continually and generically targeted by chromatin modifying complexes containing ZF-CXXC proteins via the interaction between this domain and non-methylated CpG dinucleotides. KDM2B has been proposed to recruit PRC1.1 complexes to sample the transcriptional state of CGI promoters. If an inactive promoter is encountered then this may lead to the nucleation of a polycomb domain to reinforce transcriptional repression (Farcas et al. 2012; Blackledge et al. 2014). In addition, H3K36me2 demethylation by KDM2B may facilitate the formation of a repressive

polycomb-occupied CGI state, as H3K36me2 is inhibitory to H3K27me3 deposition by PRC2 (Section 1.1.8). Indeed, promoter H3K36 demethylation appears to be linked to transcriptional repression in some conditions (Section 1.3.3), despite the apparently paradoxical link between H3K36 methylation and the inhibition of transcription initiation. H3K36 demethylation by KDM2B may, therefore, contribute to the bistability of CGI promoters, by facilitating both transcription initiation and the formation of repressive polycomb domains. I have revealed that the broad targeting of KDM2B to all CGI promoters is coupled to another transcription-limiting function, which appears to be independent of the polycomb system and demethylase activity, and may instead rely on E3 ubiquitin ligase activity of KDM2B. Therefore, KDM2B appears to play multifaceted roles in transcriptional regulation, as its broad targeting to CGI promoters is linked to diverse mechanisms that modulate transcription. These three activities may cooperate to promote gene repression of a subset of lowly transcribed genes, by restricting their already limited transcription and facilitating polycomb domain formation.

Why would KDM2B regulate transcription by three different mechanisms? A simple explanation is that these pathways act redundantly to ensure robust transcriptional repression. However, this explanation does not seem to be sufficient, as the different modes of transcriptional regulation appear to regulate distinct sets of genes and may predominate under different conditions. Whereas KDM2B is involved in the formation of polycomb domains at the CGI promoters of a subset of lowly expressed or inactive genes to enforce their repressed state, KDM2B also appears to constitutively dampen the transcription of both lowly and highly expressed genes by an independent mechanism. In contrast, H3K36 demethylation by KDM2B does not appear to contribute to the regulation of steady state gene expression in mESCs, but may play a more important role in establishing new gene expression patterns during cell fate transitions. Perhaps flexibility is key to these multifaceted roles of KDM2B, enabling this transcriptional regulator to exert overlapping but distinct regulatory activities. The evolution of KDM2 paralogues KDM2A and KDM2B in vertebrates and of short isoforms of these proteins which lack demethylase activity may have contributed to this flexibility. Developmental control of the relative expression of the long and short isoforms may allow demethylase-dependent or -independent roles to predominate at different stages, and the differential

expression of KDM2A and KDM2B may modulate the relative contributions of their polycomb-dependent and -independent modes of transcriptional repression.

Interestingly, all three modes of transcriptional regulation exerted by KDM2B appear to either facilitate or restrict transcription, rather than to play an instructive role. While H3K36me2 demethylation by KDM2B at CGI promoters does not appear to play an important role in transcription regulation in mESCs, it may broadly contribute to the generation of a transcriptionally permissive chromatin state under certain conditions, or alternatively may facilitate the formation of repressive polycomb domains by enabling PRC2-mediated H3K27 trimethylation. KDM2B also acts to broadly constrain transcription by a demethylase-independent mechanism. Furthermore, KDM2B-PRC1.1 is thought to reinforce, rather than initiate, gene repression. The broad targeting of KDM2B to CGI gene promoters would be inconsistent with a dominant role in enforcing transcriptional activation or repression, as this would lead to a widespread perturbation of gene expression. Instead, KDM2B appears to act to broadly modulate transcription, which may have different functional consequences depending on the transcriptional status and chromatin environment.

Why do KDM2A/B modestly constrain the transcription of the majority of genes? One possibility is that this functions to increase the threshold of activating signal from lineage-specific transcription factors that is required to achieve robust gene induction. This could prevent inappropriate gene activation as a result of low levels of transcription factor binding, contributing to the robustness of cell lineage specification during development. Indeed, the ATP-dependent chromatin remodeller CHD4 has been suggested to play such a role through its action at promoters to constrain gene expression (O'Shaughnessy-Kirwan et al. 2015). CHD4 knockout leads to the simultaneous expression of multiple lineage markers, which may underlie a failure to establish lineage-specific gene expression programmes. Genes with low expression are more strongly upregulated following loss of the transcription constraining activity of KDM2A/B. These lowly transcribed genes may be susceptible to the reduction in the threshold of low-level activation signals that are required to establish their more robust transcription, whereas genes that are already highly expressed may be relatively unperturbed as they are already subject to robust transcription activatory signals. Therefore, while KDM2A/B

act to broadly repress transcription, this may have different functional consequences depending on the transcriptional status of the CGI promoter. This offers a potential explanation for how demethylase- and polycomb-independent transcriptional repression by KDM2A/B might particularly affect a subset of genes, despite the generic targeting of KDM2A/B to all CGI promoters.

An emerging theme in the field of chromatin regulation is that chromatin modifying enzymes exert important activities in addition to their well characterised roles in regulating chromatin structure (Biggar and Li 2014; Dimitrova et al. 2015). These alternative activities may, in some instances, play a more important role in transcriptional regulation. This study contributes further evidence of this phenomenon, revealing a novel and widespread role for KDM2A/B in constraining transcription that, in contrast to H3K36 demethylase activity, is important for the regulation of gene expression in mESCs. The embryonic lethality phenotypes of knockouts of chromatin modifiers are often ascribed to loss of their chromatin modifying activity, however these studies are confounded by the loss of the other activities of these multifunctional proteins. Systematic dissection of the contribution of their chromatin modifying activity using conditional catalytic point mutants is therefore required to truly understand its contribution to gene regulation. Such studies may reveal that, like KDM2A/B, a variety of chromatin modifiers play overlapping and multifaceted roles in transcriptional regulation. This may contribute to their ability to fine-tune gene expression according to cellular context.

A

Appendices

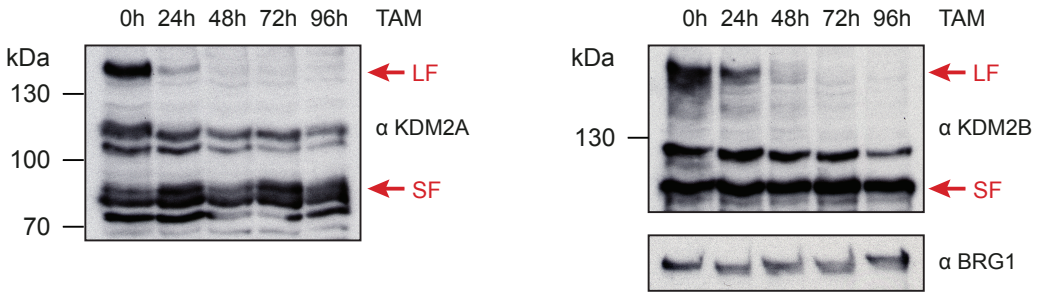


Figure A.1: Tamoxifen treatment time-course for *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs
Western blot showing loss of KDM2A/B-LF but not KDM2A/B-SF following 96 h tamoxifen (TAM) treatment of *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs. A large proportion of KDM2A/B-LF protein is lost after 24 h TAM treatment. BRG1 is a loading control for both KDM2A and KDM2B westerns.

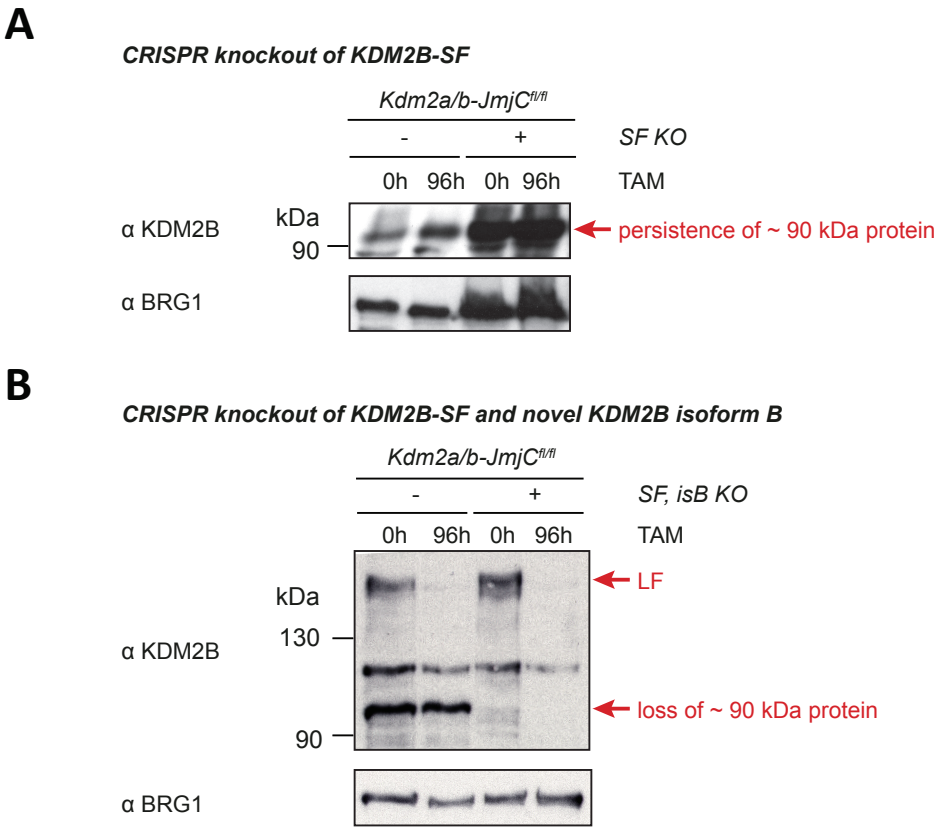


Figure A.2: A constitutive CRISPR knockout study confirms the expression of novel KDM2B isoform B in mESCs
(A) Western blot showing the persistence of a prominent band at around 90 kDa that is detected by the KDM2B-specific antibody following constitutive knockout of KDM2B-SF. BRG1 is a loading control.
(B) Western blot showing loss of the approximately 90kDa band following combined CRISPR knockout of KDM2B-SF and KDM2B isoform B. This indicates that continued expression of KDM2B isoform B masked loss of KDM2B-SF in (A), suggesting that isoform B is relatively highly expressed in mESCs. Combination of CRISPR knockout of KDM2B-SF and isoform B with 96 h tamoxifen (TAM) treatment of *Kdm2a/b-JmjC*$^{fl/fl}$ mESCs to induce loss of isoforms incorporating the JmjC-encoding exons leads to complete loss of all KDM2B protein. BRG1 is a loading control.

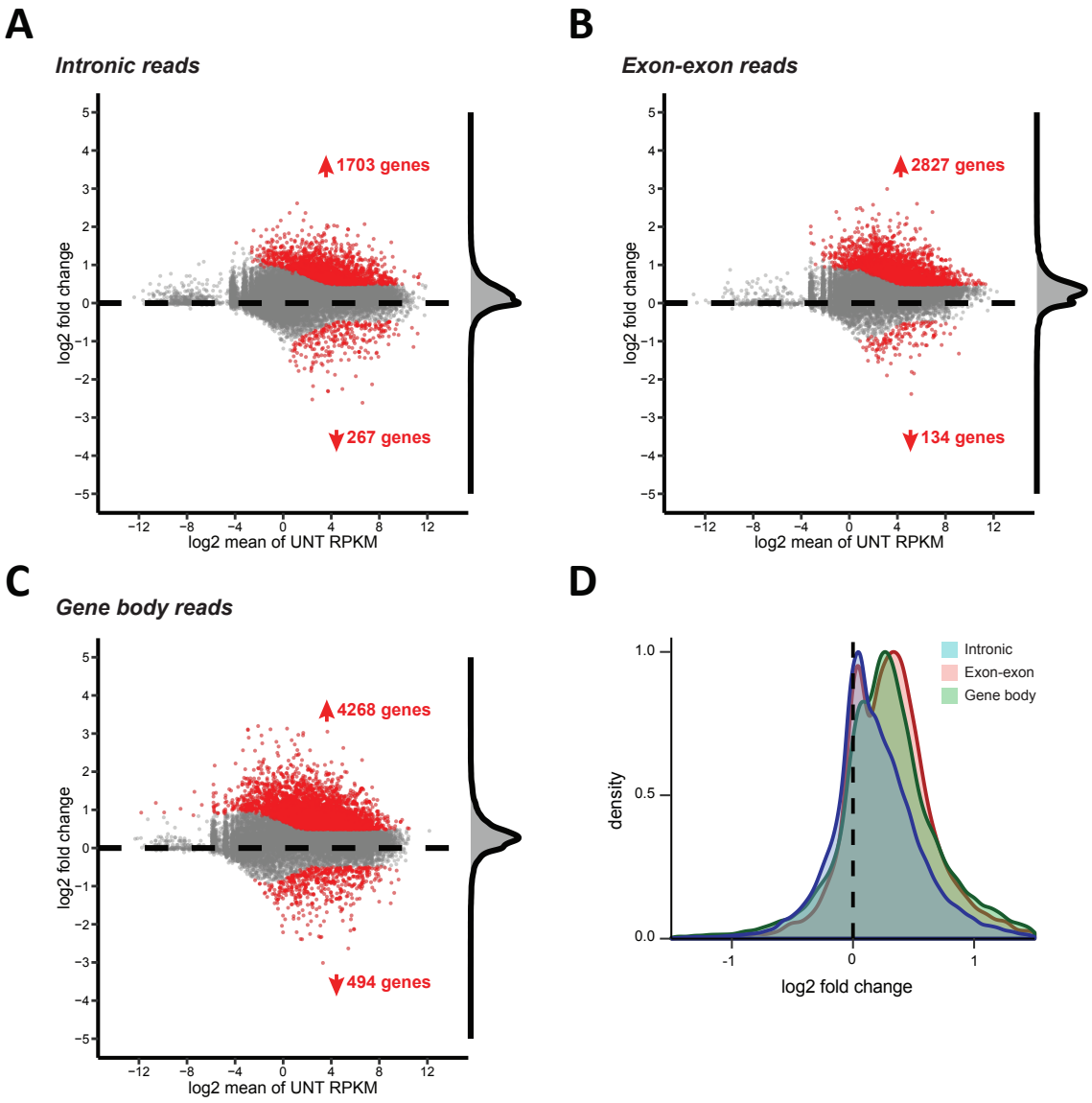


Figure A.3: Removal of KDM2A/B leads to widespread upregulation of nascent transcripts
(A) Scatter plot of log2 mean nucRNA-seq paired reads per kilobase per million mapped reads (RPKM) that map to intronic regions in *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs for a non-redundant mm10 gene set (n=20633, four biological replicates), plotted against log2 fold change in intronic nucRNA-seq signal following 96 h tamoxifen treatment to induce KDM2A/B removal. Libraries were calibrated to *Drosophila* spike-in. Significant changes were identified using DESeq2 and are coloured red ($p<0.05$, fold change>1.4). A density plot of the log2 fold change in nucRNA-seq signal is shown to the right.
(B) As (A), but for nucRNA-seq paired reads that map across exon-exon junctions, reflecting spliced mRNA.
(C) As (A), but for nucRNA-seq paired reads that map throughout the gene body region, irrespective of intronic or exonic regions.
Reads were assigned to intronic, exon-exon and gene body regions by Angelika Feldmann, using a custom pipeline.

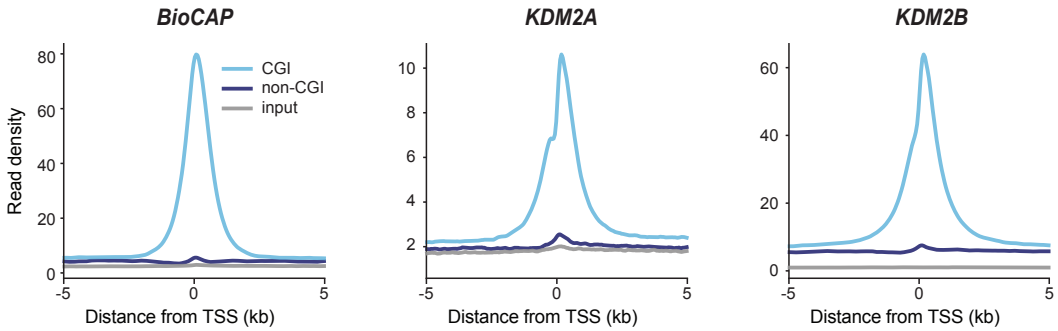


Figure A.4: Minimal enrichment of KDM2A, KDM2B and BioCAP signal at non-CGI TSS
Metaplots of BioCAP, KDM2A and KDM2B levels at all CGI (n=14106) or non-CGI TSS (n=6527). Published BioCAP and ChIP-seq datasets were used: Long et al. 2013; Farcas et al. 2012.

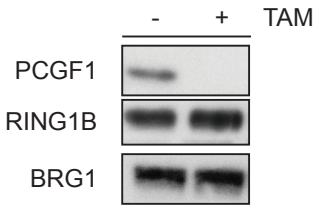


Figure A.5: Conditional knockout of PCGF1 in *Pcgf1*$^{fl/fl}$ mESCs
Western blot showing knockout of PCGF1 following tamoxifen (TAM) treatment of *Pcgf1*$^{fl/fl}$ mESCs. RING1B expression does not change following tamoxifen treatment. BRG1 is a loading control. Figure courtesy of Nadya Fursova, Klose lab.

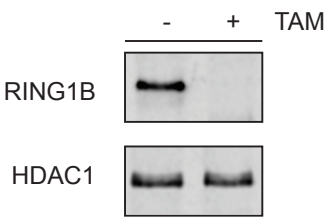


Figure A.6: Conditional knockout of RING1B in *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ mESCs
Western blot showing knockout of RING1B following tamoxifen (TAM) treatment of *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ mESCs. HDAC1 is a loading control. Figure courtesy of Neil Blackledge, Klose lab.



Figure A.7: Tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs does not perturb RING1B or SUZ12 expression
Western blot showing protein levels of SUZ12, RING1B and PCGF1 before and after 96 h tamoxifen (TAM) treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs to induce removal of KDM2A/B. BRG1 is a loading control.

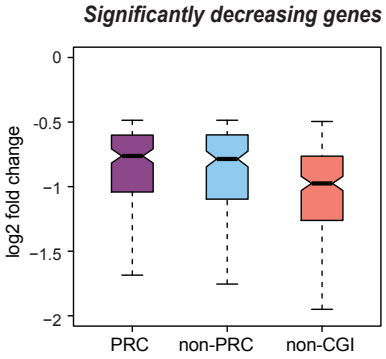


Figure A.8: Non-CGI genes that significantly decrease in expression following KDM2A/B removal are relatively strongly downregulated
Box plot showing log2 fold change in gene expression following 96 h tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, for significantly decreasing genes. Genes are subdivided into polycomb-target (PRC, n=161) or non-polycomb target (non-PRC, n=166) CGI genes and non-CGI genes (n=205).

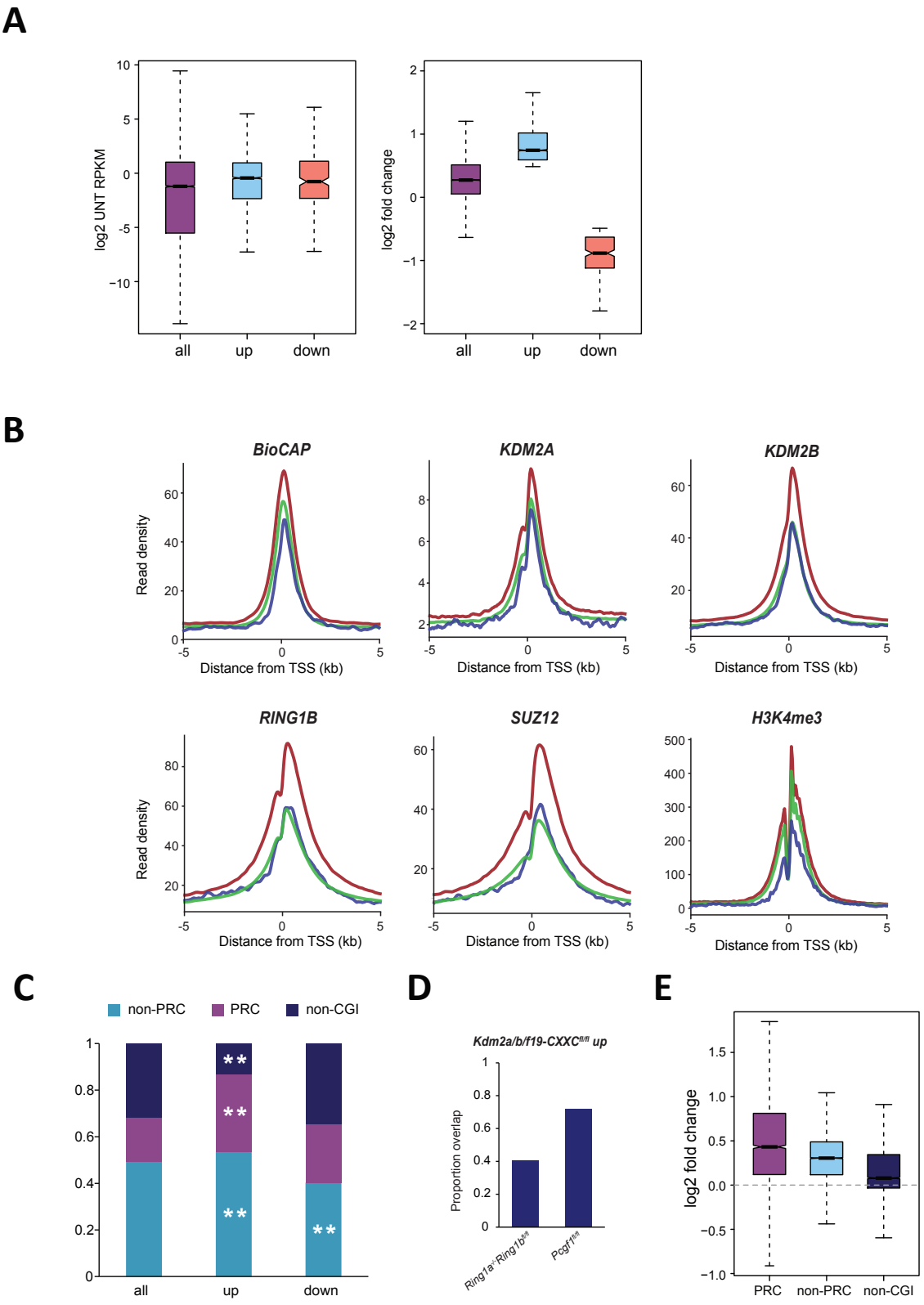


Figure A.9: Gene expression changes following removal of FBXL19 and KDM2A/B parallel those following loss of KDM2A/B

(A) Box plots showing starting expression level (log2 UNT RPKM) and log2 fold change in gene expression following 96 h tamoxifen treatment of *Kdm2a/b/f19-CXXC*$^{fl/fl}$ mESCs to induce removal of KDM2A/B and FBXL19, for all genes (n=20633, all) and genes that significantly increased (n=4277, up) or decreased (n=243, down) in expression.
(B) Metaplots of KDM2A, KDM2B, RING1B, SUZ12 and H3K4me3 ChIP signal and BioCAP signal at the TSS for all genes, and for significantly upregulated or downregulated subsets (Long et al. 2013; Farcas et al. 2012; Rose et al. 2016; Brown et al. 2017).
(C) Bar chart showing the proportion of genes that belong to non-CGI and either polycomb-target (PRC) or non-polycomb target (non-PRC) CGI classes, for all genes and the significantly upregulated and downregulated subsets. PRC target genes were defined by RING1B and SUZ12 occupancy at the TSS (ChIP-seq datasets: Rose et al. 2016). Asterisks represents empirical p-value (* represents p<0.01, ** represents p<0.001), relative to all transcripts.
(D) Bar chart showing the proportion of significantly upregulated genes following tamoxifen treatment of either *Ring1a*$^{-/-}$*Ring1b*$^{fl/fl}$ or *Pcgf1*$^{fl/fl}$ mESCs that were also significantly upregulated following tamoxifen treatment of *Kdm2a/b/f19-CXXC*$^{fl/fl}$ mESCs.
(E) Box plots showing log2 fold change in gene expression following simultaneous removal of FBXL19 and KDM2A/B, for polycomb target (PRC), non-polycomb target (non-PRC) and non-CGI genes.

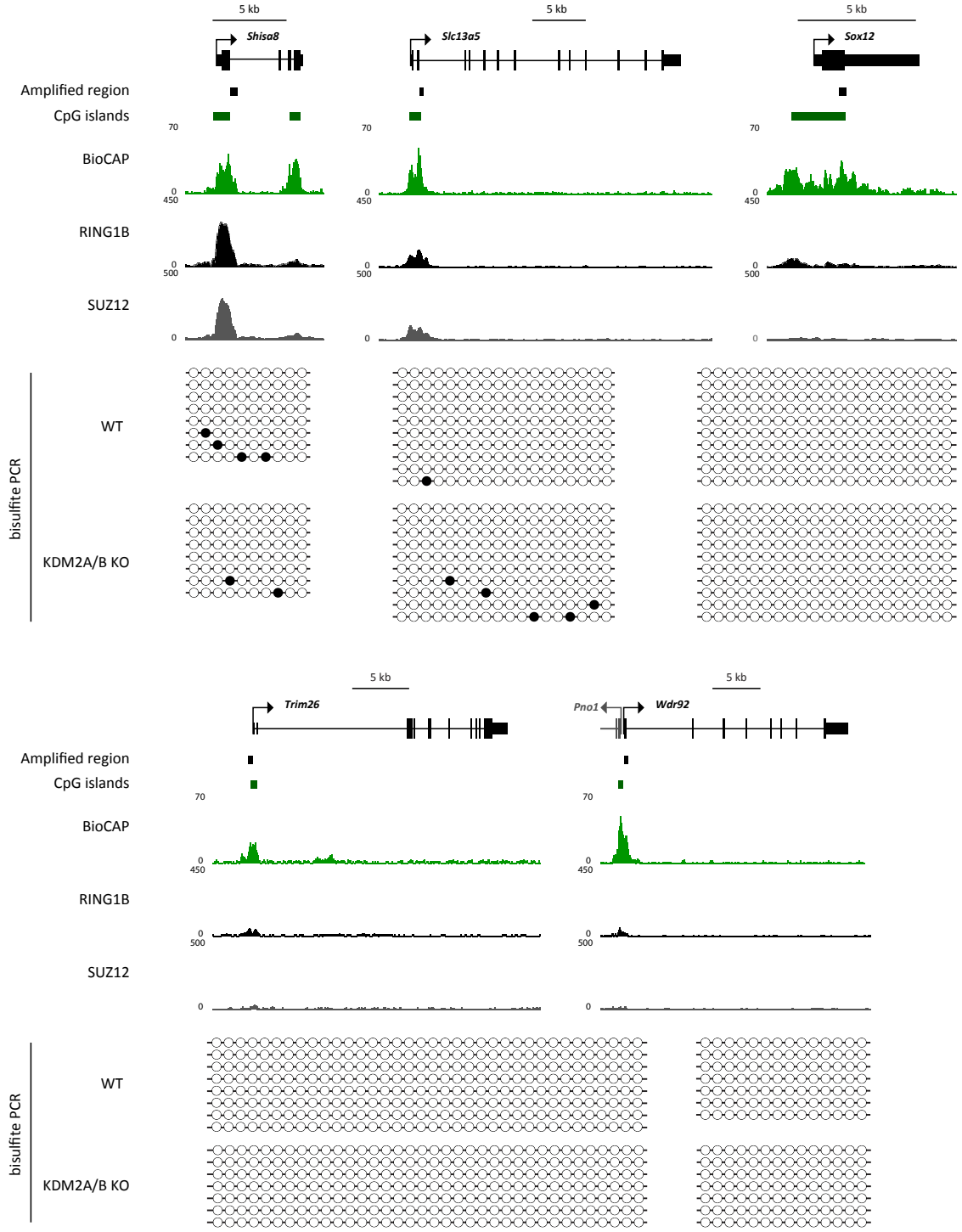


Figure A.10: KDM2A/B knockout is not accompanied by an increase in CGI methylation
Genome snapshots of a panel of CGI-associated genes, including polycomb targets and non-targets. BioCAP, RING1B and SUZ12 ChIP-seq signal are shown for reference (Published datasets: Long et al. 2013; Rose et al. 2016). The promoter region amplified for bisulfite PCR is indicated by a black bar, and the methylation state of individual CpG dinucleotides in this region is represented below, using genomic DNA from either wild-type mESCs or KDM2A/B constitutive knockout mESCs. Empty and filled circles represent non-methylated and methylated CpG dinucleotides, respectively. The TSS of the *Sox12* gene was previously shown to increase in DNA methylation following constitutive KDM2B knockout (Boulard et al. 2015).

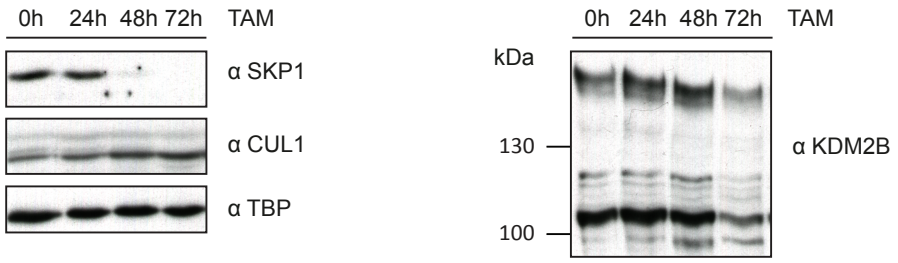


Figure A.11: Tamoxifen treatment of *SKP1*$^{fl/fl}$ mESCs leads to SKP1 knockout
Western blot showing knockout of SKP1 following 72 h tamoxifen (TAM) treatment of *Skp1*$^{fl/fl}$ mESCs. CUL1 and KDM2B are shown for reference: KDM2B expression decreases following SKP1 knockout due to reduced protein stability. TBP is a loading control.

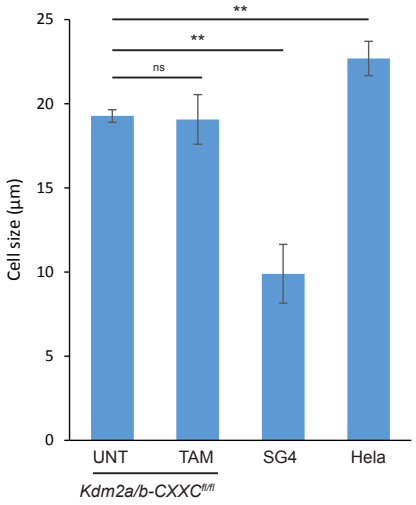


Figure A.12: Tamoxifen treatment of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs does not lead to a significant change in cell size
Bar graph showing the average cell diameter of *Kdm2a/b-CXXC*$^{fl/fl}$ mESCs, before (UNT) or after (TAM) 96 h tamoxifen treatment to induce removal of KDM2A/B. SG4 and HeLa cells were used as positive controls, due to their smaller and larger diameter respectively. Error bars show standard deviation of three biological replicates. Asterisks represent significant changes (Student's T-test, *p<0.05, ** p<0.001, ns: not significant.

-log10 p-value	log2 fold change	Gene	Summary of Uniprot description
6.46	-7.65867	Usp5	Cleaves multiubiquitin polymers, involved in unanchored 'Lys-48'-linked polyubiquitin disassembly. Binds linear and 'Lys-63'-linked polyubiquitin with a lower affinity. Knock-down causes increased p53 accumulation and activity as accumulated unanchored polyubiquitin competes with ubiquitinated p53 for proteasomal recognition.
6.35	-6.64011	Uimc1	Central role in the BRCA1-A complex by specifically binding 'Lys-63'-linked ubiquitinated histones H2A and H2AX at DNA lesions sites, targeting the BRCA1-BARD1 heterodimer to sites of DNA damage at DSBs.
6.26	-5.61044	Hax1	Recruits the Arp2/3 complex to the cell cortex and regulates reorganization of the cortical actin cytoskeleton.
5.69	-6.06478	Cct3	Molecular chaperone, assists the folding of proteins upon ATP hydrolysis.
5.60	-5.92587	Smarcd2	Component of SWI/SNF chromatin remodelling complexes. Critical regulator of myeloid differentiation.
5.46	-8.23717	Cul5	Core component of multiple SCF-like ECS (Elongin-Cullin 2/5-SOCS-box protein) E3 ubiquitin ligase complexes, implicated in ubiquitylation of JAK2 and p53.
5.45	-5.52919	Plaa	Plays a role in protein ubiquitination and degradation of ruptured lysosomes through its association with VCP.
5.33	-10.049	Hspa8	Molecular chaperone with a pivotal role in the protein quality control system. Participates in the ER-associated degradation (ERAD) quality control pathway with E3 ligase STUB1.
5.31	-7.64591	Setx	Probable RNA/DNA helicase with a role in transcription regulation. Modulates RNA Polymerase II binding to chromatin, contributes to mRNA splicing efficiency and splice site selection. Required for efficient RNA polymerase II transcription termination, via resolution of R-loop RNA-DNA hybrid formation at G-rich pause sites downstream of the poly(A) site.
5.08	-6.76523	Cops7a	Component of the COP9 signalosome complex which mediates the deneddylation of the cullin subunits of SCF-type E3 ligase complexes, leading to decreased ubiquitin ligase activity. Also involved in phosphorylation of transcription factors including p53 and JUN, which promotes and protects degradation by the ubiquitin ligase system, respectively.
4.75	-5.77043	Cpsf3	Component of the cleavage and polyadenylation specificity factor complex that plays a key role in pre-mRNA 3'-end formation. Functions as mRNA 3'-end-processing endonuclease.
4.70	-5.22314	Samhd1	Host restriction nuclease involved in viral defence. dNTPase activity reduces cellular dNTP level below that required for retroviral reverse transcription.
4.67	-5.95934	Slc28a1	Sodium-dependent and pyrimidine-selective nucleoside transport.
4.67	-5.77384	Eprs	Multifunctional protein which is primarily part of the aminoacyl-tRNA synthetase complex. Suppresses translation of inflammatory mRNAs in response to interferon gamma.
4.54	-5.92823	Cdc5l	Component of the PRP19-CDC5L complex that forms an integral part of the spliceosome and is required for activating pre-mRNA splicing, and may also play a role in the response to DNA damage.
4.28	-6.63303	Mov10	Putative RNA helicase. Required for miRNA-mediated gene silencing by the RNA-induced silencing complex.
4.24	-8.64773	Mtf2	Component of Polycomb Repressive Complex 2.1. Methylation sensitive DNA binding activity required for PRC2 recruitment to target CGIs in mESCs. Facilitates PRC2 recruitment to the inactive X chromosome and PRC2-mediated Hox repression.
4.17	-4.40491	Apip	Functions in the methionine salvage pathway, which plays a key role in cancer, apoptosis, microbial proliferation and inflammation.
4.06	-4.92474	Rbpms	Acts as a coactivator of transcriptional activity. Required to increase TGFB1/Smad-mediated transactivation.
3.93	-5.45111	Armcx1	Enhances mitochondria movement and neurite growth in both adult axons and embryonic neurons.
3.79	-4.82983	Prelid1	Involved in the modulation of the mitochondrial apoptotic pathway by ensuring the accumulation of cardiolipin in mitochondrial membranes.
3.66	-3.91579	Zmym2	May function as a transcription factor.
3.36	-6.30551	Sumo2	Ubiquitin-like protein that can be covalently attached to proteins by an E3 ligase. Regulates processes such as nuclear transport, DNA replication and repair, mitosis and signal transduction. Polymeric SUMO2 chains are susceptible to polyubiquitination, which functions as a signal for proteasomal degradation.
3.34	-4.66413	Pop1	Component of ribonuclease P, a protein complex that generates mature tRNA molecules by cleaving their 5'-ends.
3.27	-4.9404	Lars	Leucine tRNA ligase.
3.09	-1.20041	Rpl23a	Component of the ribosome. May promote p53 degradation, possibly through the stimulation of MDM2-mediated p53 polyubiquitination.
3.09	-4.22189	Mcl1	Involved in the regulation of apoptosis versus cell survival, interacts with a number of other apoptosis regulators.
3.01	-0.69662	Phf10	Involved in transcription activity regulation by chromatin remodeling. Belongs to the neural progenitors-specific chromatin remodeling complex (npBAF complex) and is required for the proliferation of neural progenitors.

Figure A.13: Details of peptides that significantly decrease in ubiquitylation following removal of KDM2A/B

Table showing the gene name, -log10 p-value, log2 fold change, and a condensed Uniprot description for peptides that signficantly decreased in ubiquitylation following loss of KDM2A/B from chromatin (See Figure 6.4). Colouring corresponds to Figure 6.4: peptides involved in ubiquitylation pathways are coloured green, MTF2 is coloured purple, and peptides corresponding to SWI/SNF subunits are coloured red.

Bibliography

Adelman, K. and Lis, J. T. (2012). "Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans". *Nat Rev Genet* 13: pp. 720–31.

Aillet, F., Lopitz-Otsoa, F., Egana, I., et al. (2012). "Heterologous SUMO-2/3-ubiquitin chains optimize IkappaBalpha degradation and NF-kappaB activity". *PLoS One* 7: e51672.

Albert, M. and Helin, K. (2010). "Histone methyltransferases in cancer". *Semin Cell Dev Biol* 21: pp. 209–20.

Alekseyenko, A. A., Peng, S., Larschan, E., et al. (2008). "A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome". *Cell* 134: pp. 599–609.

Allfrey, V. G. and Mirsky, A. E. (1964). "Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis". *Science* 144: p. 559.

Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964). "Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis". *Proc Natl Acad Sci U S A* 51: pp. 786–94.

Allis, C. D. and Jenuwein, T. (2016). "The molecular hallmarks of epigenetic control". *Nat Rev Genet* 17: pp. 487–500.

Almeida, S. F. de, Grosso, A. R., Koch, F., et al. (2011). "Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36". *Nat Struct Mol Biol* 18: pp. 977–83.

Amir, R. E., Veyver, I. B. Van den, Wan, M., et al. (1999). "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2". *Nat Genet* 23: pp. 185–8.

An, S., Yeo, K. J., Jeon, Y. H., et al. (2011). "Crystal Structure of the Human Histone Methyltransferase ASH1L Catalytic Domain and Its Implications for the Regulatory Mechanism". *The Journal of Biological Chemistry* 286: pp. 8369–8374.

Anders, S. and Huber, W. (2010). "Differential expression analysis for sequence count data". *Genome Biol* 11: R106.

Anderson, J. D., Lowary, P. T., and Widom, J. (2001). "Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites". *J Mol Biol* 307: pp. 977–85.

Andricovich, J., Kai, Y., Peng, W., et al. (2016). "Histone demethylase KDM2B regulates lineage commitment in normal and malignant hematopoiesis". *J Clin Invest* 126: pp. 905–20.

Antequera, F. (2003). "Structure, function and evolution of CpG island promoters". *Cell Mol Life Sci* 60: pp. 1647–58.

Antequera, F., Macleod, D., and Bird, A. P. (1989). "Specific protection of methylated CpGs in mammalian nuclei". *Cell* 58: pp. 509–17.

Antequera, F., Boyes, J., and Bird, A. (1990). "High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines". *Cell* 62: pp. 503–14.

Aravind, L., Iyer, L. M., and Koonin, E. V. (2003). "Scores of RINGS but no PHDs in ubiquitin signaling". *Cell Cycle* 2: pp. 123–6.

Arents, G., Burlingame, R. W., Wang, B. C., et al. (1991). "The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix". *Proc Natl Acad Sci U S A* 88: pp. 10148–52.

Audia, J. E. and Campbell, R. M. (2016). "Histone Modifications and Cancer". *Cold Spring Harb Perspect Biol* 8: a019521.

Ayoub, N., Noma, K., Isaac, S., et al. (2003). "A novel jmjC domain protein modulates heterochromatization in fission yeast". *Mol Cell Biol* 23: pp. 4356–70.

Bach, C., Mueller, D., Buhl, S., et al. (2009). "Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2". *Oncogene* 28: pp. 815–23.

Badenhorst, P., Xiao, H., Cherbas, L., et al. (2005). "The Drosophila nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis". *Genes Dev* 19: pp. 2540–5.

Bai, C., Sen, P., Hofmann, K., et al. (1996). "SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box". *Cell* 86: pp. 263–74.

Ballaré, C., Lange, M., Lapinaite, A., et al. (2012). "Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity". *Nature structural & molecular biology* 19: pp. 1257–1265.

Bannister, A. J., Schneider, R., Myers, F. A., et al. (2005). "Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes". *J Biol Chem* 280: pp. 17732–6.

Barski, A., Cuddapah, S., Cui, K., et al. (2007). "High-resolution profiling of histone methylations in the human genome". *Cell* 129: pp. 823–37.

Bartke, T., Vermeulen, M., Xhemalce, B., et al. (2010). "Nucleosome-interacting proteins regulated by DNA and histone methylation". *Cell* 143: pp. 470–84.

Baubec, T., Colombo, D. F., Wirbelauer, C., et al. (2015). "Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation". *Nature* 520: pp. 243–7.

Baylin, S. B., Herman, J. G., Graff, J. R., et al. (1998). "Alterations in DNA methylation: a fundamental aspect of neoplasia". *Adv Cancer Res* 72: pp. 141–96.

Bell, A. C. and Felsenfeld, G. (2000). "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene". *Nature* 405: pp. 482–5.

Bell, O., Wirbelauer, C., Hild, M., et al. (2007). "Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila". *The EMBO Journal* 26: pp. 4974–4984.

Ben-Saadon, R., Zaaroor, D., Ziv, T., et al. (2006). "The polycomb protein Ring1B generates self atypical mixed ubiquitin chains required for its in vitro histone H2A ligase activity". *Mol Cell* 24: pp. 701–711.

Bender, L. B., Suh, J., Carroll, C. R., et al. (2006). "MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the C. elegans germ line". *Development* 133: pp. 3907–17.

Berdasco, M., Ropero, S., Setien, F., et al. (2009). "Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma". *Proc Natl Acad Sci U S A* 106: pp. 21830–5.

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., et al. (2005). "Genomic maps and comparative analysis of histone modifications in human and mouse". *Cell* 120: pp. 169–81.

Bestor, T. H. and Ingram, V. M. (1983). "Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA". *Proc Natl Acad Sci U S A* 80: pp. 5559–63.

Biggar, K. K. and Li, S. S. C. (2014). "Non-histone protein methylation as a regulator of cellular signalling and function". *Nature Reviews Molecular Cell Biology* 16: p. 5.

Bird, A. (2002). "DNA methylation patterns and epigenetic memory". *Genes Dev* 16: pp. 6–21.

Bird, A., Taggart, M., Frommer, M., et al. (1985). "A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA". *Cell* 40: pp. 91–9.

Bird, A. P. and Wolffe, A. P. (1999). "Methylation-induced repression–belts, braces, and chromatin". *Cell* 99: pp. 451–4.

Birke, M., Schreiner, S., Garcia-Cuellar, M. P., et al. (2002). "The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation". *Nucleic Acids Res* 30: pp. 958–65.

Blackledge, N. P. and Klose, R. (2011). "CpG island chromatin: a platform for gene regulation". *Epigenetics* 6: pp. 147–52.

Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., et al. (2010). "CpG islands recruit a histone H3 lysine 36 demethylase". *Mol Cell* 38: pp. 179–90.

Blackledge, N. P., Long, H. K., Zhou, J. C., et al. (2012). "Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA". *Nucleic Acids Res* 40: e32.

Blackledge, N. P., Rose, N. R., and Klose, R. J. (2015). "Targeting Polycomb systems to regulate gene expression: modifications to a complex story". *Nat Rev Mol Cell Biol* 16: pp. 643–649.

Blackledge, N. P., Farcas, A. M., Kondo, T., et al. (2014). "Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation". *Cell* 157: pp. 1445–1459.

Bonhoure, N., Bounova, G., Bernasconi, D., et al. (2014). "Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization". *Genome Res* 24: pp. 1157–68.

Borgel, J., Tyl, M., Schiller, K., et al. (2017). "KDM2A integrates DNA and histone modification signals through a CXXC/PHD module and direct interaction with HP1". *Nucleic Acids Res* 45: pp. 1114–1129.

Bostick, M., Kim, J. K., Esteve, P. O., et al. (2007). "UHRF1 plays a role in maintaining DNA methylation in mammalian cells". *Science* 317: pp. 1760–4.

Bottger, A., Islam, M. S., Chowdhury, R., et al. (2015). "The oxygenase Jmjd6–a case study in conflicting assignments". *Biochem J* 468: pp. 191–202.

Bottomley, M. J., Stier, G., Pennacchini, D., et al. (2005). "NMR structure of the first PHD finger of autoimmune regulator protein (AIRE1). Insights into autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) disease". *J Biol Chem* 280: pp. 11505–12.

Boulard, M., Edwards, J. R., and Bestor, T. H. (2015). "FBXL10 protects Polycomb-bound genes from hypermethylation". *Nat Genet* 47: pp. 479–85.

Brechalov, A. V., Georgieva, S. G., and Soshnikova, N. V. (2014). "Mammalian cells contain two functionally distinct PBAF complexes incorporating different isoforms of PHF10 signature subunit". *Cell Cycle* 13: pp. 1970–9.

Brien, G. L., Gambero, G., O'Connell, D. J., et al. (2012). "Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation". *Nat Struct Mol Biol* 19: pp. 1273–81.

Brookes, E., de Santiago, I., Hebenstreit, D., et al. (2012). "Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs". *Cell Stem Cell* 10: pp. 157–170.

Brown, D. A., Di Cerbo, V., Feldmann, A., et al. (2017). "The SET1 Complex Selects Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin". *Cell Rep* 20: pp. 2313–2327.

Brown, S. J., Stoilov, P., and Xing, Y. (2012). "Chromatin and epigenetic regulation of pre-mRNA processing". *Hum Mol Genet* 21: R90–6.

Buchwald, G., Stoop, P. van der, Weichenrieder, O., et al. (2006). "Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b". *Embo j* 25: pp. 2465–74.

Bueno, M. T. D., Baldascini, M., Richard, S., et al. (2018). "Recruitment of lysine demethylase 2A to DNA double strand breaks and its interaction with 53BP1 ensures genome stability". *Oncotarget* 9: pp. 15915–15930.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., et al. (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position". *Nat Methods* 10: pp. 1213–8.

Buenrostro, J. D., Wu, B., Chang, H. Y., et al. (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide". *Curr Protoc Mol Biol* 109: pp. 21.29.1–9.

Buschhausen, G., Wittig, B., Graessmann, M., et al. (1987). "Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene". *Proc Natl Acad Sci U S A* 84: pp. 1177–81.

Cai, L., Rothbart, S. B., Lu, R., et al. (2013). "An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting". *Mol Cell* 49: pp. 571–82.

Camporeale, G., Shubert, E. E., Sarath, G., et al. (2004). "K8 and K12 are biotinylated in human histone H4". *Eur J Biochem* 271: pp. 2257–63.

Cao, R., Wang, L., Wang, H., et al. (2002). "Role of histone H3 lysine 27 methylation in Polycomb-group silencing". *Science* 298: pp. 1039–43.

Cao, R., Tsukada, Y., and Zhang, Y. (2005). "Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing". *Mol Cell* 20: pp. 845–54.

Capili, A. D., Schultz, D. C., Rauscher, I. F., et al. (2001). "Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains". *Embo j* 20: pp. 165–77.

Cardozo, T. and Pagano, M. (2004). "The SCF ubiquitin ligase: insights into a molecular machine". *Nat Rev Mol Cell Biol* 5: pp. 739–51.

Carrozza, M. J., Li, B., Florens, L., et al. (2005). "Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription". *Cell* 123: pp. 581–92.

Carvalho, S., Raposo, A. C., Martins, F. B., et al. (2013). "Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription". *Nucleic Acids Research* 41: pp. 2881–2893.

Casper, J., Zweig, A. S., Villarreal, C., et al. (2018). "The UCSC Genome Browser database: 2018 update". *Nucleic Acids Res* 46: pp. D762–d769.

Chang, B., Chen, Y., Zhao, Y., et al. (2007). "JMJD6 is a histone arginine demethylase". *Science* 318: pp. 444–7.

Chang, P. Y., Hom, R. A., Musselman, C. A., et al. (2010). "Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription". *J Mol Biol* 400: pp. 137–44.

Chedin, F., Lieber, M. R., and Hsieh, C. L. (2002). "The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a". *Proc Natl Acad Sci U S A* 99: pp. 16916–21.

Chen, J. and Archer, T. K. (2005). "Regulating SWI/SNF subunit levels via protein-protein interactions and proteasomal degradation: BAF155 and BAF170 limit expression of BAF57". *Mol Cell Biol* 25: pp. 9016–27.

Chen, K., Xi, Y., Pan, X., et al. (2013). "DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing". *Genome Research* 23: pp. 341–351.

Chen, Z., Zang, J., Whetstine, J., et al. (2006). "Structural insights into histone demethylation by JMJD2 family members". *Cell* 125: pp. 691–702.

Cheng, Z., Cheung, P., Kuo, A. J., et al. (2014). "A molecular threading mechanism underlies Jumonji lysine demethylase KDM2A regulation of methylated H3K36". *Genes & Development* 28: pp. 1758–1771.

Chi, P., Allis, C. D., and Wang, G. G. (2010). "Covalent histone modifications–miswritten, misinterpreted and mis-erased in human cancers". *Nat Rev Cancer* 10: pp. 457–69.

Choi, J. K. (2010). "Contrasting chromatin organization of CpG islands and exons in the human genome". *Genome Biol* 11: R70.

Chu, Y., Simic, R., Warner, M. H., et al. (2007). "Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes". *Embo j* 26: pp. 4646–56.

Chuang, L. S., Ian, H. I., Koh, T. W., et al. (1997). "Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1". *Science* 277: pp. 1996–2000.

Cierpicki, T., Risner, L. E., Grembecka, J., et al. (2010). "Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia". *Nat Struct Mol Biol* 17: pp. 62–8.

Cloos, P. A., Christensen, J., Agger, K., et al. (2006). "The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3". *Nature* 442: pp. 307–11.

Clouaire, T., Webb, S., Skene, P., et al. (2012). "Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells". *Genes Dev* 26: pp. 1714–28.

Clouaire, T., Webb, S., and Bird, A. (2014). "Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells". *Genome Biol* 15: p. 451.

Cockerill, P. N. (2011). "Structure and function of active chromatin and DNase I hypersensitive sites". *Febs j* 278: pp. 2182–210.

Comb, M. and Goodman, H. M. (1990). "CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2". *Nucleic Acids Res* 18: pp. 3975–82.

Connelly, C. and Hieter, P. (1996). "Budding yeast SKP1 encodes an evolutionarily conserved kinetochore protein required for cell cycle progression". *Cell* 86: pp. 275–85.

Cooper, D. N., Taggart, M. H., and Bird, A. P. (1983). "Unmethylated domains in vertebrate DNA". *Nucleic Acids Res* 11: pp. 647–58.

Cooper, S., Dienstbier, M., Hassan, R., et al. (2014). "Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment". *Cell Rep* 7: pp. 1456–1470.

Cope, G. A. and Deshaies, R. J. (2003). "COP9 signalosome: a multifunctional regulator of SCF and other cullin-based ubiquitin ligases". *Cell* 114: pp. 663–71.

Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters". *Science* 322: pp. 1845–8.

Coscoy, L., Sanchez, D. J., and Ganem, D. (2001). "A novel class of herpesvirus-encoded membrane-bound E3 ubiquitin ligases regulates endocytosis of proteins involved in immune recognition". *J Cell Biol* 155: pp. 1265–73.

Costello, J. F., Fruhwald, M. C., Smiraglia, D. J., et al. (2000). "Aberrant CpG-island methylation has non-random and tumour-type-specific patterns". *Nat Genet* 24: pp. 132–8.

Coulondre, C., Miller, J. H., Farabaugh, P. J., et al. (1978). "Molecular basis of base substitution hotspots in Escherichia coli". *Nature* 274: pp. 775–80.

Cross, S. H., Meehan, R. R., Nan, X., et al. (1997). "A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins". *Nat Genet* 16: pp. 256–9.

Czermin, B., Melfi, R., McCabe, D., et al. (2002). "Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites". *Cell* 111: pp. 185–96.

Das, P. P., Shao, Z., Beyaz, S., et al. (2014). "Distinct and combinatorial functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in mouse embryonic stem cells identity". *Molecular cell* 53: pp. 32–48.

Deaton, A. M. and Bird, A. (2011). "CpG islands and the regulation of transcription". *Genes Dev* 25: pp. 1010–22.

Debril, M. B., Gelman, L., Fayard, E., et al. (2004). "Transcription factors and nuclear receptors interact with the SWI/SNF complex through the BAF60c subunit". *J Biol Chem* 279: pp. 16677–86.

Del Rizzo, P. A., Krishnan, S., and Trievel, R. C. (2012). "Crystal structure and functional analysis of JMJD5 indicate an alternate specificity and function". *Mol Cell Biol* 32: pp. 4044–52.

Denissov, S., Hofemeister, H., Marks, H., et al. (2014). "Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant". *Development* 141: pp. 526–37.

Dhar, S. S., Alam, H., Li, N., et al. (2014). "Transcriptional Repression of Histone Deacetylase 3 by the Histone Demethylase KDM2A Is Coupled to Tumorigenicity of Lung Cancer Cells". *The Journal of Biological Chemistry* 289: pp. 7483–7496.

Dhayalan, A., Rajavelu, A., Rathert, P., et al. (2010). "The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation". *Journal of Biological Chemistry* 285: pp. 26114–26120.

Di Croce, L. and Helin, K. (2013). "Transcriptional regulation by Polycomb group proteins". *Nat Struct Mol Biol* 20: pp. 1147–55.

Dieuleveult, M. de, Yen, K., Hmitou, I., et al. (2016). "Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells". *Nature* 530: pp. 113–6.

Dillon, S. C., Zhang, X., Trievel, R. C., et al. (2005). "The SET-domain protein superfamily: protein lysine methyltransferases". *Genome Biol* 6: p. 227.

Dimitrova, E., Kondo, T., Feldmann, A., et al. (2018). "FBXL19 recruits CDK-Mediator to CpG islands of developmental genes priming them for activation during lineage commitment". *Elife* 7.

Dimitrova, E., Turberfield, A. H., and Klose, R. J. (2015). "Histone demethylases in chromatin biology and beyond". *EMBO Reports* 16: pp. 1620–1639.

Dobin, A., Davis, C. A., Schlesinger, F., et al. (2013). "STAR: ultrafast universal RNA-seq aligner". *Bioinformatics* 29: pp. 15–21.

Dong, R., Yao, R., Du, J., et al. (2013). "Depletion of histone demethylase KDM2A enhanced the adipogenic and chondrogenic differentiation potentials of stem cells from apical papilla". *Exp Cell Res* 319: pp. 2874–82.

Dong, S., Zhao, J., Wei, J., et al. (2014). "F-box protein complex FBXL19 regulates TGFbeta1-induced E-cadherin down-regulation by mediating Rac3 ubiquitination and degradation". *Mol Cancer* 13: p. 76.

Dorighi, K. M. and Tamkun, J. W. (2013). "The trithorax group proteins Kismet and ASH1 promote H3K36 dimethylation to counteract Polycomb group repression in Drosophila". *Development* 140: pp. 4182–92.

Drouin, S., Laramee, L., Jacques, P. E., et al. (2010). "DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes". *PLoS Genet* 6: e1001173.

Du, H. N., Fingerman, I. M., and Briggs, S. D. (2008). "Histone H3 K36 methylation is mediated by a trans-histone methylation pathway involving an interaction between Set2 and histone H4". *Genes Dev* 22: pp. 2786–98.

Du, J., Ma, Y., Ma, P., et al. (2013). "Demethylation of epiregulin gene by histone demethylase FBXL11 and BCL6 corepressor inhibits osteo/dentinogenic differentiation". *Stem Cells* 31: pp. 126–36.

Dul, B. E. and Walworth, N. C. (2007). "The plant homeodomain fingers of fission yeast Msc1 exhibit E3 ubiquitin ligase activity". *J Biol Chem* 282: pp. 18397–406.

Duncan, B. K. and Miller, J. H. (1980). "Mutagenic deamination of cytosine residues in DNA". *Nature* 287: pp. 560–1.

Eberl, H. C., Spruijt, C. G., Kelstrup, C. D., et al. (2013). "A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics". *Mol Cell* 49: pp. 368–78.

Edmunds, J. W., Mahadevan, L. C., and Clayton, A. L. (2008). "Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation". *Embo j* 27: pp. 406–20.

ENCODE (2012). "An integrated encyclopedia of DNA elements in the human genome". *Nature* 489: pp. 57–74.

External, R. N.A.C. C. (2005). "Proposed methods for testing and selecting the ERCC external RNA controls". *BMC Genomics* 6: p. 150.

Fang, D., Gan, H., Lee, J.-H., et al. (2016). "The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas". *Science (New York, N.Y.)* 352: pp. 1344–1348.

Fang, J., Hogan, G. J., Liang, G., et al. (2007). "The Saccharomyces cerevisiae Histone Demethylase Jhd1 Fine-Tunes the Distribution of H3K36me2". *Molecular and Cellular Biology* 27: pp. 5055–5065.

Fang, R., Barbera, A. J., Xu, Y., et al. (2010). "Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation". *Mol Cell* 39: pp. 222–33.

Fanti, L. and Pimpinelli, S. (2008). "HP1: a functionally multifaceted protein". *Curr Opin Genet Dev* 18: pp. 169–74.

Farcas, A. M., Blackledge, N. P., Sudbery, I., et al. (2012). "KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands". *Elife* 1: e00205.

Feng, Q., Wang, H., Ng, H. H., et al. (2002). "Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain". *Curr Biol* 12: pp. 1052–8.

Fenouil, R., Cauchy, P., Koch, F., et al. (2012). "CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters". *Genome Res* 22: pp. 2399–408.

Ferrari, K. J., Scelfo, A., Jammula, S., et al. (2014). "Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity". *Mol Cell* 53: pp. 49–62.

Flajollet, S., Lefebvre, B., Cudejko, C., et al. (2007). "The core component of the mammalian SWI/SNF complex SMARCD3/BAF60c is a coactivator for the nuclear retinoic acid receptor". *Mol Cell Endocrinol* 270: pp. 23–32.

Fodor, B. D., Kubicek, S., Yonezawa, M., et al. (2006). "Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells". *Genes & Development* 20: pp. 1557–1562.

Forrest, A. R., Kawaji, H., Rehli, M., et al. (2014). "A promoter-level mammalian expression atlas". *Nature* 507: pp. 462–70.

Frescas, D., Guardavaccaro, D., Bassermann, F., et al. (2007). "JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes". *Nature* 450: pp. 309–13.

Frescas, D., Guardavaccaro, D., Kuchay, S. M., et al. (2008). "KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state". *Cell Cycle* 7: pp. 3539–47.

Fukuda, T., Tokunaga, A., Sakamoto, R., et al. (2011). "Fbxl10/Kdm2b deficiency accelerates neural progenitor cell death and leads to exencephaly". *Mol Cell Neurosci* 46: pp. 614–24.

Furuhashi, H., Takasaki, T., Rechtsteiner, A., et al. (2010). "Trans-generational epigenetic regulation of C. elegans primordial germ cells". *Epigenetics & Chromatin* 3: pp. 15–15.

Gao, R., Dong, R., Du, J., et al. (2013). "Depletion of histone demethylase KDM2A inhibited cell proliferation of stem cells from apical papilla by de-repression of p15INK4B and p27Kip1". *Molecular and Cellular Biochemistry* 379: pp. 115–122.

Gao, Z., Zhang, J., Bonasio, R., et al. (2012). "PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes". *Mol Cell* 45: pp. 344–56.

Garcia-Ramirez, M., Rocchini, C., and Ausio, J. (1995). "Modulation of chromatin folding by histone acetylation". *J Biol Chem* 270: pp. 17923–8.

Gardiner-Garden, M. and Frommer, M. (1987). "CpG islands in vertebrate genomes". *J Mol Biol* 196: pp. 261–82.

Garneau, J. E., Dupuis, M. E., Villion, M., et al. (2010). "The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA". *Nature* 468: pp. 67–71.

Gasiunas, G., Barrangou, R., Horvath, P., et al. (2012). "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria". *Proc Natl Acad Sci U S A* 109: E2579–86.

Gaydos, L. J., Rechtsteiner, A., Egelhofer, T. A., et al. (2012). "Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells". *Cell Rep* 2: pp. 1169–77.

Gearhart, M. D., Corcoran, C. M., Wamstad, J. A., et al. (2006). "Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets". *Mol Cell Biol* 26: pp. 6880–9.

Goljanek-Whysall, K., Mok, G. F., Fahad Alrefaei, A., et al. (2014). "myomiR-dependent switching of BAF60 variant incorporation into Brg1 chromatin remodeling complexes during embryo myogenesis". *Development* 141: pp. 3378–87.

Govind, C. K., Qiu, H., Ginsburg, D. S., et al. (2010). "Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes". *Mol Cell* 39: pp. 234–46.

Gowher, H., Liebert, K., Hermann, A., et al. (2005). "Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L". *J Biol Chem* 280: pp. 13341–8.

Greer, E. L. and Shi, Y. (2012). "Histone methylation: a dynamic mark in health, disease and inheritance". *Nat Rev Genet* 13: pp. 343–57.

Gregory, G. D., Vakoc, C. R., Rozovskaia, T., et al. (2007). "Mammalian ASH1L Is a Histone Methyltransferase That Occupies the Transcribed Region of Active Genes". *Molecular and Cellular Biology* 27: pp. 8466–8479.

Guenther, M. G., Jenner, R. G., Chevalier, B., et al. (2005). "Global and Hox-specific roles for the MLL1 methyltransferase". *Proc Natl Acad Sci U S A* 102: pp. 8603–8.

Guenther, M. G., Levine, S. S., Boyer, L. A., et al. (2007). "A chromatin landmark and transcription initiation at most promoters in human cells". *Cell* 130: pp. 77–88.

Guzzo, C. M. and Matunis, M. J. (2013). "Expanding SUMO and ubiquitin-mediated signaling through hybrid SUMO-ubiquitin chains and their receptors". *Cell Cycle* 12: pp. 1015–7.

Guzzo, C. M., Berndsen, C. E., Zhu, J., et al. (2012). "RNF4-dependent hybrid SUMO-ubiquitin chains are signals for RAP80 and thereby mediate the recruitment of BRCA1 to sites of DNA damage". *Sci Signal* 5: ra88.

Hajkova, P. (2011). "Epigenetic reprogramming in the germline: towards the ground state of the epigenome". *Philos Trans R Soc Lond B Biol Sci* 366: pp. 2266–73.

Han, X. R., Zha, Z., Yuan, H. X., et al. (2016). "KDM2B/FBXL10 targets c-Fos for ubiquitylation and degradation in response to mitogenic stimulation". *Oncogene* 35: pp. 4179–90.

Harris, M. (1982). "Induction of thymidine kinase in enzyme-deficient Chinese hamster cells". *Cell* 29: pp. 483–92.

Hashimoto, H., Horton, J. R., Zhang, X., et al. (2008). "The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix". *Nature* 455: pp. 826–9.

Hay, R. T. (2005). "SUMO: a history of modification". *Mol Cell* 18: pp. 1–12.

He, J., Kallin, E. M., Tsukada, Y., et al. (2008). "The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b)". *Nat Struct Mol Biol* 15: pp. 1169–75.

He, J., Nguyen, A. T., and Zhang, Y. (2011). "KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia". *Blood* 117: pp. 3869–80.

He, J., Shen, L., Wan, M., et al. (2013). "Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes". *Nat Cell Biol* 15: pp. 373–84.

Hebbes, T. R., Thorne, A. W., and Crane-Robinson, C. (1988). "A direct link between core histone acetylation and transcriptionally active chromatin". *Embo j* 7: pp. 1395–402.

Hendrich, B. and Bird, A. (1998). "Identification and characterization of a family of mammalian methyl-CpG binding proteins". *Mol Cell Biol* 18: pp. 6538–47.

Hewitt, E. W., Duncan, L., Mufti, D., et al. (2002). "Ubiquitylation of MHC class I by the K3 viral protein signals internalization and TSG101-dependent degradation". *Embo j* 21: pp. 2418–29.

Hindson, B. J., Ness, K. D., Masquelier, D. A., et al. (2011). "High-throughput droplet digital PCR system for absolute quantitation of DNA copy number". *Anal Chem* 83: pp. 8604–10.

Hindson, C. M., Chevillet, J. R., Briggs, H. A., et al. (2013). "Absolute quantification by droplet digital PCR versus analog real-time PCR". *Nature Methods* 10: p. 1003.

Ho, L., Jothi, R., Ronan, J. L., et al. (2009). "An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network". *Proc Natl Acad Sci U S A* 106: pp. 5187–91.

Ho, M. S., Tsai, P. I., and Chien, C. T. (2006). "F-box proteins: the key to protein degradation". *J Biomed Sci* 13: pp. 181–91.

Holliday, R. (1987). "The inheritance of epigenetic defects". *Science* 238: pp. 163–70.

Holliday, R. and Pugh, J. E. (1975). "DNA modification mechanisms and gene activity during development". *Science* 187: pp. 226–32.

Hopkinson, R. J., Walport, L. J., Munzel, M., et al. (2013). "Is JmjC oxygenase catalysis limited to demethylation?" *Angew Chem Int Ed Engl* 52: pp. 7709–13.

Hosogane, M., Funayama, R., Nishida, Y., et al. (2013). "Ras-induced changes in H3K27me3 occur after those in transcriptional activity". *PLoS Genet* 9: e1003698.

Hsiao, P. W., Fryer, C. J., Trotter, K. W., et al. (2003). "BAF60a mediates critical interactions between nuclear receptors and the BRG1 chromatin-remodeling complex for transactivation". *Mol Cell Biol* 23: pp. 6210–20.

Hu, B., Petela, N., Kurze, A., et al. (2015). "Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq". *Nucleic Acids Res* 43: e132.

Hu, M., Sun, X.-J., Zhang, Y.-L., et al. (2010). "Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling". *Proceedings of the National Academy of Sciences* 107: pp. 2956–2961.

Huang, N., Baur, E. vom, Garnier, J. M., et al. (1998). "Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators". *The EMBO Journal* 17: pp. 3398–3412.

Huang, Y., Fang, J., Bedford, M. T., et al. (2006). "Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A". *Science* 312: pp. 748–51.

Huang, Y., Liu, Y., Yu, L., et al. (2015). "Histone demethylase KDM2A promotes tumor cell growth and migration in gastric cancer". *Tumour Biol* 36: pp. 271–8.

Huang, Z., Wu, H., Chuai, S., et al. (2013). "NSD2 is recruited through its PHD domain to oncogenic gene loci to drive multiple myeloma". *Cancer Res* 73: pp. 6277–88.

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009a). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists". *Nucleic Acids Res* 37: pp. 1–13.

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009b). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources". *Nat Protoc* 4: pp. 44–57.

Huh, J. W., Wu, J., Lee, C. H., et al. (2012). "Multivalent di-nucleosome recognition enables the Rpd3S histone deacetylase complex to tolerate decreased H3K36 methylation levels". *Embo j* 31: pp. 3564–74.

Iguchi-Ariga, S. M. and Schaffner, W. (1989). "CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation". *Genes Dev* 3: pp. 612–9.

Illingworth, R. S. and Bird, A. P. (2009). "CpG islands–'a rough guide'". *FEBS Lett* 583: pp. 1713–20.

Illingworth, R., Kerr, A., DeSousa, D., et al. (2008). "A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci". *PLoS Biology* 6: e22.

Inagaki, T., Iwasaki, S., Matsumura, Y., et al. (2015). "The FBXL10/KDM2B Scaffolding Protein Associates with Novel Polycomb Repressive Complex-1 to Regulate Adipogenesis". *The Journal of Biological Chemistry* 290: pp. 4163–4177.

Ipenberg, I., Guttmann-Raviv, N., Khoury, H. P., et al. (2013). "Heat shock protein 90 (Hsp90) selectively regulates the stability of KDM4B/JMJD2B histone demethylase". *J Biol Chem* 288: pp. 14681–7.

Ito, T., Yamauchi, M., Nishina, M., et al. (2001). "Identification of SWI.SNF complex subunit BAF60a as a determinant of the transactivation potential of Fos/Jun dimers". *J Biol Chem* 276: pp. 2852–7.

Iwamori, N., Zhao, M., Meistrich, M. L., et al. (2011). "The testis-enriched histone demethylase, KDM4D, regulates methylation of histone H3 lysine 9 during spermatogenesis in the mouse but is dispensable for fertility". *Biol Reprod* 84: pp. 1225–34.

Jaenisch, R., Schnieke, A., and Harbers, K. (1985). "Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues". *Proc Natl Acad Sci U S A* 82: pp. 1451–5.

Jaju, R. J., Fidler, C., Haas, O. A., et al. (2001). "A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia". *Blood* 98: pp. 1264–7.

Janzer, A., Stamm, K., Becker, A., et al. (2012). "The H3K4me3 histone demethylase Fbxl10 is a regulator of chemokine expression, cellular morphology, and the metabolome of fibroblasts". *J Biol Chem* 287: pp. 30984–92.

Jeltsch, A. (2006). "On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme". *Epigenetics* 1: pp. 63–6.

Jin, B., Ernst, J., Tiedemann, R. L., et al. (2012). "Linking DNA Methyltransferases (DNMTs) to Epigenetic Marks and Nucleosome Structure Genome-Wide in Human Tumor Cells". *Cell reports* 2: pp. 1411–1424.

Jin, J., Cardozo, T., Lovering, R. C., et al. (2004). "Systematic analysis and nomenclature of mammalian F-box proteins". *Genes Dev* 18: pp. 2573–80.

Jinek, M., Chylinski, K., Fonfara, I., et al. (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity". *Science* 337: pp. 816–21.

Johmura, Y., Sun, J., Kitagawa, K., et al. (2016). "SCFFbxo22-KDM4A targets methylated p53 for degradation and regulates senescence". *Nature Communications* 7: p. 10574.

Johnson, E. S. (2004). "Protein modification by SUMO". *Annu Rev Biochem* 73: pp. 355–82.

Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond". *Nat Rev Genet* 13: pp. 484–92.

Jones, P. L., Veenstra, G. J., Wade, P. A., et al. (1998). "Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription". *Nat Genet* 19: pp. 187–91.

Jones, P. A. and Baylin, S. B. (2007). "The Epigenomics of Cancer". *Cell* 128: pp. 683–692.

Jonkers, I. and Lis, J. T. (2015). "Getting up to speed with transcription elongation by RNA polymerase II". *Nat Rev Mol Cell Biol* 16: pp. 167–77.

Joshi, A. A. and Struhl, K. (2005). "Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation". *Mol Cell* 20: pp. 971–8.

Josse, J., Kaiser, A. D., and Kornberg, A. (1961). "Enzymatic synthesis of deoxyribonucleic acid. VIII. Frequencies of nearest neighbor base sequences in deoxyribonucleic acid". *J Biol Chem* 236: pp. 864–75.

Jung, H. R., Pasini, D., Helin, K., et al. (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". *Mol Cell Proteomics* 9: pp. 838–50.

Kalb, R., Latwiel, S., Baymaz, H. I., et al. (2014). "Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression". *Nat Struct Mol Biol* 21: pp. 569–71.

Karmodiya, K., Krebs, A. R., Oulad-Abdelghani, M., et al. (2012). "H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells". *BMC Genomics* 13: p. 424.

Kass, S. U., Goddard, J. P., and Adams, R. L. (1993). "Inactive chromatin spreads from a focus of methylation". *Mol Cell Biol* 13: pp. 7372–9.

Katoh, M. and Katoh, M. (2004). "Identification and characterization of FBXL19 gene in silico". *Int J Mol Med* 14: pp. 1109–14.

Kawakami, E., Tokunaga, A., Ozawa, M., et al. (2015). "The histone demethylase Fbxl11/Kdm2a plays an essential role in embryonic development by repressing cell-cycle regulators". *Mech Dev* 135: pp. 31–42.

Kawazu, M., Saso, K., Tong, K. I., et al. (2011). "Histone Demethylase JMJD2B Functions as a Co-Factor of Estrogen Receptor in Breast Cancer Proliferation and Mammary Gland Development". *PLoS ONE* 6: e17830.

Kempe, H., Schwabe, A., Crémazy, F., et al. (2015). "The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise". *Molecular Biology of the Cell* 26: pp. 797–804.

Kent, W. J., Sugnet, C. W., Furey, T. S., et al. (2002). "The human genome browser at UCSC". *Genome Res* 12: pp. 996–1006.

Keogh, M. C., Kurdistani, S. K., Morris, S. A., et al. (2005). "Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex". *Cell* 123: pp. 593–605.

Kidder, B. L., Palmer, S., and Knott, J. G. (2009). "SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells". *Stem Cells* 27: pp. 317–28.

Kim, D. H., Tang, Z., Shimada, M., et al. (2013). "Histone H3K27 trimethylation inhibits H3 binding and function of SET1-like H3K4 methyltransferase complexes". *Mol Cell Biol* 33: pp. 4936–46.

Kim, J.-Y., Kee, H. J., Choe, N.-W., et al. (2008). "Multiple Myeloma-Related WHSC1/MMSET Isoform RE-IIBP Is a Histone Methyltransferase with Transcriptional Repression Activity". *Molecular and Cellular Biology* 28: pp. 2023–2034.

Kim, S., Kim, H., Fong, N., et al. (2011). "Pre-mRNA splicing is a determinant of histone H3K36 methylation". *Proc Natl Acad Sci U S A* 108: pp. 13564–9.

Kim, S. M., Kee, H. J., Eom, G. H., et al. (2006). "Characterization of a novel WHSC1-associated SET domain protein with H3K4 and H3K27 methyltransferase activity". *Biochem Biophys Res Commun* 345: pp. 318–23.

Kim, T. and Buratowski, S. (2007). "Two Saccharomyces cerevisiae JmjC domain proteins demethylate histone H3 Lys36 in transcribed regions to promote elongation". *J Biol Chem* 282: pp. 20827–35.

Kim, T. D., Shin, S., and Janknecht, R. (2016). "ETS transcription factor ERG cooperates with histone demethylase KDM4A". *Oncol Rep* 35: pp. 3679–88.

King, H. W., Fursova, N. A., Blackledge, N. P., et al. (2018). "Polycomb repressive complex 1 shapes the nucleosome landscape but not accessibility at target genes". *Genome Res*.

Kizer, K. O., Phatnani, H. P., Shibata, Y., et al. (2005). "A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation". *Mol Cell Biol* 25: pp. 3305–16.

Klose, R. J., Yamane, K., Bae, Y., et al. (2006). "The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36". *Nature* 442: pp. 312–6.

Klose, R. J., Cooper, S., Farcas, A. M., et al. (2013). "Chromatin sampling–an emerging perspective on targeting polycomb repressor proteins". *PLoS Genet* 9: e1003717.

Klose, R. J., Gardner, K. E., Liang, G., et al. (2007). "Demethylation of Histone H3K36 and H3K9 by Rph1: a Vestige of an H3K9 Methylation System in Saccharomyces cerevisiae?" *Molecular and Cellular Biology* 27: pp. 3951–3961.

Knezetic, J. A. and Luse, D. S. (1986). "The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro". *Cell* 45: pp. 95–104.

Kodzius, R., Kojima, M., Nishiyori, H., et al. (2006). "CAGE: cap analysis of gene expression". *Nature Methods* 3: p. 211.

Komander, D. (2009). "The emerging complexity of protein ubiquitination". *Biochem Soc Trans* 37: pp. 937–53.

Komarnitsky, P., Cho, E. J., and Buratowski, S. (2000). "Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription". *Genes Dev* 14: pp. 2452–60.

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

Kouzarides, T. (2007). "Chromatin modifications and their function". *Cell* 128: pp. 693–705.

Koyama-Nasu, R., David, G., and Tanese, N. (2007). "The F-box protein Fbl10 is a novel transcriptional repressor of c-Jun". *Nat Cell Biol* 9: pp. 1074–80.

Kristjuhan, A. and Svejstrup, J. Q. (2004). "Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo". *Embo j* 23: pp. 4243–52.

Krogan, N. J., Kim, M., Tong, A., et al. (2003). "Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II". *Mol Cell Biol* 23: pp. 4207–18.

Ku, M., Koche, R. P., Rheinbay, E., et al. (2008). "Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains". *PLoS Genet* 4: e1000242.

Kuo, A. J., Cheung, P., Chen, K., et al. (2011). "NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming". *Molecular cell* 44: pp. 609–620.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., et al. (2002). "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein". *Genes Dev* 16: pp. 2893–905.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y. M., et al. (2008). "dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing". *Genes Dev* 22: pp. 2799–810.

Lainović, D., Novotná, J., Jakšová, S., et al. (2017). "A demethylation deficient isoform of the lysine demethylase KDM2A interacts with pericentromeric heterochromatin in an HP1a-dependent manner". *Nucleus* 8: pp. 563–572.

Landry, J., Sutton, A., Hesman, T., et al. (2003). "Set2-Catalyzed Methylation of Histone H3 Represses Basal Expression of GAL4 in Saccharomyces cerevisiae". *Molecular and Cellular Biology* 23: pp. 5972–5978.

Langmead, B. and Salzberg, S. L. (2012). "Fast gapped-read alignment with Bowtie 2". *Nat Methods* 9: pp. 357–9.

Larschan, E., Alekseyenko, A. A., Gortchakov, A. A., et al. (2007). "MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism". *Mol Cell* 28: pp. 121–33.

Lauberth, S. M., Nakayama, T., Wu, X., et al. (2013). "H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation". *Cell* 152: pp. 1021–36.

Lee, B. K., Lee, J., Shen, W., et al. (2017). "Fbxl19 recruitment to CpG islands is required for Rnf20-mediated H2B mono-ubiquitination". *Nucleic Acids Res* 45: pp. 7151–7166.

Lee, C. H., Wu, J., and Li, B. (2013). "Chromatin remodelers fine-tune H3K36me-directed deacetylation of neighbor nucleosomes by Rpd3S". *Mol Cell* 52: pp. 255–63.

Lee, J. H. and Skalnik, D. G. (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". *J Biol Chem* 280: pp. 41725–31.

Lee, J. H., Voo, K. S., and Skalnik, D. G. (2001). "Identification and characterization of the DNA binding domain of CpG-binding protein". *J Biol Chem* 276: pp. 44669–76.

Lee, J. H., Tate, C. M., You, J. S., et al. (2007). "Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex". *J Biol Chem* 282: pp. 13419–28.

Lee, J., Thompson, J. R., Botuyan, M. V., et al. (2008). "Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor". *Nature structural & molecular biology* 15: pp. 109–111.

Leeb, M., Pasini, D., Novatchkova, M., et al. (2010). "Polycomb complexes act redundantly to repress genomic repeats and genes". *Genes Dev* 24: pp. 265–76.

Lehnertz, B., Ueda, Y., Derijck, A. A., et al. (2003). "Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin". *Curr Biol* 13: pp. 1192–200.

Lei, H., Oh, S. P., Okano, M., et al. (1996). "De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells". *Development* 122: pp. 3195–205.

Lessard, J., Wu, J. I., Ranish, J. A., et al. (2007). "An essential switch in subunit composition of a chromatin remodeling complex during neural development". *Neuron* 55: pp. 201–15.

Lewis, E. B. (1978). "A gene complex controlling segmentation in Drosophila". *Nature* 276: pp. 565–70.

Lewis, J. D., Meehan, R. R., Henzel, W. J., et al. (1992). "Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA". *Cell* 69: pp. 905–14.

Li, B., Howe, L., Anderson, S., et al. (2003). "The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II". *J Biol Chem* 278: pp. 8897–903.

Li, B., Carey, M., and Workman, J. L. (2007). "The role of chromatin during transcription". *Cell* 128: pp. 707–19.

Li, E., Bestor, T. H., and Jaenisch, R. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality". *Cell* 69: pp. 915–26.

Li, H., Ilin, S., Wang, W., et al. (2006). "Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF". *Nature* 442: pp. 91–5.

Li, J., Moazed, D., and Gygi, S. P. (2002). "Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation". *J Biol Chem* 277: pp. 49383–8.

Li, X., Moon, G., Shin, S., et al. (2018). "Cooperation between ETS variant 2 and Jumonji domaincontaining 2 histone demethylases". *Mol Med Rep* 17: pp. 5518–5527.

Li, Y., Trojer, P., Xu, C. F., et al. (2009). "The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate". *J Biol Chem* 284: pp. 34283–95.

Li, Z., Schug, J., Tuteja, G., et al. (2011). "The nucleosome map of the mammalian liver". *Nat Struct Mol Biol* 18: pp. 742–6.

Liang, G., He, J., and Zhang, Y. (2012). "Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming". *Nat Cell Biol* 14: pp. 457–66.

Lickert, H., Takeuchi, J. K., Von Both, I., et al. (2004). "Baf60c is essential for function of BAF chromatin remodelling complexes in heart development". *Nature* 432: pp. 107–12.

Lin, L. J., Minard, L. V., Johnston, G. C., et al. (2010). "Asf1 can promote trimethylation of H3 K36 by Set2". *Mol Cell Biol* 30: pp. 1116–29.

Lister, R., Pelizzola, M., Dowen, R. H., et al. (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences". *Nature* 462: pp. 315–22.

Liu, H., Liu, L., Holowatyj, A., et al. (2016). "Integrated genomic and functional analyses of histone demethylases identify oncogenic KDM2A isoform in breast cancer". *Mol Carcinog* 55: pp. 977–90.

Liu, W., Ma, Q., Wong, K., et al. (2013). "Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release". *Cell* 155: pp. 1581–1595.

Long, H. K., Blackledge, N. P., and Klose, R. J. (2013). "ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection". *Biochemical Society transactions* 41: pp. 727–740.

Lorch, Y., LaPointe, J. W., and Kornberg, R. D. (1987). "Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones". *Cell* 49: pp. 203–10.

Lores, P., Visvikis, O., Luna, R., et al. (2010). "The SWI/SNF protein BAF60b is ubiquitinated through a signalling process involving Rac GTPase and the RING finger protein Unkempt". *Febs j* 277: pp. 1453–64.

Love, M. I., Huber, W., and Anders, S. (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome Biol* 15: p. 550.

Lu, C., Jain, S. U., Hoelper, D., et al. (2016). "Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape". *Science* 352: pp. 844–9.

Lu, L., Gao, Y., Zhang, Z., et al. (2015). "Kdm2a/b Lysine Demethylases Regulate Canonical Wnt Signaling by Modulating the Stability of Nuclear beta-Catenin". *Dev Cell* 33: pp. 660–74.

Lu, T., Jackson, M. W., Wang, B., et al. (2010). "Regulation of NF-B by NSD1/FBXL11-dependent reversible lysine methylation of p65". *Proceedings of the National Academy of Sciences of the United States of America* 107: pp. 46–51.

Lu, Z., Xu, S., Joazeiro, C., et al. (2002). "The PHD domain of MEKK1 acts as an E3 ubiquitin ligase and mediates ubiquitination and degradation of ERK1/2". *Mol Cell* 9: pp. 945–56.

Lucio-Eterovic, A. K., Singh, M. M., Gardner, J. E., et al. (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". *Proc Natl Acad Sci U S A* 107: pp. 16952–7.

Luco, R. F., Pan, Q., Tominaga, K., et al. (2010). "Regulation of Alternative Splicing by Histone Modifications". *Science (New York, N.Y.)* 327: pp. 996–1000.

Luger, K., Mader, A. W., Richmond, R. K., et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution". *Nature* 389: pp. 251–60.

Luo, W., Chang, R., Zhong, J., et al. (2012). "Histone demethylase JMJD2C is a coactivator for hypoxia-inducible factor 1 that is required for breast cancer progression". *Proc Natl Acad Sci U S A* 109: E3367–76.

Lyapina, S., Cope, G., Shevchenko, A., et al. (2001). "Promotion of NEDD-CUL1 conjugate cleavage by COP9 signalosome". *Science* 292: pp. 1382–5.

Lynch, M. D., Smith, A. J., De Gobbi, M., et al. (2012). "An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment". *Embo j* 31: pp. 317–29.

Lyst, M. J., Ekiert, R., Ebert, D. H., et al. (2013). "Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor". *Nature neuroscience* 16: 10.1038/nn.3434.

Mallette, F. A., Mattiroli, F., Cui, G., et al. (2012). "RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites". *Embo j* 31: pp. 1865–78.

Maltby, V. E., Martin, B. J. E., Schulze, J. M., et al. (2012). "Histone H3 Lysine 36 Methylation Targets the Isw1b Remodeling Complex to Chromatin". *Molecular and Cellular Biology* 32: pp. 3479–3485.

Mansouri, M., Bartee, E., Gouveia, K., et al. (2003). "The PHD/LAP-domain protein M153R of myxomavirus is a ubiquitin ligase that induces the rapid internalization and lysosomal destruction of CD4". *J Virol* 77: pp. 1427–40.

Mantri, M., Krojer, T., Bagg, E. A., et al. (2010). "Crystal structure of the 2-oxoglutarate- and Fe(II)-dependent lysyl hydroxylase JMJD6". *J Mol Biol* 401: pp. 211–22.

Marango, J., Shimoyama, M., Nishio, H., et al. (2008). "The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor". *Blood* 111: pp. 3145–54.

Marcon, E., Ni, Z., Pu, S., et al. (2014). "Human-chromatin-related protein interactions identify a demethylase complex required for chromosome segregation". *Cell Rep* 8: pp. 297–310.

Marguerat, S. and Bahler, J. (2012). "Coordinating genome expression with cell size". *Trends Genet* 28: pp. 560–5.

Mark, K. G., Loveless, T. B., and Toczyski, D. P. (2016). "Isolation of ubiquitinated substrates by tandem affinity purification of E3 ligase-polyubiquitin-binding domain fusions (ligase traps)". *Nat Protoc* 11: pp. 291–301.

Matic, I., Hagen, M. van, Schimmel, J., et al. (2008). "In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy". *Mol Cell Proteomics* 7: pp. 132–44.

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., et al. (2010). "Conserved role of intragenic DNA methylation in regulating alternative promoters". *Nature* 466: pp. 253–7.

McDaniel, S. L., Fligor, J. E., Ruan, C., et al. (2016). "Combinatorial Histone Readout by the Dual Plant Homeodomain (PHD) Fingers of Rco1 Mediates Rpd3S Chromatin Recruitment and the Maintenance of Transcriptional Fidelity". *J Biol Chem* 291: pp. 14796–802.

McDaniel, S. L. and Strahl, B. D. (2017). "Shaping the Cellular Landscape with Set2/SETD2 Methylation". *Cellular and molecular life sciences : CMLS* 74: pp. 3317–3334.

Meehan, R. R., Lewis, J. D., and Bird, A. P. (1992). "Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA". *Nucleic Acids Research* 20: pp. 5085–5092.

Meissner, A., Mikkelsen, T. S., Gu, H., et al. (2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells". *Nature* 454: pp. 766–70.

Mendenhall, E. M., Koche, R. P., Truong, T., et al. (2010). "GC-rich sequence elements recruit PRC2 in mammalian ES cells". *PLoS Genet* 6: e1001244.

Michel, B. C. and Kadoch, C. (2017). "A SMARCD2-containing mSWI/SNF complex is required for granulopoiesis". *Nat Genet* 49: pp. 655–657.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells". *Nature* 448: pp. 553–60.

Milne, T. A., Dou, Y., Martin, M. E., et al. (2005). "MLL associates specifically with a subset of transcriptionally active target genes". *Proc Natl Acad Sci U S A* 102: pp. 14765–70.

Milne, T. A., Kim, J., Wang, G. G., et al. (2010). "Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis". *Mol Cell* 38: pp. 853–63.

Min, J., Zhang, Y., and Xu, R. M. (2003). "Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27". *Genes Dev* 17: pp. 1823–8.

Miyazaki, H., Higashimoto, K., Yada, Y., et al. (2013). "Ash1l methylates Lys36 of histone H3 independently of transcriptional elongation to counteract polycomb silencing". *PLoS Genet* 9: e1003897.

Mohandas, T, Sparkes, R., and Shapiro, L. (1981). "Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation". *Science* 211: pp. 393–396.

Mohn, F., Weber, M., Rebhan, M., et al. (2008). "Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors". *Mol Cell* 30: pp. 755–66.

Mohrmann, L. and Verrijzer, C. P. (2005). "Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes". *Biochim Biophys Acta* 1681: pp. 59–73.

Morselli, M., Pastor, W. A., Montanini, B., et al. (2015). "In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse". *Elife* 4: e06205.

Muller, J. and Verrijzer, P. (2009). "Biochemical mechanisms of gene regulation by polycomb group protein complexes". *Curr Opin Genet Dev* 19: pp. 150–8.

Muller, J., Hart, C. M., Francis, N. J., et al. (2002). "Histone methyltransferase activity of a Drosophila Polycomb group repressor complex". *Cell* 111: pp. 197–208.

Muntean, A. G., Tan, J., Sitwala, K., et al. (2010). "The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis". *Cancer Cell* 17: pp. 609–21.

Murray, K. (1964). "The occurrence of epsilon-N-methyl lysine in histones". *Biochemistry* 3: pp. 10–5.

Nabet, B., Roberts, J. M., Buckley, D. L., et al. (2018). "The dTAG system for immediate and target-specific protein degradation". *Nat Chem Biol* 14: pp. 431–441.

Nakayama, J., Rice, J. C., Strahl, B. D., et al. (2001). "Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly". *Science* 292: pp. 110–3.

Nan, X., Campoy, F. J., and Bird, A. (1997). "MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin". *Cell* 88: pp. 471–81.

Nan, X., Ng, H. H., Johnson, C. A., et al. (1998). "Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex". *Nature* 393: pp. 386–9.

Napoles, M. de, Mermoud, J. E., Wakao, R., et al. (2004). "Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation". *Dev Cell* 7: pp. 663–76.

Nelson, C. J., Santos-Rosa, H., and Kouzarides, T. (2006). "Proline isomerization of histone H3 regulates lysine methylation and gene expression". *Cell* 126: pp. 905–16.

Neri, F., Rapelli, S., Krepelova, A., et al. (2017). "Intragenic DNA methylation prevents spurious transcription initiation". *Nature* 543: pp. 72–77.

Ng, H. H., Zhang, Y., Hendrich, B., et al. (1999). "MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex". *Nat Genet* 23: pp. 58–61.

Ng, H. H., Robert, F., Young, R. A., et al. (2003). "Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity". *Mol Cell* 11: pp. 709–19.

Nimura, K., Ura, K., Shiratori, H., et al. (2009). "A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome". *Nature* 460: pp. 287–91.

Nishimura, K., Fukagawa, T., Takisawa, H., et al. (2009). "An auxin-based degron system for the rapid depletion of proteins in nonplant cells". *Nat Methods* 6: pp. 917–22.

Norton, V. G., Imai, B. S., Yau, P., et al. (1989). "Histone acetylation reduces nucleosome core particle linking number change". *Cell* 57: pp. 449–57.

Nottke, A., Colaiacovo, M. P., and Shi, Y. (2009). "Developmental roles of the histone lysine demethylases". *Development* 136: pp. 879–89.

Okamoto, I. and Heard, E. (2009). "Lessons from comparative analysis of X-chromosome inactivation in mammals". *Chromosome Res* 17: pp. 659–69.

Okano, M., Bell, D. W., Haber, D. A., et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development". *Cell* 99: pp. 247–57.

Ooi, S. K., Qiu, C., Bernstein, E., et al. (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA". *Nature* 448: pp. 714–7.

Orlando, D. A., Chen, M. W., Brown, V. E., et al. (2014). "Quantitative ChIP-Seq normalization reveals global modulation of the epigenome". *Cell Rep* 9: pp. 1163–70.

O'Shaughnessy-Kirwan, A., Signolet, J., Costello, I., et al. (2015). "Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification". *Development (Cambridge, England)* 142: pp. 2586–2597.

Padovan-Merhar, O., Nair, G. P., Biaesch, A., et al. (2015). "Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms". *Molecular cell* 58: pp. 339–352.

Paik, W. K. and Kim, S. (1974). "Epsilon-alkyllysinase. New assay method, purification, and biological significance". *Arch Biochem Biophys* 165: pp. 369–78.

Paik, W. K. and Kim, S. (1973). "Enzymatic demethylation of calf thymus histones". *Biochemical and Biophysical Research Communications* 51: pp. 781–788.

Park, S., Osmers, U., Raman, G., et al. (2010). "The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression". *Biochemistry* 49: pp. 6576–86.

Pascual, J., Martinez-Yamout, M., Dyson, H. J., et al. (2000). "Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor". *J Mol Biol* 304: pp. 723–9.

Pasini, D., Bracken, A. P., Hansen, J. B., et al. (2007). "The polycomb group protein Suz12 is required for embryonic stem cell differentiation". *Mol Cell Biol* 27: pp. 3769–79.

Payer, B. and Lee, J. T. (2008). "X chromosome dosage compensation: how mammals keep the balance". *Annu Rev Genet* 42: pp. 733–72.

Pedersen, M. T. and Helin, K. (2010). "Histone demethylases in development and disease". *Trends Cell Biol* 20: pp. 662–71.

Pedersen, M. T., Agger, K., Laugesen, A., et al. (2014). "The demethylase JMJD2C localizes to H3K4me3-positive transcription start sites and is dispensable for embryonic development". *Mol Cell Biol* 34: pp. 1031–45.

Pedersen, M. T., Kooistra, S. M., Radzisheuskaya, A., et al. (2016). "Continual removal of H3K9 promoter methylation by Jmjd2 demethylases is vital for ESC self-renewal and early development". *Embo j* 35: pp. 1550–64.

Penzo, M., Casoli, L., Pollutri, D., et al. (2015). "JHDM1B expression regulates ribosome biogenesis and cancer cell growth in a p53 dependent manner". *Int J Cancer* 136: E272–81.

Perino, M., Mierlo, G. van, Karemaker, I. D., et al. (2018). "MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding". *Nat Genet* 50: pp. 1002–1010.

Peters, A. H.F. M., Kubicek, S., Mechtler, K., et al. (2003). "Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin". *Molecular Cell* 12: pp. 1577–1589.

Peterson, C. L. and Laniel, M. A. (2004). "Histones and histone modifications". *Curr Biol* 14: R546–51.

Pfau, R., Tzatsos, A., Kampranis, S. C., et al. (2008). "Members of a family of JmjC domain-containing oncoproteins immortalize embryonic fibroblasts via a JmjC domain-dependent process". *Proc Natl Acad Sci U S A* 105: pp. 1907–12.

Pokholok, D. K., Harbison, C. T., Levine, S., et al. (2005). "Genome-wide map of nucleosome acetylation and methylation in yeast". *Cell* 122: pp. 517–27.

Ponnaluri, V. K., Vavilala, D. T., Putty, S., et al. (2009). "Identification of non-histone substrates for JMJD2A-C histone demethylases". *Biochem Biophys Res Commun* 390: pp. 280–4.

Portela, A. and Esteller, M. (2010). "Epigenetic modifications and human disease". *Nat Biotechnol* 28: pp. 1057–68.

Pradhan, S., Bacolla, A., Wells, R. D., et al. (1999). "Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation". *J Biol Chem* 274: pp. 33002–10.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., et al. (2001). "The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor". *Genes Dev* 15: pp. 1613–8.

Qiao, Q., Li, Y., Chen, Z., et al. (2011). "The Structure of NSD1 Reveals an Autoregulatory Mechanism Underlying Histone H3K36 Methylation". *The Journal of Biological Chemistry* 286: pp. 8361–8368.

Qin, S., Guo, Y., Xu, C., et al. (2013). "Tudor domains of the PRC2 components PHF1 and PHF19 selectively bind to histone H3K36me3". *Biochem Biophys Res Commun* 430: pp. 547–53.

Quinlan, A. R. and Hall, I. M. (2010). "BEDTools: a flexible suite of utilities for comparing genomic features". *Bioinformatics* 26: pp. 841–2.

Rahman, S., Sowa, M. E., Ottinger, M., et al. (2011). "The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3". *Mol Cell Biol* 31: pp. 2641–52.

Ram, O., Goren, A., Amit, I., et al. (2011). "Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells". *Cell* 147: pp. 1628–1639.

Ramirez, F., Ryan, D. P., Gruning, B., et al. (2016). "deepTools2: a next generation web server for deep-sequencing data analysis". *Nucleic Acids Res* 44: W160–5.

Ramirez, J., Dege, C., Kutateladze, T. G., et al. (2012). "MBD2 and multiple domains of CHD4 are required for transcriptional repression by Mi-2/NuRD complexes". *Mol Cell Biol* 32: pp. 5078–88.

Ramirez-Carrozzi, V. R., Braas, D., Bhatt, D. M., et al. (2009). "A Unifying Model for the Selective Regulation of Inducible Transcription By CpG Islands and Nucleosome Remodeling". *Cell* 138: pp. 114–128.

Ran, F. A., Hsu, P. D., Wright, J., et al. (2013). "Genome engineering using the CRISPR-Cas9 system". *Nat Protoc* 8: pp. 2281–2308.

Rao, B., Shibata, Y., Strahl, B. D., et al. (2005). "Dimethylation of Histone H3 at Lysine 36 Demarcates Regulatory and Nonregulatory Chromatin Genome-Wide". *Molecular and Cellular Biology* 25: pp. 9447–9459.

Rayasam, G. V., Wendling, O., Angrand, P. O., et al. (2003). "NSD1 is essential for early post-implantation development and has a catalytically active SET domain". *Embo j* 22: pp. 3153–63.

Razin, A. and Cedar, H. (1994). "DNA methylation and genomic imprinting". *Cell* 77: pp. 473–6.

Rea, S., Eisenhaber, F., O'Carroll, D., et al. (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases". *Nature* 406: pp. 593–9.

Reitsma, J. M., Liu, X., Reichermeier, K. M., et al. (2017). "Composition and Regulation of the Cellular Repertoire of SCF Ubiquitin Ligases". *Cell* 171: 1326–1339.e14.

Reynolds, N., Salmon-Divon, M., Dvinge, H., et al. (2012). "NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression". *Embo j* 31: pp. 593–605.

Riggs, A. D. (1975). "X inactivation, differentiation, and DNA methylation". *Cytogenet Cell Genet* 14: pp. 9–25.

Riising, E. M., Comet, I., Leblanc, B., et al. (2014). "Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide". *Mol Cell* 55: pp. 347–60.

Robertson, K. D. and Wolffe, A. P. (2000). "DNA methylation in health and disease". *Nat Rev Genet* 1: pp. 11–9.

Robin, P., Fritsch, L., Philipot, O., et al. (2007). "Post-translational modifications of histones H3 and H4 associated with the histone methyltransferases Suv39h1 and G9a". *Genome Biology* 8: R270–R270.

Rondelet, G., Dal Maso, T., Willems, L., et al. (2016). "Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B". *J Struct Biol* 194: pp. 357–67.

Rose, N. R., King, H. W., Blackledge, N. P., et al. (2016). "RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes". *Elife* 5.

Ruan, C., Lee, C. H., Cui, H., et al. (2015). "Nucleosome contact triggers conformational changes of Rpd3S driving high-affinity H3K36me nucleosome engagement". *Cell Rep* 10: pp. 204–15.

Sabo, P. J., Humbert, R., Hawrylycz, M., et al. (2004). "Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries". *Proc Natl Acad Sci U S A* 101: pp. 4537–42.

Saha, A. and Deshaies, R. J. (2008). "Multimodal activation of the ubiquitin ligase SCF by Nedd8 conjugation". *Mol Cell* 32: pp. 21–31.

Sanchez, C., Sanchez, I., Demmers, J. A., et al. (2007). "Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor". *Mol Cell Proteomics* 6: pp. 820–34.

Sanchez, R. and Zhou, M. M. (2011). "The PHD finger: a versatile epigenome reader". *Trends Biochem Sci* 36: pp. 364–72.

Sankar, A., Kooistra, S. M., Gonzalez, J. M., et al. (2017). "Maternal expression of the histone demethylase Kdm4a is crucial for pre-implantation development". *Development* 144: pp. 3264–3277.

Sankaran, S. M., Wilkinson, A. W., Elias, J. E., et al. (2016). "A PWWP Domain of Histone-Lysine N-Methyltransferase NSD2 Binds to Dimethylated Lys-36 of Histone H3 and Regulates NSD2 Function at Chromatin". *J Biol Chem* 291: pp. 8465–74.

Sato, M. H., Ura, K., Hohmura, K. I., et al. (1999). "Atomic force microscopy sees nucleosome positioning and histone H1-induced compaction in reconstituted chromatin". *FEBS Lett* 452: pp. 267–71.

Saxonov, S., Berg, P., and Brutlag, D. L. (2006). "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters". *Proc Natl Acad Sci U S A* 103: pp. 1412–7.

Schaft, D., Roguev, A., Kotovic, K. M., et al. (2003). "The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation". *Nucleic Acids Res* 31: pp. 2475–82.

Scheel, H. and Hofmann, K. (2003). "No evidence for PHD fingers as ubiquitin ligases". *Trends Cell Biol* 13: 285–7; author reply 287–8.

Schilling, E. and Rehli, M. (2007). "Global, comparative analysis of tissue-specific promoter CpG methylation". *Genomics* 90: pp. 314–23.

Schimmel, J., Larsen, K. M., Matic, I., et al. (2008). "The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle". *Mol Cell Proteomics* 7: pp. 2107–22.

Schmidt, D., Wilson, M. D., Spyrou, C., et al. (2009). "ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions". *Methods* 48: pp. 240–8.

Schmitges, F. W., Prusty, A. B., Faty, M., et al. (2011). "Histone methylation by PRC2 is inhibited by active chromatin marks". *Mol Cell* 42: pp. 330–41.

Schneider, R., Bannister, A. J., Myers, F. A., et al. (2004). "Histone H3 lysine 4 methylation patterns in higher eukaryotic genes". *Nat Cell Biol* 6: pp. 73–7.

Schoeftner, S., Sengupta, A. K., Kubicek, S., et al. (2006). "Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing". *Embo j* 25: pp. 3110–22.

Schotta, G., Sengupta, R., Kubicek, S., et al. (2008). "A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse". *Genes Dev* 22: pp. 2048–61.

Schroeder, S. C., Schwer, B., Shuman, S., et al. (2000). "Dynamic association of capping enzymes with transcribing RNA polymerase II". *Genes Dev* 14: pp. 2435–40.

Shen, C., Ipsaro, J. J., Shi, J., et al. (2015). "NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler". *Mol Cell* 60: pp. 847–59.

Shen, L., Kondo, Y., Guo, Y., et al. (2007). "Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters". *PLoS Genet* 3: pp. 2023–36.

Shi, L., Sun, L., Li, Q., et al. (2011). "Histone demethylase JMJD2B coordinates H3K4/H3K9 methylation and promotes hormonally responsive breast carcinogenesis". *Proceedings of the National Academy of Sciences of the United States of America* 108: pp. 7541–7546.

Shi, X., Hong, T., Walter, K. L., et al. (2006). "ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression". *Nature* 442: pp. 96–9.

Shi, Y. and Whetstine, J. R. (2007). "Dynamic regulation of histone lysine methylation by demethylases". *Mol Cell* 25: pp. 1–14.

Shi, Y., Lan, F., Matson, C., et al. (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1". *Cell* 119: pp. 941–53.

Shiio, Y. and Eisenman, R. N. (2003). "Histone sumoylation is associated with transcriptional repression". *Proc Natl Acad Sci U S A* 100: pp. 13225–30.

Shogren-Knaak, M., Ishii, H., Sun, J. M., et al. (2006). "Histone H4-K16 acetylation controls chromatin structure and protein interactions". *Science* 311: pp. 844–7.

Simon, J. A. and Kingston, R. E. (2013). "Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put". *Mol Cell* 49: pp. 808–24.

Simone, C. (2006). "SWI/SNF: the crossroads where extracellular signaling pathways meet chromatin". *J Cell Physiol* 207: pp. 309–14.

Sims R. J., r., Chen, C. F., Santos-Rosa, H., et al. (2005). "Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains". *J Biol Chem* 280: pp. 41789–92.

Sinha, K. M., Yasuda, H., Coombes, M. M., et al. (2010). "Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase". *The EMBO Journal* 29: pp. 68–79.

Skourti-Stathaki, K., Proudfoot, N. J., and Gromak, N. (2011). "Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination". *Mol Cell* 42: pp. 794–805.

Smith, S. T., Petruk, S., Sedkov, Y., et al. (2004). "Modulation of heat shock gene expression by the TAC1 chromatin-modifying complex". *Nat Cell Biol* 6: pp. 162–7.

Smolle, M., Venkatesh, S., Gogol, M. M., et al. (2012). "Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange". *Nature structural & molecular biology* 19: pp. 884–892.

Sohn, D. H., Lee, K. Y., Lee, C., et al. (2007). "SRG3 interacts directly with the major components of the SWI/SNF chromatin remodeling complex and protects them from proteasomal degradation". *J Biol Chem* 282: pp. 10614–24.

Song, J., Rechkoblit, O., Bestor, T. H., et al. (2011). "Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation". *Science* 331: pp. 1036–40.

Stalder, J., Larsen, A., Engel, J. D., et al. (1980). "Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I". *Cell* 20: pp. 451–60.

Steffen, P. A. and Ringrose, L. (2014). "What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory". *Nat Rev Mol Cell Biol* 15: pp. 340–56.

Strahl, B. D. and Allis, C. D. (2000). "The language of covalent histone modifications". *Nature* 403: pp. 41–5.

Strahl, B. D., Grant, P. A., Briggs, S. D., et al. (2002). "Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression". *Mol Cell Biol* 22: pp. 1298–306.

Streubel, G., Watson, A., Jammula, S. G., et al. (2018). "The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells". *Molecular Cell* 70: 371–379.e5.

Struhl, K. (1999). "Fundamentally different logic of gene regulation in eukaryotes and prokaryotes". *Cell* 98: pp. 1–4.

Suetake, I., Shinozaki, F., Miyagawa, J., et al. (2004). "DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction". *J Biol Chem* 279: pp. 27816–23.

Sun, X. J., Wei, J., Wu, X. Y., et al. (2005). "Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase". *J Biol Chem* 280: pp. 35261–71.

Sun, Y., Ding, L., Zhang, H., et al. (2006). "Potentiation of Smad-mediated transcriptional activation by the RNA-binding protein RBPMS". *Nucleic Acids Res* 34: pp. 6314–26.

Sural, T. H., Peng, S., Li, B., et al. (2008). "The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila X chromosome". *Nature structural & molecular biology* 15: pp. 1318–1325.

Suzuki, T., Minehata, K., Akagi, K., et al. (2006). "Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice". *Embo j* 25: pp. 3422–31.

Sved, J. and Bird, A. (1990). "The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model". *Proc Natl Acad Sci U S A* 87: pp. 4692–6.

Swartz, M. N., Trautner, T. A., and Kornberg, A. (1962). "Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids". *J Biol Chem* 237: pp. 1961–7.

Sykes, P. J., Neoh, S. H., Brisco, M. J., et al. (1992). "Quantitation of targets for PCR by use of limiting dilution". *Biotechniques* 13: pp. 444–9.

Tan, M. K., Lim, H. J., and Harper, J. W. (2011). "SCF(FBXO22) regulates histone H3 lysine 9 and 36 methylation levels by targeting histone demethylase KDM4A for ubiquitin-mediated proteasomal degradation". *Mol Cell Biol* 31: pp. 3687–99.

Tan, M. K., Lim, H. J., Bennett, E. J., et al. (2013). "Parallel SCF adaptor capture proteomics reveals a role for SCFFBXL17 in NRF2 activation via BACH1 repressor turnover". *Mol Cell* 52: pp. 9–24.

Tan, Y., Jin, C., Ma, W., et al. (2018). "Dismissal of RNA Polymerase II Underlies a Large Ligand-Induced Enhancer Decommissioning Program". *Mol Cell* 71: 526–539.e8.

Tanaka, Y., Katagiri, Z., Kawahashi, K., et al. (2007). "Trithorax-group protein ASH1 methylates histone H3 lysine 36". *Gene* 397: pp. 161–8.

Tanaka, Y., Okamoto, K., Teye, K., et al. (2010). "JmjC enzyme KDM2A is a regulator of rRNA transcription in response to starvation". *The EMBO Journal* 29: pp. 1510–1522.

Tatarskiy, V. V., Simonov, Y. P., Shcherbinin, D. S., et al. (2017). "Stability of the PHF10 subunit of PBAF signature module is regulated by phosphorylation: role of beta-TrCP". *Sci Rep* 7: p. 5645.

Tatham, M. H., Jaffray, E., Vaughan, O. A., et al. (2001). "Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9". *J Biol Chem* 276: pp. 35368–74.

Tavares, L., Dimitrova, E., Oxley, D., et al. (2012). "RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3". *Cell* 148: pp. 664–78.

Taverna, S. D., Li, H., Ruthenburg, A. J., et al. (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers". *Nat Struct Mol Biol* 14: pp. 1025–1040.

Tazi, J. and Bird, A. (1990). "Alternative chromatin structure at CpG islands". *Cell* 60: pp. 909–20.

Teif, V. B., Vainshtein, Y., Caudron-Herger, M., et al. (2012). "Genome-wide nucleosome positioning during embryonic stem cell development". *Nat Struct Mol Biol* 19: pp. 1185–92.

Testoni, S., Bartolone, E., Rossi, M., et al. (2012). "KDM2B is implicated in bovine lethal multi-organic developmental dysplasia". *PLoS One* 7: e45634.

Thomson, J. P., Skene, P. J., Selfridge, J., et al. (2010). "CpG islands influence chromatin structure via the CpG-binding protein Cfp1". *Nature* 464: pp. 1082–6.

Tie, F., Banerjee, R., Stratton, C. A., et al. (2009). "CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing". *Development* 136: pp. 3131–41.

Trizzino, M., Barbieri, E., Petracovici, A., et al. (2018). "The Tumor Suppressor ARID1A Controls Global Transcription via Pausing of RNA Polymerase II". *Cell Rep* 23: pp. 3933–3945.

Trojer, P., Zhang, J., Yonezawa, M., et al. (2009). "Dynamic Histone H1 Isotype 4 Methylation and Demethylation by Histone Lysine Methyltransferase G9a/KMT1C and the Jumonji Domain-containing JMJD2/KDM4 Proteins". *J Biol Chem* 284: pp. 8395–405.

Tropberger, P. and Schneider, R. (2010). "Going global: novel histone modifications in the globular domain of H3". *Epigenetics* 5: pp. 112–7.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., et al. (2006). "Histone demethylation by a family of JmjC domain-containing proteins". *Nature* 439: pp. 811–6.

Turner, B. M. (1993). "Decoding the nucleosome". *Cell* 75: pp. 5–8.

Turner, B. M. (2000). "Histone acetylation and an epigenetic code". *Bioessays* 22: pp. 836–45.

Tykocinski, M. L. and Max, E. E. (1984). "CG dinucleotide clusters in MHC genes and in 5' demethylated genes". *Nucleic Acids Res* 12: pp. 4385–96.

Tzatsos, A., Pfau, R., Kampranis, S. C., et al. (2009). "Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus". *Proc Natl Acad Sci U S A* 106: pp. 2641–6.

Tzatsos, A., Paskaleva, P., Ferrari, F., et al. (2013). "KDM2B promotes pancreatic cancer via Polycomb-dependent and -independent transcriptional programs". *J Clin Invest* 123: pp. 727–39.

Uchida, D., Hatakeyama, S., Matsushima, A., et al. (2004). "AIRE functions as an E3 ubiquitin ligase". *J Exp Med* 199: pp. 167–72.

Ulrich, H. D. (2005). "Mutual interactions between the SUMO and ubiquitin systems: a plea of no contest". *Trends Cell Biol* 15: pp. 525–32.

Vardimon, L., Kuhlmann, I., Doerfler, W., et al. (1981). "Methylation of adenovirus genes in transformed cells and in vitro: influence on the regulation of gene expression?" *Eur J Cell Biol* 25: pp. 13–5.

Venkatesh, S. and Workman, J. L. (2015). "Histone exchange, chromatin structure and the regulation of transcription". *Nat Rev Mol Cell Biol* 16: pp. 178–89.

Venkatesh, S., Smolle, M., Li, H., et al. (2012). "Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes". *Nature* 489: pp. 452–5.

Venolia, L., Gartler, S. M., Wassman, E. R., et al. (1982). "Transformation with DNA from 5-azacytidine-reactivated X chromosomes". *Proc Natl Acad Sci U S A* 79: pp. 2352–4.

Vermeulen, M., Mulder, K. W., Denissov, S., et al. (2007). "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4". *Cell* 131: pp. 58–69.

Vertino, P. M., Sekowski, J. A., Coll, J. M., et al. (2002). "DNMT1 is a component of a multiprotein DNA replication complex". *Cell Cycle* 1: pp. 416–23.

Vogelstein, B. and Kinzler, K. W. (1999). "Digital PCR". *Proc Natl Acad Sci U S A* 96: pp. 9236–41.

Voo, K. S., Carlone, D. L., Jacobsen, B. M., et al. (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: pp. 2108–2121.

Vorobyeva, N. E., Soshnikova, N. V., Nikolenko, J. V., et al. (2009). "Transcription coactivator SAYP combines chromatin remodeler Brahma and transcription initiation factor TFIID into a single supercomplex". *Proc Natl Acad Sci U S A* 106: pp. 11049–54.

Voss, T. C. and Hager, G. L. (2014). "Dynamic regulation of transcriptional states by chromatin and transcription factors". *Nat Rev Genet* 15: pp. 69–81.

Wade, P. A., Pruss, D., and Wolffe, A. P. (1997). "Histone acetylation: chromatin in action". *Trends Biochem Sci* 22: pp. 128–32.

Waechter, D. E. and Baserga, R. (1982). "Effect of methylation on expression of microinjected genes". *Proc Natl Acad Sci U S A* 79: pp. 1106–10.

Wagner, E. J. and Carpenter, P. B. (2012). "Understanding the language of Lys36 methylation at histone H3". *Nat Rev Mol Cell Biol* 13: pp. 115–26.

Wagner, K. W., Alam, H., Dhar, S. S., et al. (2013). "KDM2A promotes lung tumorigenesis by epigenetically enhancing ERK1/2 signaling". *The Journal of Clinical Investigation* 123: pp. 5231–5246.

Walport, L. J., Hopkinson, R. J., Chowdhury, R., et al. (2016). "Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases". *Nature Communications* 7: p. 11974.

Wang, G. G., Cai, L., Pasillas, M. P., et al. (2007). "NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis". *Nat Cell Biol* 9: pp. 804–12.

Wang, H., Wang, L., Erdjument-Bromage, H., et al. (2004). "Role of histone H2A ubiquitination in Polycomb silencing". *Nature* 431: pp. 873–8.

Wang, J., Muntean, A. G., Wu, L., et al. (2012). "A subset of mixed lineage leukemia proteins has plant homeodomain (PHD)-mediated E3 ligase activity". *J Biol Chem* 287: pp. 43410–6.

Wang, J., Tadeo, X., Hou, H., et al. (2013). "Epe1 recruits BET family bromodomain protein Bdf2 to establish heterochromatin boundaries". *Genes Dev* 27: pp. 1886–902.

Wang, L. Y., Hung, C. L., Chen, Y. R., et al. (2016). "KDM4A Coactivates E2F1 to Regulate the PDK-Dependent Metabolic Switch between Mitochondrial Oxidation and Glycolysis". *Cell Rep* 16: pp. 3016–3027.

Wang, T., Chen, K., Zeng, X., et al. (2011). "The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner". *Cell Stem Cell* 9: pp. 575–87.

Wang, Z., Song, J., Milne, T. A., et al. (2010). "Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression". *Cell* 141: pp. 1183–94.

Watt, F. and Molloy, P. L. (1988). "Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter". *Genes Dev* 2: pp. 1136–43.

Weber, M., Hellmann, I., Stadler, M. B., et al. (2007). "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome". *Nat Genet* 39: pp. 457–66.

Wei, J., Mialki, R. K., Dong, S., et al. (2013). "A new mechanism of RhoA ubiquitination and degradation: roles of SCF(FBXL19) E3 ligase and Erk2". *Biochim Biophys Acta* 1833: pp. 2757–2764.

Wei, N. and Deng, X. W. (1999). "Making sense of the COP9 signalosome. A regulatory protein complex conserved from Arabidopsis to human". *Trends Genet* 15: pp. 98–103.

Whetstine, J. R., Nottke, A., Lan, F., et al. (2006). "Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases". *Cell* 125: pp. 467–81.

Wilkins, S. E., Islam, M. S., Gannon, J. M., et al. (2018). "JMJD5 is a human arginyl C-3 hydroxylase". *Nature Communications* 9: p. 1180.

Williams, L. H., Fromm, G., Gokey, N. G., et al. (2015). "Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks". *Mol Cell* 58: pp. 311–322.

Wittschieben, B. O., Otero, G., Bizemont, T. de, et al. (1999). "A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme". *Mol Cell* 4: pp. 123–8.

Witzel, M., Petersheim, D., Fan, Y., et al. (2017). "Chromatin-remodeling factor SMARCD2 regulates transcriptional networks controlling differentiation of neutrophil granulocytes". *Nature genetics* 49: pp. 742–752.

Wong, S. J., Gearhart, M. D., Taylor, A. B., et al. (2016). "KDM2B Recruitment of the Polycomb Group Complex, PRC1.1, Requires Cooperation between PCGF1 and BCORL1". *Structure* 24: pp. 1795–1801.

Wu, X., Johansen, J. V., and Helin, K. (2013). "Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation". *Mol Cell* 49: pp. 1134–46.

Wysocka, J., Swigut, T., Xiao, H., et al. (2006). "A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling". *Nature* 442: pp. 86–90.

Xiao, T., Hall, H., Kizer, K. O., et al. (2003). "Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast". *Genes Dev* 17: pp. 654–63.

Xie, L., Pelz, C., Wang, W., et al. (2011). "KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription". *Embo j* 30: pp. 1473–84.

Xu, C., Bian, C., Lam, R., et al. (2011). "The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain". *Nat Commun* 2: p. 227.

Xu, C., Liu, K., Lei, M., et al. (2018). "DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants". *Structure* 26: 85–95.e3.

Yamagishi, T., Hirose, S., and Kondo, T. (2008). "Secondary DNA structure formation for Hoxb9 promoter and identification of its specific binding protein". *Nucleic Acids Res* 36: pp. 1965–75.

Yang, L., Lin, C., Liu, W., et al. (2011). "ncRNA- and Pc2 Methylation-Dependent Gene Relocation between Nuclear Structures Mediates Gene Activation Programs". *Cell* 147: pp. 773–788.

Yang, S., Zheng, X., Lu, C., et al. (2016). "Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase". *Genes Dev* 30: pp. 1611–6.

Yang, W.-S., Campbell, M., and Chang, P.-C. (2017). "SUMO modification of a heterochromatin histone demethylase JMJD2A enables viral gene transactivation and viral replication". *PLoS Pathogens* 13: e1006216.

Yoh, S. M., Lucas, J. S., and Jones, K. A. (2008). "The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation". *Genes Dev* 22: pp. 3422–34.

Yonashiro, R., Ishido, S., Kyo, S., et al. (2006). "A novel mitochondrial ubiquitin ligase plays a critical role in mitochondrial dynamics". *Embo j* 25: pp. 3618–26.

Yoon, H. G., Chan, D. W., Reynolds, A. B., et al. (2003). "N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso". *Mol Cell* 12: pp. 723–34.

Youdell, M. L., Kizer, K. O., Kisseleva-Romanova, E., et al. (2008). "Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36". *Mol Cell Biol* 28: pp. 4915–26.

Youn, M. Y., Yokoyama, A., Fujiyama-Nakamura, S., et al. (2012). "JMJD5, a Jumonji C (JmjC) domain-containing protein, negatively regulates osteoclastogenesis by facilitating NFATc1 protein degradation". *J Biol Chem* 287: pp. 12994–3004.

Yu, G., Wang, J., Lin, X., et al. (2016). "Demethylation of SFRP2 by histone demethylase KDM2A regulated osteo-/dentinogenic differentiation of stem cells of the apical papilla". *Cell Prolif* 49: pp. 330–40.

Yuan, G., Ma, B., Yuan, W., et al. (2013). "Histone H2A Ubiquitination Inhibits the Enzymatic Activity of H3 Lysine 36 Methyltransferases". *The Journal of Biological Chemistry* 288: pp. 30832–30842.

Yuan, W., Xie, J., Long, C., et al. (2009). "Heterogeneous nuclear ribonucleoprotein L Is a subunit of human KMT3a/Set2 complex required for H3 Lys-36 trimethylation activity in vivo". *J Biol Chem* 284: pp. 15701–7.

Yuan, W., Wu, T., Fu, H., et al. (2012). "Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation". *Science* 337: pp. 971–5.

Yuan, W., Xu, M., Huang, C., et al. (2011). "H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation". *The Journal of Biological Chemistry* 286: pp. 7983–7989.

Zeleznik-Le, N. J., Harden, A. M., and Rowley, J. D. (1994). "11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene". *Proc Natl Acad Sci U S A* 91: pp. 10610–4.

Zhang, H., Kobayashi, R., Galaktionov, K., et al. (1995). "p19Skp1 and p45Skp2 are essential elements of the cyclin A-CDK2 S phase kinase". *Cell* 82: pp. 915–25.

Zhang, S., Roche, K., Nasheuer, H. P., et al. (2011). "Modification of histones by sugar beta-N-acetylglucosamine (GlcNAc) occurs on multiple residues, including histone H3 serine 10, and is cell cycle-regulated". *J Biol Chem* 286: pp. 37483–95.

Zhang, Y. and Reinberg, D. (2001). "Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails". *Genes Dev* 15: pp. 2343–60.

Zhang, Y., Shan, C.-M., Wang, J., et al. (2017). "Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation". *Scientific Reports* 7: p. 43906.

Zhao, J., Wei, J., Mialki, R. K., et al. (2012). "F-box protein FBXL19-mediated ubiquitination and degradation of the receptor for IL-33 limits pulmonary inflammation". *Nat Immunol* 13: pp. 651–8.

Zhao, J., Mialki, R. K., Wei, J., et al. (2013). "SCF E3 ligase F-box protein complex SCF(FBXL19) regulates cell migration by mediating Rac1 ubiquitination and degradation". *Faseb j* 27: pp. 2611–9.

Zheng, W., Ibanez, G., Wu, H., et al. (2012). "Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2". *J Am Chem Soc* 134: pp. 18004–14.

Zheng, Y., Xue, Y., Ren, X., et al. (2018). "The Lysine Demethylase dKDM2 Is Non-essential for Viability, but Regulates Circadian Rhythms in Drosophila". *Frontiers in Genetics* 9.

Zhou, C., Wee, S., Rhee, E., et al. (2003). "Fission yeast COP9/signalosome suppresses cullin activity through recruitment of the deubiquitylating enzyme Ubp12p". *Mol Cell* 11: pp. 927–38.

Zhou, J. C., Blackledge, N. P., Farcas, A. M., et al. (2012). "Recognition of CpG island chromatin by KDM2A requires direct and specific interaction with linker DNA". *Mol Cell Biol* 32: pp. 479–89.

Zofall, M. and Grewal, S. I. (2006). "Swi6/HP1 recruits a JmjC domain protein to facilitate transcription of heterochromatic repeats". *Mol Cell* 22: pp. 681–92.