

A New Family of Vertebrate-Specific Polycombs Encoded by the *LCOR/LCORL* Genes Balance PRC2 Subtype Activities.

Eric M. Conway

Thesis submitted to Trinity College Dublin

for the degree of Doctor of Philosophy

2017

Thesis supervisor: Prof. Adrian P. Bracken

Cancer Epigenetics Laboratory, Department of Genetics,

University of Dublin, Trinity College, Dublin 2

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work, except where otherwise acknowledged.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Eric Conway, September 2017

Acknowledgements

First of all, I would like to thank Adrian Bracken for giving me the opportunity to do a PhD in his lab, and most importantly for his fantastic mentorship and guidance throughout the four years. I would also like to thank the range of collaborators who made the work in this thesis possible, Gerard Cagney and Giorgio Oliviero in UCD, Raphael Margueron and Daniel Holoch in the Curie Institute, Haruhiko Koseki and colleagues in RIKEN, Luciano Di Croce at the CRG in Barcelona, and finally Aoife McLysaght, Karsten Hokamp, Darren Fitzpatrick and Alan Rice at the Smurfit Institute of Genetics in TCD.

In addition, I am very grateful to the Irish Cancer Society and BREAST-PREDICT for providing the funding that allowed me to carry out my PhD projects.

I am deeply indebted to all members of the Bracken laboratory, past and present, both for my training in the early years and more importantly, for the vibrant collaborative environment that we have cultivated. Without the help of Orla Deevy, Eleanor Glancy, Marlena Mucha and Evan Healy this project would be unrecognisable from what it is today. I am particularly grateful to Evan, for listening to my constant ramblings (both scientific and otherwise), and Emilia Jerman for allowing me to inherit a fantastic project built upon her rock solid foundations. More importantly, I am sorry that all of you have had to listen to my choices in music for four years!

I am very lucky to have my parents, and my brother Ben, for their constant love and support. I would also like to thank my girlfriend Aisling, who doubles as a source of both emotional and scientific support. Finally, I am grateful to have Frankie and Jonah Conway and of course Quilliam B. Yeats in my life to remind me there is life outside of the lab!

Table of Contents

Chapter 1: Introduction	14
1.1 General introduction: Transcriptional memory	15
1.2 Chromatin and histone post-translation modifications.	16
1.3 Readers, writers and erasers and their role in transcriptional memory.....	19
1.4 Polycomb group proteins	21
1.5 Canonical and non-canonical PRC1 complexes.....	24
1.6 The emerging paradigm of variant PRC2 complexes	30
1.7 The genes encoding PRC2 and Histone 3 are frequently mutated in cancer.	33
1.8 The discovery of PALI1.....	36
Chapter 2: Materials and methods	40
2.1 Cell culture.	41
2.2 FLAG immunoprecipitations for mass spectrometric analyses.	42
2.3 Mass spectrometry analysis.....	42
2.4 Cloning and plasmid generation.	44
2.5 Generation of overexpression cell lines.	45
2.6 Preparation of whole cell protein lysates.	46
2.7 RNA preparation and RT- analysis.	46
2.8 FLAG immunoprecipitations.....	46
2.9 Evolutionary analysis.	47
2.10 Endogenous IPs.....	47
2.11 Chromatin Immunoprecipitations.	48
2.12 RNA-seq and bioinformatics analysis.	49
2.13 Generation of Pali1/2 double knockout cells.....	50
2.14 Accession codes	51
2.15 Mouse dissection	51
2.16 3T3 growth assay.....	51

2.17 Antibodies	51
2.18 Oligonucleotides	52
Chapter 3. Proteomic analysis of the PRC1 complex	55
3.1 Introduction	56
3.2 Results	58
3.2.1 Evaluation of “in gel” and “in solution” tryptic digest LC-MS/MS methods for identification of novel PRC1 associated proteins.	58
3.2.2 <i>In vivo</i> BMI1 co-immunoprecipitation.	60
3.2.3 HOX-9 proteins function as transcriptional repressors but do not recruit PRC1.	61
3.2.4 Ectopic expression of HOX-9s does not repress <i>INK4A</i> and has no effect on cellular proliferation.	63
3.2.6 BMI1 AF27 antibody non-specifically immunoprecipitates HOX-9 proteins.	65
3.2.7 Proteomic analysis of RYBP, CBX8 and PCGF2.	66
3.3 Discussion.....	69
Chapter 4: Biochemical and evolutionary analysis of PALI1.....	71
4.1 Introduction	72
4.2 Results	74
4.2.1 Mapping the minimal region of PALI1 required for interaction with PRC2. ..	74
4.2.2 Identification of a novel isoform transcribed from the <i>LCORL</i> gene locus (<i>PALI2</i>) that shares the Polycomb interaction region of PALI1.	75
4.2.3 The PIP domain of PALI1 and PALI2 arose during vertebrate evolution	76
4.2.4 Identification of two aromatic residues that are essential for the interaction between the PIP domain and PRC2.....	78
4.3 Discussion	80
Chapter 5. Characterisation of the functional and biochemical antagonism between Pali1 and Aebp2	82
5.1 Introduction	83
5.2 Results	85

5.2.1 Pali1 knockout strategy in mouse embryonic stem cells.	85
5.2.2 Pali1 is essential for mouse development and promotes H3K27me3 <i>in vivo</i> .85	
5.2.3 Pali2 knockout strategy using CRISPR-Cas9 mediated gene editing.	86
5.2.4 Pali1 and Aebp2 antagonise each other biochemically for protein stability..	86
5.2.5 Pali1 and Aebp2 occupy mutually exclusive sub-types of PRC2 complexes.	88
5.2.6 Consequences of knockout of PRC2 subunits on H3K27 methylation.....	88
5.2.7 Loss of Pali1/2 or Aebp2 has opposing effects on PRC2 activity <i>in vitro</i> and <i>in vivo</i>	89
5.2.8 The balance of PRC2.1 and PRC2.2 activities is essential for the proper regulation of Polycomb target genes during ESC differentiation.	91
5.2.9 Pali1/2 and Aebp2 competition model.....	93
5.3 Discussion.....	94
Chapter 6: General Discussion.....	97
6.1 Discussion.....	98
6.2 Conclusions	103
References	104

Table of figures

Chapter 1

Figure 1.1: The nucleosome and histone post-translational modifications.

Figure 1.2: Readers, writers and erasers of histone post-translational modifications.

Figure 1.3: The role of Polycomb group proteins in cell fate decisions.

Figure 1.4: The Variable compositions of the PRC1 and PRC2 complexes.

Table 1.5: Mutations in genes encoding PRC2 subunits and Histone H3 in cancer.

Figure 1.6: PALI1 associates with CTBP co-repressor proteins and PRC2.1 and promotes HKMT activity *in vitro*.

Chapter 3

Figure 3.1: Evaluation of in gel and in solution tryptic digest LC-MS/MS methods for identification of novel PRC1 associated proteins.

Figure 3.2: *In vivo* BMI1 co-immunoprecipitation.

Figure 3.3: HOX-9 proteins function as transcriptional repressors but do not recruit PRC1.

Figure 3.4: Ectopic expression of HOX-9s does not repress *INK4A* and has no effect on cellular proliferation.

Figure 3.5: HOX-9 proteins do not co-immunoprecipitate BMI1.

Figure 3.6: BMI1 AF27 antibody non-specifically immunoprecipitates HOX-9 proteins.

Figure 3.7: Proteomic analysis of RING1B, RYBP, CBX8 and PCGF2.

Chapter 4

Figure 4.1: Mapping the minimal region of PALI1 required for interaction with PRC2.

Figure 4.2: Identification of a novel isoform transcribed from the *LCORL* gene locus (*PALI2*) that shares the Polycomb interaction region of PALI1.

Figure 4.3: The PIP domain of PALI1 and PALI2 arose during vertebrate evolution.

Figure 4.4: Identification of two aromatic residues that are essential for the interaction between the PIP domain and PRC2.

Chapter 5

Figure 5.1: Pali1 knockout strategy in mouse embryonic stem cells.

Figure 5.2: Pali1 is essential for mouse development and promotes H3K27me3 *in vivo*.

Figure 5.3: Pali2 knockout strategy using CRISPR-Cas9 mediated gene editing.

Figure 5.4: Pali1 and Aebp2 antagonise each other biochemically for protein stability.

Figure 5.5: Pali1 and Aebp2 occupy mutually exclusive sub-types of PRC2 complexes.

Figure 5.6: Consequences of knockout of PRC2 subunits on H3K27 methylation.

Figure 5.7: Loss of Pali1/2 or Aebp2 has opposing effects on PRC2 activity *in vitro* and *in vivo*.

Figure 5.8: The balance of PRC2.1 and PRC2.2 activities is essential for the proper regulation of Polycomb target genes during ESC differentiation.

Figure 5.9: Pali1/2 and Aebp2 competition model.

Summary

The Polycomb Repressive Complex 2 (PRC2) is composed of the core subunits SUZ12, EED, RBBP4/7 and EZH1/2, which together are responsible for all di- and tri- methylation of lysine 27 on Histone H3 (H3K27me_{2/3}). In addition, several highly conserved sub-stoichiometric accessory proteins associate with PRC2, including AEBP2, JARID2, EPOP and PCL1-3. However, little is known about the division of labour amongst these cofactors or how they confer the diverse roles of the PRC2 complex. Here we report the discovery of a new family of vertebrate specific PRC2 associated proteins; ‘Polycomb associated LCOR isoform 1’ (PALI1) and PALI2, which are novel isoforms of the LCOR and LCORL genes, respectively. PALI1 promotes PRC2 methyltransferase activity *in vitro* and shares a unique PRC2 interaction domain with PALI2 that emerged during vertebrate evolution. The PALI proteins define a unique subset of PRC2 complexes that lack AEBP2 and JARID2. We uncover a biochemical and genetic antagonism between the Pali1/2 proteins and Aebp2 in mouse embryonic stem cells, consistent with an essential role for Pali1/2 during differentiation. In summary, we identify a new family of vertebrate specific PRC2 proteins and reveal an unexpected antagonism between sub-stoichiometric subunits of the PRC2 complex suggesting that alternative complexes may have opposing effects on transcriptional regulation.

Abbreviations

Aebp2-AE binding protein

Apc-Anaphase promoting complex

Aut2-Autism Susceptibility Candidate 2

Bcor-BCL6 Co-repressor

Bcorl1-BCL6 Co-repressor Like 1

Bmi1-B-cell-specific Moloney murine leukaemia virus integration site 1

C10ORF12-Chromosome 10 open reading frame 12

Cbfb-Core Binding Factor β

Cbx-Chromobox

Ccna2-Cyclin A2

Cdkn2a-Cyclin Dependent Kinase inhibitor 2A

cDNA-complementary DNA

ChIP-Chromatin immunoprecipitation

CpG-Cytosine-phosphate-Guanine

cPRC1-canonical Polycomb Repressive Complex 1

CRISPR-Clustered Regularly Interspaced Short Palindromic Repeats

Csnk2a-Casein Kinase 2A

Csnk2b-Casein Kinase 2B

Ctbp-C-terminal binding protein

E2f6-E2f transcription factor 6

E3 ligase-Ubiquitin Ligase

Eed-Embryonic Ectoderm Development

Epop-Elongin BC and Polycomb Repressive Complex 2 associated protein

ESC-Embryonic Stem Cell

Ezh1/2-Enhancer of Zeste 1/2

Fbrs-Fibrosin

Fbrsl1-Fibrosin Like 1

H2A-Histone 2A

H2AK119ub1-Histone 2A lysine 119 mono-ubiquitination

H3-Histone 3

Hdac-Histone Deacetylase

HEK293T-Human Embryonic Kidney cells (with SV40 T antigen)

HMEC-Human Mammary Epithelial Cells

Hox genes-Homeobox genes

IP-immunoprecipitation

Ink4a-Inhibitor of cyclin dependent Kinase 4

Kdm2b-Lysine demethylase 2b

LC-MS/MS-Liquid chromatography mass spectrometry

Lcor-Ligand dependent Co-Repressor

Lcorl-Ligand dependent Co-Repressor like

me1-mono-methylation

me2-di-methylation

me3-tri-methylation

Mga-Max gene associated protein

mRNA-messenger RNA

ncPRC1- non-canonical Polycomb Repressive Complex 1

NR-nuclear receptor

ORF-open reading frame

PALI1-Polycomb associated *LCOR* isoform 1

Pcgr/ Psc-Polycomb Group Ring Finger/ Posterior Sex Combs

PCR-polymerase chain reaction

Ph/ Phc-Polyhomeotic

Pho-Pleiohomeotic

PIP domain-Pali1 interaction with PRC2 domain

PRC1-Polycomb Repressive Complex 1

PRC2-Polycomb Repressive Complex 2
PRE-Polycomb Response Element
RAWUL-Ring-finger And WD40 associated Ubiquitin Like domain
Rest-RE1 silencing transcription factor
Ring1A/B-Really Interesting New Gene 1A/B
RT-qPCR-Reverse Transcription quantitative PCR
Runx1-Runt-related transcription factor 1
Rybp-Ring1 and Yy1 binding protein
Skp1-S-Phase Kinase Associated Protein 1
Suz12-Suppressor of Zeste 12 homolog
Usp7-Ubiquitin Specific Peptidase 7
Yaf2-Yy1 Associated Factor 2
Zfp277-Zinc Finger Protein 277

Publications

Brien, G.L., Healy, E., Jerman, E., **Conway, E.**, Fadda, E., O'Donovan, D., Krivtsov, A.V., Rice, A.M., Kearney, C.J., Flaus, A., *et al.* (2015). A chromatin-independent role of Polycomb-like 1 to stabilize p53 and promote cellular quiescence. *Genes Dev* 29, 2231-2243.

Conway, E*, Healy, E*, and Bracken, A.P. (2015). PRC2 mediated H3K27 methylations in cellular identity and cancer. *Curr Opin Cell Biol* 37, 42-48.

Conway, E. and Bracken, A.P. (2017) Unraveling the Roles of Canonical and Non-canonical PRC1 complexes. *Polycomb Group Proteins* 1, 57-80.

Conway, E*, Jerman, E*, Healy, E., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D.J., Watson, A., Rice, A.M, Mucha, M., Pratt-Kelly, I., Koseki, Y., Nakayama, M., Ishikura, T., Streubel, G., Wynne, K., Hokamp, K., McLysaght, A., Di Croce, L., Margueron, R., Cagney, G., Koseki, H., Bracken, A.P. (2017) A new family of vertebrate-specific Polycombs encoded by the LCOR/LCORL genes balance PRC2 subtype activities. *Molecular Cell*, 70, 408-421.

Streubel, G*, Watson, A*, Jamulla, S.G., Oliviero, G., Luca Negri, G., Dillon, E., Wynne, K., **Conway, E.**, McCole, R., Glancy, E., Pasini, D., Krogan, N.J., Bracken, A.P., Cagney, G. (2017) The H3K36 methyltransferase Nsd1 demarcates PRC2-mediated H3K27me2 and H3K27me3 domains in Embryonic Stem Cells. *Molecular Cell*, 70, 371-379.

*Joint first author

Chapter 1: Introduction

1.1 General introduction: Transcriptional memory

Transcriptional regulation is the essential process of controlling which genes are actively transcribed into RNA, which is, often, subsequently translated into proteins. Careful control of transcription allows all organisms, both single and multi-cellular, to carry out a wide variety of vital processes; from adapting to external stresses (Kops et al., 2002), control of metabolism (Jacob and Monod, 1961) to the regulation of mating type of yeast (Haber, 2012). Differentiation is a complex, multi-dimensional process that is tightly regulated by precise transcriptional regulation (Moris et al., 2016).

The capacity of multi-cellular organisms to form different tissues with a variety of functions, shapes and sizes from a nearly identical genetic code is a highly complex process. A key concept in differentiation is that not all genes are actively transcribed in all cells. Each cell type has its own specific transcriptional signature, for example a stem cell will not have any tissue specific gene expression but will have high expression of stem cell associated transcription factors. Differentiation occurs when a progenitor or stem cell receives a stimulus to change its function/ shape and size to fulfil a particular role in a tissue (Moris et al., 2016). Genes that are associated with maintaining the progenitor or stem cell population will become repressed and certain tissue specific genes (i.e *MyoD* in muscle cells) will become activated (Bracken and Helin, 2009).

Transcriptional memory is a vital part of this differentiation process as it helps to maintain stem cell and tissue homeostasis. Transcriptional memory is the inheritance of either active or inactive transcription of particular sets of genes following cell division (Alabert et al., 2015). For instance, in a differentiated muscle

cell, the *MyoD* gene and other muscle specific genes are transcriptionally active. After cell division the two daughter cells need to restore the transcriptional profile that had been established in the parent cell, in this case muscle specific genes like *MyoD* will be active in the daughter cells, while stem cell specific genes such as *Nanog* will be repressed (Hansen et al., 2008). This process of transcriptional memory is tightly regulated through a number of mechanisms. Broadly speaking, chemical modifications of DNA, or of the histone proteins around which DNA is wrapped, can dictate whether a gene is capable of being activated or repressed (Jenuwein and Allis, 2001). Although how exactly histone modifications facilitate transcription or repression is still unclear, it has been shown that some of them can contribute to long and short range chromatin looping interactions (Isono et al., 2013; Kundu et al., 2017; Lau et al., 2017). Importantly, some of these DNA and histone modifications are heritable during mitotic cell division, thereby allowing transcriptional memory between parent and daughter cells (Alabert et al., 2015; Hansen et al., 2008).

1.2 Chromatin and histone post-translation modifications.

In eukaryotes, DNA is organised into chromatin, which is composed of several histone proteins assembled into nucleosomes. Each nucleosome is comprised of a core histone octamer, in which there are two copies each of Histone 2A, Histone 2B, Histone 3 and Histone 4 around which 147 base pairs of DNA are wrapped (Figure 1.1A) (Luger et al., 1997). Histone 1 binds to 25bp of linker DNA each side of the nucleosome and is essential for higher order chromatin structures (Bednar et al., 2017). These nucleosomes, which are embedded throughout the genome, have been described as “beads on a string” (Olins et al., 1976; Woodcock et al.,

1976). Outside of this fundamental subunit, the nucleosomes, depending on chromatin modifying proteins, are capable of compacting DNA and regulating chromatin looping. This allows for either transcriptional repression (heterochromatin), or alternatively the nucleosomes can form a loose conformation that is conducive to active transcription (euchromatin) (Kieffer-Kwon et al., 2017).

The histone proteins all possess hydrophobic C termini which are buried in the interaction surface of the nucleosome octamer. Importantly, they each also have hydrophilic N-terminal tails which protrude out from the octamer allowing them to be post-translationally modified (Luger et al., 1997). These post-translational modifications form the basis of what was initially termed the histone code hypothesis (Figure 1.1A) (Jenuwein and Allis, 2001). This hypothesis suggests that the transcriptional state of a particular gene can be predicted by the covalently modified residues on the surrounding histone tails (Jenuwein and Allis, 2001). For example, Histone 3 can be acetylated at lysine 27 on its N terminal tail by the p300 histone acetyltransferase (Figure 1.1A) (Ogryzko et al., 1996). In general, acetylated histone modifications are associated with open regions of chromatin such as those involved in transcriptional activation or newly replicated chromatin, as the negatively charged acetyl group renders the histone tails with a slight negative charge, resulting in repulsion of the negatively charged phosphate backbone of DNA (Alabert et al., 2015; Kouzarides, 2007). While this hypothesis is broadly accurate it is not as well defined as a 'code', such as how specific triplet codons code for an amino acid (Crick, 1968). For instance there is an ambiguously high frequency of co-occurrence of both the active and repressive histone modifications H3K4me3 and H3K27me3 at repressed developmental promoters in

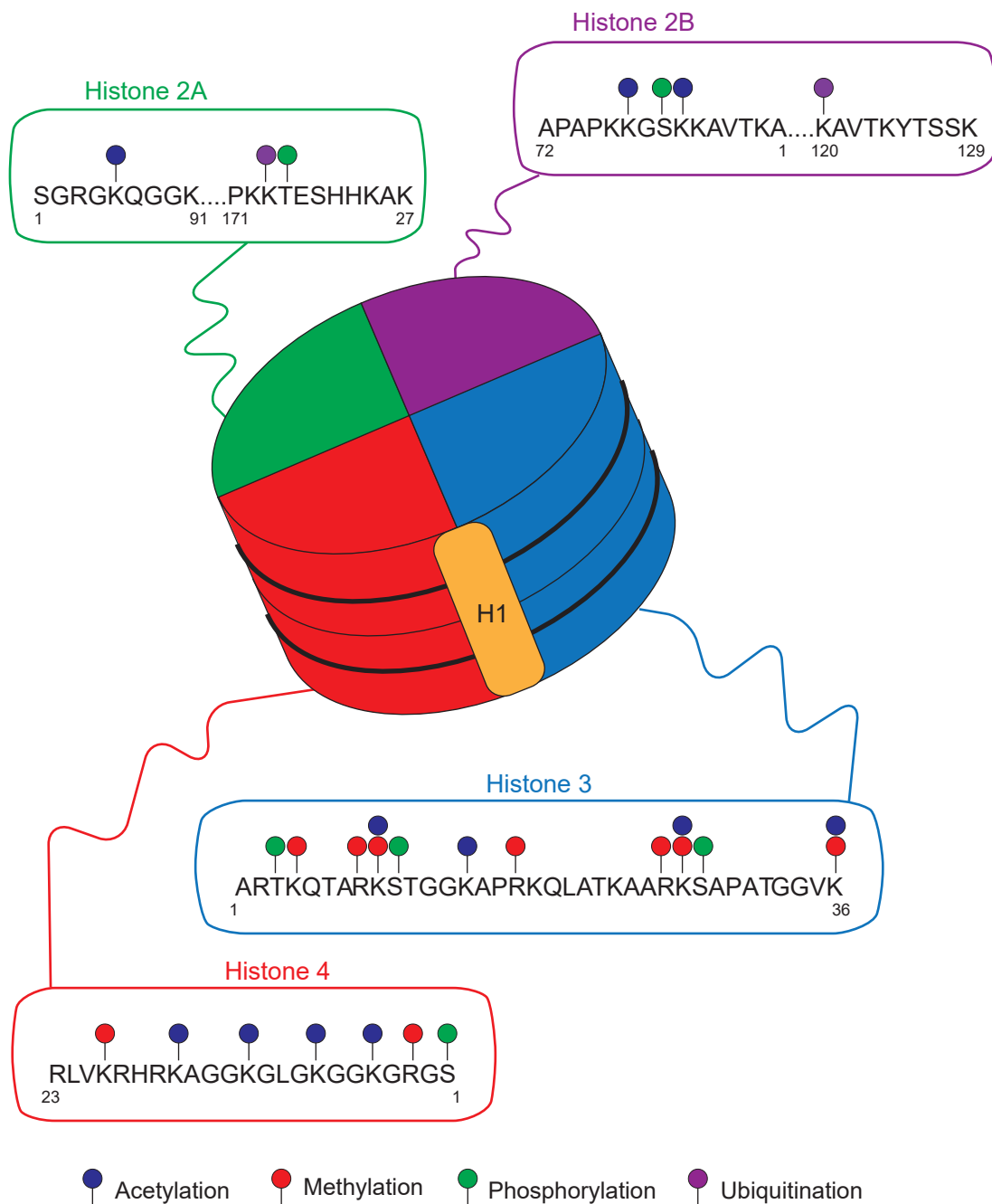


Figure 1.1. The nucleosome and histone post-translational modifications.

Illustration of the core subunit of chromatin, the nucleosome. The nucleosome octamer consists of two copies each of H2A, H2B, H3 and H4. Histone 1 (yellow) links the DNA that is wrapped around the octamer. The protruding Histone tails can be subject to a wide range of covalent post-translational modifications, some of which are depicted here.

embryonic stem cells (Azuara et al., 2006; Bernstein et al., 2006). These bivalent genes demonstrate that neither modification is completely definitive at predicting transcriptional status of a particular gene.

There is a wide range of modifications that occur on histones tails, including acetylation, methylation, ubiquitination and phosphorylation (Figure 1.1A) (Kouzarides, 2007). These post-translational modifications (PTMs) or “marks” confer effects on gene expression at their associated genomic sites. For example, H3K4me3 is a mark associated with the promoters of actively transcribed genes, while H3K36me3 is commonly found along the gene body of actively transcribed genes (Kouzarides, 2007). The localisation of the histone PTMs is not limited to promoters and gene bodies alone; they can even regulate the condensation of large regions of chromatin. Supporting this, in quiescent cells most transcriptional activity is ceased, coordinated in part by the di-methylation of H4K20 (Everitts et al., 2013). Regions of chromatin featuring this mark are tightly packed into repressive chromatin conformations to prevent any latent transcriptional activity (Kieffer-Kwon et al., 2017). Similarly, centromeric chromatin is another example of a region of DNA that needs to be transcriptionally silenced; despite not containing many genes. Failure to correctly silence centromeres can result in mitotic defects (Hill and Bloom, 1987). Therefore, these centromeres are marked with H3K9me3 (Nakayama et al., 2001), a histone mark found at constitutive heterochromatin. This mark is catalysed by the Suv39H1 enzyme, which was one of the first histone methyltransferases to be discovered (Rea et al., 2000).

1.3 Readers, writers and erasers and their role in transcriptional memory

Histone post-translational modifications are catalysed by different classes of enzymes. The first histone modification enzyme to be discovered was Suv39H1, which, as described above, catalyses H3K9me3 (Figure 1.2A) (Nakayama et al., 2001; Rea et al., 2000). Other examples of histone modifying enzymes include EZH2, which catalyses H3K27me3, and p300, which mediates H3K27ac (Kouzarides, 2007). Collectively, these enzymes are referred to as “writer” enzymes (Figure 1.2A). Another group of chromatin regulatory proteins called “readers” have conserved reader domains such as chromodomains, bromodomains, tudor or PHD domains (Musselman et al., 2012b). In the case of Suv39H1 mediated H3K9me3, it is bound specifically by the HP1 γ protein that reads H3K9me3 via its chromodomain (Figure 1.2A) (Bannister et al., 2001; Lachner et al., 2001), which leads to chromatin compaction and condensation (Cheutin et al., 2003). Remarkably other chromodomain containing proteins such as CBX8 read other methylated lysines, such as H3K27me3 in the case of CBX8, with high specificity and can then recruit associated effector complexes which subsequently alter gene expression (Fischle et al., 2003; Min et al., 2003). The “writing” of histone post-translational modifications is not thought to be permanent, since most are reversible through specific “eraser” enzymes, such as the histone demethylases, for instance UTX and JMJD2A, and the histone deacetylases, such as HDAC1/2 (Klose et al., 2006; Kouzarides, 2007). The UTX enzyme removes methyl groups from H3K27me3 (Figure 1.2A) (Agger et al., 2007), this process is thought to confer the activation of genes, such as Hox genes, during differentiation (Agger et al., 2007). The histone deacetylase (HDACs) family form a part of variety

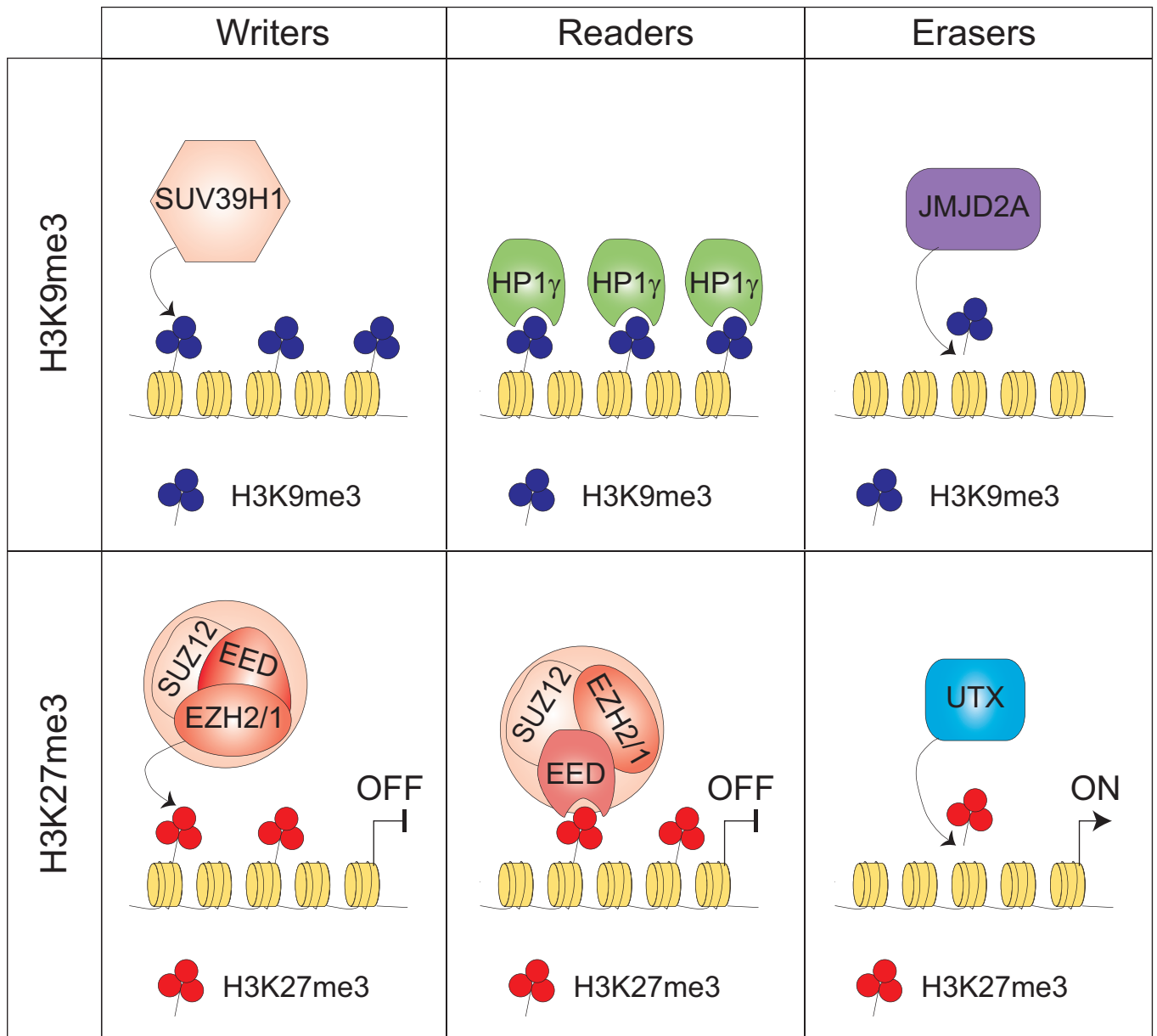


Figure 1.2 Readers, writers and erasers of histone post-translational modifications.

Illustration of the histone modification “writing” of the SUV39H1/EZH2 enzymes (left), the “reading” function of HP1 γ /EED proteins (middle) and the demethylase “eraser” function of the JMJD2A/UTX enzymes (right).

of multi-protein chromatin regulatory complexes, such as NuRD and Sin3a (Heinzel et al., 1997; Laherty et al., 1997; Xue et al., 1998).

The Histone modifications can in some cases be heritable during mitotic cell division, thereby providing a mechanism for the transmission of transcriptional memory from one cell to its daughter cells. During the process of DNA replication the existing histone octamers, and their accompanying histone modifications are diluted out by 50% to each new DNA strand (Alabert et al., 2015). However, the histone modification complexes have diverse mechanisms for the re-establishment of some of the histone modifications which have been diluted. New H3 and H4 histones that are incorporated into the daughter strand nucleosomes of newly replicated DNA are rich in acetyl modifications at various residues (Alabert et al., 2015). These acetyl residues, associated with newly synthesised histones, which are found on the *de novo* histone octamers, are removed rapidly after incorporation into the nucleosome (Alabert et al., 2015). Concurrently, histone writers such as the Suv39H1 complex are enriched on nascent DNA (Alabert et al., 2014). The Suv39H1 containing complex contains the reader HP1 γ which recruits the complex back to heterochromatin sites, where the mature histone octamers are still modified with H3K9me₃, to restore the mark on the surrounding *de novo* incorporated histones (Alabert et al., 2015; Bannister et al., 2001; Fritsch et al., 2010). A further example of the restoration of histone post-translation modification on *de novo* incorporated histones is the post-mitotic re-establishment of the H3K27me₃ mark (Hansen et al., 2008). This modification is mediated by the Polycomb Repressive Complex 2 (PRC2), which is also enriched at the replication fork (Alabert et al., 2014; Hansen et al., 2008). Among the many PRC2 subunits, the complex contains

the reader protein EED, which is capable of specifically binding the H3K27me3 modification through its aromatic cage (Margueron et al., 2009). This provides a mechanism through which the complex can be recruited back to sites with H3K27me3 modified mature histones to catalyse its mark on the *de novo* histones on nascent chromatin (Alabert et al., 2015; Alabert et al., 2014).

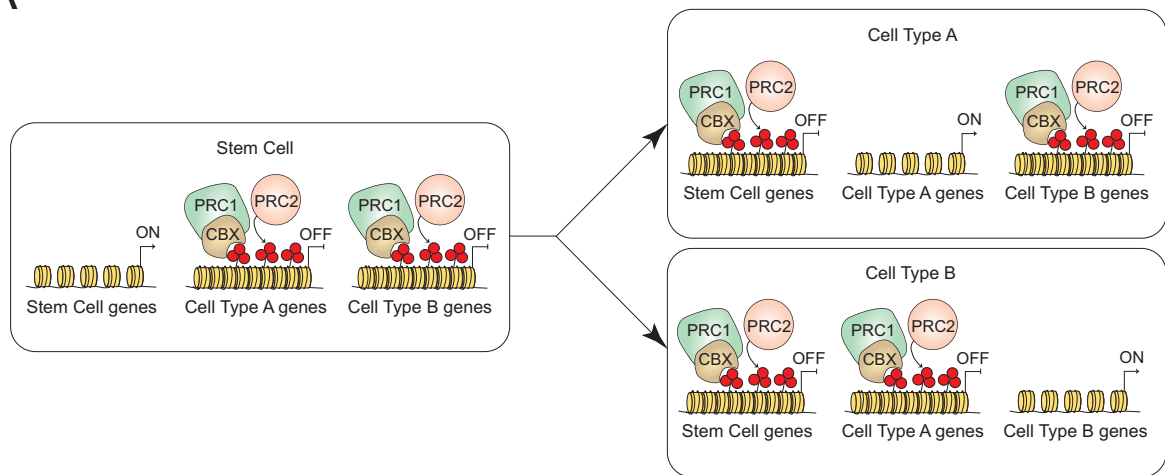
1.4 Polycomb group proteins

Polycombs were first discovered through mutagenesis studies in *Drosophila*, which identified the *Polycomb* (*Pc*) gene as being an essential regulator of anterior-posterior axis segmentation during embryogenesis (Lewis, 1978). In this study, and many later studies in *Drosophila* and mouse, loss of function of genes encoding Polycomb group proteins leads to homeotic transformations. These transformations take the form of anteriorisation of body segments and features such as sex combs and vertebrae (Adler et al., 1991; Akasaka et al., 1996; Akasaka et al., 2001; Dura et al., 1987; Isono et al., 2005). The cause of these transformations is a failure of the developing embryo to correctly maintain repression of *Hox* genes in segments of the embryo in which they are not normally repressed. The precise regulation of these *Hox* transcription factors is essential given their important role in specification of body segments along the anterior-posterior axis (McGinnis and Krumlauf, 1992). Genetic knockout of several Polycomb group proteins in multiple studies has shown that their loss also leads to embryonic lethality, demonstrating their essential role during development (Adler et al., 1991; Akasaka et al., 2001; Faust et al., 1995; Pasini et al., 2004).

The Polycomb group proteins function primarily as two distinct multi-subunit chromatin regulatory complexes, the Polycomb Repressive Complex 1 (PRC1),

and Polycomb Repressive Complex 2 (PRC2) (Cao et al., 2005; Cao et al., 2002). The mammalian PRC2 core complex is made up primarily of a enzymatic subunit Ezh1 or Ezh2 along with essential subunits Suz12, Eed and Rbbp4/7 (Margueron and Reinberg, 2011). The PRC2 complex is responsible for all H3K27 di and tri-methylation of the cell; these histone modifications are associated with transcriptional repression (Bracken et al., 2006; Ferrari et al., 2014; Lee et al., 2015). Both the PRC1 and PRC2 complexes are essential regulators of cell fate as they maintain the transcriptional repression of developmental associated genes during differentiation (Figure 1.3A) For example, in a stem cell they maintain repression of differentiation associated genes, but upon differentiation to cell type A (for example a neuron) PRC1 and PRC2 leave the promoter of neuronal specific genes, facilitating their activation by other factors, and instead occupy the promoters of stem cell associated transcription factors to maintain their repression (Figure 1.3A). It is unclear whether the PRC2 complex is instructive in turning off transcription of its targets genes, or whether it simply maintains their repression in the absence of an activating signal. It seems more likely that its main role is in the maintenance of repression as upon the deletion of Suz12 or Ezh2 in ESCs there is little to no activation of PRC2 target genes. However, when the ESCs are induced to differentiate, then the PRC2 targets can become aberrantly activated (Riising et al., 2014). In fact, it has been demonstrated that the absence of transcription alone was sufficient to recruit PRC2 to unmethylated CpG islands (Riising et al., 2014), suggesting that PRC2 occupies its target genes after the onset of repression to maintain that state rather than actively shutting off transcription itself.

A



B

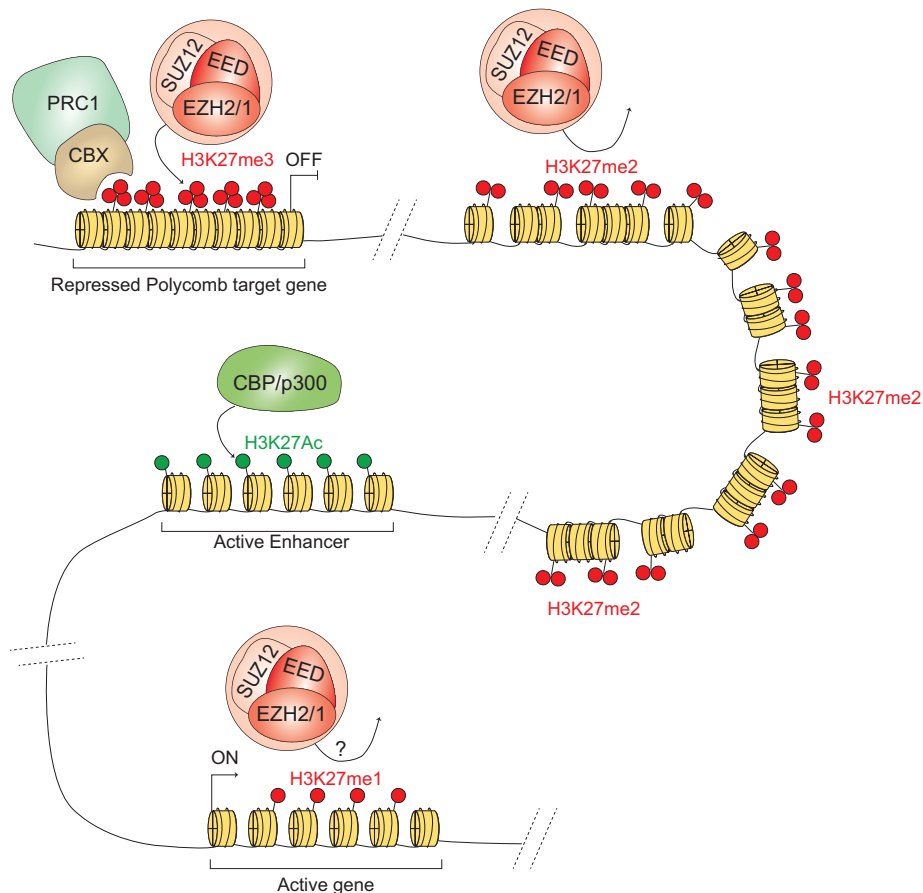


Figure 1.3 The role of Polycomb group proteins in cell fate decisions.

(A) Left. In pluripotent stem cells stem cell genes are active; as such their promoters have a loose chromatin structure that facilitates active transcription. However, tissue/lineage specific genes such as muscle or neural transcription factors are not actively transcribed. The repression of these genes is maintained by the enzymatic activities of the PRC1 and PRC2 complexes. Right. In differentiated tissues stem cell genes are silenced, this silencing is maintained by the PRC1 and PRC2 complexes, along with alternative lineage specific genes. For example in cell type A (muscle) cell type B associated transcription factors genes are silenced by PRC1 and PRC2 and vice versa. Figure adapted from Conway et al, 2015.

(B) Illustration of the H3K27 modifications (H3K27me1/2/3 and H3K27ac). This graphic illustrates the enzymes responsible for each modification, and where each modification is predominantly located in the genome. Figure adapted from Conway et al, 2015.

The PRC2 complex mediated H3K27me2 seems to be a “default” modification found on histone octamers on nascent DNA shortly after replication (Alabert et al., 2015), and is highly enriched at intergenic regions and inactive enhancers (Figure 1.3B) (Ferrari et al., 2014; Lee et al., 2015). On the other hand, H3K27me3 is a histone modification that is primarily enriched at promoters of developmentally repressed genes (Figure 1.3B) (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006). There is also some evidence that PRC2 regulates H3K27me1, a modification that is found on the gene bodies of actively transcribed genes (Figure 1.3B) (Ferrari et al., 2014; Hojfeldt et al., 2018). PRC2 histone methyltransferase kinetics are interesting in this context as the complex rapidly catalyses H3K27me1, while H3K27me2 and H3K27me3 are accumulated slower (Alabert et al., 2015; Hojfeldt et al., 2018; McCabe et al., 2012a). These kinetics are consistent with PRC2 being stably tethered at unmethylated CpG islands of its target genes allowing the complex sufficient time to catalyse H3K27me3, however PRC2 is not stably bound at H3K27me1 and H3K27me2 sites suggesting a ‘hit and run’ mechanism for H3K27 methyltransferase activity at these sites. In addition to the core complex the PRC2 core complex also has a number of accessory substoichiometric subunits including Jarid2, Aebp2, Pcl1/2/3 (aka Phf1, Mtf2 and Phf19 respectively) and Epop which play varying roles in PRC2 complex function (Holoch and Margueron, 2017).

The PRC1 complexes contain a core heterodimer of one of six Pcgf proteins (Pcgf1-6) and one of either Ring1A or Ring1B (Cao et al., 2005; Gao et al., 2012). This Pcgf-Ring heterodimer represents the E3 ubiquitin ligase core of PRC1 and its primary substrate is Histone 2A at lysine 119 (residue 118 in *Drosophila*) (Cao et

al., 2005; McGinty et al., 2014; Pengelly et al., 2015; Wang et al., 2004a). Like PRC2 mediated H3K27me3, PRC1 mediated mono-ubiquitinated H2AK119 is a histone modification that is enriched at promoters of developmentally repressed genes, together with the PRC1 and PRC2 complexes (Morey et al., 2013; Rose et al., 2016; Wu et al., 2013). Variant combinations of PRC1 complexes exist, each defined by a different Pcgf subunit with a distinct mechanism of action, depending on its complex members (Junco et al., 2013). The six Pcgf paralogs share a highly conserved Ring domain which allows them to dimerise with Ring1A or Ring1B, while the RAWUL (Ring-finger and WD40 associated ubiquitin-like) domain of each Pcgf protein has diverged throughout vertebrate evolution to the extent that each can form a unique PRC1 sub-type (Conway and Bracken, 2017; Junco et al., 2013). Therefore the RAWUL domain of each Pcgf protein present in a particular complex provides the specificity for the remaining composition of that PRC1 complex (Figure 1.4A). Generally speaking, these variant combinations of PRC1 can be divided into two separate complexes, canonical and non-canonical. The main function of the canonical PRC1 complex is to physically condense chromatin into repressive Polycomb bodies (Isono et al., 2013), while the primary function of the non-canonical PRC1 complex is linked to its ability to mediate H2AK119ub1 (Blackledge et al., 2014; Endoh et al., 2017; Rose et al., 2016).

1.5 Canonical and non-canonical PRC1 complexes

The core heterodimer of canonical PRC1 (cPRC1) complex consists of either Pcgf2 or Pcgf4 (also known as Bmi1) and Ring1A or Ring1B. Pcgf2 and Bmi1 are the most conserved within their RAWUL domains (Junco et al., 2013), and as such, they share most of their interacting partners, including the Phc (Polyhomeotic) and

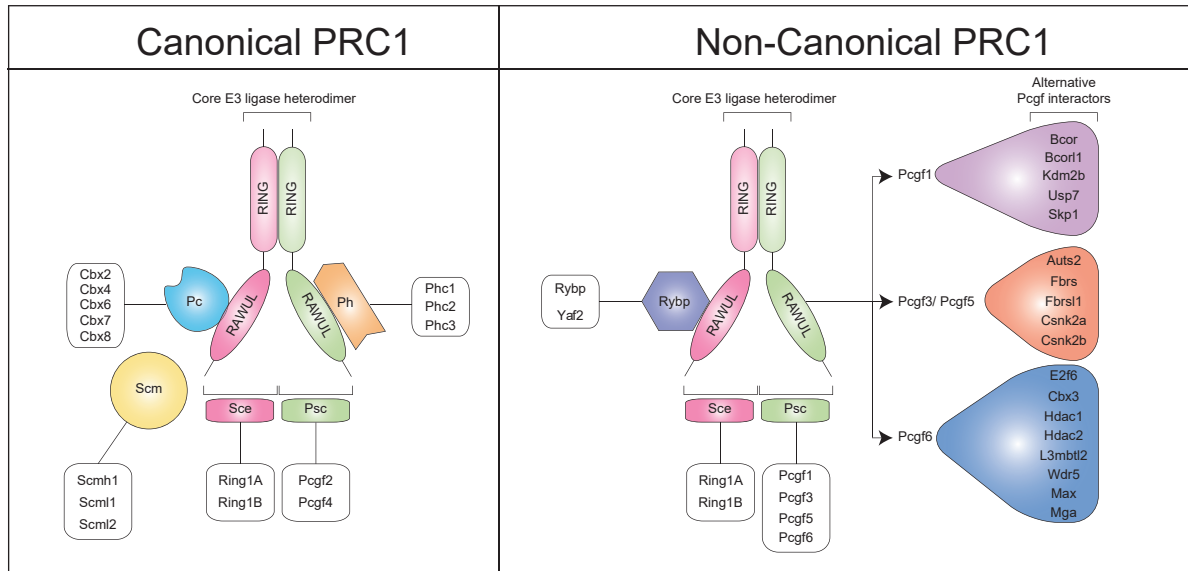
Cbx (Chromobox) subunits, each of which has multiple paralogs, Phc1/2/3 and Cbx2/4/6/7/8, respectively (Figure 1.4A) (Gao et al., 2012; Hauri et al., 2016). The unique features of the Phc and Cbx subunits are what make the cPRC1 and ncPRC1 complexes differ so drastically in their mechanism of action. The Cbx subunit is capable of reading H3K27me3 via its chromodomain, and thereby recruiting the cPRC1 complex to chromatin sites marked by PRC2 mediated H3K27me3 (Fischle et al., 2003; Min et al., 2003), leading to the co-localisation of the PRC2 complex and the cPRC1 complex (Bracken et al., 2006).

The Phc subunits of cPRC1 complexes contains a sterile alpha motif (SAM) that facilitates oligomerisation with Phc subunits in other nearby cPRC1 complexes on the promoters of neighbouring and distal genes (Isono et al., 2013). This oligomerisation allows Phc subunits at different promoters to create loops of promoter-promoter contacts that help maintain repression by grouping Polycomb marked genes into repressive Polycomb bodies (Boettiger et al., 2016; Francis et al., 2004; Kundu et al., 2017; Schoenfelder et al., 2015). The way this condensation of chromatin leads to gene repression is not yet fully understood but it is possibly mediated simply by steric hindrance of RNA polymerase II, supporting this STORM (stochastic optical reconstruction microscopy) studies have shown that Polycomb bodies are among the most condensed regions of chromatin in the genome (Boettiger et al., 2016). The cPRC1 complex does not seem to rely on its ubiquitination activity to form Polycomb bodies (Eskeland et al., 2010), particularly as it seems to have greatly weaker catalytic activity than ncPRC1 (Blackledge et al., 2014; Rose et al., 2016).

The loss of cPRC1 complex subunits in both *Drosophila* and mouse leads to early embryonic lethality and derepression of *Hox* genes. The derepression of *Hox* genes in these mutants is consistent with the classical model of Polycomb mediated repression of *Hox* genes (Lewis, 1978). This classical Polycomb phenotype is seen in cPRC1 specific subunit knockouts, for example, in *dPh*, *mPhc1/2* or *mPcgf2/Bmi1* null organisms (Akasaka et al., 2001; Dura et al., 1987; Isono et al., 2005). In contrast, the non-canonical PRC1 (ncPRC1) complex has emerging active and repressive roles in the control of transcription and, intriguingly, seems to have both shared and unique target genes with PRC2 and cPRC1 (Endoh et al., 2017; Morey et al., 2013). Loss of function of non-canonical PRC1 specific subunits, such as mRybp and dKdm2, also leads to early embryonic lethal phenotypes in both mouse and *Drosophila*. However, these phenotypes are independent of changes in *Hox* gene expression (Gonzalez et al., 2008; Lagarou et al., 2008; Purity et al., 2005). This suggests that although the cPRC1 complex is essential for maintaining the spatio-temporal pattern of *Hox* gene expression, the ncPRC1 complex is not.

The non-canonical PRC1 complex can feature any of the six Pcgf proteins together with Ring1A or Ring1B (Gao et al., 2012). They are distinguished from cPRC1 by their lack of Phc or Cbx subunits, instead they contain one of two paralogous proteins, Rybp or Yaf2 (Figure 1.4A) (Gao et al., 2012). These subunits increase the H2AK119ub1 activity of the ncPRC1 complex through co-ordination of conformational changes of the core heterodimer (Rose et al., 2016). The additional ncPRC1 complex composition is determined by the RAWUL domain of the particular Pcgf protein present (Figure 1.4A) (Junco et al., 2013). For example,

A



B

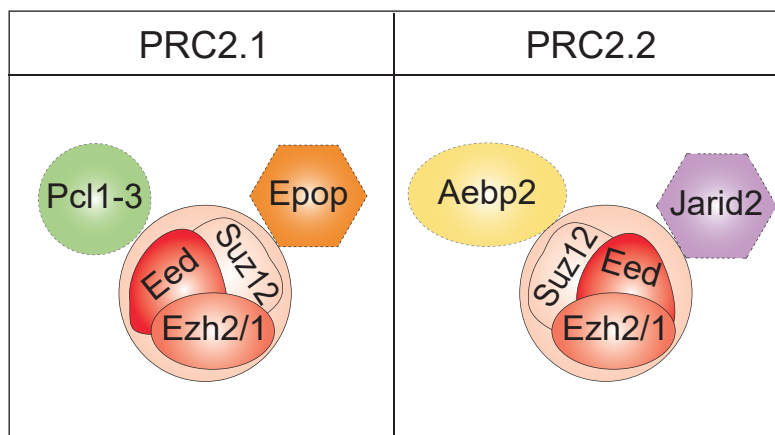


Figure 1.4 The Variable compositions of the PRC1 and PRC2 complexes.

(A) Depiction of the variable composition of the canonical and non-canonical PRC1 complexes, along with the names of their *Drosophila* orthologs, as determined by the binding partners associated with the RAWUL domain of each Pcgf protein. Figure adapted from Conway and Bracken, 2017.

(B) Depiction of the newly defined sub-types (PRC2.1 and PRC2.2) of PRC2 complex composition.

Pcgf1-PRC1 complexes contain Kdm2B, Bcor, Bcorl1, Skp1 and Usp7 (Farcas et al., 2012), while Pcgf3/5-PRC1 consists of Auts2, Csnk2A/B, Fbrs and Fbrsl1 (Gao et al., 2014), while Pcgf6-PRC1 contains E2f6, Max, Mga, Wdr5, Hdac1/2 and L3mbtl2 (Endoh et al., 2017). These divergent ncPRC1 compositions are thought to determine the more specialised mechanisms of action of each non-canonical PRC1 complex. The Pcgf1 and Pcgf6 ncPRC1 complexes have been primarily linked with repression of Polycomb target genes through H2AK119ub1 activity (Blackledge et al., 2014; Endoh et al., 2017; Wu et al., 2013). However, less is known about the mechanism of action of the Pcgf3 and Pcgf5 containing ncPRC1 complexes. The studies to date suggest that they may have a role in transcriptional activation, despite their ability to mediate H2AK119ub1 (Blackledge et al., 2014; Gao et al., 2014).

Since non-canonical PRC1 complexes lack Cbx and Phc subunits, their ability to regulate transcription cannot occur through the same chromatin condensation mechanism as for the canonical PRC1 complexes. It appears that each type of non-canonical PRC1 complex has a unique mechanism of recruitment to chromatin. For example, the Pcgf1-PRC1 complex is targeted to chromatin via its Kdm2b subunit (Farcas et al., 2012; Wu et al., 2013). The Kdm2b protein binds to unmethylated CpG islands throughout the genome, primarily present at promoters (Blackledge et al., 2014; Wu et al., 2013). It recognises unmodified CpG islands through its CxxC motif which, together with its Pcgf1 interacting Leucine Rich Repeat (LRR), are essential for the association of Pcgf1-PRC1 to chromatin (Blackledge et al., 2014; Wu et al., 2013). The absence of Pcgf1-PRC1 at its target

genes, in Kdm2b depleted cells, leads to a global reduction in H2AK119ub1 in mouse ESCs (Blackledge et al., 2014).

Pcgf3 and Pcgf5 are highly conserved in their RAWUL domains (Junco et al., 2013) and as a result Pcgf3-PRC1 and Pcgf5-PRC1 share accessory subunits, such as Aust2 and Fbrs (Gao et al., 2014; Gao et al., 2012). The mechanism of action of Pcgf3/5-PRC1 complexes is not yet understood but it has been reported that Pcgf5 can act as a transcriptional activator (Gao et al., 2014). The final ncPRC1 variant, Pcgf6-PRC1, has recently been shown to be involved in transcriptional repression in mouse ES cells. It shares many target genes with the cPRC1 complex, but also has many unique targets (Endoh et al., 2017). Pcgf6-PRC1 is reported to be recruited to these sites by two of its associated proteins, Max and Mga (Endoh et al., 2017). These two proteins form a heterodimeric transcription factor complex that allows Pcgf6-PRC1 to get to its target genes and mediate its repressive activity (Endoh et al., 2017).

The mechanism by which H2AK119ub1 leads to transcriptional repression has not yet been fully elucidated. A number of studies over the last few years have reported a role for H2AK119ub1 in recruiting the PRC2 complex (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014). For example, artificial recruitment of Pcgf1, Pcgf3, Pcgf5 and Pcgf6 to chromatin has been shown to recruit the PRC2 complex in a H2AK119ub1 dependent manner (Blackledge et al., 2014; Endoh et al., 2017). It has been recently reported that Jarid2, a sub-stoichiometric subunit of PRC2, can bind the H2AK119ub1 mark through an ubiquitin interaction motif (UIM) near its N terminus (Cooper et al., 2016). Supporting the potential role for H2AK119ub1 in recruiting the PRC2 complex it has been shown that nucleosomes

featuring the H2AK119ub1 mark stimulate stronger catalytic activity of Jarid2-PRC2 more than unmodified nucleosomes (Kalb et al., 2014). These findings suggest a mechanism in which Jarid2 can read the H2AK119ub1 mark and recruit PRC2 to these sites.

In spite of this, the mechanism of recruitment of all PRC2 complexes to chromatin cannot be explained as simply as through the reading function of Jarid2. For instance, while the deletion of Rybp and Yaf2, components of ncPRC1, leads to a reduction in H2AK119ub1 and H3K27me3, it does not completely deplete H3K27me3 levels, and it has a minimal effect on PRC2 occupancy at its target genes (Rose et al., 2016). Supporting this, studies which deplete cells of either Pcgf1 or Pcgf6 exhibit a modest reduction, but never a complete loss of H3K27me3 at Polycomb target genes (Blackledge et al., 2014; Endoh et al., 2017; Farcas et al., 2012; Wu et al., 2013). In addition, Ring1A/B double knockout intestinal crypts are completely devoid of H2AK119ub1 but do not feature global changes in H3K27me3 levels (Chiacchiera et al., 2016a). Despite this, knock in of a catalytic mutant of Ring1B in mouse ES cells showed a partial loss of H3K27me3 at PRC2 targets, suggesting that the catalytic activity of Ring1B does in fact contribute to promoting H3K27me3 deposition (Illingworth et al., 2015). Taken together, it seems that H2AK119ub1 may boost the catalysis of H3K27me3, but it does not seem to be the sole cause of PRC2 complex recruitment and activity.

However, H2AK119ub1 is vital for development. Supporting this, mouse embryos with a constitutive knock in allele of the Ring1B catalytic mutant do not survive until birth (Illingworth et al., 2015). While the majority of Ring1B mutant embryos survived further into development than Ring1B null embryos, this still delineates an

essential role for the catalytic activity of Ring1B in development, which is independent of its role within canonical PRC1 complexes. In addition, mutagenesis of the ncPRC1 substrate lysines on H2A in *Drosophila* results in pupal lethality demonstrating further the importance of this modification during development (Pengelly et al., 2015). These studies illustrate the important role of ncPRC1 complex activity in regulating gene expression. However, this activity appears to primarily act independently of classic Polycomb target genes such as the *Hox* genes.

1.6 The emerging paradigm of variant PRC2 complexes

An emerging theme is that while the PRC2 complex does not exhibit quite as much diversity in its sub-stoichiometric subunits as the PRC1 complex, different forms of PRC2 complex exist. The sub-stoichiometric subunits of PRC2 include Jarid2, Aebp2, Pcl1/2/3, C10ORF12 and Epop (Holoch and Margueron, 2017). The variations of PRC2 can be divided into two distinct complexes, PRC2.1 and PRC2.2 (Figure 1.4B). The defining characteristic of PRC2.1 is the presence of a Pcl1/2/3 protein, while PRC2.1 can have additional accessory subunits such as C10ORF12/PALI1 or Epop (Alekseyenko et al., 2014; Hauri et al., 2016; Holoch and Margueron, 2017). Jarid2 and Aebp2 are the sub-stoichiometric subunits that make up the PRC2.2 complex. These subunits are mutually exclusive to the Pcl proteins and the other PRC2.1 subunits (Grijzenhout et al., 2016; Hauri et al., 2016; Holoch and Margueron, 2017). Interestingly Jarid2, Pcl2, Pcl3 and Aebp2 all seem to occupy the promoters of PRC2 target genes alongside core complex members and the H3K27me3 mark (Beringer et al., 2016; Brien et al., 2012; Grijzenhout et al., 2016; Liefke et al., 2016). There is no distinctive binding pattern

of the individual sub-stoichiometric subunits that suggests they may have different mechanisms of function or different sets of target genes.

Aside from the aforementioned H2AK119ub1 reading ability of Jarid2, multiple studies have found that Jarid2 is capable of promoting H3K27me3 both *in vivo* and *in vitro* (Cooper et al., 2016; Kalb et al., 2014; Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010a; Peng et al., 2009). Depletion of Jarid2 leads to a reduction in PRC2 complex occupancy and catalytic activity at Polycomb target genes accompanied by impaired differentiation of ESCs (Landeira et al., 2010; Pasini et al., 2010a). A recent study showed that loss of Aebp2 correlated with increases in H3K27me3 and Suz12 levels at PRC2 targets in mouse ESCs (Grijzenhout et al., 2016). These opposing effects of Jarid2 and Aebp2 on H3 methyltransferase activity suggests that, although Aebp2 can occupy the same complex as Jarid2, perhaps catalytically active Jarid2-PRC2 complexes may exist independent of Aebp2 (Grijzenhout et al., 2016).

The Pcl1/2/3 proteins have also been shown to promote H3K27me3 activity *in vitro* and *in vivo*, and have been linked with a role in recruiting the PRC2 complex to its target genes (Ballare et al., 2012; Brien et al., 2012; Li et al., 2011; Sarma et al., 2008). The Pcl proteins read the H3K36me3 modification through their Tudor domain (Brien et al., 2012; Brien et al., 2015; Musselman et al., 2012a). These Tudor domains are essential for the ability of Pcl proteins to read chromatin, and recruit the PRC2.1 complex to its target genes (Brien et al., 2012; Brien et al., 2015). It has also been reported that Pcl proteins are capable of binding to CpG repeats through an extended helix domain (Li et al., 2017). Pcl2 with mutations in this CpG binding domain fails to recruit PRC2 to its target genes (Li et al., 2017),

providing further evidence for a role of Pcl proteins in recruiting PRC2.1 to chromatin.

The Pcl proteins can form a part of two distinct types of PRC2.1 complexes. They can either form a complex with Epop or C10ORF12 (Figure 1.4B) (Alekseyenko et al., 2014; Hauri et al., 2016). These two subtypes of PRC2 are also mutually exclusive since Epop is never in the same complex as C10ORF12, and vice versa (Alekseyenko et al., 2014). Knockout of Epop leads to an increase in H3K27me3 and PRC2 occupancy (Beringer et al., 2016), similarly to Aebp2, despite evidence from *in vitro* assays suggesting that both Aebp2 and Epop can promote methyltransferase activity (Cao and Zhang, 2004; Zhang et al., 2011). These studies contribute to an emerging pattern in which the sub-stoichiometric subunits of PRC2 confer specific functions upon the complex and also an intriguing potential self-regulatory function of PRC2 catalytic activity through its own subunits (Holoch and Margueron, 2017). These discoveries also point to a divergence in the mechanism of PRC2 complex recruitment during vertebrate evolution. Although *Drosophila* share Aebp2, Jarid2 and Pcl orthologs with mammals, they lack Epop and C10ORF12. Interestingly, the *Drosophila* PRC2 complex is targeted to its 'Polycomb response elements' (PRE) by the DNA binding Pho factor (Mohd-Sarip et al., 2005; Muller and Kassis, 2006; Wang et al., 2004b). However, the mammalian ortholog of Pho, Yy1, does not co-localise with PRC2 in mouse embryonic stem cells (Vella et al., 2012).

1.7 The genes encoding PRC2 and Histone 3 are frequently mutated in cancer.

In recent years, the sequencing of cancer genomes has led to the realisation that a large proportion of cancers carry mutations in genes encoding chromatin regulators (Dawson and Kouzarides, 2012). A seminal review by Vogelstein and colleagues highlighted that out of 140 cancer ‘driver’ genes, 30 encoded chromatin regulators, clearly illustrating that a better understanding of their function is a pressing need (Vogelstein et al., 2013). This set of chromatin regulating genes includes well established driver genes such as *MLL2/3* which are part of the Trithorax complex that antagonise Polycomb function through its catalysis of H3K4me3 (Milne et al., 2002). Another prominent driver gene highlighted in this analysis is the *SNF5* tumour suppressor gene (Vogelstein et al., 2013), whose protein product is a component of the BAF chromatin remodelling complex, which is also associated with active transcription and has been reported to have an antagonistic relationship with Polycomb function (Wilson et al., 2010). Deletion of *SNF5* is considered to be the main driver event in malignant rhabdoid tumours (Reisman et al., 2009).

The paradigm in which the regulation of chromatin dynamics and transcription has aberrant function in cancer expands to include dysregulation of the PRC2 complex and its catalytic function. Recurrent mutations of the catalytic SET domain of EZH2 have been reported in several cancer types (Table 1.5) (Bodor et al., 2013; Hodis et al., 2012; Huether et al., 2014) most frequently in diffuse large B cell lymphomas (DLBCL) (Morin et al., 2010; Morin et al., 2011). Some of these SET domain mutations have since been found to be “change of function mutations” which lose the ability to use unmodified H3 as a substrate, but have enhanced catalytic activity

Mutations in PRC2 members and Histone H3 coding genes in cancer			
Gene	Aberration	Cancer type (frequency %)	H3K27 methylation status
EZH2 'change of function' mutations			
EZH2	pTyr641X	Lymphoma (9–24%), parathyroid adenoma (1%), ALL (2%), melanoma (2%)	Elevated H3K27me3 Reduced H3K27me2
	pAla677Gly	Lymphoma (1–2%), Ewing sarcoma (5%)	Elevated H3K27me3 Reduced H3K27me2
	pAla687Val	Lymphoma (1–2%)	Elevated H3K27me3 Reduced H3K27me2
PRC2 loss of function mutations			
EZH2	Homozygous mutation	Leukemia (4%), myeloid disorders (1–3%)	Reduced H3K27me3
	Heterozygous mutation	Leukemia (1%), myeloid disorders (6%)	Reduced H3K27me3
SUZ12	Mutation	MDS/MPN (1–3%), leukemia (2–3%)	Reduced H3K27me3
	Heterozygous deletion	MPNST (15–22%)	Reduced H3K27me3
	Heterozygous deletion and mutation	MPNST (16–26%)	Reduced H3K27me3
	Homozygous deletion	MPNST (12%)	Reduced H3K27me3
EED	Heterozygous deletion and mutation	MPNST (2–5%)	Reduced H3K27me3
	Heterozygous deletion	MPNST (3–14%)	Reduced H3K27me3
	Homozygous deletion	MPNST (10%)	Reduced H3K27me3
AEBP2	Mutation	Leukemia (1%), MPNST (7%)	Reduced H3K27me3
Histone H3 mutations			
H3F3A	pLys27Met mutation	High grade glioma (18–71%), low grade glioma (1–2%), leukemia (1%)	Reduced H3K27me3 Reduced H3K27me2
HIST1H3B	pLys27Met mutation	High grade glioma (3–18%)	Reduced H3K27me3 Reduced H3K27me2

Table 1.5 Mutations in genes encoding PRC2 subunits and Histone H3 in cancer.

Table of reported cancer mutations resulting in EZH2 change of function mutations, PRC2 complex member loss of function mutations and Histone 3 coding gene mutations along with the effect on PRC2 mediated histone modifications. Figure adapted from Conway et al 2015.

on nucleosomes containing H3K27me2. (McCabe et al., 2012a; Sneeringer et al., 2010). Consequently, these cancers have reduced levels of H3K27me2 and higher levels of H3K27me3 (Table 1.5) (Conway et al., 2015; McCabe et al., 2012a; Sneeringer et al., 2010). Several small molecule inhibitors of the catalytic activity of EZH2 (EZH2i) have been developed that have been shown to reduce cancer cell proliferation rates *in vitro* in DLBCLs with SET domain mutations (Knutson et al., 2012; McCabe et al., 2012b; Qi et al., 2012). Interestingly, these EZH2i drugs have also been shown to have efficacy in MRTs as these tumours have aberrantly high levels of PRC2 activity following the loss of SNF5 (Knutson et al., 2013). This exhibits as a proof of principle that inhibition of PRC2 for therapeutic benefit is not limited to tumours harbouring EZH2 SET domain mutations. Currently, EZH2i drugs are being tested in clinical trials in both DLBCL and MRTs (Brien et al., 2016). Aside from these change of function mutations in EZH2, many loss of function mutations in PRC2 complex members have been found in cancer. These mutations include frameshift and nonsense mutation of EZH2 in leukaemias (Ernst et al., 2010; Nikoloski et al., 2010) and recurrent deletion of EED and SUZ12 in malignant peripheral nerve sheath tumours (MPNST) (Table 1.5) (Lee et al., 2014; Zhang et al., 2014). Tumours with these genetic lesions exhibit reduced levels of H3K27me3 (Conway et al., 2015; Lee et al., 2014; Zhang et al., 2014).

Dysregulation of the PRC2 function in cancer also extends to recurrent mutations in Histone 3 at lysine 27 in pediatric diffuse intrinsic pontine gliomas (DIPG) (Table 1.5) (Sturm et al., 2012; Wu et al., 2012). Multiple copies of histone 3 encoding genes exist, and remarkably highly recurrent mutations occur in two of these genes (H3F3A and HIST1H3B) which alter the lysine residue at position 27 to a

methionine (Sturm et al., 2012; Wu et al., 2012). Surprisingly, this lysine to methionine mutation, in just a single copy of a histone 3 coding gene, is sufficient to deplete levels of H3K27me_{2/3} (Lewis et al., 2013). Several studies reported that this lysine to methionine mutation resulted in the sequestering of the PRC2 complex at mutant nucleosomes via the H3K27M residue binding the SET domain of EZH2 (Justin et al., 2016; Lewis et al., 2013). However, this result has recently been challenged by a study that found mutant H3 in fact prevented PRC2 complex occupancy at mutant nucleosome sites (Piunti et al., 2017). While the mechanism of how H3K27M mutations contribute to cancer remains elusive, a comprehensive study recently found that EZH2i drugs could have efficacy in inhibiting H3K27M mutant tumour growth through derepression of the *INK4A* gene (Mohammad et al., 2017).

Despite the promising development and early clinical trial results from EZH2i drugs, two in vitro studies have reported acquired resistance in DLBCL cell lines treated with EZH2i (Baker et al., 2015; Gibaja et al., 2016). These mutations, which occur outside of the SET domain of EZH2, prevent it interacting with the drug through changes in the structure of the protein (Brooun et al., 2016). These “proof of principle” resistance mutations pose a potential problem in EZH2i based cancer therapies (Holohan et al., 2013). Several types of cancer treatment are currently in the form of combination therapies to target multiple pathways or therapeutic targets simultaneously to reduce the risk of drug resistant mutations occurring in one pathway leading to recurrence of tumour growth (Holohan et al., 2013). The discovery of EZH2 resistance mutations highlights the potential need to inhibit PRC2 through alternative mechanisms. Towards this, drugs inhibiting the ability of

EED to read and propagate the H3K27me3 modification through its WD40 repeats have recently been developed (He et al., 2017; Qi et al., 2017) but their mechanism of action and the efficacy of their use in vivo has not been well characterised yet.

1.8 The discovery of PALI1.

Due to the recent discovery of variant subtypes of PRC2 complex previous work from our laboratory (Jerman & Bracken, unpublished data) sought to identify novel PRC2.1 associated proteins. To this end, mass spectrometric (MS) analysis was performed of in-gel tryptic digestions from FLAG immunoprecipitations of FLAG-PCL1/PHF1 stably expressed in HEK293T cells (Figure 1.6A, Jerman & Bracken, unpublished data). As expected, this approach identified peptides from the PCL1, EZH2, SUZ12 and EED proteins, as well as p53, which our lab has previously reported to associate with PCL1 (Brien et al., 2015). In addition, peptides of the LCOR (Ligand dependent co-repressor) protein and the predicted C10ORF12 protein were also detected (Figure 1.6A, Jerman & Bracken, unpublished data). LCOR is a widely expressed transcriptional co-repressor known to associate with CTBP proteins and nuclear receptors via conserved N-terminal domains (Fernandes et al., 2003; Palijan et al., 2009) it has recently been identified as a PRC2 associated protein in several mass spectrometry analyses (Hauri et al., 2016; Smits et al., 2013). Interestingly, the peptides of both proteins were detected at the top of the gel at a molecular weight corresponding to approximately 260kDa, which is larger than the predicted ~50 kDa for LCOR and ~137 kDa for C10ORF12 (Figure 1.6A, Jerman & Bracken, unpublished data). To investigate this, an antibody was generated against C10ORF12 and Western blots were performed on

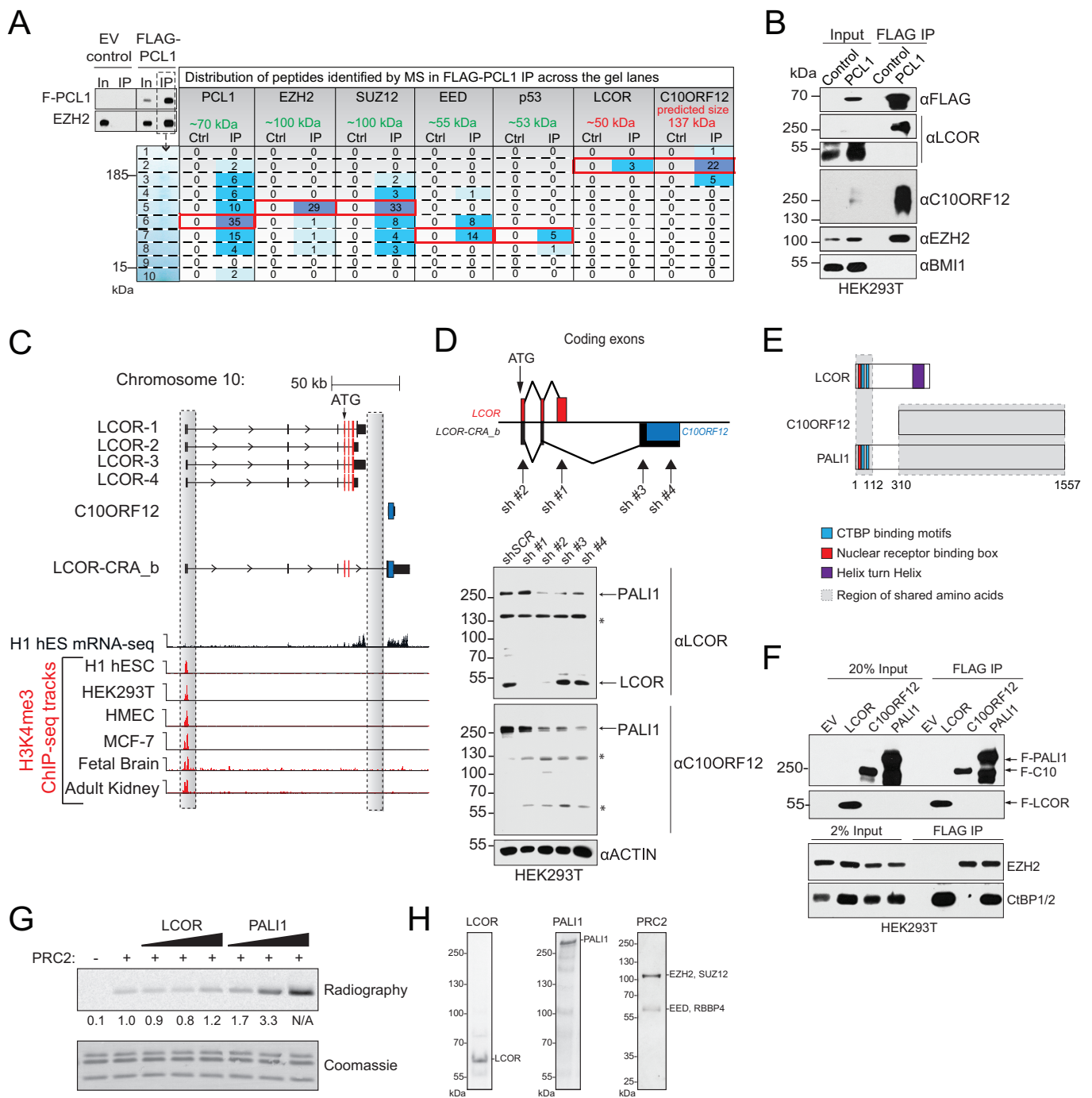


Figure 1.6 PALI1 associates with CTBP co-repressor proteins and PRC2.1 and promotes HKMT activity in vitro.

(A) Left: Western blot analysis and Coomassie staining of FLAG immunoprecipitations (IP) of nuclear lysates of HEK293T cells stably expressing FLAG-PCL1 or empty vector (EV). The Coomassie stained gel was subjected to in-gel tryptic digestion as 10 separate slices and subjected to LC-mass spectrometry. Right: The MaxQuant peptide counts of the indicated proteins are listed in the table adjacent to the respective gel band that they were identified in.

(B) Western blot analysis with the indicated antibodies of FLAG-IPs of nuclear lysates from control and FLAG-PCL1 stably expressing HEK293T cells.

(C) A genomic view of the LCOR gene locus showing five alternative splicing variants (LCOR1-4, LCOR-CRA_b) and C10ORF12. RNA-seq of human embryonic stem cells and H3K4me3 ChIP-seq tracks in various cell lines, downloaded from ENCODE and Epigenetic Roadmap projects (Kundaje et al., 2015), are presented. Grey boxes highlight the predicted transcription start site of LCOR and C10ORF12.

an independent FLAG-PCL1 immunoprecipitation and it was found that both anti-C10ORF12 and anti-LCOR antibodies detected bands at about 260 kDa (Figure 1.6B, Jerman & Bracken, unpublished data). Supporting this result, a previous study reported that the *LCOR* gene locus encodes an isoform called *LCOR-Cra_b* (Hauri et al., 2016), which includes *C10ORF12* (Figure 1.6C, Jerman & Bracken, unpublished data). It was decided to name this 260kDa protein 'PALI1', for 'PRC2 Associated LCOR Isoform 1'.

It was essential to confirm that the PALI1 protein is encoded by the *LCOR-Cra_b* alternative splice form of *LCOR*. Firstly, consistent with C10ORF12 being part of PALI1 and not an independent gene, ChIP-Seq tracks of H3K4me3 from a number of different human cell lineages were examined (Figure 1.6C, Jerman & Bracken, unpublished data). H3K4me3 is a histone post-translational modification associated with gene promoters (Ernst et al., 2011). While H3K4me3 is enriched at the *LCOR* gene promoter, as expected, it is absent immediately upstream of the *C10ORF12* coding region. This suggests that *C10ORF12* does not possess an independent promoter. Next the 4,674 nucleotide PALI1 coding sequence was cloned by RT-PCR on cDNA from primary human mammary epithelial cells (Jerman & Bracken, unpublished data). This confirmed that PALI1 is encoded by *LCOR-Cra_b*, a transcript comprising the first two coding exons of *LCOR* spliced to a third larger exon that contains *C10ORF12* (Figure 1.6C-D, Jerman & Bracken, unpublished data). To characterise the anti-C10ORF12 (from here on referred to as anti-PALI1) and anti-LCOR antibodies and further substantiate the existence of the PALI1 protein, four shRNA constructs were designed to target specific coding exons, which would either deplete *LCOR* on its own, *LCOR* and PALI1 together, or

PAL11 on its own, in stably infected HEK293T cells (Figure 1.6D, Jerman & Bracken, unpublished data). The shRNA targeting the LCOR specific exon (sh#1) reduced the protein levels of LCOR, compared to control, scrambled shRNA, but did not affect levels of PAL11 (Figure 1.6D, Jerman & Bracken, unpublished data). The shRNA designed to target the predicted first common coding exon of the *LCOR* and *LCOR-Cra_b* mRNA transcripts (sh#2), reduced the protein levels of both LCOR and PAL11 (Figure 1.6D, Jerman & Bracken, unpublished data). Finally, two shRNAs targeting *LCOR-Cra_b*, either in the unique region between LCOR and C10ORF12 (sh#3), or within C10ORF12 (sh#4), led to reduced levels of PAL11, but did not affect LCOR (Figure 1.6D, Jerman & Bracken, unpublished data). Taken together, these data confirmed the specificity of the antibodies on endogenous proteins and that C10ORF12 forms part of the ~260KDa PAL11 protein.

Next, to explore the potential common and unique interactions of PAL11 and LCOR with PRC2 and CTBP proteins (Figure 1.6E, Jerman & Bracken, unpublished data), PAL11, C10ORF12 and LCOR were expressed as FLAG-HA-tagged proteins in HEK293T cells and FLAG-immunoprecipitations were performed (Figure 1.6F, Jerman & Bracken, unpublished data). This confirmed that, like LCOR, PAL11 immunoprecipitated CTBP1/2 corepressor proteins, while only C10ORF12 and PAL11 immunoprecipitated EZH2. Since PRC2 is a histone methyltransferase (HKMT), it was analysed whether PAL11 could modulate this activity *in vitro*. PRC2 was incubated with increasing amounts of recombinant PAL11 protein together with recombinant oligonucleosomes. This resulted in a dose-dependent increase in the methylation of histone H3, while LCOR did not promote HKMT activity, consistent with its inability to associate with the PRC2 complex (Figure 1.6G Holoch, Healy,

Margueron & Bracken, unpublished data). Taken together, it was established that PALI1 is a ~260KDa isoform of LCOR capable of binding CTBP proteins and PRC2 and promoting PRC2 activity *in vitro*. Considering that the region surrounding the LCOR gene, including C10ORF12, is recurrently deleted in Lymphoma (Figure 1.5A) (Bouska et al., 2014; Chan et al., 2015) the identification of the PALI1 spliceform could have potential implications for the mechanism of oncogenesis and the role of PRC2 in these cancers.

Chapter 2: Materials and methods

2.1 Cell culture.

HEK293T and SH-SY5Y cells were cultured in DMEM supplemented with 10% FBS (Gibco), 100 U/mL penicillin (Gibco) and 100 U/mL streptomycin (Gibco). SH-SY5Y differentiation was induced by addition of 10 μ M ATRA for 8 days. Flp-In T-Rex 293 cells (Invitrogen) were cultured in DMEM supplemented with 10% Tetracycline screened FBS (Fisher), 100 U/mL penicillin (Gibco) and 100 U/mL streptomycin (Gibco).

Embryonic stem cells (ESCs) were grown on gelatinised culture dishes in GMEM (sigma) supplemented with 10% ES cell qualified FBS (Millipore), 100 U/mL penicillin (Gibco), 100 U/mL streptomycin (Gibco), 50 μ M β -mercaptoethanol (sigma), 1:100 Glutamax (Gibco), 1:100 non-essential amino acids (Gibco), 1mM sodium pyruvate (Gibco) and 1:500 homemade leukemia inhibitory factor (LIF). For Embryoid body differentiation, 2-3x10⁶ million ESCs were washed three times with PBS (Lonza) and seeded to non-adherent petri-dishes in ESC media without LIF. The media was changed every 2 days. Insect cells (Sf9 and High Five) were cultured in Hinks TNMFH (sigma) supplemented with 10% FBS (Gibco) and 0.5% Gentamycin (Sigma).

Primary Human Mammary Epithelial Cells (HMECs) were passaged in the M87A medium: 50% MM4 medium [DMEM/F12 (Gibco), 10 μ g/ml insulin (Sigma), 10nM tri-iodothyronine (Sigma), 1nM β -estradiol (Sigma), 0.1 μ g/ml hydrocortisone (Sigma), 0.5% FBS (26140, Gibco), 5ng/ml epidermal growth factor (Peprotech), 2mM glutamine (Lonza), 1ng/ml cholera toxin (Sigma)] and 50% MCDB170 medium [MEGM media (Lonza) supplemented with 5 μ g/ml transferrin (Lonza), 10⁻⁵ isoproterenol (Sigma), and 2mM glutamine (Lonza). M87A media was further

supplemented with 0.1nM oxytocin (Bachem) and 0.1% Albu-Max 1 (Invitrogen). For passaging, ~80% confluent cells were trypsinised by STV (5.37mM KCl, 6.9mM NaHCO₃, 136.9mM NaCl, 5.55mM D-Glucose, 0.54mM EDTA, 500mg/L Trypsin (1:250)) buffer and split at a 1 to 5 ratio.

2.2 FLAG immunoprecipitations for mass spectrometric analyses.

Nuclear or whole cell pellets were re-suspended in 4-5 packed cell volumes of IPH buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 0.5mM EDTA, 0.5% NP-40 detergent, 0.5μM DTT, 2μg/mL aprotinin, 1μg/mL leupeptin, 1mM PMSF) and incubated for 30 min at 4°C with rotation. MgCl₂ was then added to a final concentration of 7mM before the digestion of chromatin with 250U/mL benzonase-nuclease overnight at 4°C with rotation. The IPH whole cell or nuclear lysates were pre-cleared to remove non-specifically binding proteins by incubation with mouse IgG agarose beads (Sigma). FLAG-tagged proteins were immunoprecipitated from these lysates using anti-FLAG-M2 affinity gel agarose (Sigma; A2220) for 4-16 hours at 4°C with rotation. The agarose beads were washed five times with IPH buffer and immunoprecipitated material was obtained in solution by competitive elution with 250μg/mL 3xFLAG peptide (Sigma) while shaking at 25°C.

2.3 Mass spectrometry analysis.

For in gel digests, after immunoprecipitation of EZH2 or BMI1, the eluates were denatured in SDS-loading buffer and then separated on 4–12% gradient Nu-PAGE gels (Novex) and stained with GelCode Blue solution (Thermo Scientific). The EZH2 and BMI1 lanes were excised and cut into 10 fragments from the top to the bottom of the gel. Each fragment was diced into small (~1 mm³) pieces and washed three times with 25mM NH₄HCO₃/50% acetone, dehydrated in

acetonitrile/ NH_4HCO_3 (3:2) and rehydrated with 50mM NH_4HCO_3 twice, followed by dehydration with undiluted acetonitrile and reduction of proteins in 10mM DTT in 25mM NH_4HCO_3 for 1 hour at 56°C and a subsequent incubation in 55mM iodoacetamide for 1 h. The gel pieces were washed with 25mM NH_4HCO_3 , then 25mM NH_4HCO_3 /50% acetonitrile and dehydrated. The peptides were then digested with one half of Trypsin Singles reaction (Sigma, T7575) in 25mM NH_4HCO_3 overnight at 37°C. Finally, peptides were extracted from the gel pieces with 50% acetone/ 5% formic acid (Sigma, 94318) and dried by vacuum concentration. The final peptide sample was resuspended in 20µl 0.1% formic acid.

Material immunoprecipitated with subsequent IPs for in solution digest, were treated with trypsin while still bound to agarose beads, as described (Wisniewski et al., 2009). The digested peptides of all immunoprecipitations were analysed on a Thermo Scientific LTQ Orbitrap. Each sample was injected onto a nanoACQUITY Symmetry C18 trap (5µm particle size, 180µm x 20mm) in buffer A (0.1% formic acid in water) at a flow rate of 4µl/min and then separated over a nanoACQUITY BEH C18 analytical column (1.7µm particle size, 100µm x 100mm) over 1 h with a gradient from 2% to 25% buffer B (99.9% acetone/0.1% formic acid) at a flow rate of 0.4µl/min. The mass spectrometer continuously collected data in a data-dependent manner, collecting a survey scan in the Orbitrap mass analyzer at 60,000 resolution with an automatic gain control (AGC) target of 1×10^6 followed by collision-induced dissociation (CID) MS/MS scans of the 10 most abundant ions in the survey scan in the ion trap with an AGC target of 5,000, a signal threshold of 1,000, a 2.0 Da isolation width, and 30ms activation time at 35% normalized collision energy. Charge state screening was employed to reject unassigned or 1+

charge states. Dynamic exclusion was enabled to ignore masses for 30 s that had been previously selected for fragmentation. Raw files were processed using version 1.1.36 of MaxQuant. For protein identification the human UniProt database (release 2013_12; 67,911 entries) was combined with the reversed sequences and sequences of widespread contaminants, such as human keratins. Carbamidomethylation was set as fixed modification. Variable modifications were oxidation (M) and N-acetyl (protein). Initial peptide mass tolerance was set to 6 ppm and fragment mass tolerance was set to 20 ppm. The peptide and protein false discovery rates (FDR) were set to 0.01.

2.4 Cloning and plasmid generation.

Full length ORFs of *PCL1*, *PCL2*, *RING1B*, *LCOR*, *C10ORF12* and *PAL11*, as well as all protein fragments used in this study, were PCR amplified (primers available upon request) from cDNA generated from either HMECs or HEK293T human cell lines. The human *JARID2* ORF was PCR amplified from Open Biosystems image clone 4520786. Codon optimised gBlocks for human *HOXA9*, *HOXC9* and *HOXD9* coding sequences were purchased (IDT). All PCR products and gBlocks were inserted into the pCR8/GW/TOPO Gateway cloning entry vector (Invitrogen). The pCR8-*AEBP2* plasmid was a kind gift from Kristian Helin. The *PCGF1-5* expression vectors were a kind gift from Robert Klose. All ORFs were subsequently sub-cloned into Gateway compatible expression vectors by recombination using LR-Clonase enzyme (Invitrogen). Single and double point mutants of *PAL11* extended PIP were generated using the GeneArt site-directed mutagenesis system (Invitrogen), in accordance with manufacturer's instructions.

2.5 Generation of overexpression cell lines.

Transient or stable ectopic expression of proteins was achieved using the calcium phosphate transfection of HEK293T with either pMINKIO FLAG-HA or pLENTI FLAG-HA expressions vectors. 1-10µg of DNA was transfected into HEK293T for 24-48 hours. For generation of stable cell lines the cells were selected with puromycin (1µg/mL) 48 hours post transfection.

Stable ectopic expression SH-SY5Y cell lines were generated by nucleofecting (Amaza cell line nucleofector kit V) 1.75×10^6 cells with 2µg of plasmid using amaza nucleofector II A-023 programme. 48 hours post-nucleofection cells were split to clonal density and selected with 1µg/mL puromycin. Clones were screened for expression of the protein of interest.

Retroviral ectopic expression in HMEC cells was achieved by generation of retrovirus in amphotrophic HEK293T phoenix cells. The phoenix cells were transfected with pMINKIO expression vectors using the calcium phosphate method. 8 hours post-transfection the phoenix cell media was replaced with HMEC culture media. Retrovirus was collected from the media of the phoenix cells 48 and 72 hours post transfection, filtered through a 0.45µm filtered and subsequently was put onto seeded HMEC cells along with 0,5µg/mL polybrene for two 8 hour periods on consecutive days. 48 hours after infection the HMEC cells were selected with 1µg/mL for 5-7 days before checking expression of the protein of interest.

Inducible GAL4 tagged protein expression HEK293T cells were generated as described (Brien et al., 2012). pCDNA5-FRT-TO-GAL4 vector containing the ORF coding sequence for the protein of interest was co-transfected with pOG44 Flp-recombinase vector into Flp-In Trex 293T cells using the calcium phosphate

transfection method. Subsequently cells were split to clonal density and subject to selection with 300µg/mL Hygromycin. Tagged protein expression was induced by addition of 1µg/mL Doxycycline for 48 hours.

2.6 Preparation of whole cell protein lysates.

Cells were scraped down to collect them, washed three times in PBS and resuspended in ice cold High Salt buffer (50mM Tris-HCl, pH 7.2, 300mM NaCl, 0.5% (v/v) NP-40, 1mM EDTA pH7.4, 2µg/mL Aprotinin, 1µg/mL Leupeptin, 1mM PMSF). Cells were then sonicated and incubated for 20 minutes at 4°C while rotating to ensure sufficient lysis. The lysates were then clarified at 14,000RPM at 4°C for 20 mins. PCR

2.7 RNA preparation and RT- analysis.

Total RNA was extracted using the RNeasy kit (Qiagen) and cDNA was generated by reverse transcription PCR using the TaqMan Reverse Transcription kit (Applied Biosystems). Relative mRNA expression levels were determined by the SYBR Green I detection chemistry (Applied Biosystems) on the ABI Prism 7500 Fast Real-Time PCR System. The levels of *RPLPO* or *GAPDH* were used as normalisers. The error bars indicate standard deviation of triplicate qPCR data.

2.8 FLAG immunoprecipitations.

Cells were resuspended in 650µL of High Salt buffer (50mM Tris-HCl, pH 7.2, 300mM NaCl, 0.5% (v/v) NP-40, 1mM EDTA pH7.4, 2µg/mL Aprotinin, 1µg/mL Leupeptin, 1mM PMSF) and sonicated once for 10 seconds. The lysates rotated at 4°C for 20 minutes before diluting the lysates with 650µL of No Salt buffer (50mM Tris-HCl, pH 7.2, 0.5% (v/v) NP-40, 1mM EDTA pH7.4, 2µg/mL Aprotinin, 1µg/mL Leupeptin, 1mM PMSF). The lysates were then clarified at 20,817g at 4°C. The

lysates were pre-cleared with equilibrated Mouse IgG coupled beads (Sigma) for 1 hour. The lysates were then incubated with 20 μ L of FLAG beads (Sigma) overnight in the presence of 250U/mL Benzonase nuclease. Beads were washed 5 times with Wash buffer (1:1 dilution of High Salt: No Salt buffer). FLAG tagged complexes were eluted by competitive elution using 250 μ g/mL 3xFLAG peptide (Sigma) while shaking at 25°C.

2.9 Evolutionary analysis.

A maximum likelihood phylogenetic tree was inferred using RaxML (Stamatakis, 2014) from a Clustal Omega (Sievers et al., 2011) alignment of LCOR and LCORL exonic orthologue sequences identified using TBLASTN (Camacho et al., 2009) and HMMER (Finn et al., 2015) in representative genomes (Yates et al., 2016).

2.10 Endogenous IPs.

Embryonic stem cells were resuspended in Buffer C (20mM HEPES pH 7.9, 0.2mM EDTA, 1.5mM MgCl₂, 20% glycerol, 420mM NaCl, 2 μ g/mL Aprotinin, 1 μ g/mL Leupeptin, 1mM PMSF), sonicated 3x 15 seconds and dounced 20 times with a tight pestle. Lysates were incubated for 20 min rotating at 4°C and clarified by centrifugation at 20,817g at 4°C for 20 min. Lysates were dialysed for 5 hours at 4°C against 50 volumes of Buffer C100 (20mM HEPES pH 7.9, 0.2mM EDTA, 1.5mM MgCl₂, 20% glycerol, 125mM KCl). Lysates were again clarified by centrifugation at 20,817g at 4°C for 20 min. 5 μ g antibody was coupled to 20 μ L packed Protein A beads (Sigma) by incubation in 1mL PBS (0.1% Tween-20) at 4°C rotating overnight. Beads were collected by centrifugation at 5,440g at room temperature and washed twice in 1mL 0.2M Sodium Borate pH 9.0. Antibodies were then crosslinked to beads by incubation in 1mL 0.2M Sodium Borate pH 9.0

(containing 20mM dimethyl pimelimidate dihydrochloride) at room temperature rotating for 30 min. Reaction was quenched by washing beads once in 1mL 0.2M Ethanolamine pH 8.0 and incubating for 2 hr at room temperature rotating in 1mL 0.2M Ethanolamine pH 8.0. Beads were washed once in Buffer C100 and blocked for 60 minutes at 4°C rotating in Buffer C100 (0.1mg/mL Insulin (Sigma), 0.2mg/mL Chicken egg albumin (Sigma), 0.1% (v/v) fish skin gelatin (Sigma)). Antibody-crosslinked beads were incubated with protein lysates, in the presence of 250U/mL Benzonase nuclease, at 4°C rotating for 3 hours and washed 5 times in Buffer C100 (+0.02% NP-40). After the final wash beads were resuspended in 100µL of SDS-PAGE sample buffer. Immunoprecipitated material was eluted by boiling for 5 min with shaking before centrifuging the beads at 20,817g for 5 minutes and keeping the resulting supernatant.

2.11 Chromatin Immunoprecipitations.

Cells were washed once with PBS before crosslinking for 10 minutes with PBS containing 1% formaldehyde (Sigma). Crosslinking was quenched with 0.125M Glycine for 5 minutes before two PBS washes. The crosslinked cells were lysed in 6mL of SDS-Lysis buffer (100mM NaCl, 50mM Tris pH8.1, 5mM EDTA pH 8.0, 0.02% NaN₃, 0.5% SDS, 2µg/mL Aprotinin, 1µg/mL Leupeptin, 1mM PMSF). Chromatin was pelleted by centrifugation at 1200RPM for 5 minutes at room temperature. The supernatant was then discarded and the chromatin was resuspended in 3mL of ChIP buffer (2:1 dilution of SDS-Lysis buffer: Triton dilution buffer [100mM Tris pH 8.6, 100mM NaCl, 5mM EDTA pH 8.0, 0.02% NaN₃, 5% Triton X-100, 2µg/mL Aprotinin, 1µg/mL Leupeptin, 1mM PMSF]). Chromatin was sheared to approximately 100bp-1000bp fragments by successive 30 second

rounds of sonication at 5-8% amplitude. Sonicated chromatin was pre-cleared for 30 minutes using equilibrated protein A beads (Sigma) that had been blocked in TE (10mM Tris pH8.1, 1mM EDTA pH 8.0) containing 0.5mg/mL BSA and 0.2mg/mL Herring Sperm DNA. 10-100µg (DNA) of chromatin was incubated overnight with antibody while rotating at 4°C. Following clarification of insoluble precipitates the chromatin was incubated for 3 hours with 50µL of blocked protein A beads. After incubation the beads were washed three times in Mixed Micelle Buffer (150mM NaCl, 20mM Tris pH 8.1, 5mM EDTA pH 8.0, 5.2% Sucrose, 0.02% NaN₃, 1% Triton X-100, 0.2% SDS), twice with Buffer 500 (0.1% Sodium Deoxycholate, 1mM EDTA pH 8.0, 50mM HEPES pH7.5, 1% Triton X-100, 0.02% NaN₃), twice with LiC detergent wash (0.5% Sodium Deoxycholate, 1mM EDTA pH 8.0, 250mM LiCl, 0.5% NP-40, 10mM Tris pH 8.0, 0.02% NaN₃) and finally one wash with TE. Immunoprecipitated material was eluted from the beads with Elution buffer (0.1M NaHCO₃, 1% SDS) while shaking for 1 hour at 65°C. The supernatant of the elution was retained and incubated overnight at 65°C while shaking to reverse the crosslinks. The eluted complexes were then subject to RNase (Thermo Fisher) and Proteinase K (Sigma) treatment prior to DNA clean up through Phenol Chloroform clean up and Ethanol precipitation. ChIP enrichment was analysed by qPCR using the SYBR Green I detection chemistry (Applied Biosystems) on the ABI Prism 7500 Fast Real-Time PCR System.

2.12 RNA-seq and bioinformatics analysis.

Libraries were prepared using the Truseq Stranded Total RNA library prep kit Ribo-Zero Human/ Mouse/ Rat (Illumina) according to manufacturer's instructions. Sequencing was performed on an Illumina HiSeq 2500 sequencer using HiSeq v4

chemistry following the manufacturers protocols. Raw sequencing reads were aligned to the mouse reference genome (mm10) using HISAT v2.0.5 (Pertea et al., 2016). The alignment files were converted to bigwig files and scaled using hits per billion with the bamTobw.sh utility (Zhu et al., 2013) for visualisation on the UCSC Genome Browser (Kent et al., 2002). Sequence reads were aggregated into a count for each gene using StringTie v1.3.3b (Pertea et al., 2016). Differentially expressed genes were identified using DESeq2 (Love et al., 2014). Data for H3K27me3 was taken from GSE81081 and processed as described previously (Streubel et al., 2017). In brief, reads were aligned to the mouse reference genome (mm10) using Bowtie2 v2.1.0 (Langmead and Salzberg, 2012) and peaks determined using macs2 v2.1.1 (Feng et al., 2012). H3K27me3 positive genes were defined as the set of genes with an H3K27me3 peak (FDR = 0.05) overlapping the promoter regions (TSS +/- 2kb) of mm10 RefSeq genes. Genes were ranked based on the proportion of the promoter region covered by an H3K27me3 peak.

2.13 Generation of Pali1/2 double knockout cells

Pali1 and Pali2 double knockout mouse ESCs were generated by transfecting Pali1 knockout ESCs with pSpCas9 (BB)-2A-eGFP vector (Addgene, px458) containing a guide RNA targeting the 5' end of the Pali2 specific exon, using Lipofectamine 2000. The same vector without a gRNA sequence was used as the negative control. 48 hours after transfection the GFP high population of cells was collected by FACs and individual cells were seeded to each well of a 96 well plate. Individual clones were expanded and genotyped by amplifying the region surround the PAM site and Sanger sequencing to identify indels.

2.14 Accession codes

RNA-seq data are available with the accession code: GSE100679.

2.15 Mouse dissection

E10.5 embryos were harvested from pregnant females and dissected according to the accompanying illustration under a stereomicroscope in ice cold PBS. Embryos sections were subsequently homogenised using a gentleMACs dissociater with C-tubes, using the “mouse brain” cycle.

2.16 3T3 growth assay

For proliferation curves 7.5×10^5 HMEC cells were seeded on 10cm plates. Three days later the total amount of cells is counted and 7.5×10^5 HMEC cells were seeded again. The population doubling of the cells is calculated by the formula $\log(N_f/N_i)/\log_2$ where N_i is the initial number of cells and N_f is the final number after each count.

2.17 Antibodies

Antibodies				
Target	Source	Western blot	IP	ChIP
Anti-FLAG	Sigma-Aldrich F3165	1 in 100		
Anti-EZH2	BD43 (Pasini et al., 2004)	1 in 10		
Anti-EZH2	AC22 (Bracken et al., 2006)		5µg	2µg
Anti-LCOR	Santa Cruz Biotechnology sc-163009	1 in 200		
Anti-C10ORF12/ PALI1	Merck ABE1367	1 in 500		
Anti-BMI1	AF27 (Bracken et al., 2007)		5µg	2µg
Anti-BMI1	DC9 (Bracken et al., 2007)	1 in 30		
Anti-BMI1	Merck F6		5µg	
Anti-CTBP	Santa Cruz Biotechnology sc-11390	1 in 200		
Anti-ACTIN	MP Biomedicals #08691001	1 in 5000		
Anti-PCL1	Proteintech 15663-1-AP	1 in 500		
Anti-G9A	Cell Signaling #3306 (C6H3)	1 in 1000		
Anti-H2Aub	Cell Signaling #8240 (D27C4)	1 in 2000		

Anti-H3K27me3	Active motif #61017	1 in 1000		
Anti-H3K27me3	Merck #07-449			1µg
Anti-H3	Abcam ab1791	1 in 10000		1µg
Anti-EED	#AA19.30 (Bracken et al., 2003)	1 in 10		
Anti-AEBP2	Cell Signaling #14129 (D7C6X)	1 in 500		
Anti-AEBP2	(Grijzenhout et al., 2016)			20µL
Anti-JARID2	Abcam ab48137	1 in 1000		
Anti-FLAG M2 affinity gel	Sigma-Aldrich A2220		20µL	
Anti-SUZ12	Cell Signaling #3737 (D39F6)	1 in 2000		
Anti-HA	Cell Signaling #3724 (C29F4)	1 in 1000		
Anti-HA	Santa Cruz Biotechnology sc-805	1 in 500		
Anti-RYBP	Sigma-Aldrich PRS2227	1 in 1000	5µg	
Anti-GAPDH	Santa Cruz Biotechnology sc-25778	1 in 1000		
Anti-H3K9me2	Abcam 1220	1 in 1000		
HRP-linked rabbit secondary	Sigma-Aldrich A0545	1 in 10000		
HRP-linked mouse secondary	Merck 401253	1 in 10000		
HRP-linked goat secondary	Santa Cruz Biotechnology sc-2020	1 in 10000		
Anti-p16	clone JC8	1 in 250		
Anti-H3K27me1	Ative motif 61015	1 in 2000		
Anti-H3K27me2	Cell signaling 9728S	1 in 2000		
Anti-H3K27ac	Merck 07-360	1 in 5000		
Anti-H3K36me3	Abcam ab9050	1 in 1000		
Anti-CBX8	(Atsuta et al., 2001; Dietrich et al., 2007)		5µg	
Anti-RING1B				
Anti-PCGF2	Merck 09-724		5µg	
Anti-HOXA9	Upstate #07-178		5µg	
Anti-HOXC9	Abcam-ab50839	1 in 100		
Anti-HOXD9	Santa Cruz Biotechnology sc-8320	1 in 200	5µg	
Anti-APC1	Bethyl A301-653A	1 in 2000		
Anti-APC10	Proteintech 15788-1-AP	1 in 1000		
Anti-GAL4	Santa Cruz Biotechnology sc-577			2µg

2.18 Oligonucleotides

PCR primers	Forward	Reverse
hPALI1 Fragment 1	ATGCAGAGTTCAGCTTTAGTAG	TCAAGCCTTTCCCAGACTC
hPALI1 Fragment 2	CTGCCAACTGTTCGTACACT	TCACAAAGTCTGCCTGGTG
hPALI1 Fragment 3	GAGGACAACCAAAGCATCAG	TCATGTTGGCTGTCCTTCTT

hPALI1 Fragment 4	ATGCTGGACAAAGAAGTCAAG	TCATGGGCATTCCATTTTCTCA
hPALI1 Fragment 5	CCAAGAGCAAGGAACAAATCA	TCACTTTGCATCCAGCCG
hPALI1 PIR1	AAAGAAGGACAGCCAA	TCATTTAAAGTTTGTCAAAA GAGCA
hPALI1 PIR2	GAGAGTCAGAAGTGTTCTCCT GTTCA	TCAGACAGGGGGAACGTC TCC
hPALI1 PIR3	GTCATTGTGAAGAAGCTCAAT ACTC	TCAGGTAGCTCCAGGAAATTT CTTC
hPALI1 PIR4	CTACAGGCTGAGCGCT	TCA GCTGACATCACTGCC ATCTTC
hPALI1 PIR5	GAGAATCCTGATCAAGTGG	TCAATATTTTCTAAGAATCCG GGTAC
hPALI1 PIR6	CATCCTCCAGCAAACCTGC	TCAACGACTTGGCTTACTTTC CACA
hPALI2 Fragment 4 orthologous region	CTTATGTCTTTAAACAATAAT GCTG	TCAGTTGTGCAGCTTTC
hJARID2	ATGAGCAAGGAAAGACCCAAG	TCATGATGAGCTCGAAGCAC
hPALI2 exon mapping 1F	GCCTCATGCACTGTGTAG	
hPALI2 exon mapping 2F	TTTATGGACCACGGCTACG	
hPALI2 exon mapping 3F	ACTGTGAACCAGAAGAGCTG	
hPALI2 exon mapping 4F	CATCTCTTGATTCTTCACAGT CAA	
hPALI2 exon mapping 5F	TCAGAACTGTGATCCTAACAT TCC	
hPALI2 exon mapping 6F	GGGATGGAGTGTTAGATCTCTC	
hPALI2 exon mapping 7F	CTCTACAGAAAATGGTCACTC AGTT	
hPALI2 exon mapping 8F	GCTGACCATCTTGGCAATAGTA	
hPALI2 exon mapping 8R	CTTGTTGCTGGTGATGTATGC	
hPALI2 exon mapping 9R	TGGGACGATTAAGGCACATC	
mGm340 mouse genotyping primers	GGGAACGCCAGTTTGAGATGC	CGGCGGCCTCTTCAGTGACAG
mGm340 ESC genotyping primers	GGGAACGCCAGTTTGAGATGC	GGGAGTTGTGCTGCGGTC

RT-qPCR primers	Forward	Reverse
<i>mRplpo</i> mRNA	TTCATTGTGGGAGCAGAC	CAGCAGTTTCTCCAGAGC
<i>mPax3</i> mRNA	AACAAGCTGGAGCCAATCA ACTGA	AGGTCTCCGACAGCTGGTA TGTTG
<i>mT</i> mRNA	CTTCGTGACGGCTGACAAC CA	GCGCCTGAGGCTCTGGTTT G
<i>mGata6</i> mRNA	CAGCAAGATGAATGGCCTC AGC	CAAGCCGCCGTGATGAAG G
<i>mPali1</i> mRNA	GATGTCAGCCTCCATCAAG ACTGC	GGCATTCTGTTATTGGTTG GCTGA
<i>mAebp2</i> short isoform mRNA	CCGGGAGATGTGGTCAGA GCTG	TGCAGCACCCACAAGGCG

<i>mAebp2</i> long isoform mRNA	TGACGAGGAGGCGGAGGC A	GCTCCGACATGGTCTCCG CC
<i>mBmi1</i> _mRNA_F1	CTTCCTGTTTGCCAGTCC CA	CTGTTGCTGGTTCCATTCA TGGTAC
<i>mHoxA9</i> mRNA	CAAGGCAGGCGGCAGCTC T	CGCCGCCTTGGACTGGAA G
<i>mHoxC9</i> mRNA	ACT GTA GCG ATT TTC CGT CCT G	CAC GAC GTA CTG AAT ACA GCC G
<i>mHoxD9</i> mRNA	CCGAGTTCGCCTCGTGTAG TTTTG	TGGTACAGGCCGCTCATC GTC
<i>hCCNA2</i> mRNA	AGCTGGCCTGAATCATTAA TACG	GGTGAAGGTCCATGAGAC AAGG
<i>hINK4A</i> mRNA	AGAGGATTTGAGGGACAG GGTC	CCTCTTTCTTCTCCGGTG C

ChIP-qPCR primers	Forward	Reverse
<i>mPax3</i> tss	AGAAAGGCGGAAAGAGATTA GGAC	GAGCCAGTGAGGATGGAAAA AG
<i>mT</i> tss	TCCGCAGAGTGACCCTTTTTTC	TACCCAACAGCCACCTTCACT TC
<i>mGata4</i> tss	TCTTTCCTCCCTACTCTCAGT GGTCC	TCACCTTCTCCTCTACCAGCC CC
<i>mGata6</i> tss	TTTTTCTCCTCCCCTGGATCC	CTCTCTACAAGCTGCTGGGTG C
<i>mEomes</i> tss	AAGAGAAAATGAACTCGCAG GCG	TTTGCTATTGGCTGTTAGAGG TGACA
<i>mNkx2.9</i> tss	TAAGGATGGAAGTGCGAGGC	TTCGCTCCAGCACTCATTCA
<i>mSox1</i> tss	TTTGACACAGTTCAGCCCTGAG TGA	GGTGCACAAACCACTTGCCAA AGA
<i>mGapdh</i> tss	AGTGTGCACCAAGGACATCCA G	CCCATTTTACTCGGGAAGCAG
<i>hCCNA2</i> tss	TGACGTCATTCAAGGCGA	GCTCAGTTTCTTTGGTTTAC
<i>Luciferase</i> tss	ACGCCAAAAACATAAAGAAAG GC	CCAGCGGTTCCATCCTCTAGA G

gRNA sequences	Target sequence	Reverse complement target sequence
gPali2 #1	CGAGGGACTTCGAGATCTAC	GTAGATCTCGAAGTCCCTCG

Chapter 3. Proteomic analysis of the PRC1 complex

3.1 Introduction

The PRC1 complexes exists in a variety of combinations of canonical and non-canonical forms (Conway and Bracken, 2017; Gao et al., 2012). The canonical PRC1 (cPRC1) complex is recruited to chromatin primarily through the Chromodomain containing subunits (CBX2/4/6/8), which bind to H3K27me3 on PRC2 target genes (Fischle et al., 2003; Min et al., 2003). However, several recent papers have suggested alternative mechanisms of recruitment for cPRC1 through interaction with DNA binding transcription factors such as ZFP277, RUNX1, REST and CBF β (Dietrich et al., 2012; Negishi et al., 2010; Ren and Kerppola, 2011; Yu et al., 2012). These studies reported physical interactions between PRC1 and these transcription factors and also functionally linked the transcription factors to the repressive activity of cPRC1 at cohorts of Polycomb target genes.

The primary aim of this chapter was to use a proteomics approach to identify potential transcription factors that can associate with PRC1 and may be involved in recruiting or coordinating its activities during differentiation. To do this I used a neuroblastoma cell lines (SH-SY5Y) in which lineage specific transcription factors are already activated, as opposed to studies in Embryonic Stem Cells (ESC) in which most lineage specific genes are “poised”. In addition, our lab has previously used this cell line as a model for dynamic changes in Polycomb activity upon induction of differentiation (Egan et al., 2013).

A technical aim for this project was to optimise our protocols for endogenous co-immunoprecipitation followed by LC-MS/MS. The basis for this aim is to test a new trypsinisation method, which has subsequently been published (Oliviero et al.,

2016; Streubel et al., 2017). This new method would allow for more cost and time effective LC-MS/MS analysis of co-IPs.

A secondary aim of this project was to identify potential non-histone targets for PRC1 mediated ubiquitination. It has previously been determined that, in *Drosophila* at least, Psc, the homolog of mammalian BMI1/PCGF4 and PCGF2, is capable of mediating ubiquitination of the Cyclin B protein and target it for degradation (Mohd-Sarip et al., 2012). Similarly, three reports in mammalian cells provided evidence that the PRC1 complex components BMI1 and RING1B can contribute to the ubiquitination of p53 and its subsequent degradation (Calao et al., 2013; Su et al., 2013; Wen et al., 2014).

In this chapter, I initially observed a potential physical link between HOX-9 transcription factors and BMI1. I show that although the HOX-9 proteins can act as repressors they do not have a role in regulating the onset of senescence and that the observed interaction with BMI1 was the result of an unfortunate artefact of non-specific antibody binding. Despite this, I generated high quality PRC1 subunit proteomics data sets in a differentiated neuroblastoma cell line, which may be a useful resource for future studies.

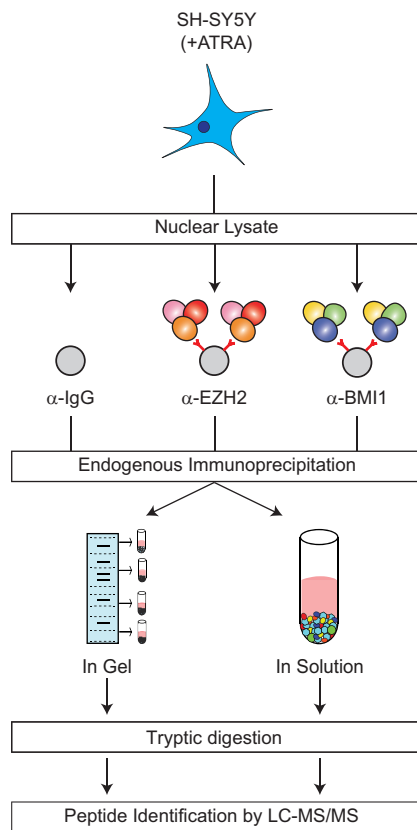
3.2 Results

3.2.1 Evaluation of “in gel” and “in solution” tryptic digest LC-MS/MS methods for identification of novel PRC1 associated proteins.

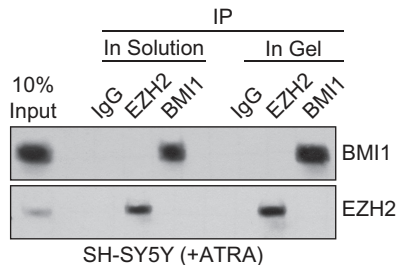
In order to identify novel PRC1 and PRC2 complex associated proteins I performed endogenous co-immunoprecipitations (IP) of BMI1 and EZH2 followed by unbiased identification of associated proteins using liquid chromatography mass spectrometry (LC-MS/MS). Additionally, I tested the quality of a new LC-MS/MS tryptic digest method in which proteins are subjected to tryptic digestion while the immunoprecipitated proteins are still bound to antibody conjugated beads (Oliviero et al., 2016; Streubel et al., 2017). This “in solution” digestion method was performed in parallel to the standard method of eluting antibody bound proteins from the beads followed by their resolution on SDS-PAGE gels and subsequent trypsinisation within the gel (Figure 3.1A) (Brien et al., 2012). The benefits of the “in solution” digest method are reduced cost of the LC-MS/MS run, shorter experimental time and a lower risk of contamination by environmentally abundant peptides, which can occur when running and cutting the SDS-PAGE gel. Most importantly, I wanted to assess whether the “in solution” method produced results of a similar quality to the “in gel” digestion method.

To do this, Western blots were performed on BMI1 and EZH2 immunoprecipitations from nuclear lysates of SH-SY5Y cells treated with ATRA (all-trans retinoic acid) for 8 days. This revealed that BMI1 and EZH2 immunoprecipitations each pulled down their respective baits (Figure 3.1B). Importantly, BMI1 and EZH2 did not co-immunoprecipitate each other showing high technical specificity. The remaining immunoprecipitated samples were then subjected to LC-MS/MS by either the “in

A



B



C

	Protein Name	IgG		EZH2		BMI1	
		Gel	Solution	Gel	Solution	Gel	Solution
Canonical PRC1 Complex	PHC1	0	0	0	0	25	67
	PHC2	0	0	4	3	61	57
	SCML2	0	0	1	0	9	50
	CBX2	0	0	0	2	12	43
	CBX8	0	0	8	1	43	42
	RING1B	0	0	6	3	35	38
	RING1A	0	0	2	1	29	32
	BMI1	0	0	1	2	17	30
	CBX4	0	0	0	0	5	29
	PHC3	0	0	0	0	7	24
	SCMH1	0	0	0	0	2	23
	PCGF2	0	0	0	0	2	19
	CBX6	0	0	0	0	0	4
Non-Canonical PRC1 Complex	USP7	0	0	5	2	6	26
	YAF2	0	0	0	0	2	9
	RYBP	0	0	0	0	2	7
	CBX3	0	0	2	6	0	6
	KDM2B	0	0	0	0	0	1
	BCORL1	0	0	0	0	0	1
	SKP1	0	0	0	2	0	1
	PCGF1	0	0	1	1	2	1
	CSNK2B	0	0	1	0	0	4
	SKP1	0	0	0	2	0	1
Hox Proteins	HOXD9	0	0	0	0	1	6
	HOXC9	0	0	0	0	0	2
	HOXD8	0	0	0	0	0	1
	HOXA9	0	0	0	0	2	1
PRC2 Complex	SUZ12	0	0	145	91	2	0
	EZH2	0	0	115	75	1	0
	JARID2	0	0	66	64	0	0
	EED	0	0	92	57	0	0
	RBBP4	0	2	59	53	2	9
	PCL2	0	0	55	37	0	0
	AEBP2	0	0	33	32	0	0
	C10ORF12	0	0	21	21	1	0
	EPOP	0	0	23	19	0	0
	RBBP7	0	0	7	19	0	4
	PCL1	0	0	23	18	0	0
	PCL3	0	0	9	17	0	0
	EZH1	0	0	0	2	0	0

Figure 3.1 Evaluation of in gel and in solution tryptic digest LC-MS/MS methods for identification of novel PRC1 associated proteins.

(A) Experimental strategy used to identify novel EZH2 and BMI1 associated proteins in SH-SY5Y nuclear lysates followed by either in gel or in solution tryptic digestion methods.

(B) Endogenous IPs of EZH2 and BMI1 in SH-SY5Y (+ATRA) nuclear lysates, followed by western blot analysis with the indicated antibodies. The resulting eluates were subject to LC-MS/MS using the indicated tryptic digestion method.

(C) Table showing the spectral counts resulting from either in gel or in solution tryptic digestion following EZH2 or BMI1 endogenous co-IPs as shown in (B).

gel” or “in solution” methods. BMI1 and EZH2 co-immunoprecipitations with both methods of tryptic digest were successful in immunoprecipitating their respective complexes (Figure 3.1C). This also confirmed the specificity of the EZH2 and BMI1 immunoprecipitations, despite negligible amounts of background signal for PRC2 complex subunits in the BMI1 immunoprecipitations and signal for PRC1 complex subunits in the EZH2 immunoprecipitations. As the PRC1 and PRC2 complexes frequently co-occupy genomic sites (Bracken et al., 2006) this low level background could be a result of incomplete nuclease digestion during the immunoprecipitation causing either experiment to pull down nearby chromatin and its associated proteins. . The EZH2 co-immunoprecipitation was enriched for several known PRC2 complex subunits including EED, AEBP2, JARID2, SUZ12, EPOP and PCL1. These subunits were enriched to very similar extents with the “in solution” and the “in gel” methods, suggesting that both approaches are equally effective in identifying PRC2 associated proteins. Surprisingly, the “in solution” method was slightly more effective than the “in gel” method in the BMI1 experiment for detecting peptides from subunits such as PHC2, RING1A and CBX2 (Figure 3.1C). Both methods were capable of detecting components of the non-canonical PRC1 complex, RYBP and YAF2. While the composition of variants of the non-canonical PRC1 complexes including PCGF1, PCGF3, PCGF5 or PCGF6 have been well established in previous studies (Endoh et al., 2017; Gao et al., 2014; Gao et al., 2012; Hauri et al., 2016), the endogenous composition of BMI1 containing non-canonical PRC1 complexes has not been defined. Here I show that BMI1 also co-immunoprecipitated USP7, SKP1, BCORL1 and KDM2B (Figure 3.1C), perhaps suggesting that it can be a part of a non-canonical complex with a similar composition to the PCGF1-ncPRC1 complex. A caveat to this, however, is

that the BMI1 IPs also identified peptides mapping to PCGF1, suggesting that the PCGF1-PRC1 complex was non-specifically immunoprecipitated and that the presence of USP7, BCORL1, SKP1 and KDM2B is a result of this indirect PCGF1 immunoprecipitation, as previously mentioned this could be due to incomplete nuclease digestion as PCGF1-PRC1 and cPRC1 have been reported to co-occupy genomic sites (Blackledge et al., 2014; Liefke et al., 2016).

One of the primary aims of this experiment was to identify novel BMI1 associated transcription factors that could perhaps provide a mechanism for how BMI1-PRC1 is targeted to its target genes as previously reported for REST, RUNX1 and ZFP277 (Dietrich et al., 2012; Negishi et al., 2010; Ren and Kerppola, 2011; Yu et al., 2012). Intriguingly, both “in gel” and “in solution” tryptic digestions of BMI1 immunoprecipitated material identified peptides mapping to multiple HOX proteins including HOXA9, HOXC9 and HOXD9 (Figure 3.1C). This observation is the first potential evidence of a physical interaction between Polycomb complexes and HOX transcription factors, which are a family of conserved homeodomain containing transcription factors with essential roles in the regulation of anterior-posterior axis specification (McGinnis and Krumlauf, 1992). I next wished to explore the possibility of a potential mechanism of recruitment for PRC1 via HOX transcription factors.

3.2.2 *In vivo* BMI1 co-immunoprecipitation.

In order to validate the potential interaction between the BMI1 and HOX-9, proteins I decided to examine a biological system in which the HOX proteins are relevant, the developing mouse embryo (McGinnis and Krumlauf, 1992; Noordermeer et al., 2011). I first established the anterior expression boundary of each of my Hox

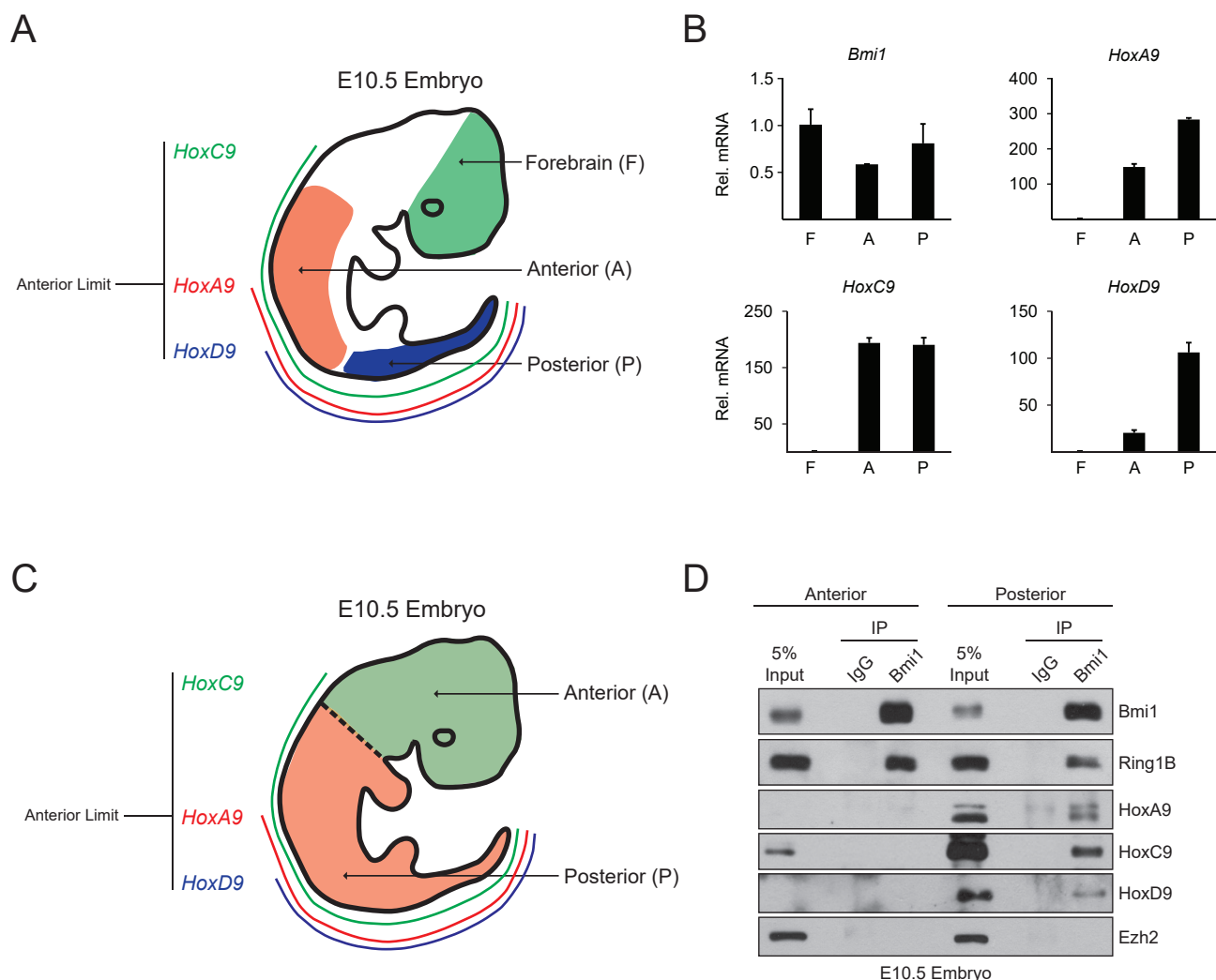


Figure 3.2 *In vivo* BMI1 co-immunoprecipitation.

(A) Illustration of an E10.5 mouse embryo and the anterior expression limits of HoxA9, HoxC9 and HoxD9.

(B) RT-qPCR of the indicated genes in either the forebrain (F), anterior (A) and posterior (P) regions of the E10.5 embryo as illustrated in (A).

(C) Illustration of the dissection performed on E10.5 embryos for endogenous co-IPs carried out in (D).

(D) Endogenous IPs of BMI1 in the anterior and posterior regions of E10.5 embryos as illustrated in (C), followed by western blot analysis with the indicated antibodies.

genes of interest by isolating anterior, posterior and forebrain sections of E10.5 mice (Figure 3.2A) and performing RT-qPCRs for *HoxA9*, *HoxC9* and *HoxD9*. In accordance with published data (Noordermeer et al., 2011), *HoxD9* expression was almost entirely restricted to the posterior section, while *HoxA9* and *HoxC9* were highly expressed in both the anterior and posterior sections (Figure 3.2B). None of the *Hox* genes tested exhibited expression in the forebrain, while *Bmi1* expression levels were constant throughout the three sections.

I decided to revise the dissection plan to minimise tissue waste. To do this, the embryos were sectioned into only two parts; anterior and posterior (Figure 3.2C). The anterior section served as a control region with high *Bmi1* expression levels but low *Hox-9* expression, while the posterior section has high *Bmi1* and *Hox-9* expression. *Bmi1* endogenous co-immunoprecipitations were performed in both the anterior and posterior sections of the embryo which revealed that Ring1B was equivalently immunoprecipitated, as expected (Figure 3.2D). However, in the posterior section the *Bmi1* antibody was capable of co-immunoprecipitating *HoxA9*, *HoxC9* and *HoxD9*, while none were detected in the anterior segment. This experiment provided additional evidence to the LC-MS/MS experiment that *Bmi1* might be capable of immunoprecipitating HOX-9 proteins, albeit both experiments were restricted to the use of the same endogenous BMI1 antibody.

3.2.3 HOX-9 proteins function as transcriptional repressors but do not recruit PRC1.

The HOX-9 proteins have previously been reported to act as transcriptional repressors (Jung et al., 2010; Jung et al., 2014). As such, the potential interaction between HOX-9s and BMI1 represented a very interesting potential mechanism

whereby HOX-9s, akin to ZFP277 and RUNX1, may be capable of recruiting the PRC1 complex to its target promoters to mediate transcriptional repression.

To test this hypothesis, we made use of an established artificial recruitment reporter assay (Brien et al., 2012; Hansen et al., 2008) in which a doxycycline inducible GAL4 tagged bait protein can be recruited to its binding sites within an artificial promoter upstream of a luciferase reporter gene (Figure 3.3A). These Flp-In TREX 293 cells were used to generate stable cell lines with inducible expression of GAL4-HOXA9, GAL4-HOXC9, GAL4-HOXD9, GAL4-BMI1 (as a positive control) and GAL4 empty vector (EV) as a negative control. These cell lines were treated for 48 hours either with DMSO, as a control, or Doxycycline to induce GAL4 tagged protein expression. The induction of GAL4-BMI1 expression led to a 68% decrease in luciferase activity, consistent with the role of BMI1 in the repressive cPRC1 complex (Figure 3.3B). Interestingly, HOXA9, HOXC9 and HOXD9 exhibited similar levels of repression (69-79%). Next, to determine if HOXD9 mediated repression correlated with recruitment of the PRC1 complex, chromatin immunoprecipitations (ChIPs) were performed. ChIP of GAL4 at the luciferase promoter showed that following DOX induction in the GAL4-EV, GAL4-HOXD9 and GAL4-BMI1 cell lines that the fusion proteins were indeed recruited to their binding sites at the luciferase promoter, but not at the CCNA2 promoter, a negative control (Figure 3.3C). The BMI1 ChIPs established that DOX induced GAL4-BMI1 cells did have BMI1 recruitment as expected, but the induced GAL4-HOXD9 cells did not result in any recruitment of BMI1. This result suggests that while exogenously expressed HOX-9s proteins can act as repressors, they do not recruit BMI1 at least in this artificial system.

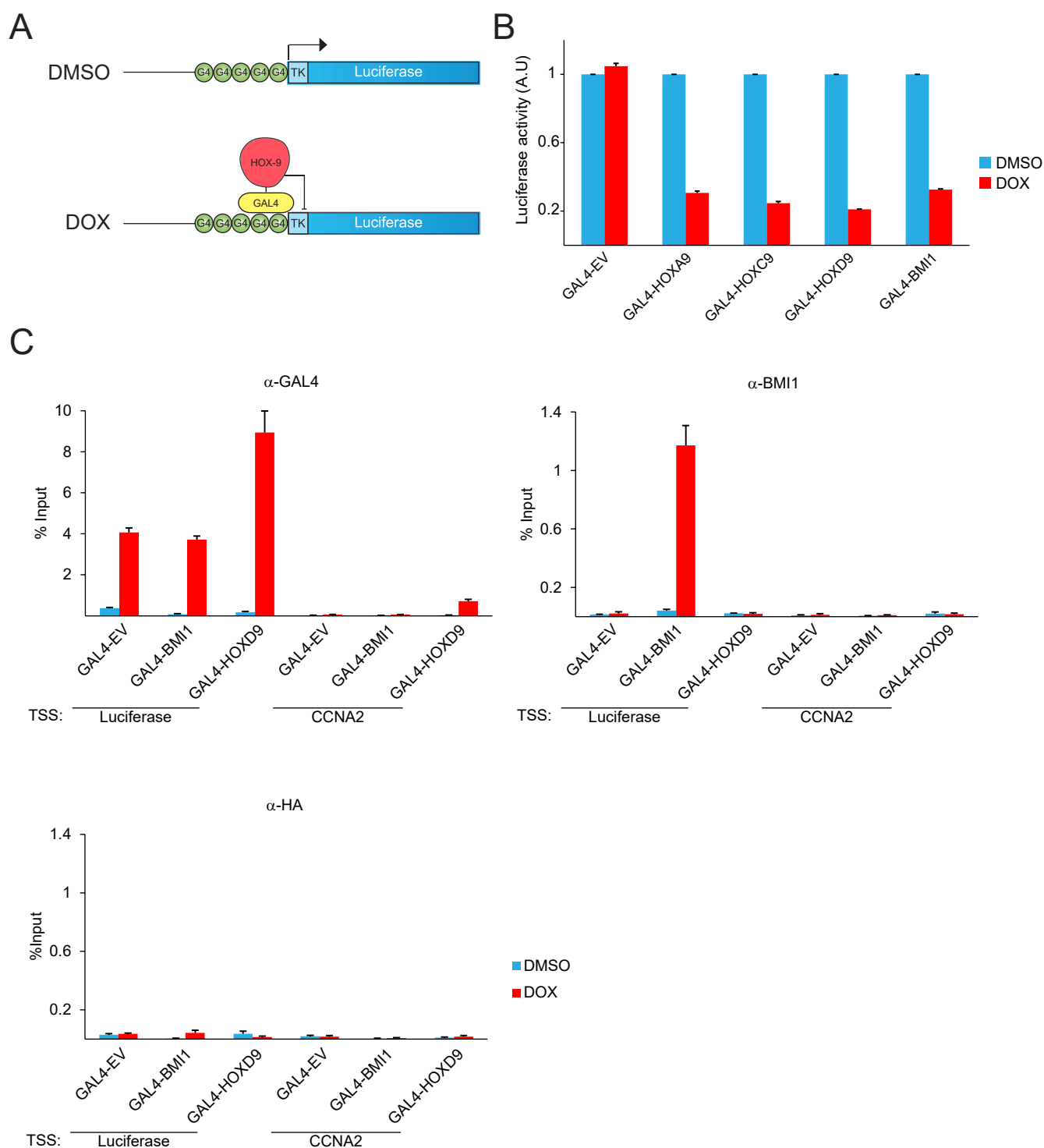


Figure 3.3 HOX-9 proteins function as transcriptional repressors but do not recruit PRC1.

(A) Schematic representation of the GAL4-TK Luc reporter system with a Luciferase gene under the control of a promoter containing GAL4 binding motifs.

(B) Quantification of Luciferase activity in cells after a 48 hour Doxycycline treatment of cells to induce expression of the indicated GAL4 tagged protein. Experiment was carried out in biological triplicates and technical triplicates.

(C) ChIP-qPCR analyses using the indicated antibodies on GAL4-EV, GAL4-HOXD9 or GAL4-BMI1 inducible cell lines after DMSO or Doxycycline treatment. Experiment was carried out in technical triplicates with one biological replicate.

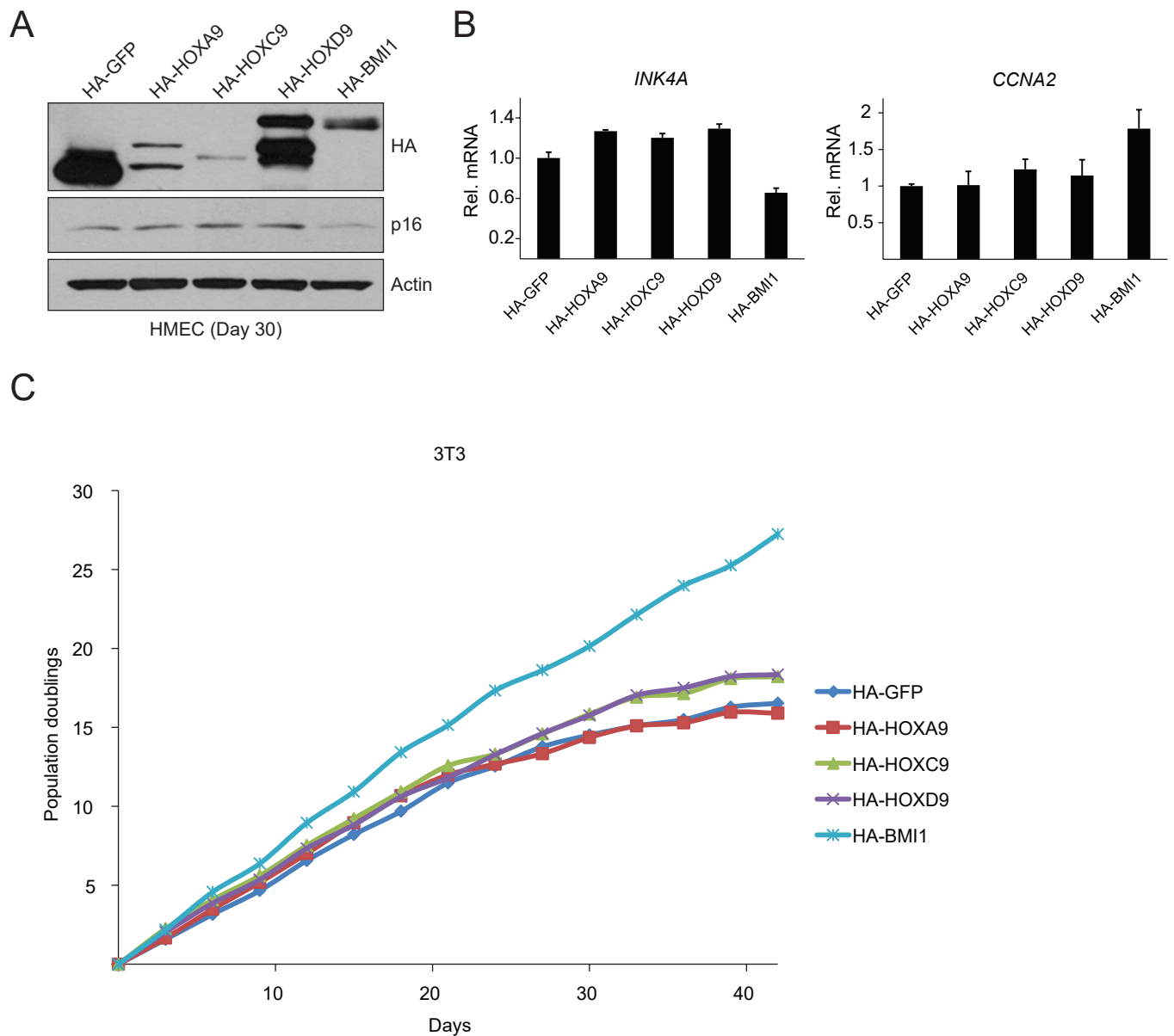


Figure 3.4 Ectopic expression of HOX-9s does not repress INK4A and has no effect on cellular proliferation.

(A) Western blot analysis on a dilution series of whole cell lysates of HMEC cells at day 30 of the 3T3 assay ectopically expressing the indicated HA tagged proteins.

(B) RT-qPCR of cDNA from HMEC cells at day 30 of the 3T3 assay ectopically expressing the indicated HA tagged protein. The results were normalised to levels of the RPLPO housekeeping gene.

(C) 3T3 growth assay on HMEC cells overexpressing the indicated HA tagged protein.

3.2.4 Ectopic expression of HOX-9s does not repress *INK4A* and has no effect on cellular proliferation.

To investigate the possibility that HOX-9s may be able to coordinate the repressive activity of the PRC1 complex at an endogenous PRC1 target gene, I next ectopically expressed HA tagged HOXA9, HOXC9, HOXD9 and BMI1 in human mammary epithelial cells (HMECs). This primary cell line, like human diploid fibroblasts, will undergo replicative senescence in culture after successive passaging. Early passage HMECs have low expression of the key senescence regulator *INK4A*, which increases with successive passages (Bracken et al., 2007; Brien et al., 2015). *INK4A* encodes the p16 protein, an essential negative regulator of the E2F pathway. High levels of *INK4A* blocks the G1 to S phase transition of the cell cycle. The expression of *INK4A* has been well established to be regulated by the PRC1 and PRC2 complexes. The *INK4A* promoter is occupied by Polycomb proteins and H3K27me3 in early passages, but they are displaced while *INK4A* is activated as cells senesce (Bracken et al., 2007). The ectopic expression of BMI1 maintains the repressed state of the *INK4A* locus, which correlates with a delayed onset of senescence (Bracken et al., 2007). Similarly, ectopic expression of HOXA9 has been previously reported to repress *INK4A* and delay the onset of senescence in IMR90 fibroblasts (Martin et al., 2013).

Here, I ectopically expressed HA tagged HOXA9, HOXC9, HOXD9 and BMI1 in HMEC cells and monitored the effects on *INK4A* expression and cellular proliferation (Figure 3.4A). While ectopic expression of BMI1 led to decreased p16 (encoded by *INK4A*) protein levels, HOXA9, HOXC9 and HOXD9 did not affect p16 levels (Figure 3.4A) These changes in protein levels were also reflected on the

mRNA level as RT-qPCRs of the *INK4A* gene showed no change following HOX-9 overexpression, but a reduction was seen following BMI1 ectopic expression, as expected (Figure 3.4B). Furthermore, while ectopic expression of BMI1 led to increased mRNA levels of *CCNA2*, the expression of the HOX-9 proteins had no effect. In addition, I performed a 3T3 growth assay which showed that, unlike BMI1, the expression of HOX-9 proteins did not affect proliferation of HMECs (Figure 3.4C). Taken together these results suggest that HOX-9s have no effect on the repression of *INK4A* or the regulation of senescence. This was unexpected considering that HOXA9 has been proposed to repress *INK4A* in IMR90 fibroblasts (Martin et al., 2013). Nonetheless, this assay failed to show coordination of PRC1 activity by the HOX-9 transcription factors at the *INK4A* locus.

3.2.5 HOX-9 proteins do not co-immunoprecipitate BMI1.

Next, to establish whether BMI1 associates with the HOX-9 proteins in the context of the PRC1 complex or independently, I performed co-immunoprecipitations for RING1B (Figure 3.5A) along with BMI1 followed by Western blot analysis. As observed previously, BMI1 antibody immunoprecipitated HOXA9 and HOXD9 with high efficiency. However, RING1B antibody was incapable of co-immunoprecipitating HOXA9 and HOXD9, but did efficiently enrich for BMI1, RYBP and CBX8. Neither BMI1 nor RING1B could co-immunoprecipitate EZH2, which confirmed the specificity of the experiment. This result supports the idea that the observed interaction between BMI1 and the HOX-9 proteins is independent of RING1B, and possibly PRC1. To further assess if this interaction is independent of the PRC1 complex I performed co-immunoprecipitations for CBX8 and RYBP (Figure 3.5B) which both immunoprecipitated BMI1 but not HOXD9. Taken

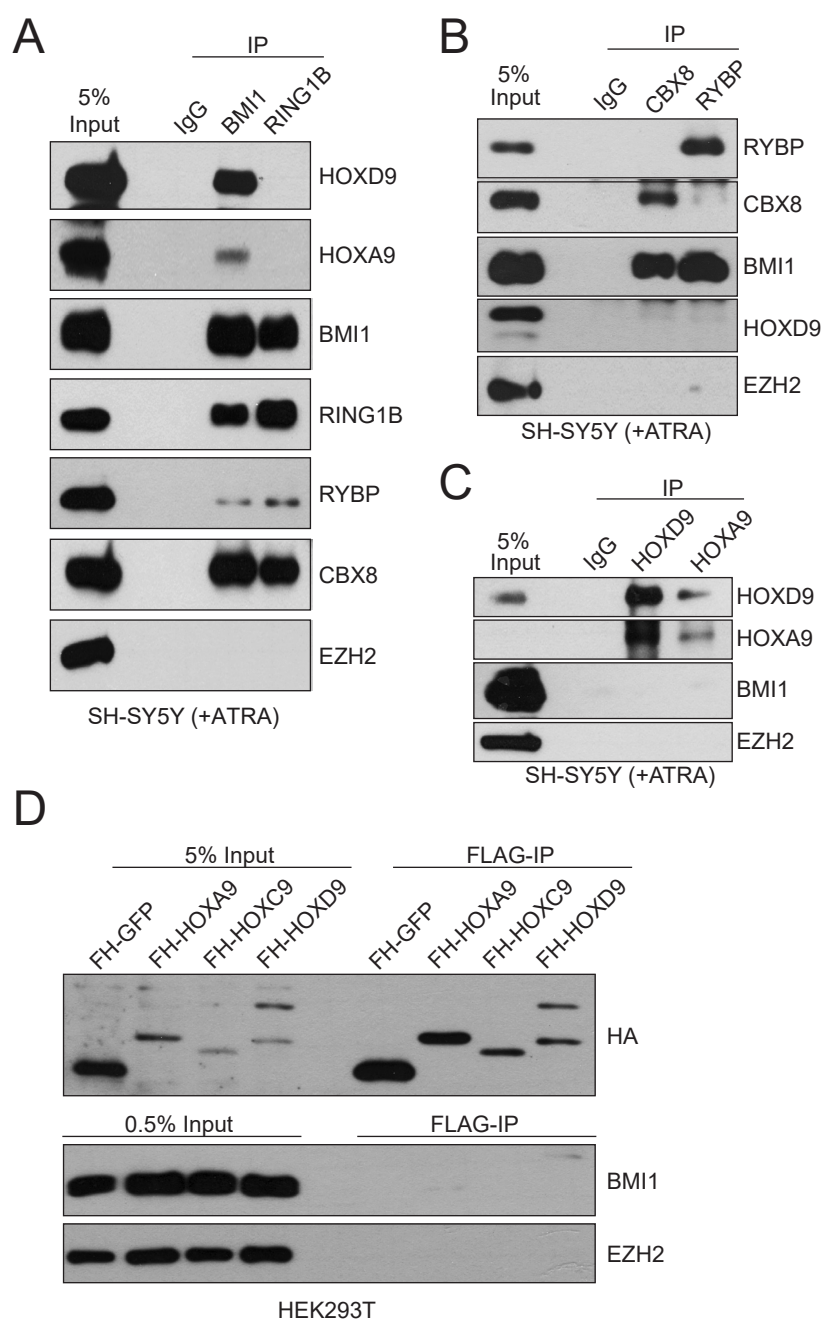


Figure 3.5 HOX-9 proteins do not co-immunoprecipitate BMI1.

(A, B and C) Endogenous IPs, using the indicated antibodies, from SH-SY5Y (+ATRA) nuclear lysates followed by Western blot analysis.

(D) FLAG IPs on whole cell lysates of FLAG-GFP, HOXA9, HOXC9 and HOXD9 transiently expressed in HEK293T cells followed by Western blot analysis.

together, these experiments suggest that HOXA9 and HOXD9 do not interact with BMI1 when it is in either the canonical or the non-canonical PRC1 complexes. An alternative possibility is that my results are artefactual because the anti-BMI1 AF27 antibody (Bracken et al., 2007) non-specifically immunoprecipitates HOX-9 proteins.

Next, I performed endogenous co-immunoprecipitations of HOXA9 and HOXD9, which suggested that they interact with each other, although this may be due to the use of polyclonal HOXA9 and HOXD9 antibodies in the co-immunoprecipitation which could result in cross reactivity between close paralogs such as HOXA9 and HOXD9 (Figure 3.5C). Strikingly, neither HOXA9 nor HOXD9 antibodies were capable of immunoprecipitating BMI1. Further to this, I cloned HOXA9, HOXC9 and HOXD9 into FLAG/ HA tagged expression vectors, transiently expressed them in HEK293T cells and carried out FLAG immunoprecipitations. Western blotting analysis showed that none of the HOX-9 proteins were capable of immunoprecipitating BMI1 (Figure 3.5D).

3.2.6 BMI1 AF27 antibody non-specifically immunoprecipitates HOX-9 proteins.

The comprehensive results showing that neither the canonical or non-canonical BMI1-PRC1 complexes can co-immunoprecipitate HOX-9 proteins, and that HOX-9 proteins cannot co-immunoprecipitate BMI1 led us to question the validity of our earlier observations. Antibodies can have non-specific affinities for background proteins in addition to their intended antigen, so I wanted to determine if the observed interaction was indeed as a result of “background” noise from the anti-BMI1 AF27 antibody. To compare BMI1 co-immunoprecipitations using alternative

antibody epitopes, I generated a stable FLAG-BMI1 expressing SH-SY5Y cell line together with a FLAG-EV control cell line. FLAG co-immunoprecipitations using two independent FLAG-BMI1 expressing clones were successful in immunoprecipitating RING1B but did not enrich for HOXC9 or HOXD9 (Figure 3.6A). Next, using an alternative BMI1 antibody (F6) to carry out endogenous co-immunoprecipitations, in parallel with the BMI1 AF27, I conclusively showed that BMI1 AF27 alone can co-immunoprecipitate HOX-9s while BMI1 F6 cannot (Figure 3.6B).

To further compare the BMI1 AF27 and FLAG-BMI1 co-immunoprecipitations, I performed LC-MS/MS on these eluates followed by plotting of the t-test difference of each experiment over its respective control. This analysis clearly shows that HOXA9, HOXC9 and HOXD9 are highly enriched in BMI1 AF27 IPs, but are not enriched in FLAG-BMI1 immunoprecipitations despite PRC1 complex components, such as RING1B and CBX8 being equivalently enriched in both (Figure 3.6C). A near identical result was observed when comparing co-IP eluates of BMI1 AF27 antibody to BMI1 F6 antibody followed by LC-MS/MS. Both antibodies were successful in pulling down the PRC1 complex, but only the BMI1 AF27 antibody showed enrichment of HOXA9, HOXC9 and HOXD9 (Figure 3.6D). These results all point to the initial interaction between BMI1 and HOX-9 proteins being an unfortunate false positive. I decided to move on to new aspects of the PRC1 proteome project.

3.2.7 Proteomic analysis of RYBP, CBX8 and PCGF2.

Another aim of this chapter was to identify potential non-histone targets of the PRC1 complex due to the reported role for the *Drosophila* homolog of PCGF1-6,

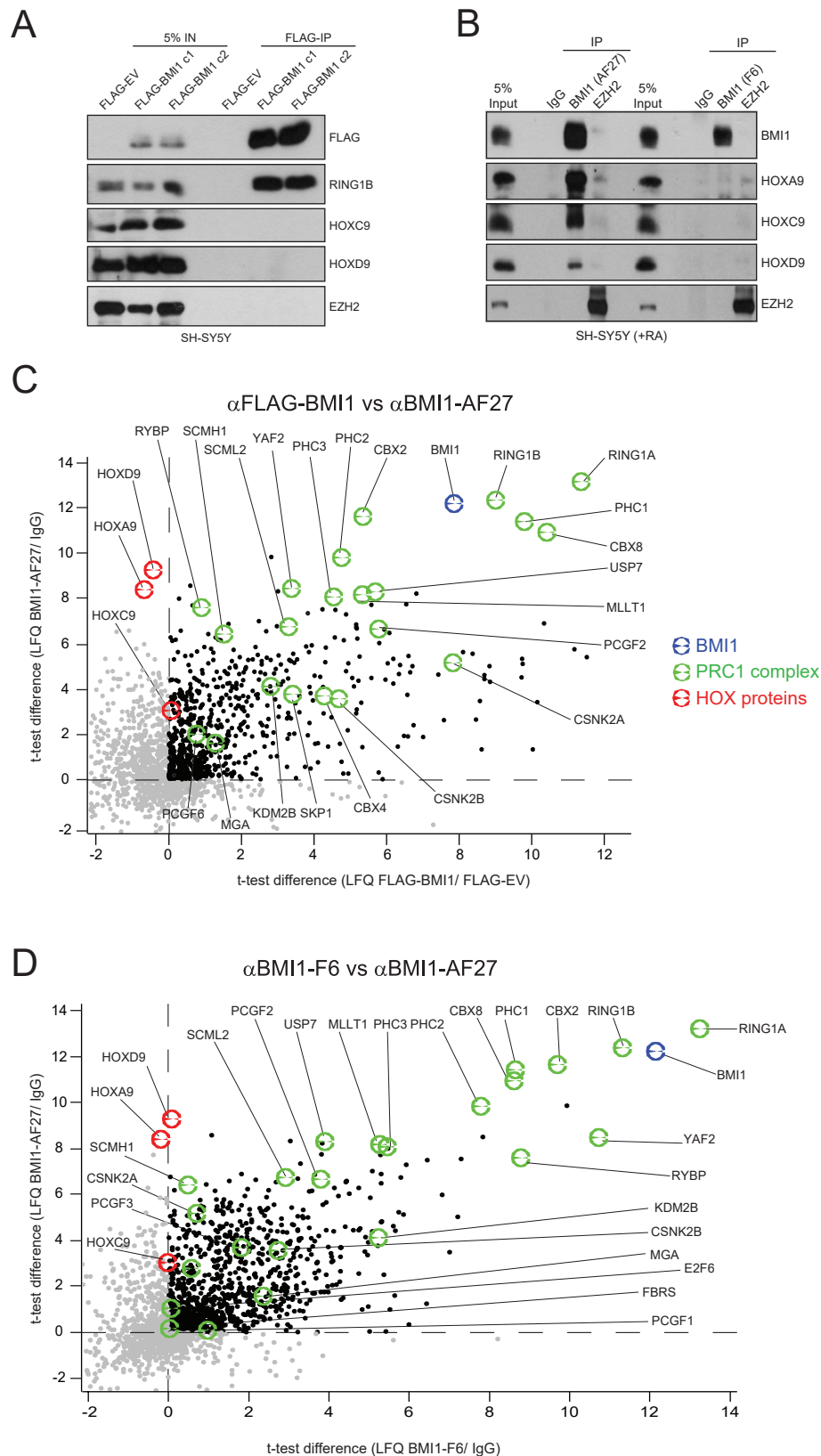


Figure 3.6 BMI1 AF27 antibody non-specifically immunoprecipitates HOX-9 proteins.

(A) FLAG IPs of FLAG-EV and two clones of FLAG-BMI1 stably expressed in SH-SY5Y cells followed by Western blot analysis.

(B) Endogenous IPs, using the indicated antibodies, from SH-SY5Y (+ATRA) nuclear lysates followed by Western blot analysis.

(C) Comparison of the LFQ intensity t-test differences of BMI1 over control IPs between BMI1 AF27 antibody and FLAG-BMI1.

(D) Comparison of the LFQ intensity t-test differences of BMI1 over control IPs between BMI1 AF27 antibody and BMI1 F6 antibody.

Psc, in regulating the cell cycle through ubiquitination mediated degradation of Cyclin B (Mohd-Sarip et al., 2012) Furthermore, BMI1 and RING1B have a possible role in the ubiquitin mediated degradation of p53 (Calao et al., 2013; Su et al., 2013; Wen et al., 2014). To try to identify novel non-histone targets for PRC1 mediated ubiquitination, I performed endogenous co-immunoprecipitations for CBX8, RYBP and PCGF2. These co-immunoprecipitations were subject to LC-MS/MS and each immunoprecipitation was successful in enriching for their respective baits along with known members of the cPRC1 and ncPRC1 complexes (Figure 3.7A-C). Surprisingly, each of these immunoprecipitations also identified members of the Anaphase promoting complex (APC) as being statistically significantly associated proteins of each bait. The APC complex is an E3 ubiquitin ligase complex involved in regulation of the eukaryotic cell cycle (Peters, 2006). This result is particularly surprising as the C terminal portion of dPsc through which it associates with CycB is not conserved in mammalian PCGF1-6 (Mohd-Sarip et al., 2012), however fly Psc does maintain some interaction with the APC, through the Lmg subunit through a region near its RING domain.

To attempt to validate the interaction between the PRC1 complex and the identified APC complex members I performed FLAG-PCGF1-5 and FLAG-RING1B co-immunoprecipitations following transient ectopic expression in HEK293T cells. Western blotting of the resulting eluates show that FLAG-PCGF2, FLAG-BMI1 and FLAG-RING1B successfully co-immunoprecipitated CBX8, as expected (Figure 3.7D). Furthermore, each of the FLAG-PCGF1-5 and FLAG-RING1B immunoprecipitations all enriched for RYBP, as expected, as each of these PRC1 subunits has been reported to be a part of the non-canonical PRC1 complex

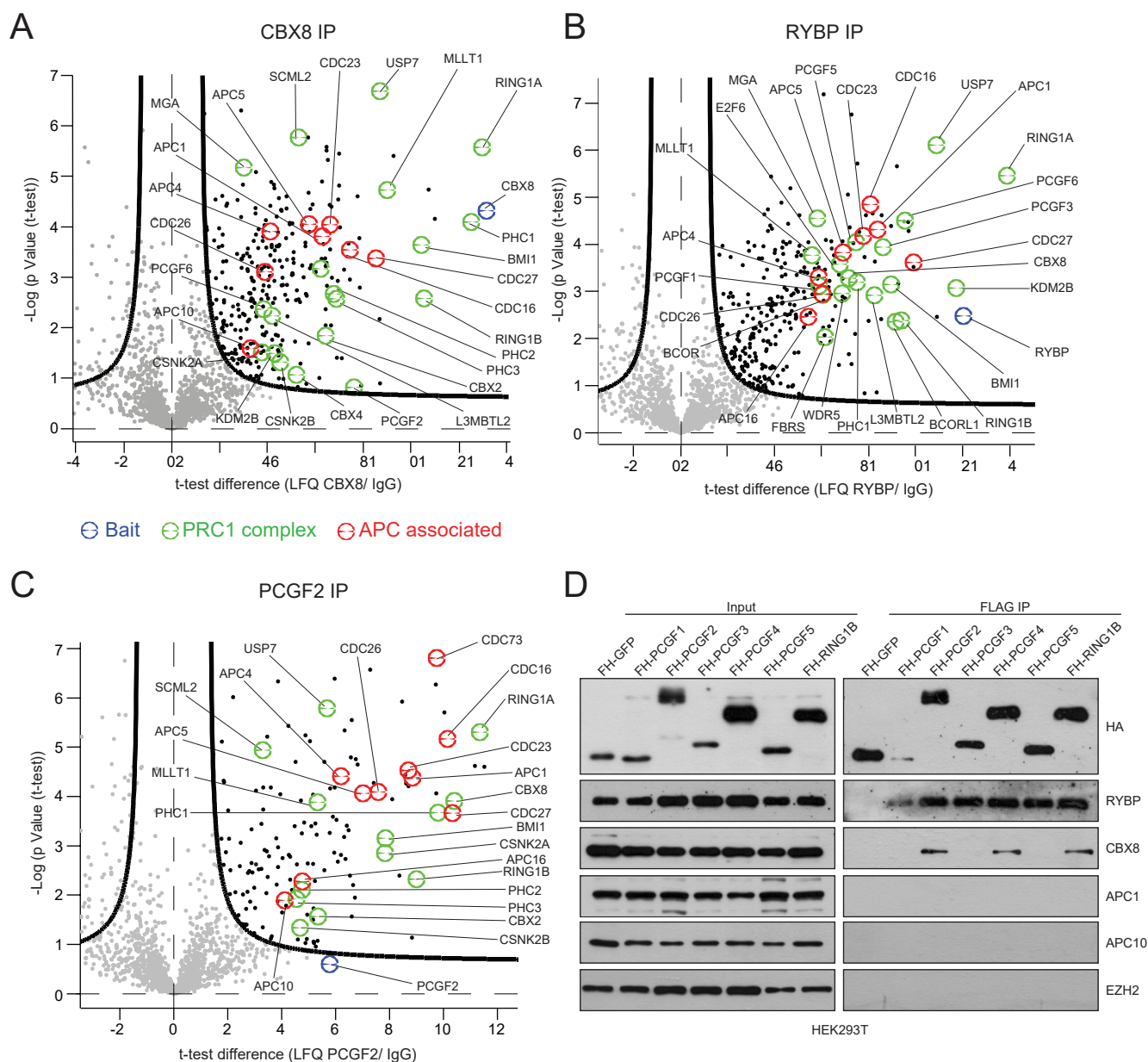


Figure 3.7 Proteomic analysis of RING1B, RYBP, CBX8 and PCGF2.

(A) Identification of RYBP associated proteins following endogenous IP from SH-SY5Y (+ATRA) nuclear lysates. LC-MS analysis was performed using permutation based FDR corrected t-test. The LFQ intensity of the RYBP purification over the EV control is plotted against the $-\log_{10}$ (P-value).

(B) Identification of CBX8 associated proteins following endogenous IP from SH-SY5Y (+ATRA) nuclear lysates. LC-MS analysis was performed using permutation based FDR corrected t-test. The LFQ intensity of the CBX8 purification over the EV control is plotted against the $-\log_{10}$ (P-value).

(C) Identification of PCGF2 associated proteins following endogenous IP from SH-SY5Y (+ATRA) nuclear lysates. LC-MS analysis was performed using permutation based FDR corrected t-test. The LFQ intensity of the PCGF2 purification over the EV control is plotted against the $-\log_{10}$ (P-value).

(D) FLAG IPs on whole cell lysates of FLAG-GFP, PCGF1-5 and RING1B transiently expressed in HEK293T cells followed by Western blot analysis.

previously (Gao et al., 2012; Hauri et al., 2016). Western blotting with antibodies for the Anaphase promoting complex components APC1 and APC10 showed that there was no enrichment of these subunits in the FLAG-PCGF1-5 or FLAG-RING1B immunoprecipitations. This result suggests that either the observed co-immunoprecipitation of PRC1 components with APC complex components is either specific to SH-SY5Y cells, or possibly that the APC subunits are a common LC-MS/MS background noise.

3.3 Discussion

In this chapter, I generated high quality proteome datasets for the cPRC1 subunits BMI1, PCGF2 and CBX8 and for the ncPRC1 subunit RYBP in differentiated SH-SY5Y cells. The resulting LC-MS/MS data is consistent with several already published large scale proteomics studies of PRC1 subunits (Gao et al., 2012; Hauri et al., 2016). The primary difference of my work and that of Hauri et al is that I used endogenous antibodies for each protein of interest, instead of exogenously expressed tagged proteins. The work presented in this chapter also established a modified LC-MS/MS method using “in solution” tryptic digestion instead of “in gel” tryptic digestion. This was done for EZH2 and BMI1 immunoprecipitations and similar results were achieved with both methods. This optimisation of the protocols has led to more cost and time effective LC-MS/MS experiments for multiple subsequent studies (Oliviero et al., 2016; Streubel et al., 2017). In addition, I optimised our existing *in vitro* endogenous co-immunoprecipitation protocol for use *in vivo* (data not shown) which may benefit future projects.

The proteomic resources I generated may also be of use to future projects, but, unfortunately for my own work, I did not detect any novel interactors that were both high scoring and had a previously defined function on chromatin that might suggest interplay with PRC1. I did identify several HOX-9 proteins in the BMI1 proteome but these turned out to be false positives as a result of non-specific recognition of HOX-9 proteins by the BMI1 AF27 antibody. However, these false positives did highlight the need to extensively validate any newly identified protein-protein interaction using several antibodies, to avoid antibody non-specificity, and also to validate interactions using epitope tagged exogenously expressed proteins. In

addition reciprocal immunoprecipitations are essential to confirm any newly identified protein-protein interactions.

In addition, my data suggests that ectopic expression of the HOX-9 proteins has little or no effect on the proliferation rate or the onset of cellular senescence in HMECs. However, ectopic expression of HOXA9 was previously reported to repress *INK4A* and delay the onset of senescence (Martin et al., 2013). It is unclear why this discrepancy between my own work and Martin et al arose. It could perhaps be due to some essential HOX-9 cofactor, such as MEIS2, not being expressed to sufficient levels in HMECs compared to IMR90s, which were used in the other study (Martin et al., 2013).

Despite ruling out a potential interaction between HOX-9s and BMI1, this chapter established that HOX-9s can act as transcriptional repressors in an exogenous assay following artificial recruitment of HOX-9s to a luciferase reporter promoter. This result is in keeping with a previous study (Jung et al., 2014) that identified a repression domain within the C terminus of HOXC9 as required for its ability to repress its target genes during development. This repression domain is independent of the Homeodomain of HOXC9 and its deletion does not prevent HOXC9 binding DNA. Therefore it will be interesting to perform proteomic studies on wild-type and mutant HOXC9 with a repression domain deletion to understand if it is required for association with co-repressors, transcription factors or chromatin remodellers.

Chapter 4: Biochemical and evolutionary analysis of PALI1

4.1 Introduction

Our group previously identified a large novel isoform of the *LCOR* gene, which we called PALI1 (PRC2 Associated LCOR Isoform 1) (Figure 1.6). We discovered that this isoform associates with both the PRC2 complex through its C-terminal region encoded by the exon previously annotated as *C10ORF12*. We also found that PALI1 is capable of promoting PRC2 catalytic activity *in vitro* (Figure 1.6).

In this chapter, my aim was to map and characterise the minimal region required for the interaction between PRC2 and PALI1. We reasoned that careful biochemical and evolutionary analyses of PALI1 would provide insights into what functional domains might be relevant in lymphoma patients with *LCOR/PALI1* locus deletions (Bouska et al., 2014; Chan et al., 2015). Narrowing down the minimal PRC2 interaction region of PALI1 could help us design targeted peptides/ inhibitors to disrupt the interaction between PALI1 and PRC2. Such inhibitors may be of use in lymphoma patients with change of function *EZH2* mutations, considering cell lines harbouring these mutations can acquire resistant to SET domain inhibitors (Baker et al., 2015; Gibaja et al., 2016). In addition narrowing down the domains required for the interaction of PRC2.1 proteins, like PALI1, with the core complex will help us understand the way the distinct PRC2.1 and PRC2.2 complexes are formed.

A further aim of this chapter was to carry out careful evolutionary analysis of *LCOR* and PALI1. *Drosophila* share Pcl (dPcl), Jarid2 (dJmj) and Aebp2 (dJing) subunits with mammalian PRC2 (Kalb et al., 2014), but they have a unique mechanism of recruiting the complex to its target sites on chromatin. The *Drosophila* Polycomb system includes a DNA binding factor called dPho which recruits PRC2 to its target Polycomb Response Elements (PREs) (Mohd-Sarip et al., 2005; Muller and Kassis,

2006). Mammals lack this Pho recruitment mechanism but still have specific recruitment of PRC2 to developmentally regulated promoters (Bracken et al., 2006; Riising et al., 2014; Vella et al., 2012). We wanted to determine if *Drosophila* and other invertebrates have any orthologous proteins to LCOR or PALI1 or if they arose later during mammalian or eutherian evolution as is the case with the PRC2.1 subunit EPOP (Beringer et al., 2016). In addition, conservation analysis of residues in the minimal interaction region of PALI1 will help to map the residues that are essential for the interaction with PRC2.

These analyses identified a highly conserved mammalian specific region of the PALI1 protein that is essential for its association with PRC2, which we called the PIP domain (PRC2 interaction with PALI1). Mutational analysis of the PIP domain identified two conserved tryptophans that have a synergistic role in maintaining the interaction between PALI1 and PRC2. The evolutionary analysis of LCOR and PALI1 uncovered a previously unannotated paralog of PALI1, which we called PALI2, which is an alternative spliceform of the *LCORL* gene locus. This PALI2 protein shares the PRC2 binding ability of PALI1 through its conserved PIP domain suggesting it can also form a part of the PRC2 complex.

4.2 Results

4.2.1 Mapping the minimal region of PALI1 required for interaction with PRC2.

To further delineate the potential interaction domains within the PALI1 protein for its various binding partners, four truncations of the C10ORF12 region of PALI1 were generated, transiently expressed them as FLAG-HA-tagged proteins in HEK293T cells and performed FLAG immunoprecipitations (Figure 4.1A, data courtesy of Emilia Jerman). As expected, EZH2, PCL1 and G9A immunoprecipitated with the C10ORF12 region, whereas BMI1 did not, while two non-overlapping fragments of C10ORF12, Fragment 1 and Fragment 4, immunoprecipitated the G9A and PRC2.1 complexes, respectively (Figure 4.1B). This indicates that the G9A and PRC2.1 complexes bind to PALI1 via distinct regions. Consistent with these G9A and PRC2.1 interaction regions being potentially important, a disorder tendency analysis of the complete PALI1 protein predicted two 'ordered' domains; one mapping to the G9A interaction region and the other to Fragment 4, which we call the PIP (Pali interacting with PRC2) domain from here on (Figure 4.1E). The identification of two distinct binding regions for these complexes is intriguing as the G9A complex and PRC2 have previously been reported to interact with each other (Mozzetta et al., 2014). This suggests that PALI1 may act as a bridging protein for these two complexes. In addition, PRC2 depends on G9A complex activity for its own recruitment and catalytic activity at a cohort of PRC2 target genes (Mozzetta et al., 2014). Our data establish that, in addition to CTBP binding motifs and a nuclear receptor binding box in its N-terminus, the PALI1 protein also contains regions within its C-terminus that allow it

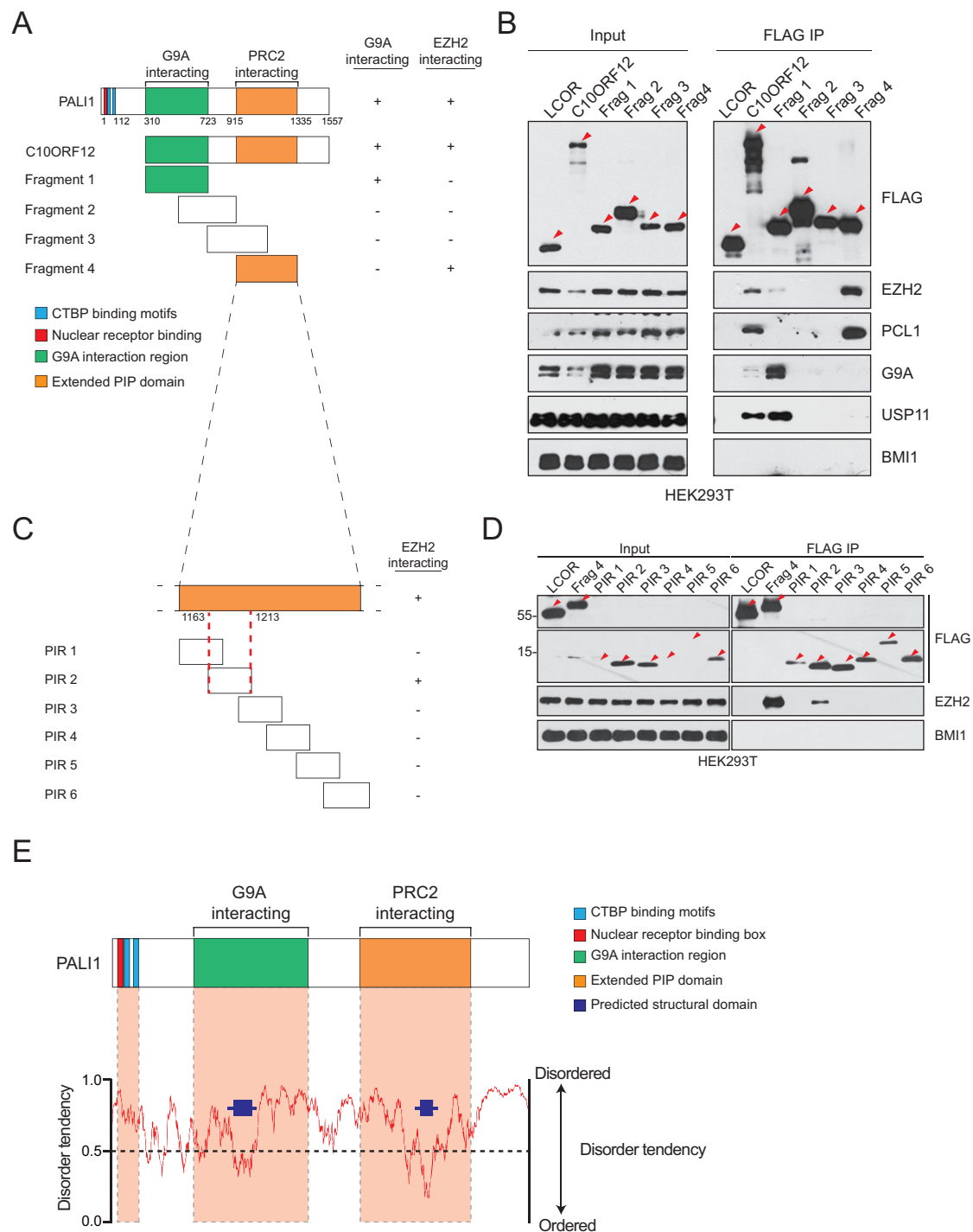


Figure 4.1 Mapping the minimal region of PALI1 required for interaction with PRC2.

(A) Left: A schematic representation of full-length PALI1 and the 4 fragments cloned and expressed as FLAG tagged proteins in HEK293T cells. Right: A summary table indicating the status of interaction of each fragment with EZH2 and G9A.

(B) FLAG-IPs of the indicated proteins (LCOR, C10ORF12 or fragments of PALI1 as shown in panel B), transiently expressed in HEK293T cells, followed by western blotting analysis for the indicated proteins (Experiment performed by Emilia Jerman).

(C) Left. Map of 6 small fragments of the PALI1 PIP domain (PRC2 interactions regions- PIR) cloned for use in FLAG-IPs. Right. Table indicating the status of interaction of each fragment with EZH2.

(D) FLAG-IPs of the indicated proteins (LCOR or small regions of PALI1 extended PIP domain as illustrated in panel A) transiently overexpressed in HEK293T cells, followed by western blot analysis.

(E) Predicted disorder tendency plot of the PALI1 amino acid sequence. The disorder plot was generated using the IUPRED disorder prediction software (<http://iupred.enzim.hu>).

to associate with several different chromatin regulators, including the SET protein, deubiquitinases and the G9A and PRC2.1 complexes.

To define the minimal PIP domain within PALI1, I generated additional truncations of the extended PIP domain (Figure 4.1C). This identified a 51 amino acid PRC2 interaction region (PIR2) sufficient to immunoprecipitate EZH2, but not BMI1 (Figure 4.1D). Although this PIR2 fragment was less efficient at immunoprecipitating EZH2 than Fragment 4.

4.2.2 Identification of a novel isoform transcribed from the *LCORL* gene locus (*PALI2*) that shares the Polycomb interaction region of PALI1.

A protein homology search of the PIR2 region identified only one other homologous human protein; an orthologous sequence within a predicted splice-form of the human *LCORL* gene (XM_011513822.2), the paralog of *LCOR*. I therefore named the predicted protein PALI2. Multiple sequence alignments of the mouse and human PALI1 and PALI2 proteins show high conservation of residues within this PIR2 region (Figure 4.2A). To validate the existence of the predicted PALI2 protein, I performed a detailed RT-PCR mapping analysis of all the *LCORL* exons on cDNA from HEK293T cells (Figure 4.2B). This confirmed that the *PALI2* specific exon 8 (with homology to the extended PIP region of PALI1) exists within an alternative spliceform of *LCORL*. This spliceform contains exons 1-6, but lacks exon 7, which contains the nucleotide sequence encoding the helix turn helix (HTH) domain of *LCORL*. Therefore, the organisation of the PALI2 coding exons within the *LCORL* gene locus is similar to that of the organisation of PALI1 within the *LCOR* gene locus.

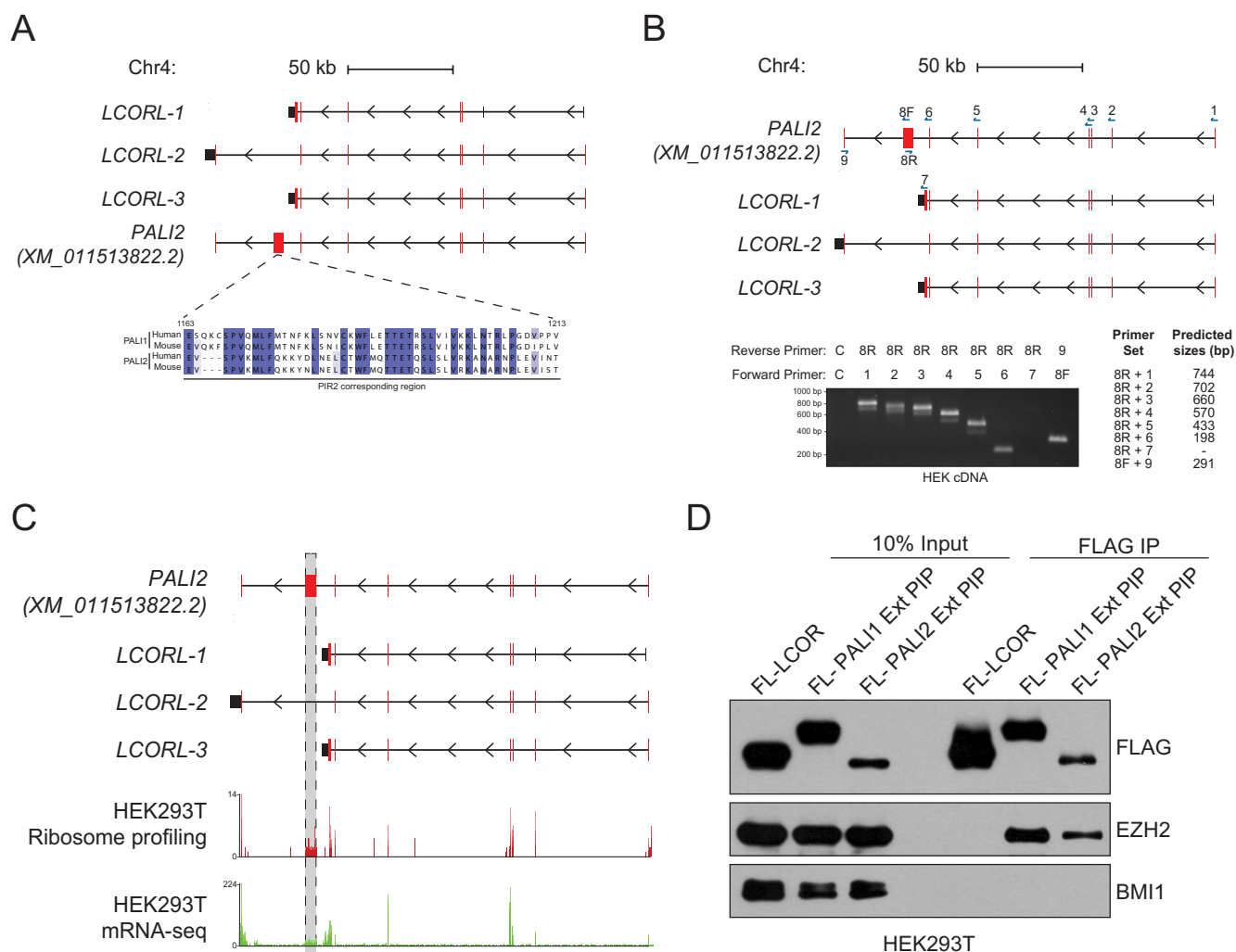


Figure 4.2 Identification of a novel isoform transcribed from the LCORL gene locus (PALI2) that shares the Polycomb interaction region of PALI1.

(A) Top. Genome view of the human chromosome 4 showing 3 annotated isoforms of the LCORL (LCOR-Like) gene and the newly identified PALI2 isoform encoded at the same gene locus. Bottom. Multiple sequence alignments of the minimal PRC2 interaction region (PIR2) in both PALI1 and PALI2. The percentage of the residues that agree with the consensus sequence is indicated by color.

(B) PCR mapping of the PALI2 isoform. Left. Geneview of the LCORL gene locus showing the primers for PCR. Right. PCR using the indicated primers using HEK293T cDNA.

(C) Ribosome profiling and mRNA seq alignments from HEK293T cells at the LCORL/ PALI2 locus. The PALI2 specific exon is highlighted in the grey box.

(D) FLAG-IPs of the indicated proteins (LCOR, extended PIP domain of PALI1 and the orthologous region of PALI2) transiently expressed in HEK293T cells, followed by western blotting analysis using the indicated antibodies.

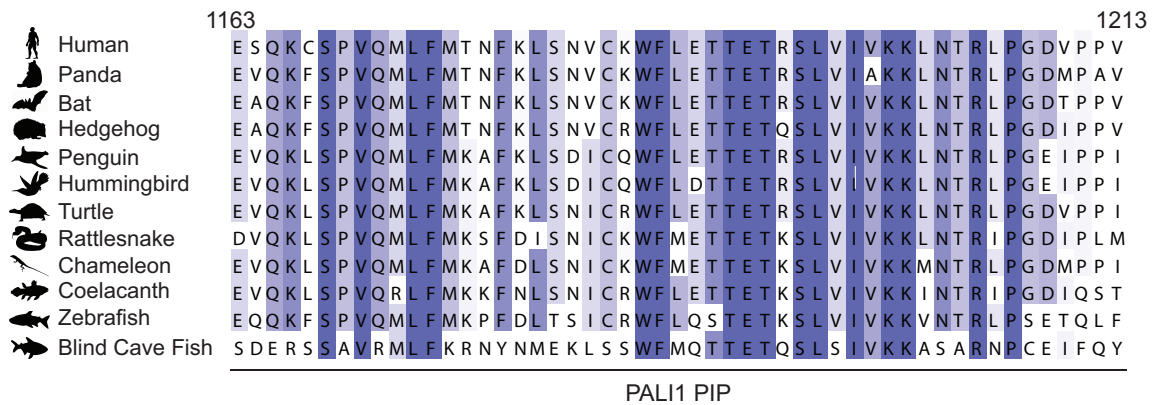
To determine whether the *PAL12* specific exon 8 of the *LCORL* gene locus undergoes translation into a protein product I analysed existing ribosome profiling data sets. Ribosome profiling is a technique that isolates mRNA following ribonuclease treatment. During this ribonuclease treatment actively translated RNAs are not degraded due to protection of the RNA by the ribosome. This is followed by isolation of ribosome complexes by density gradient centrifugation and subsequent RNA-seq of the remaining mRNA (Ingolia et al., 2012). Ribosome profiling from HEK293T cells (Iwasaki et al., 2016) shows that each of the 9 exons of the *LCORL* gene locus undergoes translation, including the *PAL12* specific exon, demonstrating that it likely translates to a protein product (Figure 4.2C).

To confirm that the *PAL12* PIP domain is also capable of associating with PRC2, I isolated the paralogous sequence of the extended *PAL11* PIP region from *PAL12* and expressed it, together with the extended PIP region of *PAL11*, as FLAG-HA-tagged proteins in HEK293T cells (Figure 4.2D). FLAG-immunoprecipitations confirmed that it is also capable of immunoprecipitating EZH2, but not BMI1.

4.2.3 The PIP domain of *PAL11* and *PAL12* arose during vertebrate evolution

I next wanted to investigate whether the PIP domain of both *PAL11* and *PAL12* were conserved to determine if they have evolutionary important functions, and also to identify potentially functionally important amino acids. Unfortunately, due to a lack of annotation of *PAL12* in many species this initial analysis was limited to *PAL11*. Using the uniprot database, I aligned vertebrate *PAL11* C terminal amino acid sequences and found that the PIP domain was highly conserved, contained potentially interesting conserved aromatic residues, and was flanked by a proline

A



B

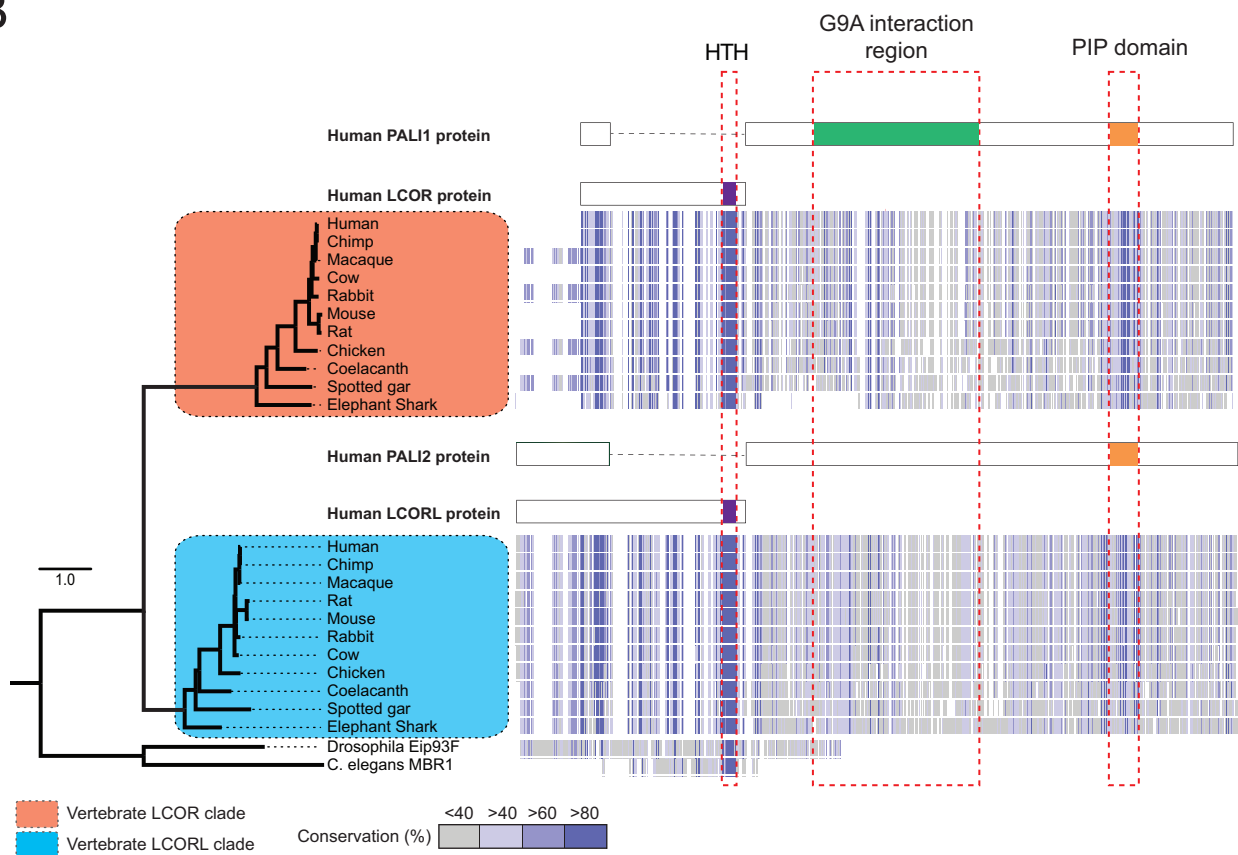


Figure 4.3 The PIP domain of PALI1 and PALI2 arose during vertebrate evolution.

(A) Amino acid alignments of the PIP domain of PALI1 orthologs from a number of vertebrate species.

(B) A maximum-likelihood phylogenetic tree (left) inferred from aligned amino acid sequences of LCOR and LCORL exons (right). The percentage of the residues that agree with the consensus sequence is indicated by color. Positions corresponding to either the human LCOR/LCORL and PALI1/2 proteins are indicated above the alignments. Analysis carried out in collaboration with Alan Rice and Aoife McLysaght.

which could potentially demarcate the end of a structured region of the protein (Figure 4.3A).

In order to analyse the conservation and evolution of the LCOR/LCORL and PALI1/2 proteins in more depth, amino acid sequences of all the possible coding exons of all four proteins across several vertebrate and invertebrate species were aligned (Figure 4.3B), in collaboration with Alan Rice and Aoife McLysaght (Smurfit Institute of Genetics, TCD). This revealed that invertebrates have a single ortholog of LCOR, which shares the highly conserved helix turn helix (HTH) region with the vertebrate LCOR and LCORL proteins. However, invertebrate species lack the PIP domain and the large C-terminus present in both the PALI1 and PALI2 proteins, suggesting these regions arose after the divergence of the vertebrate and invertebrate common ancestor. These alignments were generated after an initial search of all species within the uniprot database (506,914 total species from all kingdoms) found that there were no invertebrate orthologs of the PALI C terminus. A subsequent TBLASTN search for the PALI C terminal sequence was carried out to validate this result in a number of invertebrate species including *Drosophila*, *C. elegans*, *C. briggsae*, *Apis mellifera*, *Plasmodium*, *Ciona intestinalis* and *Amphioxus* with no ortholog of the PALI C terminus being found. As both *C. intestinalis* and *Amphioxus* regarded as relatively close to the vertebrate-invertebrate boundary this confirmed the initial finding that PALI is vertebrate specific. Interestingly, the PIP domain of PALI1 and PALI2 is highly conserved within vertebrate species, suggesting potential evolutionary importance.

4.2.4 Identification of two aromatic residues that are essential for the interaction between the PIP domain and PRC2.

In order to further functionally dissect the PIP domain of PALI1 and PALI2, I closely analysed the amino acid and charge conservation of their amino acid sequences (Figure 4.4A). Two highly conserved tryptophans (W1125 and W1186) were selected for mutation due to the known importance of aromatic residues in protein-protein interactions and for structural stability within a modular domain (Mullin et al., 2017). I reasoned the best chance of disrupting the interaction of PALI1 with PRC2 would be to mutate these aromatic residues as the PFAM analysis of the 'domain of unknown function' DUF4553 within C10ORF12 shows these two tryptophans are the two highest conserved residues among DUF4553 containing proteins. It should be noted that all of the DUF4553 containing proteins are orthologs of C10ORF12. Mutagenesis of these tryptophans to alanines individually did not disrupt the interaction with EZH2 (Figure 4.4B). However, the double mutation of W1125A and W1186A led to an almost complete loss in the ability of the PALI1 extended PIP region to immunoprecipitate EZH2, suggesting the two tryptophans are synergistically important for the interaction of the PIP domain with PRC2. Interestingly, the PIR2 fragment used to immunoprecipitate PRC2 (Figure 4.1D) did not include W1125, potentially explaining why it was weaker at immunoprecipitating EZH2 than Fragment 4. The PALI1 and PALI2 PIP domains also contain a number of other highly conserved residues, including a basic (positively charged) patch (Figure 4.4A), which may also be involved in the interaction with PRC2. Despite the shared PIP domain and CTBP binding motifs of PALI1 and PALI2, the proteins have diverged somewhat in their protein organisation. For example, PALI2 and LCORL do not share the nuclear receptor

binding box present in PALI1 and LCOR, while PALI2 has three predicted AT-hooks in its C terminus, which could confer the ability to bind to DNA (Figure 4.4C).

Taken together, I show that PALI1 is the founding member of a new family of proteins that share CTBP binding domains in their N-termini and bind PRC2.1 through a novel PIP domain in their C-termini (Figure 4.4C).

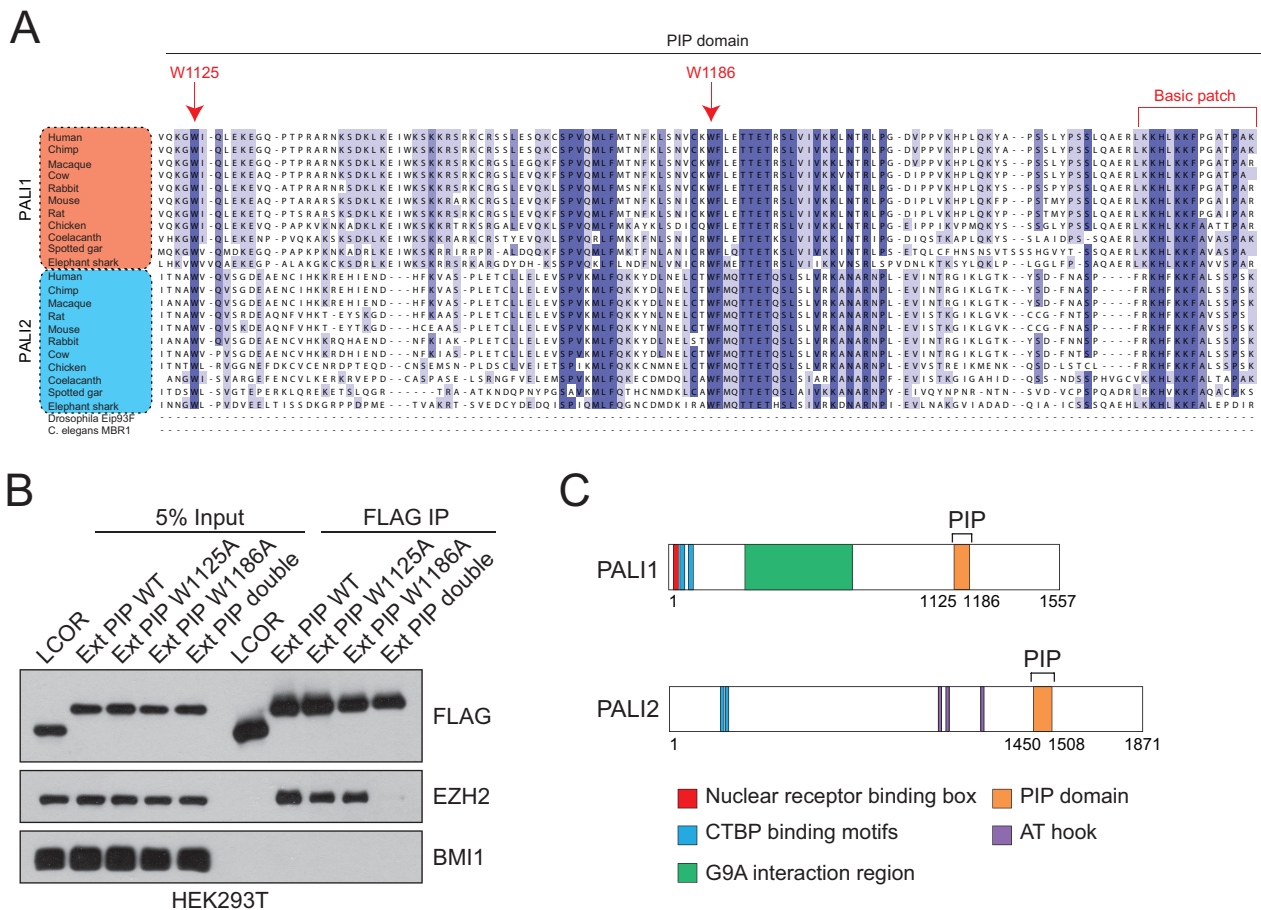


Figure 4.4 Identification of two aromatic residues that are essential for the interaction between the PIP domain and PRC2.

(A) Amino acid alignments of the PALI interaction with PRC2 (PIP) domain of PALI1 and PALI2 proteins from several vertebrate and invertebrate species. Two conserved tryptophans and a basic charged patch are indicated.

(B) FLAG-IPs of the indicated proteins (LCOR, or PALI1 extended PIP WT or with the indicated mutations) transiently expressed in HEK293T cells, followed by western blot analysis using the indicated antibodies.

(C) An illustration summarizing both the shared and unique domains and regions in the PALI1 and PALI2 proteins.

4.3 Discussion

The aim of this chapter was to biochemically characterise the interaction between PALI1 and PRC2 and explore its emergence during evolution. I mapped the minimal region of PALI1 required to bind to PRC2 (PIP domain) (Figure 4.1). Strikingly, the PIP domain is only present in one other mammalian protein, an uncharacterised *LCORL* isoform, which I called PALI2 (Figure 4.2). Therefore, I defined PALI1 and PALI2 as a new family of sub-stoichiometric subunits of the PRC2 complex. My mapping of the PIP domain is the first step towards characterising a protein-protein interaction region that could potentially be disrupted with small molecule inhibitors or stapled peptides. This is potentially important as we have already shown that PALI1 greatly increases PRC2 catalytic activity *in vitro* (Figure 1.6G). Such inhibitors of the PALI1-PRC2 interaction could provide alternative strategies to modulate PRC2 function in cancer therapies. Further steps down the route of developing PIP domain inhibitors would require in depth structural analysis of the PIP domain and its interaction with PRC2 through structural studies.

Our groups discovery of PALI1 and PALI2 added to the recent discovery of EPOP, these additional subunits further expand the Polycomb system in vertebrates. My finding that the PIP domain of PALI1 and PALI2 is exclusive to vertebrate species, despite the existence of an *LCOR* homolog in invertebrates, suggests an emergence of the PALI specific exons after the divergence of vertebrates and invertebrates. Similarly, the Di Croce group reported that EPOP is a eutherian specific PRC2 subunit (Beringer et al., 2016). This demonstrates a specific expansion of the PRC2.1 complex during vertebrate evolution which is a discovery

of potential functional interest as *Drosophila* have PCL, JARID2 and AEBP2 orthologs but not EPOP and PALI1, while the mammalian ortholog of Pho, YY1, does not target PRC2 to PREs like it does in *Drosophila* (Vella et al., 2012). It is possible that the PALI1/2 proteins and/ or EPOP could act as functional, not genetic, orthologs of the Pho transcription factor to coordinate PRC2 recruitment at target loci.

The discovery of PALI1/2 also opens up a potential link between PRC2 and CTBP and/ or nuclear receptor signalling. LCOR has previously been linked to nuclear receptors through its N terminal nuclear binding box (Fernandes et al., 2003), which is shared with PALI1. Furthermore, since PALI1 and PALI2 both have two CTBP binding motifs, it is possible that they act as physical links for PRC2 to interplay with these co-repressors and nuclear receptors. Supporting this it has been reported that CTBP proteins can associate with PRC2 (Kim et al., 2015b). It will be interesting to explore this potential link between PRC2 and CTBP/ nuclear receptor complexes and their potential roles in targeting it to chromatin.

Chapter 5. Characterisation of the functional and biochemical antagonism between Pali1 and Aebp2

5.1 Introduction

Although little is known about the divergent functions of the PRC2.1 and PRC2.2 complex subtypes, the sub-stoichiometric subunits of each have all been characterised as having roles in either promoting their histone methyltransferase activity or facilitating their interactions with other proteins. For example, the Polycomb-like proteins contribute to PRC2.1 association with chromatin via their tudor and extended helix domains, and stimulate its methyltransferase activity *in vitro* (Ballare et al., 2012; Brien et al., 2012; Brien et al., 2015; Cai et al., 2013; Li et al., 2017; Musselman et al., 2012a; Sarma et al., 2008). The recently discovered Epop subunit of PRC2.1 bridges its interaction with ElonginB/C proteins, but, unlike Polycomb-like proteins, it does not appear to be essential for the complexes association with chromatin (Beringer et al., 2016; Liefke et al., 2016).

In terms of the PRC2.2 complex, its two specific subunits, Aebp2 and Jarid2, promote its histone methyltransferase activity *in vitro* (Cao and Zhang, 2004; Li et al., 2010). While the precise role of Aebp2 in PRC2.2 is unclear, Jarid2 has recently been proposed to contribute to targeting the complex to chromatin through its ability to bind to PRC1 mediated H2AK119ub1 via its Ubiquitin interaction motif (Cooper et al., 2016). Consistent with this, the *in vitro* histone methyltransferase activity of PRC2.2 is enhanced in the presence of H2AK119ub1 modified nucleosomes (Kalb et al., 2014). However, key open questions include why there are at least two forms of the PRC2 complex and if they can act with each other to regulate Polycomb target genes during cell fate transitions. In this context, it became interesting for me to explore the role of the newest members of PRC2.1, PALI1 and PALI2.

Some intriguing and potentially illuminating clues as to the roles of PRC2.1 and PRC2.2 complexes have come from recent studies of Epop and Aebp2 in mouse embryonic stem cells (ESCs) (Beringer et al., 2016; Grijzenhout et al., 2016). Intriguingly, while depletion of Polycomb-like and Jarid2 in ESCs correlate with partial reductions in the levels of both Ezh2 and H3K27me3 at Polycomb target genes (Ballare et al., 2012; Brien et al., 2012; Holoch and Margueron, 2017; Pasini et al., 2010a; Peng et al., 2009), the loss of either Epop or Aebp2, members of PRC2.1 and PRC2.2 respectively, correlate with increased H3K27me3 (Beringer et al., 2016; Grijzenhout et al., 2016). These paradoxical results are inconsistent with the fact that recombinant Epop and Aebp2 proteins promote core PRC2 histone methyltransferase activity *in vitro* (Cao and Zhang, 2004; Zhang et al., 2011). However, supporting the observations in Aebp2 null ESCs, *Aebp2* null mice have a “Trithorax” phenotype, which is considered the opposite of a Polycomb phenotype, and classically associated with decreased expression of *Hox* genes (Grijzenhout et al., 2016). These new findings highlight how little is currently known about the respective roles of the non-core PRC2 components and also points to potential opposing activities within PRC2 complexes. In this chapter I characterise the role of PALI1 during early mouse development, investigate its mechanistic role within PRC2.1 and determine its effect on histone methyltransferase activity *in vivo*.

5.2 Results

5.2.1 Pali1 knockout strategy in mouse embryonic stem cells.

To investigate the importance of Pali1 for mouse development, conditional knockout mouse ESCs in which the Pali1 specific exon (aka *Gm340*; the mouse ortholog of *C10ORF12*) is flanked by LoxP sites were generated (Figure 5.1A) by targeting each allele for recombination with a homologous transgene with additional LoxP sites, FRT sites and antibiotic selection cassettes as shown. Recombination between these LoxP sites, when induced by the addition of TAT-Cre, led to the deletion of the Pali1 specific (*Gm340*) exon, but did not affect the coding exons of *Lcor*. The exclusive deletion of the Pali1 specific exon was confirmed by genotyping PCR and RT-qPCR (Figure 5.1B, C and D). This gene targeting and subsequent *in vivo* work was conducted in collaboration with Haruhiko Kosekis and his group (RIKEN institute, Japan).

5.2.2 Pali1 is essential for mouse development and promotes H3K27me3 *in vivo*.

Mice heterozygous for the Pali1 specific exon deletion (*Pali1*^{+/-}) (Figure 5.2A and B) were crossed and the resulting embryos were genotyped and counted at various stages of development (Figure 5.2C). At E11.5 (11.5 days post-coitus) *Pali1*^{-/-} mice had the expected Mendelian ratio compared to matched wild-type and heterozygous mice. However, none of the *Pali1*^{-/-} mice were viable at birth. Protein extracts from E11.5 embryos from *Pali1*^{-/-} mice were next analysed by Western blotting to determine if the lethality could be related to changes in the levels of PRC2 components or H3K27me3. Whereas levels of Ezh2, Lcor, H3K9me2 and PRC1 mediated H2AK119ub1 were unchanged in *Pali1*^{-/-} E11.5 embryos, a global

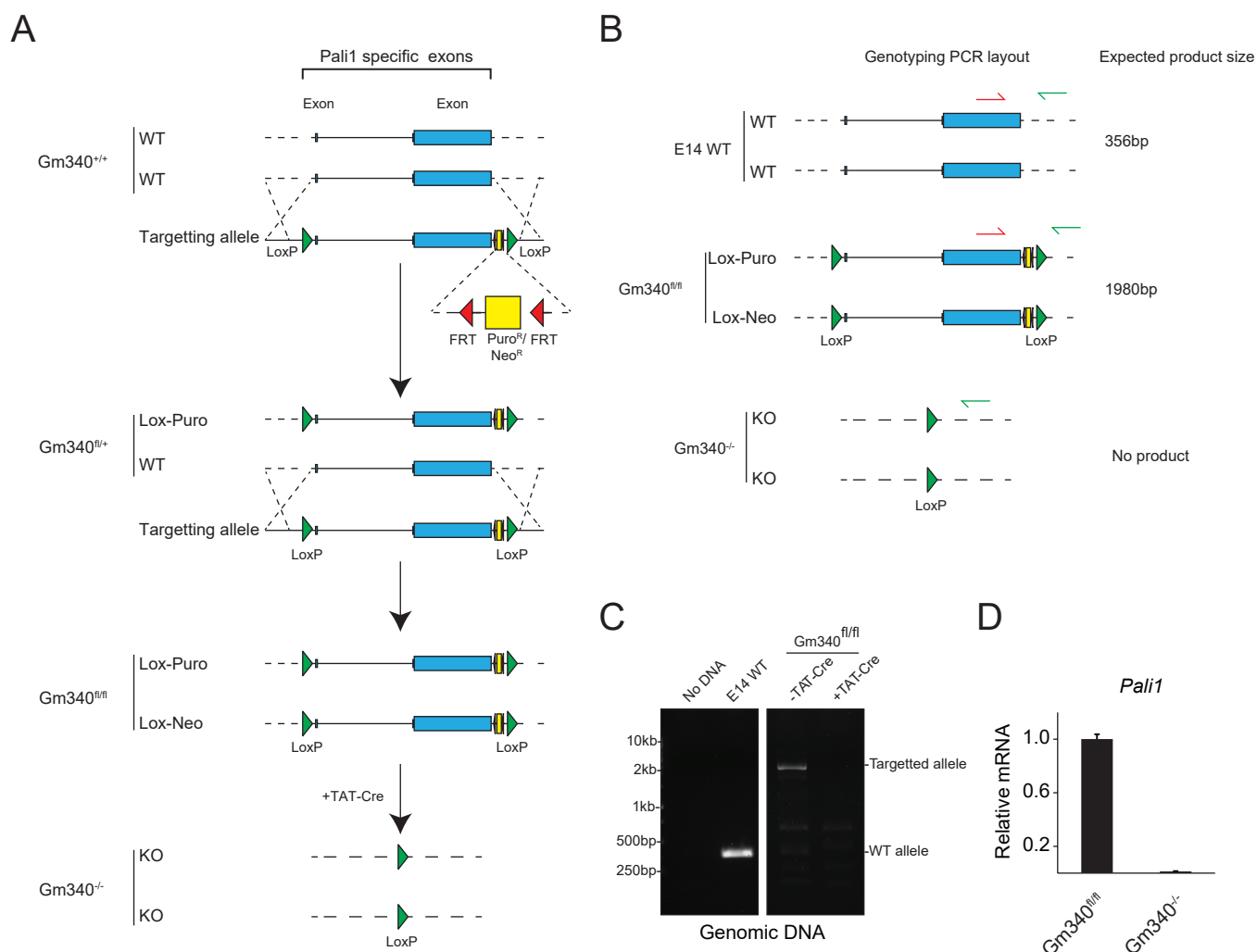


Figure 5.1 Pali1 knockout strategy in mouse embryonic stem cells.

(A) Illustration of the strategy used to target the Pali1 specific exons (aka Gm340) for Cre-Lox mediated knockout in mouse embryonic stem cells. Gene targeting carried out in collaboration with Haruhiko Koseki.

(B) Illustration of the location of PCR primers for genotyping Gm340^{fl/fl} and Gm340^{-/-} cells from genomic DNA, along with expected PCR product sizes.

(C) Genotyping PCR using the primers indicated in (B).

(D) RT-qPCR of cDNA from Gm340^{fl/fl} and Gm340^{-/-} cells to confirm knockout of Pali1 specific exons. Pali1 primers are located within the floxed region. RT-qPCR results are normalised to RPLPO cDNA levels.

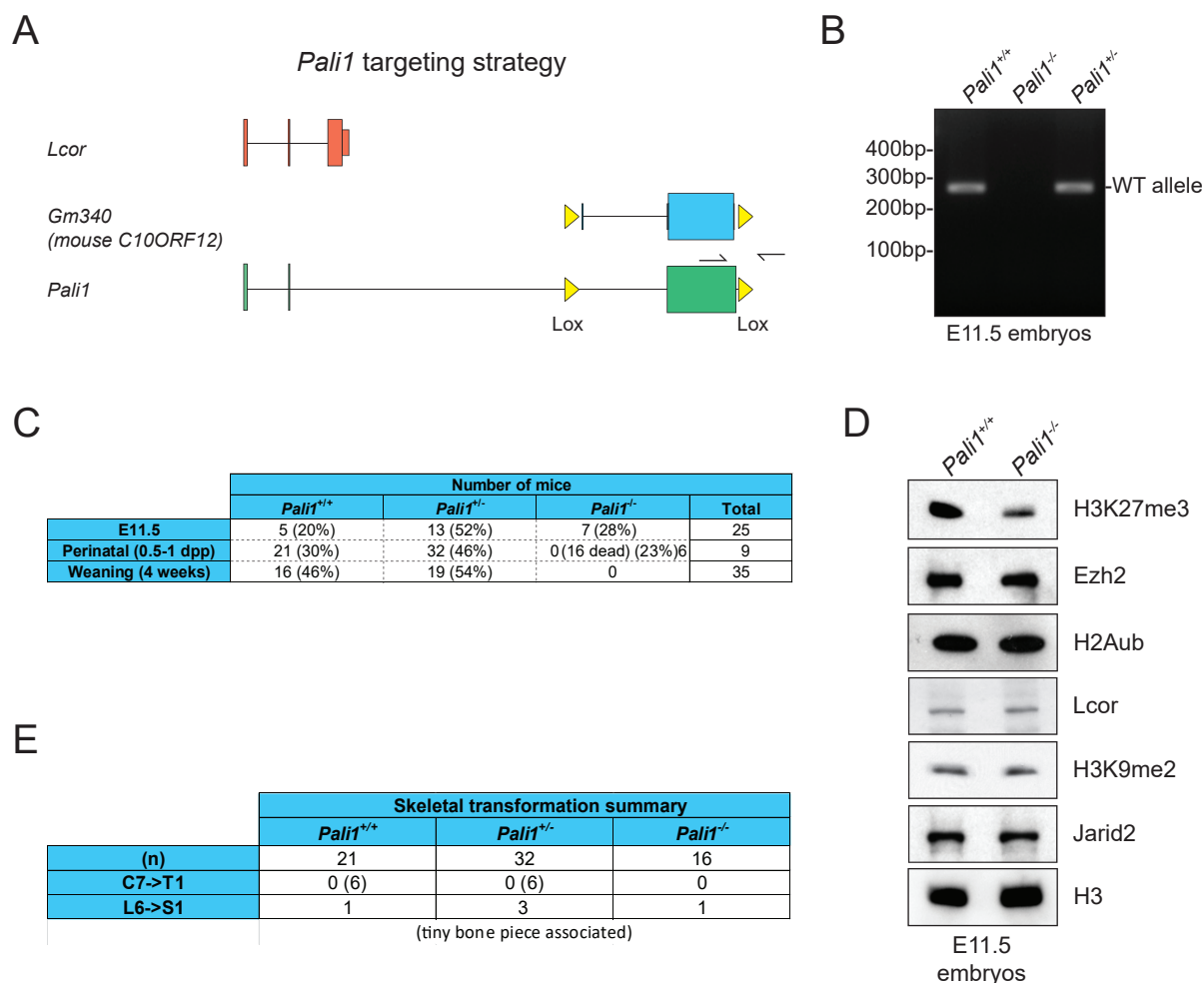


Figure 5.2 *Pali1* is essential for mouse development and promotes H3K27me3 in vivo.

(A) The strategy used to target the *Pali1* specific exon (aka Gm340) for Cre-Lox mediated knockout in mouse ESCs

(B) Genotyping PCR, using the indicated primers in (A), shows deletion of the *Pali1* allele in *Pali1*^{-/-} mice embryos (E11.5).

(C) Table showing number of mice at various stages of mouse development following a cross of *Pali1*^{+/-} mice.

(D) Western blots on whole cell lysates harvested from E11.5 *Pali1*^{+/+} or *Pali1*^{-/-} embryos, with the indicated antibodies.

(E) Summary table of skeletal transformations observed in *Pali1*^{+/-} cross progeny.

In vivo experiments performed in collaboration with Shinsuki Ito, Yoko Koseki, Manubu Nakayama, Tomoyuki Ishikura and Haruhiko Koseki.

reduction in the levels of H3K27me3 was observed compared to wild-type (Figure 5.2D), consistent with Pali1 being an important component of the PRC2.1 complex. However, no recurrent skeletal transformations were observed in *Pali1*^{-/-} embryos at the E11.5 stage, potentially due to functional redundancy between the Pali1 and Pali2 proteins (Figure 5.2E). This lethal phenotype could be linked to the potential role of Pali1 in G9a/ Glp and Ctbp2 containing complexes as deletion of any of the genes encoding these three proteins, similarly to Eed and Suz12 knockout, leads to embryonic lethality (Faust et al., 1995; Hildebrand and Soriano, 2002; Pasini et al., 2004; Tachibana et al., 2002; Tachibana et al., 2005). Nonetheless, our results establish Pali1 as an essential new protein required for development and link reduced levels of PRC2 mediated H3K27me3 with its perinatal lethal phenotype.

5.2.3 Pali2 knockout strategy using CRISPR-Cas9 mediated gene editing.

In order to study the function of Pali1 and Pali2 in ESCs without any potential masking of phenotypes by functional redundancy between the two proteins I generated *Pali1*^{-/-}, *Pali2*^{-/-} (*Pali1/2* dKO from here) ESCs using CRISPR/Cas9 gene editing to introduce out of frame deletions in the Pali2 specific exon, upstream of the PIP domain in the *Pali1*^{-/-} (*Pali1* sKO from here) ESCs (Figure 5.3A). The out of frame deletions were confirmed by Sanger sequencing of the region surrounding the PAM site (Figure 5.3A).

5.2.4 Pali1 and Aebp2 antagonise each other biochemically for protein stability.

In order to understand how loss of Pali1 could lead to decreased H3K27me3, we set out to dissect its role within the PRC2 complex in mouse ESCs. Previous studies established that loss of core PRC2 subunits, such as Eed or Suz12, leads

A

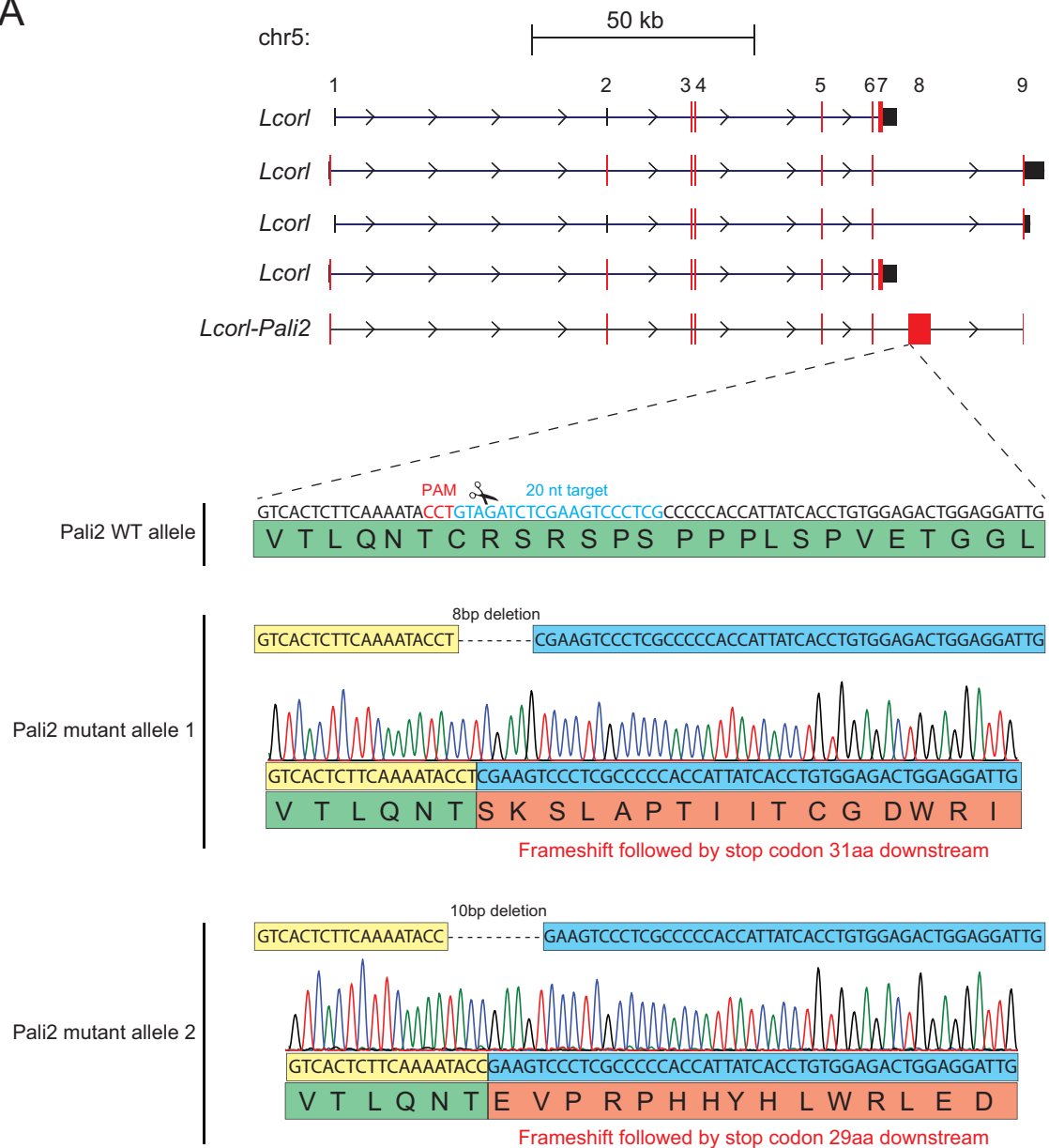


Figure 5.3 Pali2 knockout strategy using CRISPR-Cas9 mediated gene editing.

(A) The CRISPR targeting strategy for Pali2 in mouse ESCs. Top: A view of the mouse *Lcorl* and Pali2 isoforms. A guide RNA was targeted to cut DNA 5' of the Pali2 specific exon (exon 8), which contains the Polycomb interaction domain. The base (as determined by sanger sequencing) and amino acid sequence of the regions surrounding the PAM site of the Pali2 targeting gRNA is shown for the two mutant alleles present in Pali1/2 dKO ESCs following the knockout strategy.

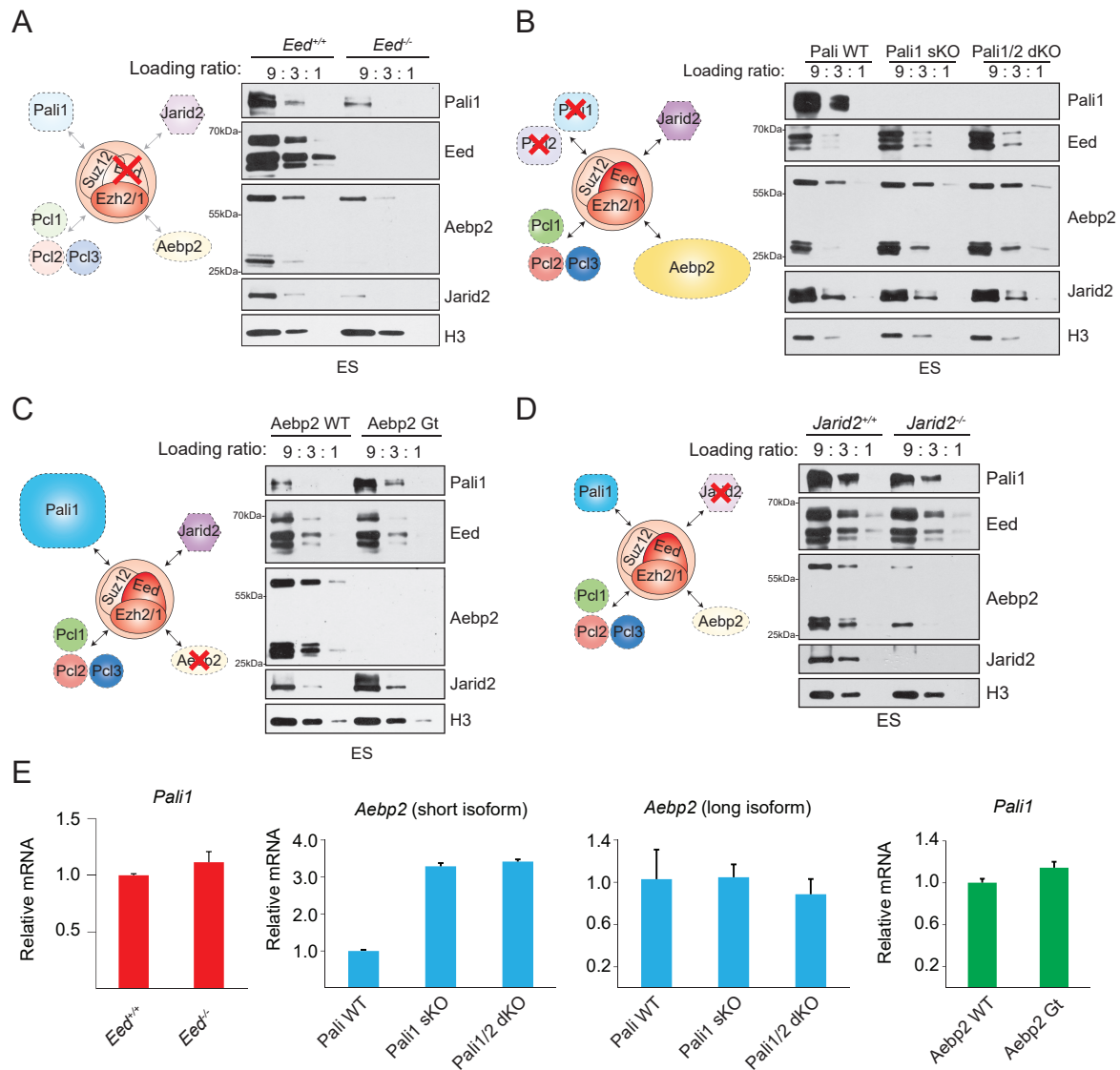


Figure 5.4 Pali1 and Aebp2 antagonise each other biochemically for protein stability.

(A, B, C and D) Western blot analysis on a dilution series of whole cell lysates of *Eed*^{-/-}, Pali1^{-/-} (Pali1 sKO), Pali1^{-/-}, Pali2^{-/-} (Pali1/2 dKO), Aebp2^{gt/gt} (Aebp2 Gt), *Jarid2*^{-/-} and matched wild-type ESCs.

(E) RT-qPCR of cDNA from *Eed*^{-/-}, Pali1^{-/-} (Pali1 sKO), Pali1^{-/-}, Pali2^{-/-} (Pali1/2 dKO), Aebp2^{gt/gt} (Aebp2 Gt), *Jarid2*^{-/-} and matched wild-type ESCs. The results were normalised to levels of the RPLPO housekeeping gene.

to destabilisation of other PRC2 proteins (Montgomery et al., 2005; Pasini et al., 2004). To determine if loss of core PRC2 would affect Pali1, I performed Western blot analysis on total lysates of *Eed*^{-/-} and matched wild-type ESCs, and found that the levels of the Pali1 protein, like Aebp2 and Jarid2, was reduced in the absence of Eed (Figure 5.4A). This result, combined with the observation that Pali1 mRNA levels were unchanged in *Eed*^{-/-} ESCs (Figure 5.4E), suggests that, consistent with being a PRC2.1 associated protein, Pali1 protein stability is reliant on the integrity of the complex.

I next determined if the stability of PRC2 complex proteins is affected in Pali1 sKO and Pali1/2 dKO ESCs. Western analysis of total lysates from the Pali1 sKO and Pali1/2 dKO ESCs showed no significant changes in the levels of Eed or Jarid2 proteins (Figure 5.4B). However, I observed a 2-3 fold increase in the amount of the short isoform of Aebp2 (Figure 5.4B), which correlated with a similar increase in the mRNA levels of the corresponding short, but not long mRNA transcript of *Aebp2* (Figure 5.4E). As a control, I investigated the consequences of loss of the other PRC2.2 component, Jarid2. Western analyses on total lysates from *Jarid2*^{-/-} ESCs showed that its loss does not affect Pali1 protein levels, but did lead to a destabilisation of Aebp2 protein, consistent with Aebp2 and Jarid2 being part of the same PRC2.2 subtype (Figure 5.4D).

Intriguingly, Pali1 protein levels were ~3 fold increased in *Aebp2*^{gt/gt} (*Aebp2* Gt) ESCs compared to matched wild-type control cells (Figure 5.4C), while its mRNA levels were unchanged (Figure 5.4E). Taken together with the fact that Pali1 is destabilised in *Eed*^{-/-} ESCs (Figure 5.4A), this suggests that Pali1 is stabilised by increased association with the core PRC2 complex in the absence of Aebp2.

5.2.5 Pali1 and Aebp2 occupy mutually exclusive sub-types of PRC2 complexes.

The data thus far are consistent with a model of competition between Pali1 and Aebp2 for association with PRC2. AEBP2 and JARID2 have been previously reported to be in the PRC2.2 variant of the complex, which is exclusive of the PRC2.1 specific subunits PCL1-3. C10ORF12 has also been reported to be a part of the PRC2.1 complex with PCL proteins. To confirm the exclusivity of subunits between these two variants of the PRC2 complex IPs were performed with FLAG-HA tagged PALI1/ PCL2/ JARID2 and AEBP2. Western blotting of the resulting eluates confirmed that AEBP2 co-IPs with JARID2 and that PCL2 co-IPs with PALI1, while PALI1 never co-IPs with AEBP2 (Figure 5.5A). To determine how this fits into our competition model between Pali1 and Aebp2, I performed endogenous immunoprecipitations of Ezh2 in Pali1 sKO, Aebp2 Gt and matching wild-type ESCs (Figure 5.5B). This confirmed our hypothesis, revealing that loss of Pali1 leads to an increase of the short isoform of Aebp2 in the PRC2 complex and reciprocally, loss of Aebp2 led to an increase in the amount of Pali1 in PRC2. These results uncover an antagonistic interplay between two protein subunits of the different PRC2.1 and PRC2.2 subtypes (Figure 5.5C).

5.2.6 Consequences of knockout of PRC2 subunits on H3K27 methylation.

Following the assessment of the effect of PRC2 sub-stoichiometric subunits on their fellow subunit stabilities I next wanted to determine what effect the loss of the sub-stoichiometric subunits would have on the PRC2 complex mediated histone modifications H3K27me3, H3K27me2 and H3K27me1 along with the functionally related modifications H3K27ac and H3K36me3. As expected, *Eed*^{-/-} ESCs featured

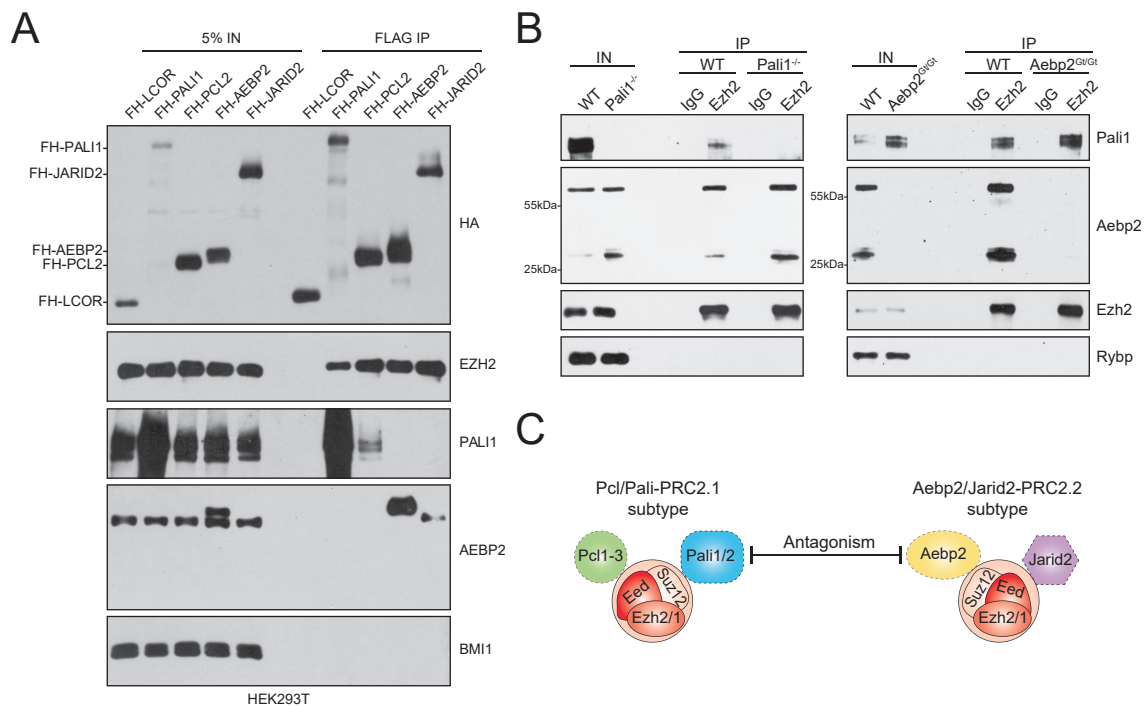


Figure 5.5 Pali1 and Aebp2 occupy mutually exclusive sub-types of PRC2 complexes.

(A) Western blots of FLAG-IP eluate of the indicated proteins following transient ectopic expression in HEK293T cells.

(B) Endogenous IPs of Ezh2 in Pali1 sKO or Aebp2 Gt ESCs, compared to matched wild-type controls, followed by western blot analysis with the indicated antibodies.

(C) Model representing the antagonism between Pcl/Pali-PRC2.1 and Aebp2/Jarid2-PRC2.2 complexes.

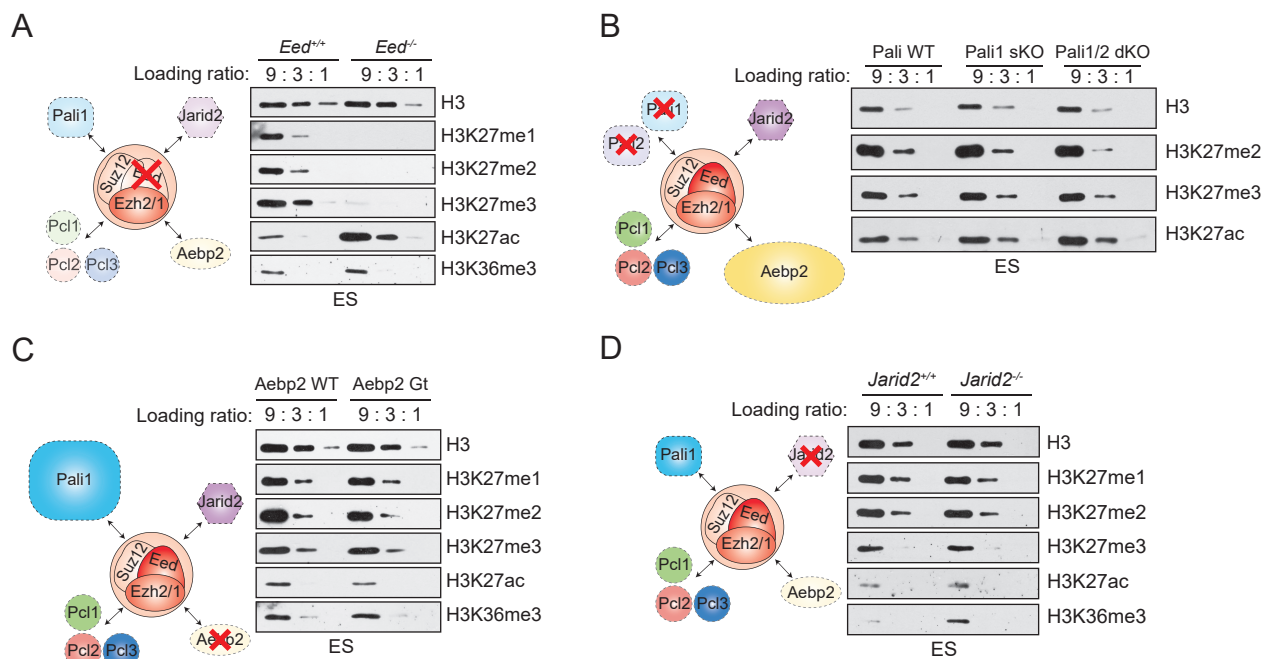


Figure 5.6 Consequences of knockout of PRC2 subunits on H3K27 methylation.

(A, B, C and D) Western blot analysis on a dilution series of whole cell lysates of *Eed*^{-/-}, Pali1^{-/-} (Pali1 sKO), Pali1^{-/-}, Pali2^{-/-} (Pali1/2 dKO), Aebp2^{gt/gt} (Aebp2 Gt), *Jarid2*^{-/-} and matched wild-type ESCs.

global losses of H3K27me1/2/3 which coincided with an increase in H3K27ac (Figure 5.6A). The careful balance between H3K27ac and H3K27me3 has been well established (Pasini et al., 2010b) and these Western blots served as a positive control for our analysis of the effect of sub-stoichiometric subunit knockout on these modifications.

Surprisingly, knockout of *Pali1/2*, *Aebp2* or *Jarid2* had no apparent effect on the global levels of H3K27me1/2/3, H3K27ac or H3K36me3 (Figure 5.6B, C and D). It has previously been shown that loss of *Aebp2* leads to increased levels of H3K27me3 at Polycomb target genes by ChIP-seq, and depletion of *Jarid2* has been shown to result in reduced H3K27me3 at these targets (Grijzenhout et al., 2016; Pasini et al., 2010a) but the global levels of H3K27me3 have not been checked in these studies. Moreover, the *Pali1* null mice embryos exhibited reduced global levels of H3K27me3 (Figure 5.2D). Broadly speaking, these results might indicate that the specific changes observed in PRC2 subunit null ESCs, in ChIP studies, at Polycomb target genes are not necessarily reflected at the global level. This suggests that disruption of any one sub-stoichiometric subunit is not sufficient to completely disrupt the complexes catalytic activity. These results become particularly important considering that PRC2 exists in two distinct variants, PRC2.1 and PRC2.2. Therefore knockout of any one sub-stoichiometric subunit at a time always leaves a variant PRC2 complex completely intact. So, it is possible that to completely abrogate PRC2 activity through its sub-stoichiometric sub-subunits that both PRC2.1 and PRC2.2 need to be disrupted separately.

5.2.7 Loss of *Pali1/2* or *Aebp2* has opposing effects on PRC2 activity *in vitro* and *in vivo*.

I next wished to explore how the antagonism between Pali1 and Aebp2 might affect PRC2 activity on Polycomb target genes. I performed Chromatin Immunoprecipitations (ChIPs) in Pali1 sKO, Pali1/2 dKO, and matched wild-type ESCs and observed a reduction in H3K27me3 at the promoters of several Polycomb target genes in the absence of the Pali proteins (Figure 5.7A). In contrast, Aebp2 Gt ESCs have increased enrichments of H3K27me3 on Polycomb target genes compared to matched wild-type ESCs (Figure 5.7B), as reported previously (Grijzenhout et al., 2016). Consistent with the idea that the interaction of Pali1 and Aebp2 with PRC2 is antagonistic, ChIPs of Aebp2 in Pali1 sKO and Pali1/2 dKO, compared to matched wild-type ESCs, established that in the absence of Pali proteins, Aebp2 increases on Polycomb target genes (Figure 5.7C), consistent with a general increase in the Aebp2 short isoform protein levels. I determined that the Aebp2 antibody was specific in ChIP by carrying out Aebp2 ChIPs in Aebp2 Gt cells and found that the Aebp2 enrichment at Polycomb target genes was substantially depleted (Figure 5.7D). Taken together, these data support the idea that the Pali1 and Aebp2 proteins are central to an antagonism between the PRC2.1 and PRC2.2 complexes and their absence disrupts the balance of PRC2 mediated H3K27me3 on Polycomb target genes.

To explore the potential mechanism behind the changes in PRC2 mediated H3K27me3 upon loss of Aebp2 or Pali1/2, I speculated that PALI1-PRC2.1 has a stronger histone methyltransferase activity compared to AEBP2-PRC2.2. To test this, histone methyltransferase assays were performed with the core PRC2 complex together with either recombinant human AEBP2 or PALI1 proteins. This was performed in collaboration with Raphael Margueron and Daniel Holloch (Institut Curie, France). This established that even a 30 fold excess of AEBP2 was not

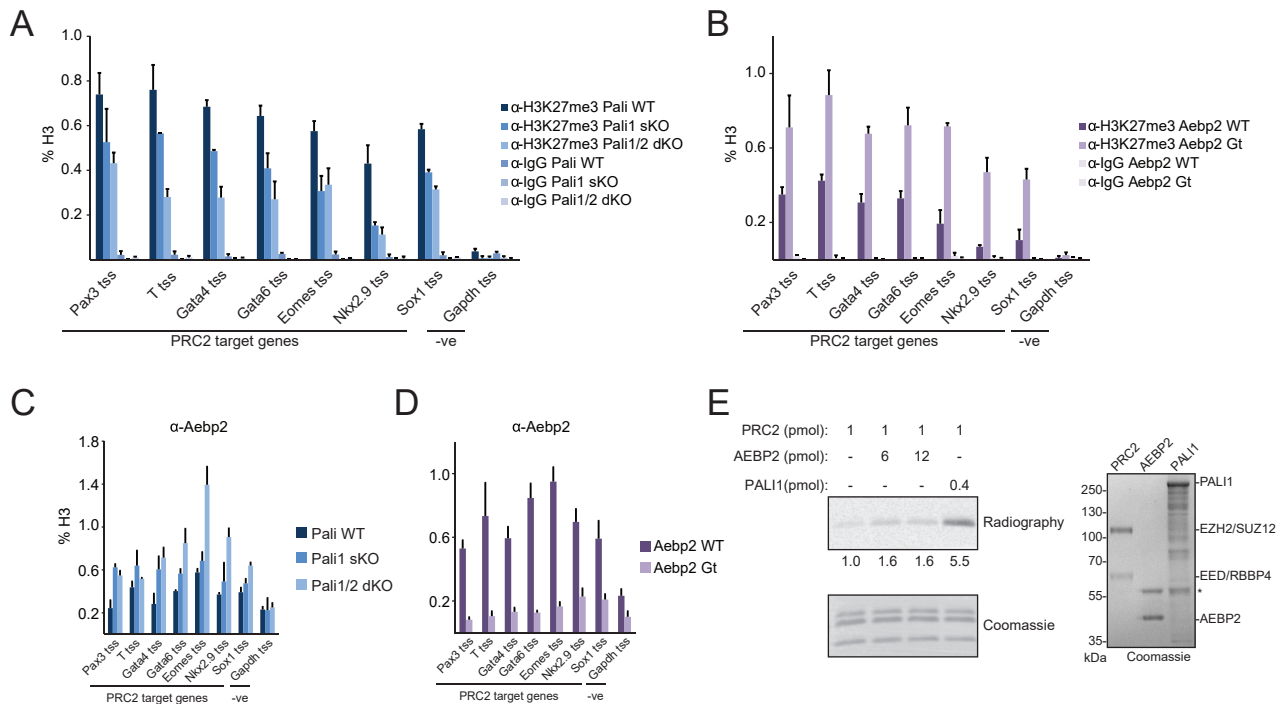


Figure 5.7 Loss of Pali1/2 or Aebp2 has opposing effects on PRC2 activity in vitro and in vivo.

(A) ChIP-qPCRs at Polycomb target genes using the indicated antibodies in Pali1 sKO and Pali1/2 dKO and matched wild-type ESCs. The Gapdh promoter is included as a negative control.

(B) ChIP-qPCRs at Polycomb target genes using the indicated antibodies in Aebp2 Gt and matched wild-type ESCs. The Gapdh promoter is included as a negative control.

(C) ChIP-qPCRs at Polycomb target genes using Aebp2 antibody in Pali1 sKO, Pali1/2 dKO and matched wild-type ESCs. The Gapdh promoter is included as a negative control.

(D) ChIP-qPCRs at Polycomb target genes using Aebp2 antibody in Aebp2 Gt and matched wild-type ESCs. The Gapdh promoter is included as a negative control.

(E) Left. Histone methyltransferase assay using recombinant PRC2 with addition of either AEBP2 or PALI1 protein, as indicated. Recombinant oligonucleosomes were used as a substrate. The PRC2 complex used contained FLAG-EED, SUZ12, EZH2 and RBBP4. Incorporation of 3H from radiolabeled SAM was used as a read-out of methyltransferase activity. Quantification shown is relative signal normalized to PRC2 only. A representative image is shown from one of four biological replicates. Right. Coomassie staining of recombinant complex/ proteins used for HMT assay. Daniel Holoch and Raphael Margueron performed this experiment.

comparable to the stronger ability of PALI1 to promote PRC2 methyltransferase activity *in vitro* (Figure 5.7E). A potential caveat here is that the *in vitro* activity of AEBP2 could be enhanced by the presence of JARID2, especially in the context of nucleosomes containing H2AK119ub1 modified histones as a substrate (Kalb et al., 2014). Also, the ‘add-back’ approach of adding proteins not co-purified with the complex, to the PRC2 complex has drawbacks of its own. For instance the kinetics of PALI1-PRC2 and AEBP2-PRC2 interactions may be different making it difficult to be certain our findings are an accurate representation of the *in vivo* context. Additionally, use of protein-protein interaction mutants, such as PALI1 W1125 and W1186 double mutant, in these assays would provide assurance that the observed increases in methyltransferase activity in the presence of PALI1 are due to an increase in PRC2 activity and not PRC2 independent methyltransferase activity. Nonetheless, our data are consistent with the idea that the increases and decreases in H3K27me3 levels in the absence of either Aebp2 or Pali1, respectively, can be explained by consequent imbalances in the proportions of the Pali-PRC2.1 and Aebp2-PRC2.2 complexes, which have different degrees of histone methyltransferase activity.

5.2.8 The balance of PRC2.1 and PRC2.2 activities is essential for the proper regulation of Polycomb target genes during ESC differentiation.

Next, I evaluated the consequences of the deregulated H3K27me3 levels in the Pali1 sKO, Pali1/2 dKO and Aebp2 Gt ESCs on the regulation of Polycomb target genes during induction to differentiate to embryoid bodies (EBs). I did not investigate whether there are any transcriptional changes between wild-type, Aebp2 Gt, Pali1 sKO and Pali1/2 dKO ESCs as it has been previously shown that

there is little to no change in transcription in Polycomb group protein knockout ESCs (Riising et al., 2014). This may be due to a lack of gene activation cues, as in contrast, large scale changes in transcription have been reported in Polycomb knockout cells upon differentiation (Pasini et al., 2010b; Riising et al., 2014). RNA-seq analysis of knockout and matched wild-type ESCs, both before and after induction to differentiate to EBs, revealed that loss of *Aebp2* led to an inability to properly activate a cohort of 796 Polycomb target genes associated with differentiation and germ layer development, including *Wnt3*, *Fgf8*, *Olig3*, *Msx3*, *Pou3f2*, *Otx1*, *FoxA2*, *T* and *Gata6* (Figure 5.8A). I focussed this analysis on Polycomb target genes as this is the simplest way to focus on direct changes of *Aebp2* loss, rather than looking at global transcription changes which may be caused by indirect downstream effects of *Aebp2* loss. The opposite phenotype was observed in the *Pali1* sKO and *Pali1/2* dKO ESCs, which had significantly higher levels of these genes, compared to matched wild type ESCs, 8 days post-induction to differentiate (Figure 5.8A). I found that there was a large correlation between genes (247) that have increased expression in *Pali1* sKO and *Pali1/2* dKO with decreased expression in *Aebp2* Gt EBs (Figure 5.8B). This RNA-seq analysis was limited in that it focussed only on PRC2 target gene expression changes, thus it is possible that some non-PRC2 related transcriptional changes may have been omitted from the analysis. However, as *Pali1* is destabilised in the absence of *Eed* (Figure 5.4A) it seems that the primary function of *Pali1* is as part of the PRC2 complex. I validated these RNA-seq observations for two representative Polycomb target genes, *Gata6* and *T*, in an independent experiment by RT-qPCR over a time course of ESC to EB differentiation (Figure 5.8C and D) These results are consistent with the changes in the levels of H3K27me3 on the promoters of these

Polycomb target genes in the knockout ESCs. Whereas the Aebp2 Gt ESCs have increased H3K27me3 levels, consistent with their impaired ability to activate Polycomb target genes appropriately during differentiation, the Pali1 sKO and Pali1/2 dKO ESCs have decreased H3K27me3, consistent with the propensity to activate these genes to higher levels during differentiation.

5.2.9 Pali1/2 and Aebp2 competition model

Taken together, these results are in line with a model in which the Pali1/2 and Aebp2 proteins are central to balancing the respective activities of PRC2.1 and PRC2.2 and that the loss of either leads to the dysregulation of Polycomb target genes during differentiation (Figure 5.9A).

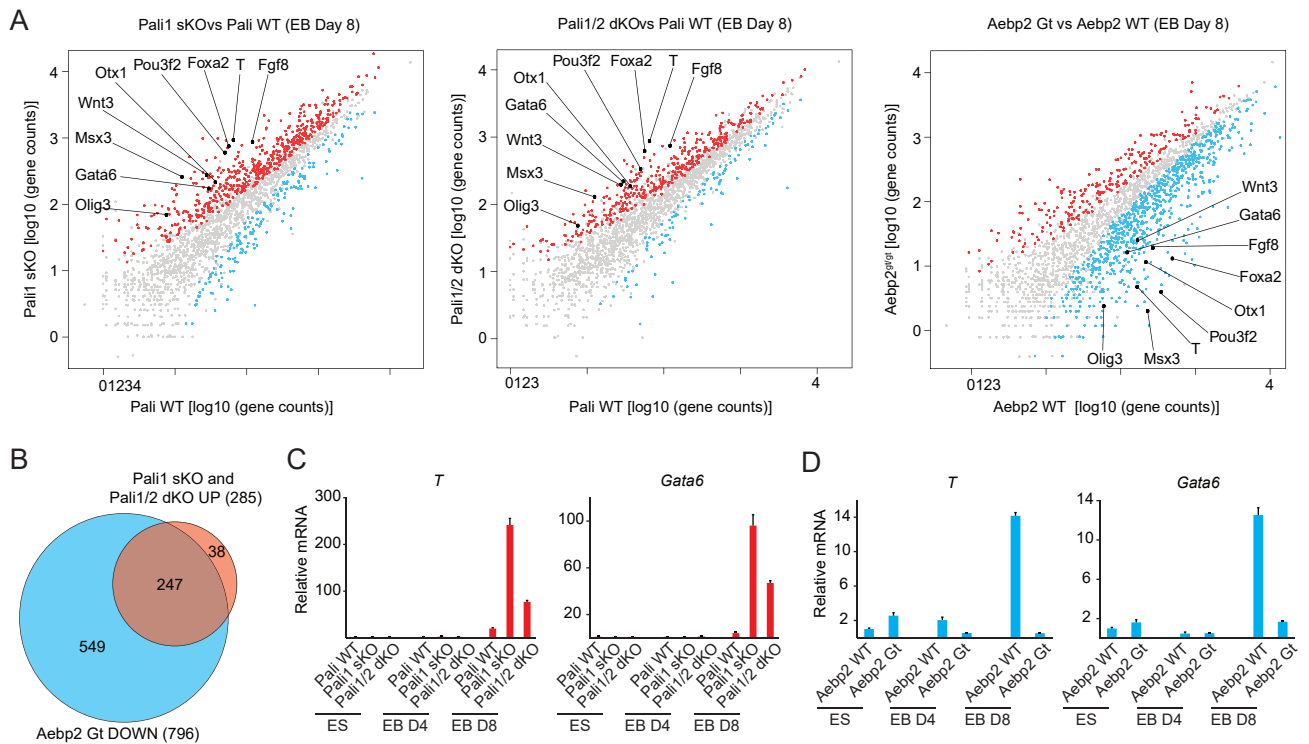


Figure 5.8 The balance of PRC2.1 and PRC2.2 activities is essential for the proper regulation of Polycomb target genes during ESC differentiation.

(A) RNA-seq data for Pali1 sKO, Pali1/2 dKO, Aebp2 Gt and matched wild-type embryoid bodies (EBs) at day 8 of differentiation. Log₁₀ of the normalised gene count data are represented for the top 3,000 H3K27me3 positive genes in ESCs (Riising et al., 2014).

(B) Venn diagram showing the overlap of H3K27me3 positive genes that are increased in both Pali1 sKO and Pali1/2 dKO, and decreased in Aebp2 Gt EBs (D8).

(C) RT-qPCR analyses of Pali1 sKO, Pali1/2 dKO and matched wild-type ESCs during induction to differentiate to embryoid bodies (EBs). The results are normalized to the levels of the Gapdh housekeeping gene.

(D) RT-qPCR analyses of Aebp2 Gt and matched wild-type ESCs during induction to differentiate to embryoid bodies (EBs). The results are normalized to the levels of the Gapdh housekeeping gene.

A

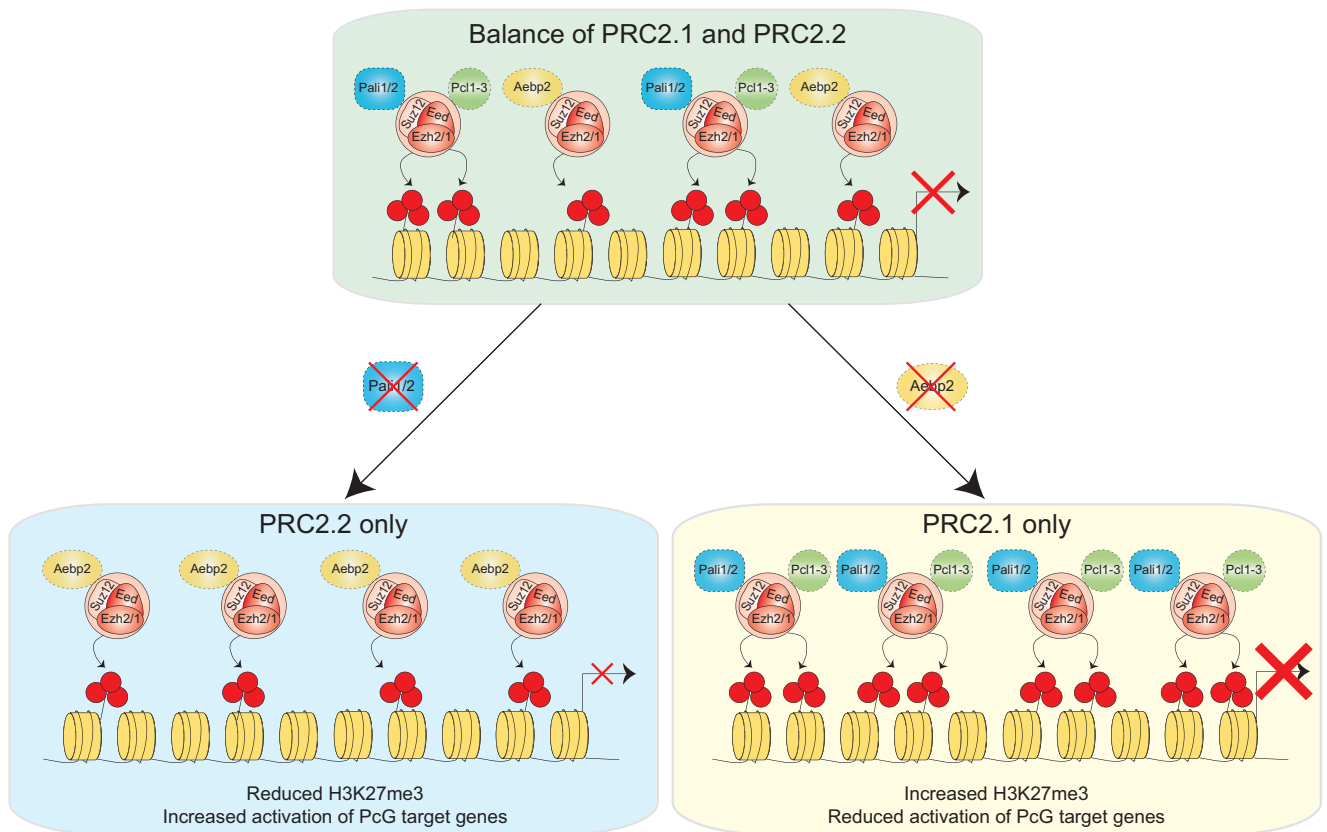


Figure 5.9 Pali1/2 and Aebp2 competition model.

(A) Depiction of the role Pali1/2 and Aebp2 proteins play in maintaining the balance of PRC2.1 and PRC2.2 activity by regulating H3K27me3 levels and PRC2 target gene expression.

5.3 Discussion

Here I have uncovered an essential role for the *Pali1* specific exon (*Gm340*) within the *Lcor* gene locus, in mouse embryonic development with lethality being evident between the E11.5 stage and birth. The knockout of a number of genes encoding the core PRC2 proteins *Ezh2*, *Eed* and *Suz12* leads to an early embryonic lethal phenotype (~E7.5-8.5) and are associated with defects in gastrulation (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004). In contrast, the disruption of the genes that encode sub-stoichiometric subunits of PRC2 such as *Jarid2*, *Pcl2* and *Aebp2* have weaker phenotypic consequences. For instance, *Pcl2* gene trap mice have viable offspring, potentially due to functional redundancy between it and its paralogs *Pcl1* and *Pcl3* (Li et al., 2011). The precise phenotype of *Jarid2* and *Aebp2* null mice is less clear, since the results vary depending on background strain of mice (Grijzenhout et al., 2016; Kim et al., 2011; Lee et al., 2000; Motoyama et al., 1997; Takeuchi et al., 1999; Takeuchi et al., 1995). However, the consistent observation is that they survive until later embryonic stages (E11.5-E15.5 for *Jarid2* null mice and E10.5-E18.5 for *Aebp2* null mice).

These results suggest that the deletion of any of the core PRC2 subunits has an immediate effect in development while the loss of any one sub-stoichiometric subunit is not sufficient to disrupt the function of the PRC2 complex as a whole (Grijzenhout et al., 2016; Kim et al., 2011; Lee et al., 2000; Motoyama et al., 1997; Takeuchi et al., 1999; Takeuchi et al., 1995). This may be due to a potential division of labour between the two PRC2 sub-complexes, PRC2.1 and PRC2.2, such that one variant compensates for the loss of the other during early development.

This potential for redundancy or compensation between the PRC2.1 and PRC2.2 variants is supported by the finding that in *Eed*^{-/-} ESCs the catalytic activity of the complexes is lost completely, but disruption of the genes encoding Pali1/2, Jarid2 or Aebp2 has little to no effect on the global level of H3K27 modifications by Western blot. Despite the absence of global differences in H3K27me3, discrete changes have been observed at PRC2 targets in ChIP experiments following disruption of the sub-stoichiometric subunits (Beringer et al., 2016; Brien et al., 2012; Grijzenhout et al., 2016; Pasini et al., 2010a). In addition, the importance of Aebp2 and Pali1/2 in ESCs only became evident from a transcriptional perspective following the induction of differentiation. This is in keeping with the hypothesis that the sub-stoichiometric subunits may have greater importance in specific lineages following differentiation rather than early in development or in ESCs.

The discovery of an apparent competition between the Aebp2 and Pali1 subunits of PRC2 is a striking observation. While we do not yet understand the mechanism at the heart of this competition, I have established clearly that the absence of either leads to an imbalance in the levels of the other and that this affects the overall activity of PRC2 at its target genes. On the one hand loss of Pali1 leads to an increase in transcription of the short isoform of Aebp2, which suggests a relatively indirect mechanism. However, a previous study has shown that Aebp2 is itself a PRC2 target gene (Kim et al., 2015a) and as the short and long isoforms of Aebp2 have distinct promoters it is quite possible that the increase in short isoform Aebp2 expression, in the absence of Pali1, is a result of derepression. On the other hand, in Aebp2 null cells, the Pali1 protein is stabilised, suggesting a direct mechanism of competition between the two proteins. It has been established that both Aebp2 and

Jarid2 directly associates with the PRC2 complex through an interaction with the middle region of Suz12, surrounding its zinc finger (Cao and Zhang, 2004; Ciferri et al., 2012; Kasinath et al., 2018). Our discovery of competition between Pali1 and Aebp2 suggests that Pali1 potentially also interacts with Suz12 near the zinc finger if they do in fact sterically hinder the binding of one another to the PRC2 complex. Future experiments aimed at determining the mechanism of this competition, should focus on deleting or mutating this region in Suz12 and investigate whether the interaction between both Pali1 and Aebp2 is lost. This experiment could potentially elucidate the region of PRC2 for which Pali1 and Aebp2 compete, which may act as a molecular “switch” between PRC2.1 and PRC2.2.

Chapter 6: General Discussion

6.1 Discussion

Here I report the discovery of a new family of vertebrate-specific proteins, PALI1 and PALI2, encoded by the *LCOR* and *LCORL* gene loci, respectively. These multi-domain proteins are defined by a novel PIP domain, which confers the ability to bind to the PRC2 complex. Several additional motifs link these proteins with additional epigenetic regulators, nuclear receptors and CTBP co-repressors. I show that Pali1 is essential for mouse development and promotes PRC2 histone methyltransferase activity *in vitro* and *in vivo*. Moreover, Pali proteins define a distinct form of the PRC2.1 subtype and are essential to balance the activity of Aebp2-PRC2.2 in ESCs and during differentiation. I anticipate that this discovery will open up new avenues of research linking Polycomb function with CTBP co-repressors and nuclear receptors in the regulation of cellular identity during development and in cancer.

The discovery of PALI1 and PALI2 raises interesting new questions about how the PRC2 complex might be targeted to chromatin in mammalian cells, the mechanisms for which remains poorly defined. As well as their PIP domain, the PALI1 and PALI2 proteins also contain several additional motifs and domains, thereby notably expanding the potential mechanisms by which the PRC2 complex could be targeted to specific chromatin sites. For instance, PALI1 shares its N-terminus with the LCOR protein, which contains both CTBP binding motifs and a nuclear receptor binding box that has been characterised to interact with the Estrogen, Progesterone and Thyroid hormone receptors (Fernandes et al., 2003; Palijan et al., 2009; Song et al., 2012). In addition to possessing two CTBP binding motifs, the PALI2 protein also contains three AT hooks, with the potential for direct

interaction with DNA. It is possible that these domains contribute to the targeting of PRC2 complex activity. In mammalian cells, an elegant model has recently emerged in which H2AK119ub1, mediated by the catalytic activity of non-canonical PRC1 complexes, can contribute to the recruitment of PRC2 to unmethylated CpG islands (Almeida et al., 2017; Blackledge et al., 2014; Endoh et al., 2017; Rose et al., 2016). The ability of the PRC2.2 specific Jarid2 subunit to bind H2AK119ub1 via its UIM domain has been proposed as the mechanism for this (Cooper et al., 2016; Kalb et al., 2014). However, the complete loss of all H2AK119ub1 in both fibroblasts and tissue specific stem cells, upon Ring1A/B deletion, does not affect global levels of H3K27me3 (Chiacchiera et al., 2016a; de Napoles et al., 2004). Furthermore, Ring1B mutant ESCs, which are incapable of catalysing H2AK119ub1, do not exhibit changes in the global levels of H3K27me3 (Illingworth et al., 2015). These results strongly support the idea that PRC2 must be targeted to chromatin by additional mechanisms, independent of H2AK119ub1. Therefore, my discovery of the multi-domain PALI1 and PALI2 proteins provides a potential mechanism by which CTBP1/2 co-repressors and nuclear receptors could contribute to targeting or regulating PRC2 activity. It will be important to delineate the respective functions of the several PALI1 and PALI2 domains and their requirement for their interplay with co-repressors, nuclear receptors and other chromatin regulators during development.

While the knockout of core PRC2 members such as *Eed*, *Suz12* and *Ezh2* all result in early embryonic lethality (~E7.5-8.5) and defects in gastrulation (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004), the *Pali1*^{-/-} mice generated here have a less pronounced phenotype, resulting in lethality between E11.5 and birth.

Furthermore, while the *Eed* and *Suz12* knockout mice lack any H3K27me3, we find that the *Pali1* knockout mice have an overall reduction, but not a total loss of H3K27me3 at E11.5. These data are consistent with the fact that there are two subtypes of the PRC2 complex with potential redundancy between subtypes. Supporting this, *Pcl2* knockout mice have viable offspring (Li et al., 2011) while *Jarid2* and *Aebp2* null mice survive until mid to late stage embryonic stages (Grijzenhout et al., 2016; Kim et al., 2011; Lee et al., 2000; Motoyama et al., 1997; Takeuchi et al., 1999; Takeuchi et al., 1995). It is also possible that the paralogous Pali2 protein might compensate for the loss of Pali1. These data all support the idea that upon the impairment of the activity of one subtype of the PRC2 complex, embryos can remain viable until later in development due to the activity of the other PRC2 subtype. However, it is clear that for proper development in the mid to late stage embryo, the balance of both PRC2.1 and PRC2.2 is essential. In addition to spatial development, Polycombs have been reported to have roles in the regulation of the cell cycle, senescence and the maintenance of stem cell niches such as intestinal stem cells and haematopoietic stem cells (Bracken et al., 2007; Bracken et al., 2003; Chiacchiera et al., 2016b; Park et al., 2003). Although we have not yet investigated the role of Pali1 in these processes it is likely to play a role in some, if not all of them. Intriguingly, we have observed a kidney differentiation defect in Pali1 null embryos (data not shown) which may explain the embryonic lethal phenotype. However, this is an area that needs further investigation to determine what role Pali1 plays in kidney development.

The discovery of the PALI1 and PALI2 proteins provides additional mechanisms by which the activity of the PRC2 complex could be deregulated in cancer and, I

anticipate, new opportunities to treat it. PRC2 function is frequently deregulated in several cancer types due to genetic perturbation (Comet et al., 2016; Conway et al., 2015). These include recurrent heterozygous 'change-of-function' mutations in *EZH2*, which occur in 22% of B-cell lymphomas, loss of function mutations in *EZH2*, *SUZ12*, *EED* and *AEBP2* in T-cell leukaemias and malignant peripheral nerve sheath tumors, and H3K27M mutations in ~78% of paediatric gliomas (Ernst et al., 2010; Jankowska et al., 2011; Lee et al., 2014; Morin et al., 2010; Morin et al., 2011; Nikoloski et al., 2010; Sturm et al., 2012; Wu et al., 2012; Zhang et al., 2014). Remarkably, the *LCOR/PAL1* gene locus is recurrently deleted in 9% of diffuse large B-cell lymphomas and 11% of transformed follicular lymphomas (Bouska et al., 2014; Chan et al., 2015). These *LCOR/PAL1* mutant lymphomas do not have a significant overlap with those carrying the 'change of function' *EZH2* mutations, suggesting they are from an independent cohort of patients (Chan et al., 2015). It will be important to determine if the levels and position of PRC2 mediated H3K27me2 and H3K27me3 modifications are altered in these cancers. Based on our work, those cancers lacking *PAL1* would also have an imbalance in the ratio of PRC2.1 to PRC2.2, which should be investigated to determine whether restoring the balance could have potential therapeutic benefit. More broadly, while I have not yet characterised the potential interplay between *PAL1* and nuclear receptors, it will be exciting to investigate this link, considering their role as key drivers in hormone dependent cancers, including prostate and breast cancer (Deroo and Korach, 2006; Heinlein and Chang, 2004).

In summary, this study adds to the growing evidence of an expanded and sub-functionalised Polycomb system in vertebrates. I believe this work will also

contribute to a better mechanistic understanding of the interplay between Polycombs, transcriptional co-repressors and nuclear receptors during cell fate decisions and carcinogenesis in higher eukaryotic cells.

6.2 Conclusions

- Pali1 and Pali2 form a new family of vertebrate-specific proteins that bind PRC2 through a highly conserved PIP domain.
- Pali1 promotes PRC2 methyltransferase activity *in vitro* and *in vivo*.
- Pali1 defines a distinct PRC2.1 subtype essential for mouse development.
- PRC2 subtype balance is essential for proper regulation of Polycomb target genes.

References

Adler, P.N., Martin, E.C., Charlton, J., and Jones, K. (1991). Phenotypic consequences and genetic interactions of a null mutation in the *Drosophila* Posterior Sex Combs gene. *Dev Genet* 12, 349-361.

Agger, K., Cloos, P.A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 449, 731-734.

Akasaka, T., Kanno, M., Balling, R., Mieza, M.A., Taniguchi, M., and Koseki, H. (1996). A role for mel-18, a Polycomb group-related vertebrate gene, during the anteroposterior specification of the axial skeleton. *Development* 122, 1513-1522.

Akasaka, T., van Lohuizen, M., van der Lugt, N., Mizutani-Koseki, Y., Kanno, M., Taniguchi, M., Vidal, M., Alkema, M., Berns, A., and Koseki, H. (2001). Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development* 128, 1587-1597.

Alabert, C., Barth, T.K., Reverón-Gómez, N., Sidoli, S., Schmidt, A., Jensen, O.N., Imhof, A., and Groth, A. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes and Development* 29, 585-590.

Alabert, C., Bukowski-Wills, J.-C., Lee, S.-B., Kustatscher, G., Nakamura, K., Alves, F.d.L., Menard, P., Mejlvang, J., Rappsilber, J., and Groth, A. (2014). Nascent chromatin capture proteomics determines chromatin dynamics during

DNA replication and identifies unknown fork components. *Nature Cell Biology* 16, 281-291.

Alekseyenko, A.A., Gorchakov, A.A., Kharchenko, P.V., and Kuroda, M.I. (2014). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc Natl Acad Sci U S A* 111, 2488-2493.

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., *et al.* (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* 356, 1081-1084.

Atsuta, T., Fujimura, S., Moriya, H., Vidal, M., Akasaka, T., and Koseki, H. (2001). Production of monoclonal antibodies against mammalian Ring1B proteins. *Hybridoma* 20, 43-46.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkschlager, M., *et al.* (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* 8, 532-538.

Baker, T., Nerle, S., Pritchard, J., Zhao, B., Rivera, V.M., Garner, A., and Gonzalez, F. (2015). Acquisition of a single EZH2 D1 domain mutation confers acquired resistance to EZH2-targeted inhibitors. *Oncotarget* 6, 32646-32655.

Ballare, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., *et al.* (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat Struct Mol Biol* 19, 1257-1265.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120-124.

Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C., *et al.* (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol Cell* 66, 384-397.e388.

Beringer, M., Pisano, P., Di Carlo, V., Blanco, E., Chammas, P., Vizan, P., Gutierrez, A., Aranda, S., Payer, B., Wierer, M., *et al.* (2016). EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. *Mol Cell* 64, 645-658.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315-326.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., *et al.* (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* 157, 1445-1459.

Bodor, C., Grossmann, V., Popov, N., Okosun, J., O'Riain, C., Tan, K., Marzec, J., Araf, S., Wang, J., Lee, A.M., *et al.* (2013). EZH2 mutations are frequent and represent an early event in follicular lymphoma. *Blood* 122, 3165-3168.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T., and Zhuang, X. (2016). Super-resolution

imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418-422.

Bouska, A., McKeithan, T.W., Deffenbacher, K.E., Lachel, C., Wright, G.W., Iqbal, J., Smith, L.M., Zhang, W., Kucuk, C., Rinaldi, A., *et al.* (2014). Genome-wide copy-number analyses reveal genomic abnormalities involved in transformation of follicular lymphoma. *Blood* 123, 1681-1690.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349-353.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* 20, 1123-1136.

Bracken, A.P., and Helin, K. (2009). Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat Rev Cancer* 9, 773-784.

Bracken, A.P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., Theilgaard-Monch, K., Minucci, S., Porse, B.T., Marine, J.C., *et al.* (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev* 21, 525-530.

Bracken, A.P., Pasini, D., Capra, M., Prosperini, E., Colli, E., and Helin, K. (2003). EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *Embo J* 22, 5323-5335.

Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., *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, 1273-1281.

Brien, G.L., Healy, E., Jerman, E., Conway, E., Fadda, E., O'Donovan, D., Krivtsov, A.V., Rice, A.M., Kearney, C.J., Flaus, A., *et al.* (2015). A chromatin-independent role of Polycomb-like 1 to stabilize p53 and promote cellular quiescence. *Genes Dev* 29, 2231-2243.

Brien, G.L., Valerio, D.G., and Armstrong, S.A. (2016). Exploiting the Epigenome to Control Cancer-Promoting Gene-Expression Programs. *Cancer Cell* 29, 464-476.

Brooun, A., Gajiwala, K.S., Deng, Y.L., Liu, W., Bolanos, B., Bingham, P., He, Y.A., Diehl, W., Grable, N., Kung, P.P., *et al.* (2016). Polycomb repressive complex 2 structure with inhibitor reveals a mechanism of activation and drug resistance. *Nat Commun* 7, 11384.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D., *et al.* (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol Cell* 49, 571-582.

Calao, M., Sekyere, E.O., Cui, H.J., Cheung, B.B., Thomas, W.D., Keating, J., Chen, J.B., Raif, A., Jankowski, K., Davies, N.P., *et al.* (2013). Direct effects of Bmi1 on p53 protein stability inactivates oncoprotein stress responses in embryonal cancer precursor cells at tumor initiation. *Oncogene* 32, 3616-3626.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.

Cao, R., Tsukada, Y., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* 20, 845-854.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298, 1039-1043.

Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* 15, 57-67.

Chan, F.C., Telenius, A., Healy, S., Ben-Neriah, S., Mottok, A., Lim, R., Drake, M., Hu, S., Ding, J., Ha, G., *et al.* (2015). An RCOR1 loss-associated gene expression signature identifies a prognostically significant DLBCL subgroup. *Blood* 125, 959-966.

Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B., and Misteli, T. (2003). Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* 299, 721-725.

Chiacchiera, F., Rossi, A., Jammula, S., Piunti, A., Scelfo, A., Ordonez-Moran, P., Huelsken, J., Koseki, H., and Pasini, D. (2016a). Polycomb Complex PRC1 Preserves Intestinal Stem Cell Identity by Sustaining Wnt/beta-Catenin Transcriptional Activity. *Cell Stem Cell* 18, 91-103.

Chiacchiera, F., Rossi, A., Jammula, S., Zanotti, M., and Pasini, D. (2016b). PRC2 preserves intestinal progenitors and restricts secretory lineage commitment. *Embo j* 35, 2301-2314.

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. *Elife* 1, e00005.

Comet, I., Riising, E.M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat Rev Cancer* 16, 803-810.

Conway, E., Healy, E., and Bracken, A.P. (2015). PRC2 mediated H3K27 methylations in cellular identity and cancer. *Curr Opin Cell Biol* 37, 42-48.

Conway, E.M., and Bracken, A.P. (2017). Unraveling the Roles of Canonical and Non-canonical PRC1 complexes. In *Polycomb Group Proteins*, Elsevier, pp. 57-80.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., *et al.* (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep* 7, 1456-1470.

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., *et al.* (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat Commun* 7, 13661.

Crick, F.H. (1968). The origin of the genetic code. *J Mol Biol* 38, 367-379.

Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* 150, 12-27.

de Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., *et al.* (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* 7, 663-676.

Deroo, B.J., and Korach, K.S. (2006). Estrogen receptors and human disease. *J Clin Invest* 116, 561-570.

Dietrich, N., Bracken, A.P., Trinh, E., Schjerling, C.K., Koseki, H., Rappsilber, J., Helin, K., and Hansen, K.H. (2007). Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus. *Embo J* 26, 1637-1648.

Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Sodersten, E., and Hansen, K. (2012). REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genet* 8, e1002494.

Dura, J.M., Randsholt, N.B., Deatrick, J., Erk, I., Santamaria, P., Freeman, J.D., Freeman, S.J., Weddell, D., and Brock, H.W. (1987). A complex genetic locus, polyhomeotic, is required for segmental specification and epidermal development in *D. melanogaster*. *Cell* 51, 829-839.

Egan, C.M., Nyman, U., Skotte, J., Streubel, G., Turner, S., O'Connell, D.J., Rraklli, V., Dolan, M.J., Chadderton, N., Hansen, K., *et al.* (2013). CHD5 is required for

neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression. *Dev Cell* 26, 223-236.

Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T., Onaga, N., *et al.* (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *eLife* 6.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., *et al.* (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43-49.

Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A.V., Waghorn, K., Zoi, K., Ross, F.M., Reiter, A., *et al.* (2010). Inactivating mutations of the histone methyltransferase gene *EZH2* in myeloid disorders. *Nat Genet* 42, 722-726.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultschi, A.I., Wutz, A., *et al.* (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* 38, 452-464.

Everitts, A.G., Manning, A.L., Wang, X., Dyson, N.J., Garcia, B.A., and Collier, H.A. (2013). H4K20 methylation regulates quiescence and chromatin compaction. *Mol Biol Cell* 24, 3025-3037.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., *et al.* (2012). KDM2B links the

Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* *1*, e00205.

Faust, C., Schumacher, A., Holdener, B., and Magnuson, T. (1995). The *eed* mutation disrupts anterior mesoderm production in mice. *Development* *121*, 273-285.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. *Nat Protoc* *7*, 1728-1740.

Fernandes, I., Bastien, Y., Wai, T., Nygard, K., Lin, R., Cormier, O., Lee, H.S., Eng, F., Bertos, N.R., Pelletier, N., *et al.* (2003). Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and -independent mechanisms. *Mol Cell* *11*, 139-150.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stutzer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol Cell* *53*, 49-62.

Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A., and Eddy, S.R. (2015). HMMER web server: 2015 update. *Nucleic Acids Res* *43*, W30-38.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* *17*, 1870-1881.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. *Science* *306*, 1574-1577.

Fritsch, L., Robin, P., Mathieu, J.R., Souidi, M., Hinaux, H., Rougeulle, C., Harel-Bellan, A., Ameyar-Zazoua, M., and Ait-Si-Ali, S. (2010). A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. *Mol Cell* 37, 46-56.

Gao, Z., Lee, P., Stafford, J.M., von Schimmelfmann, M., Schaefer, A., and Reinberg, D. (2014). An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature* 516, 349-354.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* 45, 344-356.

Gibaja, V., Shen, F., Harari, J., Korn, J., Ruddy, D., Saenz-Vash, V., Zhai, H., Rejtar, T., Paris, C.G., Yu, Z., *et al.* (2016). Development of secondary mutations in wild-type and mutant EZH2 alleles cooperates to confer resistance to EZH2 inhibitors. *Oncogene* 35, 558-566.

Gonzalez, I., Aparicio, R., and Busturia, A. (2008). Functional characterization of the dRYBP gene in *Drosophila*. *Genetics* 179, 1373-1388.

Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M.R., Szumska, D., McGouran, J.F., Bhattacharya, S., Kessler, B.M., Brockdorff, N., and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. *Development* 143, 2716-2723.

Haber, J.E. (2012). Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. *Genetics* 191, 33-64.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol* 10, 1291-1300.

Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep* 17, 583-595.

He, Y., Selvaraju, S., Curtin, M.L., Jakob, C.G., Zhu, H., Comess, K.M., Shaw, B., The, J., Lima-Fernandes, E., Szewczyk, M.M., *et al.* (2017). The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex. *Nat Chem Biol* 13, 389-395.

Heinlein, C.A., and Chang, C. (2004). Androgen receptor in prostate cancer. *Endocr Rev* 25, 276-308.

Heinzel, T., Lavinsky, R.M., Mullen, T.M., Soderstrom, M., Laherty, C.D., Torchia, J., Yang, W.M., Brard, G., Ngo, S.D., Davie, J.R., *et al.* (1997). A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature* 387, 43-48.

Hildebrand, J.D., and Soriano, P. (2002). Overlapping and unique roles for C-terminal binding protein 1 (CtBP1) and CtBP2 during mouse development. *Mol Cell Biol* 22, 5296-5307.

Hill, A., and Bloom, K. (1987). Genetic manipulation of centromere function. *Mol Cell Biol* 7, 2397-2405.

Hodis, E., Watson, I.R., Kryukov, G.V., Arold, S.T., Imielinski, M., Theurillat, J.P., Nickerson, E., Auclair, D., Li, L., Place, C., *et al.* (2012). A landscape of driver mutations in melanoma. *Cell* 150, 251-263.

Hojfeldt, J.W., Laugesen, A., Willumsen, B.M., Damhofer, H., Hedehus, L., Tvardovskiy, A., Mohammad, F., Jensen, O.N., and Helin, K. (2018). Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nat Struct Mol Biol* 25, 225-232.

Holoch, D., and Margueron, R. (2017). Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends Biochem Sci* 42, 513-542.

Holohan, C., Van Schaeybroeck, S., Longley, D.B., and Johnston, P.G. (2013). Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer* 13, 714-726.

Huether, R., Dong, L., Chen, X., Wu, G., Parker, M., Wei, L., Ma, J., Edmonson, M.N., Hedlund, E.K., Rusch, M.C., *et al.* (2014). The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun* 5, 3630.

Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. *Genes Dev* 29, 1897-1902.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 7, 1534-1550.

Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell* 26, 565-577.

Isono, K., Fujimura, Y., Shinga, J., Yamaki, M., J, O.W., Takihara, Y., Murahashi, Y., Takada, Y., Mizutani-Koseki, Y., and Koseki, H. (2005). Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of Hox genes. *Mol Cell Biol* 25, 6694-6706.

Iwasaki, S., Floor, S.N., and Ingolia, N.T. (2016). Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. *Nature* 534, 558-561.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3, 318-356.

Jankowska, A.M., Makishima, H., Tiu, R.V., Szpurka, H., Huang, Y., Traina, F., Visconte, V., Sugimoto, Y., Prince, C., O'Keefe, C., *et al.* (2011). Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood* 118, 3932-3941.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074-1080.

Junco, S.E., Wang, R., Gaipa, J.C., Taylor, A.B., Schirf, V., Gearhart, M.D., Bardwell, V.J., Demeler, B., Hart, P.J., and Kim, C.A. (2013). Structure of the polycomb group protein PCGF1 in complex with BCOR reveals basis for binding selectivity of PCGF homologs. *Structure* 21, 665-671.

Jung, H., Lacombe, J., Mazzoni, E.O., Liem, K.F., Jr., Grinstein, J., Mahony, S., Mukhopadhyay, D., Gifford, D.K., Young, R.A., Anderson, K.V., *et al.* (2010). Global control of motor neuron topography mediated by the repressive actions of a single hox gene. *Neuron* 67, 781-796.

Jung, H., Mazzoni, E.O., Soshnikova, N., Hanley, O., Venkatesh, B., Duboule, D., and Dasen, J.S. (2014). Evolving Hox activity profiles govern diversity in locomotor systems. *Dev Cell* 29, 171-187.

Justin, N., Zhang, Y., Tarricone, C., Martin, S.R., Chen, S., Underwood, E., De Marco, V., Haire, L.F., Walker, P.A., Reinberg, D., *et al.* (2016). Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun* 7, 11316.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W., Muller, C.W., Vermeulen, M., and Muller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* 21, 569-571.

Kasinath, V., Faini, M., Poepsel, S., Reif, D., Feng, X.A., Stjepanovic, G., Aebersold, R., and Nogales, E. (2018). Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science* 359, 940-944.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res* 12, 996-1006.

Kieffer-Kwon, K.R., Nimura, K., Rao, S.S.P., Xu, J., Jung, S., Pekowska, A., Dose, M., Stevens, E., Mathe, E., Dong, P., *et al.* (2017). Myc Regulates Chromatin

Decompaction and Nuclear Architecture during B Cell Activation. *Mol Cell* 67, 566-578.e510.

Kim, H., Ekram, M.B., Bakshi, A., and Kim, J. (2015a). AEBP2 as a transcriptional activator and its role in cell migration. *Genomics* 105, 108-115.

Kim, H., Kang, K., Ekram, M.B., Roh, T.Y., and Kim, J. (2011). Aebp2 as an epigenetic regulator for neural crest cells. *PLoS One* 6, e25174.

Kim, T.W., Kang, B.H., Jang, H., Kwak, S., Shin, J., Kim, H., Lee, S.E., Lee, S.M., Lee, J.H., Kim, J.H., *et al.* (2015b). Ctbp2 Modulates NuRD-Mediated Deacetylation of H3K27 and Facilitates PRC2-Mediated H3K27me3 in Active Embryonic Stem Cell Genes During Exit from Pluripotency. *Stem Cells* 33, 2442-2455.

Klose, R.J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* 442, 312-316.

Knutson, S.K., Warholic, N.M., Wigle, T.J., Klaus, C.R., Allain, C.J., Raimondi, A., Porter Scott, M., Chesworth, R., Moyer, M.P., Copeland, R.A., *et al.* (2013). Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc Natl Acad Sci U S A* 110, 7922-7927.

Knutson, S.K., Wigle, T.J., Warholic, N.M., Sneeringer, C.J., Allain, C.J., Klaus, C.R., Sacks, J.D., Raimondi, A., Majer, C.R., Song, J., *et al.* (2012). A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nat Chem Biol* 8, 890-896.

Kops, G.J., Dansen, T.B., Polderman, P.E., Saarloos, I., Wirtz, K.W., Coffey, P.J., Huang, T.T., Bos, J.L., Medema, R.H., and Burgering, B.M. (2002). Forkhead transcription factor FOXO3a protects quiescent cells from oxidative stress. *Nature* 419, 316-321.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol Cell* 65, 432-446.e435.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116-120.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev* 22, 2799-2810.

Laherty, C.D., Yang, W.M., Sun, J.M., Davie, J.R., Seto, E., and Eisenman, R.N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell* 89, 349-356.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., *et al.* (2010). Jarid2 is a

PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat Cell Biol* 12, 618-624.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359.

Lau, M.S., Schwartz, M.G., Kundu, S., Savol, A.J., Wang, P.I., Marr, S.K., Grau, D.J., Schorderet, P., Sadreyev, R.I., Tabin, C.J., *et al.* (2017). Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. *Science* 355, 1081-1084.

Lee, H.G., Kahn, T.G., Simcox, A., Schwartz, Y.B., and Pirrotta, V. (2015). Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome Res* 25, 1170-1181.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301-313.

Lee, W., Teckie, S., Wiesner, T., Ran, L., Prieto Granada, C.N., Lin, M., Zhu, S., Cao, Z., Liang, Y., Sboner, A., *et al.* (2014). PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. *Nat Genet* 46, 1227-1232.

Lee, Y., Song, A.J., Baker, R., Micales, B., Conway, S.J., and Lyons, G.E. (2000). Jumonji, a nuclear protein that is necessary for normal heart development. *Circ Res* 86, 932-938.

Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565-570.

Lewis, P.W., Muller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* 340, 857-861.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev* 24, 368-380.

Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., *et al.* (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* 549, 287-291.

Li, X., Isono, K., Yamada, D., Endo, T.A., Endoh, M., Shinga, J., Mizutani-Koseki, Y., Otte, A.P., Casanova, M., Kitamura, H., *et al.* (2011). Mammalian polycomb-like Pcl2/Mtf2 is a novel regulatory component of PRC2 that can differentially modulate polycomb activity both at the Hox gene cluster and at Cdkn2a genes. *Mol Cell Biol* 31, 351-364.

Liefke, R., Karwacki-Neisius, V., and Shi, Y. (2016). EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape. *Mol Cell* 64, 659-672.

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, 550.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251-260.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., *et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461, 762-767.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343-349.

Martin, N., Popov, N., Aguilo, F., O'Loughlen, A., Raguz, S., Snijders, A.P., Dharmalingam, G., Li, S., Thymiakou, E., Carroll, T., *et al.* (2013). Interplay between Homeobox proteins and Polycomb repressive complexes in p16INK(4)a regulation. *Embo J* 32, 982-995.

McCabe, M.T., Graves, A.P., Ganji, G., Diaz, E., Halsey, W.S., Jiang, Y., Smitheman, K.N., Ott, H.M., Pappalardi, M.B., Allen, K.E., *et al.* (2012a). Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc Natl Acad Sci U S A* 109, 2989-2994.

McCabe, M.T., Ott, H.M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G.S., Liu, Y., Graves, A.P., Della Pietra, A., 3rd, Diaz, E., *et al.* (2012b). EZH2 inhibition

as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* 492, 108-112.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* 68, 283-302.

McGinty, R.K., Henrici, R.C., and Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* 514, 591-596.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell* 10, 1107-1117.

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, 1823-1828.

Mohammad, F., Weissmann, S., Leblanc, B., Pandey, D.P., Hojfeldt, J.W., Comet, I., Zheng, C., Johansen, J.V., Rapin, N., Porse, B.T., *et al.* (2017). EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat Med* 23, 483-492.

Mohd-Sarip, A., Cleard, F., Mishra, R.K., Karch, F., and Verrijzer, C.P. (2005). Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes Dev* 19, 1755-1760.

Mohd-Sarip, A., Lagarou, A., Doyen, C.M., van der Knaap, J.A., Aslan, U., Bezstarosti, K., Yassin, Y., Brock, H.W., Demmers, J.A., and Verrijzer, C.P. (2012).

Transcription-independent function of Polycomb group protein PSC in cell cycle control. *Science* 336, 744-747.

Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr Biol* 15, 942-947.

Morey, L., Aloia, L., Cozzuto, L., Benitah, S.A., and Di Croce, L. (2013). RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. *Cell Rep* 3, 60-69.

Morin, R.D., Johnson, N.A., Severson, T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., Kuchenbauer, F., *et al.* (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* 42, 181-185.

Morin, R.D., Mendez-Lago, M., Mungall, A.J., Goya, R., Mungall, K.L., Corbett, R.D., Johnson, N.A., Severson, T.M., Chiu, R., Field, M., *et al.* (2011). Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 476, 298-303.

Moris, N., Pina, C., and Arias, A.M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nat Rev Genet* 17, 693-703.

Motoyama, J., Kitajima, K., Kojima, M., Kondo, S., and Takeuchi, T. (1997). Organogenesis of the liver, thymus and spleen is affected in jumonji mutant mice. *Mech Dev* 66, 27-37.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol Cell* 53, 277-289.

Muller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr Opin Genet Dev* 16, 476-484.

Mullin, N.P., Gagliardi, A., Khoa, L.T.P., Colby, D., Hall-Ponsele, E., Rowe, A.J., and Chambers, I. (2017). Distinct Contributions of Tryptophan Residues within the Dimerization Domain to Nanog Function. *J Mol Biol* 429, 1544-1553.

Musselman, C.A., Avvakumov, N., Watanabe, R., Abraham, C.G., Lalonde, M.E., Hong, Z., Allen, C., Roy, S., Nunez, J.K., Nickoloff, J., *et al.* (2012a). Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nat Struct Mol Biol* 19, 1266-1272.

Musselman, C.A., Lalonde, M.E., Cote, J., and Kutateladze, T.G. (2012b). Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* 19, 1218-1227.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* 292, 110-113.

Negishi, M., Saraya, A., Mochizuki, S., Helin, K., Koseki, H., and Iwama, A. (2010). A novel zinc finger protein Zfp277 mediates transcriptional repression of the Ink4a/arf locus through polycomb repressive complex 1. *PLoS One* 5, e12373.

Nikoloski, G., Langemeijer, S.M., Kuiper, R.P., Knops, R., Massop, M., Tonnissen, E.R., van der Heijden, A., Scheele, T.N., Vandenberghe, P., de Witte, T., *et al.* (2010). Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet* 42, 665-667.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011). The dynamic architecture of Hox gene clusters. *Science* 334, 222-225.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The polycomb-group gene *Ezh2* is required for early mouse development. *Mol Cell Biol* 21, 4330-4336.

Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* 87, 953-959.

Olins, A.L., Carlson, R.D., Wright, E.B., and Olins, D.E. (1976). Chromatin nucleosomes: isolation, subfractionation and physical characterization. *Nucleic Acids Res* 3, 3271-3291.

Oliviero, G., Brien, G.L., Waston, A., Streubel, G., Jerman, E., Andrews, D., Doyle, B., Munawar, N., Wynne, K., Crean, J., *et al.* (2016). Dynamic Protein Interactions of the Polycomb Repressive Complex 2 during Differentiation of Pluripotent Cells. *Mol Cell Proteomics* 15, 3450-3460.

Palijan, A., Fernandes, I., Verway, M., Kourelis, M., Bastien, Y., Tavera-Mendoza, L.E., Sacheli, A., Bourdeau, V., Mader, S., and White, J.H. (2009). Ligand-

dependent corepressor LCoR is an attenuator of progesterone-regulated gene expression. *J Biol Chem* 284, 30275-30287.

Park, I.K., Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., and Clarke, M.F. (2003). Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* 423, 302-305.

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *Embo J* 23, 4061-4071.

Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010a). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464, 306-310.

Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N., *et al.* (2010b). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res* 38, 4958-4969.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139, 1290-1302.

Pengelly, A.R., Kalb, R., Finkl, K., and Muller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev* 29, 1487-1492.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11, 1650-1667.

Peters, J.M. (2006). The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nat Rev Mol Cell Biol* 7, 644-656.

Pirity, M.K., Locker, J., and Schreiber-Agus, N. (2005). Rybp/DEDAF is required for early postimplantation and for central nervous system development. *Mol Cell Biol* 25, 7193-7202.

Piunti, A., Hashizume, R., Morgan, M.A., Bartom, E.T., Horbinski, C.M., Marshall, S.A., Rendleman, E.J., Ma, Q., Takahashi, Y.H., Woodfin, A.R., *et al.* (2017). Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. *Nat Med* 23, 493-500.

Qi, W., Chan, H., Teng, L., Li, L., Chuai, S., Zhang, R., Zeng, J., Li, M., Fan, H., Lin, Y., *et al.* (2012). Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation. *Proc Natl Acad Sci U S A* 109, 21360-21365.

Qi, W., Zhao, K., Gu, J., Huang, Y., Wang, Y., Zhang, H., Zhang, M., Zhang, J., Yu, Z., Li, L., *et al.* (2017). An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED. *Nat Chem Biol* 13, 381-388.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., *et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406, 593-599.

Reisman, D., Glaros, S., and Thompson, E.A. (2009). The SWI/SNF complex and cancer. *Oncogene* 28, 1653-1668.

Ren, X., and Kerppola, T.K. (2011). REST interacts with Cbx proteins and regulates polycomb repressive complex 1 occupancy at RE1 elements. *Mol Cell Biol* 31, 2100-2110.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* 55, 347-360.

Rose, N.R., King, H.W., Blackledge, N.P., Fursova, N.A., Ember, K.J., Fischer, R., Kessler, B.M., and Klose, R.J. (2016). RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. *Elife* 5.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., and Reinberg, D. (2008). Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol* 28, 2718-2731.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., *et al.* (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet* 47, 1179-1186.

Seibler, J., Zevnik, B., Kuter-Luks, B., Andreas, S., Kern, H., Hennek, T., Rode, A., Heimann, C., Faust, N., Kauselmann, G., *et al.* (2003). Rapid generation of inducible mouse mutants. *Nucleic Acids Res* 31, e12.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7, 539.

Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A., and Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* 41, e28.

Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci U S A* 107, 20980-20985.

Song, Y., Shan, S., Zhang, Y., Liu, W., Ding, W., Ren, W., Xia, H., Li, X., Zhang, Q., Zhao, L., *et al.* (2012). Ligand-dependent corepressor acts as a novel corepressor of thyroid hormone receptor and represses hepatic lipogenesis in mice. *J Hepatol* 56, 248-254.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.

Streubel, G., Fitzpatrick, D.J., Oliviero, G., Scelfo, A., Moran, B., Das, S., Munawar, N., Watson, A., Wynne, K., Negri, G.L., *et al.* (2017). Fam60a defines a variant Sin3a-Hdac complex in embryonic stem cells required for self-renewal. *Embo J*, pii: e201696307.

Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.A., Jones, D.T., Konermann, C., Pfaff, E., Tönjes, M., Sill, M., Bender, S., *et al.* (2012). Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell* 22, 425-437.

Su, W.J., Fang, J.S., Cheng, F., Liu, C., Zhou, F., and Zhang, J. (2013). RNF2/Ring1b negatively regulates p53 expression in selective cancer cell types to promote tumor development. *Proc Natl Acad Sci U S A* 110, 1720-1725.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., *et al.* (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* 16, 1779-1791.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. In *Genes Dev*, pp. 815-826.

Takeuchi, T., Kojima, M., Nakajima, K., and Kondo, S. (1999). jumonji gene is essential for the neurulation and cardiac development of mouse embryos with a C3H/He background. *Mech Dev* 86, 29-38.

Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., and Higashinakagawa, T. (1995). Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev* 9, 1211-1222.

Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T., and Pasini, D. (2012). Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res* 40, 3403-3418.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* 339, 1546-1558.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004a). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873-878.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004b). Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* 14, 637-646.

Wen, W., Peng, C., Kim, M.O., Ho Jeong, C., Zhu, F., Yao, K., Zykova, T., Ma, W., Carper, A., Langfald, A., *et al.* (2014). Knockdown of RNF2 induces apoptosis by regulating MDM2 and p53 stability. *Oncogene* 33, 421-428.

Wilson, B.G., Wang, X., Shen, X., McKenna, E.S., Lemieux, M.E., Cho, Y.J., Koellhoffer, E.C., Pomeroy, S.L., Orkin, S.H., and Roberts, C.W. (2010). Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* 18, 316-328.

Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat Methods* 6, 359-362.

Woodcock, C.L., Safer, J.P., and Stanchfield, J.E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. *Exp Cell Res* 97, 101-110.

Wu, G., Broniscer, A., McEachron, T.A., Lu, C., Paugh, B.S., Becksfort, J., Qu, C., Ding, L., Huether, R., Parker, M., *et al.* (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet* 44, 251-253.

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, 1134-1146.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Cote, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* 2, 851-861.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., *et al.* (2016). Ensembl 2016. *Nucleic Acids Res* 44, D710-716.

Yu, M., Mazor, T., Huang, H., Huang, H.T., Kathrein, K.L., Woo, A.J., Chouinard, C.R., Labadorf, A., Akie, T.E., Moran, T.B., *et al.* (2012). Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Mol Cell* 45, 330-343.

Zhang, M., Wang, Y., Jones, S., Sausen, M., McMahon, K., Sharma, R., Wang, Q., Belzberg, A.J., Chaichana, K., Gallia, G.L., *et al.* (2014). Somatic mutations of SUZ12 in malignant peripheral nerve sheath tumors. *Nat Genet* 46, 1170-1172.

Zhang, Z., Jones, A., Sun, C.W., Li, C., Chang, C.W., Joo, H.Y., Dai, Q., Mysliwiec, M.R., Wu, L.C., Guo, Y., *et al.* (2011). PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. *Stem Cells* 29, 229-240.

Zhu, S., Xiang, J.F., Chen, T., Chen, L.L., and Yang, L. (2013). Prediction of constitutive A-to-I editing sites from human transcriptomes in the absence of genomic sequences. *BMC Genomics* 14, 206.

(D) Top: A schematic representation of the coding exons of LCOR (red), C10ORF12 (blue) and LCOR-Cra_b (black). Bottom: Western blot analysis on lysates of HEK293T cells stably expressing shRNAs targeting the indicated coding exons.

(E) A schematic of the LCOR and PALI1 proteins showing their shared N-terminal nuclear receptors box and CTBP1/2 binding motifs. The grey shaded box indicates regions of identical protein sequence.

(F) Western blot analysis using the indicated antibodies of FLAG IPs of whole cell lysates of HEK293T cells ectopically expressing LCOR, C10ORF12 or PALI1 protein.

(G) Histone methyltransferase assay using recombinant PRC2 with a titrated addition of either LCOR or PALI1 protein, as indicated. Recombinant oligonucleosomes were used as a substrate. The PRC2 complex used contained FLAG-EED, SUZ12, EZH2 and RBBP4. Incorporation of ³H from radiolabeled SAM was used as a readout of methyltransferase activity. Quantification shown is relative signal normalized to PRC2 only. N/A denotes saturated signal that was deemed unquantifiable. A representative image is shown from one of three biological replicates.

(H) Coomassie staining of recombinant complex/ proteins used for HMT assay.

Experiments performed by Emilia Jerman and Daniel Holoch.