Understanding the Diverse Roles of Polycomb-like Proteins in the Regulation of Chromatin and Transcriptional Landscapes

Evan Healy

Thesis submitted to Trinity College Dublin
for the degree of Doctor of Philosophy

2019

Thesis Supervisor: Prof. Adrian 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.

_____

Evan Healy, September 2018

Acknowledgements

Firstly, I would like to offer a huge thank you to my supervisor, Prof. Adrian Bracken, for providing me with the wonderful opportunity to do a PhD in his lab. His mentorship and guidance has been second to none over the years. He has taught me a great deal during the course of my PhD, the most important of which being the difference between apples and oranges! I'm sure I will carry these teachings with me throughout the rest of my career. I would also like to thank all the collaborators without whom the work in this PhD thesis would not have been possible; Elisa Fadda from NUI Maynooth, Andrew Flaus in NUI Galway, David O'Connell in UCD, Haruhiko and Yoko Koseki from RIKEN, and finally Aoife McLysaght, Alan Rice and Darren Fitzpatrick from the Smurfit Institute of Genetics here in Trinity College. I am extremely grateful to Darren for all of his hard work in analysing my next generation sequencing datasets, I would have been completely and utterly lost without him.

In addition, this PhD thesis would not have been possible without the funding provided by the Irish Research Council and Science Foundation Ireland.

To the members of the Bracken Lab, past and present, I am deeply indebted to you. The welcoming, warm and collaborative atmosphere you guys have created makes for an incredibly productive and fun lab environment, one which I always enjoy coming into. I would like to personally thank Gerry Brien for taking me under his wing at the start of my PhD, Carol O' Brien for being a wonderful PhD companion and friend, as well as Eleanor Glancy, Rachel McCole, Marlena Mucha and Orla Deevy for your kindness in helping me through everything. I am particularly grateful to Eric Conway, who has supported me so much over the years, and from whom I have learned a great deal. I look forward to having you as a friend, both in and out of the lab for many years to come.

To Sarah Nolan, a long time ago you told you me I would be a doctor one day! This is maybe not the real thing, but it's close! I am extremely grateful the love, support and encouragement you have given me over the years.

I would also like to thank my aunt and uncle Sylvia and Mike O'Grady, who put a roof over my head and food in my belly when I needed it most throughout the course of my PhD!

And finally I would like to express my deepest gratitude to my loving family. To my mother, Yvonne, my father, John, and my sister, Jane, I am beholden to you for your constant and unwavering love. You have gone above and beyond to support me, not just with my studies and PhD, but with every aspect of my life. I cherish your love and can say nothing other than, Thank You, with all of my heart.

Table of Contents

Table of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Tables

Summary

The Polycomb repressive complex 2 (PRC2) is a multiprotein chromatin modifying complex that is essential for vertebrate development and frequently deregulated in cancer. In higher eukaryotes, PRC2 is responsible for all mono-, di- and tri-methylation of Histone H3 at Lysine 27, and is composed of core subunits EED, SUZ12, and EZH1/2, as well as accessory components, Polycomb-like 1-3 (PCL1-3), EPOP, PALI1/2, AEBP2 and JARID2. Two distinct subtypes of PRC2 exist, termed PRC2.1 (containing PCL1-3, EPOP and PALI1/2) and PRC2.2 (containing AEBP2 and JARID2), however, little is known about their respective overlapping and divergent functions. As critical components of the PRC2.1 complex, PCL1-3 represent an excellent avenue to further explore the roles of PRC2 in cell fate decisions during differentiation and development, but also holds promise for future targeted cancer therapies. This PhD thesis addresses the various roles of PCL proteins in both differentiated adult cells and pluripotent embryonic stem cells (ESCs). Firstly, I explore the biochemical functions of PCL1 in the maintenance of cellular quiescence through a novel PCL1-p53 regulatory axis. Secondly, through the analysis of all three PCL proteins in cycling and quiescent human cells, I have defined a novel catalytically inactive form of PRC2, lacking SUZ12, that exists in quiescent cells. Thirdly, by analysing the effects of genetic knockout of all three *Pcl* genes in ESCs I characterise roles for distinct classes of PRC2 subcomplex assemblies, PRC2.1 and PRC2.2. Taken together, I believe these data will contribute to a better mechanistic understanding of the roles of Polycomb proteins during cell fate decisions and in complex biological processes such as carcinogenesis and embryogenesis.

Abbreviations

Aebp2-AE binding protein

Apc- Anaphase promoting complex

Bcor- BCL6 Co-repressor

Bcorl1- BCL6 Co-repressor Like 1

Bmi1- B Lymphoma Mo-MLV Insertion Region 1 Homolog

BRD4- Bromodomain-containing protein 4

C10ORF12- Chromosome 10 open reading frame 12

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

CSC- Cancer Stem Cell

DIPG- Diffuse Intrinsic Pontine Glioma

DNMT1/3A/3B/3C- DNA Methyltransferase 1/3A/3B/3C

EED- Embryonic Ectoderm Development

Epop- Elongin BC and Polycomb Repressive Complex 2 associated protein

ESC- mouse Embryonic Stem Cell

EZH1/2- Enhancer of Zeste 1/2

GST- Glutathione S-Transferase

G0- Quiescence

H2A- Histone 2A

H2AK119ub1- Histone 2A lysine 119 mono-ubiquitination

H3-Histone 3

Hdac- Histone Deacetylase

***Hox* genes**- Homeobox genes

HP1- Heterochromatin Protein 1

IP-immunoprecipitation

INK4A- Inhibitor of cyclin dependent Kinase 4

JARDI2- Jumonji and AT-Rich Interaction Domain Containing 2

JMJD3/KDM6B- Lysine Demethylase 6B

Kdm2b- Lysine demethylase 2b

LC-MS/MS- Liquid chromatography mass spectrometry

MDM2- MDM2 proto-oncogene

me1- mono-methylation

me2- di-methylation

me3- tri-methylation

MGA- Max gene associated protein

MTF2- Metal Response Elements Binding Transcription Factor 2

mRNA- messenger RNA

ncPRC1- non-canonical Polycomb Repressive Complex 1

NSD1- Nuclear Receptor Binding SET Domain Protein 1

ORF- open reading frame

PALI1- Polycomb associated *LCOR* isoform 1

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

PcG- Polycomb group Proteins

PCR- polymerase chain reaction

Ph/ Phc- Polyhomeotic

PHD- Plant Homeodomain

PHF1- PHD Finger Protein 1

PHF19- PHD Finger Protein 19

Pho- Pleiohomeotic

PRC1- Polycomb Repressive Complex 1

PRC2- Polycomb Repressive Complex 2

PRE- Polycomb Response Element

PTM- Post-translational Modification

RT-qPCR- Reverse Transcription quantitative PCR

RYBP- RING1 and Yy1 binding protein

SCML2A/B- SCM Polycomb Group Protein Like 2 A/B

SNF5/SMARCB1- SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily B, Member 1

SKP1- S-Phase Kinase Associated Protein 1

SWI/SNF- SWItch/Sucrose Non-Fermentable

SUZ12- Suppressor of Zeste 12 homolog

UIM- Ubiquitin Interacting Motif (JARID2)

USP7- Ubiquitin Specific Peptidase 7

UTX/KDM6A- Lysine Demethylase 6A

YAF2- Yy1 Associated Factor 2

Xist- X-inactive Specific Transcript

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., Jerman, E., **Healy, E**., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D.J., Mucha, M.*, et al.* (2018). A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. *Mol Cell 70*, 408-421.

Healy, E., Brien G.L., O'Brien C., and Adrian P. Bracken (2018). A novel catalytically inactive PRC2 complex unique to non-dividing quiescent cells. *In preparation.*

Healy, E., Marlena, M., Fitzparick, D.J., Conway, E., Koseki, Y., Koseki, H., and Adrian P. Bracken (2018). PRC2.1 and PRC2.2 synergise to co-ordinate H3K27 tri-methylation. *In preparation.*

*- Authors contributed equally

Chapter 1

Introduction

1.1 General Introduction: Cell fate transitions

The ability of multi-cellular organisms to form a vast array of cell types and tissues with a variety of functions and specificities is a highly regulated and complex process. This diversity of tissues arises from a single fertilised egg, or zygote, with a single underlying DNA sequence, or genetic code. The full development of a complex organism requires a relationship between two important processes, an increase in cell number coupled with the ability to diversify from the initial zygotic cell. This phenotypic diversification of cell types is called differentiation, and the fidelity of this process is essential for the correct development of any multi-cellular organism (Moris et al., 2016). Differentiation occurs when a stem cell receives a stimulus to divide or to specify its function, resulting in a cell fate transition. Importantly, differentiated cells arising from cell fate transitions retain the identical genomic sequence of the cell from which it descended. Hence, cell fate determination can be said to be governed by a switch in the transcriptional programme of a certain cell type. For example, an adult muscle stem cell may receive a cue to differentiate into a more specified cell type. It achieves this through simultaneously repressing transcription of genes involved in the maintenance of “stemness”, while activating lineage specific genes (Bracken and Helin, 2009). This is a highly reproducible process borne out by the ability of mouse embryonic stem cells to recapitulate all tissues of an adult mouse after injection into a host blastocyst (Bradley et al., 1984; Keller, 2005).

One of the enduring models about how a single stem cell makes these critical cell fate decisions, or transitions is Waddington’s “epigenetic landscape”, famously represented as a pebble following an existing, diverging path down a hill (Goldberg et al., 2007). This is a simplified depiction of a complex series of events and realistically, this process is not as stochastic as it may seem. In order for a cell fate transition to be effectively maintained, all daughter cells must inherit complex transcriptional programmes through many cell divisions, and the faithful inheritance of these programmes is governed by numerous mechanisms. Generally speaking, transcription is regulated through covalent modification of

the underlying DNA or of the histone proteins around which DNA is wrapped. These modifications contribute to determining whether a gene will be active or repressed and importantly can be inherited through cell division events (Hansen et al., 2008; Smith and Meissner, 2013). The heritable regulation of transcription in this way provides the basis for how an undefined stem cell with a single copy of the genetic code, can make critical, temporal cell fate decisions that contribute to the specification of tissues and eventual complexity of a multi-cellular organisms.

1.2 Covalent methylation of DNA

Although regulation of gene expression does occur in *cis* through the function of transcription factors, changes in transcription resulting in cell fate transitions are more generally regulated by complementary processes such as DNA methylation. In mammals, the methylation of DNA occurs at cytosine nucleotides and is generally considered to be a repressive mark associated with broad regions of the genome. Interestingly, the 5'-end of genes associated with promoters contain a frequency of cytosine/guanine dinucleotides that is approximately 10 times higher than the genome average (Bird, 1986). DNA methylation is curiously excluded from these particular potions of the genome, which are termed CpG islands. (Bird et al., 1985; Smith and Meissner, 2013). Intriguingly, the lack of CpG methylation and high CG content at these islands are evolutionarily coupled. The mechanism proposed to have resulted in this high frequency of unmethylated CpG's at promoter regions involves the increased mutability of 5-methyl-cytosine (5mC) due to due to inaccurate repair of deaminated 5mC's, resulting in the introduction of thymine bases following DNA replication (Bird, 1980). This results in increased CpG mutability which causes methylated regions to lose CpGs rapidly. In contrast, unmethylated CpGs sustain higher CpG content since they are not prone to hypermutability by 5mC deamination (Bird, 1986; Cohen et al., 2011).

Covalent methylation of cytosine is performed by four essential conserved DNA methyltransferase enzymes, DNMT1, DNMT3A, DNMT3B and DNMT3C (Barau

et al., 2016; Smith and Meissner, 2013). During early embryogenesis, the developing embryo undergoes rapid DNA demethylation (Kafri et al., 1992). Following this event, *de novo* DNA methylation is deposited by DNMT3A/B/C at CG dense regions of the genome, but importantly not at CpG islands. Modification of DNA in this way is critical for establishing the methylation pattern of the genome that will be inherited by all daughter cells (Bird, 2002). This inheritance of DNA methylation states is regulated during S-phase by DNMT1, which has a preference for hemi-methylated DNA left behind following DNA replication. DNMT1 is recruited to the replication fork through an interaction with PCNA and reconstitutes the full methylation status of the parent strand (Smith and Meissner, 2013). In this way transcriptional memory can be passed from parent to daughter cell.

1.3 Histones and Chromatin

Aside from direct covalent modification of DNA, there is a further layer of complexity allowing for mitotically heritable patterns of gene expression, post-translational modification of histone proteins. In humans, when DNA is fully extended, the three billion bases would stretch to over two metres. This represents a major challenge for a single cell, to organise this molecule into a space of less than 10μm (McGinty and Tan, 2015). Compaction of this scale is established through formation of a DNA-protein complex, called chromatin. The nucleosome is the fundamental unit of chromatin and is composed of an octameric core of highly conserved histone proteins. This histone octamer is made up of two copies each of histone proteins, H2A, H2B, H3 and H4, around which 147bp of DNA are wrapped (Luger et al., 1997). This nucleosome core particle is connected to adjacent DNA through a segment of linker DNA (~25bp), which is often associated with another histone protein, H1, and is essential for higher order chromatin structures (Bednar et al., 2017).

A unique feature of core histones is that they possess unstructured N-terminal tails that protrude from the nucleosome subunit allowing for modulation of chromatin compaction through addition or removal of covalent modifications

(Figure 1.1). Post-translational modification (PTM) of specific residues within the histone tails can facilitate formation of open regions of chromatin conducive to transcription termed, euchroamtin, or condensed heterochromatin which is refractory to transcription. There is scope for a vast array of possible combinations of histone PTMs giving rise to an enormous potential for functional output. For example, acetylation of Lysine residues on the histone H3 tail by histone acetyltransferase enzymes (HATs) facilitates a more open chromatin structure and is often found at promoters of active genes. Of the many histone PTMs, acetylation of Lysine possesses the greatest ability to directly unfold chromatin, as it can neutralise basic charge of the residue. As well as directly affecting chromatin states, histone PTMs can also provide a template to recruit proteins with specific modular binding domains, thereby allowing for a further degree of complexity to the regulation of transcription (Figure 1.2) (Kouzarides, 2007). BRD4, a member of the BET family proteins, represents an elegant example of this process. BET family proteins are acetyl-lysine readers, and BRD4 through its bromo domain can specifically read acetylated Lysine 27 of histone H3, which in turn can recruit positive regulators of transcription such as the transcription elongation factor (pTEFb) and Mediator complexes to chromatin (Figure 1.2) (Jang et al., 2005; Jiang et al., 1998).

These histone PTMs, or marks, are not always directly involved in the regulation of transcription at promoters or enhancers of genes. They are also important for delineating distinct functional regions of the genome such as centromeres. For example, tri-methylation of histone H3 at Lysine 9 (H3K9me3) is critical for marking centromeric heterochromatin (Rea et al., 2000). The H3K9me3 PTM is essential to maintain transcriptional repression at structural regions of the genome such as centromeres, as spurious activity from these regions can cause mitotic defects (Hill and Bloom, 1987). The diversity of histone PTMs means that they have the potential to influence a huge array of biological processes from embryonic development to becoming dysregulated in many human pathologies such as cancer.

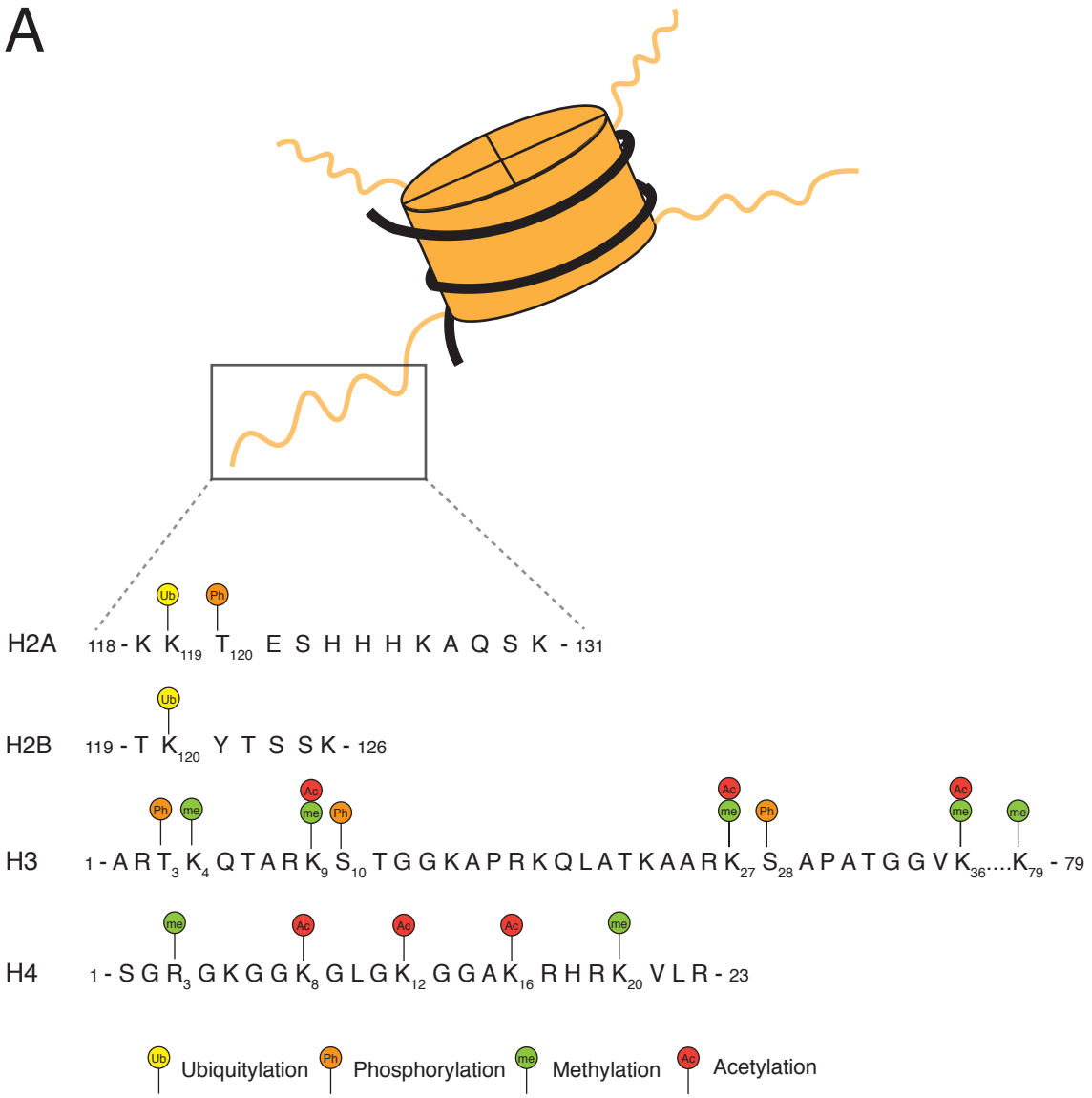


Figure 1.1 Post-translational modifications of N-terminal histone tails.
(A) (Top) Illustration of the nucleosome, the basic subunit of chromatin. The nucleosome octamer contains two copies each of H2A, H2B, H3 and H4, around which 147bp of DNA is wrapped. N-terminal histone tails protrude from the octamer. (Bottom) The histone tails are subject to a wide variety of covalent post-translational modifications which have varying effects on chromatin and transcription.

-

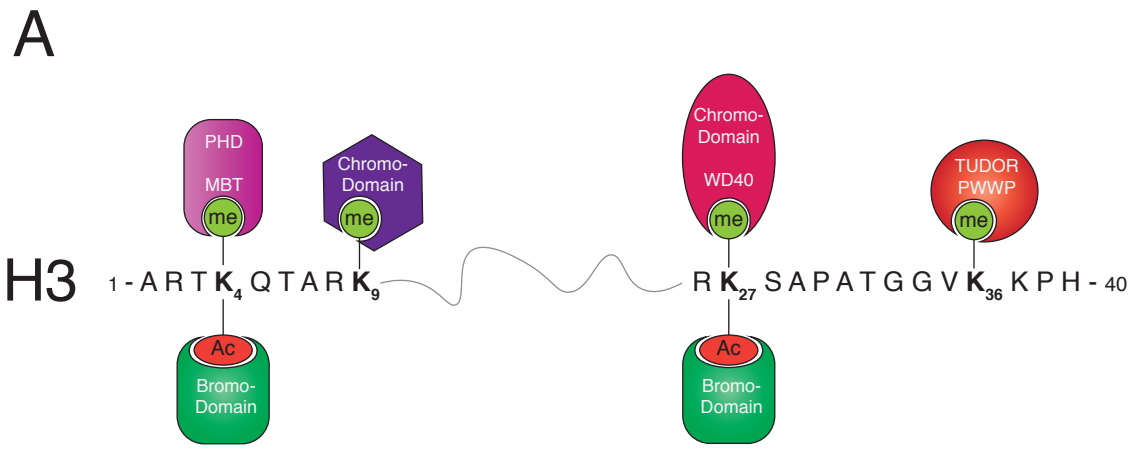


Figure 1.2 Chromatin associated epigenetic reader domains.
(A) Recognition of N-terminal histone H3 post-translational modifications by reader domains, that are contained within many chromatin associated protein complexes. These domains allow for chromatin associated complexes to perceive the epigenetic landscape of a cell**.**

1.4 Chromatin modifying proteins

The combinatorial nature of histone PTMs and their roles in the maintenance of chromatin states led to the development of a “histone code” hypothesis, whereby the transcriptional potential of a given gene could be predicted by the PTMs associated with that gene (Jenuwein and Allis, 2001). The different classes of proteins involved in regulating histone PTMs can be grouped into, writers, readers, and erasers. Writer proteins are the enzymes that catalyse the deposition of specific histone PTMs. For example, the histone methyltransferase (HMT) enzyme, SUV39H1 selectively methylates Lysine 9 of histone H3 (H3K9me), through an evolutionarily conserved SET domain (Rea et al., 2000). As mentioned previously, H3K9me3 is associated with heterochromatin formation, and this is achieved through the action of a member of the second class of proteins, reader proteins. Reader proteins possess modular structural domains capable of interacting with specific histone PTMs (Figure 1.2). In the case of H3K9me3, a reader protein HP1 contains a specific class of domain, called a chromo domain. The HP1 chromo domain has high-affinity for H3K9me3 and is capable of homo-dimerising with itself, thereby facilitating chromatin compaction and heterochromatin formation (Lachner et al., 2001). Other classes of reader domains with affinities for different PTMs include, Bromo domains (Acetyl-Lysine), Tudor domains (methyl-Lysine) and PHD domains (methyl-Lysine) (Figure 1.2) (Musselman et al., 2012b).

Interestingly, a defining feature of these histone PTMs is that they are dynamic in nature and their occupancy at chromatin is not necessarily permanent. Demethylation, deacetylation, dephosphorylation, deubiquitination and histone turnover all contribute to the transient nature of histone PTMs. Owing to this, another class of protein involved in the regulation of histone modifications are, eraser proteins. Examples of these are, JMJD3 and UTX, which remove repressive tri-methylation at Lysine 27 of H3 (H3K27me3) thereby facilitating gene activation during differentiation (Agger et al., 2007).

Chromatin accessibility can be regulated by an additional process that does not include direct catalysis, removal or recognition of a histone PTM. This mechanism involves large ATP-dependent remodelling complexes that permit compaction or decompaction of chromatin through eviction or sliding of nucleosomes (Mohrmann and Verrijzer, 2005). These complexes are usually involved in the promotion of transcription by opening chromatin or exposing transcription factor binding sites at gene promoters or enhancers (Clapier et al., 2017). The most well studied complex is the SWI/SNF complex and is one of two-dozen predicted chromatin remodelling complexes in mammals (Kadoch and Crabtree, 2015). These classes of chromatin modifying proteins and complexes are critical for maintaining transcriptional homeostasis in normal healthy cells. Their importance is demonstrated by reports that chromatin centred processes are tightly linked to the development of cancer. Indeed, the advent of next generation sequencing has established that nearly all cancers have mutations in genes encoding chromatin regulators. For this reason, there has been a huge effort to develop drugs targeting specific proteins involved in these processes (Brien et al., 2016; Kadoch and Crabtree, 2015).

1.5 Polycomb Group Proteins

Polycomb group proteins (PcG) are among the most studied chromatin modifying complexes and were first identified in *Drosophila* mutants that displayed improper body segmentation (Lewis, 1978). Initial studies proposed the *Drosophila* gene, Polycomb (*Pc*), as an essential negative regulator of homeotic gene (HOX) expression, required for correct anterior-posterior segmentation during embryogenesis. Subsequent studies have shown that PcG proteins are evolutionarily conserved from *Drosophila* to mammals, with their critical roles in development highlighted by knockout studies in mice showing that their loss results in early embryonic lethality (Laugesen and Helin, 2014). As well as embryogenesis, PcG proteins have essential roles in maintaining the correct identities of stem, progenitor and differentiated cells, achieving this primarily through the transcriptional repression of key developmental genes during differentiation (Bracken et al., 2006; Lee et al., 2006). In undifferentiated stem

cells, PcG proteins are bound at lineage specific genes, where they repress transcription and thereby maintain cellular identity. Once a stimulus to differentiate is received, PcG proteins are displaced from these lineage genes, and re-located to stem cell specific promoters, allowing for changes in transcriptional profiles during differentiation (Figure 1.3) (Bracken and Helin, 2009; Pasini et al., 2007). To date, the mechanisms of PcG recruitment and displacement remain relatively poorly understood.

1.6 Polycomb Repressive Complex 1 and Polycomb Repressive Complex 2

Since their initial discovery, a huge amount of research has focused on understanding the exact mechanisms of action of PcG proteins. Primarily, they function as two large multimeric protein complexes, Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2) (Figure 1.4). Both complexes function as writer enzymes with distinct catalytic activities. While PRC1 can ubiquitinate H2A at Lysine 119 (H2AK119ub), through a Ring-PCGF E3 ligase heterodimer, PRC2 is responsible for all mono-, di-, and tri-methylation of histone H3 at Lysine 27 (H3K27me1/2/3) through the methyltransferase activity of conserved SET domain containing enzymes, EZH1 and EZH2 (Cao et al., 2005; Ferrari et al., 2014; Hojfeldt et al., 2018; Muller et al., 2002). The H3K27me3 and H2AK119ub histone PTMs are associated with transcriptional repression of target genes (Figure 1.5).

The mammalian PRC2 complex is composed primarily of catalytic subunit, EZH1 or EZH2, along with essential core members, SUZ12 and EED (Margueron and Reinberg, 2011). All three of these proteins are critical for the stability of EZH1/2 and methyltransferase activity, while, the histone binding proteins RBBP4/7, seem to be important for complex activity *in vivo* (Margueron and Reinberg, 2011). While PRC2 along with its H3K27me3 mark are found at the promoters of developmentally repressed genes and are responsible for the maintenance of cellular identity, the H3K27me2 PTM is enriched on nascent histones shortly after DNA replication and is stably associated genome-wide with intergenic regions and inactive enhancers. (Alabert et al., 2015; Ferrari et al., 2014; Lee et al., 2015)

A

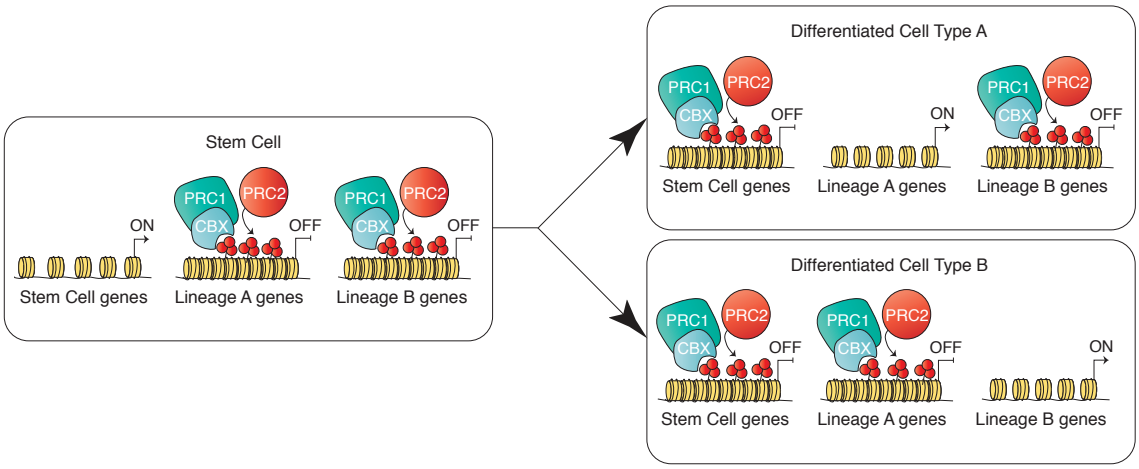


Figure 1.3 The function of Polycomb group proteins in cell fate transitions and lineage commitment.

(A) (Left) In pluripotent stem cells, stem cell genes governing pluripotency are active, and as such, the chromatin structure at their promoter regions is open and facilitates transcription. However, lineage specific genes such as muscle or neural cell transcription factors are switched off, and are not actively transcribed. The repression of these genes is governed by the enzymatic activities of the PRC1 and PRC2 complexes. (Right) In more differentiated cell types, stem cell genes are silenced, this repressed state is maintained by the PRC1 and PRC2 complexes. For example, in differentiated cell type A, Lineage A genes are active, whereas lineage B genes are silenced by PRC1 and PRC2, and vice versa. Figure adapted from Conway, Healy and Bracken 2015.

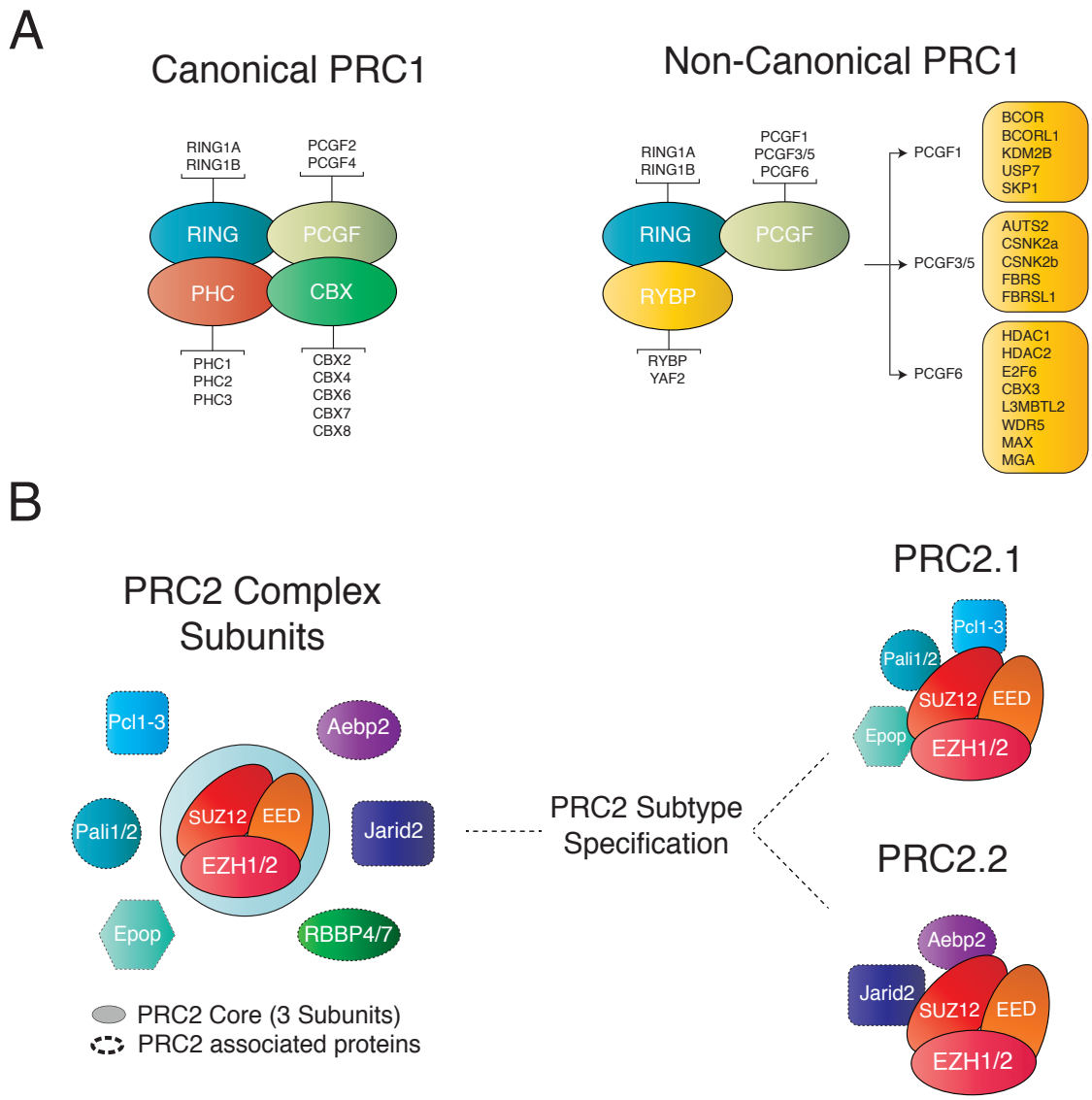


Figure 1.4 PRC1 and PRC2 exhibit variable subunit composition.
(A) Depiction of the variable subunit composition of canonical and non-canonical PRC1 complexes.
(B) Depiction of the core and accessory PRC2 subunits, as well as the newly described PRC2.1 and PRC2.2 subcomplexes.

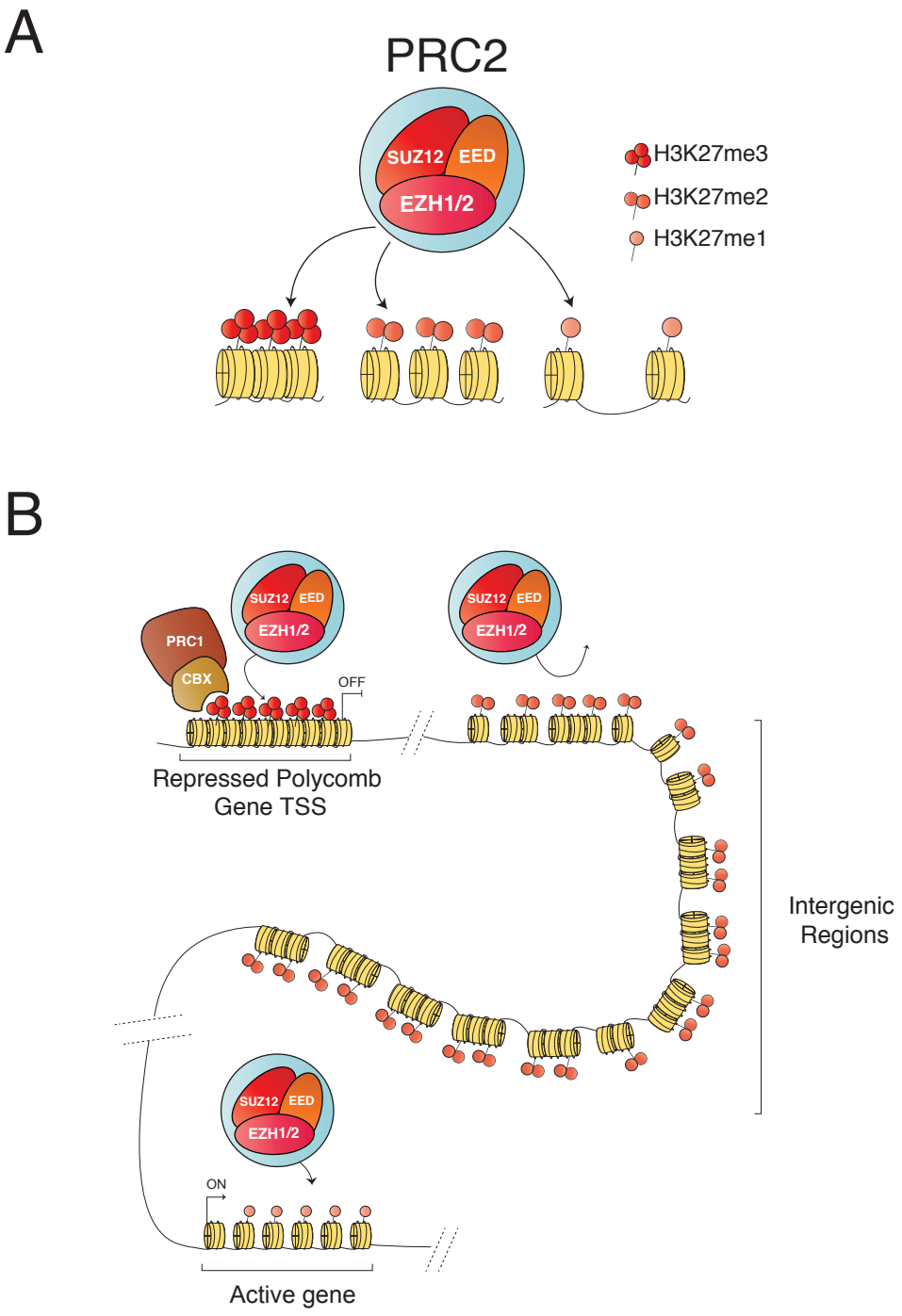


Figure 1.5 Divergent roles of PRC2 mediated Histone H3 Lys 27 methylations. (A) The PRC2 complex is capable of catalysing mono-, di, or tri-methyl groups at Lysine 27 of histone H3.

(B) Illustration of the genomic distribution of Histone H3 Lys 27 methylations. H2K27me1 is found at active gene bodies, H3K27me3 is spread broadly across 70% of the genome at intergenic regions, whereas H3K27me3 is found at the promoters of developmentally repressed genes. Figure adapted from Conway, Healy and Bracken, 2015.

(Figure 1.5). Intriguingly, PRC2 is not co-localised with H3K27me2 at these intergenic sites (Conway et al., 2015). This has been rationalised by the fact that H3K27me2 is relatively easier for PRC2 to generate and is thereby potentially mediated by a transient interaction of the complex with chromatin (Sneeringer et al., 2010). Furthermore, it is thought that the H3K27me2 modification is the default mode for H3K27 and its genome-wide occupancy represents a "repressive blanket" to prevent the misfiring of cell-type specific enhancers and promoters of alternative lineages (Conway et al., 2015). H3K27me1, on the other hand is associated with the gene bodies of actively transcribed genes (Figure 1.5). Recent studies have shown that PRC2 is responsible for all H3K27me1, which raises intriguing questions about the roles of PRC2 at these sites (Hojfeldt et al., 2018).

How PRC2 and other PcG complexes are targeted to their specific genomic loci remains a key open question. In *Drosophila*, PcG complexes are recruited to chromatin via *cis*-regulatory DNA sequences called Polycomb Response Elements (PREs) (Schuettengruber et al., 2017; Simon et al., 1993). However, apart from an unmethylated CG dinucleotide enriched at CpG islands, no defined DNA signature has been attributed to PcG localisation in mammals, leading to a focus on the specific targeting abilities of RNA. RNA mediated recruitment and localisation of PcG proteins to their respective genomic sites has been extensively studied and the best example of this is during the process of X chromosome inactivation in female mammals during development, a process mediated by a *cis*-acting RNA called *Xist* (Brockdorff et al., 1992). *Xist* RNA can directly recruit PRC2 to the inactive X, which then serves to establish and maintain the inactive state in all subsequent cell divisions (Brockdorff et al., 1992; de Napoles et al., 2004; Silva et al., 2003; Zhao et al., 2008). A recent elegant study suggested that Polycomb recruitment to the inactive X can also be initiated by PCGF3/5 containing PRC1 complexes through the deposition of H2AK119ub following *Xist* expression (Almeida et al., 2017; Zhao et al., 2008). Establishment of H2AK119ub then acts to recruit PRC2, thereby facilitating H3K27me3 deposition and robust X chromosome inactivation (Almeida et al., 2017). These

elegant observations led to the hypothesis that non-coding RNAs (ncRNAs) may provide specific targeting information for PcG complexes at other genomic regions in somatic cells. Intriguingly, PRC2 has recently been shown to promiscuously bind nascent RNAs at essentially all active genes, proposing a mechanism whereby, instead of providing targeting specificity, RNA competes with chromatin for binding to PRC2 and that this mutually exclusive relationship underpins the genome wide occupancy of PRC2 (Beltran et al., 2016).

Emerging evidence suggests, it is more likely that specific PRC2 recruitment at the majority of its genomic sites comes from an associated PRC2 interacting protein (Holoch and Margueron, 2017). The core PRC2 complex is known to associate with a number of accessory subunits in a substoichiometric manner, including JARID2, AEBP2, EPOP, PALI1/2 and PCL1/2/3 (Figure 1.4) (Ballare et al., 2012; Beringer et al., 2016; Brien et al., 2012; Conway et al., 2018; Grijzenhout et al., 2016; Liefke et al., 2016; Pasini et al., 2010). These factors can modulate PRC2 activity *in vitro* and *in vivo* (Holoch and Margueron, 2017*)*, and may also define independent mechanisms of PRC2 recruitment, in which PCL proteins target the complex to CpG islands as well as H3K36me2/3 marked chromatin, while JARID2 can bring PRC2 to genomic regions marked by H2AK119ub (Choi et al., 2017; Cooper et al., 2016; Li et al., 2011). Moving forward, it will be critically important to delineate the relative targeting contributions, if any, of each accessory PRC2 component.

The PRC1 complex exhibits a more variable biochemical composition, however, all PRC1 complexes share an essential core module consisting of a Ring-PCGF heterodimer. This Ring-PCGF core is an E3 ligase enzyme, capable of catalysing the deposition of H2AK119ub (Wang et al., 2004). There are two Ring variants, RING1A and RING1B, which dimerise with one of six PCGF proteins (PCGF1-6) (Figure 1.4). The H2AK119ub PTM and components of the PRC1 complex co-localise at the promoters of developmentally repressed genes, and are required for effective Polycomb mediated repression and chromatin compaction (Rose et al., 2016; Wang et al., 2004). The six PCGF paralogs allow for different possible

combinations of the PRC1 complex. While dimerization occurs through a conserved RING domain, other regions of the PCGF proteins are more divergent and provide the specificity for arrangement of the remaining PRC1 complex components (Junco et al., 2013). The large degree of possible PRC1 combinations can be broadly separated into two functionally distinct complexes called, canonical (cPRC1) and non-canonical (ncPRC1) PRC1. Whereas, cPRC1 functions in the physical compaction of chromatin (Isono et al., 2013; Lau et al., 2017), the primary role of ncPRC1 is catalysis of H2AK119ub (Blackledge et al., 2014; Endoh et al., 2017; Rose et al., 2016). In addition to the Ring-PCGF core, cPRC1 complexes contain Chromobox (CBX), and Polyhomeotic (PHC) subunits, however, the exact composition of cPRC1 composition appears to vary depending on cell type (Gao et al., 2012). For example, in mouse embryonic stem cells (ESCs), PCGF2 (MEL18) and CBX7 are predominantly expressed, yet when ESCs are then induced to differentiate, PCGF4 (BMI1) and CBX8 become expressed, while PCGF2 and CBX7 are downregulated (Kloet et al., 2016; Morey et al., 2012). A “hierarchical” model of Polycomb recruitment can explain the genome-wide localisation of cPRC1 complexes. This model dictates that PRC2-mediated deposition of H3K27me3 is recognised or “read” by the chromo domain of the CBX subunit which leads to subsequent recruitment of cPRC1, and rationalises the co-localisation of cPRC1 complexes with PRC2 and H3K27me3 (Bracken et al., 2006; Fischle et al., 2003; Min et al., 2003). Functionally, it is clear cPRC1 complexes play important roles in chromatin compaction. This has been elegantly illustrated by studies showing cPRC1 complexes can form dense regions of chromatin, known as Polycomb-bodies (Eskeland et al., 2010). The PHC1-3 subunits are key in Polycomb-body formation by cPRC1 complexes (Isono et al., 2013). They importantly contain sterile alpha motif (SAM) domains, with further interacting regions contained within this domain. An end helix and midloop helix allow for oligomerisation of adjacent PHC subunits resulting in the formation of a “chain” of PHC subunits. Through this elegant biochemical mechanism, cPRC1 complexes can form higher-order chromatin structures, allowing for long range, repressive interactions (Boettiger et al., 2016; Kundu et al., 2017; Schoenfelder et al., 2015). Polycomb body formation is thought to be

the primary mechanism for cPRC1 mediated transcriptional repression, and not H2AK119ub deposition by the complex.

Non-canonical PRC1 complexes are functionally and biochemically distinct from cPRC1 as they lack the CBX and PHC subunits, which are essential for targeting and mediating polycomb body formation (Figure 1.4). Instead, ncPRC1 complexes feature one of two paralogous proteins, RYBP or YAF2. Both RYBP and YAF2 compete with CBX proteins for the same binding pocket on RING1A/B, defining a mutually exclusive cPRC1 and ncPRC1 complexes (Wang et al., 2010). These ncPRC1 specific subunits increase the H2AK119ub catalytic activity of the PRC1 complex by facilitating conformational changes in the core Ring-PCGF heterodimer (Rose et al., 2016). Another defining feature of ncPRC1 complexes is that they can associate with any one of six PCGF proteins (PCGF1-6), with each of these subunits interacting with its own unique set of binding partners (Figure 1.4). For example, PCGF6-ncPRC1 complexes are composed of several accessory transcription factors (E2F6, MAX, MGA), chromatin modifiers (HDAC1/2) and the chromatin reader, CBX3 (Gao et al., 2012; Trimarchi et al., 2001). PCGF6-ncPRC1 contains its own unique set of target genes that are not bound by cPRC1, and is recruited to these sites by the transcription factors mentioned above, Max and Mga (Endoh et al., 2017). Loss of PCGF6 results in a depression of these unique genes, however, it does not lead to a derepression of *Hox* genes, consistent with reports that loss of ncPRC1 subunits during development do not lead to homeotic transformations (Gonzalez et al., 2008). The best characterised ncPRC1 complex is PCGF1-ncPRC1, which contains exclusive interactors such as, KDM2bB, BCOR, BCORL1, SKO1 and USP7 (Figure 1.4). Intriguingly, KDM2B can recruit PCGF1-ncPRC1 complexes to chromatin (Farcas et al., 2012; Wu et al., 2013). KDM2B achieves this targeting capacity by binding to unmethylated CpG islands via its CxxC motif, thereby recruiting ncPRC1 to these sites where it can catalyse H2AK119ub (Farcas et al., 2012; Wu et al., 2013). This leads to the possibility that ncPRC1 complexes are recruited to chromatin independently of PRC2, adding a further layer of complexity to the various mechanisms of PcG recruitment in mammals

(Figure 1.3). Supporting this, RYBP and H2AK119ub levels at PcG targets are unaffected by loss of EED and H3K27me3 (Tavares et al., 2012). Furthermore, artificial tethering of ncPRC1 to chromatin is sufficient to recruit PRC2 core components and leads to the accumulation of H3K27me3 (Blackledge et al., 2014). This ncPRC1 mediated mechanism to promote PRC2 chromatin occupancy and activity is a topic of debate in the Polycomb field. Recently, the JARID2 subunit of PRC2 was shown to be capable of reading the H2AK119ub mark through an N-terminal ubiquitin interaction domain (UIM) (Cooper et al., 2016), and that H2AK119ub containing nucleosomes are a much better substrate for PRC2 to catalyse H3K27me3 than unmodified nucleosomes (Kalb et al., 2014). Although this represents an elegant model for how ncPRC1 may precede PRC2 at PcG targets, it must be said that the mechanism of PRC2 recruitment to chromatin cannot be explained as simply as JARID2 "reading" pre-existing H2AK119ub. Supporting this, deletion of RYBP or YAF2 leads to a large reduction of H2AK119ub, but has little effect on H3K27me3 and no effect on PRC2 occupancy at these sites (Rose et al., 2016).

Loss of function studies in both *Drosophila* and mice have established that depletion of cPRC1 subunits (dPh, dPsc, mPHC1-3, mBmi1) leads to early embryonic lethality and derepression of Hox genes, a classical polycomb phenotype (Adler et al., 1991; Akasaka et al., 2001; Laugesen and Helin, 2014). Intriguingly, although ncPRC1 seems to share many target genes with cPRC1 and PRC2, it also has its own set of unique target genes (Endoh et al., 2017; Morey et al., 2013). This may explain why loss of ncPRC1 subunits such as RYBP also leads to early embryonic lethal phenotypes but, this phenotype is independent of *Hox* gene derepression (Gonzalez et al., 2008). This suggests that cPRC1 is critical for the spatio-temporal regulation of *Hox* genes during development, but ncPRC1 is not.

1.7 PRC2.1 versus PRC2.2

Similar to loss of function PRC1 studies, the knockout of core PRC2 components leads to early embryonic lethality (E7.5-8.5) (Laugesen and Helin, 2014). Hence,

the PRC2 trimeric core and its associated functions are very obviously critical for normal mammalian development. A unique difference between the respective complex architecture of PRC1 and PRC2 is that PRC2 does not exhibit as much diversity in its accessory subunits. However, recent studies have provided the first clear evidence that different combinations of PRC2 can have varying mechanistic and biological functions (Conway et al., 2018; Grijzenhout et al., 2016; Holoch and Margueron, 2017). The PRC2 complex can be divided into two distinct complexes, PRC2.1 and PRC2.2 (Figure 1.4). JARID2 and AEBP2 have been shown to determine the PRC2.2 subcomplex and are found to be mutually exclusive to PRC2.1 subunits, such as PCL proteins, EPOP and PALI1 (Alekseyenko et al., 2014; Conway et al., 2018; Hauri et al., 2016; Holoch and Margueron, 2017). Interestingly, all these auxiliary PRC2 subunits co-occupy the promotors of developmentally repressed genes together with core PRC2 members and the H3K27me3 mark. To date, no unique chromatin association pattern has been determined for either the PRC2.1 or PRC2.2 specific subcomplexes (Beringer et al., 2016; Brien et al., 2012; Grijzenhout et al., 2016). Numerous loss of function experiments in mouse embryonic stem cells (ESCs) have shown that these auxiliary subunits are essential for normal PRC2 function. For example, JARID2, a critical component of PRC2.2, has been shown to be vitally important for PRC2 function in ESCs. Loss of JARID2 leads to eviction of core PRC2 and depletion of H3K27me3 on chromatin, which is accompanied by a compromised ability for successful differentiation of ESCs (Landeira et al., 2010; Pasini et al., 2010). Intriguingly, it has recently been reported that another PRC2.2 component, AEBP2, actually antagonises PRC2 by inhibiting H3K27me3 catalysis, even though it can co-localise with, and occupy the same genomic regions as JARID2 (Grijzenhout et al., 2016). Supporting this, AEBP2$^{GT/GT}$ mice were shown to exhibit a trithorax phenotype *in vivo*, and to antagonise PRC2.1 function in ESCs (Conway et al., 2018; Grijzenhout et al., 2016). While the mechanism whereby AEBP2 can oppose normal PRC2 function remains unclear, it has been reported to compete with PRC2.1 components for the same binding pocket on SUZ12 (Chen et al., 2018). This antagonism possibly

allows for the assembly of distinct classes of PRC2 complexes with independent activities.

1.8 Polycomb-like Proteins

Phf1, *Mtf2* and *Phf19* are mammalian homologues of *Drosophila* Polycomb-like (*dPcl*) and hence, are also referred to as PCL1, PCL2 and PCL3 respectively (Margueron and Reinberg, 2011). Initial studies revealed that mutations in the *dPcl* gene resulted in ectopic homeotic gene expression, phenocopying deletion of Polycomb (*Pc*) (Lonie et al., 1994). The three paralogous mammalian PCL genes arose through two gene duplication events during mammalian evolution (Makino and McLysaght, 2010). An open question surrounding the study of these proteins, is why three genes with apparent functional redundancy exist in mammals. All three mammalian PCL proteins contain a number of putative chromatin and DNA binding domains, including an N-terminal Tudor domain, two PHD domains and a C-terminal winged helix domain (WH) (Figure 1.6). It is due to this unique domain architecture of PCL proteins, that a great deal of research has focused on their potential roles in mediating the recruitment and targeting of PRC2 to specific genomic loci.

The roles of Polycomb-like proteins have been primarily studied in ESCs, however, PHF1 is not typically expressed in pluripotent stem cells and hence has been more extensively studied in both primary and transformed human cell lines. In HeLa cells, PHF1 has been shown to co-sediment with PRC2 and occupy the same PcG target genes as EZH2 (Sarma et al., 2008). In the same study, PHF1 was reported to stimulate PRC2 catalytic activity *in vitro*, as well being required to efficiently convert H3K27me2 to H3K27me3 *in vivo* (Sarma et al., 2008). Adding to this, an independent study recently showed that as well as H3K27me3, PHF1 is also capable of generating H3K27me2 *in vitro* (Choi et al., 2017). However, the effects of Polycomb-like proteins on H3K27me2 *in vivo* are still unknown. Similar to PHF1, MTF2 and PHF19 have been shown to promote recruitment of PRC2 (Brien et al., 2015), with their loss resulting in reduced chromatin enrichment of H3K27me3 (Ballare et al., 2012; Brien et al., 2012;

Casanova et al., 2011; Li et al., 2017). Overall, it seems as though PCL proteins act as transcriptional repressors by positively regulating the enzymatic activity of PRC2 . Divergent roles for the three Polycomb-like proteins have also been characterised in primary human fibroblasts. Brien *et al,* showed that while *MTF2* and *PHF19* are E2F regulated genes highly expressed in cycling cells (similar to *EZH2* and *SUZ12*), *PHF1* is a p53 target gene and is predominantly expressed in non-dividing, quiescent cells (Brien et al., 2015). While ectopic expression of all three PCL proteins recruits PRC2 to repress *INK4A* gene (a classical PcG target gene in fibroblasts), only MTF2 and PHF19 confer an independent INK4A growth advantage. Interestingly, PHF1 was shown to have gained a PRC2- and chromatin-independent function to bind to and stabilise p53, negatively regulating cell proliferation and promoting cellular quiescence (Brien et al., 2015). This function of PHF1 is achieved through two unique Serine residues in the N-terminal PHD domain, that are not present in MTF2 and PHF19. These unique serine residues in PHF1 were acquired during recent vertebrate evolution and will be discussed at length in Chapter 3.

The molecular functions of MTF2 and PHF19 have primarily been studied in ESCs with research focusing, in particular on the mechanisms of PRC2 recruitment to target loci and their influences on H3K27me3 patterns. As mentioned previously, PCL proteins exhibit a unique domain architecture with several embedded chromatin and DNA binding moieties. These domains, namely a Tudor, two PHD domains and a winged helix domain, are key to the mechanisms whereby PCL proteins facilitate PRC2 function, and likely interact with chromatin through independent mechanisms. A study reporting the MTF2 PHD2 domain to be required for PRC2 recruitment to the inactive X-chromosome in undifferentiated and differentiating ESCs was the first to suggest a role for a PCL domain in PRC2 genomic targeting (Casanova et al., 2011). Furthermore, the Tudor domain of all three PCL proteins have been shown to be critical in promoting PRC2 function *in vivo* (Ballare et al., 2012; Brien et al., 2012; Musselman et al., 2012a). Interestingly, this Tudor domain has been shown to read the H3K36me3 PTM, a mark of actively transcribed genes (Ballare et al.,

2012; Brien et al., 2012; Musselman et al., 2012a). PHF19 was also reported to interact with N066, a H3K36me3 demethylase, which lead to the proposal that it recruits PRC2 to active regions of the genome to initiate their repression (Brien et al., 2012). Intriguingly, the Tudor domain of PHF1 and PHF19 has also been reported to recognise H3K36me2 (Brien et al., 2012; Musselman et al., 2012a). Relatively little is known about the roles and genome wide profiles of this particular histone PTM, however it has recently been reported that the H3K36me2 methyltansferase, NSD1, directly interacts with PRC2 (Streubel et al., 2018). NSD1 mediated H3K36me2 co-locates with H3K27me2 at broad intergenic regions of the genome, thereby facilitating demarcation of H3K27me2 and H3K27me3 domains in ESCs. PRC2 catalyses all H3K27me2, but is not found to be enriched at the intergenic sites that harbour this PTM (Ferrari et al., 2014; Hojfeldt et al., 2018). The ability of PCL proteins to recognise H3K36me2 may speculatively point to a role in the promotion and/or maintenance of H3K27me2 and the demarcation of H3K27me2/3 regions genome wide, which is absolutely essential to support proper gene expression profiles. MTF2 and PHF19 both have both been reported to be important for differentiation of ESCs to embryoid bodies (EB bodies) as knockdown of either gene perturbed normal differentiation. This differentiation phenotype as well as the widespread loss of H3K27me3 associated with loss of PCL in ESCs are only rescued by proteins carrying intact, functional Tudor domains (Brien et al., 2012; Cai et al., 2013).

PCL proteins also contain two other highly conserved PHD domains with putative chromatin binding abilities, whose functions have yet to be established. However, recently a highly conserved extended homologous region C-terminal to the second PHD domain has been reported to exhibit a winged-helix like transcription factor structure, and to exhibit binding affinity for DNA (Choi et al., 2017; Li et al., 2017; Perino et al., 2018) (Figure 1.6). One particular recent study reported that PHF1 can prolong the residency time of PRC2 on DNA and chromatin (Choi et al., 2017). The DNA binding ability of PHF1 accounts for this extended PRC2 residence as an elegantly selected combination of mutations in the winged helix (WH) domain abolished this binding. This increased tethering of PRC2 to DNA

and chromatin also makes PHF1-PRC2 a more effective methyltransferase than PRC2 on its own (Choi et al., 2017). This well designed study establishes that the interaction of PCL proteins with DNA provide the predominant tethering affinity for PRC2 with chromatin and suggest a mechanism for how PCL proteins boost the methyltransferase activity of PRC2. There have been contrasting reports regarding the sequence specificity of this PCL-WH-DNA interaction, with two studies suggesting that PCL proteins bind to a non-methlyated CpG sequence motif. Studies by Li *et al,* and Perino *et al,* have reported that the PHD2-WH of PCL proteins preferentially bind DNA sequences centred on a CpG dinucleotide (Li et al., 2017; Perino et al., 2018), however contrasting results have been published suggesting that PCL binds DNA in a sequence non-specific fashion (Choi et al., 2017). Indeed, Perino *et al* suggested that the PCL-WH domains have a preference for particular DNA structural and helical properties (Perino et al., 2018). Given that the PCL-WH-DNA interaction is relatively weak (k_d = ~30μM), and that DNA sequence and its associated helical shape are difficult to disentangle, the binding is unlikely to be affected by specific DNA structural conformations. Furthermore, this phenomenon remains to be independently validated. In either case, what is clear is that this highly conserved WH domain increases tethering of PRC2 to chromatin allowing for a longer residency time and increased catalytic activity. This potentially provides a targeting mechanism to enhance random genomic sampling by core PRC2 components.

The roles of Polycomb-like genes during mammalian development have been highlighted by two independent *in vivo* knockout studies (Li et al., 2011; Rothberg et al., 2018). These independent studies report contrasting *Mtf2 in vivo* knockout phenotypes. Li *et al* suggest that mice carrying an *Mtf2* gene trap cassette that abolishes all *Mtf2* isoforms do not present skeletal homeotic transformations, are viable and produce viable offspring (Li et al., 2011). However, a separate study by, Rothberg *et al* describe an MTF2 knockout mouse that results in embryonic lethality at E15.5 due to severe anaemia (Rothberg et al., 2018). It is difficult to reconcile these contrasting studies, but perhaps they could be explained by the

A

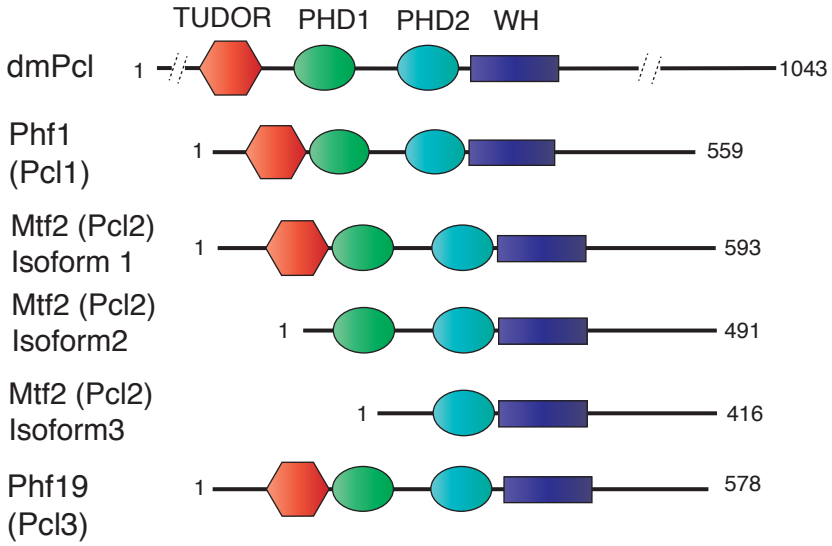


Figure 1.6 Domain architecture of the Polycomb-like proteins (Pcl).
(A) Illustration of the domain architecture of Pcl proteins. Three Pcl paralogues exist in mammals, that arose from a single ancestral gene, represented bt Drosophila Pcl (dmPcl). Pcl proteins uniquely contain a Tudor domain, two PHD domains, and a recently characterised winged helix (WH) DNA binding domain.

differing knockout strategies adopted in the studies. While Li *et al* employ a genetrap strategy inserted at exon 4 to abolish all MTF2 activity *in vivo* (Li et al., 2011), Rothberg *et al* generated MTF2 null mice using LoxP gene targeted ESCs (Rothberg et al., 2018). There could also be some system specific redundancy exhibited between *Phf1* and *Phf19.* In either case, although these *in vivo* studies present differing results, they both highlight the need for further comprehensive knockout studies on all three PCL proteins and re-affirm the biological importance of Polycomb-like proteins in normal PcG function during development.

It is well established that PCL proteins contribute to normal PRC2 biology as individual loss of each of these proteins by shRNA mediated knockdown or knockout leads to a reduction in H3K27me3 a reduction of core PRC2 components from chromatin (Ballare et al., 2012; Brien et al., 2012; Li et al., 2017; Perino et al., 2018). However, the question of functional redundancy between the three PCL proteins still remains in all of the studies performed to date. Even though the PCL genes are differentially expressed in various cells types, low level expression of the unperturbed PCL genes may partially compensate. Further genome-wide *in vivo* studies will need to be performed in order to establish the relevant contributions of each PCL proteins and their individual domains to overall PRC2 function and biology.

1.9 PRC2 and H3K27 methylations in cancer

Over the past decade, the emergence of next generation sequencing technologies have revealed the genomic landscapes of many forms of human cancers (Cancer Genome Atlas, 2015). This led to the remarkable realisation that a large proportion of cancers have mutations in genes encoding chromatin regulators (Dawson and Kouzarides, 2012). A seminal comprehensive analysis by Vogelstein and colleagues in 2013 examined the most common types of mutations and biological processes most frequently associated with cancer “driver” genes (Vogelstein et al., 2013). This highlighted that out of 140 cancer “driver” genes, 30 of these genes encode chromatin regulatory proteins. A prominent chromatin regulator whose functions is perturbed in cancer is the

SNF5 (*SMARCB1*) tumour suppressor gene. *SNF5* is found be recurrently biallelically inactivated in an aggressive paediatric malignancy known as malignant rhabdoid tumours (MRTs). The protein product of *SNF5* is a component of the BAF complex, a large multimeric nucleosome remodelling complex. The BAF complex is known to be associated with active enhancers and H3K27Ac, as such it opposes and has an antagonistic relationship with Polycomb function (Wilson et al., 2010). In addition to SNF5, several studies have reported frequent inactivation of other BAF complex members in other tumour types including breast, lung and colon carcinomas (Kadoch et al., 2013).

Multiple cancer genome sequencing studies have revealed that PRC2 and associated H3K27 methylations are frequently disrupted in human cancers (Helin and Dhanak, 2013) (Table 1.1). *EZH2* was reported in 2003 to be regulated by E2F downstream of the pRB pathway, and overexpressed in many human tumours (Bracken et al., 2003). Since then, recurrent heterozygous point mutations in the EZH2 catalytic SET domain have been reported in several types of human cancers, including B-cell lymphoma and follicular lymphoma (Bodor et al., 2013; Huether et al., 2014; McCabe et al., 2012a; Morin et al., 2010) (Table 1.1). These specific mutations have been described as “change-of-function” mutations, which confer an enhanced ability of PRC2 to convert H3K27me2 to H3K27me3 (Sneeringer et al., 2010). As a consequence, cancers harbouring these mutations have aberrantly high levels of H3K27me3 and reduced H3K27me2. How exactly this change of EZH2 function contributes to cancer progression is not fully understood but it has been reported that although the abundancy of H3K27me3 increases, it is also widely redistributed across the genome resulting in a massively reorganised transcriptional and chromatin landscape (Souroullas et al., 2016). The discovery that *EZH2*, *EED* and *SUZ12* are deleted and contain inactivating mutations in leukaemia’s and malignant peripheral nerve sheath tumours (MPNST) was initially surprising, as PRC2 function was considered to be oncogenic (Ernst et al., 2010; Lee et al., 2014; Nikoloski et al., 2010; Zhang et al., 2014a). Tumours with these particular “loss

Table 1.1 Mutations of PRC2 members and histone H3 genes in cancer.

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	pTyr646X	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%) MPNST (4–25%)	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
Other PRC2 associated genetic lesions			
EZH1	pGln571Arg	Autonomous Thyroid Adneoma (ATA) (27%)	Increased H3K27me3 Decreased H3K27me2
PHF1	Gene fusion with JAZF1 [t(6:7)(p21;p22)]	Endometrial stromal sarcomas (ESS)	Unknown
SUZ12	Gene fusion with JAZF1 [t(7:17)(p15;q21)]	Endometrial stromal sarcomas (ESS)	Decreased 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

of function" PRC2 genetic lesions exhibit reduced levels of both H3K27me2 and H3K27me3 (Conway et al., 2015) (Table 1.1).

Due to the prevalence of PRC2 dysregulation in cancer, several small molecule SAM competitive catalytic inhibitors have been developed specifically targeting EZH1/2 methyltransferase activity (EZH2i). EZH2i drugs have been shown to have efficacy *in vitro* reducing proliferation rates in B-cell lymphoma cancer cell lines with EZH2 SET domain mutations (Knutson et al., 2012; McCabe et al., 2012b; Qi et al., 2012). These targeted inhibitors showed initial promise and as such are the subject of major clinical trials in lymphomas and MRTs, which have an increased dependency on PRC2 in the absence of SNF5 and BAF complex perturbation (Brien et al., 2016). Phase 1 clinical trials have yielded great initial promise in the treatment of diffuse large B-cell lymphoma (DLBCL), with 38% (8/21) of patients showing an objective response (Italiano et al., 2018). However, it has been in follicular lymphoma, previously thought to be "incurable", where targeted EZH2i has shown the greatest efficacy (Makita and Tobinai, 2018). Preliminary results of a phase two clinical trial of Tazemetostat in follicular lymphoma have reported that 92% (12/13) of patients with EZH2 mutations had an objective response to treatment, compared with 26% (14/54) in those who did not carry an EZH2 mutation (Makita and Tobinai, 2018). These very promising preliminary results suggest that targeted therapies against "change-of-function" EZH2 mutations may have real benefit in the treatment of previously incurable cancers. However, a potential caveat to consider with these targeted therapies, is that EZH2 is deleted in T-cell leukemias (Ntziachristos et al., 2012) and so care will have to be taken to avoid secondary cancer formation.

Another example of PRC2 dysregulation in cancer are the recurrent point mutations at lysine 27 of Histone H3 (to methionine – H3K27M) in pediatric diffuse intrinsic pontine gliomas (DIPGs) (Sturm et al., 2012; Wu et al., 2012). Multiple copies of the Histone H3 gene exist in humans, however, the highly recurrent heterozygous, dominant negative H3K27M mutations occur in only two of fifteen H3 genes, which is particularly remarkable considering these gliomas

exhibit global reductions in H3K27me2 and H3K27me3 (Lewis et al., 2013). While the mechanism of how these mutations affect PRC2 activity remains somewhat elusive, initial studies suggested that H3K27M acts as a dominant negative to sequester and block the activity of PRC2 (Lewis et al., 2013). This mechanism was backed up by structural studies which report that the H3K27M peptide occupies the active site of EZH2 and exhibits tighter PRC2 binding than a canonical H3K27 peptide (Justin et al., 2016). However, this model has recently been challenged by reports that H3K27M point mutations excludes PRC2 from chromatin (Piunti et al., 2017). While, the mechanism of exactly how these mutations contribute to cancer remains debatable, a recent comprehensive study showed that H3K27M-expressing gliomas require PRC2 for proliferation, and that EZH2i abolishes tumour growth through a $p16^{INK4A}$ dependent mechanism (Mohammad et al., 2017). This suggests that inhibition of PRC2 is a potential therapeutic strategy for treatment of pediatric DIPGs (Mohammad et al., 2017).

Despite the promising preliminary results from EZH2i clinical trials, two recent studies have found that B-cell lymphoma cell lines with *EZH2* "change-of-function" mutations grown under the selective pressure of EZH2i develop specific drug resistant point mutations (Baker et al., 2015; Gibaja et al., 2016). Interestingly, these mutations occur on the wild-type allele and have been reported to alter affinity of the drug for the active site (Brooun et al., 2016). Therefore, as with many targeted therapies, resistance mutations similar to this are a major problem to be overcome in the treatment of cancers (Holohan et al., 2013). This is elegantly illustrated by the fact that drugs targeting the activity *BCR-ABL* fusion oncogene in Chronic Myeloid Leukemia (CML), are now in their second and third generation of targeted therapies, due to the fact that single amino acid changes occur in the kinase domain of resistant cancer clones (Holohan et al., 2013; Shah et al., 2004). Drugs targeting the H3K27me3 binding function of EED have recently been developed and represent a very promising pathway for targeting PRC2 outside of the catalytic activity of EZH2 (He et al., 2017; Qi et al., 2017). The mechanism of action of these allosteric inhibitors is not yet fully understood and will be explored in detail in Chapter 4.

1.10 New ways to target the PRC2 complex in cancer

Given the reports of EZH2i resistance mutations *in vitro,* it is anticipated there will be similar consequences for patients with diffuse large B-cell lymphoma, follicular lymphoma and MRTs. Therefore, there is a pressing need to find alternative strategies to inhibit EZH2 and PRC2 in various cancer types. As mentioned above, PRC2 can associate with a number of different accessory proteins in a substoichiometric manner, such as JARID2, AEBP2, EPOP, PALI1/2 and PCL1-3. These proteins are known to modulate PRC2 function and could be exploited to find novel ways of targeting PRC2 activities. Recently it has been reported that specific combinations of these accessory components compete for binding pockets on core subunit SUZ12 (Chen et al., 2018; Youmans et al., 2018). Small molecules targeting these binding pockets on SUZ12 may represent a way of modulating PRC2 subcomplex association in cells to favour a less or more active PRC2 depending on the underlying genetic background of the cancer. I believe that Polycomb-like1-3 represent and excellent avenue to explore alternative PRC2 targeted therapies. Therefore, in addition to their roles in cell fate decisions during differentiation and development, the comprehensive study of Polycom-like proteins may hold promise for future targeted cancer therapies.

1.11 Aims of thesis

The main aims of this thesis were to further study sub-functionalisation of Polycomb group proteins in vertebrates using both human and mouse cell systems. I aimed to do this by examining the various roles of the key PRC2 accessory components, Polycomb-like proteins, in both differentiated human cells and pluripotent mouse ESCs. Polycomb-like proteins are essential conserved regulators of PRC2 activities and represent a unique avenue to study how the Polycomb system has evolved and acquired additional functions from *Drosophila* to higher eukaryotes.

Firstly, I explore the biochemical functions of PCL1 in the maintenance of cellular quiescence through a novel PCL1-p53 regulatory axis. Secondly, through the analysis of all three Polycomb-like proteins in cycling and quiescent human cells, I have defined a novel catalytically inactive form of PRC2 lacking SUZ12, that exists exclusively in quiescent cells. Thirdly, by analysing the effects of genetic knockout of all three Polycomb-like genes in ESCs, I characterise the roles of distinct classes of PRC2 subtype assemblies, PRC2.1 and PRC2.2.

Chapter 2

Materials and Methods

2.1 Cell Culture

2.1.1 Culture conditions for various cell lines

IMR90 primary fibroblasts were cultured in DMEM supplemented with 10% FBS (Gibco), 100 U/mL penicillin (Gibco) and 100 U/mL streptomycin (Gibco). For serum starvation experiments, serum was removed from fibroblasts at ~75% confluence for 120h. Media was changed at 24h intervals until 120h, at which point cells were re-stimulated to enter the cell cycle by addition of media containing 10% or 20% (v/v) FBS for 24h.

Insect cells (Sf9 and High Five) were cultured in Hinks TNMFH (sigma) supplemented with 10% FBS (Gibco) and 0.5% Gentamycin (Sigma). p53-GOF breast cancer cell lines (BT-549 and MDA-MB-469) were cultured in DMEM supplemented with 10% FBS (Gibco), 100 U/mL penicillin (Gibco) and 100 U/mL streptomycin (Gibco) and 1:100 non-essential amino acids (Gibco).

Embryonic stem cells were grown on gelatinized culture dishes in GMEM (Sigma) supplemented with 10% ES qualified FBS (Millipore), 100 U/mL penicillin, 100 U/mL streptomycin (Gibco), 50 μM β-mercaptoethanol (sigma), 1:100 GlutaMax, 1:100 non-essential amino acids (Gibco), 1mM sodium pyruvate (Gibco) and 1:500 homemade leukemia inhibitory factor (LIF). For Embryoid body differentiation experiments, $2x10^6$ ESCs were washed three times in DPBS (Lonza) and seeded onto non-adherent petri dishes in ESC media without LIF. The media wash changed every two days (-LIF). For LIF/2i growth conditions, ESCs were cultured in 1:1 Neurobasal:DMEM/12 media, 10% ES qualified FBS, 100 U/mL penicillin, 100 U/mL streptomycin (Gibco), N-2 supplement (0.5%)(Gibco 17502-048), B27 (1%)(Gibco 17504-044) and 1:500 homemade leukemia inhibitory factor (LIF). The GSK inhibitor CHIRON99021 (Millipore) and MEK inhibitor (PD0325901) were also added at final concentrations of 3 μM and 1 μM respectively. Knockout of *Pcl1/3* genomic loci was induced by addition of 0.5 μM 4-OHT. After 72 hours cells were harvested and PCRs on genomic DNA performed to confirm ablation of targeted genomic regions.

2.1.2 3T3 growth assays

3T3 growth assays were conducted as follows. $1x10^6$ cells were plated on 150mm plates, 3 days later, the total number of cells was counted and $1x10^6$ cells were plated again. The cumulative increase in cell number was calculated according to the formula 'Log(Nf/Ni)/Log2' where Ni and Nf are the initial and final numbers of cells plated and counted after 3 days, respectively.

2.1.3 RNA interference

Cells were seeded 24h prior to transfection, and transfected at 30–50% confluence with 20μM siRNA–Ctrl (SIC001), siRNA–TP53#1 (custom, target sequence –TGTTCCGAGAGCTGAATGA), siRNA–TP53#2 (custom, target sequence –GTGCAGCTGTGGGTTGATT), siRNA–PCL1#1 (custom, target sequence –CACACACCGGCACTTTCATAC), siRNA–PCL1#2 (custom, target sequence – GCAACCGACAGCAGAGTTA). Cells were transfected using Lipofectamine RNAi MAX (Invitrogen) in accordance with the manufacturers in instructions. For experiments involving serum starvation, serum was removed from transfected cells 16h post transfection.

2.2 Recombinant Proteins

2.2.1 Cloning and plasmid generation

Full length ORFs of *PCL1, EZH1, EED and SUZ12*, were PCR amplified (primers available upon request) from cDNA generated from either HMECs or HEK293T human cell lines. All PCR products were inserted into the pCR8/GW/TOPO Gateway cloning entry vector (Invitrogen). All ORFs were subsequently sub-cloned into Gateway compatible expression vectors by recombination using LR-Clonase enzyme (Invitrogen).

2.2.2 Purification of GST-fusion proteins

PCL1/2/3-PHD1 wild-type and mutant fragments were cloned into the pGEX6P1 expression vector. Sequence verified clones were transformed into protease deficient *E. coli* strain BL21–DE3. Colonies were picked and grown as started cultures overnight at 37°C in 10mls of TP media. The next day the 10ml started

cultures were inoculated into 500mls of pre-warmed TP media. These cultures were grown for 3 hours at 37°C until OD600 was between 0.4- 0.6. Protein expression was induced with 0.5 mM IPTG and cultures were grown overnight at 20°C. Cell pellets were harvested and lysed in PBS containing 0.25% Triton X-100 and protease inhibitors. Lysates were precleared by centrifugation at 20000rpm for 45mins, and pre-cleared lysates were incubated with Glutathione-agarose beads (Pierce) overnight at 4°C. Beads were washed extensively in wash buffer (PBS, 350 mM NaCl, 0.25% Triton X-100) and bound GST-fusion proteins were eluted using the same buffer following the addition of 20 mM Glutathione (Sigma).

2.2.3 Baculovirus production and purification of G0-PRC2, PCL1-PRC2 and EED-PRC2 recombinant complexes

Recombinant baculoviruses for Flag/His-PCL1 and Flag/His-EED (N-terminal tag) as well as untagged EZH1, SUZ12, and EED were generated in *Spodoptera frugiperda,* Sf9 cells. Briefly, a baculovirus transfer vector (pVL1392) containing either PALI1 or LCOR coding ORFs and a linearised baculovirus DNA construct (Allele Biotech #ABP-BVD-10001) were co-transfected into Sf9 cells. The supernatants were collected from transfected cells 10 days after transfection (passage 1 [P1]) and passaged to P3, which were used for subsequent recombinant protein purifications. Defined combinations of Flag/His-PCL1 or Flag/His-EED and untagged PRC2 components were then expressed in *Trichoplusia ni*, High Five, cells. The cells were incubated at 28°C and harvested 44-48hr post infection, washed twice in PBS and lysed in BPL2B buffer (20mM Tris, 500 mM NaCl, 20% Glycerol, 4 mM MgCl2, 3 mM β-mercaptoethanol, 0.05% NP40, 2 µg/mL Aprotonin, 1 µg/mL Leupeptin, 1 mM PMSF). Lysates were sonicated and cleared by centrifugation for 30min at 20,000g. The supernatant was loaded onto Ni-NTA His resin (Novagen #70666) and incubated for 3hrs rotating at 4°C. Ni-NTA His resin was then washed a number of times in BPL2B buffer to remove non-specific interactions. The His-tagged protein complexes were subsequently eluted in Elution buffer (350 mM Imidazole, 50 mM Tris, 5 mM MgCl2, 150 mM NaCl, 0.05% NP40). Eluted fractions were

analysed by SDS-PAGE and were subsequently snap frozen in liquid nitrogen (+20% glycerol).

2.3 RNA and Protein Analysis

2.3.1 RNA preparation and RT-PCR 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* was used as a normaliser. The data and graphs presented here a representative sample of three biological replicates. Error bars on graphs indicate standard deviation of individual triplicate qPCR data.

2.3.2 mRNA expression analysis during hematopoiesis

Gene expression analysis of PRC2 components in hematopoietic lineages was performed essentially as described previously. (Seita et al., 2012; Xu et al., 2015). https://gexc.riken.jp/

2.3.3 RNA-Seq and bioinformatic analysis

RNA-Seq libraries were prepared using the NEBNext Ultra RNA library prep kit for Illumina (E7770L) according to manufacturer's instructions. Prior to starting library prep RNA concentrations were measured on a Qubit 3.0 and RIN scores calculated using an RNA ScreenTape (Agilent) (RIN >8.5). A total 1μg of high quality RNA was used for library preparation. Following adaptor ligation, DNA was PCR amplified for 8 cycles. DNA purification was then performed using NEBNext Sample Purification Beads (E7767S). The quality of cDNA libraries was analysed on a High Sensitivity D1000 Screen Tape (Agilent). The resulting libraries were then used for cluster generation and sequencing using an Illumina NextSeq 500 (ID: NB501524), with 75bp read length. Sequencing reads were quantified by pseudo-aligning to the mouse reference transcriptome

(GRCm38/mm10) using kallisto (Bray et al., 2016). Sequence reads were aggregated into a count for each gene using tximport (Soneson et al., 2015). Differentially expressed genes were identified using DESeq2 (Love et al., 2014) . Sequence tracks were generated by aligning reads to the mm10 reference genome using HISAT v2.0.5 (Pertea et al., 2016). Resulting bedGraph files were converted to bigwig format and scaled using hits per billion with the bamTobw.sh utility (Zhu et al., 2013) for visualization on the UCSC Genome Browser. AEBP2 positive genes were determined in the same manner as for the ChIP-seq analysis above. To visualize the relationship between changes in ChIP enrichment of H3K27me3, JARID2, AEBP2 and EPOP, in ESCs, and gene expression, in EBs at Day 4 and 8, in PCL WT and KO cells, the log2 fold change of each ChIP was computed for the promoter regions (TSS ± 2.5kb) and plotted against the log2 fold change in gene expression. All bioinformatic analyses of RNA-Seq datasets was performed by Dr. Darren Fitzpatrick of the Bracken Lab.

2.3.4 Preparation of whole cell protein lysates and Western blotting

Cells were scraped down to collect them, washed three times in PBS and resuspended in ice cold High Salt buffer (50 mM Tris-HCl, pH 7.2, 300 mM NaCl, 0.5% (v/v) NP-40, 1 mM EDTA pH7.4, 2 µg/mL Aprotonin, 1 µg/mL Leupeptin, 1 mM 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 14,000 RPM at 4°C for 25 mins. Protein lysates were then separated on SDS-PAGE gels and transferred to nitrocellulose membranes. Membranes were subsequently probed using the relevant primary (overnight at 4°C) and secondary (1 hr at room temperature) antibodies. Relative protein levels determined by chemiluminescence.

2.3.5 Antibodies

Antibody Target	Source	Western Blot	IP	ChIP
PCL1	Abgent (AT3294a)	1:500	5 ug	2 ug
PCL1	ProteinTech (15663-1-AP)	1:500		
PCL2	ProteinTech (16208-1AP)	1:750	5 ug	5 ug
PCL3	Brien *et al.,* 2015	1:500	5 ug	
EZH1	Merck (ABE2821)	1:1000		
EZH2	BD43 (Pasini *et al.,* 2004)	1:10		
EZH2	AC22 (Bracken *et al*., 2006)		5 ug	2 ug
SUZ12	Cell Siganlling (3737)	1:2000		1 ug
EED	AA19 (Bracken *et al*., 2003)	1:10		
BMI1	DC9 (Bracken *et al*., 2007)	1:20		
RYBP	Sigma-Aldrich (PRS2227)	1:1000		
H3K27me1	Active Motif (61015)	1:2000		5 ug
H3K27me2	Cell Signalling (9728S)	1:2000		5 ug
H3K27me3	Active Motif (61017)	1:2500		
H3K27me3	Cell Signalling (custom)			5 ug
H2AK119Ub	Cell Signalling (8240)	1:2000		5 ug
H3K27Ac	Merck (07-360)	1:1000		
H3K36me2	Cell Signalling (2901)	1:5000		
H3K36me3	Cell Signalling (4909)	1:1000		
H3K4me3	Abcam (ab8580)	1:1000		
H4K20me3	Abcam (ab9053)	1:1000		
Histone H3	Abcam (ab1791)	1:50000		1ug
p53	D01	1:50		
GST	SantaCruz (B-14, sc-138)	1:500		
CCNA2	BD Biosciences (611268)	1:1000		
p16	JC8	1:8		
GAPDH	SantaCruz (sc-25778)	1:1000		
PALI1	Merck (ABE1367)	1:500		
JARID2	Cell Signalling (13594)	1:500		5 ug
AEBP2	Cell Signalling (14129)	1:500		1 ug
EPOP	Beringer *et al*., 2017	1:500		2ug
Mouse-IgG	Millipore (12-371)		5 ug	2 ug
HA	Cell Signalling (3724)	1:1000		

2.3.6 Cellular Fractionations

Human diploid fibroblasts or ESCs were lysed in 400 μL pre-extraction buffer (20 mM HEPES pH 7.2, 0.5% Triton X-100, 50mM NaCl, 3 mM MgCl2, 300 mM Sucrose, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF). Incubate on ice

for 30 minutes. 200 μL of suspension was removed and labelled “Total extract”, 200 µL of 2X SDS-PAGE sample buffer was added to the “Total extract”. The remaining lysate was clarified at 20,817 g in a 4°C centrifuge for 10 minutes. Supernatant was kept and labelled “Soluble”. 200 µL of 2X SDS-PAGE sample buffer was added to this “Soluble” fraction. The insoluble pellet was washed once in 1mL of pre-extraction buffer before resuspension in 200 µL of pre-extraction buffer and 200 μL of 2X SDS-PAGE sample buffer. All samples were boiled at 99°C for 5 minutes before sonicating 3 times for 10 seconds at 60% amplitude.

2.4 Immunoprecipitations

2.4.1 Endogenous Co-Immunoprecipitations

Human diploid fibroblasts or Embryonic stem cells were resuspended in Buffer C (20 mM HEPES pH 7.9, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 20% glycerol, 420 mM NaCl, 2 µg/mL Aprotonin, 1 µg/mL Leupeptin, 1 mM 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 (20 mM HEPES pH 7.9, 0.2 mM EDTA, 1.5 mM MgCl2, 20% glycerol, 125 mM KCl). Lysates were again clarified by centrifugation at 20,817g at 4°C for 20 min. 5µg antibody was coupled to 20µL packed Protein A beads (Sigma) by incubation in 1 mL 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 1 mL 0.2 M Sodium Borate pH 9.0. Antibodies were then crosslinked to beads by incubation in 1 mL 0.2 M 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 0.2 M 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 (0.1 mg/mL Insulin (Sigma), 0.2 mg/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 250 U/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 100 μL 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.

2.4.2 Endogenous immunoprecipitations coupled with Mass Spec

Endogenous EZH2 IPs were performed as described above. After final wash step beads were flash frozen until ready to complete mass spectrometry analysis. Label free quantification (LFQ) LC-MS/MS analysis was performed in collaboration with the Vermeulen lab, as described previously (Kloet et al., 2016). Mass Spectrometry analysis was performed in collaboration with the lab of Michiel Vermeulen (Radboud Institute for Molecular Life Sciences).

2.4.3 Chromatin Immunoprecipitations (ChIP)

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 6 mL of SDS-Lysis buffer (100 mM NaCl, 50 mM Tris pH8.1, 5 mM EDTA pH 8.0, 0.02% NaN_3, 0.5% SDS, 2 µg/mL Aprotonin, 1 µg/mL Leupeptin, 1 mM 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 [100 mM Tris pH 8.6, 100 mM NaCl, 5 mM EDTA pH 8.0, 0.02% NaN_3, 5% Triton X-100, 2 µg/mL Aprotonin, 1 µg/mL Leupeptin, 1 mM PMSF]). Chromatin was sheared to approximately 200bp-800bp 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 (10 mM Tris pH 8.1, 1 mM EDTA pH 8.0) containing 0.5 mg/mL BSA and 0.2 mg/mL Herring Sperm DNA. 10-100 μg (DNA) of chromatin was incubated overnight with antibody while rotating at 4°C. Following clarification, the chromatin was incubated for 3 hours with 50 μL of blocked protein A beads.

After incubation the beads were washed three times in Mixed Micelle Buffer (150 mM NaCl, 20 mM Tris pH 8.1, 5 mM EDTA pH 8.0, 5.2% Sucrose, 0.02% NaN_3, 1% Triton X-100, 0.2% SDS), twice with Buffer 500 (0.1% Sodium Deoxycholate, 1 mM EDTA pH 8.0, 50 mM HEPES pH 7.5, 1% Triton X-100, 0.02% NaN_3), twice with LiCl detergent wash (0.5% Sodium Deoxycholate, 1 mM EDTA pH 8.0, 250 mM LiCl, 0.5% NP-40, 10 mM Tris pH 8.0, 0.02% NaN_3) and finally one wash with TE. Immunoprecipitated material was eluted from the beads with Elution buffer (0.1 M $NaHCO_3$, 1% SDS) while shaking for 1 hour at 65°C. The supernatant 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 by, Phenol Chloroform clean up and Ethanol precipitation. ChIP enrichment was analysed by qPCR using the SYBR Green I detection chemistry (M3003E NEB) on an Applied Biosystems Quant Studio 3 platform. The data and ChIP enrichments presented here a representative sample of three biological replicates. Error bars on graphs indicate standard deviation of individual triplicate qPCR data.

2.4.4 ChIP-Rx and library preparation

I also performed quantitative chromatin immunoprecipitation relative to a reference exogenous genome (ChIP-Rx) coupled with massively parallel DNA sequencing for the genome-wide mapping of histone modifications and PRC2 components, as described previously (Orlando et al., 2014). For histone modification ChIPs, a total of 1.67% *Drosophila* chromatin was added to each lysate. Similarly, for PRC2 ChIPs, a total of 10% human chromatin (NT2) was added to each ESC lysate. Spike in chromatin was added at the beginning of the workflow and once exogenous and ESC chromatin were combined, the sample was treated as a single ChIP-Seq experiment until completion of DNA sequencing. Following the ChIP experiment, the precipitated DNA was quantified using the Qubit dsDNA High Sensitivity Assay Kit (ThermoFisher Q32854). A Total of 2-10 ng of DNA from each ChIP-Rx experiment was used for library preparation using the NEBNext Ultra II DNA Library Kit for Illumina (E7645) and

NEBNext Multiplex Oligos for Illumina (Set#1, NEB #7335). Following adaptor ligation, DNA was PCR amplified for 5-9 cycles, depending on amount of input DNA. DNA purification was then performed using NEBNext Sample Purification Beads (E7767S). The quality of DNA libraries was analysed on a High Senstivity D1000 Screen Tape (Agilent). The resulting libraries were then used for cluster generation and sequencing using an Illumina NextSeq 500 (ID: NB501524), with 75bp read length. Sequencing reads were aligned to the mouse reference genome (mm10) using Bowtie v2.1.0 (Langmead and Salzberg, 2012). Only unique alignments were retained for downstream analyses. Sequencing reads were also aligned to the *Drosophila* and human genomes (dm6 and hg38) and normalisation factors calculated (Orlando et al., 2014). Ambiguous reads, i.e., reads that aligned to both reference and exogenous genomes were removed from all downstream analyses. Duplicate reads were removed using Picard. Bigwig files were generated at a resolution of 10bp using the bamCoverage utility from the deepTools suite (Ramirez et al., 2016) and data were subsequently visualised as ChIP-Rx normalised tracks using the UCSC genome browser. Peaks were called using MACS2 (Zhang et al., 2008) with FDR <0.05. All average plots were made using ngsplot (Shen et al., 2014).

PRC2.1 only regions were defined as the set of PCL2 peaks that overlapped with EPOP and SUZ12 peaks but did not overlap with AEBP2 and JARID2 peak. Conversely PRC2.1 and 2.2 positive regions were defined as the intersection of PCL2, EPOP, JARID2, AEBP2 and SUZ12 peaks. All peak overlaps were computed using bedtools (Quinlan and Hall, 2010). PRC2.1 only regions were also filtered to include only regions that had a maximum of 50 AEBP2 or JARID2 reads per region. To control for variability in antibody efficiency, PRC2.1 and PRC2.2 regions were normalised to one and those scaling factors applied to PRC2.1 only regions.

Polycomb (PcG) targets were defined as those refSeq genes whose TSS +/- 2.5 kb intersected with a PCL2 peak. All other genes were designated as non-targets. The numbers of reads mapping to PcG targets, non-targets and intergenic

regions were determined using featureCounts (Liao et al., 2014) and the distribution of these ChIP-Rx normalised read counts for PCL2, EPOP, AEBP2, JARID2, SUZ12 and H3K27me1/2/3 are represented as boxplots.

To define narrow and broad peaks in each wild-type experiment, peaks within 10kb of each other were merged into a single region. The distribution of the lengths of the resulting regions was partitioned into quartiles. Narrow and broad peaks were defined as those regions falling within the first and fourth quartiles, respectively. The number of reads mapping to narrow and broad peaks in the wild-type and knock-out conditions was determined using featureCounts (Liao et al., 2014) and the log2 fold change calculated using the spike-in normalised counts. The distributions of the fold change of ChIP signal in narrow and broad peaks was compared using a one-tailed Wilcoxon test. All bioinformatic analyses presented was performed by Dr. Darren Fitzpatrick of the Bracken Lab.

Peak numbers for Broad/Narrow Domains		
ChIP Antibody	**Broad**	**Narrow**
Epop	2336	2336
H3K27me3	1320	1319
Jarid2	1719	1722
Aebp2	708	709
Pcl2	1609	1610
Suz12	1668	1670

Gene numbers for Broad/Narrow Domains		
ChIP Antibody	**Broad**	**Narrow**
Epop	2225	922
H3K27me3	1151	435
Jarid2	1545	534
Aebp2	687	317
Pcl2	1569	805
Suz12	1632	829

2.5 *In vitro* Binding Assays

2.5.1 In vitro peptide binding assays

A custom made biotinylated p53 C-terminal domain peptide (residues 363-393, 1 μg), was incubated with bound to streptavidin-agarose beads (Invitrogen) for 2h at 4°C. GST-PCL1/2/3-PHD1 fragments (5 μg) were subsequently incubated with peptide bound beads in binding buffer (50 mM Tris-HCl, pH 7.5, 650 mM NaCl, 0.5% (v/v) NP-40 and 1 mM EDTA) for 30 min at 4 °C. Beads were then washed extensively in binding buffer, and bound protein eluted using 2× Laemlli dye. Eluted protein was analysed by western blotting.

2.5.2 Surface Plasmon Resonance (SPR)

SPR experiments were performed at 25 °C using a series S sensor chip SA with a BiaCore T200 SPR instrument (GE Healthcare). All experiments were performed in HBS-EP running buffer (10 mM HEPES, pH 7.4, 150 mM NaCl, 3mM EDTA, 0.05% (v/v) surfactant P20). Custom made biotinylated p53 C-terminal peptide was diluted in running buffer to 1 μM and immobilized at a flow rate of 10 μl min-1 to a density of 1714 response units (RU). GST-PCL fusion-proteins or GST control protein (at concentrations 10–160 nM) were injected onto the chip surface for 180 s at a flow rate of 30 μl min-1 . The dissociation phase was monitored for 180 s, and the chip surface was then regenerated between each consecutive cycle with a 120 s pulse of regeneration buffer (0.25% [w/v] SDS, 10 mM Glycine, pH 2). Individual sensorgrams were double-referenced against injection onto an empty flow cell, and GST-alone injections at equivalent concentrations. Data were fitted to a 1:1 Langmuir model using BIAevaluation analysis software (GE Healthcare). The observed results and apparent KD values were highly reproducible in replicate experiments. This analysis was conducted in collaboration with Dr. Darragh O'Donovan and Dr. David O'Connell of the UCD Conway Institute.

2.6 Structural Modelling and Evolutionary Analysis

2.6.1 Structural modelling of the PCL1 PHD1 domain

The conformational propensity of the PCL1 PHD1 domain was studied by molecular dynamics (MD) simulations and clustering analysis. All these calculations were run with the GROMACS simulation package (Hess et al., 2008), version 4.6.3. The PHD1 domain was built by homology modelling (HM) with the MODELLER software package (Eswar et al., 2006). The structure of the PHD domain of TRIM24 (PDBid 1O37) was used as single template due to its optimal structure and sequence alignment to the target. Additionally, I was particularly interested in the comparison to the TRIM24 PHD domain as this was co-crystallized with its unmodified H3(1-10)K4 peptide substrate (Tsai et al., 2010), providing clues to the interaction between the PCL1 PHD1 domain and its

p53 substrate. The HM procedure produced 5 structures and the 2 structures with the lowest discrete optimized protein energy (DOPE) score were chosen as starting structures for MD. The proteins termini were capped with acetyl (ACE) and N-methylamide (NME) residues to avoid endcharge effects. The AMBER99SB-ILDN force field (Lindorff-Larsen et al., 2010) was chosen to represent the protein atoms and counterions in all MD simulations, while water molecules were represented by the TIP4P-Ew (Horn et al., 2004) potential. During the MD simulations the temperature was held constant at 300 K by a Langevin thermostat (Grest and Kremer, 1986) with coupling time constant of 0.1 ps. The Berendsen barostat (was used to hold the pressure constant at 1 bar, with a time constant of 0.5 ps. The equations of motion were integrated using a leap-frog stochastic dynamics integrator with a 2 fs timestep. The linear constraint solver (LINCS) was used to constrain all bonds with hydrogen atoms (Blomberg and Siegbahn, 2012). Long range electrostatics were treated with the Particle Mesh Ewald (PME) method (York et al., 1993; Essmann et al., 1995). The maximum spacing for the Fast Fourier Transform (FFT) grid was chosen as 1 Å. In all simulations cutoff values for Coulomb were set to 12 Å and van der Waals interactions were switched off also at 12 Å. The protein was initially centred in a cubic periodic box with minimum distances between the protein and the box edges equal to 1.3 Å. The total charge was neutralized by the addition of two Na+ counterions with the genion tool available in GROMACS. The positions of hydrogen atoms, counterions and water molecules was optimized through 500000 steps of steepest descent algorithm. Then the system was equilibrated for 500 ps in the NVT and in the NPT ensembles, respectively. During both of these equilibration steps only water molecules and counterions were left unconstrained. The protein sidechains, except for the sidechains coordinating the two Zn2+ atoms, were equilibrated for 5 ns and subsequently all backbone atoms were also equilibrated for 5 ns, except for the backbone of the residues coordinating the two Zn2+ atoms and the residues immediately adjacent to these. The position of the Zn2+ atoms have been constrained during all MD simulations and the coordination sphere was maintained through distance restraints, where the corresponding distances in the TRIM24 PHD were used as equilibrium

values. Structure clustering was performed with the GROMACS tool g_cluster. Clusters were identified by means of the GROMOS algorithm (Daura et al., 1999) with a RMSD cut-off value of 1.5 Å. This structural modelling of the PCL1-PHD1 domain was conducted in collaboration with Dr. Elisa Fadda of National University of Ireland, Maynooth.

2.6.2 Evolutionary analysis

A maximum likelihood phylogenetic tree of PCL proteins was generated using MEGA6 (Tamura et al., 2013) from a Clustal Omega alignment (Sievers et al., 2011) of *Drosophila* and vertebrate sequences (Cunningham et al., 2014; Venkatesh et al., 2014). This evolutionary analysis was conducted in collaboration with Dr. Alan Rice and Prof. Aoife McLysaght of the Smurfit Institute of Genetics, Trinity College Dublin.

2.7 Primers and Oligonucleotides

2.7.1 Oligonucleotides

PCR Primers	Forward (5'-3')	Reverse (5'-3')
h*PHF1*_PHD1	ATGGAACTCCTCTGTTGTGTC	CTAGGTGGCGATCGCAAAGAC
h*PHF1*_BD2	ATGGCACTCACCAGCTTCCCT	CTATGCGGTGCTGGCAGAAGG
h*PHF19*_PHD1	ATGGAGCCCAAGTGCAACATC	CTACACAGCCAGTGCGAAGAT
h*PHF19*_BD2	ATGAAGCTGCTGCCTGACAAA	CTATCTGGAGTCATAGGAGAG
h*PHF1*_S106I	GTCCCTGGGAACCGGCTGGTCATCTGTGAGAAGTGTCGCCATGCT	AGCATGGCGACACTTCTCACAGATGACCAGCCGGTTCCCAGGGAC
h*PHF1*_S95G	CTCCTCTGTTGTGTCTGTCGCGGTGAGACTGTGGTCCCTGGGAAC	GTTCCCAGGGACCACAGTCTCACCGCGACAGACACAACAGAGGAG

2.7.2 RT-qPCR mRNA Primers

RT-qPCR Primers	Forward (5'-3')	Reverse (5'-3')
m*Rplpo* mRNA	TTCATTGTGGGAGCAGAC	CAGCAGTTTCTCCAGAGC
m*T* mRNA	CTTCGTGACGGCTGACAACCA	GCGCCTGAGGCTCTGGTTT G
m*Gata6* mRNA	CAGCAAGATGAATGGCCTCAGC	CAAGCCGCCGTGATGAAG G
m*Gata4* mRNA	GCGGAGTGGGCACGTAGACG	GCGGAGTGGGCACGTAGACG
m*Fgf5* mRNA	ACAGCGCTTGGGCTCACGG	GGAGTCTCCCGGGTTCCTAGGA
m*HoxA1* mRNA	AACAAGTACCTTACACGAGCGCGC	CTTCACCTGGGTCTCATTGAGCTG
m*Fgf17* mRNA	CGCCTGCTGCCTAACCTTAC	GCCCTGGTCCCTCACGTAC
m*Phf1* Exon1	GGCTGAGCCGTTTGGGTGC	GGAGCAGGGGAAGCTGGATCC
m*Phf1* Exon2	GGACAGTGCTCGAGAGGTGTGTCT	CCAGAAACTGGGAATCGTCCTCAA
m*Phf1* Exon9	TGGATGTGGCCCATCTTGTCC	GGGAGGATCTCTCGGTCAAAATCAA
m*Phf1* Exon12	CTGGGGGAGGGGTCTCACGT	TCCGACCTCCATCGTTTCCC
m*Phf19* Exon1	CCAGGGACTCTGGAAGCCTTTGG	AGGCCCCCCTTGTTAGGACTG
m*Phf19* Exon2	CAGGTCAGCAGTCCTAAGCAAAGC	CTTCCACAGGACCCAGTATTTGGA
m*Phf19* Exon7	AGGTGGTACCTTCGGATGCTACAGT	CCTCATGGAACCACTGCCTACACC
m*Phf19* Exon9	TTCACCTGGCTCTCTATAACTTGGGA	TCAAAGTCAAAGTACCGCTTCTTGCT
h*PHF1* mRNA	CTTTGGGAGGGTCAAGATGTG	TGGTACCCAAGTATAGCAGCCC
h*MTF2* mRNA	CCAGAAAAAGAACGCGTACAGG	CTTCTCCGCAAATGTGGTATTG
h*PHF19* mRNA	CAATGAGCCCATGATGTTTGG	CTGGTTACACACGGAGCAGAAG
h*EZH2* mRNA	GCCAGACTGGGAAGAAATCTGAGAA	AGCTGTCTCAGTCGCATGTACTCTGA
h*SUZ12* mRNA	GTGAAGAAGCCGAAAATGGAGCAC	TCAAAGGCCTGGAGGAAAAGCT
h*EZH1* mRNA	TAAATTGCACGCGTTTAGGCTG	TCAGATACCCTCTGCCAGTGTG
h*CDKN1A* (p21) mRNA	GACTTTGTCACCGAGACACCACTGG	GGTAGAGCTTGGGCAGGCCA
h*CCNA2* mRNA	TGCAGAAAGTATTGGGTAA	GCATCTAGTTTTGAAAGTCCTT
hTP53 mRNA	CGCTTCGAGATGTTCCGAG	GGCATCCTTGAGTTCCAAGG
h*EED* mRNA	GGTACAAACACTGAACGCCCTGAT	TTCCCCAACTTTTCCTTCCAGG
h*CDKN2A* (p16) mRNA	AGAGGATTTGAGGGACAGGGTC	CCTCTTTCTTCCTCCGGTGC

2.7.3 ChIP-qPCR Primers

ChIP-qPCR Primers	Forward (5'-3')	Reverse (5'-3')
h*INK4A*_#1	AGAGGGTCTGCAGCGG	TCGAAGCGCTACCTGATTCC
h*INK4A*_#2	GCCAAGGAAGAGGAATGAGGAG	CCTTCAGATCTTCTCAGCATTCG
h*INK4A*_#3	CAAGCTTCCTTTCCGTCATGC	GCCAGAGAGAACAGAATGGTCAGAGCCA
h*CCNA2*_ChIP	TGACGTCATTCAAGGCGA	GCTCAGTTTCCTTTGGTTTAC
h*CDX4*prom	TCCTGTGAAAGTGAAATGGCC	GGGCTTCAGGCTTTTACATAGC
h*FLT3*_prom	TCTCTTAAGGATGCGCGTCAC	CCCCTTCCACTTTGCACCAG
h*MEOX2*_ChIP	GTTCCAGGCAGAAGACTTCACG	AGTGAAAAAGTGACAGAGGGTGG
h*OLIG2*_ChIP	AGCCACGGCCATCTCCTAGAC	CAGGCACAAAGTCCCCACTATC
h*RARB*_prom	CATCCCAGTCCTCAAACAGCTC	GGGTCTATTCTTTGCCAAAGGG
ATF3 Promoter	TGTTTTTTCTTTTGCGTTTGGC	TCGTGGCAACCAAATCTAAACAG
h*HOXD11*_Prom-187F	AGGACATTTCTCTTCATGGCGTC	ATTCATCTTGATTGATTCTGGTGG
m*Gata4* tss	TCTTTCCTCCCTACTCTCAGT GGTCC	TCACCTTCTCCTCTACCAGCCCC
m*T* tss	TCCGCAGAGTGACCCTTTTTC	TACCCAACAGCCACCTTCACTTC
m*Nkx2.9* tss	TAAGGATGGAAGTGCGAGGC	TTCGCTCCAGCACTCATTCA
*Meis2*_denovo2.2_site1	TTATCCTGCCATCTGTCTTCAAGAGC	TGGCTAGTTAGTGATCTG
*Meis2*_denovo2.2_site2	TGCCATTACTTGAGACAGAGCACCAC	GAGGGAACATGAGTGGTC
*Gm3942*_denovo2.2_site1	CCTCATTTCTTCTCTCGCTC	AATACCCTAGGGCAGTTG
m*Hoxb13* INT	TCAATGTGCTGCCTTCTTCC	TTCGGCTTGTGACTTGCAAG
m*Pou4f3* INT	ATATGCAGCTTCACTCCTCAGG	AATGAAGGCACCACACATGC
m*Lrg1* INT	TTTGGCTGGACACTGTGTTG	ACAGGATGGATTTGCTGTGC
*Utp6*_GeneB	TCTATGGCCTTACCCACTGC	TGACACGTTTCTGCTTCCAG
*Yy1*_GeneB	GAAGACCCAACTTCGGTTCA	AGCTAGGAAGGGCGAAAGAC
*Fbxl11*_GeneB	GGTGCACATAGAAGCCCTGT	TACCTCAGCCGCAAAGAACT
*Ccnd2*_TSS	GGCAGGTCCAGAGCTGTGCATAC	CCGCGTTGGCACTTTGGC
*Prmt8*_TSS	CGTCCTCCGCGACAATCGAG	TGCGAAGGGATGCGAGGC
HoxA1	GAGAAGGGGCACTGGGCGAG	TGGAGATTGCCGCGTCCC
Hoxd12	CCTGAGAGCCGCCCAATTGG	GTGTTGTGGGAGCCCGCG
Gapdh TSS	AGTGTGCACCAAGGACATCCAG	CCCATTTTACTCGGGAAGCAG

Chapter 3

Polycomb-like 1 specifically binds to p53 through a divergent N-terminal PHD domain

3.1 Introduction

Polycomb group complexes PRC1 and PRC2 have well defined chromatin associated roles in regulating embryonic development and the cell cycle. However, PcG proteins also have less well characterised roles in the control of cellular proliferation independent of chromatin association. An early example of this is the *Drosophila* PCGF homologue, Psc, which ubiquitinates and mediates the proteosomal degradation of the anaphase promoting complex (APC) member and mitotic regulator, Cyclin-B (Mohd-Sarip et al., 2012). In addition, the mammalian *SCML2* gene encodes two isoforms; the chromatin and PRC1-associated SCML2A and the predominantly nucleoplasmic, SCML2B. SCML2B has been shown to associate with CDK/CYCLIN/p21 and p27 complexes and participates in the G1/S cell cycle checkpoint by stabilising p21, resulting in decreased kinase activity and inhibited progression through the cell cycle (Lecona et al., 2013). These reports suggest that PcG group proteins do play a role in the regulation of proliferation, beyond canonical chromatin-associated gene repression.

The first components of PRC2 to be linked with chromatin-independent roles in the regulation of cellular proliferation were the Polycomb-like proteins (PCL1-3) (Brien et al., 2015; Yang et al., 2013). In 2015, we characterised the functional bifurcation of PCL1-3 in regulating proliferation. We demonstrated that while *PCL2* and *PCL3* are E2F target genes, *PCL1* is a p53 target gene and is highly expressed in non-cycling, quiescent cells. While all three PCL proteins are capable of recruiting PRC2 to directly repress *INK4A,* only PCL2 and PCL3 confer an *INK4A* dependent growth advantage. In contrast, we showed that PCL1 has an *INK4A*- and PRC2-independent role in mediating proliferative arrest by binding to and stabilising the p53 protein (Brien et al., 2015).

Previous work reported a direct interaction between PCL1 and p53 in cancer cell lines, and demonstrated that an N-terminal region encompassing the first PHD domain (PHD1) and a C-terminal region, termed “BD2”, mediate the ability of PCL1 to bind p53 and block MDM2 ubiquitination (Yang et al., 2013). A

comparison of the amino acid sequences of all three PCL proteins indicated that the PHD1 and BD2 regions of PCL1 have diverged from the equivalent regions of PCL2 and PCL3 (Brien et al., 2015). The main aim of this research chapter was to fully characterise the interaction of PCL1 and p53 and to identify the critical amino acids that mediate this binding. To achieve this, I employed a combinatorial approach including; structural modelling of the PHD1 domain, *in vitro* biochemical binding assays, as well as an evolutionary analysis of *PCL1-3.*

It is known that prevalent p53 missense mutations abrogate its tumour suppressive function and lead to gain-of-function (GOF) mutations that promote cancer progression (Zhu et al., 2015). And so, the secondary aim of this chapter was to investigate the roles of PCL1 in p53-GOF dependent breast cancer cell lines, as a potential mechanism for disrupting the oncogenic activity of p53 in these cancers.

3.2 Results

3.2.1 Structural modelling of the PCL1-PHD1 domain

I sought to further explore the nature of the physical interaction between PCL1 and p53. In particular, I was interested in investigating how PCL1 can specifically bind to p53, while PCL2 and PCL3 lack this ability. The PCL1 protein contains well characterised N-terminal structural domains, whereas its C-terminal region is less well characterised. Analysis of PCL1 using software that predicts the intrinsic order of a protein sequence (IUPred) reveals that the BD2 region of PCL1, known to play a role in binding to p53, is likely to be largely unstructured (Dosztanyi et al., 2005) (Figure 3.1A). In contrast, the Tudor and two PHD domains, known to function as well folded structural moieties, are predicted to be highly ordered. For this reason, I decided to focus our analysis on the well-structured PCL1-PHD1 domain.

The PCL1-PHD1 domain shares 38% sequence homology with the PHD domain of TRIM24. The PHD domain of TRIM24 is known to bind to unmethylated histone H3 (aa1-10) and the crystal structure of this interaction has been solved (Tsai et al., 2010). Intriguingly, the CTD domain of p53, the minimal region required for the interaction with PCL1, bears sequence similarity to Histone H3 (aa1-10). In order to elucidate potential amino acids involved in the interaction between PCL1 and p53, I collaborated with structural biologists who performed molecular dynamics simulations of a homology model of PCL1-PHD1, based on the TRIM24-PHD-H3 (aa1-10) complex (Figure 3.1B). This analysis revealed several amino acids that could be essential for the stability of the complex and determining the specificity of PCL1 for p53. Among these amino acids were two Serine residues (S95 and S106). A multiple sequence alignment of PCL1-3 protein sequences revealed that these two Serine residues are unique to PCL1 and are lacking in PCL2/3 (Figure 3.1C).

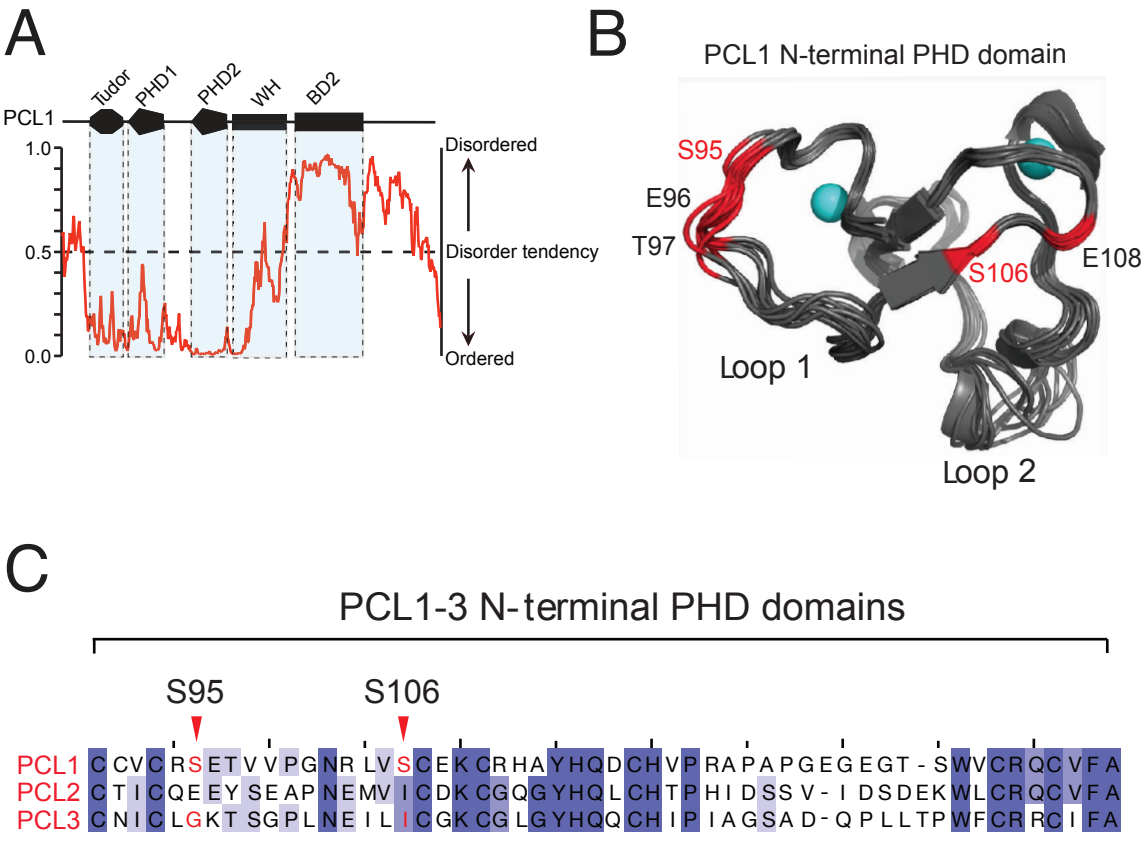


Figure 3.1 PCL1 N-terminal PHD domain structure.
(A) Protein structure disorder plot of PCL1 generated using the IUPred program, a web based server for the prediction of protein disorder based on estimated energy content.
(B) Molecular dynamics (MD) simulations of ten of the dominant conformations of PCL1-PHD1 (grey). The PCL1-PHD1 structure was determined by homology modelling based on the TRIM-24-PHD domain in complex with unmodified H3 (1-10) peptide. Zn atoms shown in cyan. Performed in collaboration with Dr. Elisa Fadda of NUI, Maynooth.
(C) Multiple sequence alignment of the PHD domains of human PCL1-3 proteins. The divergent serine residues at position 95 and 106 identified from the MD simulation are highlighted in red.

3.2.2 Serine95 and Serine106 of PCL-PHD1 are required to associate with the p53-CTD *in vitro*

I next wished to validate the potential importance of these two Serine residues in the PCL1-PHD1 domain for binding to the p53-CTD. In order to do this, I purified recombinant glutathione-S-transferase (GST) tagged fusion protein fragments representing the PHD1 domain of PCL1-3. In addition to this, I also generated single point mutant PCL1-PHD1-GST fragments in which S95 and S106 were converted to their equivalent residues in PCL3, Glycine (S95G) and Isoleucine (S106I), respectively. I also purified a double mutant fragment in which both residues were mutated (S95G/S106I) (Figure 3.2A). Next, I performed *in vitro* peptide pulldown assays using a custom made biotinylated peptide representing all 30 thirty amino acids of the unmodified p53-CTD, and analysed the precipitated fragments by western blot (Figure 3.2B). This revealed that wild-type PCL1-PHD1 bound the unmodified p53-CTD with greater strength than the equivalent wild-type PHD domains of PCL2/3. The p53-CTD is extensively post-translationally modified *in vivo*, and different combinations of PTMs are thought to refine p53 activity (Beckerman and Prives, 2010). Since our peptide binding assays were performed with an unmodified p53-CTD, they suggest that PCL1 binding to p53 may not be reliant on the presence of any particular PTM within this region (Figure 3.2B). Importantly, the binding of mutant PCL1-PHD1 to the p53-CTD was impaired, directly implicating these residues as mediators of the PCL1-p53 interaction (Figure 3.2B).

In order to, accurately quantify the *in vitro* binding of the recombinant GST-PHD1 fragments to the p53-CTD, surface plasmon resonance (SPR) was performed in collaboration with Dr. David O'Connell and Dr. Darragh O'Donovan from the UCD, Conway Institute. This analysis importantly validated the peptide pull-down assays, confirming that both Serine residues (S95 & S106) are required for the *in vitro* interaction between GST-PCL1-PHD1 and the p53-CTD peptide, with the double mutant exhibiting little to no binding (Figure 3.3B & C). This data also revealed that the binding of GST-PCL1-PHD1 to p53 is highly reproducible with

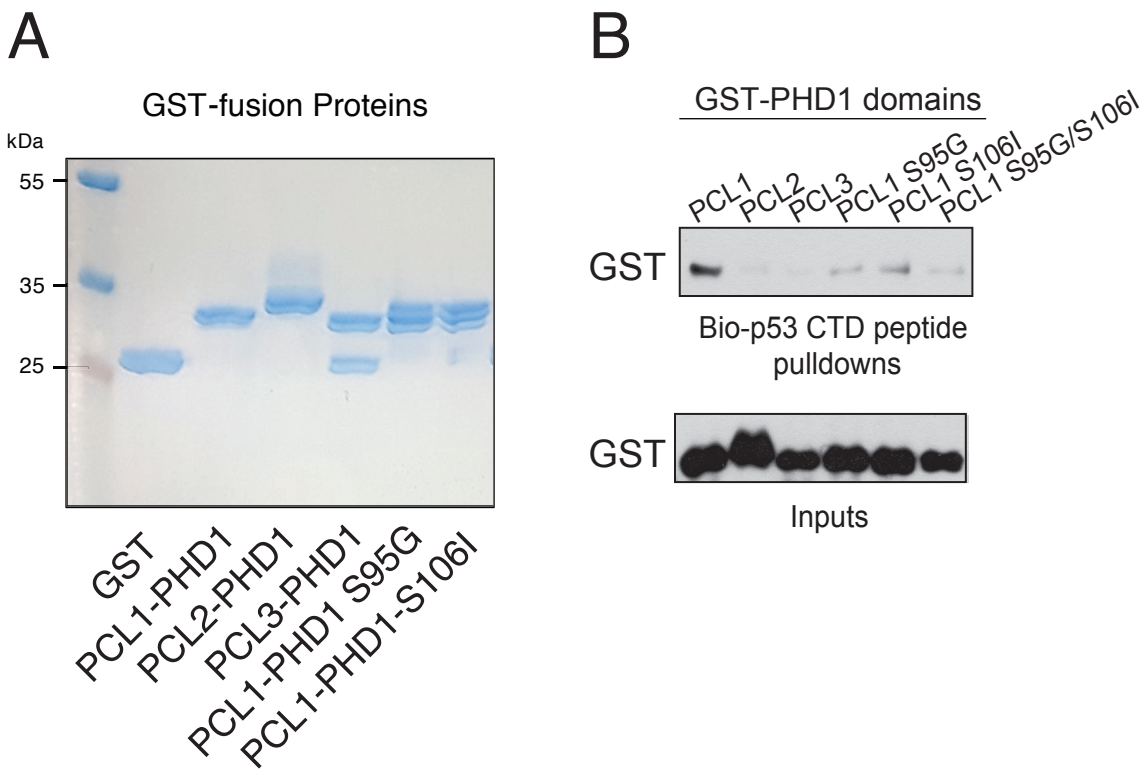
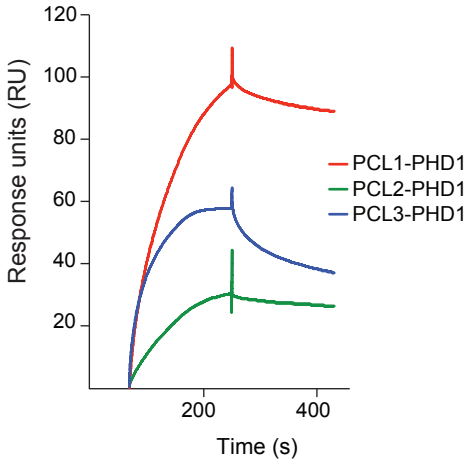
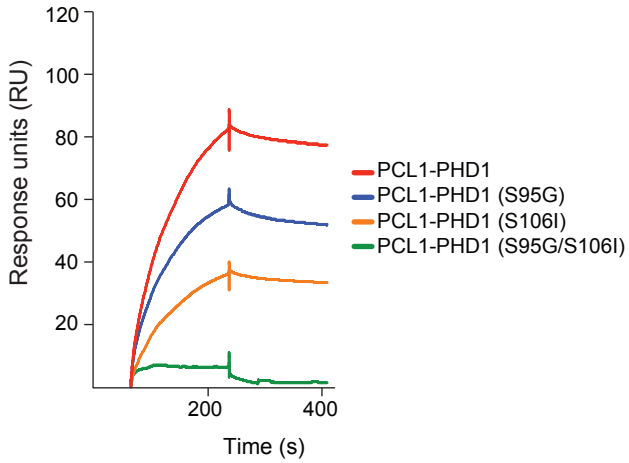


Figure 3.2 Serine 95 and Serine 106 of PCL1-PHD1 are required to associate with the p53 CTD in vitro.
(A) Representative SDS-PAGE gel image of purified recombinant GST eluted fragments for PCL1-3-PHD1 wild-type and mutant GST fusion proteins. Gel is stained with Coomassie Brilliant Blue.
(B) Peptide pulldowns of recombinant GST-tagged PCL1-3 PHD1 domains and mutant PCL1-PHD1 domains with a biotinylated p53 CTD peptide.

A



B



C

WT PCL1 PHD1	K_d (1/s)	K_a (1/Ms)	K_D (nM)
PCL1-WT	6.13E-04	5.33E+04	26.38
PCL2-WT	9.55E-04	5.51E+04	33.75
PCL3-WT	1.96E-04	9.75E+04	51.93
Point mutant PCL1 PHD1			
S95G	8.64E-08	6.40E+04	32.88
S106I	8.45E-08	4.97E+04	38.59
S95G/S106I	0	0	0

Figure 3.3 PCL1-PHD1 association with the p53 CTD is disrupted by mutating critical Serine residues.

(A) Representative SPR sensograms for wild-type PCL1-3 PHD domains binding to biotin-p53 CTD

(B) Representative SPR sensograms for wild-type and mutant PCL1-PHD domains binding to biotin-p53 CTD.

(C) Table showing the binding affinities of wild-type PCL1-3 PHD1 and point mutant PCL1-PHD fragments. Affinity rate constants were determined from a concentration series of each protein binding to the biotin—p53 CTD peptide. All SPR analysis was performed in collaboration with Dr. Darragh O'Donovan and Dr. David O'Connell of the UCD Conway Institute.

an apparent K_d of 26.4nM, as determined from independent measurements at four different concentrations of the protein (Figure 3.3C).

3.2.3 PCL1-PHD1 contains two divergent Serine residues that arose during mammalian evolution

In order to fully understand the evolutionary origins of the two unique PCL1 serine residues a multiple sequence alignment and evolutionary tree analysis of the three vertebrate PCL proteins was performed in collaboration with evolutionary biologists, Dr. Alan Rice and Prof. Aoife McLysaght from the Smurfit Institute of Genetics, (Figure 3.4). This confirmed that the three *PCL* genes are present in all vertebrates and linked through two whole genome duplications at the base of the vertebrate tree (Makino and McLysaght, 2010). I was specifically interested in exploring whether the ability to bind p53 is a newly evolved function of PCL1. In other words, is the binding a "neofunctionalised" event in which PCL1 has specifically gained this ability or alternatively, is it an ancestral function that was lost from the other two PCL proteins. The pattern of sequence divergence at sites corresponding to human S95 and S106 indicates that the substitutions giving rise to each of these residues occurred uniquely in PCL1, most likely after the whole-genome duplication events (Figure 3.4). In contrast, the equivalent residues in PCL2 and PCL3 have undergone little to no divergence from the ancestral sequence. Importantly, these two serine residues are completely conserved across all mammalian PCL1 genes, and furthermore, uniquely co-occur in mammals. This suggests that PCL1 has undergone a neofunctionalization event during recent vertebrate evolution in which it has specifically selected for the ability to bind the p53-CTD in mammals.

3.2.4 Crystal structure of human PCL1-PHD1 reflects initial molecular dynamics simulations

My initial analysis of the PCL1-PHD1 domain used a molecular dynamics approach based on a homology model of the TRIM24-PHD-H3 (aa1-10) ternary

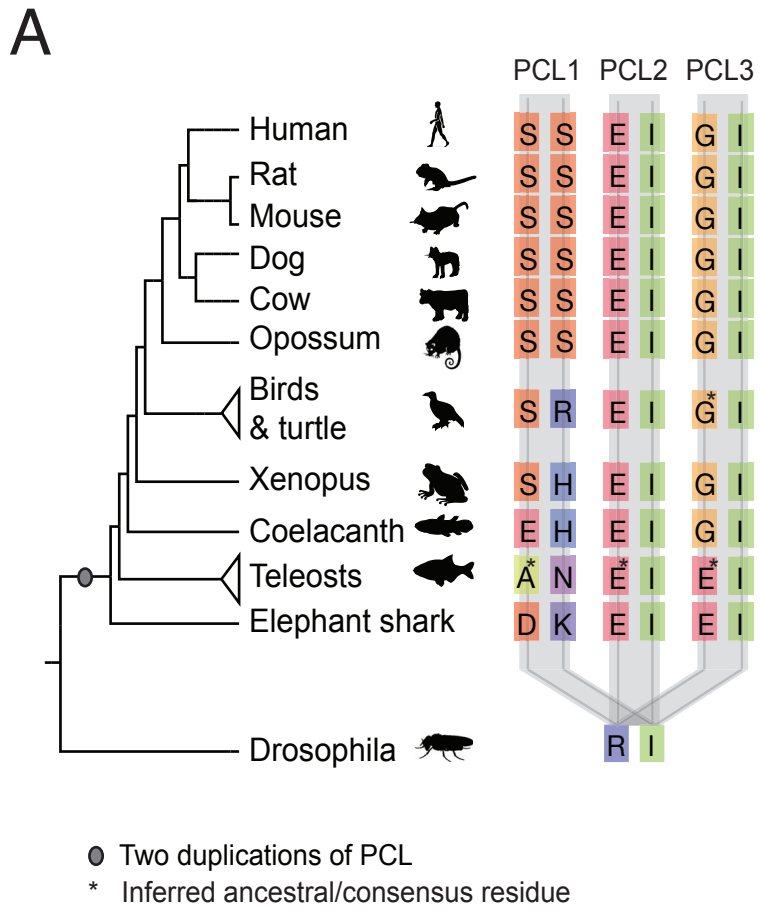


Figure 3.4 The N-terminal PHD domain of human PCL1 contains two divergent Serine residues that arose during evolution.
(A) Segment of alignment of vertebrate PCL proteins showing residues aligned to Homo Sapiens PCL1 Ser95 and Ser106. There is a single Pcl in non-vertebrates, represented by Drosophila. The phylogenetic tree indicates species relationships; branch lengths are not to scale. All evolutionary analysis was performed in collaboration with Dr. Alan Rice and Prof. Aoife McLysaght of the Smurfit Institute of Genetics, Trinity College Dublin.

complex to simulate the binding of PCL1-PHD1 to the p53-CTD (Figure 3.1B). I took advantage of this *in silico* approach due to the lack of structural data for any of the PCL proteins. However, since publication of our results (Brien et al., 2015), in-depth structural and crystallographic studies focused on PHF1 and MTF2 have been published (Choi et al., 2017; Li et al., 2017). These studies mainly concentrated on the C-terminal WH domain of PCL1 and its ability to bind DNA. However, importantly, one of these studies crystallised a fragment containing the TUDOR/PHD1/PHD2/WH domains of PHF1 (Li et al., 2017). I was interested in analysing the structural data from this study to see test the accuracy of our previous *in silico* simulations. Therefore, I downloaded the crystallographic data (PDB: 5XFN, 5XFO), isolated the PHD1 domain of PCL1 and observed an accurate overlap with our initial molecular dynamics model (Figure 3.5B - orange). The positioning of S95 and S106 in the simulation is accurately reflected in the PHD1 crystal structure, confirming the efficacy and objectivity of our original approach. It is also worth noting that the residues corresponding to S96 and S106 in the PCL2-PHD1 domain (E110/I121) (Figure 3.5B blue), are positioned relatively close together when the PHD domains of the two proteins are overlapped. This may give us an insight into the electrostatic nature of the PCL1-p53 interaction given the opposing polarity of these amino acids, particularly S106 (PCL1) and I121 (PCL2). The p53-CTD contains a high number of positively charged Lysine and Arginine residues and it will interesting to explore the specific amino acids of the p53 required for the interaction.

3.2.5 Purification of the BD2 region of PCL1

We and others have demonstrated that it is not only the PHD1 domain of PCL1 that contributes to the binding with p53 (Brien et al., 2015; Yang et al., 2013). In addition to the PCL1-PHD1 domain, a C-terminal fragment termed, BD2, is also required for the interaction. The BD2 region encompasses amino acids 349-458 of PCL1 and is highly divergent in terms of sequence homology from PCL2/3. I had previously attempted to purify full-length PCL1-3 from *E. Coli* and Sf9 insect cells. However, all three proteins were insoluble in both expression systems.

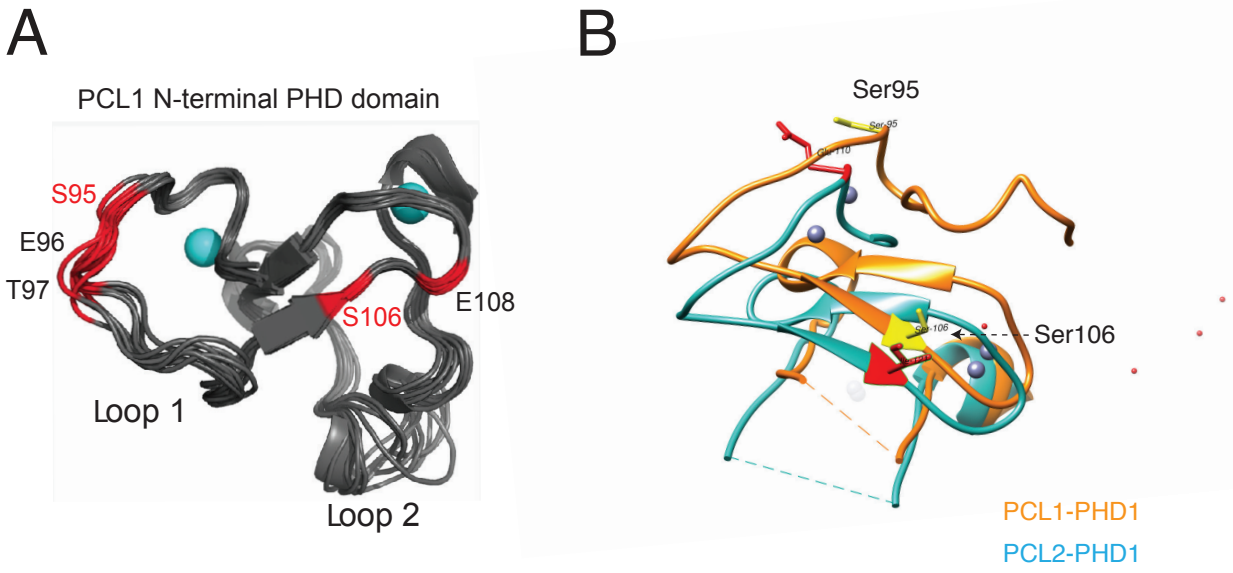


Figure 3.5 Crystal structure of PCL1-PHD1 reflects the initial molecular dynamics simulation of the same domain.
(A) Molecular dynamics (MD) simulations of ten of the dominant conformations of PCL1-PHD1 (grey). The PCL1-PHD1 structure was determined by homology modelling based on the TRIM-24-PHD domain in complex with unmodified H3 (1-10) peptide. Zn atoms shown in cyan.
(B) Crystal structure of PCL1-PHD1 (5XFN) and PCL2-PHD1 (5XFR) with highlighted Ser95 and Ser106 of PCL1 (Li et al., 2017). The equivalent residues of PCL2 are also indicated, Glu110 and Iso121 respectively. Overlapped images were created using the UCSF Chimera software for interactive analysis and visualisation of molecular structures.

Hence, I attempted a similar strategy to the PHD1 domain of PCL1, and attempted to purify recombinant GST tagged fusion protein fragments representing only the BD2 regions of PCL1 and PCL3 (Figure 3.6), and investigate their capacity to bind the p53-CTD, through *in vitro* binding assays. As demonstrated previously, I was able to purify PCL1-PHD1 GST-fusion fragments (Figure 3.6A). However, while I could purify a small amount of the PCL1-BD2-, the GST eluted fragments contained a large amount of breakdown products (Figure 3.6B). I then attempted to isolate the fragments of interest (~38 kDa) by running the eluted fragments through a size exclusion column with a 30 kDa molecular weight cut off, but this did not clean up the respective sample preparations (Figure 3.6C). Given the fact that this region is predicted to be highly disordered and the relative ease with which the N-terminal domains can be purified, I believe the C-terminal region of PCL proteins to be the source of the insolubility. I am now working on purifying a larger fragment encompassing the PHD2, WH and BD2 regions of PCL1, with expectations that could will help with purification issues. However it might be the BD2 region is just too disordered to exist in a soluble form.

3.2.6 PCL1 in gain-of-function p53 breast cancer

I next wondered what could be the potential cancer relevance of the PCL1-p53 interaction. Interestingly, while the loss of the tumour suppressive function of the *TP53* gene is required for progression of most cancers, remarkably, this gene can also acquire oncogenic functions in many cancer types due to neomorphic (gain-of-function, GOF) mutations (Dittmer et al., 1993; Muller and Vousden, 2014). These missense mutations occur at “hotspots” within the core domain of the p53 protein and include, R175H, R248Q/R248W, R249S and R273H, and are particularly common in triple negative breast cancers (Shah et al., 2012). Interestingly, the R175H mutation has been linked with the initiation of mammary tumorigenesis and promotion of more aggressive breast cancers (Lu et al., 2013). The mechanism of action of these p53-GOF mutations remains largely unknown, however, it is thought that they co-operate with other transcription factors to re-

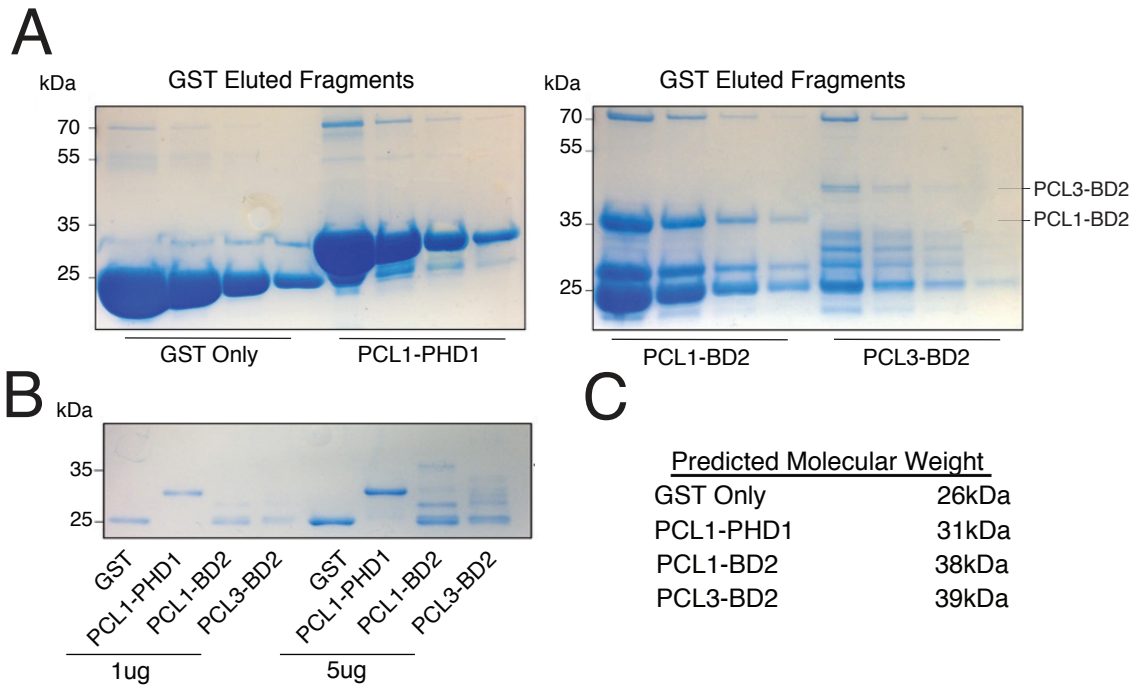


Figure 3.6 Expression and purification of GST-PCL1/3-BD2 fusion proteins.

(A) Representative SDS-PAGE gel image of purified recombinant GST eluted fragments for GST-only and GST-PCL1-PHD1 (left), GST-PCL1-BD2 and GST-PCL3-BD2 (right) fusion proteins. Breakdown products for both PCL1- and PCL3-BD2 are seen below expected size of the fusion proteins. Gel is stained with Coomassie Brilliant Blue.

(B) Representative SDS-PAGE gel image of purifications for GST-PCL1-PHD1, GST-PCL1-BD2 and GST-PCL3-BD2 fusion proteins following size exclusion (30kDa MWCO).

(C) Table of predicted molecular weights for PCL1/3-BD2 fragments.

direct p53 to novel target genes, upregulating oncogenic pathways (Do et al., 2012; Zhu et al., 2015). Importantly, since the majority of p53-GOF mutations occur outside the CTD region, I hypothesised that they are unlikely to affect the PCL1-p53 interaction. Hence, I was interested in exploring the roles of PCL1 breast cancer cell lines harbouring p53-GOF mutations.

Towards this, I decided to deplete the levels of PCL1 and p53 with previously validated siRNAs (Brien et al., 2015) in two p53-GOF triple negative breast cancer cell lines, MDA-MB-468 (R273H) and BT-549 (R249S). While the knockdown of both PCL1 and p53 levels was effective in each cell line, only the p53 knockdown resulted in decreased cellular proliferation, as illustrated by a reduction in the protein levels of cell cycle regulated transcription factor E2F1 (Figure 3.7). We have previously shown that knockdown of PCL1 results in destabilisation of p53 in quiescent human fibroblasts (Brien et al., 2015). Therefore, I was interested in investigating if PCL1 acts to stabilise p53-GOF mutants. While I efficiently knocked down PCL1, there was no observable decrease in p53-GOF protein levels (Figure 3.7). Taken together, these data suggest that while PCL1 plays an important role in maintaining wild-type p53 levels in primary quiescent fibroblasts, it does not seem to have the same function in the p53-GOF triple negative breast cancers tested here.

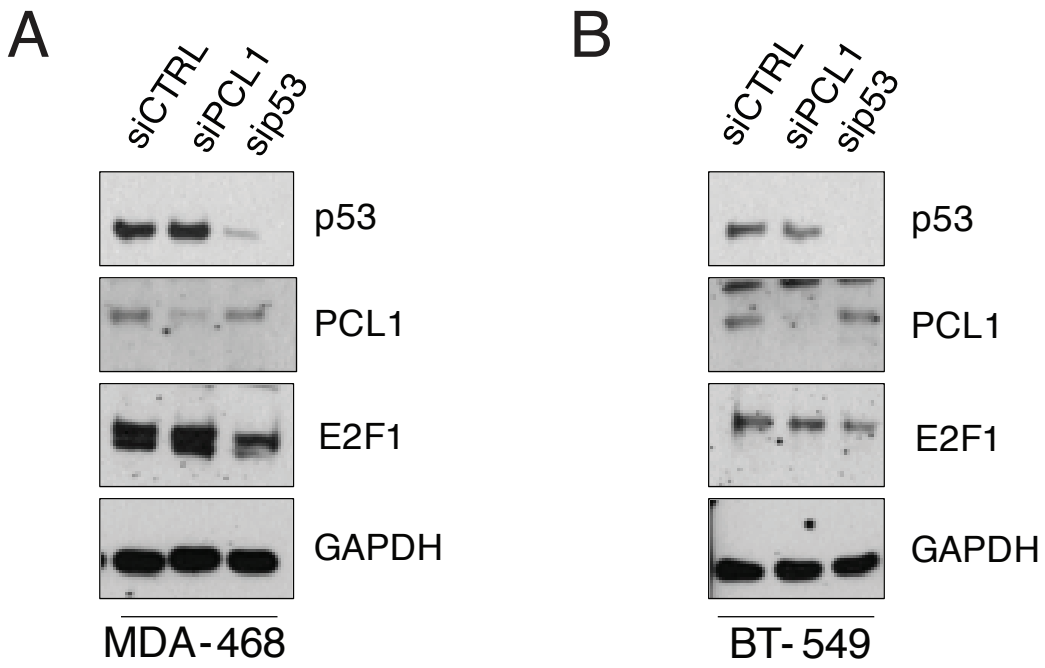


Figure 3.7 Knockdown of PCL1 in gain-of-function p53 dependent breast cancer cells does not phenocopy loss of p53.
(A) & (B) Western blot analyses using the indicated antibodies on whole cell lysates from either MDA-468 or BT-549 gain-of-function p53 dependent breast cancer cell lines transfected with siRNA constructs targeting PCL1 or TP53. Cell were harvested 72 hours after transfection with siRNA. Levels of E2F1 are used indicate proliferative capacity.

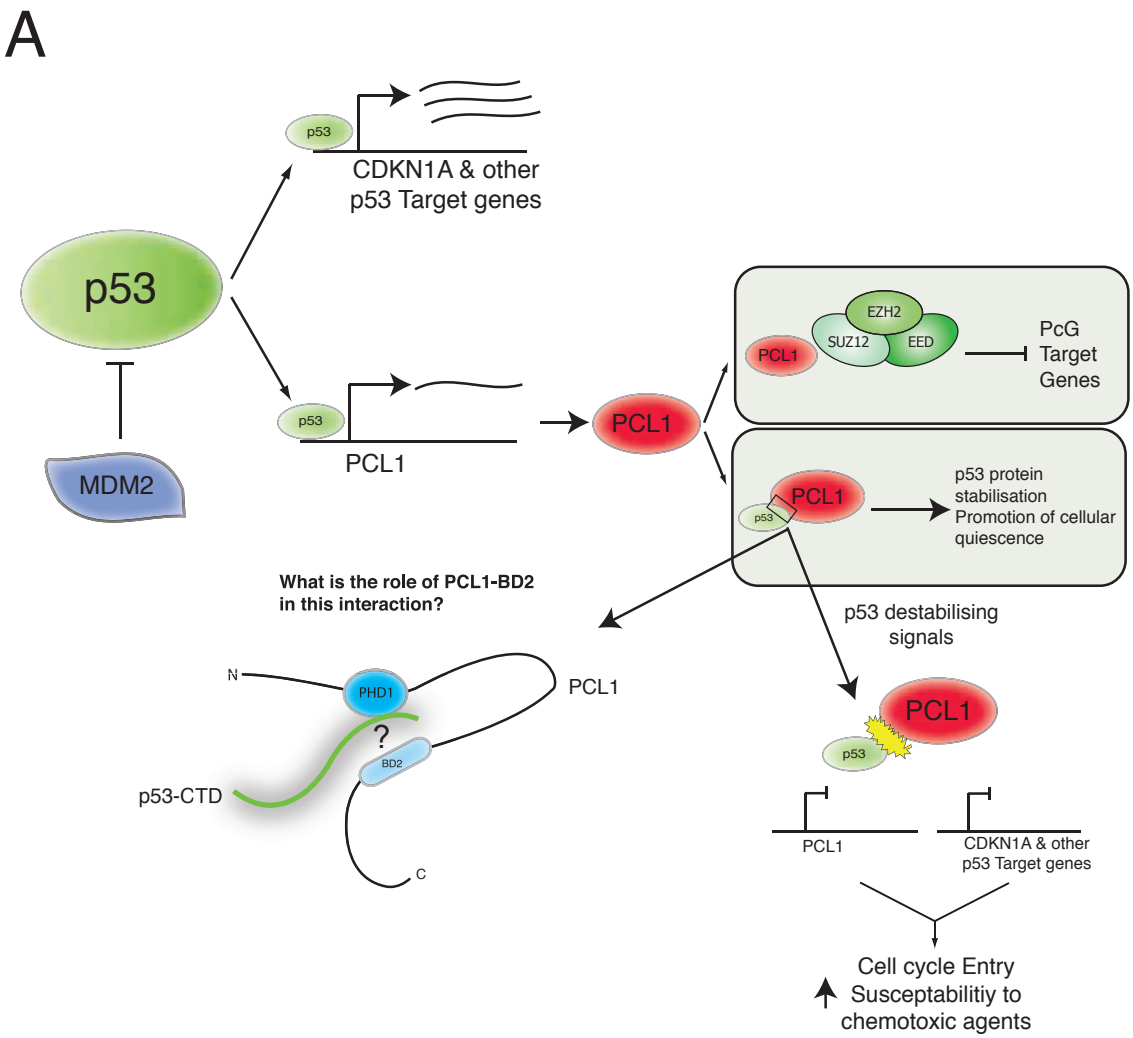


Figure 3.8 Future perspectives – Targeting the PCL1-p53 interaction in cancer.

(A) Key open questions that remain; What is the size and stoichiometry of the PCL1-p53 complex? How do the PHD1 and BD2 domains cooperate to facilitate binding to p53? How can we exploit this knowledge towards developing new therapeutics directed against quiescent cancer stem cells?

3.3 Discussion

Initial observations illustrated that *PCL1*, but not *PCL2* and *PCL3*, is highly expressed in quiescence, suggested *PCL1* may have a specific role in non-dividing cells. Pursuing this, we unravelled an essential PCL1-p53 regulatory axis in primary quiescent human fibroblasts (Brien et al., 2015). This was the first evidence of a direct link between a Polycomb group protein and the regulation of cellular quiescence, and raises the intriguing possibility that PCL1 may be required to maintain the quiescent state of tissue specific stem cells *in vivo.* In this regard, it is worth noting that loss of PRC2 core components such as *EZH1* or *EED* in hematopoietic stem cells (HSCs) leads to aberrant activation of Polycomb target genes and exhaustion of the stem cell pool (Hidalgo et al., 2012; Xie et al., 2014).

3.3.1 PCL1 has recently acquired the ability to bind p53

Here, I have provided evidence that mammalian *PCL1* has recently evolved the ability to bind to and stabilise p53 through the acquisition of two unique Serine residues in its N-terminal PHD domain. These serines are not present in PCL2 and PCL3, and are critical for the interaction between PCL1 and p53. I have demonstrated that the *PCL1* gene has undergone a neofunctionalization event to select for these amino acids and the ability to bind p53. However, while mammalian *PCL1* has acquired this relatively new function, it importantly, retains its ancestral function within PRC2 . This functional divergence of *PCL1* goes some way to explaining the selection and ultimate maintenance of *PCL* genes in vertebrate genomes. They are not redundant copies of an ancestral gene, as might have seemed the case but have novel functionality that has evolved along the ancestral one. The concept of neofunctionalization has been popular for over 40 years, but there are relatively few examples of this where the sequence substitutions have been clearly described and characterised (Conant and Wolfe, 2008; Ohno, 1970). Therefore, the *PCL1* gene represents an archetypal example of neofunctionalization, and further studies into its functional interplay with p53

will likely shed new light on the evolution of the molecular mechanisms governing cellular quiescence.

3.3.2 PCL1 and its potential roles in cancer

The *PCL1* gene is frequently translocated in endometrial stromal sarcomas, and the fusion protein products have been proposed to directly contribute to the pathogenesis of these cancers (Micci et al., 2006). However, the mechanisms of how PCL1 fusion proteins could contribute to oncogenesis are so far unexplored. Interestingly, while the PCL1 fusion genes retain the vast majority of the wild-type PCL1 coding sequence, the promoter region is lost, implying that accurate transcriptional regulation of PCL1 is lost in these cancers. This study suggests that PCL1 fusion proteins would preserve the ability of wild-type PCL1 to interact with p53 raising interesting questions about the status of the *TP53* genes in these cancers.

I also provide evidence that PCL1 does not have any major role in stabilising oncogenic p53 mutants in two triple negative breast cancer cell lines (Figure 3.7). However, this analysis was not exhaustive and used only RNAi mediated knockdown of *PCL1*. A more thorough characterisation of the function of *PCL1* in these cancers needs to be performed, employing the use of CRISPR-Cas9 mediated genetic manipulation. A challenge here will be the fact that a straightforward knockout of *PCL1* will likely abolish any PRC2 associated function of PCL1 in these cells. Therefore, to further characterise the roles of PCL1 in these cancers it will be necessary to generate a conditional allele producing a PCL1 protein incapable of interacting with p53 but still retaining the ability to bind PRC2. This will require a more extensive molecular characterisation of the PCL1-p53 interaction (Figure 3.8). Towards this and with the help of collaborators, I am in the process of conducting nuclear magnetic resonance (NMR) analysis of the PCL1-PHD1 domain in complex with an unmodified p53-CTD peptide to try and elucidate the amino acids in the p53 CTD critical for the interaction. I am also attempting to generate high quality

recombinant PCL1-p53 complexes to carry-out in-depth structural analysis of the interaction using Cryo-Electron Microscopy (Cryo-EM).

Further structural characterisation of the PCL1-p53 complex could potentially be exploited to interrupt this interaction for therapeutic gain in cancer (Figure 3.8). For example, it will be interesting to examine the relationship between PCL1 and wild-type p53 in quiescent, slowly proliferating quiescent cancer stem cells *in vivo*. Disrupting the PCL1-p53 interaction in this context could force these non-dividing cancer cells into the cell cycle, potentially rendering them more sensitive to chemotherapies that only target proliferating cells (Figure 3.8) (Brien and Bracken, 2016).

Chapter 4

A novel catalytically inactive PRC2 complex unique to non-dividing quiescent cells

4.1 Introduction

Polycomb group proteins have been extensively linked to regulation of the cell cycle. For example, the EZH2 gene is regulated by E2F transcription factors and is consequently highly expressed in rapidly proliferating cells such as cancer cells. (Bracken et al., 2003). Polycomb proteins are also key upstream repressors of the *INK4A-ARF* tumour suppressor locus in mammalian cells (Bracken et al., 2007). The *INK4A* gene encodes the p16 protein, a vital cyclin-dependent kinase (CDK) inhibitor and key regulator of cellular senescence (Gil and Peters, 2006). Furthermore, the *Drosophila* PRC1 component, Psc, has been shown to associate with CyclinB and the anaphase promoting complex (APC), independently of other PRC1 components (Mohd-Sarip et al., 2012). In addition to this, the PRC1 associated *SCML2* gene encodes two isoforms, SCML2A and SCML2B, the latter of which has been shown to associate with and enhance the function of key cell cycle regulators including, CDK inhibitors p21 and p27 (Lecona et al., 2013). These many links between Polycomb and proliferative control illustrate their potential relevance in cancer. Supporting this, EZH2 was initially thought to be an oncogene based on its high expression levels in proliferating cancer cells (Bracken et al., 2003). However, it has also been shown to be deleted in some myeloid malignancies and have recurrent gain of function mutations in many cancers including B-cell lymphomas (Kim and Roberts, 2016). In fact, small molecule inhibitors targeting EZH2 catalytic are now in clinical trials for the treatment of B-cell lymphomas and Malignant Rhabdoid tumours, and are showing early promise (Italiano et al., 2018; Makita and Tobinai, 2018).

We previously reported that, like EZH2, the PCL2 (MTF2) and PCL3 (PHF19) Polycomb-like proteins are regulated by E2F transcription factors and highly expressed in proliferating cells (Brien et al., 2015). In contrast, PCL1 (PHF1), like EZH1, is more highly expressed in non-proliferating cells. We demonstrated that PCL1 is both transcriptionally regulated by p53 and its protein product acts to stabilise p53 and promote cellular quiescence. However, we did not explore the potential for divergent assemblies of PRC2 between proliferating and non-proliferating cells. Previous work established that EZH1 and EZH2 expression

correlated with the proliferative capacity of cells during differentiation, with EZH2 expressed in proliferating progenitor cells, and EZH1 highly expressed in less proliferative and more terminally differentiated cell types (Figure 4.12) (Margueron et al., 2008; Xu et al., 2015). In this research chapter, I focus particularly on the EZH1/2 expression switch between normally cycling and quiescent cell populations. Quiescence (or G0), is a reversible and actively maintained cell state, important for many adult stem cell populations (Cheung and Rando, 2013). Important regulators of cellular quiescence include members of the p53 tumour suppressor pathway, including p53 and p21 (CDKN1A), its downstream target gene (Liu et al., 2009). The loss of p53 and p21 in neural stem cells (NSCs) and hematopoietic stem cells (HSCs) leads to an expansion in progenitor cell number stem cell number. This occurs due to reduced ability of the stem cells to remain quiescent and their consequent entry into the cell cycle (Liu et al., 2009; Meletis et al., 2006). Certain post-translational modifications of histones have also been shown to be key determinants of the quiescent state, such as H4K20me2 (Liu et al., 2013) The histone methyltransferase responsible, Suv4-20h1, has been shown to be crucial in the maintenance of adult quiescent stem cells. Depletion of Suv4-20h1 leads to the exhaustion of adult muscle stem cells, thereby providing an excellent example of how chromatin modifiers can be essential in the maintenance of this particular cell state (Boonsanay et al., 2016). Furthermore, it has been shown *in vivo* that conditional knockout of *Ezh1* and *Eed*, but not *Ezh2* induces significant loss of adult quiescent HSCs, as well as impaired self-renewal capacity (Hidalgo et al., 2012; Xie et al., 2014). This again implicates chromatin associated proteins in the regulation of cellular quiescence.

Consistent with this, the main aims of this chapter are to explore the composition and function of PRC2 in non-proliferating and quiescent cells. I show that EZH1 and PCL1 are highly expressed in quiescent populations and down regulated as cell enter the cell cycle. I also define a G0-PRC2 complex consisting of EZH1, PCL1 and EED, which importantly lacks both EZH2 and SUZ12. I show that quiescent cells are insensitive to both catalytic and allosteric inhibition of PRC2. Whereas, cycling cells treated with PRC2 inhibitors undergo premature cellular

senescence due to the activation of *INK4A*, quiescent cells treated with the same inhibitors show no adverse effects and grow at the same rate as control treated cells. These results may have important implications in the treatment of cancers with EZH2 inhibitors. The data in this chapter suggests that quiescent cancer stem cells should be insensitive to EZH2 inhibition, which may lead to tumour recurrence following initial regression (Figure 4.13).

4.2 Results

4.2.1 EZH1 and PCL1 form a divergent PRC2 complex in quiescent cells

Firstly, I analysed the mRNA and protein levels of various core and accessory PRC2 subunits in quiescent and cycling cells. Primary human fibroblasts are induced to quiesce by removing serum for 96-120 hr and stimulated to re-enter the cell cycle by addition of serum for 24 hrs. Consistent with previous observations, both *EZH1* and *PHF1* (PCL1) are significantly upregulated on both the mRNA and protein level in quiescent fibroblasts compared to proliferating (Bracken et al., 2003; Brien et al., 2015; Cheung and Rando, 2013) (Figure 4.1A & B). Next, I wished to determine if this divergent expression pattern leads to a shift in PRC2 complex composition. To test this, I performed co-immunoprecipitations of PCL1-3 in quiescent fibroblasts, and observed that PCL1 specifically pulls down only EZH1 and EED, but importantly, not EZH2 or SUZ12. (Figure 4.2). Next, I performed sucrose gradient fractionation experiments on whole cell lysates from both quiescent and cycling fibroblasts, and observed co-sedimentation of PCL1, EZH1 and EED in quiescent cells (Figure 4.3). Interestingly, PCL1 also co-sedimented with p53 at lower molecular weights in quiescent cells, consistent with our previous demonstration that PCL1 also form an independent complex with p53 (Figure 4.3) (Brien et al., 2015). Taken together, these data suggest that the predominant PRC2 complex in quiescent cells consists of PCL1, EZH1 and EED, hereafter termed, G0-PRC2.

4.2.2 G0-PRC2 chromatin association

I next decided to determine if this switch in PRC2 complex configuration in quiescent cells results in a change of PRC2 component enrichment on chromatin. In order to do this I performed chromatin enrichment western blot analysis in quiescent and cycling primary human fibroblasts. As expected, EZH1 and PCL1 are enriched in the chromatin fraction in both quiescent and cycling cells, while SUZ12 and EZH2 are only on chromatin in the cycling population (Figure 4.4A).

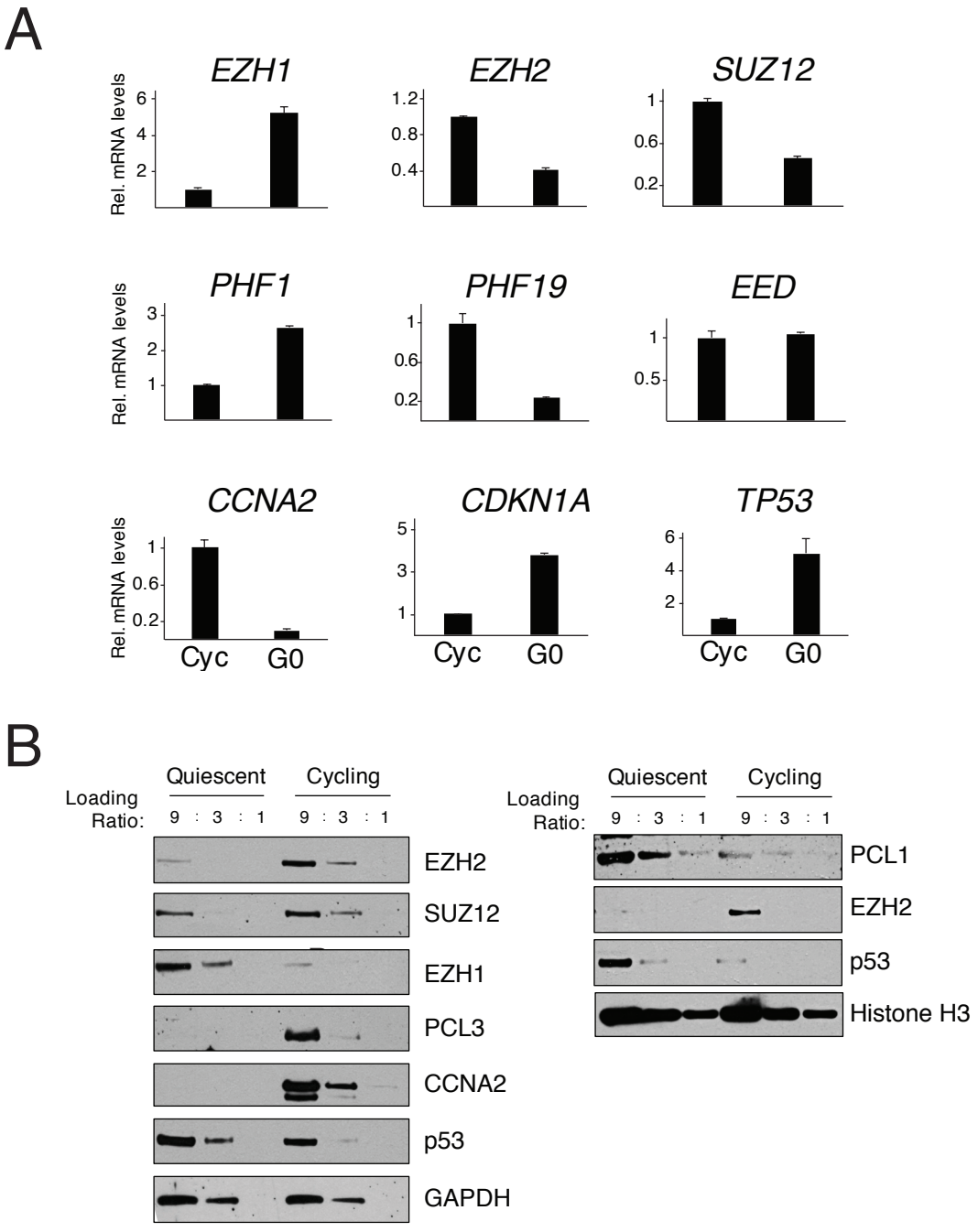


Figure 4.1 Protein and mRNA levels of PRC2 components in quiescent and cycling fibroblasts.

(A) Quantitative RT-PCR analyses of the indicated mRNA transcripts from serum starved (Quiescent - G0) or serum stimulated (Cyc) early passage human diploid fibroblasts. Cells are serum starved for 120hrs to enter quiescence and subsequently induced to re-enter the cell cycle following the addition of 20% serum for 24hrs. Data presented are a representative sample of three biological replicates and error bars indicate standard deviation of individual triplicate qPCR data.

(B) Western blot analyses using the indicated antibodies on whole cell lysates from human diploid fibroblasts. Cells were serum starved for 120hr (Quiescent) and stimulated to re-enter the cell cycle with the addition of 20% serum for 24hrs (Cyc).

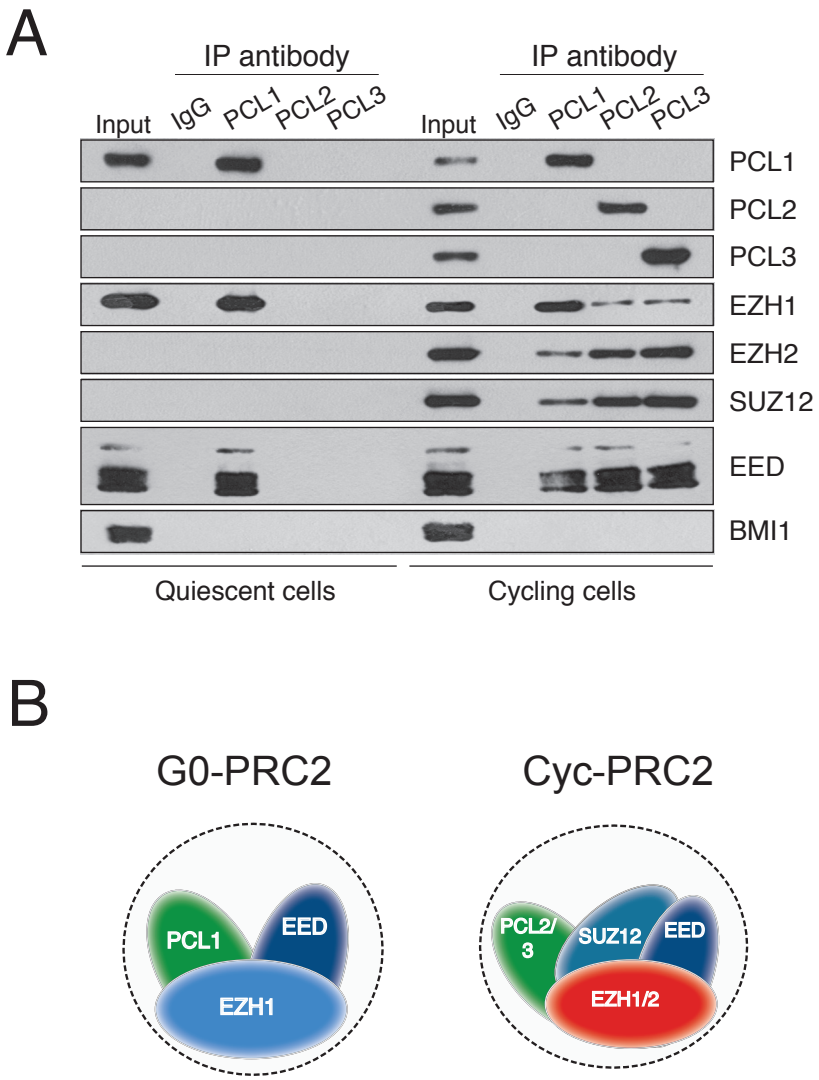


Figure 4.2 A variant "G0-PRC2" complex in quiescent cells consisting of EZH1, EED and PCL1.

(A) Western blot analyses on immunoprecipitations of the indicated proteins in quiescent and cycling populations of early passage human diploid fibroblasts.

(B) Schematic of the divergent PRC2 complexes that exist is quiescent and cycling cells. G0-PRC2 contains EZH1, PCL1 and EED, lacks SUZ12 and exists solely in quiescent cells, whereas, there are numerous combinations of the PRC2 complex present in cycling cells.

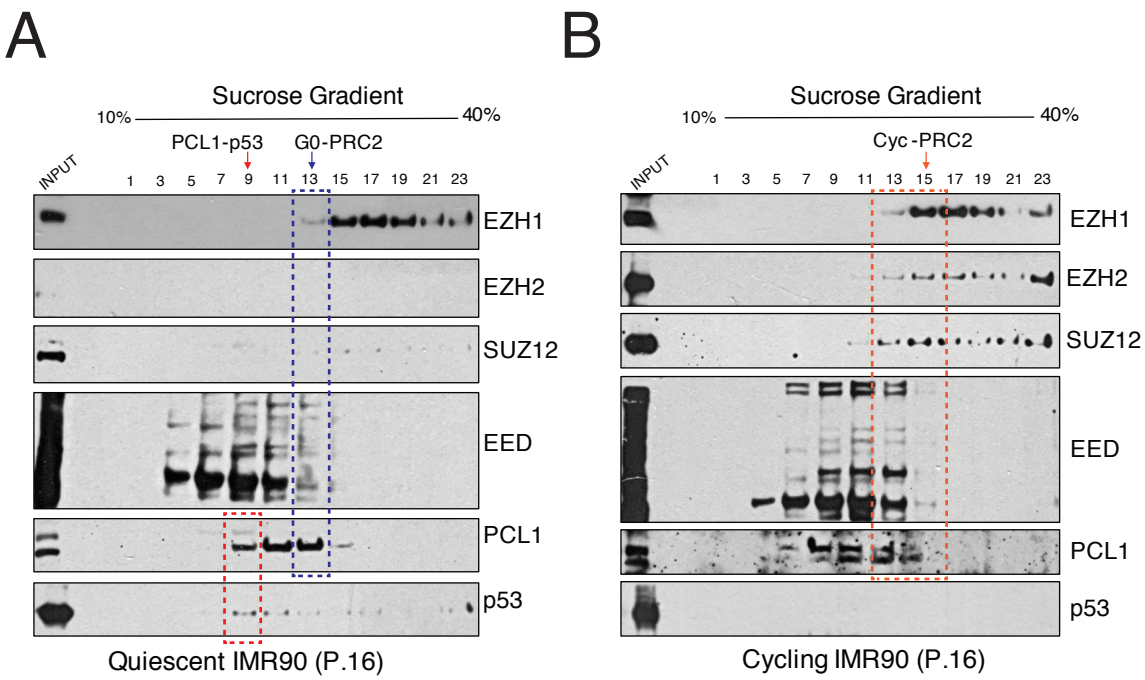


Figure 4.3 PCL1 co-sediments with EZH1 and EED, as well as p53 in quiescent fibroblasts.

(A) &(B) Sucrose gradient sedimentation analysis of quiescent (A) and serum stimulated (cycling) (B) fibroblasts, demonstrating that PCL1 co-sediments with EZH1 and EED (blue box), as well as p53 (red box) at lower molecular weights in quiescence. PRC2 sediments at higher molecular weights in cycling cells due to the presence of EZH2 and SUZ12.

Furthermore, chromatin immunoprecipitation (ChIP) experiments confirmed that PCL1 is the only Polycomb-like protein that is enriched at Polycomb targets in quiescent cells, while PCL2/3 are only present at these loci in cycling cells (Figure 4.4B). In addition, both EZH2 and SUZ12 were only bound to PcG targets in proliferating cells and were absent in quiescent cells (Figure 4.4B). I next performed ChIP followed by next generation sequencing (ChIP-Seq) of PCL1, PCL3, EZH1 and EZH2, and observed that PCL1/PCL3 and EZH1/EZH2 occupy the same genomic loci in quiescent and cycling cells respectively (Figure 4.4C). Importantly, the EED core PRC2 subunit remained unchanged in both its expression levels and chromatin occupancy in the quiescent and cycling contexts, suggesting that is an integral part of both G0- and typical PRC2 complexes of cycling cells (Figure 4.2A & 4.4A & B). Intriguingly, the global and local levels of H3K27me3 do not change between quiescent and cycling cells (Figure 4.5). This is despite the observation that neither EZH2 nor SUZ12 are highly expressed in quiescent cells. Previously, EZH1 containing PRC2 complexes were reported to be poor methyltransferase enzymes relative to EZH2 containing complexes (Lee et al., 2018; Margueron et al., 2008). It might be expected given these findings that G0-PRC2 would have little or no methyltransferase activity and would lead to a reduction in H3K27me3 in quiescent cells. However, this is not the case, as H3K27me3 levels remain constant. Taken together, these results suggest a model whereby EZH1, EED and PCL1 define a novel “G0-PRC2” complex in quiescent cells, that exists uniquely in quiescent cells and intriguingly lacks EZH2 and SUZ12.

4.2.3 Purification of a recombinant G0-PRC2 complex

I next decided to investigate whether I could purify a recombinant G0-PRC2 complex. The aim was to investigate whether it has any *in vitro* methyltransferase activity. To do this, I cloned constructs to express Flag/His-PCL1 and Flag/His-EED as well as several other untagged PRC2 components. I then generated baculovirus and purified three separate and distinct PRC2 complexes from HighFive insect cells (Figure 4.6A & B). These complexes consisted of a

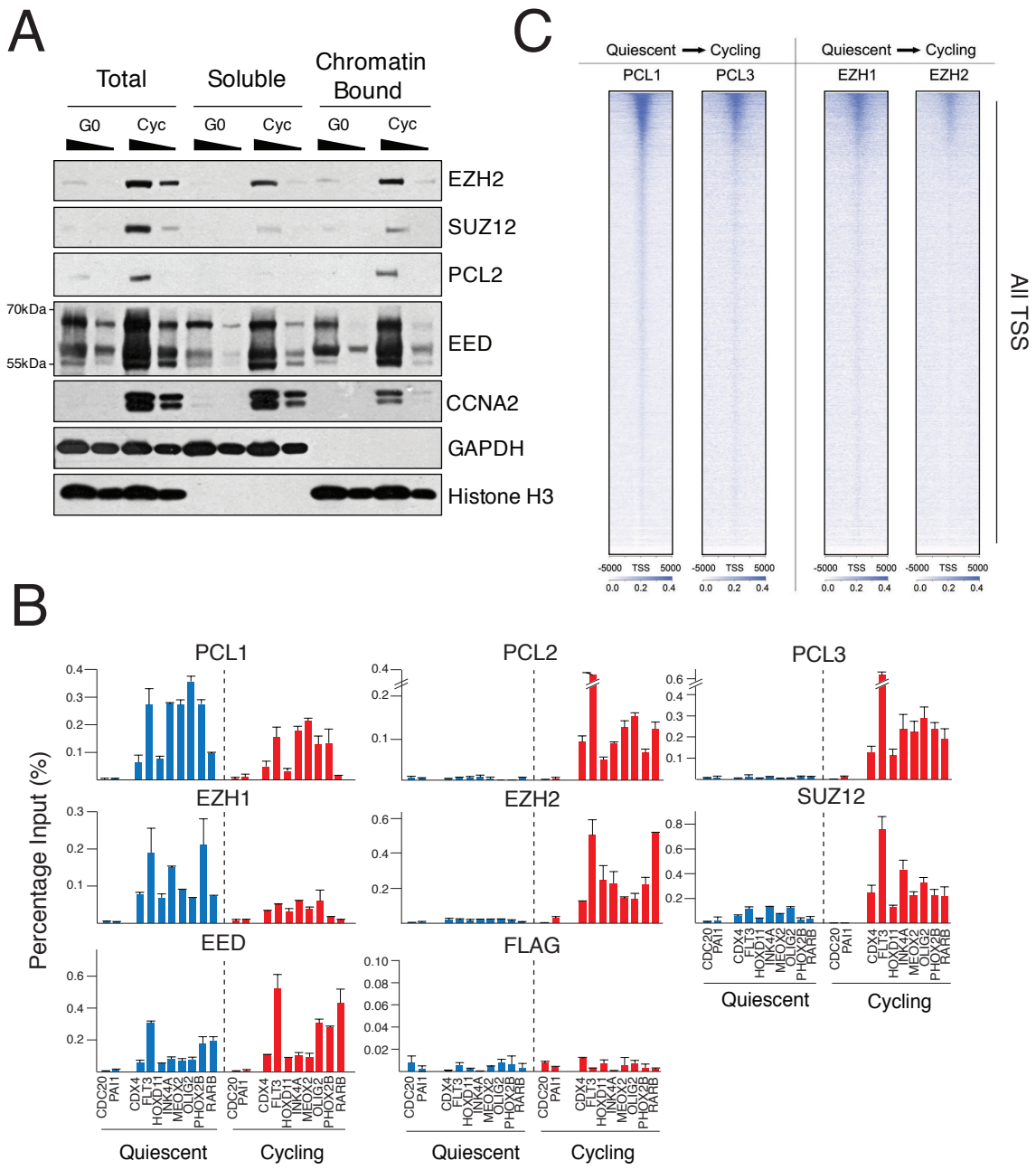


Figure 4.4 Reorganisation of PRC2 subunit chromatin enrichment in quiescence and upon re-entry into the cell cycle.

(A) Western blot analyses using the indicated antibodies on Total, Soluble and Chromatin Bound fractions prepared following serum starvation for 96-120hr (Quiescent-G0) or the same cells following treatment with 20% serum for 24hr. (Loading ratio 3:1).

(B) Chromatin immunoprecipitation (ChIP) analyses followed by next-generation sequencing using the indicated antibodies in quiescent and cycling fibroblasts, illustrating that EZH1 and PCL1 occupy the same genomic loci in quiescent cells as EZH2 and PCL3 in cycling cells, respectively.

(C) Quantitative chromatin immunoprecipitation (ChIP) analyses using the indicated antibodies in fibroblasts following 96-120hr serum starvation (Quiescent) or the same cells following treatment with 20% serum for 24hr. Quantitative PCR was performed on eluted DNA fragments for established human PRC2 target genes. ChIP enrichments are presented as the percentage of protein bound normalised to input and are a representative sample of three biological replicates.

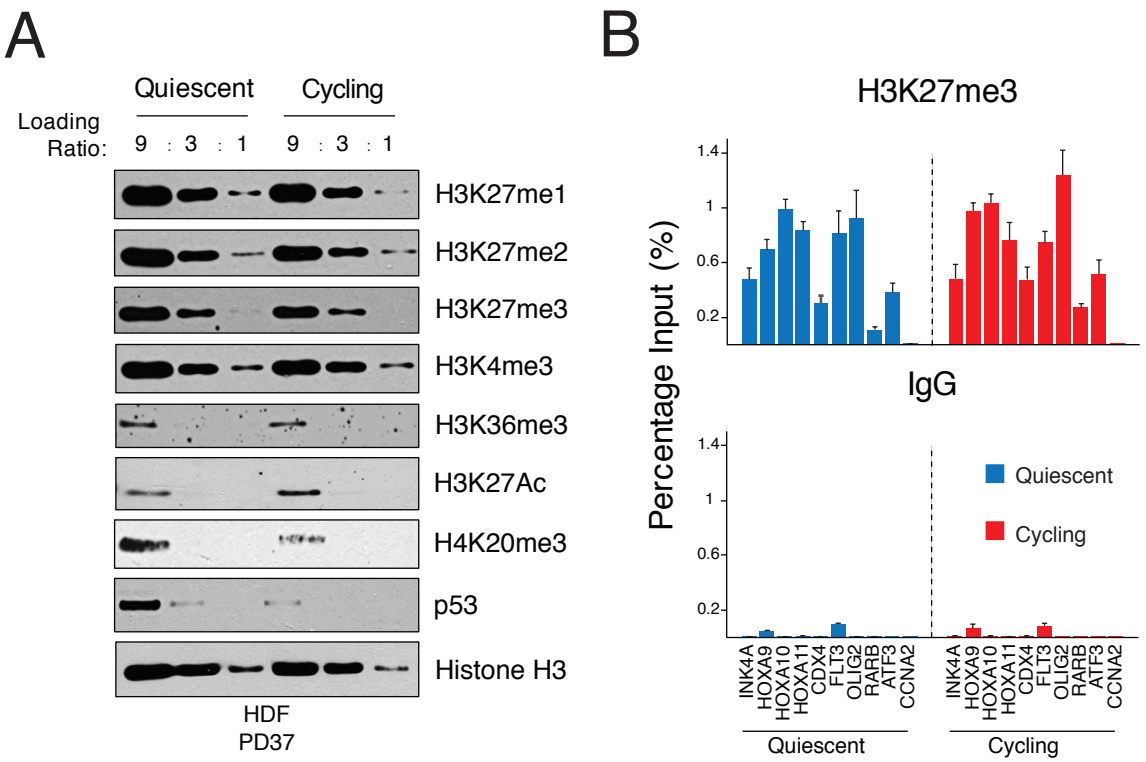


Figure 4.5 Global levels of H3K27 methylations are unchanged between quiescent and cycling fibroblasts.

(A) Semi-quantitative western blot analyses using the indicated antibodies on nuclear lysates from human diploid fibroblasts. Cells were serum starved for 120hr (Quiescent - G0) and stimulated to re-enter the cell cycle with the addition of 20% serum for 24hrs (Cycling).

(B) Chromatin immunoprecipitation (ChIP) analyses using an antibody anti-body against Histone H3 Lys-27 tri-methylation. ChIPs were performed in asyn-chronously growing cells (Cycling) or the same following serum starvation (Quiescent). Quantitative PCR was performed on eluted DNA fragments for established human PRC2 target genes. ChIP enrichments are presented as the percentage of protein bound normalised to input and are a representative sample of three biological replicates.

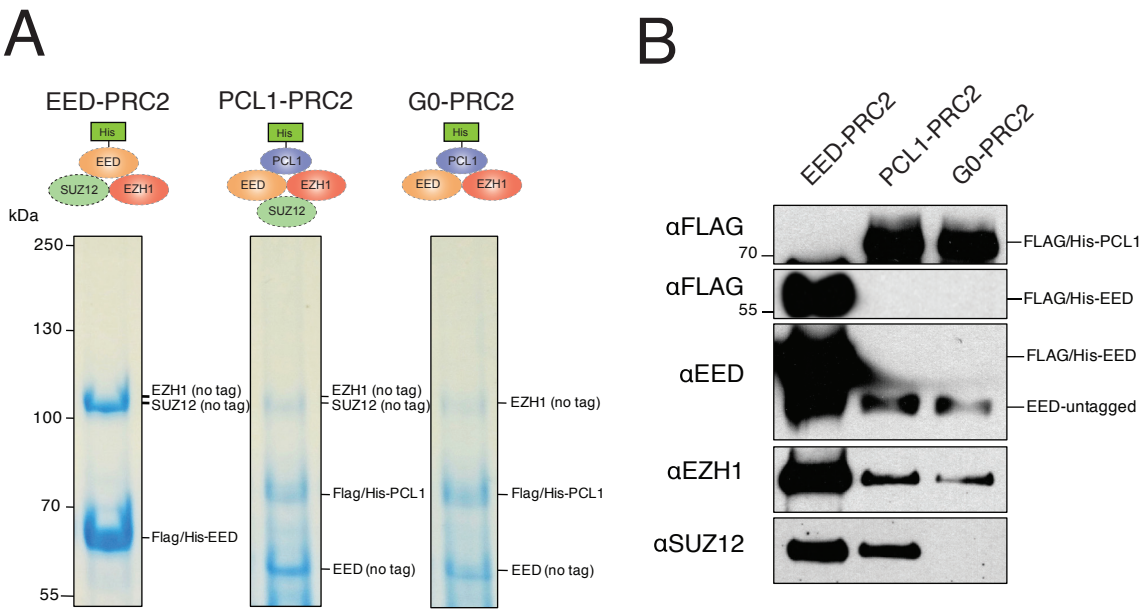


Figure 4.6 Purification of recombinant G0-PRC2 complex.
(A) Representative gel image of various Flag/His-tagged recombinant PRC2 complexes. HighFive insect cells were infected with baculovirus for either Flag/His-PCL1 or Flag/His-EED and the indicated untagged PRC2 subunit and captured using Ni2+ resin. Imidazole eluted PRC2 complexes were concentrated through 100kDa molecular weight columns, run on SDS-PAGE gels, and stained with Coomassie brilliant blue.
(B) Western blot analyses using the indicated antibodies on the various recombinant PRC2 complexes from (A).

canonical core PRC2 complex containing EZH1/EED/SUZ12, a PCL1-PRC2 complex consisting of PCL1/EZH1/EED/SUZ12, and a G0-PRC2 complex comprising of PCL1/EZH1/EED. The fact that recombinant PCL1 can exist together in a complex with EED and EZH1, without SUZ12, supports the idea that the variant G0-PRC2 complex might lack EPOP and PALI1, which require SUZ12 to incorporate into the complex (Figure 4.6B).

4.2.4 Catalytic and allosteric inhibition of PRC2 in primary human fibroblasts

Because G0-PRC2 lacks SUZ12, I was intrigued to explore the catalytic activity of this complex, since previous work has shown that SUZ12 is absolutely required for methyltransferase activity (Pasini et al., 2007). Therefore, I decided to treat quiescent and cycling cells with a catalytic inhibitor of EZH1/2 (GSK343) (McCabe et al., 2012b) and an allosteric inhibitor targeting the EED subunit (A395) (He et al., 2017). The efficacy of these small molecules has yet to be determined in primary human cells, as most of the previous experiments have been performed in highly proliferative cancer cell lines (Lindsay et al., 2017). Therefore, I first optimised the working concentrations and treatment timescale of both drugs in normally proliferating human diploid fibroblasts. Both GSK343 (EZH1/2i) and A395 (EEDi) showed a strong cellular response exhibited by an almost complete loss of H3K27me2/3, as well as derepression of the Polycomb target *INK4A* and downregulation of the proliferative marker *CCNA2* (Figure 4.7A-D). The *INK4A* tumour suppressor gene is a classic PRC2 target in fibroblasts (Bracken et al., 2007). Derepression of *INK4A* through PRC2 inhibition and subsequent onset of cellular senescence through the action of its protein product, p16, is thought to be one of the many mechanisms whereby PRC2 inhibition could work in cancer treatment (Mohammad et al., 2017).

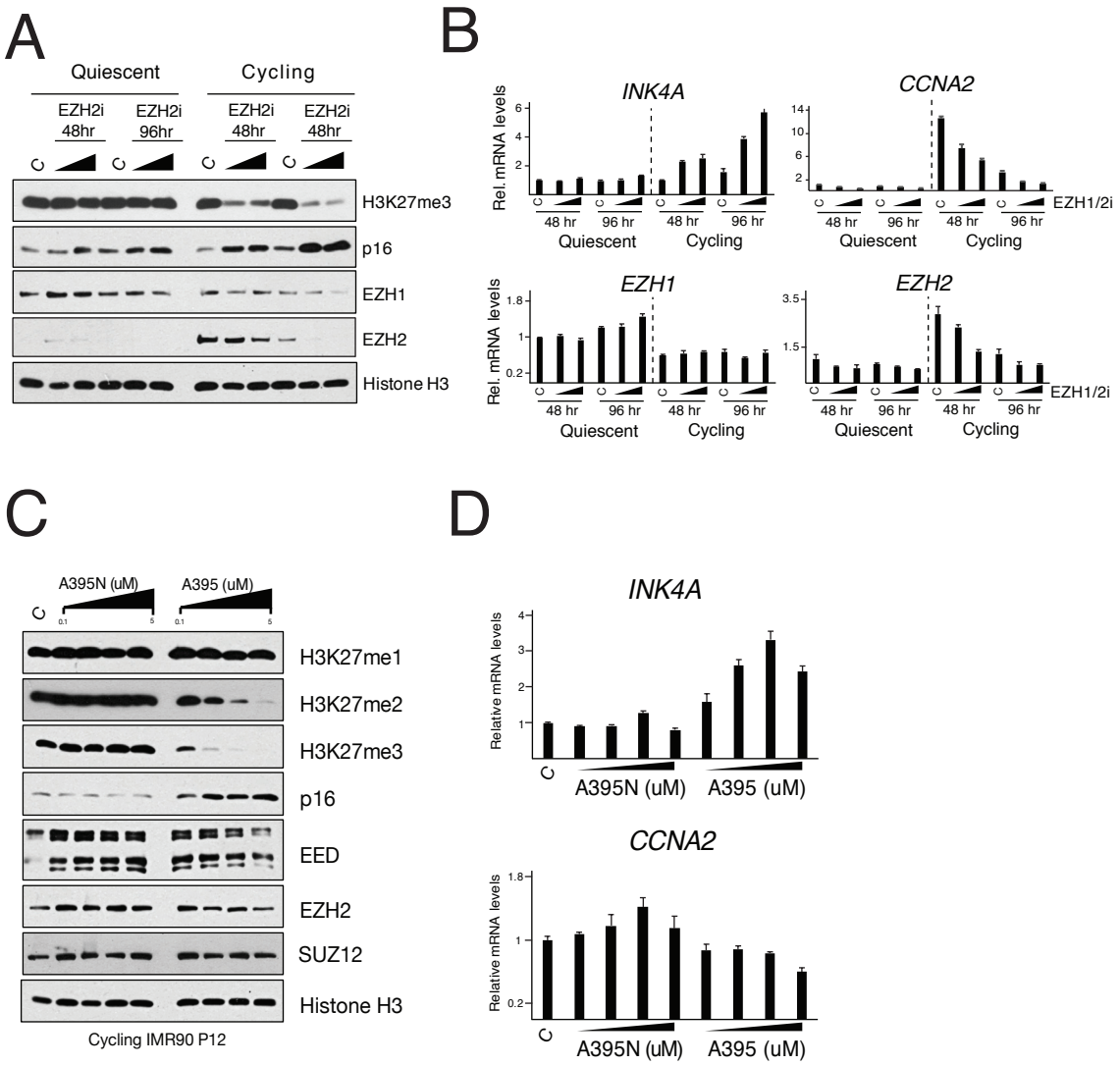


Figure 4.7 Catalytic and allosteric inhibition of PRC2 in primary human fibroblasts.

(A) Western blot analyses using the indicated antibodies on whole cell lysates from quiescent or asynchronously growing fibroblasts treated with either control (DMSO) or EZH2i (GSK343) at increasing concentrations (1-5 μM) for 48-96hr.

(B) Quantitative RT-PCR analyses of the indicated mRNA transcripts from quiescent or asynchronously growing fibroblasts treated with either control (DMSO) or EZH2i (GSK343) at increasing concentrations (1-5 μM) for 48-96hr. Data presented is a representative sample of three biological replicates and error bars indicate standard deviation of individual triplicate qPCR data.

(C) Western blot analyses using the indicated antibodies on whole cell lysates from asynchronously growing fibroblasts treated with either control (DMSO/A395N) or EEDi (A395) at increasing concentrations (0.1-5 μM) for 72hr.

(D) Quantitative RT-PCR analyses of the indicated mRNA transcripts from asynchronously growing fibroblasts treated with either control (DMSO/A395N) or EEDi (A395) at increasing concentrations (1-5 μM) for 72hr. Data presented is a representative sample of three biological replicates and error bars indicate standard deviation of individual triplicate qPCR data.

4.2.5 GO-PRC2 is catalytically inactive and therefore refractory to PRC2 inhibition

Due to the efficacy of both catalytic (EZH1/2i) and allosteric (EEDi) inhibition of PRC2 in normally proliferating primary human diploid fibroblasts, I next tested their ability to affect G0-PRC2 activity in quiescent cells. Surprisingly, neither inhibitor affected the global levels of H3K27me2/3 in quiescent cells, whereas an almost complete loss of both marks was observed in cycling cells (Figure 4.8A). Consistent with this, no increase in the expression of the p16 protein, encoded by *INK4A*, was detected in quiescent cells, whereas it was significantly activated in treated cycling cells (Figure 4.8A & B). Furthermore, *INK4A* and *CCNA2* mRNA levels were unchanged in quiescent cells treated with EZH1/2i or EEDi (Figure 4.8B). I was next interested in testing whether there were any local changes in the enrichment of H3K27me3 at Polycomb target promoters. To do this, I treated both quiescent and cycling fibroblasts with EEDi and performed ChIP for H3K27me3 followed by quantitative PCR. As expected, there was no observable decrease in H3K27me3 at the *INK4A* locus in quiescent cells, whereas there was about a 70% decrease in cycling cells (Figure 4.8C). Taken together, these data suggest that G0-PRC2 is refractory to PRC2 inhibition and indeed may have no histone methyltransferase activity.

4.2.6 PRC2i does not displace G0-PRC2 in quiescent cells

I next wished to analyse the association of G0-PRC2 with chromatin and determine whether treatment of quiescent cells with PRC2i would displace the complex from chromatin. I hypothesised that EEDi, and not EZH1/2i, would displace PRC2 components from chromatin as it targets the aromatic cage of EED, responsible for binding to H3K27me3 (Margueron et al., 2009). In order to do this, I treated both quiescent and cycling cells with EZH1/2i and EEDi and performed a subcellular fractionation followed by Western blotting of various PRC2 components. Both PRC2 inhibitors led to a significant global decrease in H3K27me2/3 levels and both SUZ12 and EED were both partially displaced from

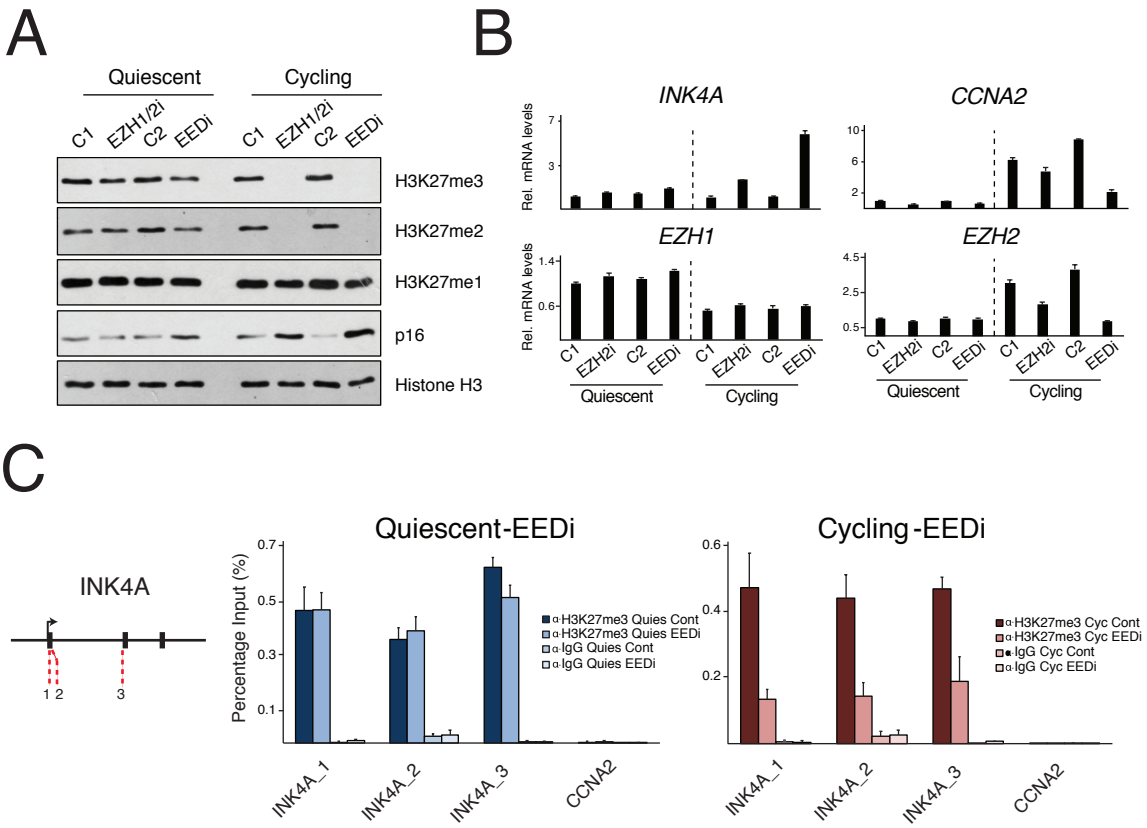


Figure 4.8 G0-PRC2 has no methyltransferase activity and is therefore refractory to PRC2 inhibition.

(A) Western blot analyses using the indicated antibodies from treated (EZH1/2i v EEDi) and control-treated (C1-DMSO, C2-A395N) quiescent and cycling fibroblasts.

(B) Quantitative RT-PCR analyses of the indicated mRNA transcripts from cells in A. Data presented is a representative sample of three biological replicates and error bars indicate standard deviation of individual triplicate qPCR data.

(C) Quantitative chromatin immunoprecipitation (ChIP) analyses from control/PRC2i treated quiescent or asynchronously growing fibroblasts using antibodies against IgG (-ve) or Histone H3 Lys-27 tri-methylation. Left/Inset - Schematic of qPCR primer locations at the INK4A gene locus. ChIP enrichments are presented as the percentage of protein bound normalised to input and are a representative sample of three biological replicates.

chromatin upon treatment with EEDi in cycling cells. The treatment of quiescent cells with both PRC2 inhibitors had no effect on H3K27me2/3 levels and did not affect the enrichment of any G0-PRC2 components on chromatin (Figure 4.9 A & B). I next performed ChIPs of SUZ12 in treated and non-treated quiescent fibroblasts, and observed weak SUZ12 enrichment at the *INK4A* locus in quiescent cells, illustrated by the levels being slightly above background, whereas it is significantly enriched at these sites in cycling cells (Figure 4.9C). This weak amount of SUZ12 observed in both chromatin fractionation and ChIP experiments could be explained by a small amount of contaminating non-quiescent cells present in the population. Nevertheless, treatment of cycling cells with EEDi partially displaced SUZ12 from the *INK4A* locus, and recapitulating the results from the cellular fractionation experiments (Figure 4.9B & C). This data again suggests that G0-PRC2 has no methyltransferase activity and might remain associated with repressive chromatin in the presence of PRC2i. These data also imply that the mechanism whereby EEDi targets PRC2 activity is mainly through allosteric inhibition and not by significantly affecting the complexes enrichment on chromatin.

4.2.7 PRC2i does not affect the ability of quiescent cells to re-enter the cell cycle

I next explored whether PRC2i, which causes senescence in cycling cells, would affect the ability of quiescent cells to re-enter the cell cycle. To do this, I designed a cell cycle re-entry assay whereby quiescent or cycling cells are treated with PRC2i for 72 hours and then stimulated to re-enter the cell cycle by the addition of serum (Figure 4.10 & 11A). Cells were harvested immediately following 72 hour treatment, or at 3 or 6 days after serum stimulation. Interestingly, the treatment of quiescent cells with EEDi did not affect their ability to re-enter the cell cycle, as demonstrated by the nearly 10-fold increase in *CCNA2* mRNA levels at Day 3 following serum addition (Figure 4.10B). This is in stark contrast to cycling cells treated with EEDi, which failed to efficiently re-enter the cycle and proliferate (Figure 4.10C). As expected, *INK4A* expression increased 2-3-fold

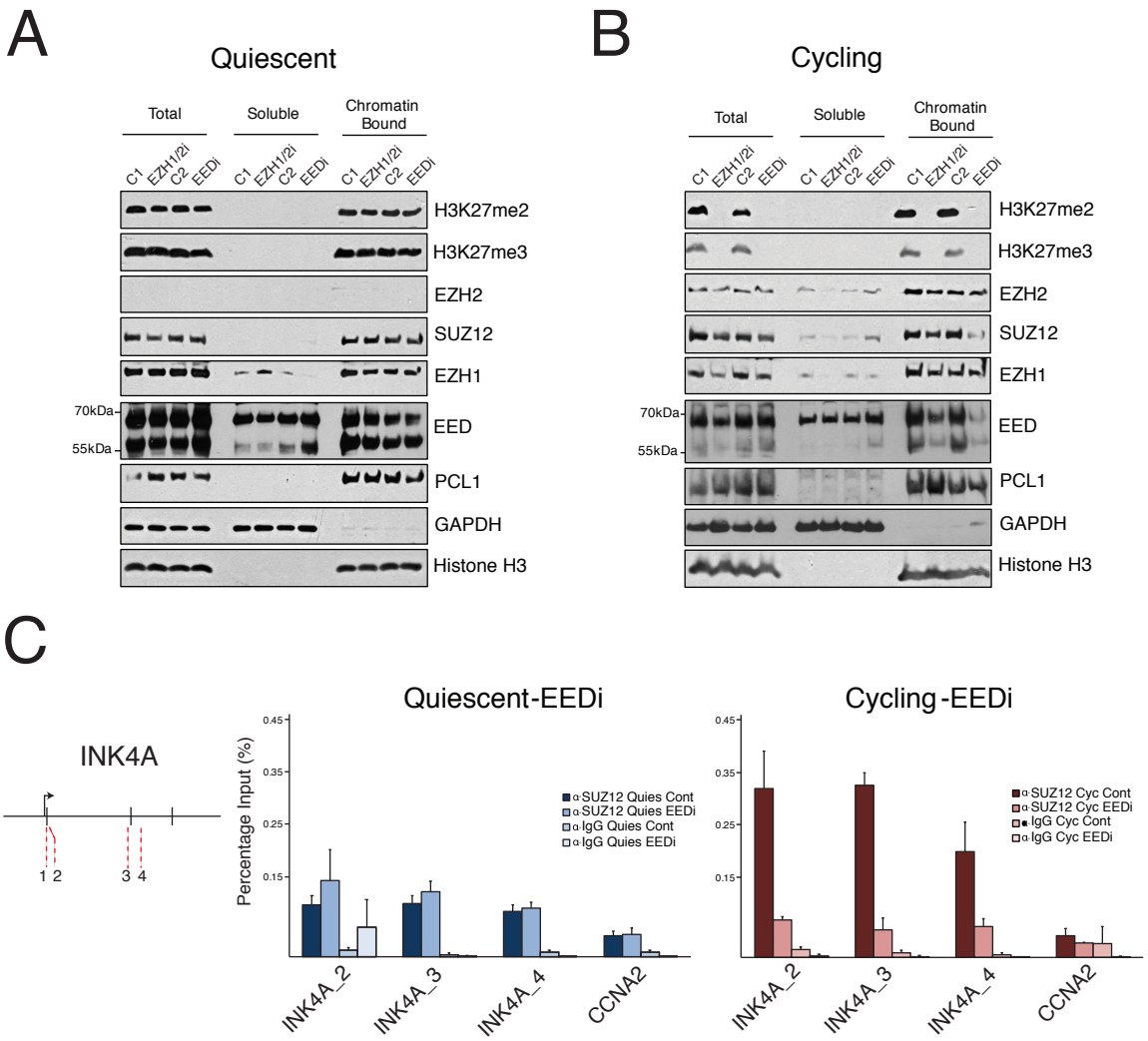


Figure 4.9 PRC2i does not displace G0-PRC2 in quiescent cells.
(A) Western blot analyses using the indicated antibodies on Total, Soluble and Chromatin Bound fractions prepared following treatment of quiescent *(A)* or asynchronously growing *(B)* fibroblasts with EZH1/2i (GSK343) or EEDi (A395) for 72 hours.
(B) See *(A).*
(C) Quantitative chromatin immunoprecipitation (ChIP) analyses from control/PRC2i treated quiescent or asynchronously growing fibroblasts using antibodies against IgG (-ve) or SUZ12. Left/Inset - Schematic of qPCR primer locations at the INK4A gene locus. ChIP enrichments are presented as the percentage of protein bound normalised to input and are a representative sample of three biological replicates.

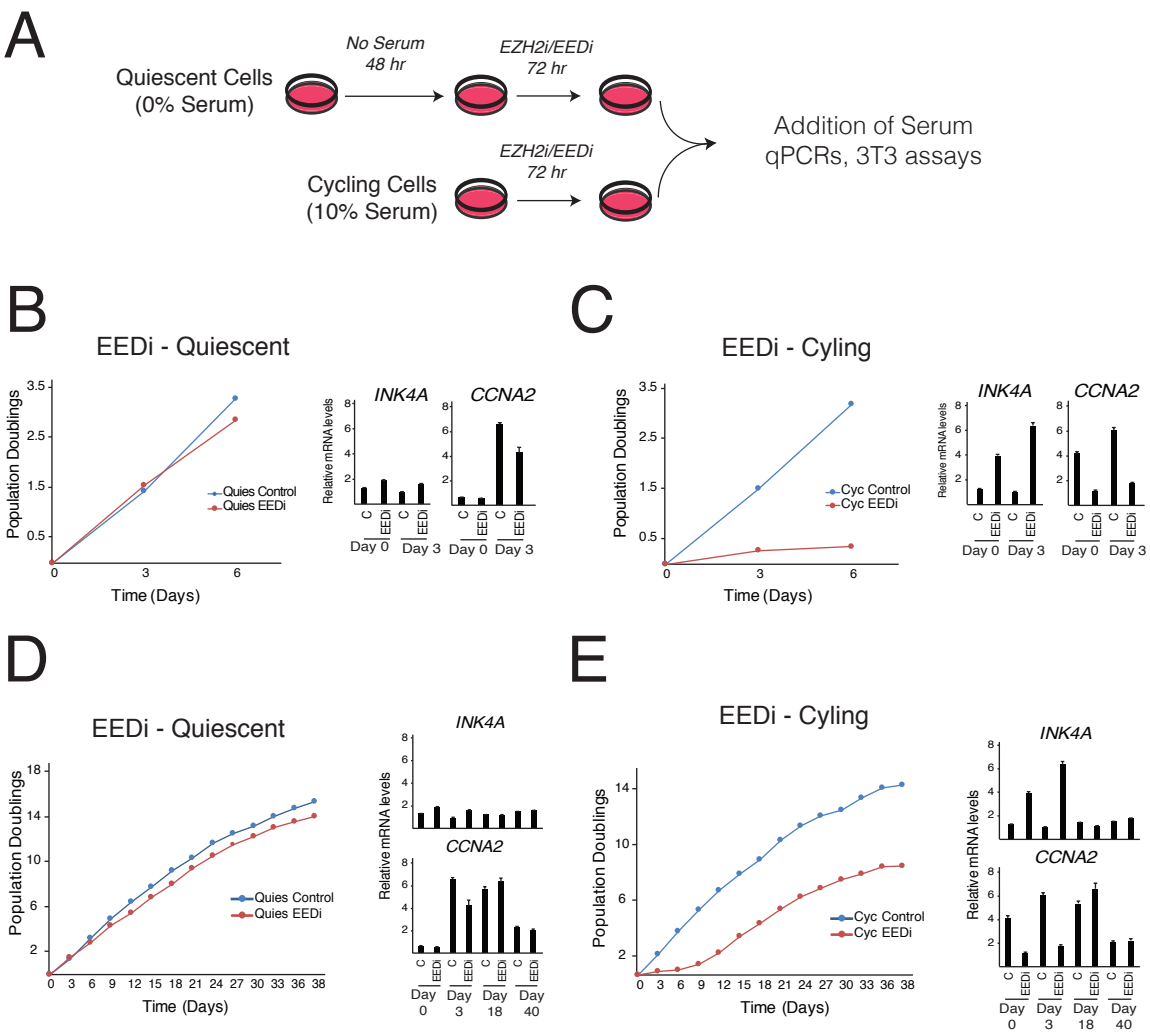


Figure 4.10 EEDi in quiescent cells does not affect their ability to re-enter the cell cycle.

(A) Schematic describing experimental work flow of treatment of cycling and quiescent fibroblasts with PRC2i.

(B) Growth assays performed on cells from treated (EEDi-A395) and control treated (A395N) quiescent fibroblasts. Assays were initiated following 72hr treatment with either EEDi or control and continued for 6 days. Right - Quantitative RT-PCR analyses of INK4A and CCNA2 mRNA transcripts from quiescent cells at indicated timepoints.

(C) Growth assays performed on cells from treated (EEDi-A395) and control treated (A395N) cycling fibroblasts. Assays were initiated following 72hr treatment with either EEDi or control and continued for 6 days. Right - Quantitative RT-PCR analyses of INK4A and CCNA2 mRNA transcripts from cycling cells at indicated timepoints.

(D) 3T3 growth assays performed on cells from treated (EEDi-A395) and control treated (A395N) quiescent fibroblasts. Assays were initiated following 72hr treatment with either EEDi or control and continued for 40 days. Right - Quantitative RT-PCR analyses of INK4A and CCNA2 mRNA transcripts from quiescent cells at indicated timepoints.

(E) 3T3 growth assays performed on cells from treated (EEDi-A395) and control treated (A395N) asynchronously growing fibroblasts. Assays were initiated following 72hr treatment with either EEDi or control and continued for 40 days. Right - Quantitative RT-PCR analyses of INK4A and CCNA2 mRNA transcripts from cycling cells at indicated timepoints.

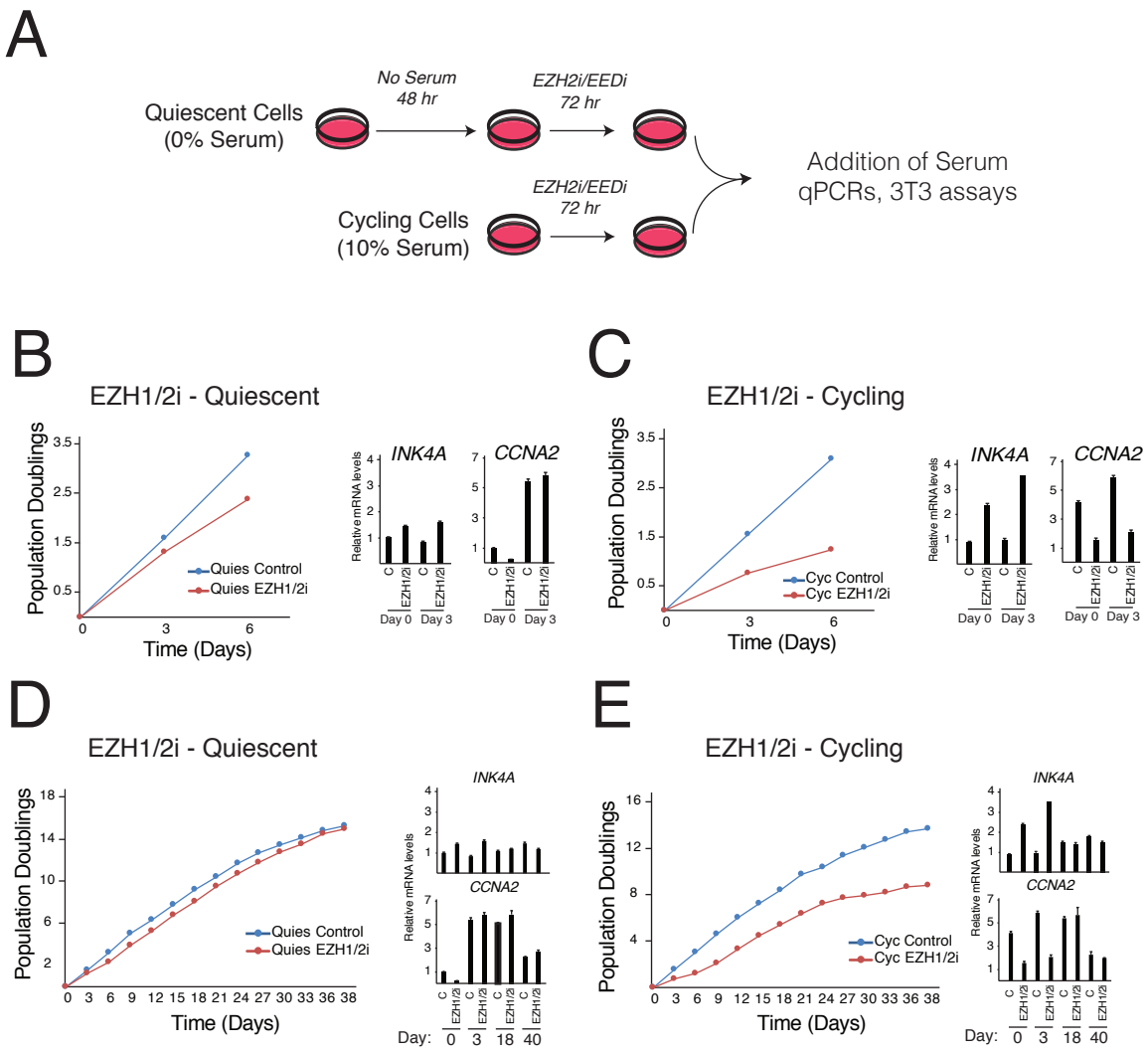


Figure 4.11 EZH1/2i in quiescent cells does not affect their ability to re-enter the cell cycle.

(A) Schematic describing experimental work flow of treatment of cycling and quiescent fibroblasts with PRC2i.

(B) Growth assays performed on cells from treated (EZH1/2i-GSK343) and control treated (DMSO) quiescent fibroblasts. Assays were initiated following 72hr treatment with either EZH1/2i or control and continued for 6 days. Right - Quantitative RT-PCR analyses of INK4A and CCNA2 mRNA transcripts from quiescent cells at indicated timepoints.

(C) Growth assays performed on cells from treated (EZH1/2i-GSK343) and control treated (DMSO) cycling fibroblasts. Assays were initiated following 72hr treatment with either EZH1/2i or control and continued for 6 days. Right - Quantitative RT-PCR analyses of INK4A and CCNA2 mRNA transcripts from cycling cells at indicated timepoints.

(D) 3T3 growth assays performed on cells from treated (EZH1/2i-GSK343) and control treated (DMSO) quiescent fibroblasts. Assays were initiated following 72hr treatment with either EZH1/2i or control and continued for 40 days. Right - Quantitative RT-PCR analyses of INK4A and CCNA2 mRNA transcripts from quiescent cells at indicated timepoints.

(E) 3T3 growth assays performed on cells from treated (EZH1/2i-GSK343) and control treated (DMSO) asynchronously growing fibroblasts. Assays were initiated following 72hr treatment with either EZH1/2i or control and continued for 40 days. Right - Quantitative RT-PCR analyses of INK4A and CCNA2 mRNA transcripts from cycling cells at indicated timepoints.

upon EEDi, accompanied by a concomitant decrease in *CCNA2* levels, indicating these cells undergo cellular senescence (Figure 4.10C). An equivalent experiment using EZH1/2i generated similar results, suggesting neither EEDi nor EZH1/2i affect the ability of quiescent cells to re-enter the cell cycle. These results suggest that G0-PRC2, the predominant PRC2 complex in quiescent cells, is refractory to PRC2i and therefore catalytically inactive.

4.2.8 PRC2i in quiescent cells does not affect long term proliferative capacity

Following the observation that PRC2i in quiescent cells does not affect their capacity to immediately re-enter the cell cycle, I next decided to test whether the long term proliferative capacity of these PRC2i treated cells was impaired. To do this, I performed a similar assay to the cell cycle re-entry experiment described in Figure 4.10. However, this time, following treatment of quiescent and cycling cells with PRC2i, I performed 3T3 growth assays over a period of 40 days. I hypothesised that by culturing the cells treated in quiescence (G0) for a longer period of time, I would potentially begin to see growth defects. However, surprisingly, cells treated with PRC2i while in quiescence proliferate normally and at the same rate as control treated cells (Figure 4.10 & 11 D &E). This is again in clear contrast with normally cycling cells treated with PRC2i. Proliferation of these cells is affected at the start of the assay due to the increase in *INK4A* expression. As a result, these cells grow at a much slower rate and senesce prematurely. Taken together, these data again suggest that inhibiting PRC2 in quiescent cells has no downstream cellular effect. Unfortunately, this leads to the possibility that quiescent cells present in any tumour or hematopoietic malignancy treated with PRC2i will not be targeted, and could lead to tumour recurrence (Figure 4.13).

4.2.9 Expression of PRC2 components in the Hematopoietic stem cell differentiation hierarchy

The *in vitro* primary human fibroblasts cell system described so far provides important insights into what might occur in quiescent cells *in vivo.* In order to analyse expression patterns of PRC2 components in the hematopoietic system, I mined pre-existing mRNA expression datasets (Seita et al., 2012). This allowed a comparison of expression for the genes encoding G0-PRC2 and canonical PRC2 components in hematopoietic stem (quiescent), progenitor (proliferative) and mature lineages (terminally differentiated, non-proliferating). Consistent with what I observed in fibroblasts, the expression pattern of *EZH1* and *EZH2/SUZ12* are inversely correlated between proliferating and non-proliferating blood cells within the hierarchy, as determined by Ki67 expression (Figure 4.12). Consistent with the fibroblasts model, *EZH1* is highly expressed in quiescent hematopoietic stem cells as well as in terminally differentiated mature lineages. In contrast, *EZH2* and *SUZ12* are expressed highest in proliferating progenitor cells. These reciprocal expression patterns are also observed for PCL1 (*PHF1*) and PCL3 (*PHF19*), with *PHF1* mirroring *EZH1* expression and *PHF19* expression similar to *EZH2/SUZ12*. These data support a model whereby G0-PRC2 may also be the predominant PRC2 complex in quiescent hematopoietic stem cells and could play a role in the maintenance of the HSC pool *in vivo*.

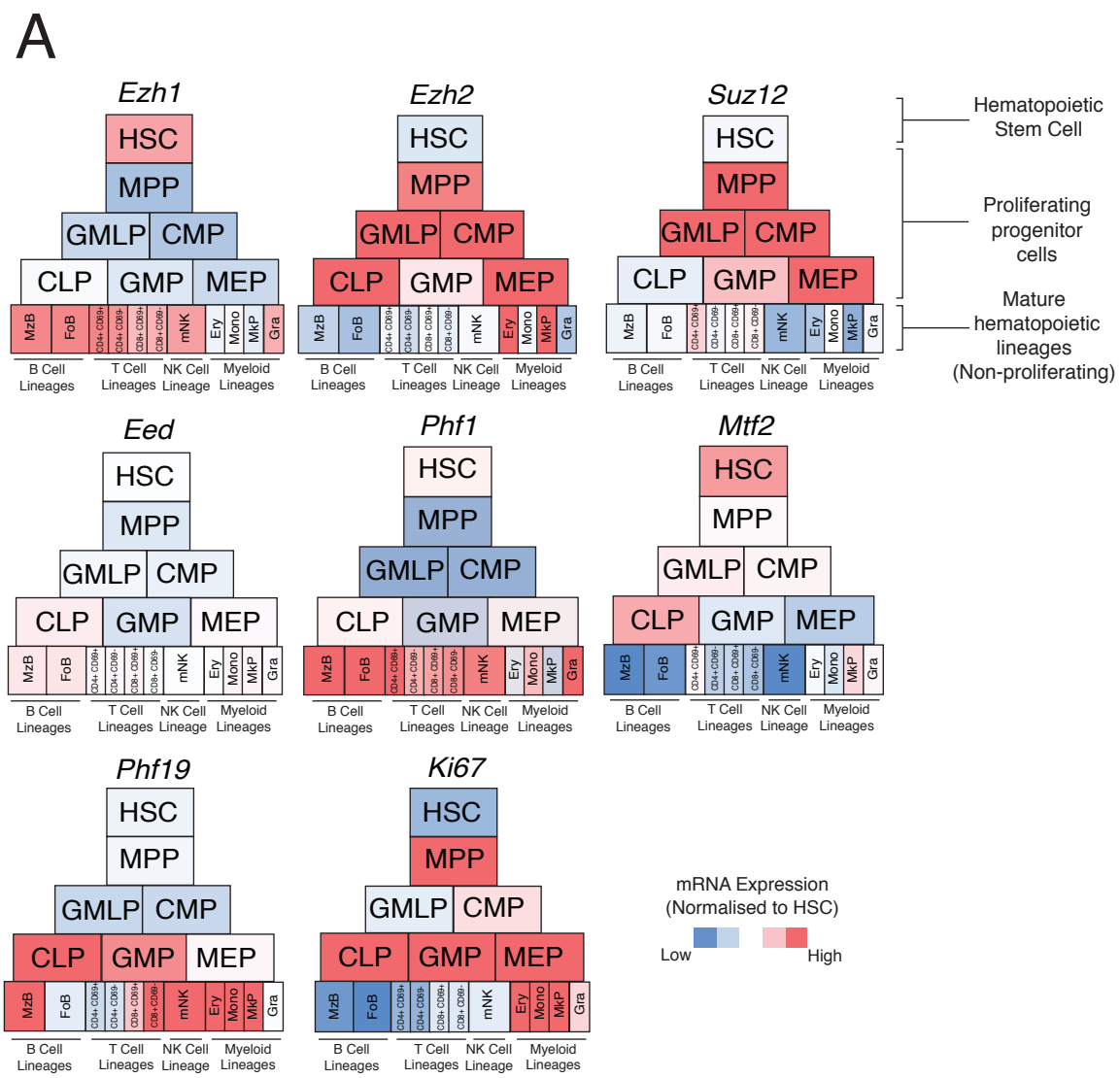


Figure 4.12 Expression of PRC2 subunit in the Hematopoietic Stem Cell (HSC) differentiation hierarchy.

(A) Expression of PRC2 components mRNA during hematopoiesis. mRNA expression values are from transcriptomic profiling using compiled microarray data (Seita et al., 2012). Relative expression values are shown in the indicated FACS-sorted hematopoeitc/progenitor/mature lineages.

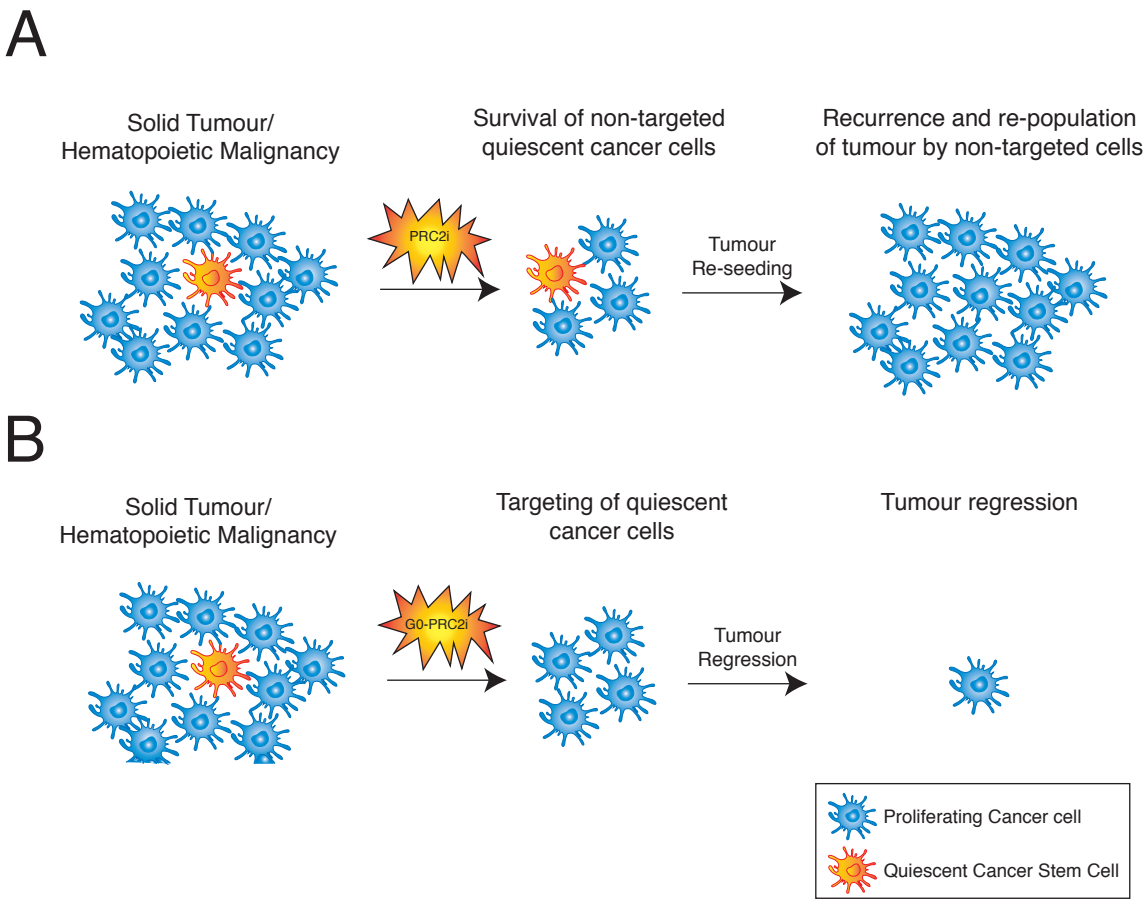


Figure 4.13 Targeting G0-PRC2 in cancer.
(A) PRC2 inhibitors target highly proliferative cancer cells in solid tumours or hematopoietic malignancies. However, PRC2i will not target quiescent cancer stem cells within these tumours and hence these cells could lead to tumour re-seeding and recurrence following initial treatment.
(B) By inhibiting G0-PRC2 through novel protein degrader therapeutics (dEED, dPCL1), this may help to target non-diving quiescent cancer stem cells, and reduce risk of tumour recurrence and re-seeding.

4.3 Discussion

The main aim of this chapter was to characterise PRC2 function in primary quiescent human fibroblasts. I was successful in defining a divergent form of PRC2 that exists exclusively in non-proliferating quiescent cells. This complex termed, G0-PRC2, contains EZH1, EED and PCL1, but intriguingly does not contain SUZ12. Importantly, previous cross-linking mass spectrometry studies support the idea that PCL proteins could incorporate in to PRC2 lacking SUZ12, via EZH1/2 (Kloet et al., 2016). Due to the fact that G0-PRC2 does not contain SUZ12, I predicted it might lack intrinsic methyltransferase activity. Consistent with this, G0-PRC2 is not sensitive to both catalytic and allosteric PRC2 inhibition, and quiescent cells treated with these inhibitors show no adverse proliferative side-effects. Given the recent emergence of PRC2 as a druggable target in multiple cancer types, these data highlight the pressing need to explore further ways to target PRC2 and have broad implications for the use of EZH2 inhibitors in cancer therapies.

4.3.1 H3K27 demethylase activity in quiescent cells

It was surprising to me that the neither the global nor local levels of H3K27me3 change between quiescent and cycling cells (Figure 4.5). This is despite the fact that quiescent cells do not express crucial core PRC2 components, EZH2 and SUZ12. I hypothesised that EZH1 might perform PRC2 mediated catalytic activities in the absence of EZH2, however, I have clearly shown that GO-PRC2 is refractory to PRC2 catalytic and allosteric inhibition, suggesting this complex lacks intrinsic methyltransferase activity. This raises the interesting question of how H3K27me3 is maintained in quiescent cells. The H3K27 demethylases, UTX and JMJD3, are known to contain highly homologous JmjC domains specifically capable of mediating the demethylation of H3K27me3 modified histones (Agger et al., 2007; Swigut and Wysocka, 2007). My data suggests that G0-PRC2 would remain associated with repressive chromatin at Polycomb targets in quiescent cells and could act to block any demethylase activity on H3K27me3. However, it

could also be that these specific demethylases are downregulated or even inactive in quiescent cells. This could potentially provide a mechanism whereby G0-PRC2 would be the sole guardian to maintain the repression of crucial Polycomb target genes during cellular quiescence. Going forward, it will be of critical importance to analyse the expression patterns of both *UTX* and *JMJD3* in quiescent and cycling cells in order to fully understand Polycomb mediated repression during quiescence.

4.3.2 Targeting G0-PRC2 in cancer

Tissues such as the hematopoietic system continually self-renew throughout the life of an organism. This self-renewal capacity is fuelled by the activity of adult stem cells that re-populate differentiated cells within the tissue. For example, tissue specific stem cells of the hematopoietic system (HSCs) are maintained in a quiescent state until activated by a stimulus to divide and differentiate (Seita and Weissman, 2010). This particular biological process is also a hallmark of cancers, whereby tumour growth is similarly fuelled by a small number of dedicated stem cells hidden in cancers, known as cancer stem cells (CSCs) (Batlle and Clevers, 2017). This concept can help to explain the almost inevitable recurrence of particular tumours following initially successful chemotherapy, and the phenomenon tumour dormancy and metastasis. CSCs are thought to exist in a non-dividing quiescent state that renders them resistant to chemotherapies designed to target highly proliferative cancer cells. Here, I have provided a proof of principle that quiescent cancer stem cells would be resistant to PRC2 inhibition *in vivo*. Considering the emergence and initial promise of EZH2 inhibitors in the clinic, the idea that there may be a population of resistant cells that could lead to tumour recurrence is worrying. A potential solution would be to develop approaches to target G0-PRC2 in combination with traditional chemotherapies. This would both reduce the bulk population of cancer cells and specifically target the non-dividing quiescent CSCs, thereby reducing the risk of recurrence (Figure 4.13). Furthermore, given the fact EZH2 inhibitor resistant mutations have

already been described (Baker et al., 2015; Gibaja et al., 2016), my data again highlights the need to explore further ways to target PRC2 in cancer.

An intriguing emerging concept in cancer drug development is the use of protein degradation strategies as opposed to chemical inhibition of a target. The degradation of the protein of interest is achieved through linking a small molecule inhibitor to a phthalimide moiety that hijacks an E3 ubiquitin ligase complex resulting in proteosomal degradation, and has already been described for BET bromodomain inhibitors (Winter et al., 2015). I hypothesise that this approach could also be used to develop a drug to target G0-PRC2. To do this, the degradation moiety could be fused to the EEDi allosteric inhibitor. Given the fact that G0-PRC2 has no intrinsic methyltransferase activity, disrupting G0-PRC2 function via degradation if EED would target both the proliferating and non-proliferating cancer cells. It will be interesting to couple this approach with detailed in-depth structural characterisation of the G0-PRC2 complex using Cryo-EM. These combinatorial strategies will hopefully provide novel therapeutic avenues that will lead to increased overall efficacy of targeting PRC2 in cancer.

4.3.3 Other *in vivo* models of cellular quiescence

I anticipate these results will also have implications in many important biological processes other than cancer. For example, in the adaptive immune system, naïve and memory T cells must be maintained in a quiescent, yet responsive state. Naïve helper T cells can survive for years in humans and, upon detection of an antigen, these cells quickly proliferate and differentiate into one of a number of effector lineages that tailors the immune response to different pathogens (Sprent and Tough, 1994). Upon resolution of infection, a proportion of these effector cells differentiate into long-lived quiescent memory T cells which can provide life-long immunity (Hammarlund et al., 2003). Epigenetic changes have a key role in controlling the distinct transcriptional profiles of memory T cells and thus, are critically important for the differentiation of naïve T cells into specific effector lineages (Weng et al., 2012). However, the epigenetic processes required to

maintain naïve and memory T cells in a quiescent state are not well understood. Interestingly, there have been several reports suggesting that EZH2 is critical for the regulation of T cell identity following activation (Tumes et al., 2013; Yang et al., 2015; Zhang et al., 2014b). However, EZH2 is not required for the maintenance of quiescent naïve and memory T cells (DuPage et al., 2015). Given the lack of EZH2 expression, it is unclear how H3K27me3 is maintained in quiescent cells. Here, I have provided some new insights toward addressing this question. I have established EZH1 displays a reciprocal expression pattern to *EZH2* and is highly expressed in quiescent cells. As part of a unique quiescent PRC2 complex (G0-PRC2), EZH1 along with PCL1 and EED maintain H3K27me3 in quiescent cells. However, this process intriguingly does not require methyltransferase activity of the complex. Consistent with a putative role for EZH1 (and G0-PRC2) in quiescent cells *in vivo*, conditional knockout of *EZH1* or *EED* in HSCs leads to exhaustion of the quiescent HSC stem cell pool (Hidalgo et al., 2012; Xie et al., 2014). Given these observations, it will be critical to explore the roles of EZH1 and G0-PRC2 in the longevity of naïve T cell quiescence *in vivo.* I believe this will provide key insights into the epigenetic mechanisms that govern T cell quiescence and how they retain the ability to proliferate and differentiate rapidly upon activation.

In relation to this, *EZH1* is curiously among the most highly upregulated genes in the embryonic diapause state of the African Turquoise Killifish (Valenzano et al., 2015) (Anne Brunet, Keystone Epigenetics and Chromatin, *Unpublished,* 2018). The turquoise killifish is found in ephemeral ponds in arid regions of Zimbabwe and Mozambique. These ponds are present for only 4-6 months of the year, and hence the killifish has developed a unique state of embryonic diapause to survive the dry season (Hu and Brunet, 2018). During this state, the embryos do not proliferate and are thought to be maintained in a quiescent like state of “suspended animation” (Hu and Brunet, 2018). This is another potential example of how epigenetic regulators can govern unique and diverse quiescent cell states. It will be very interesting to understand exactly how EZH1 and G0-PRC2

contribute to this fascinating biological process of embryonic diapause in the African Turquoise Killifish.

Chapter 5

Variant PRC2 complexes in the localisation of H3K27 methylations in ESCs

5.1 Introduction

An intriguing paradigm emerging in the Polycomb field is that PRC2 can be classified into two independent subtype assemblies (Figure 1.4). Comprehensive proteomic analyses have revealed that PRC2 assembles into two mutually exclusive subcomplexes, termed PRC2.1 and PRC2.2, which are defined by specific combinations of accessory proteins (Alekseyenko et al., 2014; Grijzenhout et al., 2016; Hauri et al., 2016). The PRC2.1 complex contains one of three PCL proteins, as well as EPOP or PALI1/2, while the PRC2.2 complex contains JARID2 and AEBP2 (Conway et al., 2018; Holoch and Margueron, 2017). Although relatively little is known about why different forms of PRC2 exist, or of any potential divergent functions, all of the PRC2 accessory components have well established roles in the regulation of either histone methyltransferase activity (Cao and Zhang, 2004; Conway et al., 2018; Sarma et al., 2008; Son et al., 2013; Zhang et al., 2011) or facilitating interactions with chromatin or other transcriptional regulators.

In the PRC2.1 complex, PCL proteins promote PRC2.1 association with chromatin via their conserved Tudor domain, which binds to H3K36me2/3 and H3K27me3 (Ballare et al., 2012; Brien et al., 2012; Musselman et al., 2012a; Sarma et al., 2008), and winged helix (WH) domain which binds to DNA (Choi et al., 2017; Li et al., 2017; Perino et al., 2018). However, PCL proteins are only one of several PRC2.1 accessory proteins. The PRC2.1 subunit, EPOP, links PRC2 with Elongin B/C proteins, yet, is not required for the complexes association with chromatin (Beringer et al., 2016; Liefke et al., 2016). Similarly, a conserved domain in the recently discovered PALI1/2 proteins provides a novel link between PRC2 and transcriptional co-repressors (Conway et al., 2018).

In the PRC2.2 complex, its two specific subunits, AEBP2 and JARID2, are widely known to promote PRC2 *in vitro* methyltransferase activity (Lee et al., 2018; Li et al., 2010; Son et al., 2013). While, the distinct roles of these proteins for PRC2.2 function and recruitment *in vivo* have yet to be fully established both have potential chromatin and DNA binding abilities. For example, AEBP2 has been

reported to bind methylated CpG DNA, however, there is a lack of *in vivo* evidence to support these claims (Kim et al., 2009; Wang et al., 2017). On the other hand, JARID2 has more well defined roles in chromatin binding, and has recently been intriguingly linked with targeting PRC2 to sites of PRC1 mediated H2AK119ub via its Ubiquitin interaction motif (UIM) (Cooper et al., 2016). Consistent with this, PRC2 *in vitro* activity is enhanced in the presence of H2AK119ub modified nucleosomes (Kalb et al., 2014). While these data all represent key findings, from a broader perspective many critical questions remain, including why there are two forms of PRC2 and if they have divergent or overlapping functions in regulating Polycomb activity *in vivo.* Towards unravelling these elusive mechanisms, here I decided to investigate the functions of PRC2.1 and PRC2.2 in ESCs and during differentiation.

Recent studies focusing on ESCs lacking various PRC2 components have provided some clues as to potential divergent roles of PRC2.1 and PRC2.2 subcomplexes. (Beringer et al., 2016; Grijzenhout et al., 2016). Intriguingly, while the loss of EPOP or AEBP2 in ESCs results in an increase of H3K27me3 at Polycomb target genes, the loss of PCL, PALI1/2 or JARID2 correlates with a depletion of H3K27me3 from chromatin (Ballare et al., 2012; Brien et al., 2012; Conway et al., 2018; Landeira et al., 2010; Pasini et al., 2010). These results are striking and inconsistent with the data suggesting recombinant AEBP2 and EPOP enhance *in vitro* PRC2 methyltransferase activity (Cao and Zhang, 2004; Zhang et al., 2011). These apparently contradictory findings serve to highlight the need for a more comprehensive *in vivo* characterisation of PRC2.1 and PRC2.2, and emphasise how little is known about their respective functions.

The number of PRC2 accessory proteins and their respective functional domains represent a large degree of complexity in terms of how they might influence PRC2 recruitment and activity. The WH DNA binding domain of PCL1-3 represents one of these PRC2.1/2.2 functional domains, and has recently been characterised by several noteworthy studies (Choi et al., 2017; Li et al., 2017; Perino et al., 2018). While, there are contrasting data about the sequence

specificity of this domain for DNA there is no doubt that it contributes to the association of PRC2 with chromatin. These studies focused mainly on MTF2 (PCL2) in ESCs, however, we and others have previously shown that PHF19 (PCL3) is indeed expressed and contributes to PRC2 activity in these cells (Ballare et al., 2012; Brien et al., 2012). Hence, the main aim of this research chapter is to investigate the effect of loss of all Polycomb-like genes on PRC2 function in ESCs, with a particular focus on the respective functions of PRC2.1 and PRC2.2, and the effects on chromatin and transcriptional landscapes in the regulation of cellular identity.

5.2 Results

5.2.1 The genome-wide binding profiles of PRC2.1 and PRC2.2 in mouse ESCs.

Given the recent emerging paradigm that PRC2 is classified into unique subcomplex assemblies (Figure 5.1) (Holoch and Margueron, 2017), I decided to determine their binding profiles in ESCs. To do this, I performed genome-wide quantitative ChIP-seq with exogenous *Homo Sapiens* reference genome spike in (ChIP-Rx) of PRC2.1 specific components, PCL2 (MTF2) and EPOP, as well as for PRC2.2 specific subunits, JARID2 and AEBP2 in wild-type ESCs. I also included core PRC2 protein, SUZ12, in my analysis. It was not possible to incorporate newly characterised PRC2.1 components, PALI1/2 in this analysis due to lack of ChIP grade antibodies. Nonetheless, my analysis revealed intriguing results as to the genomic localisation of both complexes. While, PRC2.1 and PRC2.2 share many genomic targets, a smaller cohort of genomic loci exist that are only enriched for PRC2.1 specific components (Figure 5.1B & C) (n=187). I did not find any convincing PRC2.2 only bound sites. The shared genomic targets of both subcomplexes include many classical PcG genes in ESCs and are highly enriched for H3K27me3. The PRC2.1 exclusive targets on the other hand seem to be weaker PcG targets as indicated by the relative enrichment of SUZ12 and H3K27me3 (Figure 5.1C). Interestingly however, the PRC2.1 unique genomic loci are found to overlap with CpG islands, consistent with the unique DNA and CpG binding ability of PCL1-3 (Choi et al., 2017; Li et al., 2017; Perino et al., 2018). The convergence of CpG islands and SUZ12/H3K27me3 enrichment at these distinct loci suggest they are indeed *bona fide* PRC2.1 targets. Perhaps these genes exist in a bivalent state allowing low level expression and are marked by PCL1-3 and PRC2.1 to facilitate PRC2.2 recruitment and further enrichment of H3K27me3 during particular lineage transitions. Nevertheless, this is the first analyses to indicate that PRC2.1 and PRC2.2 exhibit unique, non-overlapping genome-wide localisation patterns and may point to divergent functions for both subcomplexes in ESCs.

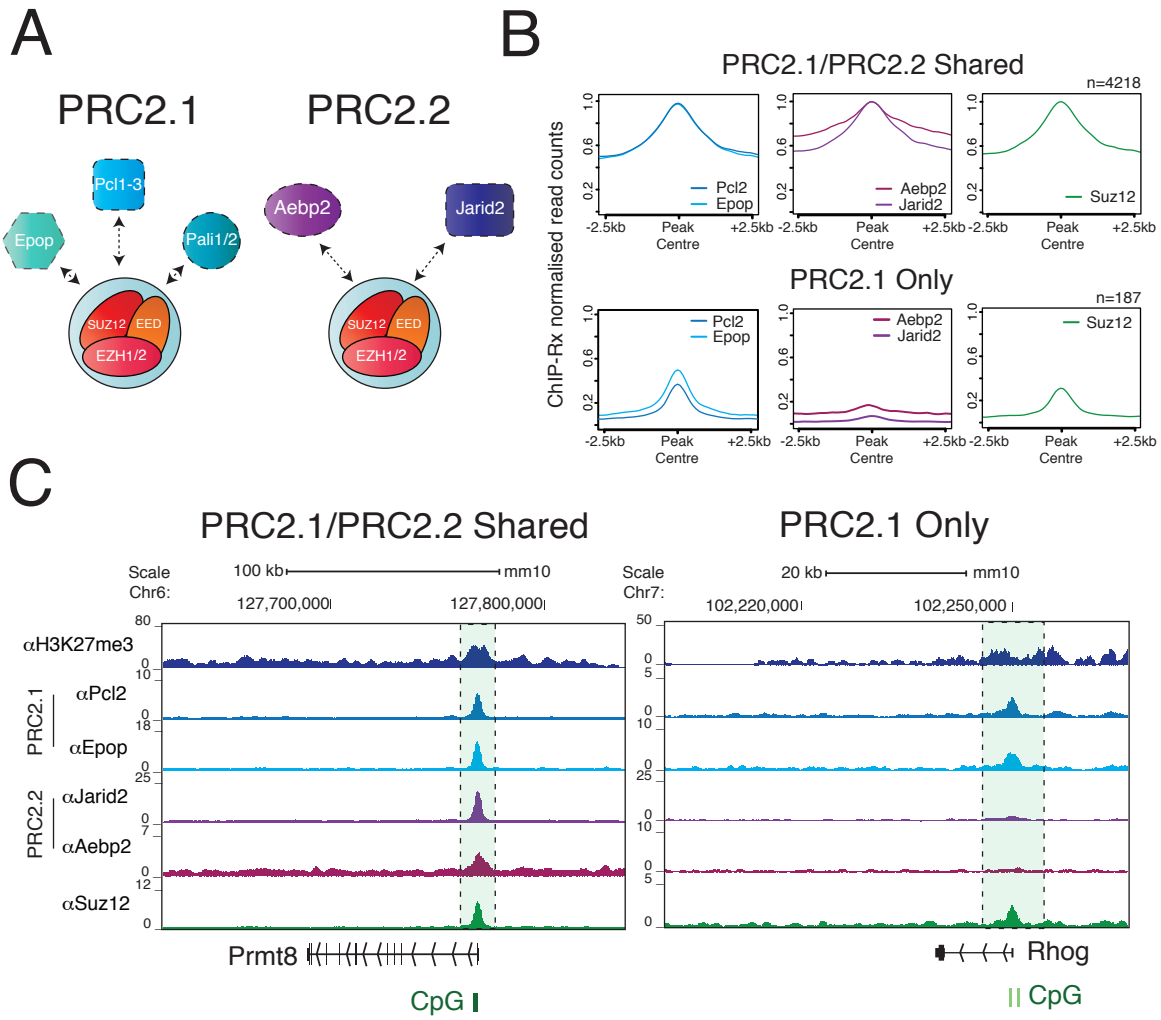


Figure 5.1 The genome-wide binding profiles of PRC2.1 and PRC2.2 in mouse ESCs.

(A) Schematic representing PRC2.1 and PRC2.2 composition and catalytic functions.

(B) Average ChIP-Rx signal profiles for PRC2.1 and PRC2.2 components at PRC2.1/PRC2.2 shared regions (n=4,218) and PRC2.1 only genomic regions (n=187) in wild-type ESCs. No reliable PRC2.2 only regions were found in our analysis.

(C) Genome browser representations of ChIP-Rx normalised reads for H3K27me3, Pcl2, Epop, Jarid2, Aebp2 and Suz12 at representative PRC2.1/PRC2.2 shared and PRC2.1 only genomic sites in wild-type ESCs. Green box highlights region around the TSS and CpG island.

5.2.2 Inducible knockout of *Pcl1* and *Pcl3* in Mtf2^{GT} ESCs

To determine the relative subunit composition of PRC2, I performed endogenous EZH2 immunoprecipitations coupled with mass spectrometry on nuclear lysates from wild-type ESCs. As expected, core PRC2 subunits were specifically pulled down (SUZ12, EED, EZH2 – bait), and importantly accessory components from both PRC2.1 and PRC2.2 were significantly enriched (Figure 5.2A). Unique peptides for PCL1 and PCL3 were found to be highly enriched in this analysis, suggesting that like PCL2, PCL1/3 also have PRC2 associated roles in wild-type ESCs. This analysis provided a clear rationale to knockout out all three *Pcl* genes in ESCs, to avoid any possibility of redundant activities. To further explore the potential divergent roles of PRC2.1 and PRC2.2, I aimed to genetically disrupt all three *Pcl* genes in ESCs. To achieve this, conditional mouse ESCs in which specific exons of the *Pcl1* and *Pcl3* genes were flanked by LoxP sites were generated by targeting each allele for recombination with a homologous transgene with additional LoxP sites, FRT sites and antibiotic resistance as shown (Figure 5.2B). These *Pcl1/3* conditional knockout cells were developed in a pre-existing cell line in which *Pcl2* (*Mtf2*) was genetically ablated through the insertion of a gene trap cassette at the third exon (Figure 5.2B) (Li et al., 2011). The addition of 4-Hydroxytamoxifen (4-OHT) to these cells induced the expression of an additional *Cre* transgene and led to recombination between these LoxP sites and knockout of specific *Pcl1/3* genomic exons (Figure 5.2C & E, 5.3A), hereafter referred to as Pcl1-3^{tKO}. The addition of DMSO to this conditional cell line resulted in a *Pcl2*/*Mtf2* only knockout cell line referred to as, Mtf2^{GT} (Figure 5.2C). Following conditional knockout and serial passaging of Pcl1-3^{tKO} cells in either LIF/2i or serum/LIF conditions there were no observable changes in overall ES cell morphology (Figure 5.2D). This gene targeting was conducted in collaboration with Haruhiko Koseki and his research group (RIKEN Institute, Japan).

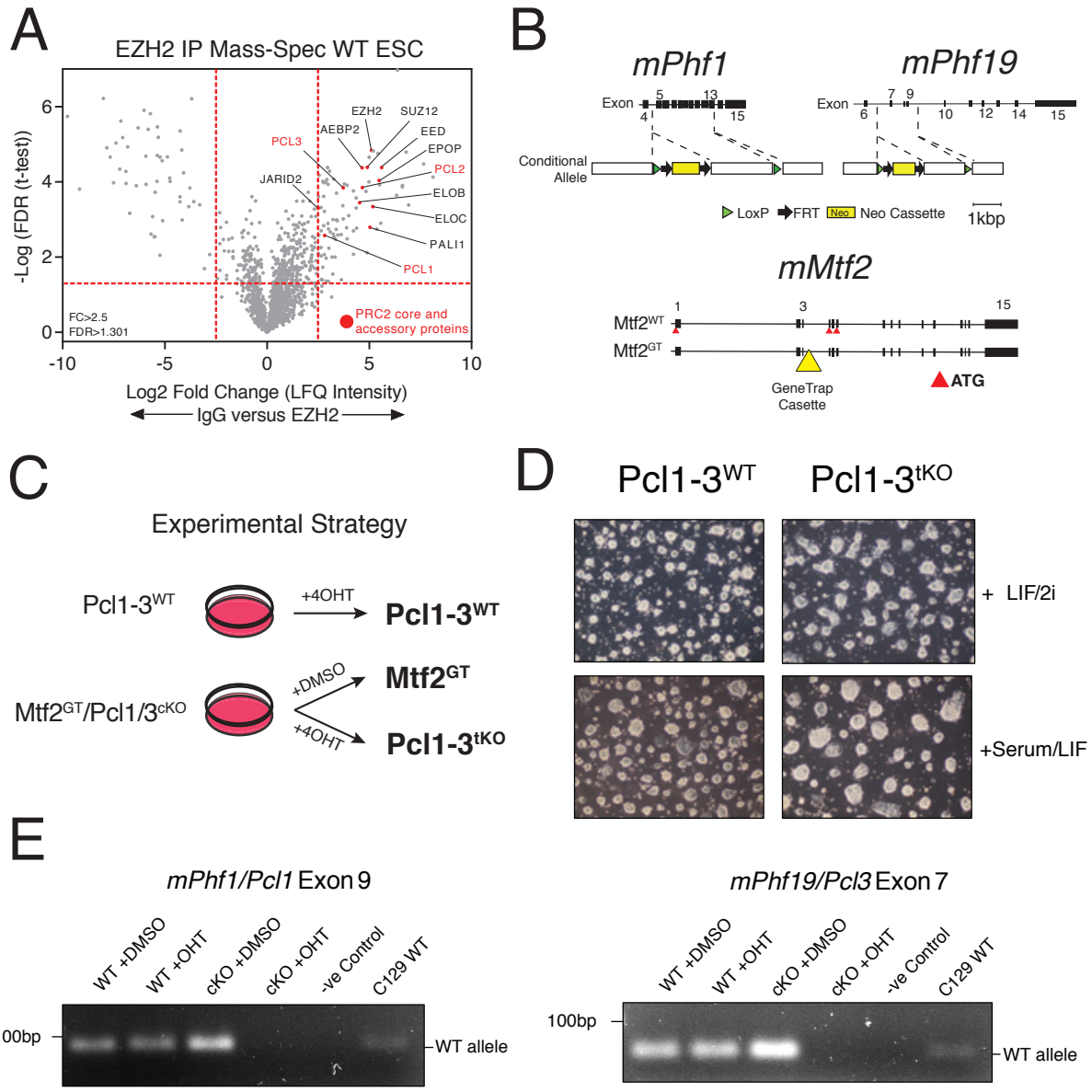


Figure 5.2 Inducible knockout of *Pcl1/Pcl3* genomic loci in Mtf2^{GT} ESCs.
(A) Identification of the Ezh2 interacting proteins in wild-type ESCs. LC-MS/MS analysis was performed using permutation-based false discovery rate (FDR)-corrected test. The label free quantification (LFQ) intensity of the bait (Ezh2) over IgG control is plotted against the -Log10 (p value).
(B) Schematic representing the *Pcl1/3* genomic targeting and knockout strategies, as well as the genetrap targeting strategy against *Pcl2 (Mtf2)*. Adapted from Li et al., 2011.Targeting vector was inserted between exon 5 and 13 of *Pcl1* and exon 7 and 9 of *Pcl3*. Conditional knockout cell lines were generated by the lab of Haruhiko Koseki, RIKEN Center for Intergrative Sciences.
(C) Conditional knockout of Pcl1/Pcl3 was induced by adding 0.5uM 4-OHT for 72 hours, resulting in three separate cell lines – Pcl1-3^{WT}, Mtf2^{GT} and Pc1-3^{tKO}.
(D) Representative images of Pcl1-3^{WT} and Pcl1-3^{tKO} ESCs cultured in both Serum/Lif and Lif/2i conditions.
(E) Representative agarose gel images confirming conditional knockout of targeted exons of both Pcl1 (Phf1) and Pcl3 (Phf19).

5.2.3 Loss of Pcl1-3 disrupts PRC2.1 in ESCs

Previous studies established that knockout out of genes encoding core PRC2 members leads to a destabilisation of PRC2 associated subunits (Hojfeldt et al., 2018; Pasini et al., 2004). To explore the effects of *Pcl1-3* knockout on core and auxiliary PRC2 subunit stability I performed western blot analysis on nuclear lysates from Pcl1-3^{WT}, Mtf2^{GT} and Pcl1-3^{tKO} ESCs. This analysis revealed, that while the protein levels of PCL2 are completely disrupted there are no significant changes in the stability of core PRC2 subunits, EZH2, EED or SUZ12 (Figure 5.3A). Interestingly, while there are no global changes in PRC2.2 auxiliary members (JARID2 & AEBP2) there is an overall decrease in the protein levels of PRC2.1 components, EPOP (Figure 5.3A) and PALI1(Figure 5.3C). It has been recently reported that PALI1 (PRC2.1) and AEBP2 (PRC2.2) have an antagonistic relationship and can define mutually exclusive PRC2 subtypes (Conway et al., 2018). To investigate changes in PRC2 complex subunit composition following loss of PCL1-3, I performed endogenous EZH2 co-immunoprecipitations (Co-IP) in Pcl1-3^{WT}, Mtf2^{GT} and Pcl1-3^{tKO} ESCs. Western blots of PRC2 components on the eluted complexes, revealed that while there was less EPOP (PRC2.1) associated with PRC2, there was a concomitant increase of AEBP2 (PRC2.2) in PRC2 following the loss of PCL1-3 (Figure 5.3B). This confirms that similar to loss of other PRC2.1 components the loss of PCL1-3 leads to an imbalance in PRC2 subcomplex composition (Figure 5.3D).

Next, I was interested in exploring the chromatin association of PRC2.1 and PRC2.2 in Pcl1-3^{tKO} ESCs. To test this, I performed western blots for PRC2.1 and PRC2.2 components on total, soluble and chromatin bound fractions from Mtf2^{GT}, Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs. Intriguingly, while PRC2.1 components EPOP and PALI1 were almost completely lost from chromatin, the levels of PRC2.2 subunit, AEBP2 in the chromatin bound fraction remain unchanged (Figure 5.3C). A slight reduction of core PRC2 complex members, EZH2 and SUZ12, from chromatin was consistent with the fact that the PRC2.2 complex remained bound. Taken together, these results suggest that PCL1-3 are

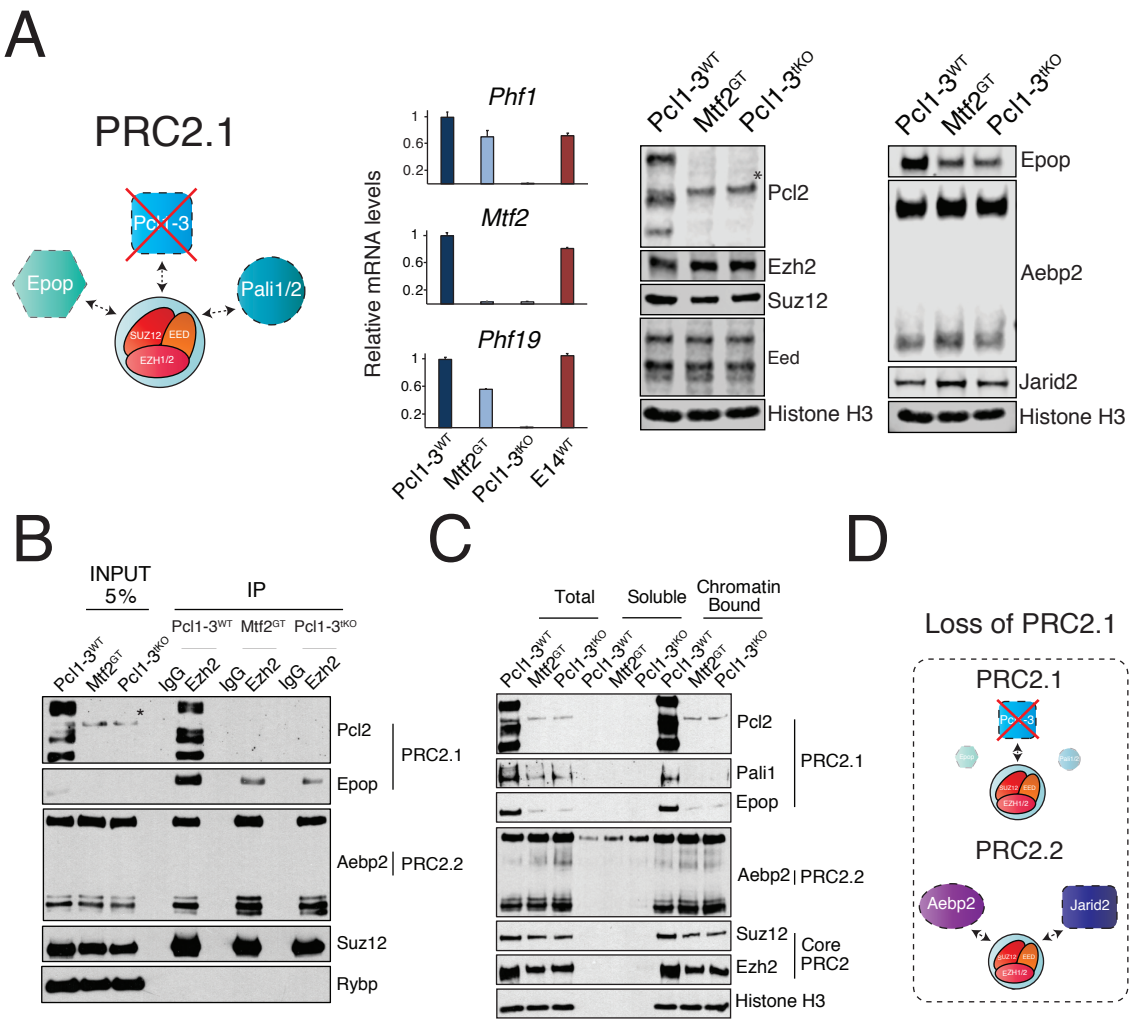


Figure 5.3 Loss of Pcl1-3 disrupts PRC2.1.
(A) Left - Schematic of PRC2.1 illustrating knockout of Pcl1-3. Middle - Quantitative RT-PCR analyses of Phf1, Mtf2 and Phf19 mRNA transcripts from Mtf2GT, Pcl1-3tKO and matched Pcl1-3WT ESCs. Right – Western blot analyses using the indicated antibodies against PRC2 core and auxiliary components on nuclear lysates from Mtf2GT, Pcl1-3tKO and matched Pcl1-3WT ESCs.
(B) Endogenous IPs of Ezh2 in Pcl1-3tKO and Mtf2GT compared to matched wild-type controls, followed by western blot with the indicated antibodies.
(C) Western blot analyses using the indicated antibodies on Total, Soluble and Chromatin Bound fractions prepared from Mtf2GT, Pcl1-3tKO and matched Pcl1-3WT ESCs; indicating that while PRC2.1 components are lost from chromatin PRC2.2 chromatin enrichment is maintained
(D) Schematic representing shift in PRC2 subcomplex balance in the absence of PRC2.1.

the defining components of the PRC2.1 subcomplex and imply that loss of these critical proteins results in overall loss of PRC2.1 association with chromatin.

5.2.4 Loss of PRC2.1 leads to a redistribution of H3K27 methylations in ESCs

I next wished to examine the effects of loss of PRC2.1, would have on PRC2 mediated histone PTMs H3K27me3, H3K27me2 and H3K27me1, as well as the functionally related modifications H2AK119ub, H3K36me3, H3K36me2, H3K27Ac and H3K4me3. Surprisingly, the complete loss of *Pcl1-3* only resulted in a relatively minor decrease in global levels of H3K27me3 (~30%), with no effect on H3K27me1/2 (Figure 5.4A). Similarly, there was no overall change in the other associated histone PTMs tested (Figure 5.4A). Previous independent studies have reported that shRNA mediated knockdown of *Pcl2* or *Pcl3* in ESCs led to global reductions of H3K27me3 (Ballare et al., 2012; Brien et al., 2012; Casanova et al., 2011) via western blot. However, that is in conflict with the histone PTM western blot data presented here and recently published report of a CRISPR-Cas9 knockout of PCL2 (Li et al., 2017). It is quite striking that knockdown and knockout studies of the same genes yield such diverse results, and it suggests that in the future, RNAi knockdown experiments should be accompanied by CRISPR-Cas9 or Cre/LoxP mediated genetic knockouts.

Two recent studies performed genome-wide ChIP-Seq for H3K27me3 in ESCs lacking PCL2 (MTF2), but present conflicting observations in relation to its localisation and relative abundance (Li et al., 2017; Perino et al., 2018). In order to improve upon these studies, I have performed quantitative genome-wide ChIP-Rx (*Drosophila* reference genome spike in) of H3K27me1/2/3 in both Pcl1-3^{tKO} and matched wild-type ESCs. This analysis revealed significant redistribution of all H3K27 methylations in cells lacking PCL1-3 (Figure 5.4B-D). Average plots profiling all H3K27me3 peaks indicate that there is ~70% reduction of H3K27me3, and a resulting concomitant increase in H3K27me1 and H3K27me2 at these sites (Figure 5.4C & D). This is a stronger reduction compared to the

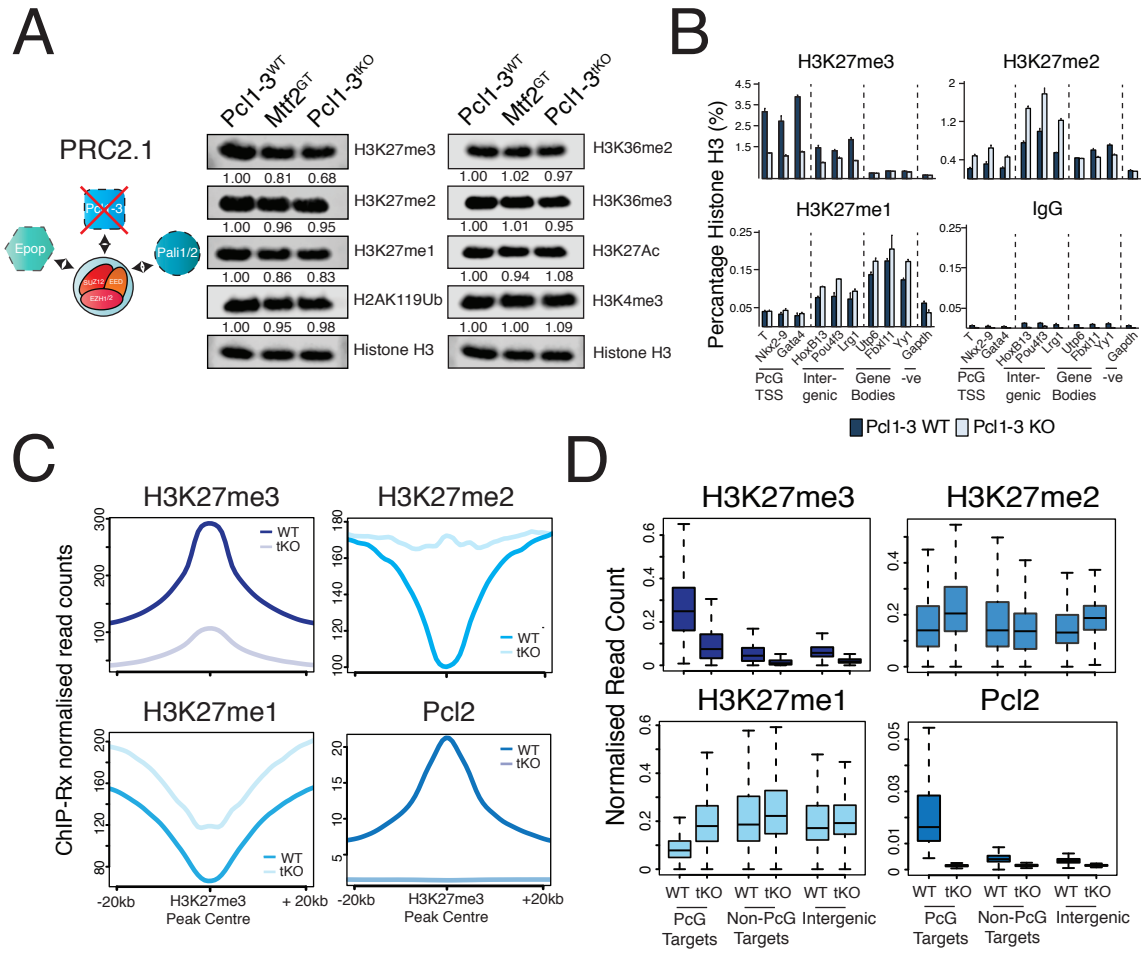


Figure 5.4 Loss of PRC2.1 results in a redistribution of H3K27methylations in ESCs.

(A) Left – Schematic of PRC2.1 illustrating knockout of Pcl1-3. Right – Western blot analyses using the indicated antibodies against Polycomb-associated histone modifications on nuclear lysates from Mtf2^{GT}, Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs.

(B) Quantitative chromatin immunoprecipitation (ChIP) analyses using the indicated antibodies from Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs. ChIP enrichments are presented as the percentage of protein bound normalised to Histone H3 and are a representative sample of three biological replicates.

(C) Average ChIP-Rx signal profiles of H3K27me3, H3K27me2, H3K27me1 and Pcl2 at all H3K27me3 peaks (+/- 20kb) (n=8526) in Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs.

(D) Distribution of H3K27me3, H3K27me2 and H3K27me1 ChIP-Rx normalised reads within PcG targets (n=5126), non-PcG targets (n=17969) (+/-2.5Kb of transcription target [TSS]) and intergenic regions (n=3893) in Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs.

western blot analysis presented in Figure 5.3, which shows only slight overall changes in the modification. This is not a new phenomenon, as it has been previously shown that loss of AEBP2 leads to an increase in H3K27me3 at PcG targets by ChIP-seq, with little to no effect on global changes of the mark by western blot (Grijzenhout et al., 2016). Taken together, these data suggest that PCL1-3 are critical for catalytic conversion of H3K27me1/2 to H3K27me3, and is consistent with a recent study reporting that PCL proteins enhance the methyltransferase efficiency of PRC2 by prolonging the residency time of the complex on chromatin (Choi et al., 2017). Furthermore, my data does not strongly suggest a role for PCL proteins in the catalysis of H3K27me1 or H3K27me2. In fact, in addition to increases in both PTMs at PcG target genes, I also observed an overall increase in H3K27me2 at intergenic regions of the genome, which may also be related to an overall decrease in H3K27me3 at these intergenic sites (Figure 5.4B-D). This average increase in H3K27me2 may be rationalised due to the fact that in the absence of PCL1-3, PRC2.1 lacks stable tethering to chromatin, and in particular at CpG islands and TSS's of PcG target genes. PRC2 is therefore potentially less restricted and free to catalyse H3K27me2 at intergenic regions (Youmans et al., 2018). The biological implications of this genome wide increase in H3K27me2 remains to be elucidated but it could play a role in more robust silencing of enhancer elements required for subsequent differentiation of ESCs. Taken together, these analyses suggests that Polycomb-like proteins are primarily important for maintaining PRC2.1 activity at CpG islands to promote H3K27me3 and that their loss leads to genome-wide redistributions of H3K27 methylations in ESCs.

5.2.5 PRC2.2 and H3K27me3 are retained at some PcG targets in Pcl1-3^{tKO} ESCs

Analysing the genome wide profiles of H3K27me1/2/3 in Pcl1-3^{tKO} ESCs revealed that although H3K27me3 is globally reduced, it is not completely lost from all PcG target genes (Figure 5.4D). In order to explore this result further, I performed quantitative genome wide ChIP-Rx for PRC2.1 (PCL2/EPOP),

PRC2.2 (JARID2/AEBP2), and core PRC2 (SUZ12) in Pcl1-3^{WT} and Pcl1-3^{tKO} ESCs. While the PRC2.1 specific member EPOP was sensitive to the loss of Polycomb-like proteins, as expected, surprisingly, both the PRC2.2 specific components, AEBP2 and JARID2 were reduced on average across all Polycomb target genes, albeit to a lesser extent (Figure 5.5A & B). This implies that PRC2.2 is at least partially dependent on PRC2.1 for binding on most Polycomb target genes. These data are consistent with the well-defined targeting roles of PCL1-3 and intriguingly suggests a co-dependence of PRC2.2 on PRC2.1. However, it also implies that loss of PCL1-3, and by extension PRC2.1, is not sufficient to completely abolish all PRC2 activity, and that the remaining activities of PRC2.2 may still be functioning in a compensatory role to maintain H3K27me3, potentially at critically important PcG target genes in ESCs.

5.2.6 PRC2.2 and H3K27me3 are maintained at "broad" PcG domains in Pcl1-3^{tKO} ESCs

I was next interested in exploring the regions at which H3K27me3 and PRC2.2 were retained in the absence of PRC2.1. When inspecting the genomic localisation patterns of PRC2.1 and PRC2.2 in Pcl1-3^{tKO} cells I observed that, while PRC2.1, PRC2.2 and H3K27me3 were significantly reduced on narrow PRC2 associated peaks they were retained at broader peak regions such as the *HoxA* locus (Figure 5.6A). Importantly, I confirmed this interesting result by qPCR in an independent ChIP experiment (Figure 5.6B). In order to determine if this pattern was recapitulated on a genome wide scale, I analysed the distribution and relative enrichment of PRC2.1, PRC2.2 and H3K27me3 at narrow and broad PcG domains in PCL1-3^{tKO} ESCs. Narrow and broad domains were defined as follows; for each ChIP-Rx experiment in Pcl1-3^{WT} cells, peaks within 10kb of each other were merged into a single region and the length distribution of the resulting regions partitioned into quartiles. Narrow and broad domains were defined as those falling within the first and fourth quartiles, respectively. Average profiles of these narrow and broad regions revealed that while PRC2.1 is equally lost from both, PRC2.2, SUZ12 and H3K27me3 were significantly retained on broad

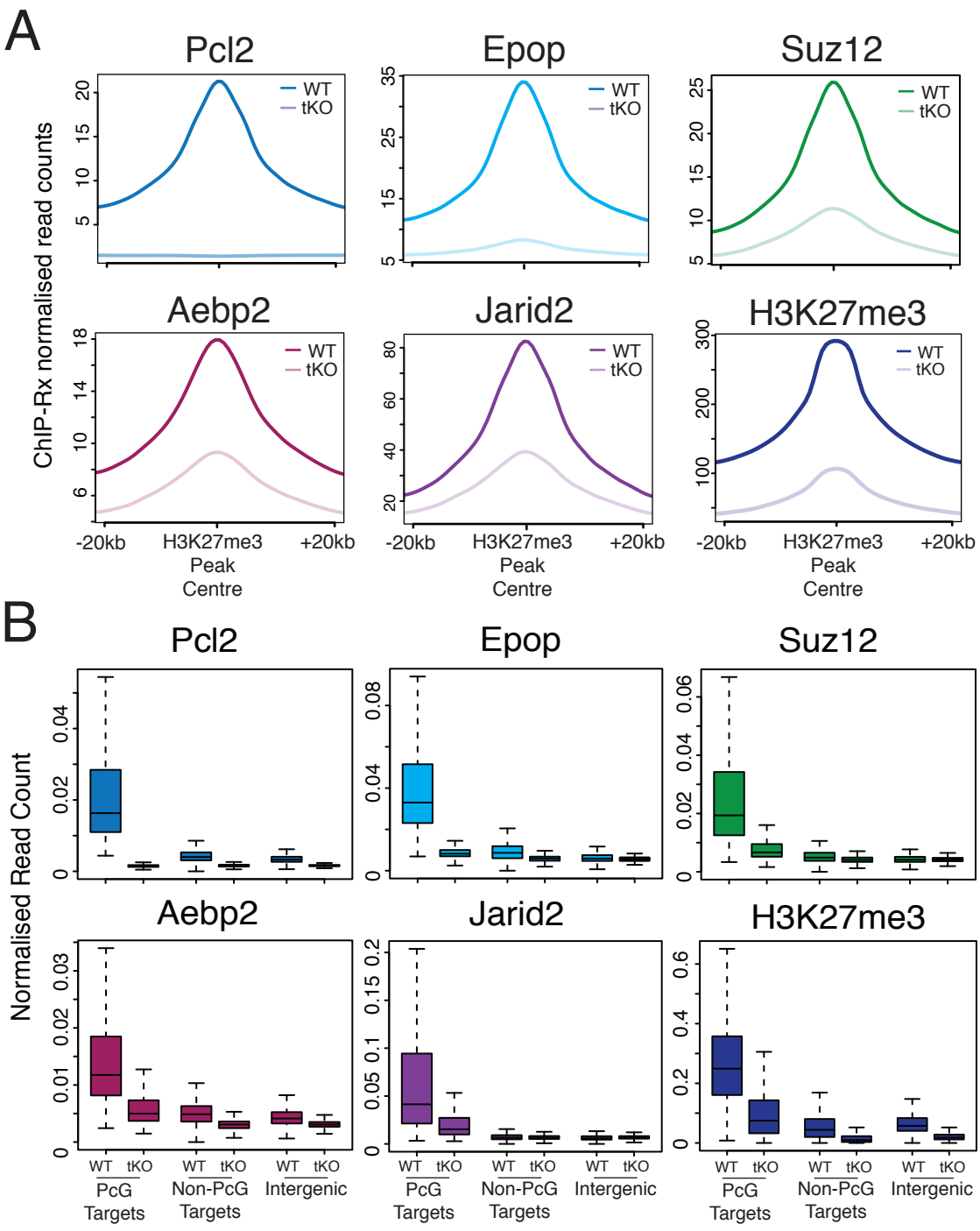


Figure 5.5 PRC2.2 and H3K27me3 are retained at some PcG targets in the absence of PRC2.1.

(A) Average ChIP-Rx signal profiles of Pcl2, Epop, Suz12, Aebp2 and Jarid2 at all H3K27me3 peaks (+/- 20kb) (n=8526) in Pcl1-3tKO and matched Pcl1-3WT ESCs.

(B) Distribution of Pcl2, Epop, Suz12, Aebp2 and Jarid2 ChIP-Rx normalised reads within PcG targets (n=5126), non-PcG targets (n=17969) (+/-2.5Kb of transcription target [TSS]) and intergenic regions (n=3893) in Pcl1-3tKO and matched Pcl1-3WT ESCs.

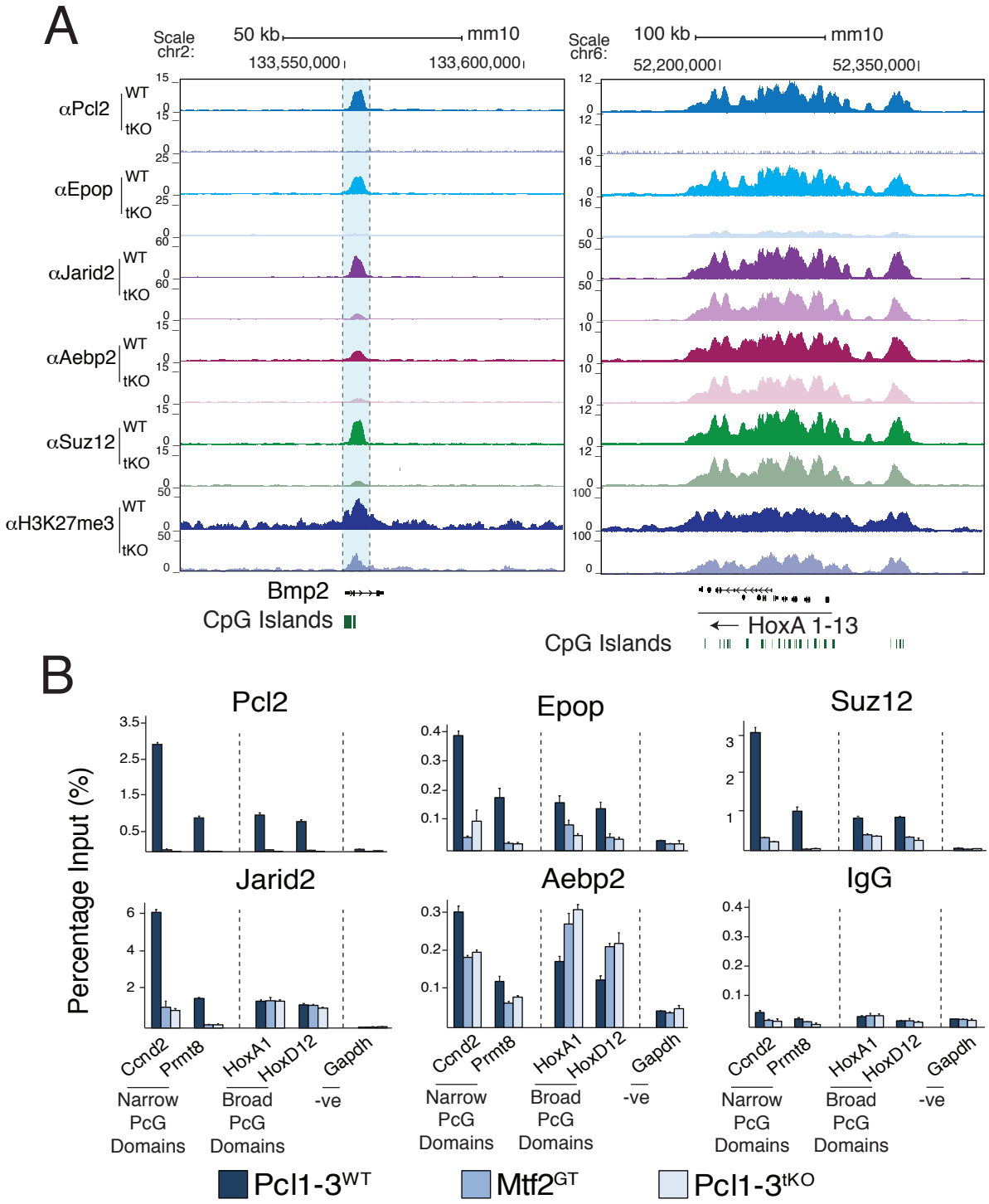


Figure 5.6 PRC2.2 and H3K27me3 are retained at broad PcG domains in the absence of PRC2.1.

(A) Genome browser representations of ChIP-Rx normalised reads for Pcl2, Epop, Jarid2, Aebp2, Suz12 and H3K27me3 at the Bmp2 gene locus (Narrow) and the HoxA locus (broad) in Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs.

(B) Quantitative chromatin immunoprecipitation (ChIP) analyses using the indicated antibodies at broad and narrow domains from Mtf2^{GT}, Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs (Analyses in A & B were performed on independent ChIP experiments). ChIP enrichments are presented as the percentage of protein bound normalised to input and are a representative sample of three biological replicates.

regions (Figure 5.7A & B). The high level of H3K27me3 maintained at a relatively small number of broad peaks might help explain the discrepancy between western blot analysis and ChIP-Rx average profile analysis presented earlier (Figure 5.4A & C). Taken together, these intriguing results suggests that knockout of a single PRC2 auxiliary component, even critical subunits such as the Polycomb-like proteins, will always leave a variant PRC2 complex intact. This variant complex can then play a compensatory role in order to maintain silencing of vital lineage specific genes contained within broad PcG repressed domains (e.g. Hox genes). It is therefore possible, that to completely abrogate PRC2 activity through its accessory subunits, both PRC2.1 and PRC2.2 need to be disrupted simultaneously.

5.2.7 Loss of PRC2.1 leads to re-targeting of PRC2.2

To further explore the reason for the apparent discrepancy between western blot and average profiles of H3K27me3 and AEBP2, I next explored the possibility that PRC2.2 components might be displaced to *de novo* sites in the Pcl1-3^{tKO} ESCs. Strikingly, I identified 185 specific genomic sites that specifically accumulate JARID2, but not PRC2.1 (EPOP) in the Pcl1-3^{tKO} condition (Figure 5.8A & B). Importantly, these *de novo* regions are marked by H3K27me3 in Pcl1-3^{WT} cells, which interestingly does not significantly change upon the loss of PRC2.1 (Figure 5.8A). JARID2 and AEBP2 accumulation at these regions was accompanied by SUZ12, suggesting the complete PRC2.2 is redistributed to these sites following the loss of PRC2.1 (Figure 5.8A & C). Critically, I validated these results by qPCR in an independent ChIP experiment to ensure that it was not a sequencing artefact from my ChIP-Rx dataset (Figure 5.8C).

Intriguingly, one of the top scoring PRC2.2 *de novo* sites is located ~500kb from the promoter of the *Meis2* gene (Figure 5.8A – right panel). *Meis2* repression during early development is known to be highly dependent on the PRC1 component RING1b, and the mechanisms of PcG mediated silencing of this gene have been elegantly studied previously (Kondo et al., 2014; Yakushiji-Kaminatsui et al., 2018). In this context, my identification of a PRC2.2 site ~500kb distal from

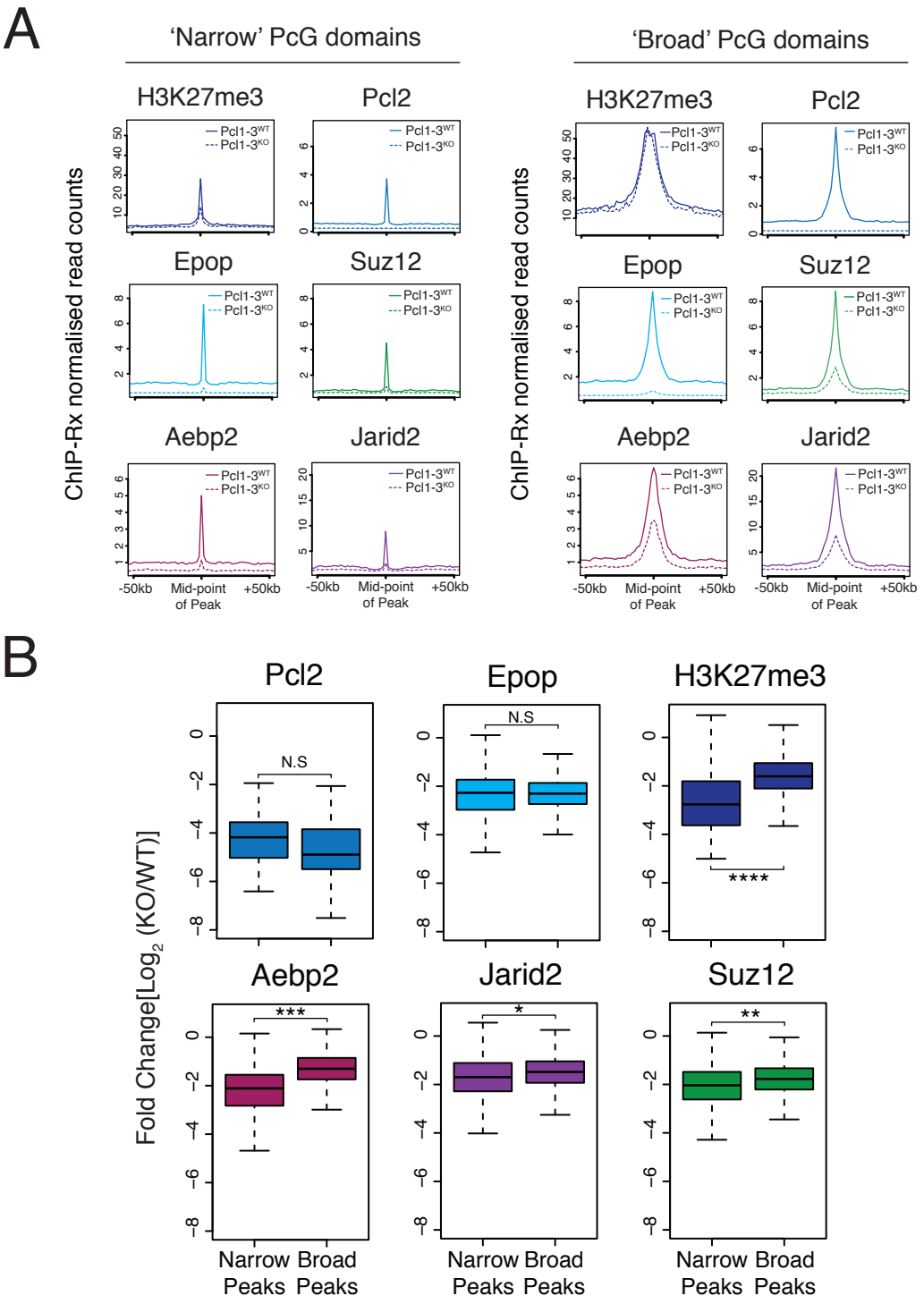


Figure 5.7 PRC2.2 and H3K27me3 are maintained at broad PcG domains in the absence of PRC2.1.

(A) Average ChIP-Rx signal profiles of H3K27me3, Pcl2, Epop, Suz12, Aebp2 and Jarid2 at "Narrow" and "Broad" PcG domains (+/- 50kb) in Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs (see methods for narrow/broad definition and peak numbers).

(B) Boxplot representations of Log2 fold change (KO/WT) for Pcl2, Epop, H3K27me3, Aebp2, Jarid2 and Suz12 at "Narrow' and "Broad" PcG domains in Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs. The distributions of the fold change of ChIP signal in narrow and broad peaks was compared using a one-tailed Wilcoxon test ("*" = $p < 0.05$ "N.S" = $p > 0.05$).

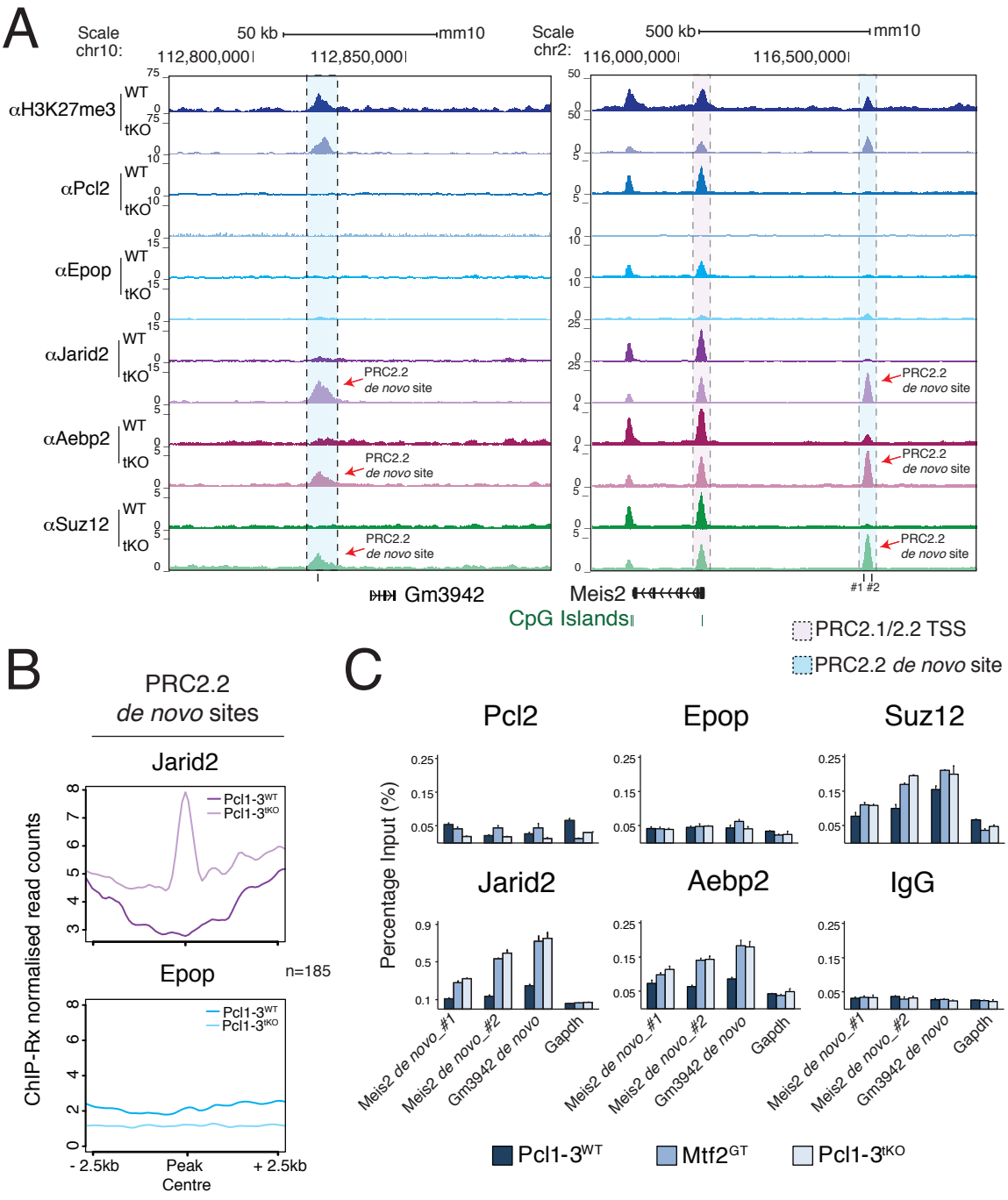


Figure 5.8 Loss of PRC2.1 leads to de novo re-targeting of PRC2.2.
(A) Genome browser representations of ChIP-Rx normalised reads for H3K27me3, Pcl2, Epop, Jarid2, Aebp2, Suz12 at sites of PRC2.2 re-targeting in the absence of PRC2.1.
(B) Average ChIP-Rx signal profiles of Jarid2 and Epop at PRC2.2 de novo sites in Pcl1-3^{tKO} (n=185) and matched Pcl1-3^{WT} ESCs.
(C) Quantitative chromatin immunoprecipitation (ChIP) analyses using the indicated antibodies at sites of PRC2.2 re-targeting in Mtf2^{GT}, Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs (A & C were performed on independent ChIP experiments). ChIP enrichments are presented as the percentage of protein bound normalised to input and are a representative sample of three biological replicates. Location of primer sets are indicated with a black bar below genome browser tracks.

the *Meis2* promoter is exciting, and it will be important to determine if this region loops to regulate *Meis2* expression during development.

5.8.9 Loss of PRC2.1 results in deregulation of Polycomb target genes during ESC differentiation

Next, I wished to evaluate the consequences of deregulated PRC2 activities in Pcl1-3^{tKO} ESCs on the expression of Polycomb target genes during induction to differentiate to embryoid bodies (EBs). Firstly, there was no observable difference in overall EB morphology at Days 2, 4 or 8 following induction of both Pcl1-3^{WT} and Pcl1-3^{tKO} ESCs to differentiate (Figure 5.9A). However, RT-qPCR analysis of key lineage specific germ layer genes, *Gata4* (endodermal) and *Fgf5* (ectodermal), revealed significantly higher expression in the PCL1-3^{tKO} compared to matched wild-type ESCs at both 4- and 8-days post induction to differentiate (Figure 5.9B). The expression of pluripotency genes, *Klf4* and *Fgf17*, were included as a positive control for the differentiation procedure (Figure 5.9B). I next expanded on these results by performing an independent RNA-seq analysis of PCL1-3^{tKO} and matched wild-type, before, during and after induction to differentiate. This revealed a cohort of genes, whose expression was significantly upregulated in the Pcl1-3^{tKO} condition, at both 4- and 8-days post induction (Figure 5.10). Among the most highly upregulated genes were key developmental regulators and known PcG target genes such as, *Gata4, Meis2, Wnt3, Fgf5, Igfbp5 and FoxA2* (Figure 5.10A & B). Analysing the top PRC2 target genes (defined by AEBP2 occupancy, n=2043) revealed the increase in expression was coincident with a loss of H3K27me3, PRC2.1 and PRC2.2 from the promoters of these genes in ESCs, however, there did not appear to be a correlation between the increase in expression and change in PRC2 occupancy at individual genes (Figure 5.11A & B). 74-76% of genes that lose either H3K27me3, EPOP, JARID2 or AEBP2 in ESCs are upregulated 8-days post induction to differentiate (Figure 5.11A & B). These results are consistent with a greater propensity of cells lacking PRC2.1 to activate PcG target genes during

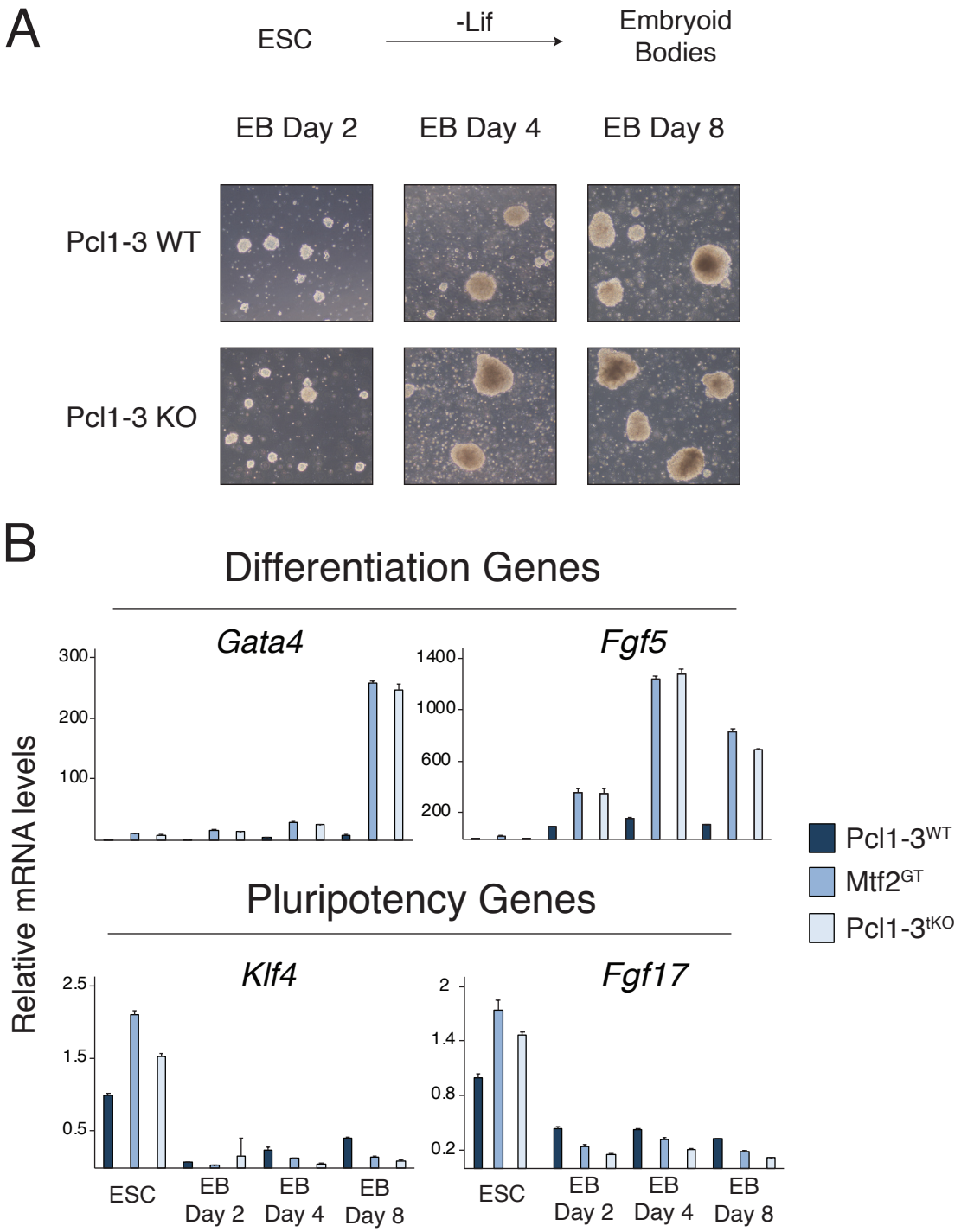


Figure 5.9 Loss of PRC2.1 leads to early activation of certain lineage specific genes during ESC differentiation.
(A) Representative images of Pcl1-3^{WT} and Pcl1-3^{tKO} at Day 2, 4 and 8 of ESC differentiation to embryoid bodies.
(B)RT-qPCR analyses of *Gata4* (endodermal), *Fgf5* (ectodermal) as well as *Klf4* and *Fgf17* (both pluripotency) mRNA transcripts from Mtf2^{GT}, Pcl1-3^{tKO} and matched Pcl1-3^{WT} ESCs at Day 2, 4 and 8 of ESC differentiation. Data presented are a representative sample of three biological replicates and error bars indicate standard deviation of individual triplicate qPCR data.

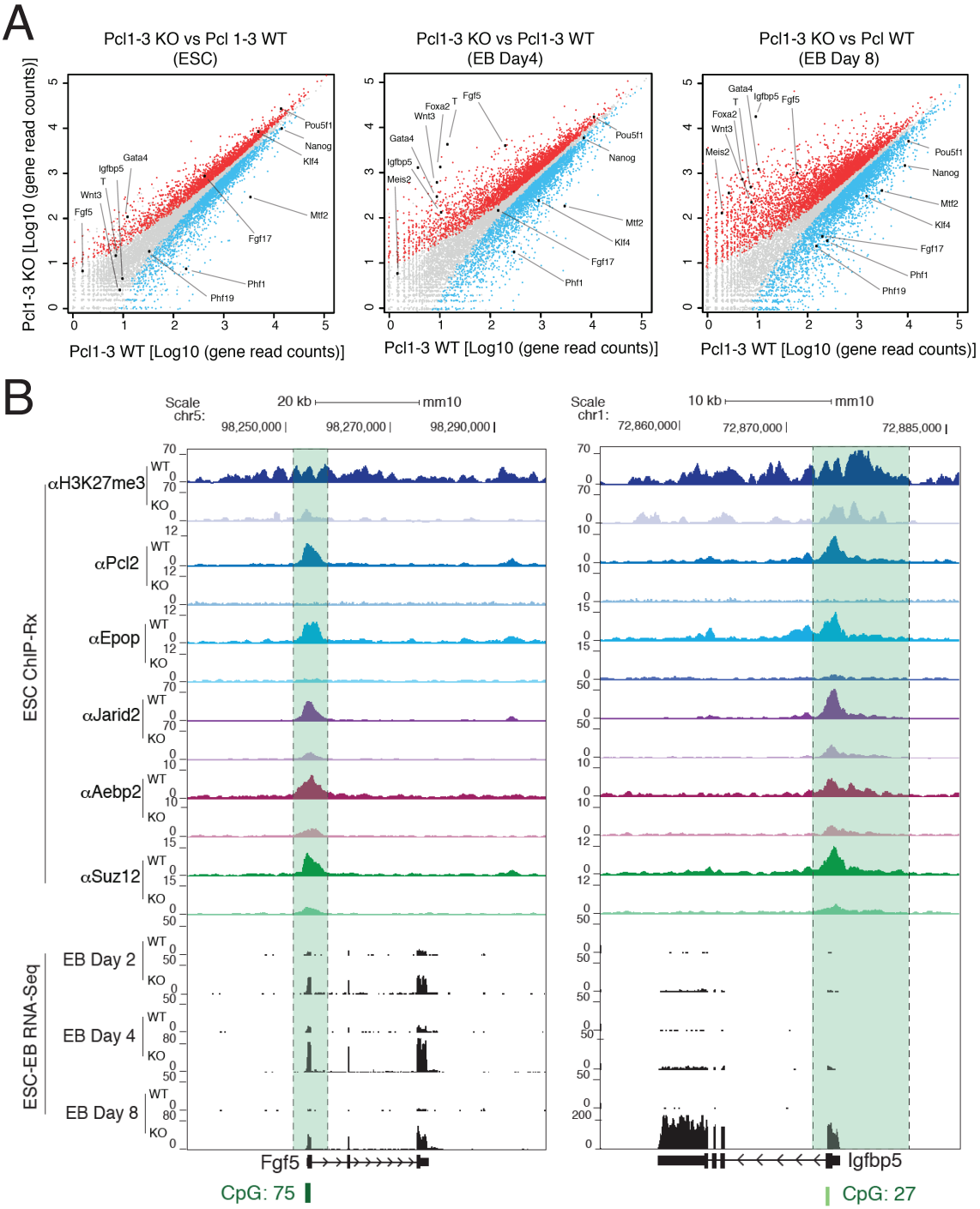


Figure 5.10 Loss of PRC2.1 results in deregulation of Polycomb target genes during ESC differentiation.

(A) Scatterplots of differentially expressed genes in Pcl1-3^{tKO} versus matched wild-type ESCs as well as embryoid bodies (EBs) at day 4 and 8 of differentiation. Log10 of the normalised gene count data are represented for all Ref-seq genes (n=19500). Genes showing significant expression changes are shown in red (increased) and blue (decreased).

(B) Genome browser representations of ChIP-Rx normalised reads for H3K27me3, Pcl2, Epop, Aebp2, Jarid2 and Suz12 in ESCs, as well as RNA-Seq reads at EB day 2, 4 and 8 at the Fgf5 and Igfbp5 gene loci in Pcl1-3^{tKO} and matching wild-type controls. The green box highlights the region around the transcription start site (TSS).

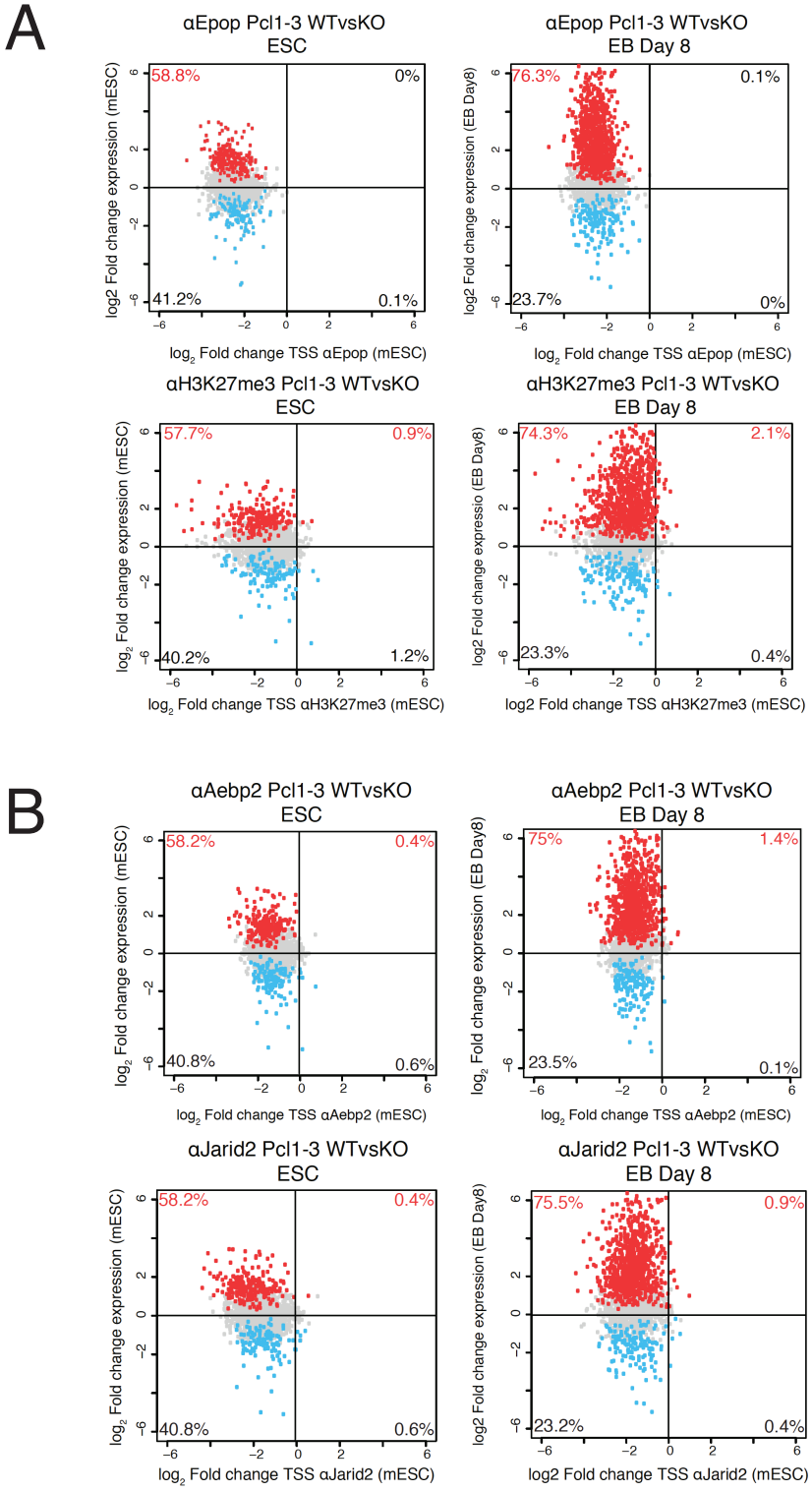


Figure 5.11 Improper regulation of PcG target genes during ESC differentiation coincides with a loss of PRC2.1, PRC2.1 and H3K27me3 from chromatin.

(A) & (B) The relationship between changes in Epop, H3K27me3, Aebp2 and Jarid2 levels in Pcl1-3^{tKO} ESCs and gene expression in ESCs and EBs (Day 8) for the top PRC2 target genes defined by Aebp2 enrichment (n=2043). Genes showing significant expression changes are shown in red (increased) and blue (decreased). The percentage of all genes in each quadrant is indicated.

differentiation and are in agreement with a similar study targeting PRC2.1 components, PALI1/2 (Conway et al., 2018).

5.8.10 Genes associated with narrow Polycomb domains exhibit more variable expression patterns

Having established that PRC2.2 and H3K27me3 are retained at broad PcG domains following the loss of PRC2.1 (Figure 5.6 & 5.7), I next wished to analyse the expression patterns of these genes compared to genes that fall within narrow PcG domains in ESCs and after induction to differentiate. An analysis of the raw RNA abundance of both categories of genes revealed that those within narrow regions exhibit a more variable expression pattern in Pcl1-3^{tKO} ESCs (Figure 5.12A). This more variable expression pattern of genes from narrow regions was also evident following induction of Pcl1-3^{tKO} ESCs to differentiate (Figure 5.12A). A representative gene associated with a narrow peak, *Sox17*, was upregulated in Pcl1-3^{tKO} at EB Day 8, while there was no detectable expression from any gene within the *HoxA* locus, located within a broad domain (Figure 5.12B). This variable expression pattern nicely coincides with either a loss or retention of H3K27me3 and SUZ12 enrichment. Taken together, I believe that while PCL proteins are critically important for PRC2 function and targeting, their loss does not confer global deregulation of the PcG system in ESCs. This data implies there are additional compensatory mechanisms at play, potentially involving and interplay between PRC2.2, ncPRC1 and H2AK119ub. These mechanisms likely facilitate the robust repression of crucial lineage specific genes involved in cell fate determination.

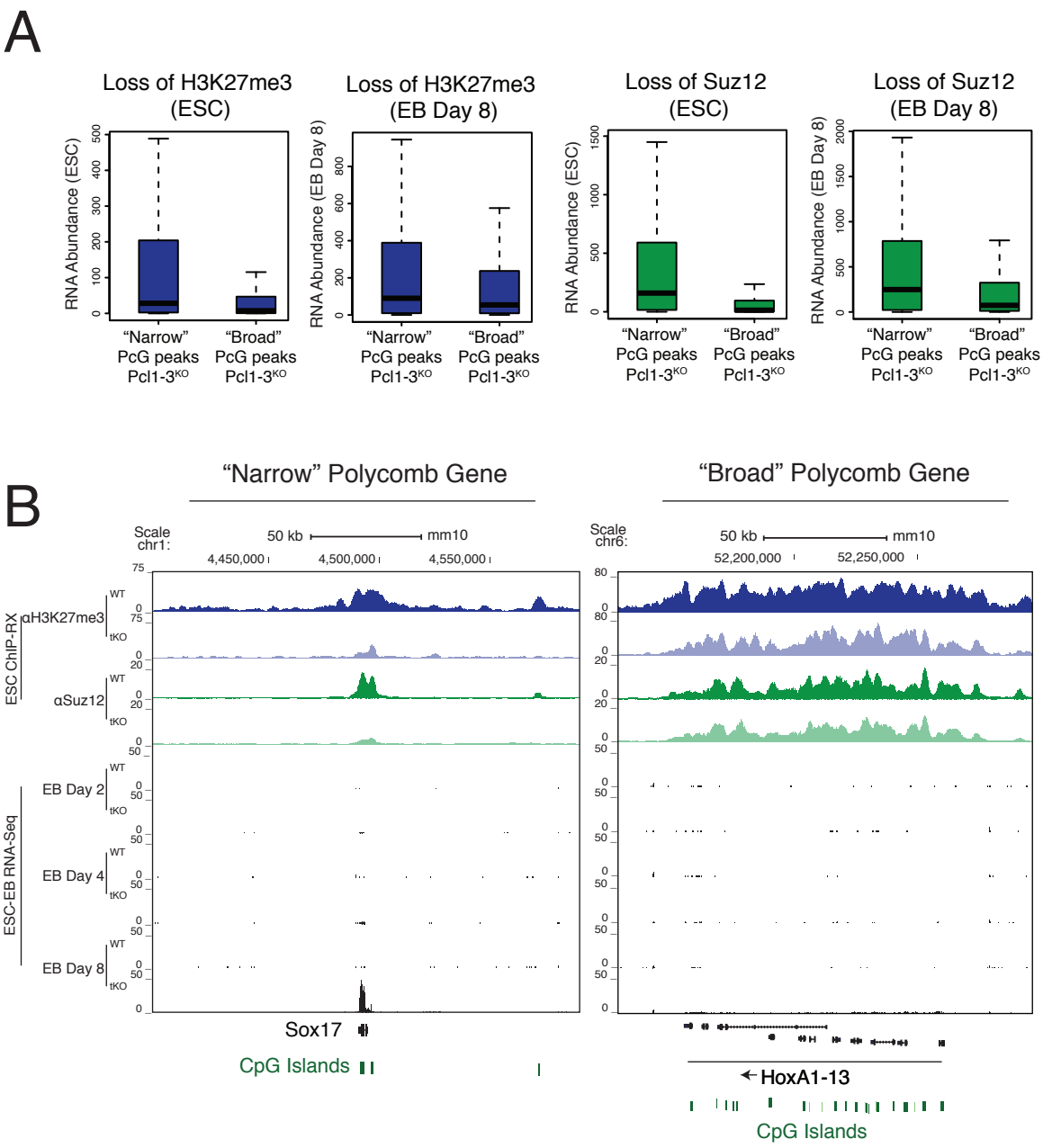


Figure 5.12 Genes associated with narrow Polycomb domains exhibit more variable expression patterns.

(A) Boxplot representations of the relationship between changes in H3K27me3 or Suz12 levels in Pcl1-3^{tKO} ESCs and the expression of genes that fall within both "narrow" and "broad" PcG domains in steady state ESCs and EB Day 8. Gene expression is determined from raw RNA-seq read counts. See methods for broad/narrow gene numbers.

(B) Genome browser representations of ChIP-Rx normalised reads for H3K27me3 and Suz12 in ESCs, as well as RNA-Seq reads at EB day 2, 4 and 8 at the Sox17 and HoxA gene loci in Pcl1-3^{tKO} and matching wild-type controls.

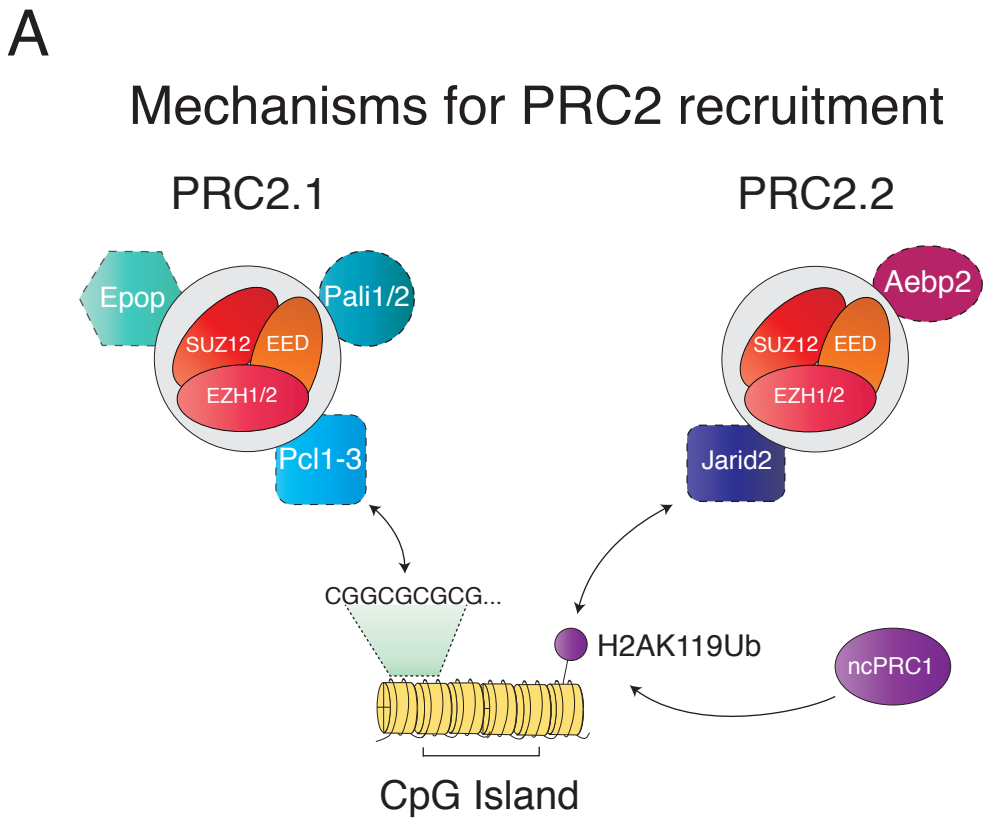


Figure 5.13 Mechanisms of PRC2 recruitment.
(A) Schematic describing two forms of PRC2 recruitment in ESCs. Pcl1-3 target PRC2.1 and PRC2.2 to CpG DNA whereas Jarid2 can bind and recruit PRC2 to CpG islands marked PRC1 mediated H2AK119Ub.

5.3 Discussion

In this chapter, I uncovered new roles for two unique PRC2 subcomplex assemblies; PRC2.1 and PRC2.2. ChIP-Rx analysis of PRC2.1 and PRC2.2 components in wild-type ESCs revealed that the genome wide localisation of these subcomplexes is largely overlapping. However, there is a cohort of PRC2.1 specific genomic loci. I also established that PCL1-3 are the essential defining components of the PRC2.1 subcomplex, and by maintaining EPOP and PALI1 on chromatin, they regulate the precise genome wide distribution of H3K27 methylations. Loss of PRC2.1 results in a decrease of PRC2.2 and H3K27me3 from chromatin, however, JARID2 and AEBP2 along with H3K27me3 are retained at broad PcG domains. Interestingly, genes that fall within these broad domains exhibit less variable expression compared to genes from narrow regions when PCL1-3^{tKO} ESCs are induced to differentiate. Coupled to this, PRC2.2 components are also intriguingly re-targeted in the absence of PRC2.1. Taken together, I have uncovered new and exciting roles for PCL1-3 in maintaining the balance of PRC2.1 and PRC2.2 subtype activities in ESCs and during differentiation.

5.3.1 Polycomb-like protein function during embryonic development.

While the knockout of core PRC2 subunits (EZH2, EED or SUZ12) leads to early embryonic lethality at E7.5-8.5 (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004), the disruption of genes encoding auxiliary components, such as PCL2, JARID2 and AEBP2 results in milder developmental phenotypes. For example, *Jarid2* and *Aebp2* null mice survive until later developmental stages (E11.5 for *Jarid2* and E10.5-18.5 for *Aebp2*) (Grijzenhout et al., 2016; Kim et al., 2011; Takeuchi et al., 1995). Similarly, despite reduction of both H3K27me2/3 at E11.5, *Pali1* knockout mice are born, but the phenotype is perinatal lethal, with death seeming to result from severe kidney defects (Conway et al., 2018). The precise phenotype of Polycomb-like knockout mice is much less clear. Two independent studies have reported contrasting results based on *Pcl2* (*Mtf2*) knockout mice (Li et al., 2011; Rothberg et al., 2018). Li *et al* suggest that *Pcl2*

null mice are viable but do present homeotic skeletal transformations, while Rothberg *et al* describe an embryonic lethal phenotype at E15.5. Reconciliation of these contrasting phenotypes is difficult, however, it might be explained by compensation of PCL2 loss by PCL1/PCL3. We have previously shown that *PCL1-3* are regulated differently at the transcriptional level (Brien et al., 2015). Hence, these genes are likely differentially expressed in various tissues and at different stages in the developing embryo. Therefore, it is very possible that *Pcl1* null mice would present a largely different phenotype to *Pcl2* or *Pcl3* knockouts, and vice versa. Taken together, I believe that to in order expand on previous *in vivo* studies and on the findings presented here, it will be necessary include all three Polycomb-like paralogues in any future murine knockout studies.

5.3.2 PRC2.1 vs PRC2.2 biochemistry

Similar to recent studies, I have shown here that the absence of a PRC2 auxiliary component (PCL1-3) leads to an imbalance in complex composition and PRC2 activity at PcG target genes (Conway et al., 2018; Grijzenhout et al., 2016). This raises interesting questions regarding the biochemistry and assembly of PRC2.1 and PRC2.2. It has been established previously that JARID2 and AEBP2 directly associate with core PRC2 via interaction with a region in SUZ12 proximal to its Zinc finger (Ciferri et al., 2012; Kasinath et al., 2018). More recently, direct *in vitro* biochemical competition has been proposed between components of PRC2.1 and PRC2.2, with SUZ12 providing unique structural platforms that can define distinct classes of complex (Chen et al., 2018). This model of steric competition between PRC2.1 and PRC2.2, suggests that a region within SUZ12 could act as a “molecular switch” capable of defining the two classes of complex and regulating divergent PRC2 activities. Future experiments should be aimed at elucidating the exact amino acid residues in SUZ12 that are important for this function, using this novel information to investigate whether PRC2 subcomplex activities can be modulated *in vivo*.

5.3.3 Mechanisms of PRC2 recruitment

Trans-acting DNA binding transcription factors such as Pho are essential for PRC2 recruitment to Polycomb response elements (PREs) in *Drosophila* (Bracken and Helin, 2009). However, biochemical evidence for this process is lacking in mammalian systems (Vella et al., 2012), with the mechanisms of how PRC2 is specifically recruited to chromatin remaining poorly defined. The data presented here, as well as in previously reports, strongly suggest that there are at least two mechanisms functioning to recruit PRC2 to chromatin in mammals (Figure 5.13). I have shown that as part of the PRC2.1 complex, PCL1-3 are critical in targeting PRC2 to CpG islands. However, an elegant model has recently emerged in which H2AK119ub modified nucleosomes, deposited by 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 mechanism for the recruitment of PRC2 to H2AK119ub is proposed to involve the ubiquitin interacting motif (UIM) of JARID2 (Cooper et al., 2016), however, the global levels of H3K27me3 remain unchanged upon complete loss of H2AK119ub via RING1a/b deletion (Chiacchiera et al., 2016; de Napoles et al., 2004). I speculate that the UIM of JARID2 may be functioning to maintain PRC2 occupancy at broad Polycomb domains in the absence of PRC2.1. It will be intriguing to investigate the relative enrichment of both cPRC1 and ncPRC1 complexes, as well as H2AK119Ub at both broad and narrow domains. Furthermore, it will be interesting to explore whether H2AK119Ub, along with JARID2 can maintain PRC2.2 activity at crucial lineage determining genes in the absence of PRC2.1. Additional mechanisms at play may involve the ability of AEBP2 to bind DNA (Kim et al., 2009; Wang et al., 2017) or an as of yet unidentified role for another PRC2 auxiliary component, such as PALI1 or EPOP.

Chapter 6

General Discussion

6.1 Discussion

The PRC2 complex is fundamentally important to lineage specification during mammalian development, with its critical functions being underscored by the embryonic lethal phenotypes of *Ezh2*$^{-/-}$, *Eed*$^{-/-}$ *and Suz12*$^{-/-}$ knockout mice (Faust et al., 1995; Margueron and Reinberg, 2011; O'Carroll et al., 2001; Pasini et al., 2004). In addition to development, deletion of PRC2 components in differentiating adult tissues generally leads to defective tissue regeneration resulting from impaired differentiation and/or loss of adult stem cell populations (Ezhkova et al., 2009; Hidalgo et al., 2012; Mochizuki-Kashio et al., 2011). Owing to the importance of PRC2 in both developing and adult tissues, the main goal of the research presented in this thesis was to delineate the relative contributions of Polycomb-like proteins (PCL1-3), essential PRC2 associated factors, to overall PRC2 biology and activity in both pluripotent stem cells and differentiated cellular systems.

Here, I uncover new mechanistic roles for PCL proteins in both pluripotent embryonic stem cells and differentiated human cells. PCL1 has been shown to be highly expressed in quiescent cells and stabilise p53 (Brien et al., 2015). I show that PCL1 has specifically gained this function through a unique evolutionary process known as, neofunctionalisation. I also show that PCL1 defines a novel PRC2 complex, lacking SUZ12, that exists exclusively in quiescent cells which is both methyltransferase inactive and refractory to PRC2 inhibition. Through a conditional knockout strategy in ESCs, I have also revealed that PCL1-3 define a unique PRC2 subcomplex, termed PRC2.1, and are critically important in the maintenance of chromatin and transcriptional landscapes in the regulation of cellular identity during differentiation. These observations suggest that PCL proteins contribute to PRC2 activity at different stages of the cell cycle and in different cell types during differentiation and development, and may potentially help to reconcile why there are three PCL genes in mammals as opposed to one ancestral copy in *Drosophila*.

6.1.1 PCL proteins in adult stem and progenitor cell populations

Adult tissue specific stem cells exist as long-lived undifferentiated cells that have the unique capacity to produce differentiating progenitor cells while maintaining their own identity by self-renewal (Fuchs and Chen, 2013). In doing so, they can maintain adult tissue homeostasis and compensate for tissue loss during the life of an organism. Adult tissue specific stem cells are generally maintained in a quiescent state for prolonged periods of time until a stimulus to differentiate and divide is received. This process yields one daughter cell that retains stem cell identify and another lineage committed progenitor cell. This process eventually results in terminally differentiated non-dividing populations of cells. My observations on the gene expression, biochemistry and chromatin enrichment patterns of PRC2 components in primary human cells imply that the predominant PRC2 complex that exists in quiescent cells consists of EZH1, EED and PCL1 (G0-PRC2). Indeed, analysis of the transcriptional profiles of several quiescent stem cell populations indicate that EZH1 and PCL1 expression are significantly enriched in quiescent stem cells *in vivo* (Cheung and Rando, 2013). My analysis of the hematopoietic stem cell (HSC) hierarchy, also reveal that EZH1 and PCL1 are both highly expressed in quiescent HSCs and non-dividing terminally differentiated cells, with EZH2, SUZ12, PCL2 and PCL3 conversely being expressed in proliferating progenitor cells (Figure 4.12). It will be critically important to confirm these findings by analysing expression and chromatin binding dynamics of both G0-PRC2 and canonical PRC2 complexes in purified stem and progenitor populations, to unravel the exact roles of these complexes *in vivo.* To date I have been unsuccessful in finding an appropriate *in vivo* model system to study G0-PRC2 function. However, analysis of quiescent naïve and memory T cells in constitutive *EZH1* knockout mice may represent an intriguing system to explore the consequences of disrupting G0-PRC2 function *in vivo*.

6.1.2 PCL proteins as essential PRC2 recruitment factors

It has been proposed that multivalent cooperative engagement of adjacent histone modifications, within or between nucleosomes, by proteins with multiple

reader domains may provide an additional layer of specificity for recruitment of chromatin associated factors (Ruthenburg et al., 2007). In light of this and given their unique domain architecture with both chromatin and DNA binding abilities', Polycomb-like proteins have been shown to be a key player in the targeting of PRC2 to specific genomic loci. The reading capabilities of the PCL-Tudor domain have been well established, as they bind to H3K36me2/3 and *de novo* recruit PRC2 to silence actively transcribed genes during differentiation (Ballare et al., 2012; Brien et al., 2012). However, the recently identified winged-helix DNA binding domain of PCL1-3 adds yet another layer of complexity to this process. This domain has been proposed to bind specifically to CpG islands and even to unique DNA structural features within these genomic regions (Li et al., 2017; Perino et al., 2018). It is likely that these domains interplay with one another to facilitate accurate recruitment. Towards exploring this further, I have designed rescue experiments to perform in our PCL1-3^{tKO} ESCs, aimed at delineating the relative contributions of the Tudor and winged-helix domains in PRC2 recruitment during differentiation.

Here, I have shown PCL1-3 are key factors in the recruitment of PRC2 and the deposition of H3K27me3 in ESCs. However, my observation that H3K27me3 and PRC2.2 are maintained at some genomic loci indicates the process of specific PRC2 targeting is likely more complex and goes beyond PCL1-3. Indeed, roles for both PCL2 (MTF2) and JARID2 in the stable *de novo* recruitment of PRC2 and in the creation of H3K27me3 domains have been described recently (Oksuz et al., 2018), suggesting that established PRC2 localisation likely requires more than one associated factor. Furthermore, a recently defined DNA binding domain in AEBP2 (Wang et al., 2017), the putative ability of PALI1/2 proteins to bind co-repressor complexes and nuclear receptors (Conway et al., 2018), as well the capacity of EPOP to interact with Elongin BC proteins (Beringer et al., 2016), all complement the hypothesis that PRC2 targeting may not be reliant on one single component. Taken together, this implies that stable PRC2 recruitment is multivalent and requires the cooperation of both PRC2.1 and PRC2.2 components. Going forward, it will be very important to further delineate the

respective functions of the several DNA and chromatin binding domains within PRC2.1 and PRC2.2 and their interplay with one another and with other chromatin regulators during lineage commitment and mammalian development.

6.1.3 PCL proteins as targets for cancer therapy

Since 2010, multiple genome-wide sequencing studies have revealed that PRC2 function and H3K27 methylations are frequently deregulated in several cancers (Comet et al., 2016; Conway et al., 2015) (Table 1.1). This has led to the development of several small molecule inhibitors for EZH2 and subsequent initiation of clinical trials (Makita and Tobinai, 2018; McCabe et al., 2012b). Here, I have described various novel mechanisms that could be exploited to target PRC2 via Polycomb-like proteins in cancer. Firstly, defining the interaction between PCL1 and p53 in quiescent cells has led to the hypothesis that disrupting this interaction in quiescent cancer stem cells may destabilise p53 and force these cells into the cell-cycle. This would then render these cells more sensitive to traditional chemotherapies that only target proliferating cells (Figure 3.8). Secondly, our observation that G0-PRC2 is the predominant PRC2 complex in quiescent cells, and that this complex is refractory to two forms of PRC2 inhibition suggests that EZH2 drugs in clinical trials at the moment will not be effective against quiescent cancer stem cells. I am now in the process of collaborating to develop protein degrader drugs targeting PRC2 core subunit, EED. By degrading the EED protein, in combination with EZH2 inhibitors, this will hopefully disrupt both G0-PRC2 and canonical PRC2 complexes, reducing the overall risk of recurrence and tumour re-seeding caused by non-targeted quiescent cancer stem cells (Figure 4.13).

It has already been well established that cancer cell lines grown under the selective pressure of EZH2 inhibitors develop resistant secondary mutations (Baker et al., 2015; Gibaja et al., 2016). This represents a worrying scenario regarding the treatment of lymphoma patients with these drugs in clinical trials at the moment. The occurrence of EZH2i resistant populations of cells in human malignancies is a real possibility and so there is an urgent and pressing need to

discover additional ways to target PRC2 in cancer. I believe that further in-depth studies focusing on PRC2 accessory subunits, similar to the one presented here, will provide new important insights into the chromatin and transcriptional dynamics of PRC2. My observations as well as the results of future studies, on the systematic functions of all PRC2 subunits, will have extremely important implications in the treatment of cancers and particularly in the hunt for second-wave PRC2 inhibitors.

In summary, this study adds to a growing body of evidence regarding sub-functionalisation of the Polycomb system in higher eukaryotes. Through focusing on Polycomb-like proteins in pluripotent and differentiated cells I have uncovered new insights into how the function of Polycomb group proteins have specialised during evolution. I believe this work will contribute to a better mechanistic understanding of the roles of Polycomb proteins during cell fate decisions and in complex biological processes such as carcinogenesis and embryogenesis.

6.2 Conclusions

- PCL1 interacts with p53 through two divergent Serine residues in its N-terminal PHD1 domain, having acquired this function during recent mammalian evolution.
- PCL1 defines a unique PRC2 complex, lacking SUZ12, that exists exclusively in non-diving quiescent cells, termed G0-PRC2.
- G0-PRC2 in methyltransferase inactive, and is refractory to PRC2 inhibition.
- PCL1-3 are the defining components of PRC2.1, a unique PRC2 subtype assembly in ESCs.
- Loss of PRC2.1 perturbs the delicate balance of PRC2 subtype activities and results in improper regulation of PcG target genes during differentiation.

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., 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., Reveron-Gomez, 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 Dev *29*, 585-590.

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.

Baker, T., Nerle, S., Pritchard, J., Zhao, B., Rivera, V.M., Garner, A., and Gonzalvez, 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.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Herault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354*, 909-912.

Batlle, E., and Clevers, H. (2017). Cancer stem cells revisited. Nature medicine *23*, 1124-1134.

Beckerman, R., and Prives, C. (2010). Transcriptional regulation by p53. Cold Spring Harb Perspect Biol *2*, a000935.

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*, 729.

Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T.*, et al.* (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res *26*, 896-907.

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.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev *16*, 6-21.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res *8*, 1499-1504.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.

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.

Boonsanay, V., Zhang, T., Georgieva, A., Kostin, S., Qi, H., Yuan, X., Zhou, Y., and Braun, T. (2016). Regulation of Skeletal Muscle Stem Cell Quiescence by Suv4-20h1-Dependent Facultative Heterochromatin Formation. Cell Stem Cell *18*, 229-242.

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.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255-256.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol *34*, 525-527.

Brien, G.L., and Bracken, A.P. (2016). The PCL1-p53 axis promotes cellular quiescence. Cell Cycle *15*, 305-306.

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.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515-526.

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.

Cancer Genome Atlas, N. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature *517*, 576-582.

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., 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.

Casanova, M., Preissner, T., Cerase, A., Poot, R., Yamada, D., Li, X., Appanah, R., Bezstarosti, K., Demmers, J., Koseki, H.*, et al.* (2011). Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. Development *138*, 1471-1482.

Chen, S., Jiao, L., Shubbar, M., Yang, X., and Liu, X. (2018). Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. Mol Cell *69*, 840-852 e845.

Cheung, T.H., and Rando, T.A. (2013). Molecular regulation of stem cell quiescence. Nat Rev Mol Cell Biol *14*, 329-340.

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

Choi, J., Bachmann, A.L., Tauscher, K., Benda, C., Fierz, B., and Muller, J. (2017). DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. Nat Struct Mol Biol *24*, 1039-1047.

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.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol *18*, 407-422.

Cohen, N.M., Kenigsberg, E., and Tanay, A. (2011). Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. Cell *145*, 773-786.

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.

Conant, G.C., and Wolfe, K.H. (2008). Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet *9*, 938-950.

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., Jerman, E., Healy, E., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D.J., Mucha, M.*, et al.* (2018). A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. Mol Cell.

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.

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.

Dittmer, D., Pati, S., Zambetti, G., Chu, S., Teresky, A.K., Moore, M., Finlay, C., and Levine, A.J. (1993). Gain of function mutations in p53. Nature genetics *4*, 42-46.

Do, P.M., Varanasi, L., Fan, S., Li, C., Kubacka, I., Newman, V., Chauhan, K., Daniels, S.R., Boccetta, M., Garrett, M.R.*, et al.* (2012). Mutant p53 cooperates with ETS2 to promote etoposide resistance. Genes Dev *26*, 830-845.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433-3434.

DuPage, M., Chopra, G., Quiros, J., Rosenthal, W.L., Morar, M.M., Holohan, D., Zhang, R., Turka, L., Marson, A., and Bluestone, J.A. (2015). The chromatin-modifying enzyme Ezh2 is critical for the maintenance of regulatory T cell identity after activation. Immunity *42*, 227-238.

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, 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. Nature genetics *42*, 722-726.

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

Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.H., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell *136*, 1122-1135.

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.

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.

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.

Fuchs, E., and Chen, T. (2013). A matter of life and death: self-renewal in stem cells. EMBO Rep *14*, 39-48.

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.

Gil, J., and Peters, G. (2006). Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. Nat Rev Mol Cell Biol *7*, 667-677.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

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.

Hammarlund, E., Lewis, M.W., Hansen, S.G., Strelow, L.I., Nelson, J.A., Sexton, G.J., Hanifin, J.M., and Slifka, M.K. (2003). Duration of antiviral immunity after smallpox vaccination. Nature medicine *9*, 1131-1137.

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. Nature chemical biology *13*, 389-395.

Helin, K., and Dhanak, D. (2013). Chromatin proteins and modifications as drug targets. Nature *502*, 480-488.

Hidalgo, I., Herrera-Merchan, A., Ligos, J.M., Carramolino, L., Nunez, J., Martinez, F., Dominguez, O., Torres, M., and Gonzalez, S. (2012). Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest. Cell Stem Cell *11*, 649-662.

Hill, A., and Bloom, K. (1987). Genetic manipulation of centromere function. Molecular and cellular biology *7*, 2397-2405.

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.

Holoch, D., and Margueron, R. (2017). Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. Trends Biochem Sci *42*, 531-542.

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.

Hu, C.K., and Brunet, A. (2018). The African turquoise killifish: A research organism to study vertebrate aging and diapause. Aging Cell *17*, e12757.

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.

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.

Italiano, A., Soria, J.C., Toulmonde, M., Michot, J.M., Lucchesi, C., Varga, A., Coindre, J.M., Blakemore, S.J., Clawson, A., Suttle, B.*, et al.* (2018). Tazemetostat, an EZH2 inhibitor, in relapsed or refractory B-cell non-Hodgkin lymphoma and advanced solid tumours: a first-in-human, open-label, phase 1 study. Lancet Oncol *19*, 649-659.

Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell *19*, 523-534.

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

Jiang, Y.W., Veschambre, P., Erdjument-Bromage, H., Tempst, P., Conaway, J.W., Conaway, R.C., and Kornberg, R.D. (1998). Mammalian mediator of

transcriptional regulation and its possible role as an end-point of signal transduction pathways. Proc Natl Acad Sci U S A *95*, 8538-8543.

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.

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.

Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci Adv *1*, e1500447.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nature genetics *45*, 592-601.

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev *6*, 705-714.

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.

Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. Genes Dev *19*, 1129-1155.

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, H., Kang, K., and Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. Nucleic Acids Res *37*, 2940-2950.

Kim, K.H., and Roberts, C.W. (2016). Targeting EZH2 in cancer. Nature medicine *22*, 128-134.

Kloet, S.L., Makowski, M.M., Baymaz, H.I., van Voorthuijsen, L., Karemaker, I.D., Santanach, A., Jansen, P., Di Croce, L., and Vermeulen, M. (2016). The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. Nat Struct Mol Biol *23*, 682-690.

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. Nature chemical biology *8*, 890-896.

Kondo, T., Isono, K., Kondo, K., Endo, T.A., Itohara, S., Vidal, M., and Koseki, H. (2014). Polycomb potentiates meis2 activation in midbrain by mediating interaction of the promoter with a tissue-specific enhancer. Dev Cell *28*, 94-101.

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.

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.

Laugesen, A., and Helin, K. (2014). Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell *14*, 735-751.

Lecona, E., Rojas, L.A., Bonasio, R., Johnston, A., Fernandez-Capetillo, O., and Reinberg, D. (2013). Polycomb protein SCML2 regulates the cell cycle by binding and modulating CDK/CYCLIN/p21 complexes. PLoS Biol *11*, e1001737.

Lee, C.H., Holder, M., Grau, D., Saldana-Meyer, R., Yu, J.R., Ganai, R.A., Zhang, J., Wang, M., LeRoy, G., Dobenecker, M.W.*, et al.* (2018). Distinct Stimulatory Mechanisms Regulate the Catalytic Activity of Polycomb Repressive Complex 2. Mol Cell *70*, 435-448 e435.

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. Nature genetics *46*, 1227-1232.

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. Molecular and cellular biology *31*, 351-364.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

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.

Lindsay, C.D., Kostiuk, M.A., Harris, J., O'Connell, D.A., Seikaly, H., and Biron, V.L. (2017). Efficacy of EZH2 inhibitory drugs in human papillomavirus-positive and human papillomavirus-negative oropharyngeal squamous cell carcinomas. Clin Epigenetics *9*, 95.

Liu, L., Cheung, T.H., Charville, G.W., Hurgo, B.M., Leavitt, T., Shih, J., Brunet, A., and Rando, T.A. (2013). Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. Cell Rep *4*, 189-204.

Liu, Y., Elf, S.E., Miyata, Y., Sashida, G., Liu, Y., Huang, G., Di Giandomenico, S., Lee, J.M., Deblasio, A., Menendez, S.*, et al.* (2009). p53 regulates hematopoietic stem cell quiescence. Cell Stem Cell *4*, 37-48.

Lonie, A., D'Andrea, R., Paro, R., and Saint, R. (1994). Molecular characterisation of the Polycomblike gene of Drosophila melanogaster, a trans-acting negative regulator of homeotic gene expression. Development *120*, 2629-2636.

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.

Lu, X., Liu, D.P., and Xu, Y. (2013). The gain of function of p53 cancer mutant in promoting mammary tumorigenesis. Oncogene *32*, 2900-2906.

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 A resolution. Nature *389*, 251-260.

Makino, T., and McLysaght, A. (2010). Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc Natl Acad Sci U S A *107*, 9270-9274.

Makita, S., and Tobinai, K. (2018). Targeting EZH2 with tazemetostat. Lancet Oncol *19*, 586-587.

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., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol Cell *32*, 503-518.

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

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.

McGinty, R.K., and Tan, S. (2015). Nucleosome structure and function. Chem Rev *115*, 2255-2273.

Meletis, K., Wirta, V., Hede, S.M., Nister, M., Lundeberg, J., and Frisen, J. (2006). p53 suppresses the self-renewal of adult neural stem cells. Development *133*, 363-369.

Micci, F., Panagopoulos, I., Bjerkehagen, B., and Heim, S. (2006). Consistent rearrangement of chromosomal band 6p21 with generation of fusion genes JAZF1/PHF1 and EPC1/PHF1 in endometrial stromal sarcoma. Cancer Res *66*, 107-112.

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.

Mochizuki-Kashio, M., Mishima, Y., Miyagi, S., Negishi, M., Saraya, A., Konuma, T., Shinga, J., Koseki, H., and Iwama, A. (2011). Dependency on the polycomb gene Ezh2 distinguishes fetal from adult hematopoietic stem cells. Blood *118*, 6553-6561.

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. Nature medicine *23*, 483-492.

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.

Mohrmann, L., and Verrijzer, C.P. (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochim Biophys Acta *1681*, 59-73.

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.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di Croce, L. (2012). Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell *10*, 47-62.

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. Nature genetics *42*, 181-185.

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

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197-208.

Muller, P.A., and Vousden, K.H. (2014). Mutant p53 in cancer: new functions and therapeutic opportunities. Cancer Cell *25*, 304-317.

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.

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. Nature genetics *42*, 665-667.

Ntziachristos, P., Tsirigos, A., Van Vlierberghe, P., Nedjic, J., Trimarchi, T., Flaherty, M.S., Ferres-Marco, D., da Ros, V., Tang, Z., Siegle, J.*, et al.* (2012).

Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nature medicine *18*, 298-301.

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. Molecular and cellular biology *21*, 4330-4336.

Ohno, S. (1970). Evolution by gene duplication (London,

New York,: Allen & Unwin;

Springer-Verlag).

Oksuz, O., Narendra, V., Lee, C.H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A.*, et al.* (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. Mol Cell *70*, 1149-1162 e1145.

Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep *9*, 1163-1170.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Molecular and cellular biology *27*, 3769-3779.

Pasini, D., Bracken, 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. (2010). JARID2

regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306-310.

Perino, M., van Mierlo, G., Karemaker, I.D., van Genesen, S., Vermeulen, M., Marks, H., van Heeringen, S.J., and Veenstra, G.J.C. (2018). MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. Nature genetics.

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.

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. Nature medicine *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. Nature chemical biology *13*, 381-388.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.

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.

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*.

Rothberg, J.L.M., Maganti, H.B., Jrade, H., Porter, C.J., Palidwor, G.A., Cafariello, C., Battaion, H.L., Khan, S.T., Perkins, T.J., Paulson, R.F*., et al.* (2018). Mtf2-PRC2 control of canonical Wnt signaling is required for definitive erythropoiesis. Cell Discov *4*, 21.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat Rev Mol Cell Biol *8*, 983-994.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., and Reinberg, D. (2008). Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Molecular and cellular biology *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. Nature genetics *47*, 1179-1186.

Schuettengruber, B., Bourbon, H.M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell *171*, 34-57.

Seita, J., Sahoo, D., Rossi, D.J., Bhattacharya, D., Serwold, T., Inlay, M.A., Ehrlich, L.I., Fathman, J.W., Dill, D.L., and Weissman, I.L. (2012). Gene Expression Commons: an open platform for absolute gene expression profiling. PLoS One *7*, e40321.

Seita, J., and Weissman, I.L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip Rev Syst Biol Med *2*, 640-653.

Shah, N.P., Tran, C., Lee, F.Y., Chen, P., Norris, D., and Sawyers, C.L. (2004). Overriding imatinib resistance with a novel ABL kinase inhibitor. Science *305*, 399-401.

Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G.*, et al.* (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature *486*, 395-399.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics *15*, 284.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell *4*, 481-495.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. Dev Biol *158*, 131-144.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat Rev Genet *14*, 204-220.

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.

Son, J., Shen, S.S., Margueron, R., and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. Genes Dev *27*, 2663-2677.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res *4*, 1521.

Souroullas, G.P., Jeck, W.R., Parker, J.S., Simon, J.M., Liu, J.Y., Paulk, J., Xiong, J., Clark, K.S., Fedoriw, Y., Qi, J.*, et al.* (2016). An oncogenic Ezh2 mutation induces tumors through global redistribution of histone 3 lysine 27 trimethylation. Nature medicine *22*, 632-640.

Sprent, J., and Tough, D.F. (1994). Lymphocyte life-span and memory. Science *265*, 1395-1400.

Streubel, G., Watson, A., Jammula, S.G., Scelfo, A., Fitzpatrick, D.J., Oliviero, G., McCole, R., Conway, E., Glancy, E., Negri, G.L.*, et al.* (2018). The

H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. Mol Cell *70*, 371-379 e375.

Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.A., Jones, D.T., Konermann, C., Pfaff, E., Tonjes, 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.

Swigut, T., and Wysocka, J. (2007). H3K27 demethylases, at long last. Cell *131*, 29-32.

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.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H.*, et al.* (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664-678.

Trimarchi, J.M., Fairchild, B., Wen, J., and Lees, J.A. (2001). The E2F6 transcription factor is a component of the mammalian Bmi1-containing polycomb complex. Proc Natl Acad Sci U S A *98*, 1519-1524.

Tsai, W.W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.Y., Shi, X., Schwarzer, D., Plunkett, W.*, et al.* (2010). TRIM24 links a non-canonical histone signature to breast cancer. Nature *468*, 927-932.

Tumes, D.J., Onodera, A., Suzuki, A., Shinoda, K., Endo, Y., Iwamura, C., Hosokawa, H., Koseki, H., Tokoyoda, K., Suzuki, Y.*, et al.* (2013). The polycomb

protein Ezh2 regulates differentiation and plasticity of CD4(+) T helper type 1 and type 2 cells. Immunity *39*, 819-832.

Valenzano, D.R., Benayoun, B.A., Singh, P.P., Zhang, E., Etter, P.D., Hu, C.K., Clement-Ziza, M., Willemsen, D., Cui, R., Harel, I.*, et al.* (2015). The African Turquoise Killifish Genome Provides Insights into Evolution and Genetic Architecture of Lifespan. Cell *163*, 1539-1554.

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. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.

Wang, R., Taylor, A.B., Leal, B.Z., Chadwell, L.V., Ilangovan, U., Robinson, A.K., Schirf, V., Hart, P.J., Lafer, E.M., Demeler, B.*, et al.* (2010). Polycomb group targeting through different binding partners of RING1B C-terminal domain. Structure *18*, 966-975.

Wang, X., Paucek, R.D., Gooding, A.R., Brown, Z.Z., Ge, E.J., Muir, T.W., and Cech, T.R. (2017). Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. Nat Struct Mol Biol *24*, 1028-1038.

Weng, N.P., Araki, Y., and Subedi, K. (2012). The molecular basis of the memory T cell response: differential gene expression and its epigenetic regulation. Nat Rev Immunol *12*, 306-315.

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.

Winter, G.E., Buckley, D.L., Paulk, J., Roberts, J.M., Souza, A., Dhe-Paganon, S., and Bradner, J.E. (2015). DRUG DEVELOPMENT. Phthalimide conjugation as a strategy for in vivo target protein degradation. Science *348*, 1376-1381.

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. Nature genetics *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.

Xie, H., Xu, J., Hsu, J.H., Nguyen, M., Fujiwara, Y., Peng, C., and Orkin, S.H. (2014). Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. Cell Stem Cell *14*, 68-80.

Xu, J., Shao, Z., Li, D., Xie, H., Kim, W., Huang, J., Taylor, J.E., Pinello, L., Glass, K., Jaffe, J.D.*, et al.* (2015). Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. Mol Cell *57*, 304-316.

Yakushiji-Kaminatsui, N., Kondo, T., Hironaka, K.I., Sharif, J., Endo, T.A., Nakayama, M., Masui, O., Koseki, Y., Kondo, K., Ohara, O.*, et al.* (2018). Variant

PRC1 competes with retinoic acid-related signals to repress Meis2 in distal forelimb bud. Development.

Yang, X.P., Jiang, K., Hirahara, K., Vahedi, G., Afzali, B., Sciume, G., Bonelli, M., Sun, H.W., Jankovic, D., Kanno, Y.*, et al.* (2015). EZH2 is crucial for both differentiation of regulatory T cells and T effector cell expansion. Sci Rep *5*, 10643.

Yang, Y., Wang, C., Zhang, P., Gao, K., Wang, D., Yu, H., Zhang, T., Jiang, S., Hexige, S., Hong, Z.*, et al.* (2013). Polycomb group protein PHF1 regulates p53-dependent cell growth arrest and apoptosis. J Biol Chem *288*, 529-539.

Youmans, D.T., Schmidt, J.C., and Cech, T.R. (2018). Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits. Genes Dev *32*, 794-805.

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

Zhang, Y., Kinkel, S., Maksimovic, J., Bandala-Sanchez, E., Tanzer, M.C., Naselli, G., Zhang, J.G., Zhan, Y., Lew, A.M., Silke, J.*, et al.* (2014b). The polycomb repressive complex 2 governs life and death of peripheral T cells. Blood *124*, 737-749.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

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.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750-756.

Zhu, J., Sammons, M.A., Donahue, G., Dou, Z., Vedadi, M., Getlik, M., Barsyte-Lovejoy, D., Al-awar, R., Katona, B.W., Shilatifard, A.*, et al.* (2015). Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. Nature *525*, 206-211.

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.