

Understanding the role of Polycomb-like proteins in cellular quiescence and cancer

Rachel McCole

A thesis submitted to Trinity College Dublin
for the degree of Doctor of Philosophy

2023

Thesis Supervisor: Prof. Adrian Bracken

Chromatin Biology Laboratory,

Department of Genetics,

University of Dublin, Trinity College, Dublin 2

Thesis 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.



Rachel McCole

31st March 2021

Date

Abstract

Polycomb proteins are a class of chromatin modifying complexes that are essential regulators of development, differentiation and maintenance of cell lineage. Polycomb proteins exist in two large, distinct, multimeric complexes, Polycomb repressive complex 1 and 2 (PRC1 & PRC2). PRC2 is responsible for all mono-, di- and trimethylation of lysine 27 of histone H3 (H3K27) and is composed of the core containing, EZH1 or EZH2, together with SUZ12, EED and RBBP4/7. Several studies have also characterised several accessory proteins, JARID2, AEBP2, PALI1, EPOP, PHF1/MTF2/PHF19 (alternatively referred to as PCL1/2/3) and recently EZHIP which interact with the core complex to regulate PRC2 activity and function in the recruitment of the complex to chromatin. EZH2 and the Polycomb-like proteins, MTF2 and PHF19, are regulated by E2F transcription factors and are highly expressed in rapidly growing cells, whereas PHF1 is reported to be a p53 target and found to be primarily expressed in slower growing or quiescent cells. Quiescence (also referred to as G0) is a reversible cellular state during which cells are not proliferating but ready to re-enter the cell cycle upon external stimulus. In this PhD thesis, I have characterised the role of Polycomb-like proteins in cellular quiescence and cancer. Firstly, I have analysed the assembly of a PHF1-p53 complex and the genome-wide localisation of p53 in quiescent human fibroblasts. I have also identified a distinct form of PRC2, called G0-PRC2, which predominates in quiescent or non-proliferating cells and is refractory to PRC2 inhibition or degradation. Finally, using a PRC2-focused CRISPR-Cas9 screen, I have identified regions within PHF1 of functional dependency. Taken together, I believe that these results will contribute to a better understanding of Polycomb proteins during cellular quiescence and raise the intriguing question of quiescent cells being insensitive to PRC2 inhibition or degradation suggesting a cautionary note in the treatment of cancers with PRC2-targeting drugs.

Publications

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

Healy, E*, **McCole, R***, Brien G.L., Wang, C., Monger, C., Neikes, H.K., Vermeulen, M., Potjewyd, F., James, L.I. and Bracken, A.P (2021). Quiescent cells are refractory to PRC2 inhibition. *In preparation*.

**Joint first author*

Acknowledgements

First and foremost, I would like to thank my supervisor Adrian Bracken for giving me the opportunity to do a PhD in his lab and for his guidance and support over the course of the last four years. I would also like to extend my thanks to Claudio Ciferri, Edward Kraft and the team at Genentech for supplying baculovirus reagents and the fantastic collaboration opportunity during this project. Thanks are also due to Hannah Neikes and Michiel Vermeulen in Radboud University for their assistance with the Mass Spectrometry experiments and analysis. To Chun-Wei (David) Chen and his lab at the City of Hope, Los Angeles for generation of the PRC2 sgRNA library plasmids in addition to the processing of the CRISPR screen data.

I wish to thank the Irish Research Council for funding the work carried out in this thesis and to the Inspiring generations fund in Trinity College Dublin for the additional financial aid in the last few months.

To all members, past and present, of the Bracken lab for all of the support, advice and helpful discussions over the last few years. To Eric Conway, James Nolan and Marlena Mucha, for all of the assistance and advice with the CRISPR screens. To Craig Monger and Anthony Doherty, for their patience in answering all of my bioinformatics-related questions! I would particularly like to thank Evan Healy for his mentorship and kindness over the years. I am indebted to you for all of the advice and help with everything from starving fibroblasts in tissue culture to making bigwigs! ☺ The quiescence-related projects wouldn't be where they are today without your support and I am so grateful for all that you've done!

A special thanks to Eleanor Glancy, Marlena Mucha and Orla Deevy, for all of the coffee & chocolate runs, memes and many discussions (both scientific or otherwise ☺) over the years. It's been an absolute pleasure to work alongside you all and I couldn't have asked for better friends during this PhD!

I also wish to thank my parents, and my sister, Natalie for their support and love throughout the years. A special mention must also go to my cat Felix. Not often does

a cat get a mention in a thesis but his love and support has been instrumental in getting me through the last few years. Last but by no means least, Stephen. I am forever grateful for your love and encouragement throughout the years.

Table of contents

THESIS DECLARATION	I
ABSTRACT	III
PUBLICATIONS	IV
ACKNOWLEDGEMENTS.....	V
TABLE OF CONTENTS	VII
LIST OF FIGURES.....	X
LIST OF TABLES.....	XIII
ABBREVIATIONS.....	XIV
CHAPTER 1: INTRODUCTION	1
1.1 CELL FATE AND IDENTITY	3
1.2 CHROMATIN STRUCTURE	5
1.3 POST TRANSLATIONAL MODIFICATIONS.....	7
1.4 CHROMATIN MODIFYING ENZYMES	9
1.5 POLYCOMB REPRESSIVE COMPLEXES.....	11
1.5.1 PRC1	12
1.5.2 PRC2	15
1.6 RECRUITMENT OF POLYCOMB REPRESSIVE COMPLEX	24
1.7 MUTATIONS IN PRC2 IN CANCER	28
1.8 STRUCTURE AND FUNCTION OF TP53.....	30
1.9 THE PRC2-INDEPENDENT ROLE OF PHF1 DURING QUIESCENCE	32
1.10 AIMS OF THESIS	35
CHAPTER 2: MATERIALS AND METHODS.....	37
2.1 TISSUE CULTURE METHODS	39
2.1.1 Cell lines and growth media.....	39
2.1.2 Production of lentivirus	40
2.1.3 Generation of stable cell lines	40
2.1.4 Calculation of Cas9 activity in cancer cell lines	41
2.1.5 PRC2 library screen in cancer cell lines.....	41
2.1.6 Validation of PRC2 library screen	43
2.2 PROTEIN METHODS.....	44
2.2.1 Baculovirus production of p53 and PRC2 components	44
2.2.2 Purification of PHF1-p53 complex.....	44
2.2.3 Purification of G0-PRC2.....	45

2.2.4	<i>Preparation of whole cell protein lysates and Western blotting</i>	45
2.2.5	<i>FLAG immunoprecipitations</i>	46
2.2.6	<i>Endogenous immunoprecipitations</i>	46
2.2.7	<i>Sucrose gradients</i>	48
2.3	DNA METHODS	48
2.3.1	<i>Chromatin immunoprecipitation (ChIP)</i>	48
2.3.2	<i>Library generation for NGS of ChIP samples</i>	49
2.4	RNA METHODS	50
2.4.1	<i>RNA preparation for RT-PCR analysis</i>	50
2.4.2	<i>RNA preparation for RNA-sequencing</i>	50
2.5	BIOINFORMATICS METHODS	50
2.5.1	<i>Analysis of ChIP-sequencing data</i>	50
2.5.2	<i>Analysis of RNA-sequencing data</i>	51

CHAPTER 3: CHARACTERISATION OF THE PHF1-P53 COMPLEX AND THE GENOME-WIDE LOCALISATION OF P53 IN QUIESCENT CELLS..... 55

3.1	INTRODUCTION	57
3.2	RESULTS	59
3.2.1	<i>The PHF1-p53 complex in human cells is smaller and less stable than PHF1-PRC2</i>	59
3.2.2	<i>Recombinant PHF1-p53 complex is estimated to be 200 kDa</i>	61
3.2.3	<i>p53 is enriched on the promoters of p53 cell cycle target genes following serum starvation of primary human fibroblasts</i>	66
3.2.4	<i>p53 is significantly enriched at target sites in DNA-damaged and Quiescent primary human fibroblasts and less so in DMSO-treated and cycling</i>	68
3.2.5	<i>p53 is significantly enriched at target sites in DNA-damaged and Quiescent TIG3-T immortalised fibroblasts and less so in DMSO-treated and cycling</i>	70
3.2.6	<i>Genes related to cell cycle inhibition are elevated in quiescent cells while cell division and DNA replication genes are elevated in cycling TIG3-T upon addition of serum</i>	70
3.2.7	<i>Serum starvation of TIG3-T cells shifts the expression of PRC2 components from EZH2/PHF19 in cycling cells to EZH1/PHF1 in quiescent cells</i>	73
3.3	DISCUSSION	77

CHAPTER 4: BIOCHEMICAL ANALYSIS OF PRC2 IN QUIESCENT CELLS..... 81

4.1	INTRODUCTION	83
4.2	RESULTS	86
4.2.1	<i>Purification of PRC2 lacking SUZ12</i>	86
4.2.2	<i>High salt buffer and low temperature improve the purification of G0-PRC2 subunits</i>	88
4.2.3	<i>The composition of PRC2 in quiescent and cycling primary human fibroblasts</i>	90
4.2.4	<i>Evaluation of EED IP-MS for identification of PRC2 accessory proteins in quiescent and cycling cells</i>	94

4.2.5	<i>Evaluation of EZH1/2 IP-MS for identification of PRC2 accessory proteins in quiescent and cycling cells</i>	96
4.2.6	<i>EZH1/PHF1-PRC2 is likely the most abundant PRC2 within quiescent TIG3-T</i>	99
4.2.7	<i>Degradation of EED destabilises G0-PRC2 in human fibroblasts</i>	101
4.2.8	<i>Quiescent human fibroblasts can re-enter the cell cycle upon inhibition of PRC2</i>	103
4.2.9	<i>Serum starvation of certain cancer cell lines is sufficient to induce quiescence</i>	105
4.2.10	<i>Degradation of PRC2 in thyroid cancer cell lines destabilise the complex but does not reduce H3K27me3 levels</i>	107
4.3	DISCUSSION	108
CHAPTER 5: ANALYSING THE DEPENDENCY OF A PANEL OF CANCER CELL LINES ON PRC2 SUBUNITS		111
5.1	INTRODUCTION	113
5.2	RESULTS	116
5.2.1	<i>Selection of PHF1/MTF2/PHF19 -dependent cancer cell lines</i>	116
5.2.2	<i>Generation of Cas9-expressing cancer cell lines and testing Cas9 cutting efficiency</i>	116
5.2.3	<i>Highlighting dependencies on core PRC2 in Ewing Sarcoma and thyroid papillary carcinoma cancer cell lines</i>	120
5.2.4	<i>PHF1 is a predicted dependency in Ewing Sarcoma and thyroid carcinoma cancer cell lines</i>	121
5.2.5	<i>CRISPR tiling sgRNA targeting amino acids 19-27 and 125-133 of PHF1 deplete in A673 and BCPAP but not CAL120 or Pfeiffer</i>	123
5.2.6	<i>Validation of dependency on amino acids 19-27 and 125-133 in A673, BCPAP but not CAL120 cells</i>	126
5.2.7	<i>The regions 19-27 and 125-133 of PHF1 are highly conserved in mammalian PHF1</i>	129
5.2.8	<i>Expression of mouse PHF1 does not rescue PHF1 sgRNA depletion in A673</i>	129
5.2.9	<i>Strategy for improved rescue of PHF1 depletion in A673 cells</i>	130
5.3	DISCUSSION	134
CHAPTER 6: GENERAL DISCUSSION		137
6.1	INTRODUCTION	138
6.1.1	<i>The role of Polycomb-like proteins in the PRC2 complex</i>	138
6.1.2	<i>The PRC2-independent role of PHF1</i>	138
6.2	GENERAL DISCUSSION	139
6.2.1	<i>Summary of findings</i>	139
6.2.2	<i>PHF1 as a p53 target</i>	140
6.2.3	<i>PRC2 complex assemblies in quiescent and cycling cells</i>	141
6.2.4	<i>The role of EZH1-PRC2 in facilitating chromatin compaction</i>	142
6.2.5	<i>PHF1 as a potential target in cancer therapies</i>	143
6.2.6	<i>Further perspectives</i>	143
6.3	CONCLUSIONS	145
REFERENCES		147

List of figures

Chapter 1

Figure 1.1 Schematic of chromatin compaction

Figure 1.2 Post translational modification of histone tails

Figure 1.3 Reader, writer and eraser proteins modify the chromatin landscape.

Figure 1.4 Composition of Polycomb repressive complexes

Figure 1.5 The structure of PRC2

Figure 1.6 Schematic of domains of PRC2 core and accessory proteins

Figure 1.7 Simplified schematic of PRC1 and PRC2 hierarchal recruitment

Figure 1.8 Schematic of the p53 signalling pathway

Chapter 3

Figure 3.1 The PHF1-p53 complex is smaller and less stable than PHF1-PRC2 in human cells.

Figure 3.2 Recombinant PHF1-p53 complexes can be expressed and co-purified from insect cells.

Figure 3.3 The PHF1-p53 complex is greater than 200 kDa in size.

Figure 3.4 Purification of PHF1-p53 complex of the correct stoichiometry.

Figure 3.5 p53 is enriched on the promoters of cell cycle target genes following DNA-damage or serum starvation of primary human fibroblasts.

Figure 3.6 p53 is significantly enriched at target sites in DNA-damaged and quiescent primary human fibroblasts and less so in DMSO-treated and cycling.

Figure 3.7 p53 is significantly enriched at target sites in DNA-damaged and quiescent immortalised human fibroblasts and less so in DMSO-treated and cycling.

Figure 3.8 Genes related to cell cycle arrest are elevated in Quiescent cells while cell division and DNA replication genes are elevated in Cycling TIG3-T upon addition of serum.

Figure 3.9 Serum starvation of TIG3-T shifts the expression of PRC2 components from EZH2/PHF19 to EZH1/PHF1.

Chapter 4

Figure 4.1 Purification of recombinant human G0-PRC2.

Figure 4.2 The stability of PHF1 in G0-PRC2 increases in high salt.

Figure 4.3 Figure 4.3 High salt and temperature are critical for G0-PRC2 purification.

Figure 4.4 Overview of EED, EZH1 and EZH2 IP-MS in quiescent and cycling TIG3-T fibroblasts.

Figure 4.5 Identification of EED-interacting proteins in quiescent and cycling TIG3-T fibroblasts.

Figure 4.6 Identification of EZH1-interacting proteins in quiescent and cycling TIG3-T fibroblasts.

Figure 4.7 Identification of EZH2-interacting proteins in quiescent and cycling TIG3-T fibroblasts.

Figure 4.8 PRC2 interactions in quiescent and cycling cells are broadly correlated between RNA and protein expression.

Figure 4.9 Degradation of PRC2 destabilises the complex but does not reduce H3K27 methylations in quiescent cells.

Figure 4.10 Inhibition of PRC2 in quiescent cells does not affect their ability to re-enter the cell cycle.

Figure 4.11 Serum starvation of BCPAP is sufficient to induce quiescence.

Chapter 5

Figure 5.1 Selection of cancer cell lines which are sensitive to PRC2 depletion.

Figure 5.2 Schematic of PRC2 CRISPR tiling screen.

Figure 5.3 PHF1 is a dependency in A673 and BCPAP cell lines but not CAL120.

Figure 5.4 CRISPR tiling sgRNA targeting codons corresponding to amino acids 19-27 and 125-133 of PHF1 deplete in A673 and BCPAP cancer cell lines.

Figure 5.5 EZH2 but not PHF1 is a dependency in CAL120 and Pfeiffer cancer cell lines.

Figure 5.6 Validation of dependency on regions 19-27 and 125-133 of PHF1 in A673 and BCPAP but not CAL120 cell lines.

Figure 5.7 The regions 19-27 and 125-133 of PHF1 are highly conserved in mammalian PHF1.

Figure 5.8 Expression of mouse PHF1 does not rescue depletion of human PHF1 in A673 cells.

Figure 5.9 Strategy for improved rescue of PHF1 depletion in A673 cells.

List of tables

Table 1.1 Domains within core and accessory PRC2 proteins

Table 2.1 PCR reaction components for the amplification of gDNA

Table 2.2 PCR cycle for the amplification of gDNA

Table 2.3 List of plasmids

Table 2.4 List of antibodies

Table 2.5 List of RT-qPCR primers

Table 2.6 List of ChIP primers

Table 5.1 Types of screening methods to identify novel dependencies

Abbreviations

AEBP2 - AE binding protein 2

BCOR - BCL6 Co-repressor

BCORL1 - BCL6 Co-repressor Like 1

BMI1 - B Lymphoma Mo-MLV Insertion Region 1 Homolog

C10ORF12 - Chromosome 10 open reading frame 12

CBX - Chromobox

CCNA2 - Cyclin A2

CDKN2A - Cyclin Dependent Kinase inhibitor 2A

cDNA - complementary DNA

ChIP - Chromatin Immunoprecipitation

CpG - Cytosine-phosphate-Guanine

cPRC1 - canonical Polycomb Repressive Complex 1

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

E2F - E2f transcription factor

E3 ligase - Ubiquitin Ligase

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

G0 - Quiescence

H2A - Histone 2A

H2AK119ub1 - Histone 2A lysine 119 mono-ubiquitination

H3 - Histone 3

Hdac - Histone Deacetylase

Hox genes - Homeobox genes

IP - Immunoprecipitation

INK4A - Inhibitor of cyclin dependent Kinase 4

JARID2 - Jumonji and AT-Rich Interaction Domain Containing 2

Kdm2b - Lysine demethylase 2b

Ki67/MKI67 - Proliferation marker proteins Ki-67

LC-MS/MS - Liquid chromatography mass spectrometry

MDM2 - MDM2 proto-oncogene

me1 - mono-methylation

me2 - di-methylation

me3 - tri-methylation

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

PAL1/2 - Polycomb associated LCOR isoform 1

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

PcG – Polycomb-Group

PCL1/2/3 – Polycomb-like proteins 1/2/3 (PHF1/MTF2/PHF19)

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

RBBP4/7 - RetinoBlastoma-Binding Protein 4/7

Ring1A/B - Really Interesting New Gene 1A/B

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

Chapter 1: **Introduction**

1.1 Cell fate and identity

Despite each cell sharing the same genetic code, multicellular organisms have the capacity to assemble a variety of different cell types and tissues ranging in size, shape and function. The differentiation of cells into various tissue types is a highly regulated and complex process and the development of any multicellular organism is dependent on the integrity of this process. A key concept in determining cell fate is that not all genes are actively transcribed at a given time, and the transcriptional signature of each cell is vastly different. For example, in stem cells the expression of stem cell-associated and pluripotency genes is higher than in differentiated cell types while the inverse is true in differentiated, lineage-committed tissues. Stimulus of progenitor or stem cells drives the transition of these cells towards a more specific cell type and results in the switch in the transcriptional profiles however, importantly, the genomic sequence within each cell remains consistent irrespective of cell type (Moris *et al*, 2016). The transcriptional signature of cells is tightly regulated; the activation or repression of specific genes can be determined by a number of mechanisms, notably post-translational modification of histone proteins and the modification of underlying DNA (Jenuwein & Allis, 2001; Kouzarides, 2007; Smith & Meissner, 2013)

Although some aspects of how histone post-translational modifications (PTMs) regulate DNA accessibility remain unclear, for the most part these modifications do not directly contribute to the transcription or repression of specific genes but rather facilitate downstream events leading to higher-order chromatin relaxation or compaction (Kouzarides, 2007). The majority of histone methylation modifications are diluted two-fold by newly deposited histones following mitotic cell division and can direct, at least in part, the *de novo* deposition of modifications on nascent histones (Alabert *et al*, 2015). However, the kinetics of re-establishment of histone marks on newly deposited histones varies depending on the modification (Alabert *et al*, 2015; Margueron & Reinberg, 2010).

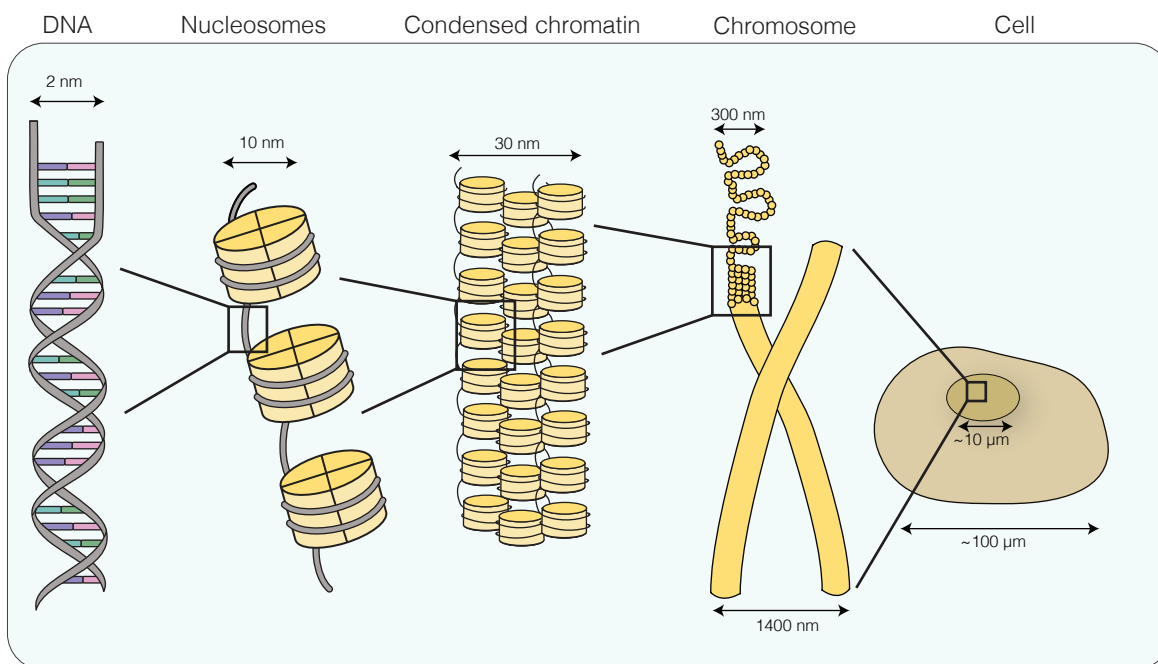


Figure 1.1 Schematic of chromatin compaction.

Negatively-charged DNA is wrapped around positively charged histones to form a DNA-protein complex known as a nucleosome. Nucleosomes are condensed into loops and folds with the resulting condensed sections of chromatin assembling into tightly compacted chromosomes.

1.2 Chromatin structure

DNA is the genetic code found in almost every organism and is responsible for carrying the instructions for cellular development and function. Transmission of this hereditary material between cells is essential for cell survival and growth. The human genome consisting of ~ 3 billion base pairs, equating to 2 metres in length, must be arranged in a highly organised fashion to fit inside the nucleus of just 5–10 µm in diameter. DNA is compacted into a DNA-protein complex known as chromatin (Figure 1.1). The “beads on a string” image of chromatin was first observed by electron microscopy and revealed a model for chromatin composition and higher levels of DNA packaging (Olins & Olins, 1974).

Chromatin is composed of repeating units of nucleosomes, which are defined as the nucleosome core particle (NCP), a flanking linker DNA of ~10-70 bp in length and a linker histone H1 (Flaus, 2011). The NCP is made up of 147 bp of DNA wrapped 1.6 times around a core histone octamer containing two copies of each histone proteins H2A, H2B, H3 and H4 and is flanked by linker DNA of varying lengths depending on tissue and species type. Histones H3 and H4 assemble into a tetramer and associate with 2 dimers of H2A and H2B to form the histone octamer core (Luger *et al*, 1997; McGinty & Tan, 2015). The linker histone H1 can bind the flanking linker DNA between nucleosomes and has been reported to induce a more structured and rigid nucleosome conformation (Bednar *et al*, 2017)

Histone H1 is not as well studied as the other histones and this may be in part due to the largely unstructured and lysine-rich sequence of H1, rendering it difficult to study in biochemical assays and has come to be known as the “forgotten histone”. Recent studies have however highlighted an important role for H1 in regulating chromatin condensation with loss of H1 associated with TAD decompaction (Willcockson *et al*, 2021; Yusufova *et al*, 2021). H1 plays an important role in promoting trimethylation of histone H3 lysine 27 (H3K27me3) with loss of H1 phenocopying knockout of EZH2 (Willcockson *et al*, 2021).

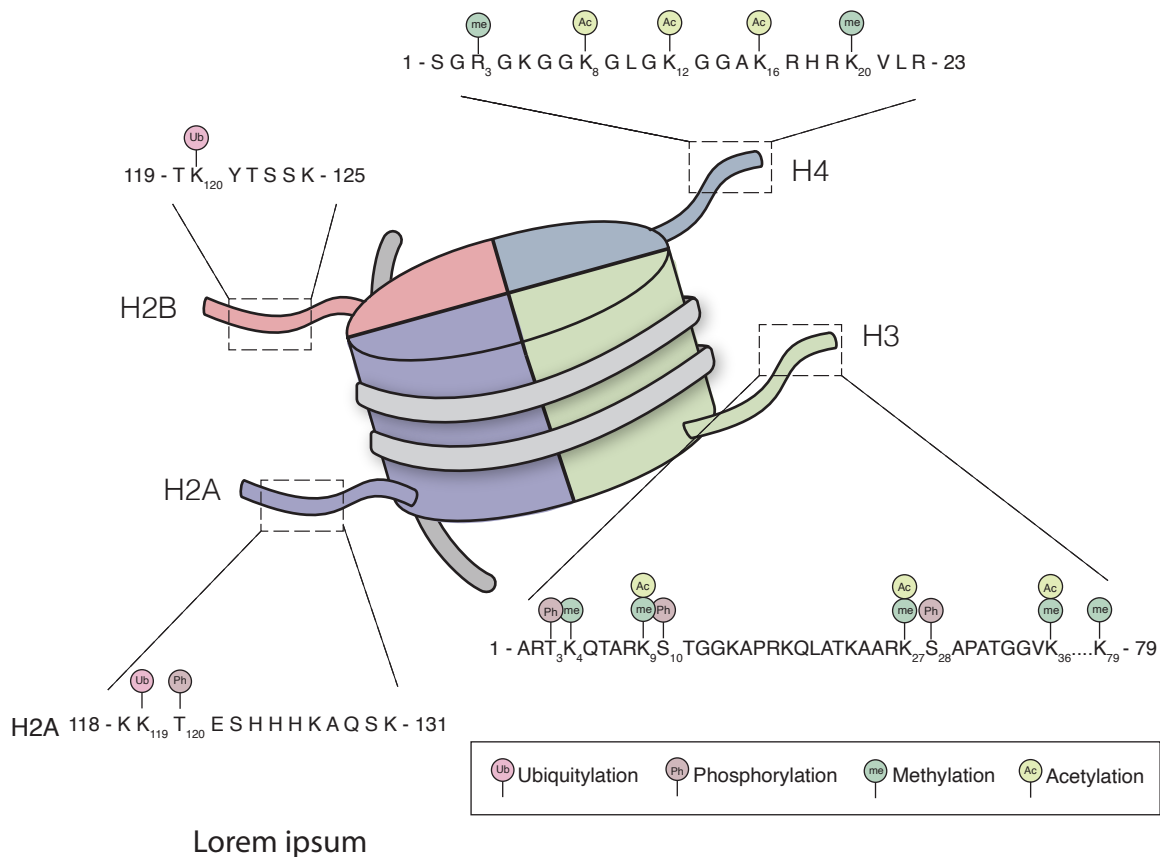


Figure 1.2 Post translational modification of histone tails.

The basic unit of chromatin is the nucleosome which consists of a histone octamer core wrapped in 147bp of DNA. The histone tails that protrude from the core unit are subject to various post translational modifications including methylation, acetylation, ubiquitylation and phosphorylation which can have a repressive or activating effect on transcription.

As mentioned above, nucleosome core particles are comprised of the four core histones enwrapped in 147 base pairs of DNA and protruding from the nucleosome core are unstructured histone tails which can be covalently modified in several places (Figure 1.2). Post translational modifications (PTMs) can modulate chromatin compaction by addition or removal of a modification which can alter gene expression by facilitating a more open, transcriptionally active chromatin, termed euchromatin, or alternatively can constrict as heterochromatin which is generally associated with reduced transcription.

1.3 Post translational modifications

Chromatin compaction can be regulated by the addition or removal of post translational modifications onto the unstructured histone tails protruding from the nucleosome. Histones can be subjected to a number of post translational modifications including, but not limited to, methylation, acetylation and phosphorylation (Figure 1.2) (Kouzarides, 2007). Acetylation of amino acid residues on histone tails neutralizes the side chain's positive charge of the amino acid and subsequently decreases the affinity for the negatively charged DNA leading to chromatin relaxation. The addition of acetylation to amino acids residues is mediated by a group of enzymatic "writer" proteins known collectively as histone acetyltransferases (HATs) whilst removal of acetyl groups is catalysed by a family of "eraser" enzyme termed histone deacetylases (HDAC) (Yang & Seto, 2007). Aside from facilitating a more open chromatin structure, acetylation of residues with histone tails can also mediate the recruitment of chromatin regulator proteins adding to the regulatory mechanism of the mark. BET family proteins, for example BRD4, are an example of chromatin readers that can recognise and to acetyl lysine via a bromo domain (Dey *et al*, 2003; Jung *et al*, 2014). Although typically associated with recognition of methylated amino acid residues, PHD finger domains have also been identified as potential readers of acetylated lysine residues (Musselman *et al*, 2009; Qiu *et al*, 2012; Zeng *et al*, 2010).

Methylation of amino acids, on the other hand, can have a transcriptive or repressive effect on chromatin depending on the target residue and the degree of methylation.

For example, monomethylation of histone lysine 27 is typically associated with open, active chromatin at gene bodies whereas trimethylation of the same residue correlates with repressed promoters of genes (Bracken *et al*, 2006; Conway *et al*, 2015; Ferrari *et al*, 2014). The addition or removal of methyl groups is performed by two classes of proteins aptly named histone methyltransferases (HMT) or histone demethyltransferases (HDMT). EZH2, the catalytic component of Polycomb repressive complex 2, is a well characterised HMTase that catalyses mono-, di- and tri-methylation of histone 3 lysine 27 (H3K27) (Højfeldt *et al*, 2018) whilst UTX, a member of the MLL3/4/COMPASS complex, removes di- and tri-methylation from H3K27 (Figure 1.3) (Agger *et al*, 2007).

Interestingly, the colocalization of both repressive and active histone marks has been noted on the promoters of developmental genes in embryonic stem cells. So-called “bivalent chromatin” is defined by the existence of two such opposing histone modifications within the same region. Normally, these opposing histone marks occur on chromatin independently of each other so their co-occurrence at promoters of developmental genes in ES cells was an unexpected observation (Azuara *et al*, 2006; Bernstein *et al*, 2006).

The H3K4me3 modification is typically associated with regions of transcriptionally active chromatin however this mark of open chromatin can also co-exist with the repressive histone modification H3K27me3 on a large cohort of developmental genes (Azuara *et al*, 2006; Bernstein *et al*, 2006; Mikkelsen *et al*, 2007). Despite the presence of H3K27me3, bivalent marked promoters maintain expression of developmental genes at low levels in ES cells until differentiation to another cell type. The simultaneous presence of opposing histone marks primes ES cells for the rapid switch in the expression of developmental genes during differentiation (Pan *et al*, 2007; Voigt *et al*, 2013; Zhao *et al*, 2007). Depletion of H3K27me3 in EED knockout ES cells resulted in premature expression of neural-specific genes in undifferentiated ES cells highlighting the importance of H3K27me3 in balancing the expression of bivalently-marked genes (Azuara *et al*, 2006).

Similar to bivalent promoters, poised enhancers are marked by presence of two transcriptionally opposing modifications for example H3K4me1 and H3K27me3 (Rada-Iglesias *et al*, 2011). Enhancer elements are short stretches of DNA that mediate interactions with transcription-regulating factors which in turn facilitates the expression of particular target genes. Enhancers exist in an active, primed or poised state; active enhancers are enriched with both H3K4me1 and H3K27Ac and assist in increasing gene expression whilst primed enhancers are marked with only H3K4me1 and represent an intermediary state neither facilitating or inhibiting transcription (Calo & Wysocka, 2013; Creyghton *et al*, 2010). Poised enhancers are similarly enriched with H3K4me1 but their chromatin signature differs from active enhancers in that there is an enrichment of tri-methylated H3K27 and absence of acetylation (Rada-Iglesias *et al*, 2011). Similar to the mechanism of bivalent promoters, poised enhancers can drive expression of proximal genes upon differentiation of ES cells to a lineage-committed cell type (Cruz-Molina *et al*, 2017). Bivalent chromatin has, for the most part, been characterised in ES cells but one study has defined a subset of cell type-specific poised domains observed solely in a differentiated neural progenitor cell lines (Mohn *et al*, 2008).

1.4 Chromatin modifying enzymes

Readers, writers and erasers are classes of proteins that have been grouped based on their ability to regulate histones. Writer proteins are enzymes that are capable of depositing PTMs onto histones. An example of a writer protein is the histone methyltransferase enzyme SUV39H1 which was one of the first histone methyltransferases and found to selectively methylates lysine 9 of histone H3 (H3K9me3) (Rea *et al*, 2000). Post translational modifications are not thought to be a permanent feature on chromatin and most can be removed by a class of histone regulatory proteins termed “erasers”. Demethylases, deubiquitinases and deacetylases are some of the classes of eraser proteins that can remove specific

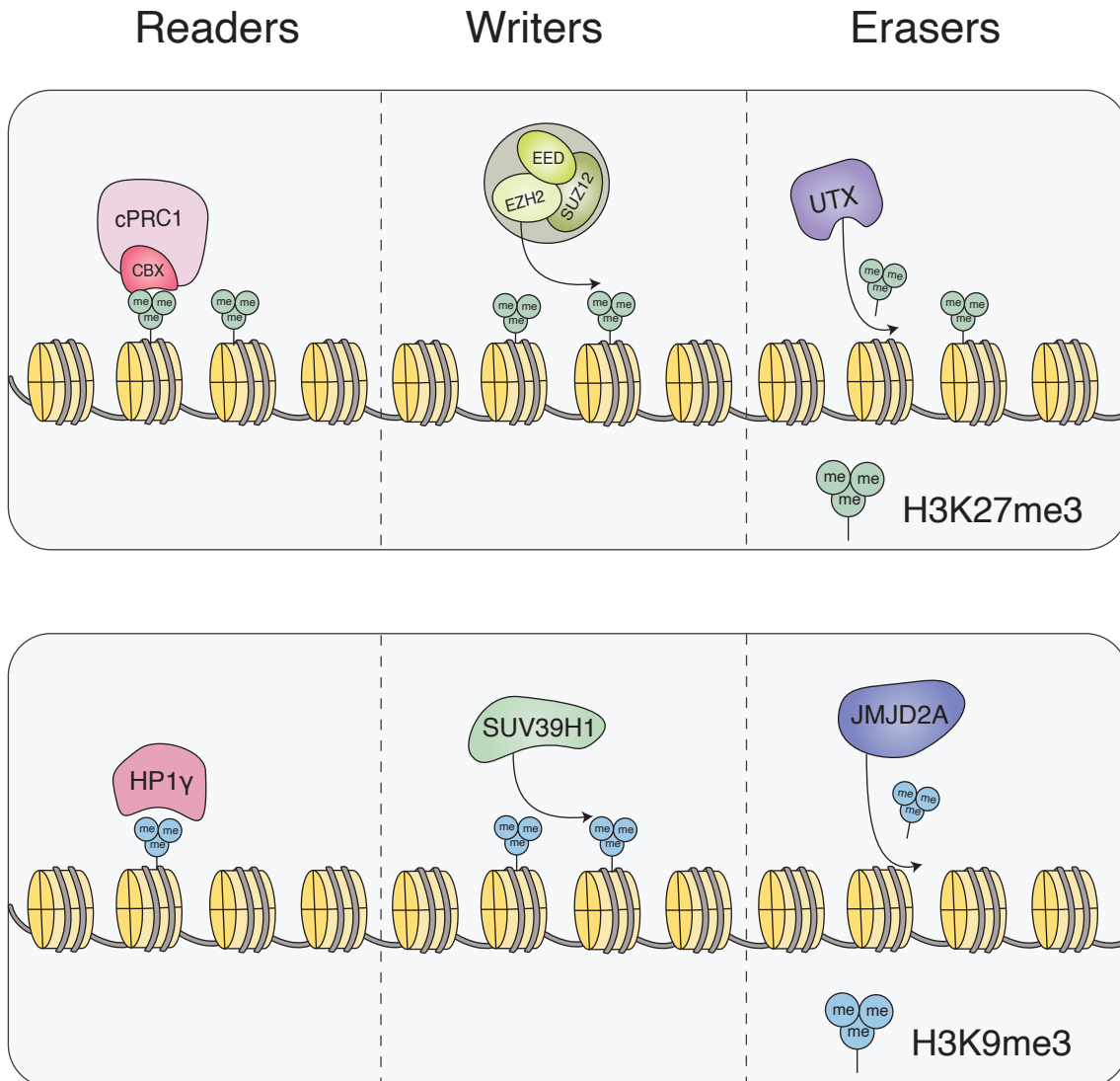


Figure 1.3 Reader, writer and eraser proteins modify the chromatin landscape.

Reader proteins (e.g. CBX/HP1 γ) contain domains capable of recognising specific post-translational modifications. Writer proteins (e.g. EZH2/SUV39H1) are catalytic enzymes responsible for addition of post translational modifications to chromatin. Eraser proteins (e.g. UTX/JMJD2A) facilitate removal of histone modifications.

PTMs from histone substrates. An example of an eraser protein is the histone demethylase enzyme UTX which demethylates tri-methylated lysine 27 on histone H3 (Agger *et al.*, 2007). Reader proteins contain “reader domains”, such as bromodomains, chromodomains and PHD fingers, which are capable of recognising and binding to specific histone marks. An example of a reader protein is CBX8, a subunit within canonical PRC1 (cPRC1), that binds to PRC2-deposited H3K37me3 via a highly conserved chromodomain (Fischle *et al.*, 2003; Min *et al.*, 2003) or the HP1 protein that recognise SUV39H1-mediated H3K9me3 via a chromodomain (Figure 1.3) (Bannister *et al.*, 2001).

1.5 Polycomb Repressive Complexes

Polycomb group proteins (PcG) were first identified as repressors of homeotic (Hox) genes in *Drosophila melanogaster* around forty years ago (Lewis, 1978). Early studies reported PcG as an essential regulator of Hox genes and more recent studies confirm that this role is conserved in other organisms including plants, mice and humans (Schuettengruber *et al.*, 2017). PcG proteins have been identified as essential regulators of development, differentiation and maintenance of cell lineage (Bracken *et al.*, 2006; Conway *et al.*, 2015; Pasini *et al.*, 2004). Cellular identity is maintained by binding of PcG proteins to specific lineage genes thus repressing transcription (Pasini *et al.*, 2007). PcG proteins exist in two large, distinct, multimeric complexes, Polycomb repressive complex 1 and 2 (PRC1 & PRC2). Both complexes have enzymatic activity that is associated with transcriptional repression of genes, with PRC1 catalysing the mono-ubiquitination of lysine 119 of histone H2A (H2AK119Ub) and PRC2 responsible for all mono- di- and tri-methylation of H3K27 (Cao *et al.*, 2005; Højfeldt *et al.*, 2018; Müller *et al.*, 2002). H3K27me3 is found at the promoters of repressed genes and co-localises with PRC2 at these sites while H3K27me2 and H3K27me1 marks are associated at intergenic regions and gene bodies of actively transcribed genes, respectively. Interestingly, while PRC2 is responsible for catalysing H3K27me2 and H3K27me1 marks, the complex does not stably associate with these marks on chromatin (Ferrari *et al.*, 2014; Lee *et al.*, 2015). H3K27me2 is thought to be the default mark for H3K27 acting as a repressive blanket across the genome to prevent incorrect activation of non-cell-type-specific

enhancers by protecting from K27 acetylation (Conway *et al.*, 2015). H3K27me1 is predominantly located within actively transcribed gene bodies however, PRC2 does not co-localise within these regions nor at regions containing H3K27me2. It has been suggested that PRC2 generates H3K27me1/2 marks more readily than H3K27me3 and the interaction with chromatin is transient (Højfeldt *et al.*, 2018; Sneeringer *et al.*, 2010).

1.5.1 PRC1

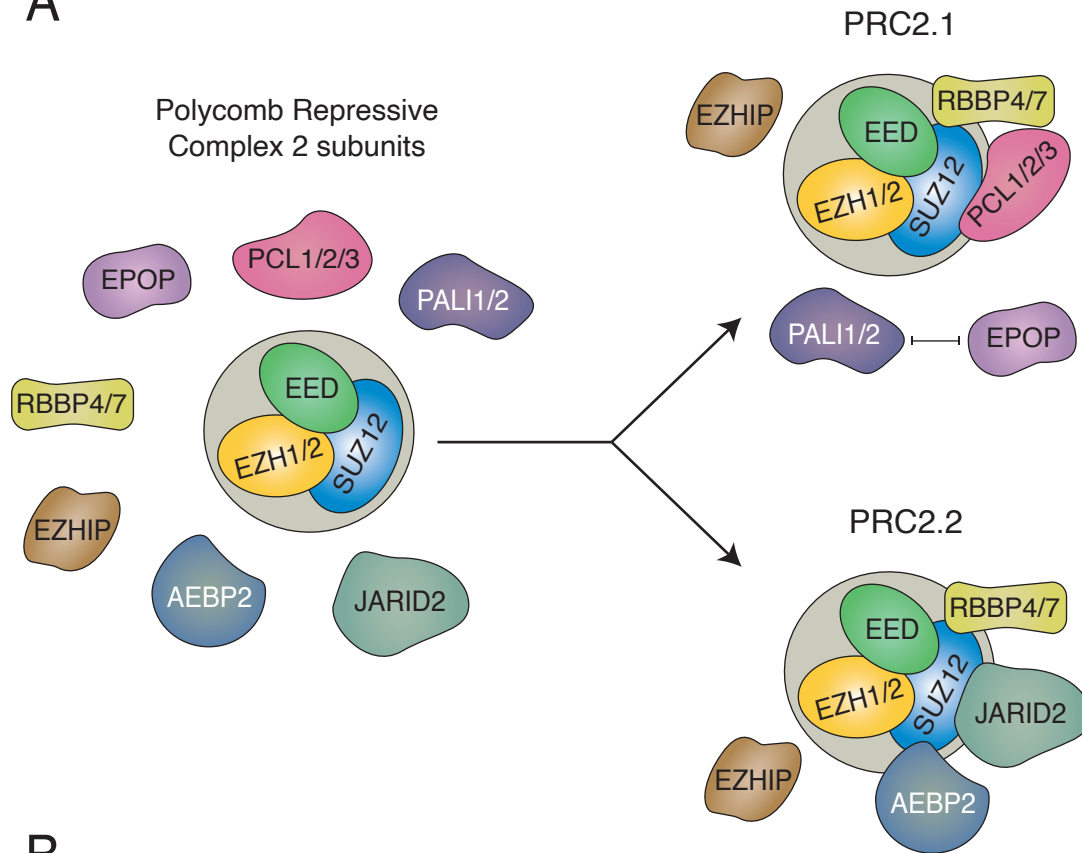
Purification of *Drosophila* PRC1 was achieved in 1999 using an exogenously-tagged approach to isolate the complex from a fly cell line. The complex was identified as a large multimeric complex consisting of the catalytic Ring/Sce protein along with Psc, Pc, Ph and Scm and acted as an inhibitor to the SWI/SNF complex *in vitro* (Shao *et al.*, 1999). Further studies in vertebrates, identified homologs of these *Drosophila* core proteins along with additional interactors (Wang *et al.*, 2004a). Mammalian PRC1 complex exists in a variety of sub-assemblies and depending on the composition of accessory proteins, can more or less catalytically active (Taherbhoy *et al.*, 2015). One of the E3-ubiquitin ligase, RING1A or RING1B, together with one of six PCGF proteins assemble as the heterodimer core of PRC1 (Gao *et al.*, 2012). All PRC1 complexes share a common Ring-PcGF core however the complex can be subdivided into six groups, PRC1.1-1.6 depending on the associated PCGF. Additionally, PRC1 can also be categorized as two forms, canonical and noncanonical (also referred to as variant) PRC1, depending on the subunit assembly. For the most part, noncanonical PRC1 (PRC1.1/3/5/6), contains one of PCGF1/3/5/6, RING1A/B, either RYBP or YAF2 along with additional subtype-specific interactors for example BCOR and BCORL1 assemble within PRC1.1 (Farcas *et al.*, 2012), whilst FBRS and FBRSL1 interact with PRC1.3/5 (Figure 1.4) (Gao *et al.*, 2012). Expression of PCGF proteins varies depending on cell-type with PCGF2/6 the most abundant of the PCGF in mouse ES cells, in contrast to PCGF4/5 which is lowly expressed (Kloet *et al.*, 2016; Scelfo *et al.*, 2019). Interestingly, recent studies have shown that while most subcomplexes of ncPRC1 are localised at CpG-rich promoter elements, the PCGF subunits mediate complex association at different target sites (Scelfo *et al.*, 2019). The inclusion of RYBP or YAF2 within ncPRC1 is mutually exclusive with the canonical subunits,

CBX/PHC/SCM, and although both RYBP/YAF2 and CBX-containing complexes can mediate compaction of chromatin, only RYBP/YAF2-PRC1 stimulates the ubiquitinating activity of the complex (Gao *et al.*, 2012).

The various subunits that assemble within ncPRC1 along with the tissue-specific expression of complex members further add to the complexity of the various roles of ncPRC1 and the composition and function of complexes can fluctuate depending on cell type. ncPRC1.1 (containing the subgroup-specific interactors BCOR, BCORL, Skp1, USP7 and Kdm2b) is the most well studied of the variant complexes and recruitment to target sites is mediated by binding to unmethylated CpG islands through the CxxC motif of Kdm2b (Farcas *et al.*, 2012; Wu *et al.*, 2013). A recent study has suggested that vPRC1.3/5 initiates recruitment of Polycomb to the inactive X chromosome (Almeida *et al.*, 2017). Following recruitment to chromatin, vPRC1 mediates deposition of ubiquitin to H2AK119 which is read by the UIM domain of JARID2 within PRC2 (Cooper *et al.*, 2016) and subsequently facilitates recruitment of canonical PRC1 via CBX-mediated recognition of PRC2-deposited H3K27me3 (Fischle *et al.*, 2003; Min *et al.*, 2003).

Conversely, despite containing the E3 ubiquitin ligase RING1A or RING1B, canonical PRC1 (cPRC1) exhibits less catalytic activity than ncPRC1 and functions primarily as a mediator of chromatin compaction (Taherbhoy *et al.*, 2015). Consisting of RING1A/B, either PCGF2/4, SCM, PHC along with any of CBX2/4/6/7/8, cPRC1 is recruited to chromatin by binding of PRC2-deposited H3K27me3 by the highly conserved N-terminal chromodomain of CBX (Figure 1.4) (Fischle *et al.*, 2003; Min *et al.*, 2003). The exact composition of cPRC1 varies depending on the cell-type for example PCGF2 along with CBX7 are upregulated in ES cells yet in a differentiated cell model, PCGF4 and CBX8 expression is elevated whilst PCGF2 and CBX7 are downregulated (Kloet *et al.*, 2016; Morey *et al.*, 2012).

A



B

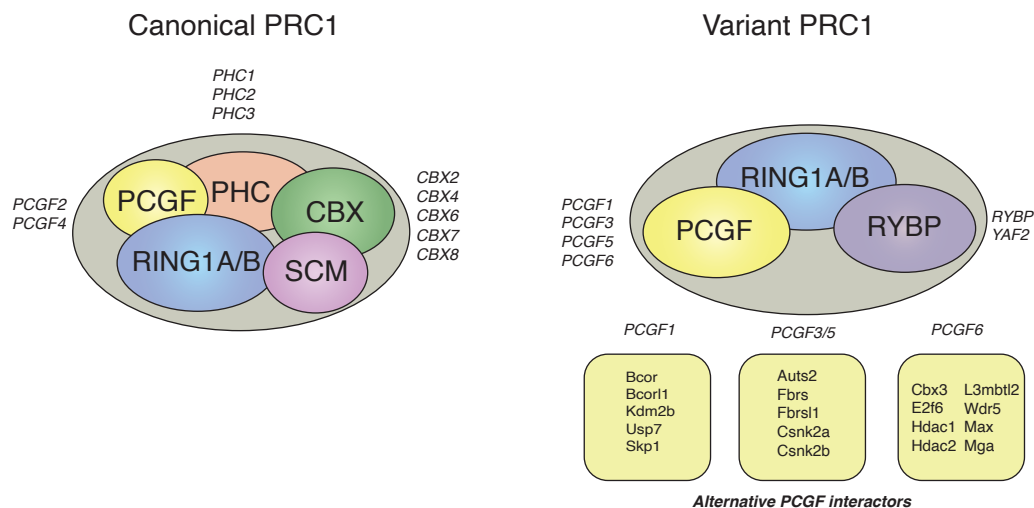


Figure 1.4 Composition of Polycomb repressive complexes.

(A) PRC2 is classified into two subcomplexes, namely PRC2.1 and PRC2.2.

(B) PRC1 exists as two distinct, multi-subunit complexes; canonical and variant PRC1 (cPRC1 & vPRC1).

Looping and condensation of chromatin are regulated via a sterile alpha motif (SAM) within PHC sub-unit containing that is sufficient for oligomerisation with PHC of adjacent cPRC1 located on promoters of nearby genes (Isono *et al*, 2013). Taken together, the multimerisation of cPRC1 results in the formation of Polycomb bodies which are microscopically visible “speckles” of condensed chromatin, enriched in Polycomb-associated proteins and PcG target genes (Isono *et al.*, 2013).

1.5.2 PRC2

PRC2-associated subunits assemble into two distinct complexes, termed PRC2.1 and PRC2.2 with the former containing Polycomb-like proteins (PHF1/MTF2/PHF19) and either PALI1/2 or EPOP and the latter consisting of JARID2 and AEBP2 (Figure 1.4) (Alekseyenko *et al*, 2014; Hauri *et al*, 2016). The catalytic core PRC2, consisting of EED, EZH1/2 and SUZ12, facilitate the methylation of lysine 27 of Histone H3 (H3K27) via the C-terminal SET domain of EZH1/2. PRC2 have been identified as the sole histone methyltransferase for all mono-, di- and trimethylation of H3K27 in mouse ES cells (Højfeldt *et al.*, 2018). It has been debated whether RBBP4/7 is considered a core component of the complex as EED, EZH2 and SUZ12 are sufficient to methylate Histone H3 in the absence of RBBP4/7 (Cao & Zhang, 2004). Depletion of one of the core subunits results in global reduction or complete loss of H3K27me1/2/3 in mouse ES cell models (Højfeldt *et al.*, 2018; Lavarone *et al*, 2019; Montgomery *et al*, 2005; Pasini *et al.*, 2007). In addition to reduced methylation activity, loss of one of the core subunits greatly impairs the overall stability of the remaining complex members (Højfeldt *et al.*, 2018; Montgomery *et al.*, 2005; Pasini *et al.*, 2004). Loss of one of the trimeric core subunits of PRC2 does not impair the morphology or proliferation of undifferentiated ES cells, however, PRC2-deficient cells fail to correctly differentiate (Lavarone *et al.*, 2019; Oksuz *et al*, 2018; Pasini *et al.*, 2007).

The histone methyltransferase components, EZH1 and EZH2, are mutually exclusive within PRC2 and exhibit distinct expression profiles with EZH2 upregulated in proliferating cells while EZH1 is more abundant in slower growing, terminally differentiated cells (Margueron *et al*, 2008; Oliviero *et al*, 2016; Xu *et al*,

2015). While EZH1 and EZH2 are both capable of generating H3K27 mono-, di- and trimethylation, the catalytic activity of EZH1 is less than EZH2 both *in vitro* and *in vivo* (Højfeldt *et al.*, 2018; Lavarone *et al.*, 2019; Margueron *et al.*, 2008). Loss of EZH1 in mouse ES cells does not impair the methylation signature of H3K27 suggesting either a potential compensatory mechanism by EZH2 (Højfeldt *et al.*, 2018; Lavarone *et al.*, 2019). EZH1-containing complexes have a higher affinity than EZH2 for both nucleosome substrates or DNA (Lee *et al.*, 2018; Margueron *et al.*, 2008; Son *et al.*, 2013) and recent studies have also highlighted an EZH1-PRC2 dimerisation model that promotes chromatin compaction of *in vitro* nucleosomal arrays (Grau *et al.*, 2021). These two paralogues arose through a gene duplication event and even though the enzymatic SET domains of EZH1 and EZH2 share approximately 94% amino acid sequence identity, overall the proteins share only 65% sequence identity suggesting that the functional difference between EZH1/2 may be derived from the regions outside of the SET domain (Grau *et al.*, 2021; Son *et al.*, 2013; Völkel *et al.*, 2019). Taken together, EZH1 may play a more direct role in transcriptional repression by mediating chromatin compaction and PcG body formation.

As mentioned previously, EED is crucial for normal differentiation of mouse ES cells to EBs and in mouse development (Faust *et al.*, 1995; Oksuz *et al.*, 2018). The WD40 repeat domains of EED assemble into a 7-bladed, β -propeller structure and act as a rigid scaffold within the complex (Figure 1.5) (Jiao & Liu, 2015). Crystal structure analysis of EED revealed an aromatic cage on the top surface that mediates interaction with H3K27me3 and allosterically enhances activity of the complex further propagating the mark on neighbouring nucleosomes (Margueron *et al.*, 2009; Xu *et al.*, 2010). The binding of EED to H3K27me3 alters the conformation of the stimulation response motif (SRM) of EZH2 from a mobile, disordered region to a more structured alpha helix which subsequently stabilises the SET domain within EZH2 thereby enhancing the activity of the complex (Jiao & Liu, 2015; Kasinath *et al.*, 2021; Kasinath *et al.*, 2018). While the EED protein itself does not display enzymatic activity, perturbation of the aromatic cage results in reduced H3K27 methylation suggesting that EED is an essential component within the

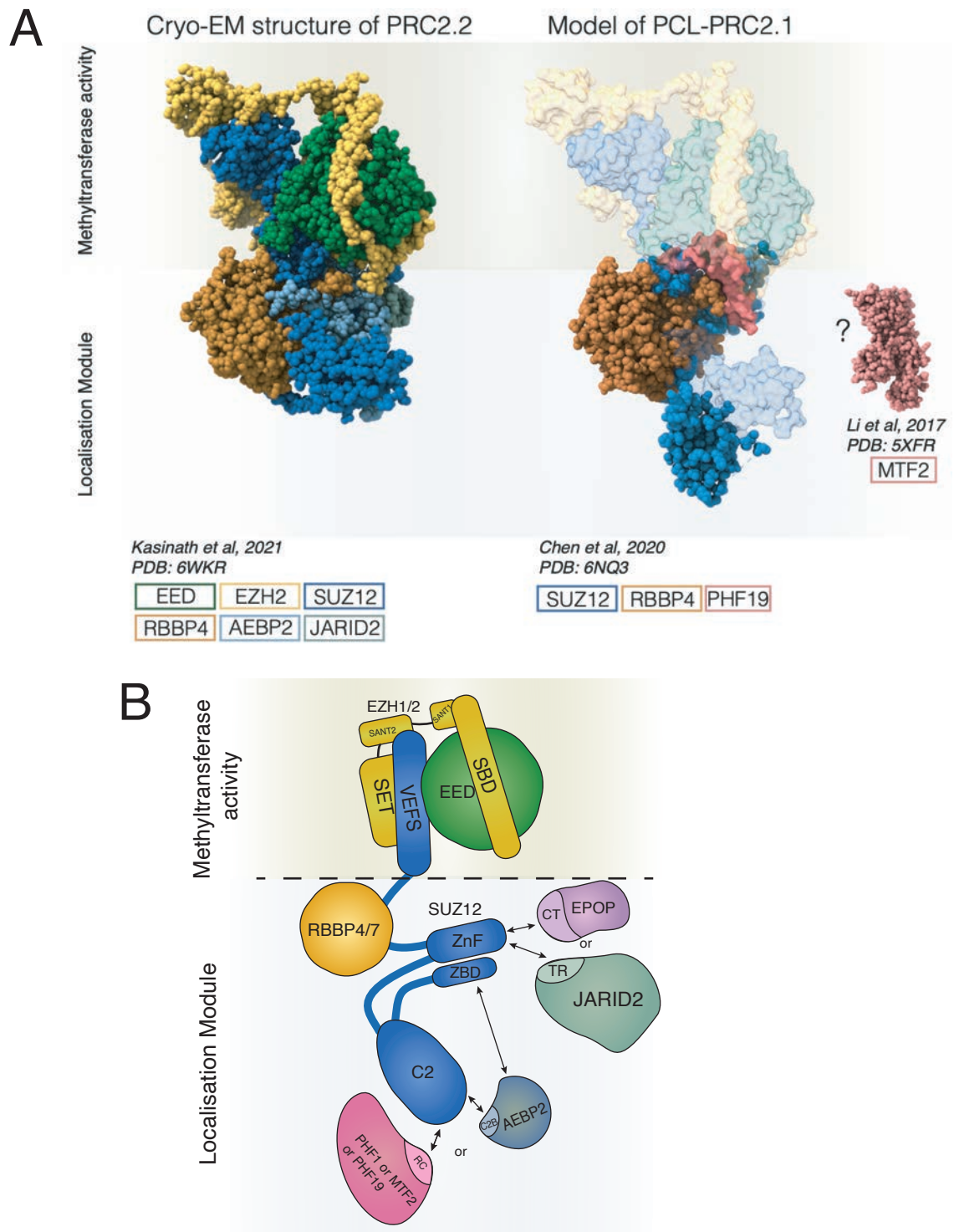


Figure 1.5 The structure of PRC2.

(A) Cryo-EM structure of PRC2.2 complex containing AEBP2 and JARID2 subunits (PDB file 6WKR: Kasinath et al, 2021). A model of PCL-PRC2 has been generated by overlaying PDB files 6WKR (PRC2.2) with 6NQR (SUZ12 RBBP4 and C-terminal domain of PHF19) (PDB file 6NQ3: Chen et al, 2020). There are no complete structures available for PRC2.1 yet however a crystal structure of amino acids 43-358 of MTF2 has been included (PDB file 5XFR: Li et al, 2017). There are no crystal or Cryo-EM structures available yet for EPOP or PALI1/2.

(B) Simplified schematic of PRC2. Highlighted within each of the accessory proteins (PHF1/MTF2/PHF19, EPOP, JARID2 and AEBP2) are the domains that have been reported to interact with SUZ12 in the literature.

complex and necessary for complete catalytic activity (Margueron *et al.*, 2009; Oksuz *et al.*, 2018).

SUZ12 plays a vital role in the structural integrity of PRC2 and acts as a platform to bridge together the enzymatic and regulatory “lobes” of the complex (Chen *et al.*, 2020; Chen *et al.*, 2018; Ciferri *et al.*, 2012; Højfeldt *et al.*, 2019; Kasinath *et al.*, 2018). The C-terminal VEFS domain of SUZ12, together with EED and EZH2, assemble as the catalytic moiety of PRC2 while domains within the N-terminal and central region of SUZ12 provide docking sites for the respective accessory proteins within the regulatory lobe (Figure 1.5) (Chen *et al.*, 2018; Youmans *et al.*, 2018). The PRC2.1-associated EPOP protein competes for binding with the PRC2.2-associated JARID2 on the N-terminally-located neck region (NR) of SUZ12 and binding of either protein is mutually exclusive (Figure 1.5). Likewise, the interaction of PRC2.1 and PRC2.2-specific proteins, PCLs and AEBP2, with SUZ12 occur independently of each other and is mediated by the central C2 domain of SUZ12. (Chen *et al.*, 2020; Chen *et al.*, 2018). The interaction between AEBP2 and SUZ12 is not only localised to the C2 domain but also within the nearby Zinc finger domain as identified by cross-linking mass spectrometry (Ciferri *et al.*, 2012; Zhang *et al.*, 2019). A binding site for PRC2.1-specific interactor, PALI1/2 has not yet been identified within SUZ12, however, given that PALI1/2 and EPOP occur independently of each other in the complex, it could be speculated that PALI1/2 competes for binding within the neck region.

The Polycomb-like proteins PHF1, MTF2, and PHF19 (alternatively known as PCL1, 2 and 3 or collectively as PCLs) are the most well studied members of PRC2.1 to date with extensive analyses highlighting mechanisms in which PCLs target the complex to chromatin. PCL proteins are a group of multidomain proteins which play an important role in both enhancing the methyltransferase activity of the complex as well as in the recruitment of the complex to chromatin. One of the earliest studies on *Drosophila* Pcl employed yeast hybridisation assays to highlight the interaction between Pcl and E(z) (O'Connell *et al.*, 2001). Expression of PCLs varies depending

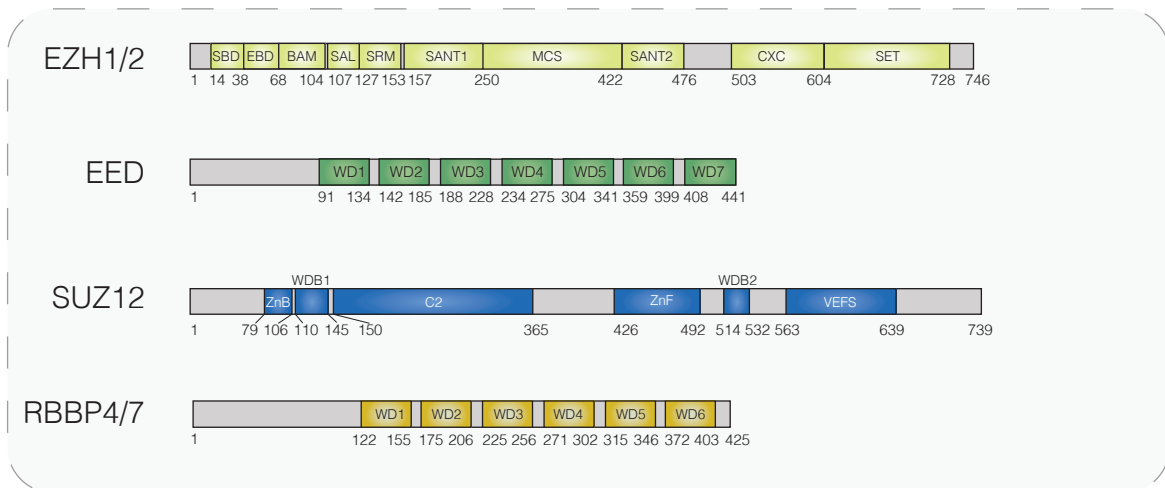
on cell type and stages of the cell cycle with the highest PHF1 expression observed in non-proliferating, quiescent and terminally differentiated cells which contrasts with MTF2 and PHF19 expression which is correlated, though not limited to, undifferentiated, proliferating ES cells (Brien *et al*, 2015; Oliviero *et al.*, 2016). Despite the extensive study of PCLs *in vivo*, the complete structure of PRC2.1 remains unsolved although the structures of various fragments of the PCL proteins have been solved (Figure 1.5). PHF1, as well as MTF2 and PHF19, form a stable complex with core members of PRC2 and were revealed to enhance the methyltransferase activity of the complex *in vitro* (Casanova *et al*, 2011; Choi *et al*, 2017; Sarma *et al*, 2008; Walker *et al*, 2010; Zhang *et al.*, 2019). PCLs contain a Tudor, two PHD domains (PHD1/2), and extended helix (EH) domain and a C-terminal reverse chromatin domain (Figure 1.6). Additionally, a binding domain 2 (BD2) has been characterised in PHF1 and has been shown to interact with the C-terminal tail of p53 and subsequently stabilise the protein thereby protecting from MDM2 degradation (Brien *et al.*, 2015). The function of this BD2 region in MTF2 or PHF19, if any, remains unknown. Structural studies and *in vitro* biochemical assays have highlighted that Tudor domain of PCLs recognises and binds to modified histone tail, namely H3K36me3 (Ballaré *et al*, 2012; Brien *et al*, 2012; Cai *et al*, 2013; Musselman *et al*, 2012). The affinity is highly specific with a single mutation of Y47, Y62 or W50 in the Tudor domain of PHF1, MTF2 or PHF19 sufficient to obliterate the interaction with H3K36me3 peptides (Brien *et al.*, 2012; Li *et al*, 2017).

EPOP (C17ORF96) has been identified as a subunit within PRC2.1 along with PHF1, MTF2 or PHF19 (Beringer *et al*, 2016; Liefke *et al*, 2016; Liefke & Shi, 2015). The interaction between EPOP and core PRC2 is mutually exclusive with that of PALI1/2 meaning these two accessory proteins cannot co-exist within the same complex (Alekseyenko *et al.*, 2014; Hauri *et al.*, 2016). Interestingly, whilst EPOP has been shown to stimulate PRC2 activity *in vitro*, it has been reported that EPOP acts as an inhibitor of the complex *in vivo* (Beringer *et al.*, 2016; Liefke *et al.*, 2016; Zhang *et al*, 2011) . PALI1 has been identified as a member of PRC2.1 and its presence in the complex is mutually exclusive with EPOP. The protein is largely disordered, however, distinct domains have been identified to interact with a number of chromatin regulators including G9A, USP11 and USP22 (Conway *et al*, 2018). It has been reported that PRC2 interacts with the G9A so PALI1 may act as the

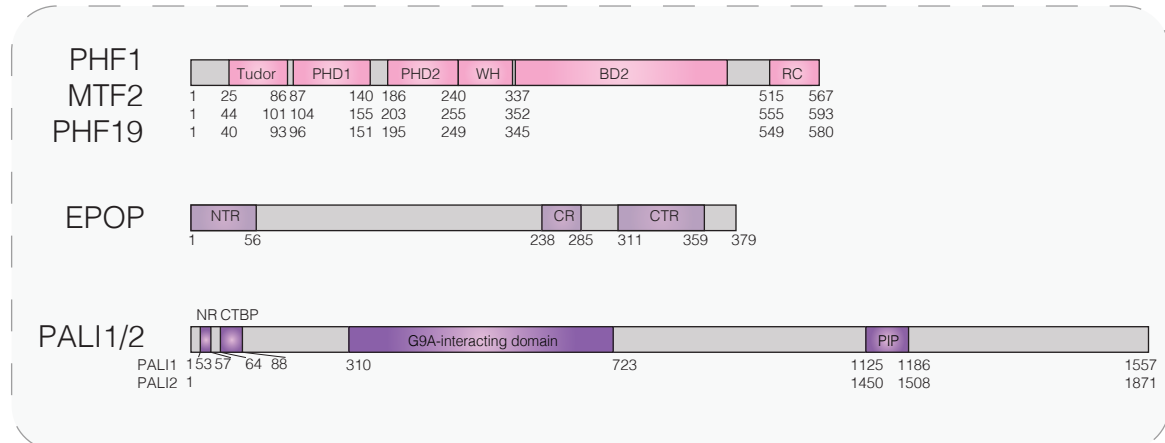
bridging protein between these two complexes. Cells lacking PALI1 show a reduction in H3K27me3 at Polycomb target sites confirming the *in vitro* observation that PALI1 contributes to the methyltransferase activity of the complex (Conway *et al.*, 2018). A recent study has also highlighted that lysine residues within the PIP domain are methylated by PRC2 and also PALI1 can facilitate PRC2 binding to DNA (Zhang *et al.*, 2020). The role of PALI2 within PRC2 remains less well studied, however, the conserved Pali interacting with PRC2 (PIP) domain of PALI2 is capable of associating with core PRC2 suggesting that PALI2 may behave similarly to PALI1 within the complex (Conway *et al.*, 2018). Loss of PALI2 in addition to PALI1 results in a slighter greater reduction in H3K27me3 levels compared with PALI1 single knock-out cells suggesting that in mouse ES cells, PALI2 may not have as great an effect on the catalytic activity of the complex. PALI2 does not share the nuclear binding domain found in PALI1 whilst DNA-binding AT-hooks are observed within the C-terminus of PALI2 suggesting that the two proteins have diverse roles within PRC2 (Figure 1.6).

JARID2 is a multidomain protein that together with AEBP2, assembles into the PRC2 subgroup known as PRC2.2. JARID2 enhances the catalytic activity of PRC2 *in vitro* and *in vivo* and together with AEBP2 stimulates activity of the complex more than each individual accessory protein (Conway *et al.*, 2018; Son *et al.*, 2013). In addition to enhancing the catalytic activity of the complex, JARID2 can also recognize H2AK119ub via the ubiquitin-interacting motif within the N-terminus and recruit PRC2 to chromatin (Cooper *et al.*, 2016). The domains within JARID2 can mediate interactions with different substrates including modified histone tails via the UIM, and binding to SUZ12 through the transrepression domain (Chen *et al.*, 2018). A DNA-binding region has also been characterised within the Jumonji N-term (JmjN) and AT-rich interaction domain (ARID) domains (Kim *et al.*, 2004). Interestingly, a JARID2 and MTF2 complex has been observed in mouse embryonic stem cells although the mechanism or function of the hybrid complex remains unclear (Grijzenhout *et al.*, 2016; Zhang *et al.*, 2011).

Core PRC2



PRC2.1 accessory proteins



PRC2.2 accessory proteins

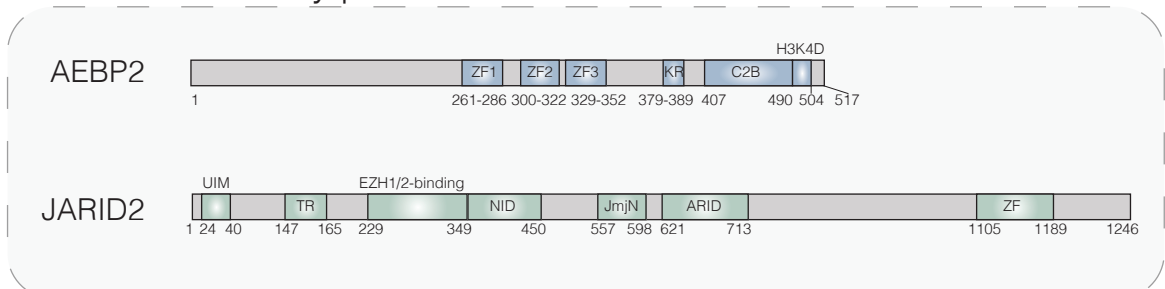


Figure 1.6 Schematic of domains of PRC2 core and accessory proteins. Representation of the domains in PRC2 core and accessory proteins that have been characterised in the literature. Amino acid positions are listed below the schematic.

Table 1.1 Domains within core and accessory PRC2 proteins

<i>Protein</i>	<i>Acronym</i>	<i>Name</i>
EZH1/2	SBD	SANT1L-binding domain
	EBD	EED-binding domain
	BAM	β-addition motif
	SAL	SET activation loop
	SRM	Stimulation response motif
	SANT1	Swi3, Ada2, N-Cor, and TFIIIB domain 1
	MCS	Motif connecting SANT1 and SANT2
	SANT2	Swi3, Ada2, N-Cor, and TFIIIB domain 2
	CXC	CXC domain
	SET	Su(var)3-9, E(z) and Trx domain
EED	WD1-7	WD-repeat
SUZ12	ZnB	Zinc finger binding domain
	WDB1	WD-domain binding 1
	C2	C2 domain
	ZnF	Zinc Finger
	WDB2	WD-domain binding 2
	VEFS	VRN2-EMF2-FIS2-Su(z)12 box
RBBP4/7	WD1-7	WD-repeat
PHF1 MTF2 PHF19	Tudor	Tudor domain
	PHD1	Plant homeodomain 1
	PHD2	Plant homeodomain 2
	WH/EH	Winged Helix/Extended homology domain
	BD2	Binding domain 2 (PHF1-specific)
	RC	Reverse Chromodomain
EPOP	NTR	N-terminal region
	CR	Central region
	CT	C-terminal region
PAL1/2	NR	Nuclear receptor binding box
	CTBP	CTBP binding motif
	G9A-interacting	G9A-interacting region
	PIP	PRC2 interaction with PAL1
AEBP2	ZF1-3	Zinc finger 1-3
	KR	Lysine/Arginine-rich motif
	C2B	C2-binding
	H3K4D	H3K4 displacement domain
JARID2	UIM	Ubiquitin-interacting motif
	TR	Transrepression domain
	EZH1/2-binding	EZH1/2-binding domain
	NID	Nucleosome interaction domain
	JmjN	Jumonji N-term
	ARID	AT-rich interaction domain
	ZF	Zinc Finger

Related to Figure 1.6

AEBP2 is a zinc finger containing protein that is associated as a member of the PRC2.2 complex along with JARID2 (Grijzenhout *et al.*, 2016; Hauri *et al.*, 2016). A 54kDa long and 32 kDa short isoform of AEBP2 is expressed in mouse ES cells with the latter expressed primarily in adult cells (Kim *et al.*, 2015). All isoforms of AEBP2 contain three conserved zinc fingers (ZF1, ZF2 and ZF3) as well as a C2-binding domain (C2B) which competes with PCLs for binding at the C2 lobe of SUZ12 (Figure 1.6 & Table 1.1) (Chen *et al.*, 2018). In addition to the C2 domain of SUZ12, AEBP2 binds an N-terminally located helix of SUZ12 (termed ABH) via hydrophobic interactions (Kasinath *et al.*, 2018; Youmans *et al.*, 2018). It has been reported that AEBP2 binds to DNA however it is unclear whether the affinity for DNA is sequence-specific (He *et al.*, 1999; Kim *et al.*, 2009; Wang *et al.*, 2017). A detailed Cryo-EM structure confirmed spatial arrangement of AEBP2 zinc finger domains was localised near EED and the SET domain of EZH2 which confirmed the importance of C-terminal AEBP2 for stability of core PRC2 by acting as a stabilising bridge between the core PRC2 members (Kasinath *et al.*, 2021; Kasinath *et al.*, 2018). Recent structural studies have also highlighted engagement of the zinc fingers of AEBP2 with a ubiquitin molecule.

It is clear that AEBP2 enhances methyltransferase activity of PRC2 in *in vitro* assays, however, whether AEBP2 contributes to methylation *in vivo* remains less well known (Cao & Zhang, 2004; Conway *et al.*, 2018; Kalb *et al.*, 2014). Loss of *Aebp2* in mouse ES cells correlates with a small increase in H3K27me3 levels at Polycomb target genes suggesting a potential inhibitory effect on activity of the complex (Conway *et al.*, 2018; Grijzenhout *et al.*, 2016). The increase in H3K27me3 may be attributable to the antagonism between PRC2.2 and PRC2.1 which have different degrees of histone methyltransferase activity.

Recent studies have identified a novel PRC2-interacting protein as an inhibitor of the enzymatic activity of the complex. Initially identified as an EZH1 interactor in mESCs (Maier *et al.*, 2015), EZHIP (EZH1/2 Inhibitory protein) remained uncharacterised until recent work by a number of groups has identified as having a H3K27M-like inhibitory mechanism on PRC2. EZHIP, alternatively known as CXORF67 or CATACOMB, is a highly disordered protein that has been shown to

interact with core PRC2 subunits (Hübner *et al*, 2019; Jain *et al*, 2019; Pajtlér *et al*, 2018; Piunti *et al*, 2019; Ragazzini *et al*, 2019). EZHIP is not expressed or very lowly expressed in many cell lines which may explain why the protein has not been reported as a PRC2-interactor in the majority of cell lines. Similar to the H3K27M substrate observed in diffuse intrinsic pontine glioma (DIPG) tumours, the highly conserved serine rich domain within the C-terminus of EZHIP directly interacts with EZH2 to inhibit catalytic activity (Hübner *et al*, 2019; Jain *et al*, 2019). Consequently, global H3K27me2/3 levels are reduced following interaction between EZHIP and EZH2. Fusion of the K27M-like peptide domain (KLM) of EZHIP with H3 results in a reduction in H3K27 di- and trimethylation, highlighting that the KLM domain sequesters PRC2 activity by emulating a H3K27M peptide (Jain *et al*, 2019).

1.6 Recruitment of Polycomb Repressive Complex

The recruitment of PcG complexes to chromatin in *Drosophila* is attributed to the long stretches of hundreds of base pairs of DNA termed Polycomb Response Elements (PREs) (Schuettengruber *et al*, 2017; Simon *et al*, 1993). While mammalian PcG complexes typically associate at unmethylated CG-rich repeats of DNA within gene promoters, a “PRE-like” DNA signature has not yet been identified (Vella *et al*, 2012). The longstanding hierarchical model of PcG recruitment was that recruitment of PRC2 to unmethylated CpG islands and catalysis of H3K27me3 was crucial in serving as a platform for the subsequent binding of chromodomain-containing CBX subunit of PRC1. Recruitment of PRC1 then mediates transcriptional repression either through mono-ubiquitination of H2AK119 or through condensation of chromatin leading to Polycomb chromatin domain formation. This model of PcG recruitment was widely adopted for many years based on evidence suggesting that loss of PRC2 impaired Pc (PRC1-associated, *Drosophila* homolog of CBX binding to PREs in fly cells (Cao *et al*, 2002; Wang *et al*, 2004b).

However, recent studies within the last 10 years have challenged this model and although the precise mechanism of Polycomb recruitment to chromatin is not yet fully understood, recent studies have highlighted an alternative hierarchical pathway

whereby ncPRC1 is recruited to chromatin to mediate deposition of H2AK119ub. The H2AK119ub histone mark is recognised by the UIM domain within PRC2-specific protein JARID2, resulting in catalysis of H3K27me3 which in turn is read by cPRC1 via the CBX protein (Figure 1.7). Supporting this model, loss of H3K27me3 in EED null ES cells yields a reduced occupancy of some PRC1 proteins but a minimal change in H2AK119ub or RYBP levels at PcG target loci suggesting PRC1 association with chromatin is independent of PRC2 (Tavares *et al.*, 2012). Recruitment of ncPRC1 is mediated through ncPRC1-exclusive interactors and the exact mechanism of ncPRC1 targeting to chromatin remains unclear for some of the subcomplexes. PCGF1-ncPRC1 is one of the best characterised complexes and targeting of the complex to unmethylated CGI is achieved by the CxxC motif of KDM2B (Farcas *et al.*, 2012). It is possible that the K36 demethylase domain of KDM2B, JmjC, may also assist in the transcriptional repression of target loci by removal of methylation marks from lysine residues and thereby mediating recruitment of the PRC1 to these sites (He *et al.*, 2008; Wu *et al.*, 2013). An elegant model for how ncPRC1 recruits PRC2 to chromatin was presented using a tethering assay to recruit the various PCGF1-6 proteins to an artificial fragment of human DNA lacking any PcG targets and recruitment of PCGF1/3/5-ncPRC1 was sufficient for the catalysis of H2AK119ub, subsequent targeting of PRC2 and methylation of H3K27 (Blackledge *et al.*, 2014). Furthermore, PRC2.2-subunit JARID2 can recognise and bind to mono-ubiquitinated H2AK119 via the N-terminal UIM (Cooper *et al.*, 2016). It has also been shown that binding of JARID2/AEBP2-PRC2 to H2AK119ub-modified nucleosomes simulates the enzymatic activity of the PRC2 *in vitro* (Kalb *et al.*, 2014) and recent Cryo-EM structural analysis of a 6-mer JARID2/AEBP2-PRC2 confirmed the affinity of PRC2.2 for H2AK119-modified nucleosomes. Additionally, a role of the zinc fingers within AEBP2 in binding to a ubiquitin molecule on H2AK119 was identified (Kasinath *et al.*, 2021).

Although the deposition of H2AK119ub has been highlighted as important for the recruitment of the PRC2.2 subcomplex, the recruitment of PcG to chromatin is not simply attributable to a simple step-wise recruitment of ncPRC1 followed by PRC2 and cPRC1 and there is emerging evidence to suggest that PRC2.1 can associate with chromatin independently of PRC1 and H2AK119ub (Glancy *et al.*, 2020; Healy *et al.*, 2019). PRC2 occupancy on chromatin and H3K27me3 levels are only

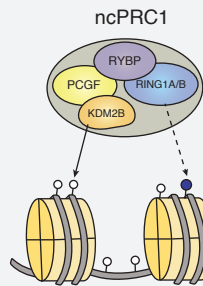
modestly reduced in the absence of H2AK119ub in mouse ES cells (Blackledge *et al*, 2020; Fursova *et al*, 2019; Healy *et al.*, 2019; Tamburri *et al*, 2020). Therefore, while it is clear that H2Aub-dependent and independent mechanisms are capable of recruiting PRC2 to chromatin, the role of PRC2.1-specific interactors, namely PHF1/PHF19/MTF2, EPOP and PAL1/2, in mediating the recruitment of PRC2 to polycomb domains is yet to be fully established.

Pcl-PRC2 may play a role in targeting mammalian PRC2 to chromatin through recognition of H3K36me2/3 by the N-terminal Tudor domain (Ballaré *et al.*, 2012; Brien *et al.*, 2012; Cai *et al.*, 2013; Musselman *et al.*, 2012) and while H3K36me3 is considered the classic substrate for PCLTudor, the Tudor domain of PHF1 and PHF19 have also been reported to bind H3K27me3 mark (and the testes-specific histone variant H3tK27me3), albeit with lesser affinity (Brien *et al.*, 2012; Dong *et al*, 2020; Kycia *et al*, 2014; Musselman *et al.*, 2012). The H3K36me3-demethylase, NO66 is capable of interacting with PHF19 in mouse ES cells and PHF19 is sufficient to recruit both PRC2 and NO66 to regions of transcriptionally active chromatin where then transcriptional repression is initiated (Brien *et al.*, 2012). Interestingly, while the Tudor domain of mammalian Pcl has the capacity to recognise modified H3K36 substrates, the Tudor domain of *Drosophila* Pcl does not show affinity for methylated lysine substrates which is attributed to the absences of several key aromatic amino acids that are crucial for the H3K36me3-binding cage (Cai *et al.*, 2013; Friberg *et al*, 2010; Gatchalian *et al*, 2015).

Within the literature, many chromatin modifying and regulatory proteins are capable of reading histone tails via PHD domain, so it seems likely that the PHD domains of Polycomb-like proteins could function in a similar manner. Towards understanding the role of Pcl PHD domains in recruitment of PRC2 to chromatin, the PHD2 domain of MTF2 has been reported to contribute to MTF2 chromatin binding *in vivo*

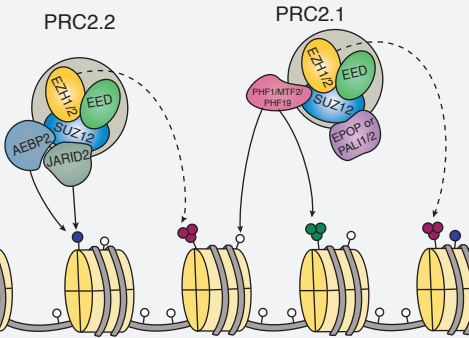
ncPRC1 recruitment

- Subunits within ncPRC1 binds unmethyl-CpG or DNA elements
- RING1A/B catalyses H2AK119ub



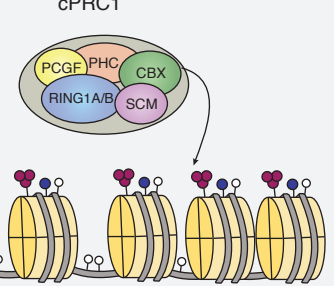
PRC2 recruitment

- JARID2-UIM binds ncPRC1 and recruits PRC2.2
- EZH2 catalyses H3K27me3
- PRC2.1 recruited via polycomb-like proteins binding CpG (or H3K36me3)
- EZH2 catalyses H3K27me3



cPRC1 recruitment

- cPRC1 binds PRC2-deposited H3K27me3
- PHC subunit facilitates multimerisation of PRC1



Chromatin compaction
Higher order structures

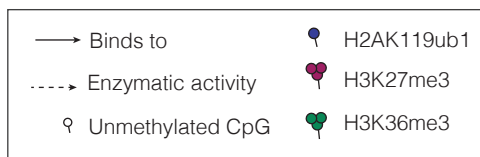


Figure 1.7 Simplified schematic of PRC1 and PRC2 hierarchal recruitment. ncPRC1 is recruited to chromatin through the recognition of CpG Islands or DNA elements via DNA-binding proteins subunits, for example, KDM2B. Deposition of ubiquitin on H2AK119 is catalysed by RING1A/B which is recognised by the PRC2.2-specific accessory proteins (AEBP2 and JARID2). PRC2 can also be recruited to chromatin independently of H2AK119ub-deposition through domains within the polycomb-like proteins (PHF1/MTF2/PHF19). PALI1 has also been recently reported to bind DNA. Catalysis of H3K27me3 by EZH1/2 subsequently serves as a docking site for cPRC1 which binds directly to this modification through CBX proteins. cPRC1 can further facilitate compaction of chromatin through multimerisation of the complex via the PHC subunit.

(Casanova *et al.*, 2011) while the PHD1 domain of PHF1 has been reported to read H4R3me2s mark *in vitro* (Liu *et al.*, 2018).

One recent study has reported that PHF1 enhances residency time of PRC2 on chromatin (Choi *et al.*, 2017) and recruitment of Pcl proteins to chromatin is speculated to be mediated by the winged helix (WH) domain which specifically binds to DNA (Choi *et al.*, 2017; Li *et al.*, 2017; Perino *et al.*, 2018). Contrasting results have been presented regarding the sequence-specificity of Pcl-WH for DNA motifs with Li *et al.* and Perino *et al.* suggesting a preference for CG-rich repeats of DNA whereas Choi *et al.* report that Pcl-WH does not bind to DNA in a sequence specific manner. Regardless, there is strong evidence to suggest a role for Pcl proteins in the targeting PRC2 to chromatin *in vitro* however there is a lack of *in vivo* evidence to support these claims and further studies are necessary to delineate the function of these proteins and their domains in PRC2 recruitment.

1.7 Mutations in PRC2 in cancer

The advancements in next generation sequencing (NGS) technologies within the last two decades have revolutionised human cancer research and has shed light on the genomic landscape of many cancer types. NGS is a valuable tool in identifying and characterising mutations within various cancer types and studies have revealed a multitude of cancer mutations occurring within the genes of chromatin regulators (Plass *et al.*, 2013; Vogelstein *et al.*, 2013). Cancer sequencing studies have revealed that PRC2 and associated H3K27 methylations are perturbed in many human cancers and mutations have been identified in a variety of cancers in PRC2-associated proteins (Chammas *et al.*, 2020; Conway *et al.*, 2015).

EZH2, an E2F-regulated gene and downstream of the pRB path, was reported to be amplified and a marker of poor prognosis in human tumours (Bracken *et al.*, 2003; Kleer *et al.*, 2003; Varambally *et al.*, 2002). However, elevated levels of EZH2 are unlikely to drive deregulation of H3K27 methylations given that the core components assemble in a 1:1:1 stoichiometry (Smits *et al.*, 2013) and elevated EZH2 does not

correlate with increased EED and SUZ12 in parallel. The majority of PRC2 mutations have been characterised within the core components and can be loosely categorised as either loss of function or change of function (Conway *et al.*, 2015). A large proportion of mutations have been noted within the SET domain of EZH2 and typically correlates with elevated H3K27me3 levels and reduced H3K27me2 (McCabe *et al.*, 2012; Morin *et al.*, 2010; Sneeringer *et al.*, 2010). One of the best characterised EZH2 mutations is tyrosine 641 (Y641), a heterozygous mutation (Y641C, Y641F, Y641H, Y641N, and Y641S) occurring in lymphoma, ALL and melanoma (Conway *et al.*, 2015). Mutation of Y641 alters substrate specificity favouring di-methylated H3K27 over mono- or un-methylated substrate hence the imbalance in H3K27me2 over me3 (Yap *et al.*, 2011).

Whilst mutation of PRC2 accessory proteins have been identified in cancer, the exact mechanism in which these mutations affect the activity or structure of the complex has been poorly characterised. For example, SUZ12 and PHF1 fusion proteins have been identified in endometrial stromal sarcomas. Although endometrial stromal sarcomas are amongst the rarest uterine malignancies, JAZF1-SUZ12 gene fusion are reported to be present in approximately 75% of benign endometrial stromal nodules, 50% of low grade and 15% of high grade endometrial stromal sarcomas (Hrzenjak, 2016). JAZF1-SUZ12, consisting of fusion of JAZF1 to the N-terminus of SUZ12, reportedly destabilises the remaining core components and decreases catalytic activity of the complex (Ma *et al.*, 2017). Recent findings have suggested that an alpha-helix within the N-terminus of SUZ12 (amino acids 80-107) is in close proximity to nucleosomal DNA (Grau *et al.*, 2021) and another separate study has suggested that mutation of single amino acid residues within the N-terminus of SUZ12 is sufficient to disrupt the balance between PRC2.1 and PRC2.2 (Youmans *et al.*, 2020). Similarly, inclusion of JAZF1-SUZ12 in *in vitro* complexes, impairs EPOP binding within the complex (Chen *et al.*, 2018). Taken together, it is tempting to speculate that SUZ12 fusion proteins may disrupt binding to chromatin or the shift the balance in complex formation however further experiments are necessary to better understand these fusion proteins.

The presence of these PRC2 mutations also provide potential targets for therapeutic strategies and several small molecule inhibitors have been developed to allosterically and catalytically inhibit the complex. GSK126 and EPZ-6438 are two catalytic inhibitors that compete with SAM to inhibit the activity of wild type and mutant EZH2 (Bratkowski *et al*, 2018). EZH2 inhibitors have varying degrees of selectivity for histone methyltransferases and despite the near identical SET domain, affinity for EZH1 is approximately 50-150 fold less than EZH2 (Knutson *et al*, 2013; McCabe *et al.*, 2012). In 2019, the EZH2 inhibitor Tazemetostat was approved by the Food and Drug Association (FDA) for the treatment of Follicular lymphomas. EZH2-inhibitor resistance has been reported in some instances highlighting the need for alternative PRC2 targeting strategies (Baker *et al*, 2015; Gibaja *et al*, 2016).

1.8 Structure and function of TP53

p53 is a tumour suppressor that plays a major role in various cell signalling pathways including but not limited to cell cycle arrest, the DNA damage and apoptosis (Fischer, 2017) and has subsequently been coined the nickname of “Guardian of the Genome” for its involvement in these biological processes. p53 was first reported in the literature in the 1970s and has been the focus of many studies over the years. Initially characterised to exhibit oncogenic properties, it was later discovered to play a tumour suppressive role within cells (Levine & Oren, 2009). TP53 is one of the most frequently mutated genes in a variety of human cancers with homozygous mutations occurring in approximately 50% of cancers. The majority of these mutations were characterised as disrupting the tumour suppressive function and instead promoting oncogenic effects (Baugh *et al*, 2018; Milner & Medcalf, 1991; Pitolli *et al*, 2019). p53 protein contains several functional domains including transactivation, DNA-binding, tetramerisation domain as well as a disordered C-terminal domain. In its active form, p53 assembles into a tetramer of four identical subunits that binds with high affinity to a DNA consensus sequence via the DNA-binding domain (Weinberg *et al*, 2004). In proliferating, unstressed cells, p53

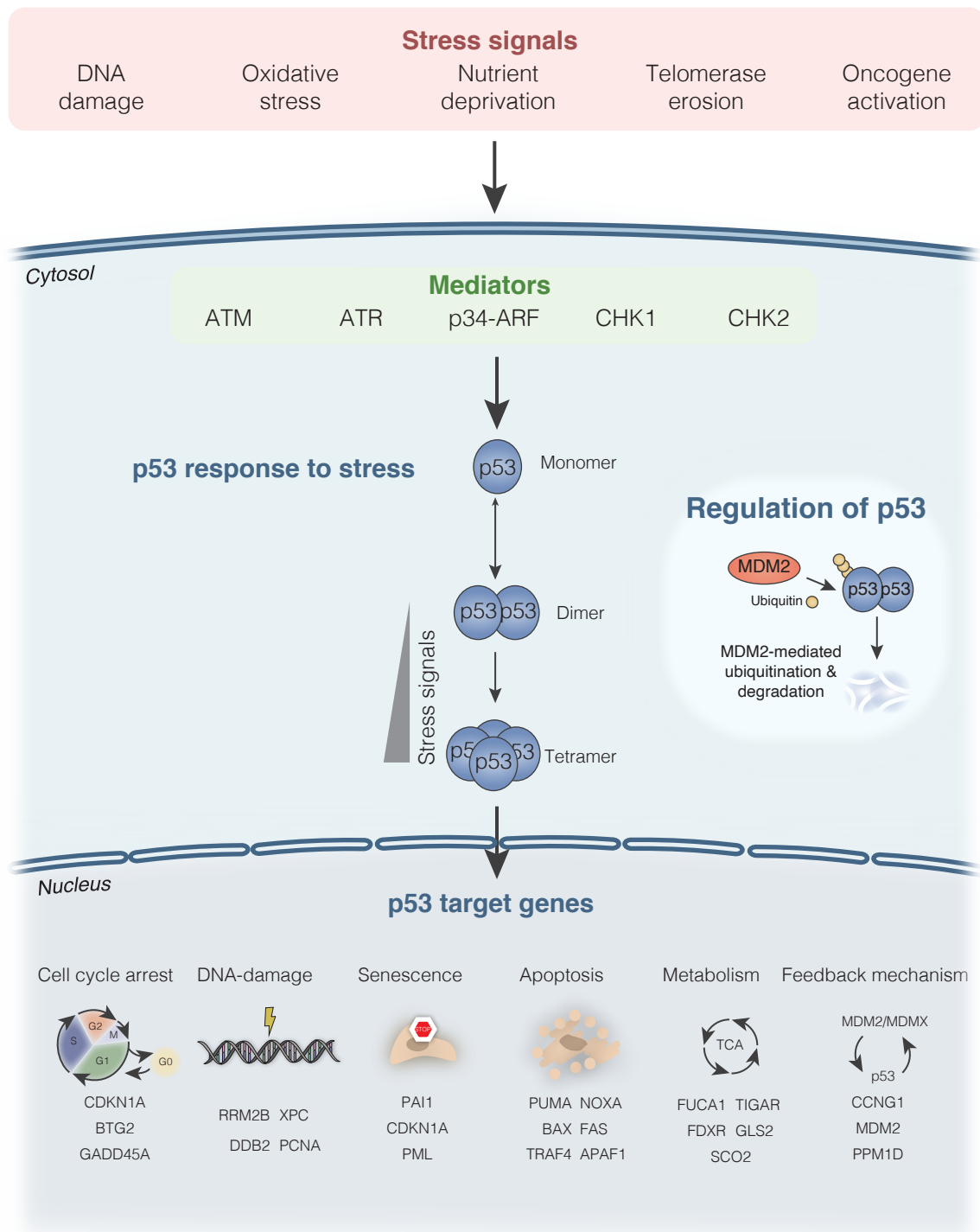


Figure 1.8 Schematic of the p53 signalling pathway

In response to cell stress, mediator proteins can post-translationally modify p53 resulting in the accumulation of p53 protein which assembles into an active tetrameric form in the cell. p53 can then bind to and upregulate a variety of target genes depending on the initial stress signal.

largely exists in a dimeric form with MDM2 mediating the degradation of p53 to maintain low levels of the protein (Katz *et al*, 2018). The p53 pathway within cells is a complex signalling cascade involving numerous intermediary proteins and ultimately resulting in upregulation of p53 target genes related to the cell stress trigger (Fischer, 2017). In response to cellular stress (e.g. DNA damage), upstream mediators (e.g. ATM/CHK2) are activated which can promote the posttranslational modification of p53. Modification of p53 can alter the protein status in a number of different ways including extending its half-life, inhibiting the MDM2-mediated degradation and also promoting the ability of p53 to bind to specific DNA sequences (Harris & Levine, 2005). As a consequence, p53 protein levels accumulate with the cell resulting in activation of the transcription factor and upregulation of p53 target genes. p53 has been established as a transcriptional activator of hundreds of target genes including cell cycle regulators such as CDKN1A and GADD45A, and apoptosis-related factors, for example PUMA and BAX (Fischer, 2017; Fischer *et al*, 2014) (Figure 1.8).

Although less well studied compared to its role in the cell cycle regulation, DNA damage response or apoptosis, p53 is also involved in the regulation of quiescence. Loss of p53 results in a reduction in quiescent cells numbers (Liu *et al*, 2009; McConnell *et al*, 2016) and furthermore, knockdown of p53 in human diploid fibroblasts, shifts cells from a quiescence state towards a more proliferative state, phenocopying loss of PHF1 (Brien *et al*, 2015). PHF1, whose canonical role has been as an accessory subunit within PRC2, has been reported to have a PRC2-independent role in the stabilisation of p53 during cellular quiescence.

1.9 The PRC2-independent role of PHF1 during quiescence

The growth, replication, and division of a single cell into two identical daughter cells is a highly complex process comprised of a sequence of orderly events termed collectively as the cell cycle. The cell cycle involves four main phases: G1 (the first gap phase), S (the synthesis phase), G2 (the second gap phase) and finally M (the mitosis phase). During these sequential stages, the cell increases in size (G1), the cellular contents duplicate (S), the cell prepares to divide (G2) and finally undergoes

division (M). Each stage of the cell cycle is tightly regulated by a numerous checkpoints to ensure that the preceding process has been successfully completed before progressing to the next phase. Progression is mediated by cyclin-dependent kinases (CDKs) and cyclin proteins with activation resulting in cell cycle arrest (Schafer, 1998; Vermeulen *et al*, 2003). In addition to the aforementioned stages of the cell cycle, cells can also exist in a dormant, non-proliferating state, termed quiescence (or G0). Quiescence is distinct from senescence which is defined as an irreversible state of cellular aging and quiescent cells can re-enter the cell cycle in response to various stimuli and begin proliferating. Quiescent cells exist in a low metabolic state and can be characterised by the absence of DNA replication, low condensed chromatin state as well as reduced transcriptional activity (McKnight *et al*, 2015; Swygert *et al*, 2019). As mentioned above, p53 and its target gene CDKN1A are important regulators of quiescence with loss of either gene impairing quiescence entry and maintenance (Liu *et al.*, 2009; Perucca *et al*, 2009). Chromatin regulatory proteins, as well as certain post-translational modifications, have also been reported as essential determinants of quiescence (Boonsanay *et al*, 2016; McKnight *et al.*, 2015; Young *et al*, 2017). Suv4-20h1, the histone methyltransferase responsible for H4K20me2, has been shown to be essential in maintaining quiescent stem cells with knockout of Suv4-20h1 resulting in the depletion of stem cell pool (Boonsanay *et al.*, 2016). Additionally, loss of Rpd3, a histone deacetylase, has been noted to impair quiescence entry and survival in yeast cells (McKnight *et al.*, 2015). Finally, while loss of the histone methyltransferase *Ezh2* does not significantly affect the quiescent status of hematopoietic stem cells (HSCs), knockout of the *Ezh1* results in a depletion of adult HSCs suggesting a role for *Ezh1*, but not *Ezh2*, in the maintenance of HSC quiescence (Hidalgo *et al*, 2012; Xie *et al*, 2014). Taken together, these highlight the importance of chromatin regulatory proteins during quiescence.

Aside from the role within PRC2, PHF1 has been reported to function off chromatin, independently of PRC2. It was reported that PHF1 is recruited to double strand breaks following irradiation where it co-localised with Ku70/Ku80 (Hong *et al*, 2008). PCL1 was reported to interact with other DNA damage repair proteins including p46, SMC1 and p53 (Hong *et al.*, 2008). The PHF1-p53 interaction was further characterized and two distinct domains of PHF1, named binding domain 1 and 2

(BD1 and 2), were shown to independently interact with the C-terminal domain (CTD) of p53. P53 is a known target of MDM2-ubiquitylation however PHF1 bound to p53 has been shown to protect from MDM2-mediated ubiquitylation and degradation (Yang *et al*, 2013). The PHF1-p53 interaction has been shown to promote cellular quiescence (Brien *et al*, 2015). PHF1 has been reported as a p53 target gene in quiescent cells and is predominantly expressed in non-cycling, quiescent cells. Depletion of PHF1 in fibroblasts phenocopies loss of p53 and a reduction in quiescent cell population (Brien *et al*, 2015). Within a tumour, there is a population of quiescent cancer stem cells which do not respond to typical cancer treatment therefore PHF1-p53 interaction is predicted to be a potential “Achilles heel” for targeting quiescence cancer cells (Brien & Bracken, 2016). One study has highlighted another potential role for PHF1 in bridging the interaction with PRMT5, an arginine methyltransferase, via the C-terminal domain and DDB1, a member of the Cullin4B-Ring E3 ligase complex, via the PHD2 domain however these observations remain to be independently validated (Liu *et al*, 2018).

1.10 Aims of thesis

The core aims of this thesis were to:

Chapter 3: Characterisation of the PHF1-p53 complex and the genome-wide localisation of p53 in quiescent cells

- Characterise the size and composition of the PHF1-p53 complex in vitro and in vivo.
- Compare the localisation and binding profile of p53 in quiescent human fibroblasts.
- Analyse the RNA expression signature in quiescent and cycling human fibroblast with a particular focus on PRC2 components.

Chapter 4: Biochemical analysis of PRC2 in quiescent cells

- Purify various combinations of recombinant PRC2.
- Characterise the composition of quiescent-specific and cycling-specific PRC2 in human fibroblasts.
- Assess the effects of PRC2 inhibition and degradation using small targeting molecules in human fibroblasts and cancer cells.

Chapter 5: Analysing the dependency of a panel of cancer cell lines on PRC2 subunits

- Identify dependencies on PRC2 subunit proteins in multiple cancer cell lines
- Validate the PHF1 dependencies observed in two cancer cell lines
- Rescue the depletion of PHF1 in a dependent human cancer cell line by stable expression of mouse PHF1

Chapter 2: **Materials and Methods**

2.1 Tissue culture methods

2.1.1 Cell lines and growth media

Primary IMR90 human fibroblasts were grown in DMEM (Lonza) supplemented with 10% FBS (Gibco), 100 U/ml penicillin and 100 U/ml streptomycin (Gibco). For serum starvation experiments, fibroblasts were grown until 60-70% confluency. Cells were subjected to serum-free media for up to 120 hours with media changes at 24 hour intervals until 120 hours. Cells were re-stimulated with media containing 20% FBS for 24 hours.

The insect cells used in Chapter 3 for the generation of a PHF1-p53 complex, (*Spodoptera frugiperda*) *Sf9* and *Hi5* (derived from *Trichoplusia ni*), were cultured on T75/T175 flasks in Hink's TNM-FH Insect Medium (Sigma) supplemented with 10% FBS (Gibco) and 0.5% Gentamycin (Sigma). The insect cells used in Chapter 4 for the generation of a G0-PRC2-associated complexes (*Trichoplusia ni* – *Tni*) were grown in suspension in a serum-free ESF 921 Insect Cell Media (Oxford Expression Technologies).

RCC4 plus vector alone kidney cancer cells were cultured in DMEM (Sigma) with 10% FBS (Gibco) plus 2 mM Glutamine, 0.5mg/ml (v/v) G418, 100 U/ml penicillin and 100 U/ml streptomycin (Gibco). BCPAP thyroid cancer cells were grown in RPMI-1640 (Gibco) with 10% heat-inactivated FBS (Gibco), 100 U/ml penicillin and 100 U/ml streptomycin (Gibco). A673 Ewing's Sarcoma cells were cultured in DMEM (Sigma) supplemented with 10% FBS (Gibco), 100 U/ml penicillin and 100 U/ml streptomycin (Gibco). HEC-1A endometrial cancer cells were grown in McCoy's 5a (Sigma) with 10% FBS (Gibco), 100 U/ml penicillin and 100 U/ml streptomycin (Gibco). NCIH196 lung cancer cells were cultured in RPMI-1640 supplemented with 10% FBS, 100 U/ml penicillin and 100 U/ml streptomycin (Gibco). HEK293T cells were grown in DMEM (Sigma) supplemented with 10% FBS (Gibco), 100 U/ml penicillin and 100 U/ml streptomycin (Gibco).

2.1.2 Production of lentivirus

HEK293T cells were split onto T75 flasks containing 12 ml of DMEM media one day prior to transfection so that they were ~80% confluent on the day of transfection. On the day of transfection, 5 µg of the plasmid of interest, 2 µg Pax8 and 3 µg VSVG were combined along with 30 µg polyethylenimine (PEI) and 500 µl of Optimem (Gibco) and left for 10-30 mins at room temperature. This was then added drop-wise to the confluent HEK293T cells. Media was collected from the HEK293 cells after 24 hours and replaced with 12ml of fresh media. Supernatant was collected after 48 hours, combined with the sample taken at 24 hours, filtered through a 0.44 µm filter and stored short-term at 4°C or long-term at -80°C. The amount of DNA/lentiviral packaging vectors/Optimem was altered depending on the scale of virus production i.e. T25 flask.

2.1.3 Generation of stable cell lines

To generate cancer cell lines stably expressing Cas9, lentivirus was generated as described in section 2.1.2 using the pSpCas9 plasmid containing blasticidin resistance. Cancer cell lines were split so that they were approximately 60-70% confluent and left to adhere slightly to the dish for 1-2 hours. Media was removed from the cells and lentivirus was added along with polybrene (final concentration 0.5µg/ml). Cells were spinoculated at 1700rpm for 1 hour at room temperature. Lentivirus was carefully removed from the cells and replaced with fresh media. 24 hours post transduction, blasticidin (Invitrogen) was added (final concentration 20 µg/ml). Cells were grown in blasticidin until the non-transduced cells dish died. Cells were not kept in culture for long periods of time (i.e. weeks/months) as a reduction in Cas9 activity has been reported.

The pCAG mouse Phf1 and Phf19 constructs used in the validation of the screen were not in a lentiviral backbone and were transfected into cells using the standard protocol of 5 µg of the plasmid of interest, 2 µg Pax8 and 3 µg VSVG were combined along with 30 µg polyethylenimine (PEI) and 500 µl of Optimem (Gibco) onto a 10

cm dish of A673 cells at approximately 80% confluency. Cells were selected with puromycin until the non-transfected cell line died.

2.1.4 Calculation of Cas9 activity in cancer cell lines

pXPR_011-sgEGFP plasmid was a lentiviral plasmid used to assay Cas9 activity in cells. Lentivirus was prepared as described in section 2.1.2. Cell lines were seeded onto 12-well tissue culture dishes and allowed to adhere for 1-2 hours prior to spinoculation with pXPR_011-sgEGFP lentivirus at varying volumes (15, 30, 60, 120, 240 μ l) plus polybrene (final concentration 0.5 μ g/ml) for 1 hour at 1700 rpm. Media was removed and replaced with fresh media and 24 hours after transduction, puromycin was added to the wells at a concentration that ensured that all non-transduced cells were completely dead after 2 days. The amount of puromycin varies between 0.5–2 μ g/ml (final concentration) depending on the cell line.

48 hours after adding puromycin, cells were trypsinised and counted. The infection efficiency was calculated by dividing the “number of cells infected and with selection” by the “number of cells in the infected with no selection”. Cells which have an infection efficiency of 15-30% were passaged for the duration of the experiment. 4 cell populations were continued for the remainder of the experiment; 1. Parental cells (No Cas9, no pXPR_011), 2. Parental cells plus pXPR_011 (No Cas9, plus pXPR_011), 3. Cas9-expressing cells (Plus Cas9, no pXPR_011) and 4. Cas9-expressing cells (Plus Cas9, plus pXPR_011). FACs analysis to check GFP levels was performed using the BD Accuri C6 Plus flow cytometer and Flowjo analysis software on all four cell lines around days 12-14. Cas9 activity was estimated based on the depletion of GFP in cell lines expressing Cas9 along with the pXPR_001 plasmid and cell lines which surpassed 70% activity were taken forward for the PRC2 library screens.

2.1.5 PRC2 library screen in cancer cell lines

PRC2 sgRNA library lentivirus was prepared as described in section 2.1.2 using the ipUSEPR plasmid containing an RFP cassette. The PRC2 sgRNA library was generated by Dr. Chun-Wei (David) Chen, City of Hope Comprehensive Cancer Center. A titer assay was performed prior to the large scale library screen to calculate the volume of virus required to give an multiplicity of infection (MOI) of 10%. The volume of lentivirus required to give an MOI of 10% varies depending on the cell line and cell numbers.

For the library screens, I used an sgRNA library containing approximately 4000 sgRNA and to maintain 1000x coverage of the guide at an MOI of approximately 10%, I need to transduce 40 million cells at the beginning of the screen. To account for variances, I seeded 50 million cancer cells in total across several 6-well dishes and spinoculated at room temperature for 1 hour at 1700 rpm with PRC2 sgRNA library lentivirus at an MOI of 10%. Lentivirus was removed from the wells and replaced with fresh media. 24 hours post spinoculation puromycin was added for two days at a concentration of 0.5–2 $\mu\text{g/ml}$ (depending on the cell line). On day three, puromycin was removed from the cells and cells were trypsinised and counted. Five million cells were reseeded onto dishes and the remaining cells were snap frozen in liquid nitrogen in five million cell pellets. Cells were counted and five million were continued every third day until day eighteen when the final pellets were harvested.

DNA extraction and sequencing were performed by the Chen group. Briefly, genomic DNA was extracted using the using HotSHOT protocol. 5 million cells were lysed in 250 μl of HotA buffer (25 mM NaOH, 0.25 mM EDTA in ddH₂O (final pH ~12)) and boiled at 95°C for 30 minutes. Genomic DNA was fragmented briefly using sonicator for 15 seconds. 500 μl of HotB buffer (40 mM Tris-HCl (pH 6.8)) and cells were briefly vortexed and centrifuged. The final yield of the recovered supernatant is approximately 1 μg of genomic DNA per 30 μl .

The PCR was set up and cycle performed as described in Table 2.1 and 2.2. A total of 25 reactions were set-up. (Per 1 million cells needed 5 tubes of PCR (100 μL each) to amplify (~5 μg gDNA). Following amplification, PCR products were pooled and

mixed well. A 50 µl PCR product was run on a 2 or 2.5% agarose gel and the 150bp product was extracted from the gel using the Qiagen Gel Extraction kit. Samples were eluted in 20ul of 1x TE buffer and stored at -20°C until sequenced.

Table 2.1 PCR reaction components for the amplification of gDNA

Component	Stock	Volume
Sample DNA	1 µg per 30 µl	30 µl
DreamTaq HotStart Green 2x Master Mix	2x	50 µl
Primer mix (DCF01/DCR03; 5 µM each)	5 µM each	10 µl
ddH ₂ O	-	10 µl
TOTAL	-	100 µl

Primer DCF01: 5'-cttgtggaaggacgaaacaccg-3'

Primer DCR03: 5'-cctaggaacagcggttataaaaagc-3'

Table 2.2 PCR cycle for the amplification of gDNA

Step	Temperature (°C)	Time (seconds)
Initial denaturation	95	60
Denaturation	95	30
Annealing	65	20
Extension	72	30
Final Extension	72	300
Hold	4	∞

35 cycles

2.1.6 Validation of PRC2 library screen

The sgRNA that were identified as a predicted dependency in PHF1 (amino acids 19-27 and 125-133) were cloned into the ipUSEPR plasmid containing an RFP

cassette and puromycin resistance. Lentivirus was generated as described in section 2.1.2 and used to transduce Cas9-expressing cells in 96-well format by spinoculation at 1700 rpm for 1 hour at room temperature. Lentivirus was removed and cells were cultured for a period of 24 days in total. Cells were trypsinised and split at the required ratio every 3 days. Every 6 days the RFP abundance was assessed using the Guava easyCyte HT system 12 and analysed using FlowJo software.

2.2 Protein methods

2.2.1 Baculovirus production of p53 and PRC2 components

Baculoviruses for FLAG-His-PHF1 and FLAG-His-p53 were generated by Dr. Gerard Brien and Dr. Evan Healy in the Bracken lab. StrepHis-PHF1 baculovirus was a gift from Dr. Claudio Ciferri of Genentech, San Francisco. Baculoviruses for untagged PHF1 and p53 were generated in *Spodoptera frugiperda*, *Sf9* cells. Baculoviruses were prepared by co-transfecting a baculovirus transfer vector (pVL1392) containing either untagged PHF1 or p53 ORFs and BacPAK6 Linearised DNA construct (Oxford Expression Technologies) into *Sf9* cells. The supernatant (passage 1 [P1]) was recovered from the cells and used to infect *Sf9* cells. The supernatant was recovered and the process was repeated until P3.

2.2.2 Purification of PHF1-p53 complex

Hi5 insect cells were infected with different combinations of baculovirus for tagged and untagged PHF1 and p53. Cells were grown at 28°C, harvested 44-48hrs after infection, washed twice in PBS and snap-frozen in liquid nitrogen. Pellets were lysed in BPL2B buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 20% Glycerol, 4 mM MgCl₂, 3 mM β-mercaptoethanol, 0.05% NP40, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF). Lysate was sonicated for 2-3 rounds at 50% amplitude for 10 seconds each round and clarified by centrifugation at 21,000g for 30 minutes. Supernatant was loaded onto pre-washed Ni-NTA His resin (Novagen #70666) and rotated for 3 hours at 4°C. The resin was washed thoroughly in BPL2B buffer to remove non-specific interactions. Proteins were eluted from the beads in elution buffer (350 mM

Imidazole, 50 mM Tris-HCl pH 8, 5 mM MgCl₂, 150 mM NaCl, 0.05% NP40). Eluted fractions were analysed by Coomassie staining on SDS-PAGE gels and subsequently snap frozen in liquid nitrogen (after addition of 20% glycerol).

2.2.3 Purification of G0-PRC2

Tni insect cells were infected in an equal ratio with Baculoviruses for each of the PRC2 components. Cells were grown in suspension at 28°C, collected after 48 hours and harvested by centrifugation at 4000 rpm. Supernatant was discarded and pellets were stored at -80°C until needed. Pellets were lysed in 20 mM Tris, 500 mM NaCl, 20% Glycerol, 4 mM MgCl₂, 1 mM TCEP, 5 mM imidazole, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF. Samples were sonicated at 2-3 times for 10 seconds at 50% amplitude. After sonication, 0.05% NP-40 (final volume) and 1 µl/ml lysate of Benzonase was added to the samples and rotated for 30 mins at 4° C. Lysate was clarified by centrifugation at 21,000g for 30 minutes at 4°C. Supernatant was incubated with pre-washed Ni-NTA His resin and rotated at 4°C for 3 hours. Beads were washed with 1ml of wash buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 20% Glycerol, 4 mM MgCl₂, 1 mM TCEP, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF) 6 times and eluted in the same buffer plus 3 50 mM imidazole for “High salt purifications”. For high to low salt purifications, beads were washed 4 times in wash buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 20% Glycerol, 4 mM MgCl₂, 1 mM TCEP, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF), twice in a lower salt wash buffer (10 mM Imidazole, 50 mM Tris-HCl pH 8, 5 mM MgCl₂, 150 mM NaCl). Elutions were performed in the lower salt buffer plus 350 mM imidazole. Elutions were aliquoted and snap-frozen in the presence of 20% glycerol.

2.2.4 Preparation of whole cell protein lysates and Western blotting

Cells were harvested by either trypsinising or scraping down, washed three times in cold PBS and snap frozen in liquid nitrogen. Pellets were lysed in 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 Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF). Cells were sonicated, clarified at 21,000 g for 25 minutes at 4°C and boiled at 95°C for 5 minutes with 2X

Laemli buffer. Lysate was separated by SDS-PAGE and transferred to nitrocellulose membrane. Membranes were blocked for 1 hour in 5% milk PBS-Tween (0.1%) followed by overnight incubation in primary antibody at 4°C. Membranes were incubated with the relevant secondary antibody for 1 hour and protein levels were determined by chemiluminescence using X-ray film or scanned using the LICOR Odyssey.

2.2.5 FLAG immunoprecipitations

Cells were resuspended in High Salt Buffer (50 mM Tris-HCl, pH 7.2, 300 mM NaCl, 0.5% (v/v) NP-40, 1 mM EDTA pH 7.4, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1mM PMSF) and sonicated. Samples were rotated at 4C for 20 minutes before addition of No Salt Buffer (50mM Tris-HCl, pH 7.2, 0.5% (v/v) NP-40, 1 mM EDTA pH 7.4, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF) and samples were clarified at 21,000 g for 30 minutes at 4°C. Samples were then pre-cleared with mouse IgG coupled beads for 1 hour followed by overnight incubation with FLAG beads (Gift of Dr. Claudio Ciferri of Genentech, San Francisco) plus Benzonase nuclease. Beads were thoroughly washed with Wash buffer (1:1 of High Salt and Low Salt buffer) and elutions were collected after addition of competitive FLAG peptide. 2X Laemli dye was added to the eluate and analysed by western blotting. FLAG-Ips performed in the presence of urea were assembled as above with the exception that samples were incubated in the presence of urea ranging from 0-2.5M overnight. Washes and elutions were performed as previously described.

2.2.6 Endogenous immunoprecipitations

Antibodies were cross-linked one day prior to performing the immunoprecipitation. 5µg of antibody was coupled to 20µl of packed Protein A/G beads in 1ml PBS-Tween (0.1%) overnight rotating at 4°C. Beads were centrifuged at 1500g (4000rpm) and washed twice with 1ml of 0.3M Sodium Borate pH 9.0. Antibodies were cross-linked to beads by addition of 1ml 20mM Dimethyl pimelimidate dihydrochloride (made up in 0.2M Sodium Borate) for 30 minutes rotating at room temperature. Beads were quenched by incubating the beads with 1ml 0.2M

Ethanolamine pH 8.0 followed by another 1ml of 0.2M Ethanolamine pH 8.0 rotating at room temperature for 2 hours. Beads were washed once with Buffer C (20 mM HEPES pH 7.9, 0.2 mM EDTA, 1.5mM MgCl₂, 20% glycerol, 420 mM NaCl, 2 µg/ml Aprotinin, 1µg/ml Leupeptin, 1mM PMSF) and blocked for 60 minutes at 4°C in Buffer C100 plus 0.1 mg/ml Insulin (Sigma), 0.2 mg/ml Chicken egg albumin (Sigma) and 0.1% (v/v) fish skin gelatin (Sigma).

Cells were resuspended in Buffer C and sonicated 3x 15 seconds using a Soniprep 150 probe sonicator set to 50% amplitude. Lysates were incubated at 4°C rotating for 30 minutes to ensure sufficient lysis followed by clarification at 20,000g at 4°C for 30 minutes. Supernatant was 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). The dialysed samples plus Benzonase nuclease were clarified again and then added to the antibody-coupled beads rotating for 3 hours at 4°C. Beads were washed 4 times with 1ml Buffer C100. A small portion of beads (20%) were put aside to confirm immunoprecipitation of target proteins. Samples were eluted using SDS-PAGE sample buffer. Beads were boiled for 5 minutes at 95°C 950rpm and the eluted material was removed following centrifugation at 20,000g for 5 minutes and analysed by SDS-PAGE and Western blot. The remainder of beads (80%) were snap-frozen in liquid nitrogen and Label free quantification (LFQ) LC-MS/MS analysis was performed in collaboration with the Vermeulen lab.

Briefly, the beads were resuspended in elution buffer (2 M Urea, 100 mM Tris pH 8, 10 mM DTT) and incubated 20 min on a shaker (1300 rpm) at room temperature. Iodoacetamide was added to a final concentration of 50 mM and incubated for 10 minutes at room temperature shaking in the dark. Partial digestion and elution from the beads was initiated by addition of 0.25 mg Trypsin (Promega; V5113) for 2 hr. The supernatant was recovered and the beads were resuspended in 50 ml of elution buffer followed by a 5 minute incubation shaking at room temperature. The supernatants were combined and 0.1 mg Trypsin was added followed by overnight incubation at room temperature. The digestion was arrested with the addition of TFA at a final concentration 0.5%. The digested samples were then desalted and purified using StageTips (Rappsilber *et al*, 2007). Buffer B (80% acetonitrile, 0.1% formic

acid) was used to elute the peptides from StageTips and concentrated to 5 ml by SpeedVac centrifugation at room temperature. The sample was brought up to 12 ml using buffer A (0.1% formic acid). Samples were measured using a gradient from 9%–32% Buffer B for 114 min followed by washes at 50% then 95% Buffer B resulting in total 140 min data collection time. Mass spectra were recorded on an LTQ-Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific). Scans were collected in data-dependent top speed mode with dynamic exclusion set at 60 s.

2.2.7 Sucrose gradients

Sucrose gradients were prepared in 5ml centrifugation tubes (Beckman Coulter) by assembling layers of varying sucrose concentrations (Sucrose 10-40%, 50 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM EDTA pH 7.4, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF). Proteins were added to the column and ultra-centrifuged at 45,000 rpm in a SWTI 55 rotor for 16 hours at 4°C. Fraction were removed from the column and analysed by western blotting.

2.3 DNA methods

2.3.1 Chromatin immunoprecipitation (ChIP)

Cells were trypsinised, counted and normalised prior to formaldehyde cross-linking. Cells were fixed in 10ml of PBS containing 1% formaldehyde for 10 mins rotating at room temperature and subsequently quenched with glycine at a final concentration of 0.125M for 5 minutes rotating at room temperature. Cells were gently washed three times with PBS and finally resuspended in 6 ml SDS-lysis buffer (100 mM NaCl, 50 mM Tris pH8.1, 5 mM EDTA pH 8.0, 0.5% SDS, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF). Samples were frozen in liquid nitrogen and stored at -20°C until needed. Samples were thawed at 37°C briefly and centrifuged at 1200rpm for 6 minutes to pellet the chromatin. The pellet was resuspended in 1 ml 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, 5% Triton X-100, 2 µg/ml Aprotinin, 1 µg/ml Leupeptin, 1 mM PMSF) and sonicated using a Soniprep 150 probe sonicator set to 50% amplitude for cycles of 1 sec on/4 secs off for a total of 5 mins total

sonication time. The resulting chromatin were 200-1000bp in size as confirmed by testing the samples on a 1% agarose gel.

Samples were diluted with ChIP buffer, divided across the required antibodies and incubated overnight at 4°C. Magnetic Protein G (or A) Dynabeads (Pierce) were added to the antibody/lysate mix for 3 hours rotating at 4°C. Beads were subjected to 8 rounds of washing with each wash consisting of 1ml of buffer rotating at 4°C for 5 minutes. Beads were washed 3 times with Mixed Micelle (150 mM NaCl, 20 mM Tris pH 8.1, 5 mM EDTA pH 8.0, 5.2% Sucrose, 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), 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% NaN₃) and finally one wash with TE. Samples were eluted from beads with Elution buffer (0.1 M NaHCO₃, 1% SDS) with shaking for 1 hour at 65°C. The supernatant was removed and incubated overnight at 65°C while shaking to reverse the crosslinks. Samples were treated with RNase (Thermo Fisher) and Proteinase K (Sigma) and DNA was recovered using the Qiagen MinElute PCR Purification Kit. ChIPs were analysed using Luna Universal qPCR Master Mix (NEB) on the QuantStudio 3 Real-Time PCR System.

2.3.2 Library generation for NGS of ChIP samples

The ChIP DNA was quantified using the Qubit dsDNA High Sensitivity Assay Kit (ThermoFisher Q32854). Samples were prepared for sequencing using the NEBNext Ultra II DNA Library Kit for Illumina (E7645) and NEBNext Multiplex Oligos for Illumina (Set#1, NEB #7335). 0.5ng-1ng of ChIP DNA for each sample was used for library preparation. After the ligation of adaptors, DNA was PCR amplified for 8-10 cycles depending on the input amount of DNA. The PCR product was purified using NEBNext Sample Purification Beads (E7767S) and the eluate was analysed on a High Sensitivity D1000 Screen Tape (Agilent). The libraries were subsequently used for cluster generation and sequencing using an Illumina NextSeq500 (ID: NB501524), with single end 75bp read length.

2.4 RNA methods

2.4.1 RNA preparation for RT-PCR analysis

Cells were harvested, washed twice in PBS and snap frozen in liquid nitrogen. RNA was extracted from the cells using the RNeasy kit (Qiagen). cDNA was generated by reverse transcription PCR using the TaqMan Reverse Transcription kit (Applied Biosystems). Relative mRNA expression levels were determined using Luna Universal qPCR Master Mix (NEB) on the QuantStudio 3 Real-Time PCR System. RPLPO was used as a normalizer and reactions were performed in triplicate.

2.4.2 RNA preparation for RNA-sequencing

Cells were harvested, washed twice in PBS and snap frozen in liquid nitrogen. RNA was extracted from the cells using the Rneasy kit (Qiagen). The concentration of RNA was measured using a Qubit 3.0 and the quality of RNA was assessed using an Agilent RNA ScreenTape. Samples with a RIN > 8.5 were carried forward for RNA-seq. 1 mg of RNA was used for each sample during library preparation and samples were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina (E7770L). Following adaptor ligation, DNA was PCR amplified for 8 cycles and the PCR product was purified using NEBNext Sample Purification Beads (E7767S). Sample quality was assessed using an Agilent High Sensitivity D1000 Screen Tape prior to sequencing and subsequently then used for cluster generation and sequencing using an Illumina NextSeq 500 (ID: NB501524), with paired end 36bp read length.

2.5 Bioinformatics methods

2.5.1 Analysis of ChIP-sequencing data

Sequencing reads were aligned to the human genome (hg38) using Bowtie v2.1.0 (Langmead & Salzberg, 2012). The aligned files were converted to bigwig files using bamCoverage, a tool within the deepTools suite (Ramírez *et al*, 2014; Ramírez *et*

et al, 2016) and peaks were called using mascs2 (FDR = 0.05) (Feng *et al*, 2012). Peaks were annotated using HOMER (Heinz *et al*, 2010). Overlapping of p53 peaks in two cell conditions was performed using bedtools intersect (Quinlan & Hall, 2010)

2.5.2 Analysis of RNA-sequencing data

Raw sequencing reads were aligned to the human genome (hg38) using HISAT2 (Pertea *et al*, 2016) and the aligned files were converted to bigwigs using bamcoverage (Ramírez *et al.*, 2014). Sequence reads were aggregated into a count for each gene using featurecounts (Liao *et al*, 2014) and differentially gene expression was identified using DESeq2 (Love *et al*, 2014). The volcano plots were generated using EnhancedVolcano (Blighe *et al*, 2019) and GO analysis was performed using Metascape (Zhou *et al*, 2019)

Table 2.3 List of plasmids

Plasmid	Bacterial res.	Mammal res.	Tag	Source
pLENTI FLAG Cas9	Amp	Blasticidin	FLAG	Bracken lab
pXPR_011_EGFP	Amp	Puromycin	GFP	Bracken lab
ipUSEPR EV	Amp	Puromycin	RFP	C. Chen lab
ipUSEPR_PRC2sgRNAlib	Amp	Puromycin	RFP	C. Chen lab
ipUSEPR_individual_PHF1 sgRNAs	Amp	Puromycin	RFP	This thesis
pCag EV	Amp&Chlor	Puromycin	FLAG/Avi	Bracken lab
pCag mPhf1	Amp	Puromycin	FLAG/Avi	Bracken lab
pCag mPhf19	Amp	Puromycin	FLAG/Avi	Bracken lab
pVL untagged EV	Amp	-	-	Bracken lab
pVL untagged-PRC2 components	Amp	-	-	Bracken lab
pVL FLAG/His EV	Amp	-	FLAG/His	Bracken lab
pVL FLAG/His-PRC2 components	Amp	-	FLAG/His	Bracken lab

Table 2.4 List of antibodies

Protein	Source	Western Blot	IP	ChIP
PHF1	Protein Tech (15663-1-AP)	1:500	-	-
EZH1	Cell Signalling (D7D5D)	1:1000		
EZH2	BD43 (Pasini et al., 2004)	1:10	-	-
EZH2	AC22 (Bracken et al., 2006)	-	5 µg	2 µg
EED	Cell Signalling (51673S)	1:500		
SUZ12	Cell Signalling (3737)	1:2000	5 µg	1 µg
BMI1	DC9 (Bracken et al., 2007)	1:20	-	-
Histone H3	Abcam (ab1791)	1:50000	-	1 µg
p53	Helin lab (DO-1)	1:50	-	2 µg
CCNA2	BD Biosciences (611268)	1:1000	-	-
GAPDH	Protein Tech (10494-1-AP)	1:10000	-	-
Beta-actin	Protein Tech (60008-1-Ig)	1:10000	-	-
Mouse-IgG	Millipore (12-371)	-	5 µg	2 µg
Rabbit-IgG	Millipore (12-370)	-	5 µg	2 µg
FLAG M2	Sigma	1:1000		
HA	Cell Signalling (3724)	1:1000		

Table 2.5 List of RT-qPCR primers

Target	Forward (5'-3')	Reverse (5'-3')
mRplpo mRNA	TCATTGTGGGAGCAGAC	CAGCAGTTTCTCCAGAGC
hPHF1 mRNA	CTTTGGGAGGGTCAAGATGT G	TGGTACCCAAGTATAGCAGCC C
hMTF2 mRNA	CCAGAAAAAGAACGCGTACA GG	CTTCTCCGCAAATGTGGTATT G
hPHF19 mRNA	CAATGAGCCCATGATGTTTG G	CTGGTTACACACGGAGCAGAA G
hEZH2 mRNA	GCCAGACTGGGAAGAAATCT GAGAA	AGCTGTCTCAGTCGCATGTAC TCTGA
hSUZ12 mRNA	GTGAAGAAGCCGAAAATGGA GCAC	TCAAAGGCCTGGAGGAAAAG CT
hEZH1 mRNA	TAAATTGCACGCGTTTAGGC TG	TCAGATACCCTCTGCCAGTGT G
hCDKN1A (p21) mRNA	GACTTTGTCACCGAGACACC ACTGG	GGTAGAGCTTGGGCAGGCCA
hCCNA2 mRNA	TGCAGAAAGTATTGGGTAA	GCATCTAGTTTTGAAAGTCCTT
hTP53 mRNA	CGCTTCGAGATGTTCCGAG	GGCATCCTTGAGTTCCAAGG
hGADD45A mRNA	TCTGCAGATCCACTTCACCC TG	CAGGATGTTGATGTGCTTCTC G
hPAI1 mRNA	AGAGAACCTGGGAATGACCG AC	TGAAAGACTCGTGAAGTCAGC C
hMDM2 mRNA	AGATGGAGCAAGAAGCCGA GC	AGCAGCAGGATCTCGGTCAG AG

Table 2.6 List of ChIP primers

Target	Forward (5'-3')	Reverse (5'-3')
hPHF1 ChIP	TAGGTCTCCTACGTCTGCCC C	CCTCCTTACCGGTGACAGCTG
hMTF2 ChIP	TTTCACTATTGTGTTAGCCCC GG	ACTCTTTAACGTTGCAGTGAG GC
hPHF19 ChIP	CAACACAGACGGGCAAGAGA C	CCACAACCTATCTGAATGCCA C
hCDKN1A -2283 ChIP	AGCAGGCTGTGGCTCTGATT	CAAAATAGCCACCAGCCTCTT CT
hCDKN1A -1391 ChIP	CTGTCCTCCCCGAGGTCA	ACATCTCAGGCTGCTCAGAGT CT
hCDKN1A -20 ChIP	AGCTGCGCCAGCTGAGG	CACAAGGAACTGACTTCGGCA
hGADD45A ChIP	GGCTGCCTTTGTCCGACTAG AG	TGATGAAACTAGCACATCTCG GC

Chapter 3: Characterisation of the PHF1-p53 complex and the genome-wide localisation of p53 in quiescent cells

3.1 Introduction

The role of Polycomb group proteins as chromatin regulators and their importance in embryonic development has been well studied over the last 20 years (Faust *et al.*, 1995; O'Carroll *et al.*, 2001; Pasini *et al.*, 2004). The Polycomb-like proteins (PCL1/PHF1, MTF2/PCL2 and PHF19/PCL3) are accessory subunits within PRC2 and serve a critical function in mediating recruitment of the complex to chromatin (Brien *et al.*, 2012; Healy *et al.*, 2019; Li *et al.*, 2017; Perino *et al.*, 2018). Additionally, biochemical in vitro assays have suggested that PCL proteins enhance PRC2 methyltransferase activity of the complex (Sarma *et al.*, 2008; Zhang *et al.*, 2019). Interestingly, several groups have reported a PRC2-independent role for PHF1 whereby it can bind to p53 (Brien *et al.*, 2015; Hong *et al.*, 2008; Yang *et al.*, 2013).

The first indication that PHF1 may serve a role outside of PRC2 was identified by Hong *et al.* (2008). Here, PHF1 was reported to localise at double strand breaks following DNA damage, an interaction thought to be mediated by the N-terminal Tudor and the poorly characterised central region (amino acids 340-431). Additionally, PHF1 was reported to physically interact with several DNA damage response-associated proteins including p53 (Hong *et al.*, 2008). Further work by Yang *et al.* (2013) later characterised two domains, PHD1 and binding domain 2 (BD2), of PHF1 that mediate binding to the C-terminal tail of p53. P53 protein stability is regulated in a ubiquitin-dependent manner. MDM2 is an E3 ligase that binds to p53 and mediates covalent conjugation of ubiquitin molecules to p53 thereby targeting it for proteasomal degradation (Haupt *et al.*, 1997; Kruse & Gu, 2009; Moll & Petrenko, 2003). It was established that binding of PHF1 to p53 was sufficient to protect p53 from MDM2-mediated ubiquitination and degradation (Yang *et al.*, 2013). Recent findings by Brien *et al.* (2015) have identified two key serine residues within the PHD1 of PHF1 to be sufficient for binding to the C-terminus of p53 and mutation of these serine residues to their equivalent residues in PHF19 (Glycine 95 and Isoleucine 106) abolished binding to p53 C-terminus in in vitro peptide pulldown assay (Brien *et al.*, 2015). Additionally, while MTF2 and PHF19 expression are regulated by E2F, PHF1 was identified as a p53 target gene in quiescent cells.

Taken together, there is strong evidence to suggest the assembly of a PHF1-p53 complex in cells. One of the aims of this chapter is to better understand the composition of the PHF1-p53 complex and to characterise it in terms of the overall size and stability between the two proteins. In my work towards reaching this goal, I perform several biochemical analyses of this complex both in vivo using an exogenously-tagged PHF1 construct stably expressed in a human cell line and in vitro using a recombinant complex expressed in insect cells.

As mentioned previously, published data has reported PHF1 as a p53 target in quiescent primary human fibroblasts (Brien *et al.*, 2015). Building on this, the second aim of this chapter was to investigate p53 target genes and gene expression profiles in various cellular conditions. To this end, I carried out chromatin immunoprecipitation, coupled to next generation sequencing (ChIP-seq) in four cellular conditions in two cell lines; DMSO, etoposide-treated, quiescent and cycling, primary IMR90 and TERT immortalised TIG3 (TIG3-T) human diploid fibroblasts. To support these observations, RNA-sequencing analyses was also performed to assess gene expression in the four stated cell conditions.

3.2 Results

3.2.1 The PHF1-p53 complex in human cells is smaller and less stable than PHF1-PRC2

It has been reported that two distinct domains, PHD1 and BD2, of PHF1 are sufficient to bind to p53 (Brien *et al.*, 2015; Yang *et al.*, 2013). However, the precise size and stoichiometry of the PHF1-p53 complex remains unknown. To explore this, I used a HEK293T cell line stably expressing FLAG-tagged PHF1 (generated previously in the lab) and performed FLAG-immunoprecipitation (FLAG-IP). P53 and EZH2 were confirmed as interactors of PHF1 by Western blot analysis (Figure 3.1A). Following FLAG-IP, the eluate was subjected to sucrose gradients ranging in sucrose density from 10-40%. Interestingly, PHF1 was observed across a broad range of fractions while p53 sedimented at a more restricted set of fractions corresponding to a lower molecular weight than that observed for EZH2 (Figure 3.1B). PRC2 is estimated to be 200-600 kDa depending on which accessory proteins are present. Since the PHF1-p53 complex sedimented higher in the sucrose gradient, this experiment suggests that the PHF1-p53 complex is smaller than PHF1-PRC2.

To characterise the relative affinity of PHF1 for p53 compared to PRC2, FLAG-Ips were performed in the presence of increasing concentrations of urea. Urea is a denaturing reagent sufficient to break protein-protein interactions. The technique of using urea to break protein-protein interactions was utilised previously to assess the stability of large protein complexes such as SWI/SNF (Kadoch & Crabtree, 2013). The PHF1-EZH2 interaction was stable in higher concentrations of urea supporting previous work demonstrating PHF1 as a stable subunit of PRC2 (Alekseyenko *et al.*, 2014; Hauri *et al.*, 2016). In contrast a much lower concentration of urea was sufficient to disrupt the PHF1-p53 interaction (Figure 3.1C). This data indicates that while PHF1 interacts with p53, the stability of the interaction is much lower than PHF1 interaction with PRC2.

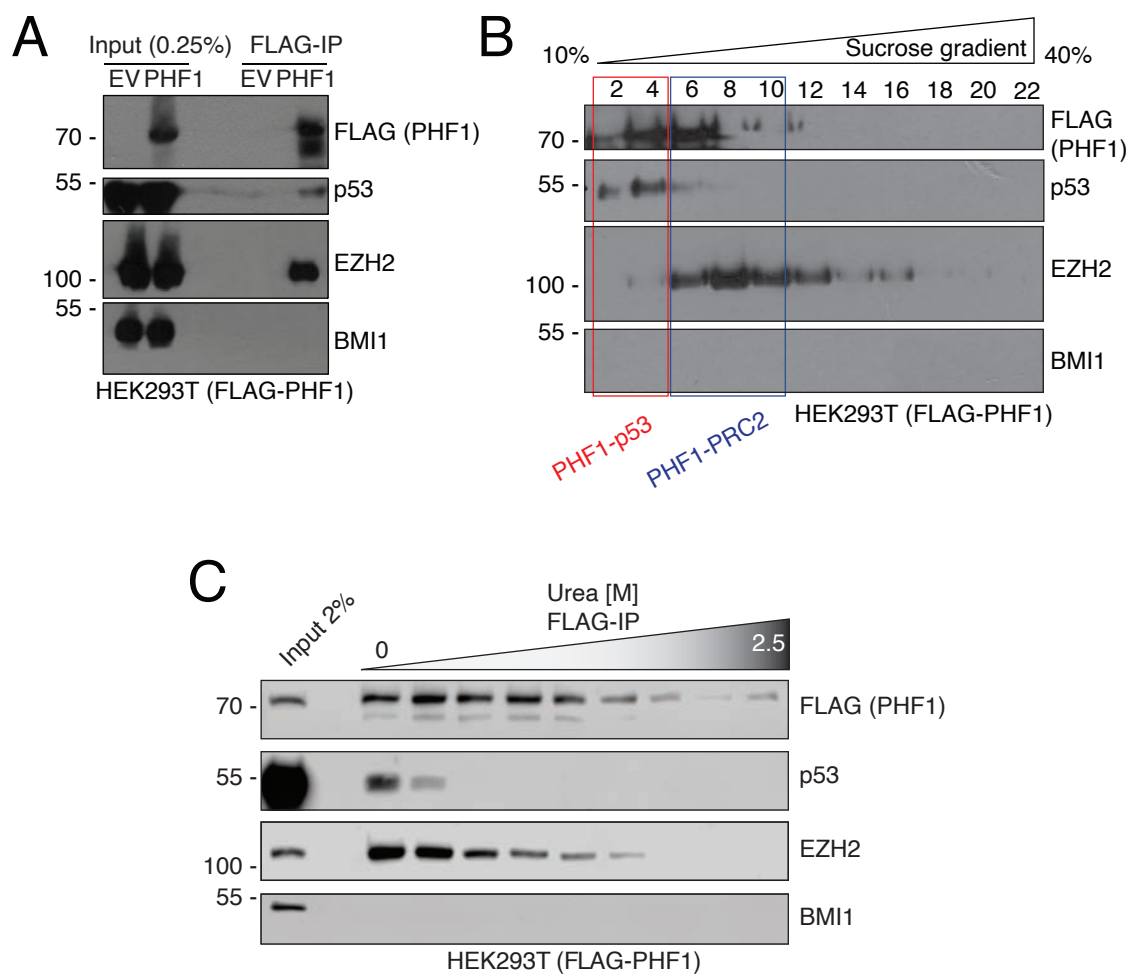


Figure 3.1 The PHF1-p53 complex is smaller and less stable than PHF1-PRC2 in human cells.

(A) Western blot analysis FLAG-PHF1 immunoprecipitation from HEK293T cells.

(B) Sucrose gradient sedimentation analysis of PHF1-p53 and PHF1-PRC2. Eluate was derived from FLAG-IPs in Panel A.

(C) Western blot analysis of FLAG-IPs performed in the presence of increasing concentrations of urea.

3.2.2 Recombinant PHF1-p53 complex is estimated to be 200 kDa

Currently, neither a crystal nor Cryo-EM structure of the PHF1-p53 complex exists and similarly, the structure for the more stable PHF1-PRC2 has not yet been solved. PHF1 is predicted to be highly disordered protein, particularly towards the C-terminus, and PHF1-PRC2 has proved challenging for structural studies (Finogenova *et al*, 2020). In an effort to better understand the biochemistry of the PHF1-p53 complex and to validate the observations made in human cells, I co-expressed the PHF1 and p53 proteins using a baculovirus expression system (BEVS) in insect cells and then purified a recombinant PHF1-p53 complex. To generate baculovirus to use for insect cell co-infection, a plasmid containing ORF of the protein of interest was co-transfected into *Spodoptera frugiperda* 9 (Sf9) cells with a linearised fragment of DNA containing the baculovirus packaging and assembly information. The resulting supernatant recovered from Sf9 cells was used to propagate additional baculovirus by re-infection of Sf9 (Figure 3.2A). Western blot analysis of Sf9 lysate confirmed successful baculovirus generation prior to larger scaler infection of Hi5 (Figure 3.2B). High 5 (or Hi5) are derived from the parental *Trichoplusia ni* cells and selected for their efficiency in the synthesis of large volumes of protein and thus, was used for the generation of recombinant PHF1-p53 complex. FLAG/6His-PHF1 and untagged p53 were co-expressed in Hi5 cells and purified using nickel-NTA beads (Figure 3.2C). Untagged PHF1 and FLAG/6His-p53 were also analysed in parallel. However, due to weaker expression of untagged PHF1 (Figure 3.2D), subsequent purifications were performed with FLAG/6His-PHF1 and untagged p53.

In order to assess the size of PHF1-p53 complex, eluate from nickel affinity purifications were subjected to sucrose gradients (Figure 3.3A&B). The results revealed PHF1 across a number of fractions. Monomeric PHF1, which is estimated to have a molecular weight of 62 kDa, sedimented at approximately 60-80 kDa in the sucrose gradient (Figure 3.3B). p53 band intensity increased across fractions 9-13 with a peak observed in fraction 13 (corresponding to ~200 kDa). PHF1 band intensity decreased in fraction 11 and increased in fraction 13. This overlaps with the peak in p53 in fraction 13 and would suggest that the PHF1-p53 is approximately

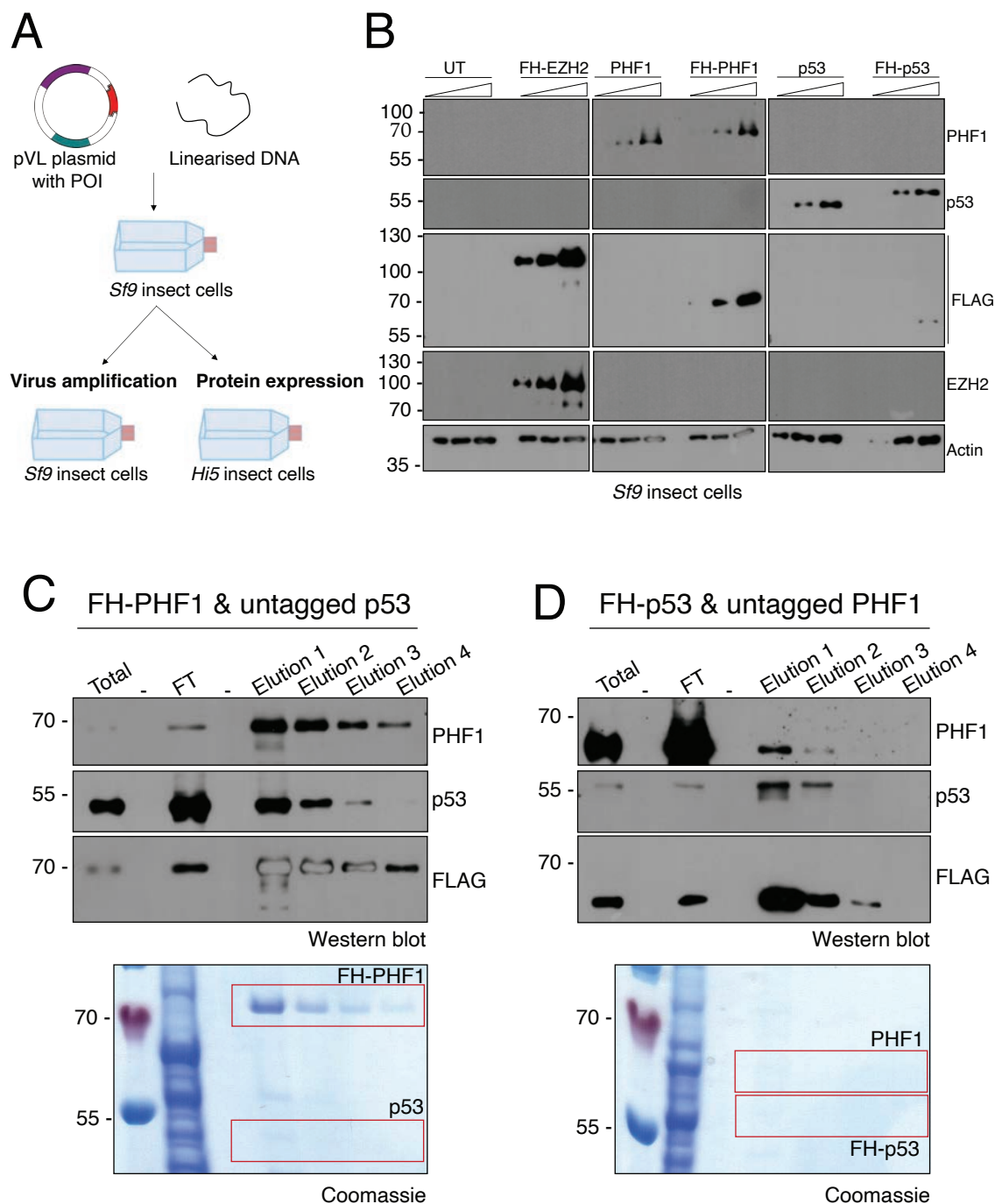


Figure 3.2 Recombinant PHF1-p53 complexes can be expressed and co-purified from insect cells.

(A) Schematic of baculovirus production and amplification in *Sf9* insect cells and protein expression in *Hi5* insect cells.

(B) Western blots confirming generation of baculovirus for FLAG/6His-tagged and untagged PHF1 and p53 in *Sf9* insect cells. FLAG/6His-tagged EZH2 was used as a positive control.

(C) Western blot analysis and Coomassie staining of Ni-NTA affinity purification of FLAG/6His-tagged PHF1 co-expressed with untagged p53.

(D) Western blot analysis and Coomassie staining of Ni-NTA affinity purification of FLAG/6His-tagged p53 co-expressed with untagged PHF1.

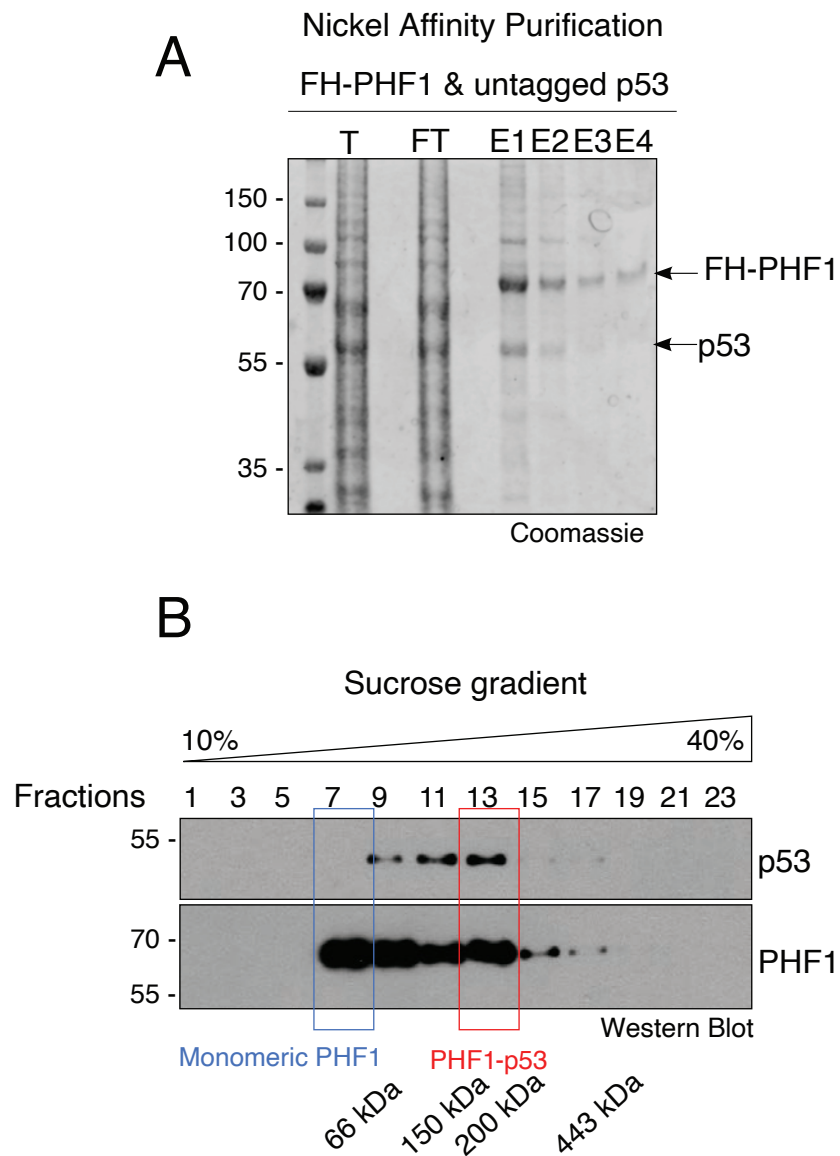


Figure 3.3 The PHF1-p53 complex is greater than 200 kDa in size.

(A) Ni-NTA affinity purification of FLAG/His-PHF1 and untagged p53 complex analysed by Coomassie staining of SDS-PAGE gel. Image was scanned using the Odessey LI-COR imaging system.

(B) Western blot analysis of FLAG/His-PHF1 and untagged p53 complex from Panel (A) subjected to sucrose gradient sedimentation. Molecular weight standards of 66, 150, 200 and 443 kDa were subjected to sucrose density gradients in parallel.

>200 kDa in molecular weight. Samples were analysed by SDS-PAGE and subsequent gels were stained with Coomassie. However, due to low protein concentration the stoichiometry of the complex could not be determined (Data not shown).

While the current approach to purify the PHF1-p53 complex using FLAG-PHF1 and untagged p53 was sufficient to recover the complex *in vitro*, the excess of monomeric PHF1 made it challenging to estimate the stoichiometry of the complex. As a means to improve the conditions to address the question of stoichiometry of the PHF1-p53 complex, a double purification system was adopted and involved StrepTactin purification followed by FLAG purification. StrepII-tagged PHF1 baculovirus (Gift of Dr. Claudio Ciferri) plus FLAG/His-p53 were used to infect Hi5 insect cells. Firstly, StrepII-PHF1/FLAG-p53 complex was purified by StrepTactin beads and the eluate was applied to FLAG beads with the intention of removing any monomeric StrepII-PHF1 (Figure 3.4A) and thus yielding a final complex of the correct stoichiometry. The PHF1-p53 complex was observed by Western blot analysis confirming that the PHF1 and p53 do interact and assembly into a complex *in vitro* and can be purified on a small scale (Figure 3.4B). I attempted to purify the complex using FLAG purification followed by StrepTactin (Figure 3.4C)(as opposed to StrepTactin followed by FLAG (Figure 3.4B)). While the complex could be recovered, the overall yield was significantly less (Figure 3.4C).

I later attempted to purify the PHF1-p53 complex on a larger scale for Cryo-EM analysis. However due to breakdown/truncation of PHF1, I was unable to purify a complex suitable for Cryo-EM (data not shown). The biochemical data presented here demonstrates that PHF1 forms a complex with p53 independently of its interaction with PRC2. Analysis of expression and purification of the PHF1-p53 complex in insect cells revealed that the complex is approximately 150 – 200 kDa supporting a 1:2 stoichiometry.

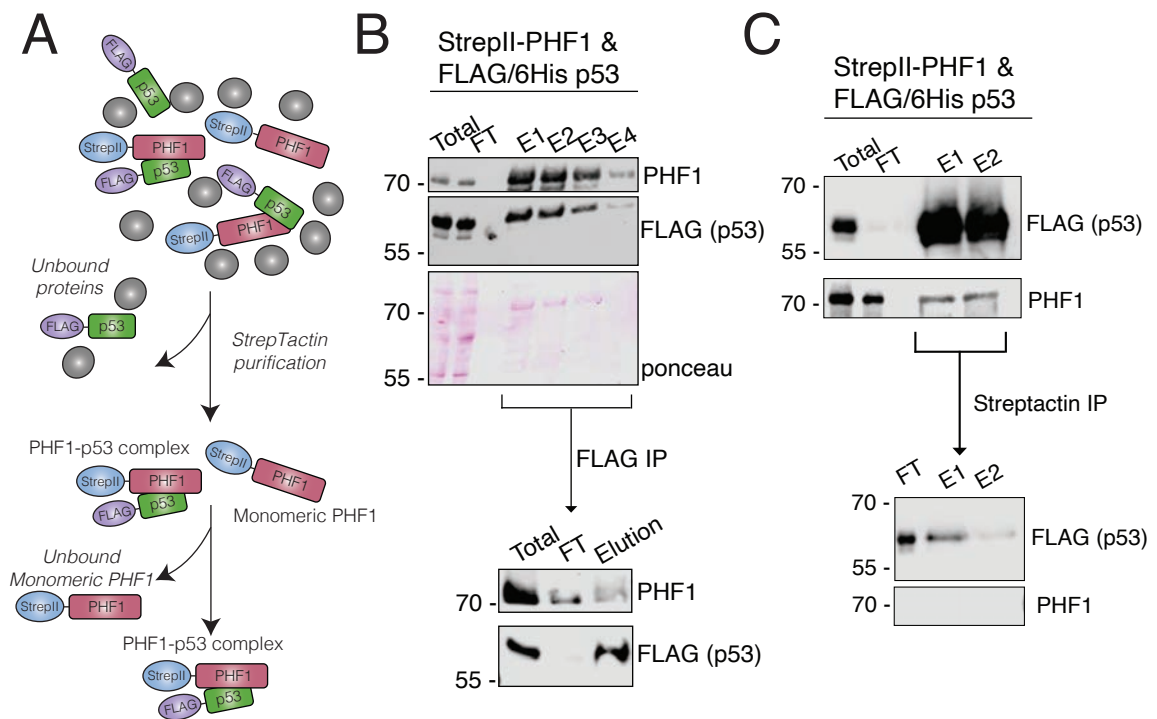


Figure 3.4 Purification of PHF1-p53 complex of the correct stoichiometry.
 (A) Schematic for purification of StrepII-PHF1 and FLAG-p53. Lysate was subjected to StrepTactin beads and the subsequent eluate was applied to FLAG beads to isolate a PHF1-p53 complex of the correct stoichiometry.
 (B) StrepTactin purification of StrepII-PHF1-FLAG-p53 complex followed by FLAG purification.
 (C) FLAG purification of StrepII-PHF1-FLAG-p53 complex with subsequent StrepTactin purification.

3.2.3 p53 is enriched on the promoters of p53 cell cycle target genes following serum starvation of primary human fibroblasts

It was previously reported that a PHF1 is a p53-target in quiescent human primary fibroblasts with a p53-like binding motif located within the PHF1 promoter whereas MTF2 and PHF19 are E2F-regulated (Brien *et al.*, 2015). Furthermore, ChIP-qPCR was used to confirm the interaction between the p53 motif within the PHF1 promoter and p53 (Brien *et al.*, 2015).

To assess whether additional quiescent-specific p53 binding sites could be identified throughout the genome along with the PHF1 promoter, I performed ChIP-sequencing using a p53 antibody in quiescent, cycling, asynchronous and DNA-damaged primary human fibroblasts (Figure 3.5A). The role of p53 in the DNA-damage response has been well established in the literature and was included in the ChIP-sequencing experiment to serve as a positive control for p53 binding (Fischer, 2017). To induce a DNA-damage response in IMR90 primary human fibroblasts, the cells were treated with etoposide at varying concentrations and durations and it was consequently decided that 1.25 μ M etoposide for 12 hours was sufficient to induce the desired response. Fibroblasts were treated with either DMSO or etoposide and harvested after 12 hours. For the quiescent and cycling cells, fibroblasts were cultured in a serum-free media for 96 hours and subsequently released into the cell cycle for 24 hours with the replenishment of 10% serum (Figure 3.5A). ChIP-qPCR analysis confirmed elevated levels of p53 bound to p53 target genes in the etoposide-treated and quiescent cells, while the PHF19 promoter served as a negative control (Figure 3.5B). ChIP-qPCR was performed at the PHF1 promoter, but no p53 was detected. Given that PHF1 was predicted to be a weaker p53 target than CDKN1A in the 2015 paper, I decided to proceed with next generation sequencing (NGS) of these p53 ChIPs to rule out any qPCR technical issues.

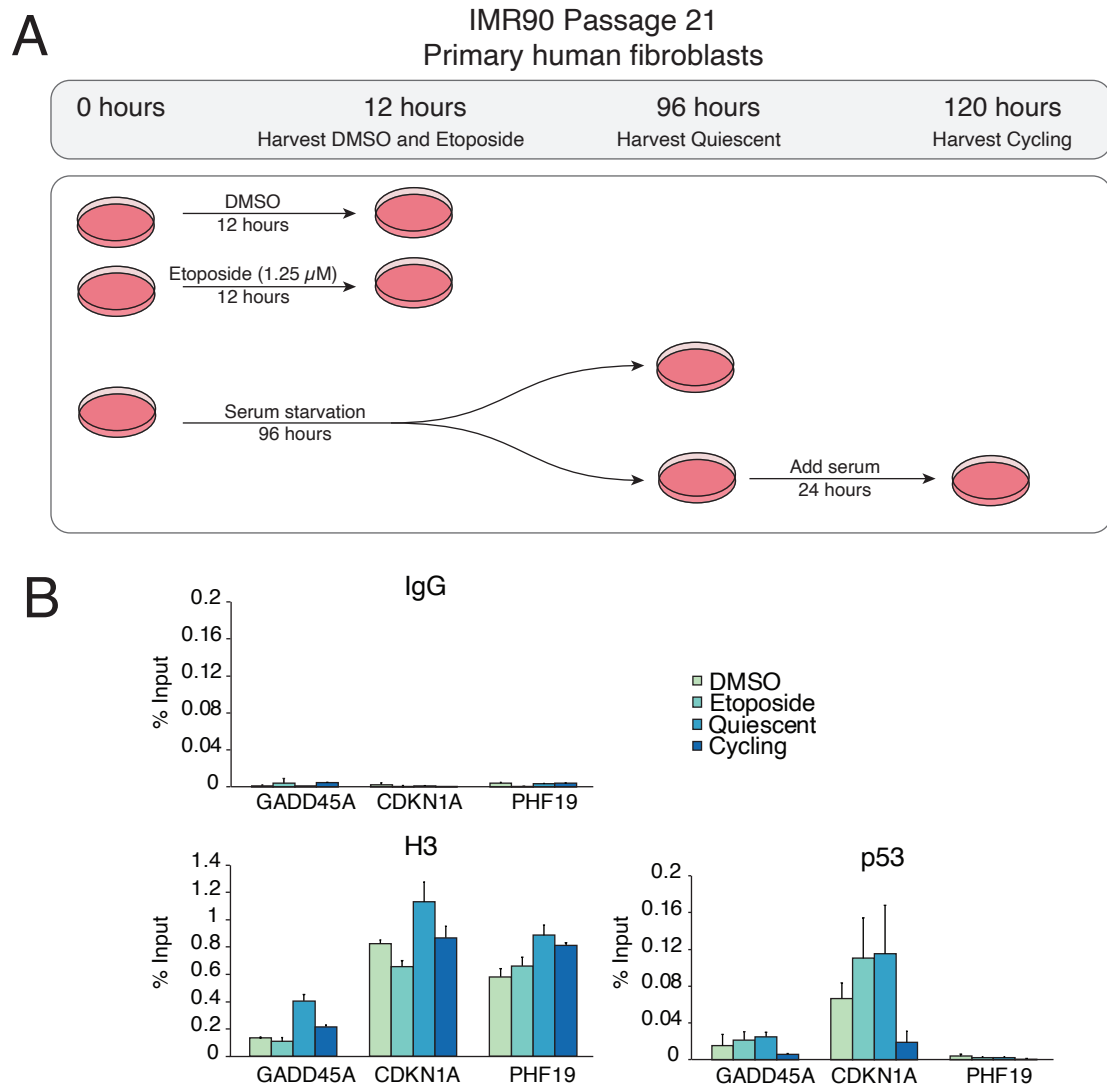


Figure 3.5 p53 is enriched on the promoters of cell cycle target genes following DNA-damage or serum starvation of primary human fibroblasts. (A) DMSO and Etoposide (at a final concentration of 1.25 μ M) were added to IMR90 primary human fibroblasts for 12 hours. Serum was removed from fibroblasts for 96 hours and subsequently harvested (Quiescent). The remaining cells were replenished with serum for 24 hours followed by harvesting (Cycling). (B) Quantitative chromatin immunoprecipitation (ChIP) analyses using the indicated antibodies in fibroblasts as described in Panel A. The eluted DNA fragments were used for Quantitative PCR analysis. p53 targets sites (GADD45A and CDKN1A) served as positive controls while PHF19 served as negative. ChIP enrichments are presented as the percentage of protein bound normalised to input.

3.2.4 p53 is significantly enriched at target sites in DNA-damaged and Quiescent primary human fibroblasts and less so in DMSO-treated and cycling

Genome wide analysis of p53 binding profiles in DMSO, etoposide-treated, quiescent and cycling cells revealed a large overlap of p53 target genes in etoposide-treated and quiescent cells (Figure 3.6A). While this was expected for these cellular conditions, analysis overlaying the target genes from the two conditions revealed that some target genes were specific to either condition (Figure 3.6A). For example, CDKN1A (alternatively known as p21) was confirmed as a p53 target common in both cell conditions (Figure 3.6B, middle panel). In contrast, etoposide-treated cells exhibited binding of p53 at DNA-damage-related targets such as DDB1 (3.6B, left panel). Following serum-starvation and subsequent cell cycle arrest at G0, several p53 targets were identified as slightly enriched in quiescent compared to etoposide-treated cells e.g. the ANKRD12 gene locus (Figure 3.6B, right panel). P53 was not localised on the extended PHF1 gene locus in any of the cell contexts studied. Overall, p53 binding at target genes was most enriched in etoposide-treated and quiescent cells. However, , it must be noted p53 was also enriched, albeit not as high, at some p53 target sites in DMSO-treated cells, for example the CDKN1A locus (Figure 3.6B). The asynchronous population of cells at various stages of the cell cycle likely results in elevated p53 in some of the cells undergoing various cell cycle checkpoints.

HOMER (Hypergeometric Optimization of Motif EnRichment) analysis of p53 peaks in the four cell conditions indicated that a large percentage of p53 peaks fall within intergenic regions (40%-62%). However, following DNA-damage or serum starvation, the number of p53 targets sites within introns/exons and TSS was increased (Figure 3.6C).

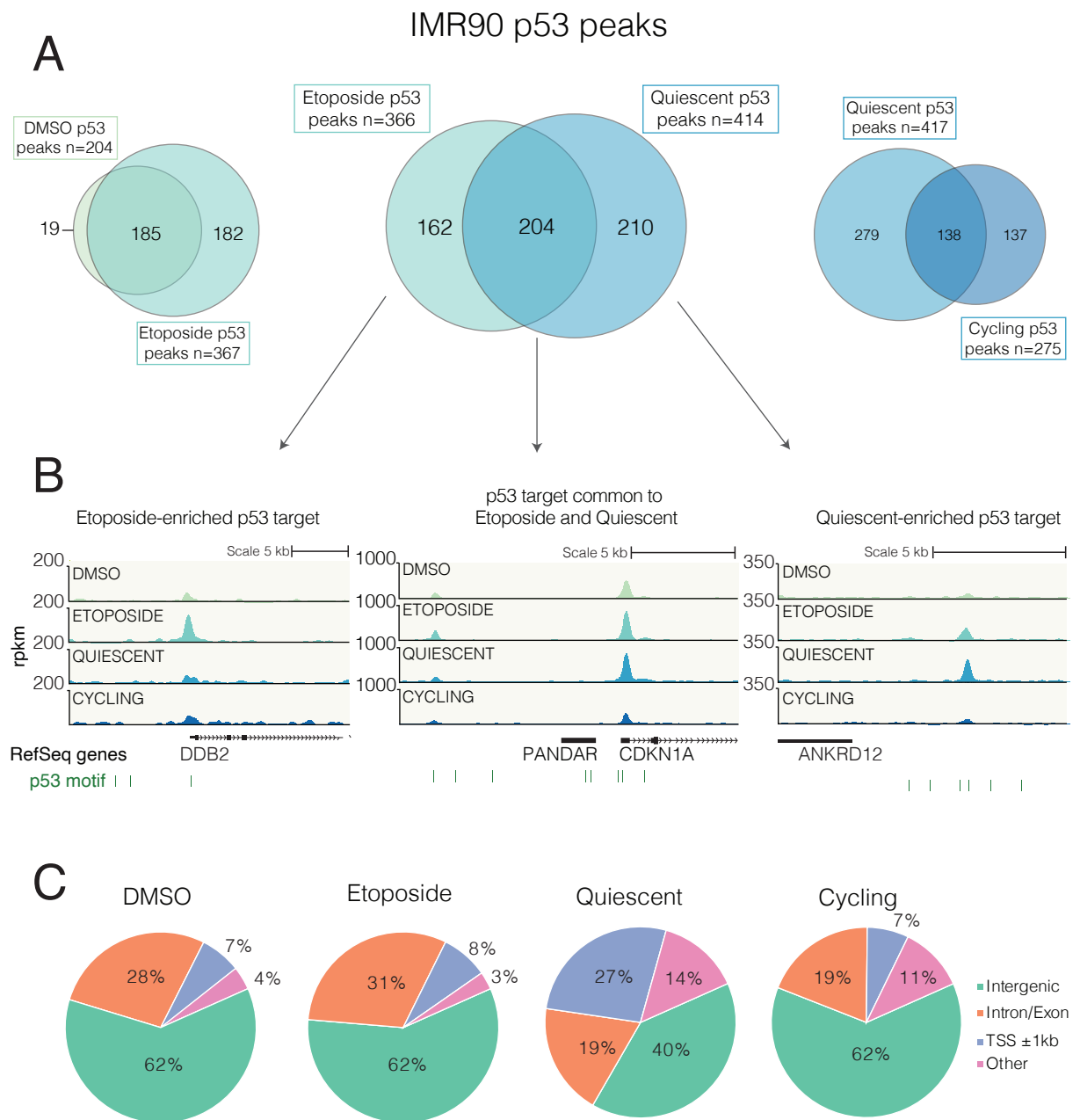


Figure 3.6 p53 is significantly enriched at target sites in DNA-damaged and Quiescent primary human fibroblasts and less so in DMSO-treated and cycling.

(A) Venn diagram comparison of p53 peaks in “DMSO vs Etoposide”, “Etoposide vs Quiescent” and “Quiescent vs Cycling” IMR90 cells.

(B) Genome browser representations of RPKM-normalised reads for p53 in DMSO, Etoposide, Quiescent and Cycling IMR90 cells. An Etoposide-specific gene locus (DDB2), a site common to both Etoposide and Quiescent (CDKN1A) and a Quiescent-specific site (ANKRD12) are represented.

(C) Proportions of p53 ChIP-seq peaks annotated to different genomic regions in DMSO, Etoposide, Quiescent and Cycling IMR90 cells.

3.2.5 p53 is significantly enriched at target sites in DNA-damaged and Quiescent TIG3-T immortalised fibroblasts and less so in DMSO-treated and cycling

Next, I performed a p53 ChIP-sequencing experiment in TIG3-T, an immortalized fibroblast cell line, in order to compare whether there was equivalent p53 binding compared with primary fibroblasts. Cells were treated as described previously (Figure 3.6A) and ChIP-qPCR analysis of p53 at the CDKN1A locus confirmed that TIG3-T were sufficiently DNA-damaged or serum-starved prior to sequencing (Figure 3.7A).

There was a large overlap in the number of p53 targets identified in etoposide-treated and quiescent cells (Figure 3.7B). As expected, addition of etoposide to the cells was sufficient to induce a DNA-damage response and p53 was bound to DNA-damage-related p53 targets such as RR2MB (Figure 3.7C, left panel). P53 was located on targets related to cell cycle, such as CDKN1A, in both etoposide and quiescent cells (Figure 3.7C, middle panel). Similar to the results in primary human fibroblasts, ANKD12 was identified as a p53 target in quiescent cells along with several peaks (Figure 3.7C, right panel), however, PHF1 was not identified as a target site in the TIG3-T cell line either. Overall, the majority of p53 targets in all cell contexts studied are located within intergenic regions (72-87%) with the number of intron/exon and TSS peaks increasing following etoposide treatment or serum starvation (Figure 3.7C). Taken together, the results presented here have identified a cohort of p53 target genes in two separate cell lines and four cell conditions (DMSO, etoposide-treated, quiescent and cycling).

3.2.6 Genes related to cell cycle inhibition are elevated in quiescent cells while cell division and DNA replication genes are elevated in cycling TIG3-T upon addition of serum

p53 binding profiles are dynamic between DMSO, etoposide-treated, quiescent and cycling fibroblasts (Figure 3.6&7). To extend on these ChIP analyses and assess the gene expression patterns in the four cell conditions, I carried out RNA-seq. As

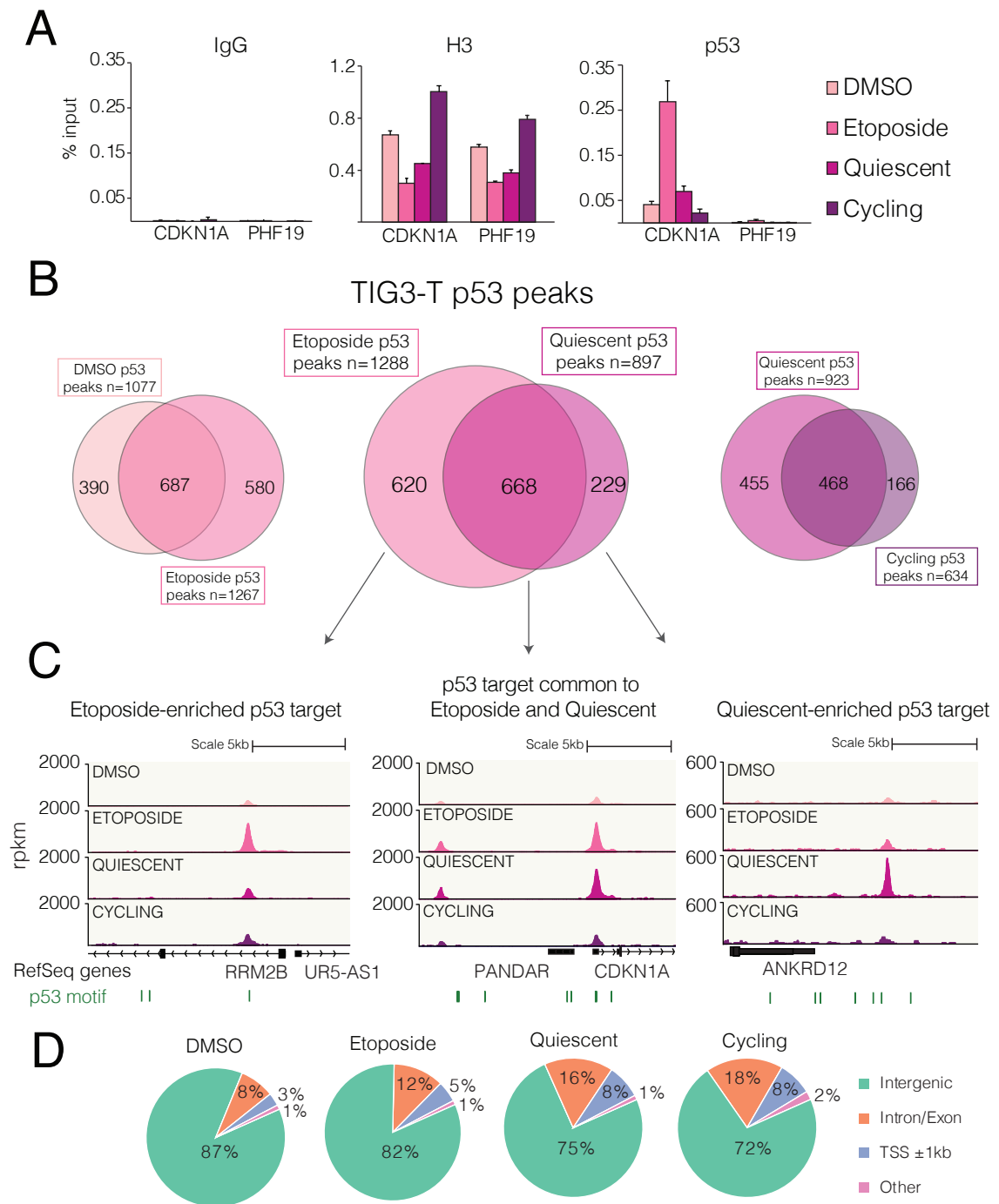


Figure 3.7 p53 is significantly enriched at target sites in DNA-damaged and quiescent immortalised human fibroblasts and less so in DMSO-treated and cycling.

(A) Quantitative chromatin immunoprecipitation (ChIP) analyses using the indicated antibodies. The eluted DNA fragments were used for Quantitative PCR analysis. ChIP enrichments are presented as a percentage of protein bound normalised to input.

(B) Venn diagram comparison of p53 peaks in “DMSO vs Etoposide”, “Etoposide vs Quiescent” and “Quiescent vs Cycling” TIG3-T cells.

(C) Genome browser representations of RPKM-normalised reads for p53 in DMSO, Etoposide, Quiescent and Cycling TIG3-T cells. An Etoposide-specific gene locus (RRM2B), a site common to Etoposide and Quiescent (CDKN1A) and a Quiescent-specific site (ANKRD12) are represented.

(D) Proportions of p53 ChIP-seq peaks annotated to different genomic regions in DMSO, Etoposide, Quiescent and Cycling TIG3-T cells.

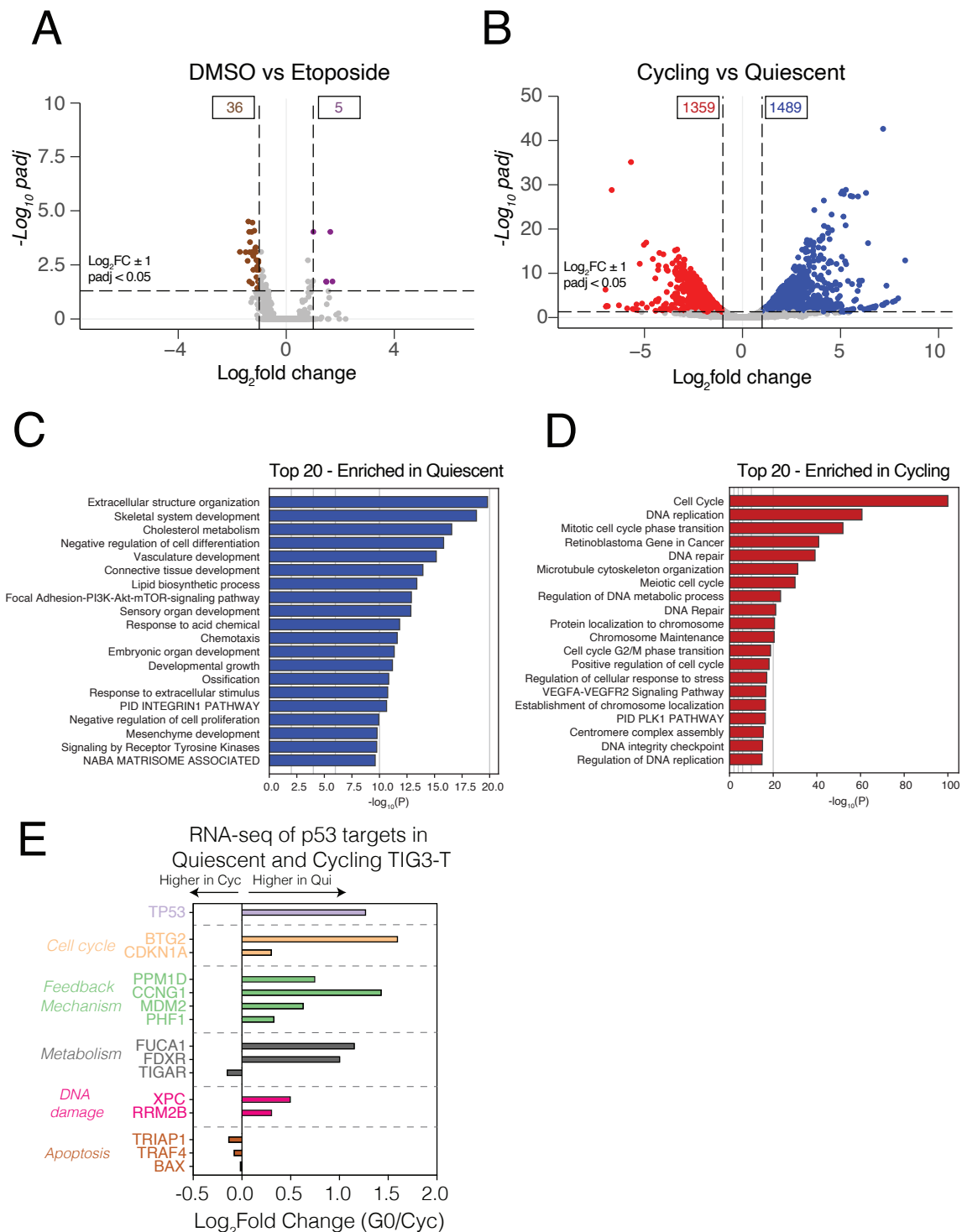


Figure 3.8 Genes related to cell cycle arrest are elevated in Quiescent cells while cell division and DNA replication genes are elevated in Cycling TIG3-T upon addition of serum.

(A) Volcano plots of RNA-sequencing data for DMSO-treated, Etoposide-treated and (B) Quiescent and Cycling TIG3-T cells. The $\log_2$ fold change of normalised gene counts for Quiescent over Cycling is represented on the X-axis and the $-\log_{10}$ adjusted p value is represented on the Y-axis.

(C) and (D) The Top 20 Gene ontology (GO) terms for genes upregulated in quiescent or cycling cells with the strongest enrichment based on p-value. GO obtained using the Metascape database [<http://metascape.org>].

(E) $\log_2$ Fold Change (G0/Cyc) of p53 target genes in Quiescent over Cycling. p53 target genes were obtained from a review by Fisher (2017).

described in the previous sections of this chapter, TIG3-T immortalised fibroblasts were cultured in a serum-free medium for a total of 96 hours to induce quiescence and following addition of serum for 24 hours, harvested as cycling cells. Asynchronously growing fibroblasts were treated with DMSO or etoposide at a concentration of 1.25 μ M and harvested after 12 hours (Figure 3.5A). Following RNA extraction, poly-adenylated mRNA was captured using oligo-dT beads and used for NGS library generation.

Surprisingly, RNA-sequencing analysis revealed relatively few gene expression changes between DMSO or etoposide-treated TIG3-T cells (Figure 3.8A). In contrast, the RNA-seq analysis between cycling and quiescent cells revealed approximately 1400 genes were upregulated in quiescent TIG3-T cells relative to cycling cells (highlighted in blue) (Figure 3.8B). A similar number of genes were found to be upregulated in cycling cells compared to quiescent (highlighted in red) (Figure 3.8B).

Gene Ontology (GO) analysis revealed that among the most enriched genes that get upregulated in quiescent cells were those relating to the negative regulation of cell differentiation and proliferation (Figure 3.8C). In contrast, genes relating to cell cycle and DNA replication were among the most upregulated in cycling cells (Figure 3.8D). The elevated expression of p53 found in quiescent cells compared to cycling cells correlated with the upregulation of several p53 target genes relating cell cycle arrest such as CDKN1A and BTG2 (Figure 3.8E).

3.2.7 Serum starvation of TIG3-T cells shifts the expression of PRC2 components from EZH2/PHF19 in cycling cells to EZH1/PHF1 in quiescent cells

While PHF1 was not identified as a p53 target in IMR90 primary cells or TIG3-T immortalised cells, expression of PHF1 was elevated in quiescent compared to cycling TIG3-T fibroblasts. Previously, it was noted that the expression of PRC2 core and accessory subunits can vary depending on cell cycle or cell type (Brien *et*

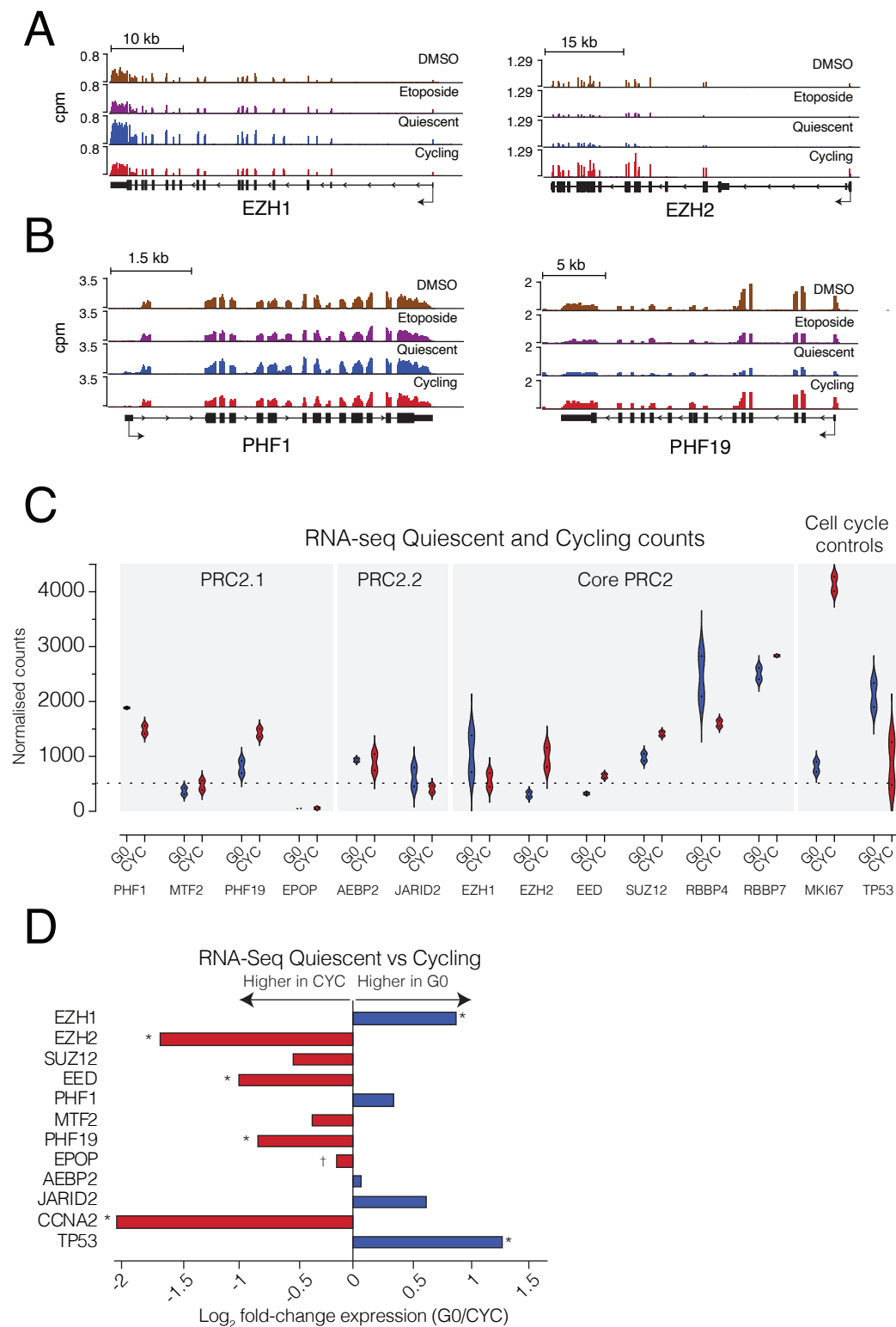


Figure 3.9 Serum starvation of TIG3-T shifts the expression of PRC2 components from EZH2/PHF19 to EZH1/PHF1.

(A) Genome browser representation of EZH1 and EZH2 and (B) PHF1 and PHF19 in DMSO-treated, Etoposide-treated, Quiescent and Cycling TIG3-T.

(C) Normalised counts for each PRC2 component in quiescent and cycling TIG3-T cells. Each replicate is represented by a single point and the median is represented by the solid line.

(D) Log₂ fold-change expression of PRC2 subunits in quiescent and cycling-TIG3-T. * $p < 0.05$, † denote read counts were less than 60

al., 2015; Oliviero *et al.*, 2016; Xu *et al.*, 2015). Therefore, I decided to assess the expression of other PRC2 components in quiescent and cycling TIG3-T cells.

The changes in expression of PRC2 components in DMSO and etoposide-treated cells were negligible (Figure 3.9A&B). However, serum-starvation and release of TIG3-T cells highlighted a switch in expression from EZH1 in quiescent cells to EZH2 in cycling cells (Figure 3.9A). A similar expression pattern was observed between PHF1 and PHF19 with elevated PHF1 in quiescent and elevated PHF19 in cycling TIG3-T (Figure 3.9B). Of the three Polycomb-like proteins, PHF1 expression in quiescent cells was highest compared to MTF2 and PHF19 in line with previous studies in human fibroblasts (Brien *et al.*, 2015). The overall expression levels of MTF2, EPOP and JARID2 were found to be significantly lower compared with PHF1, PHF19 and AEBP2 and a slight increase in EED and SUZ12 expression was observed in cycling cells (Figure 3.9C). The overall RNA-sequencing analysis highlighted a distinct switch in expression from EZH1 and PHF1 in quiescent cells to EZH2 and PHF19 in cycling (Figure 3.9D).

3.3 Discussion

While efforts were made to study the PHF1-p53 complex in vitro, the exact molecular weight, composition and overall structure of the complex still remains unclear. I observed that the recombinant PHF1-p53 complex was approximately 200 kDa based on my sucrose density gradients analysis. Based on this, I hypothesise that since PHF1 is approximately 60 kDa, it most likely interacts with a dimeric form of p53. Despite several attempts to purify the PHF1-p53 complex from insect cells, the inherent instability and degradation of PHF1 proved challenging when purifying the complex. Towards improving the purification of the PHF1-p53 complex, I have generated a PHF1 construct containing two tagged moieties (6His- and FLAG-tag) on both the N- and C-terminus of the protein, respectively. PHF1 is predicted to have a highly unstructured and disordered C-terminus. As a result, many PHF1 breakdown/truncation products were often observed following purification of the complex (Data not shown). I hypothesize that by exploiting a C-terminal tag on the protein, this will improve the purification of full length PHF1 and would aid in downstream analysis such as gel filtration to determine the size of the complex and Cryo-EM to determine the structure of the complex. Further structural characterisation of the complex would be valuable to better understand the composition of the complex and provide insight on how the disruption of this interaction could force quiescent cells to re-enter the cell cycle.

The results of the p53 ChIP-sequencing were surprising as previously published data suggested that p53 binds to the PHF1 promoter in quiescent cells, thus driving its expression (Brien et al, 2015). However, the results from the ChIP-sequencing in both primary or immortalised fibroblasts were not in agreement with these observations. Some potential explanations of these differences may be due to the (1) cell line background and passage number, (2) weaker antibody affinity than used in the 2015 paper and (3) variance in cross-linking of chromatin material for ChIP-sequencing. While the cell line used in the 2015 paper and the cell lines used in these experiments were all derived from embryonic lung tissue, PHF1 was not a target in the cell lines I used for ChIP-sequencing. Alternatively, I speculate that the p53 antibody used in these experiments may have weaker affinity compared to previous studies and it could be valuable to increase antibody concentration or try

an alternate antibody in future studies. Finally, cross-linking of chromatin for the ChIP experiments presented in this thesis was performed in suspension while cross-linking of material for the 2015 paper was performed by addition of formaldehyde directly to the cells on tissue culture dishes. It is possible that the differences observed (i.e. the absence of p53 on the PHF1 promoter) between these two sets of ChIPs may be due to cross-linking and thus it would be valuable to perform cross-linking using both methods in parallel.

Irrespective of the inability to detect p53 on the PHF1 promoter, the ChIP-sequencing of p53 in two separate cell lines are valuable datasets for studying p53 target genes in human fibroblasts. I have identified p53 target sites in four cellular conditions in two cell lines. At least half of all p53 peaks identified in IMR90 and TIG3-T cells were located in intergenic regions and these numbers are in line with previously published p53 ChIP data (Idogawa *et al*, 2014; Nguyen *et al*, 2018; Sammons *et al*, 2015). Given that these intergenic peaks make up such a large portion of p53 binding in the genome, it could be interesting to investigate the biological importance or function of these targets. Additionally, while p53 target genes have been well studied in the various asynchronous and DNA-damaged cell lines (Nguyen *et al*, 2018), p53 target genes have been poorly characterised in quiescent cells and thus it would be valuable to validate the “quiescent-enriched” p53 peaks identified in this thesis in additional quiescent primary cell models.

RNA-sequencing analysis in four cell conditions (DMSO, etoposide-treated, quiescent and cycling fibroblasts) further complements the observations made in the p53 ChIP-sequencing. The alterations in gene expression in response to etoposide treatment was significantly less than expected, yet it is important to note that the concentration of etoposide (1.25 μ M for 12hr) was lower than typically used in published RNA-sequencing studies (20 μ M for 6hr, 25 μ M for 2hr) (Boucas *et al*, 2015; Wei *et al*, 2018). The results presented in this chapter demonstrated that the concentration of etoposide was sufficient to induce a DNA-damage response, as evidenced by p53 bound to DNA damage-associated p53 targets. It is possible that a higher concentration and longer time course would result in more significant differential expression. In furthering the work presented in this chapter, it would be

interesting to continue analysis of p53 target genes identified in the p53 ChIPs and their correlation in gene expression.

The results presented in this research chapter have confirmed the *in vitro* and *in vivo* assembly of a PHF1-p53 complex. Biochemical analysis of the complex has confirmed that the complex is smaller in size and is less stable than PHF1-PRC2. P53 ChIP-seq has highlighted p53 target sites in two cell lines; IMR90 and TIG3-T, and four cell conditions; DMSO, etoposide, quiescent and cycling cells. RNA-seq analysis in matched cell conditions has confirmed the expression of several p53 target genes in quiescent cells and also highlighted a distinct switch in expression of PRC2 components from EZH1/PHF1 in quiescent cells to EZH2/PHF19 in cycling cells.

Chapter 4: **Biochemical analysis of PRC2 in quiescent cells**

4.1 Introduction

PRC2 can be further subdivided into at least two mutually exclusive sub-complexes, called PRC2.1 and PRC2.2, based on the association of the core proteins with various accessory proteins (Alekseyenko *et al.*, 2014; Hauri *et al.*, 2016). PRC2.1 is defined as containing one Polycomb-like protein (PHF1, PHF19 or MTF2) together with either EPOP or PALI1/2, while PRC2.2 comprises of AEBP2 and JARID2 (Beringer *et al.*, 2016; Conway *et al.*, 2018; Grijzenhout *et al.*, 2016; Liefke *et al.*, 2016). EZH1 and EZH2 are the two mutually exclusive H3K27 specific histone methyltransferases that drive the catalytic activity of PRC2. While the expression of EZH2 is regulated by E2F transcription factors and found to be highly expressed in proliferating cells (Bracken *et al.*, 2003), the expression of EZH1 has been reported to be higher in non-proliferating and terminally differentiated cells (Brien *et al.*, 2015; Margueron *et al.*, 2008; Xu *et al.*, 2015).

Interestingly, a recent study of the African turquoise killifish (*Nothobranchius furzeri*) identified EZH1 is amongst the most upregulated gene during the embryonic diapause state (Hu *et al.*, 2020). This diapause is a reversible quiescent-like state during killifish embryonic development during which cells are not proliferating or growing (Hu & Brunet, 2018). This killifish is native to transient ponds within regions of Zimbabwe and Mozambique and in order to survive the long spells of droughts present for 6-8 months of the year in these regions, the killifish enter a state of dormancy called diapause. However, the mechanism in which EZH1 contributes to this quiescent-like state is unclear.

Like EZH2, the Polycomb-like proteins MTF2 and PHF19 have been reported to be E2F-regulated and are highly expressed in rapidly growing cells, whereas PHF1 was reported to be a p53 target and found to be primarily expressed in slower growing or quiescent cells (Brien *et al.*, 2015; Pasini *et al.*, 2004; Tasdemir *et al.*, 2016). Quiescence (also referred to as G0) is a reversible cellular state during which cells are not proliferating but ready to re-enter the cell cycle upon external stimulus (Cheung & Rando, 2013). Due to the inherent non-proliferating state of quiescent cells and their limited availability in vivo, studying their underlying biology can be

challenging. However, it has been established that the removal of serum from asynchronously growing cells can induce quiescence in many cell types and serum starvation of fibroblasts has been used as a model of cellular quiescence in many studies (Bracken *et al.*, 2003; Brien *et al.*, 2015; Pasini *et al.*, 2004; Tasdemir *et al.*, 2016). The experiments presented within this chapter use this strategy to induce cellular quiescence in TERT immortalised TIG3 human fibroblasts (TIG3-T).

In this research chapter, the main aims are to characterise the composition of PRC2 within quiescent, non-dividing cells and contrast this to its composition in proliferating cells. While preliminary, unpublished data within the Bracken lab initially suggested a quiescent PRC2 complex existed lacking SUZ12, I have shown that in the absence of SUZ12, the remaining PRC2 components are largely unstable. Through endogenous immunoprecipitation of core PRC2 components coupled with mass spectrometry, I show that SUZ12 is present in quiescent cells and together with EZH1, EED, RBBP4 and PHF1 assemble as the predominant complex found in quiescent cells, termed G0-PRC2.1. Furthermore, I have characterised the effects of degradation of PRC2 within quiescent and cycling TIG3-T cells with a degrader which specifically targets the core subunit EED.

Cancer stem cells are a population of cells that exist in a largely quiescent state within the tumour microenvironment and are thought to be resistant to chemotherapeutic drugs (Batlle & Clevers, 2017). This small population of quiescent cancer stem cells are thought drive the resurgence of tumours. Towards understanding the implications of targeting PRC2 in quiescent cancer stem cells, I have established a quiescent cancer model by serum starvation of the thyroid papillary cancer cell line, BCPAP. It has been shown that culturing of thyroid cancer cells in serum-free media resulted in a large shift towards sub-G0 or G0/G1 phase of the cell cycle. These undifferentiated papillary thyroid carcinoma cells form cellular clusters resembling “tumour spheres” and are highly resistant to chemotherapeutic drugs (Giuffrida *et al.*, 2016). Upregulation of EZH2 has been reported to drive the proliferation and migration in thyroid cancer cells and correlates with poorer survival rate in poorly differentiated thyroid cancers and so

inhibition/degradation of PRC2 could be a potential target in thyroid cancers (Masudo *et al*, 2018; Wang *et al*, 2019; Xue *et al*, 2019).

4.2 Results

4.2.1 Purification of PRC2 lacking SUZ12

SUZ12 is considered a key component of PRC2 with the N-terminal and C-terminal VEFS region essential for the interaction with auxiliary factors and catalytic components, respectively. Previous PHF1/MTF2/PHF19 endogenous immunoprecipitation experiments performed within the Bracken lab had highlighted the assembly of a PRC2 complex potentially lacking SUZ12 in quiescent cells. I decided to purify various forms of recombinant PRC2 complexes in order to assess the structural integrity of a G0-PRC2 lacking SUZ12. I generated baculovirus for each PRC2 component in *Sf9* insect cells which was subsequently used for infection of *Tni* insect cells. The complexes consisted of “G0-PRC2” containing PHF1/EED/EZH1/RBBP4, “G0-PRC2 plus SUZ12” and “PHF1-PRC2” containing PHF1/EED/EZH2/RBBP4/SUZ12 (Figure 4.1A). I hypothesised that the lack of SUZ12 in G0-PRC2 may result in complex instability and so I included affinity tags on every PRC2 component (with the exception of RBBP4) in order to recover as much of each component as possible. Insufficient protein concentration within insect cells may be detrimental to complex formation and thus first a nickel affinity purification gives the best chance for assembly of complexes by concentrating all PRC2 components. The aim for purification of PRC2 was to perform nickel-NTA purification (to recover as much of each PRC2 subunit as possible), FLAG purification (to allow for purification of a complex containing only full length PHF1 as PHF1 was C-terminally FLAG-tagged) followed by further polishing of the complexes on a sizing column.

Unsurprisingly, as all proteins contained a 6His tag (with the exception of RBBP4), all PRC2 components were purified after the first nickel purification (Figure 4.1B). However, following FLAG purification, G0-PRC2 appeared to be largely unstable compared to G0-PRC2 plus SUZ12 or PHF1-PRC2 (Figure 4.1C). Western blot analysis did show that EZH1, EED and RBBP4 and PHF1 were recovered following FLAG purification, however, the levels of protein were significantly less than SUZ12-

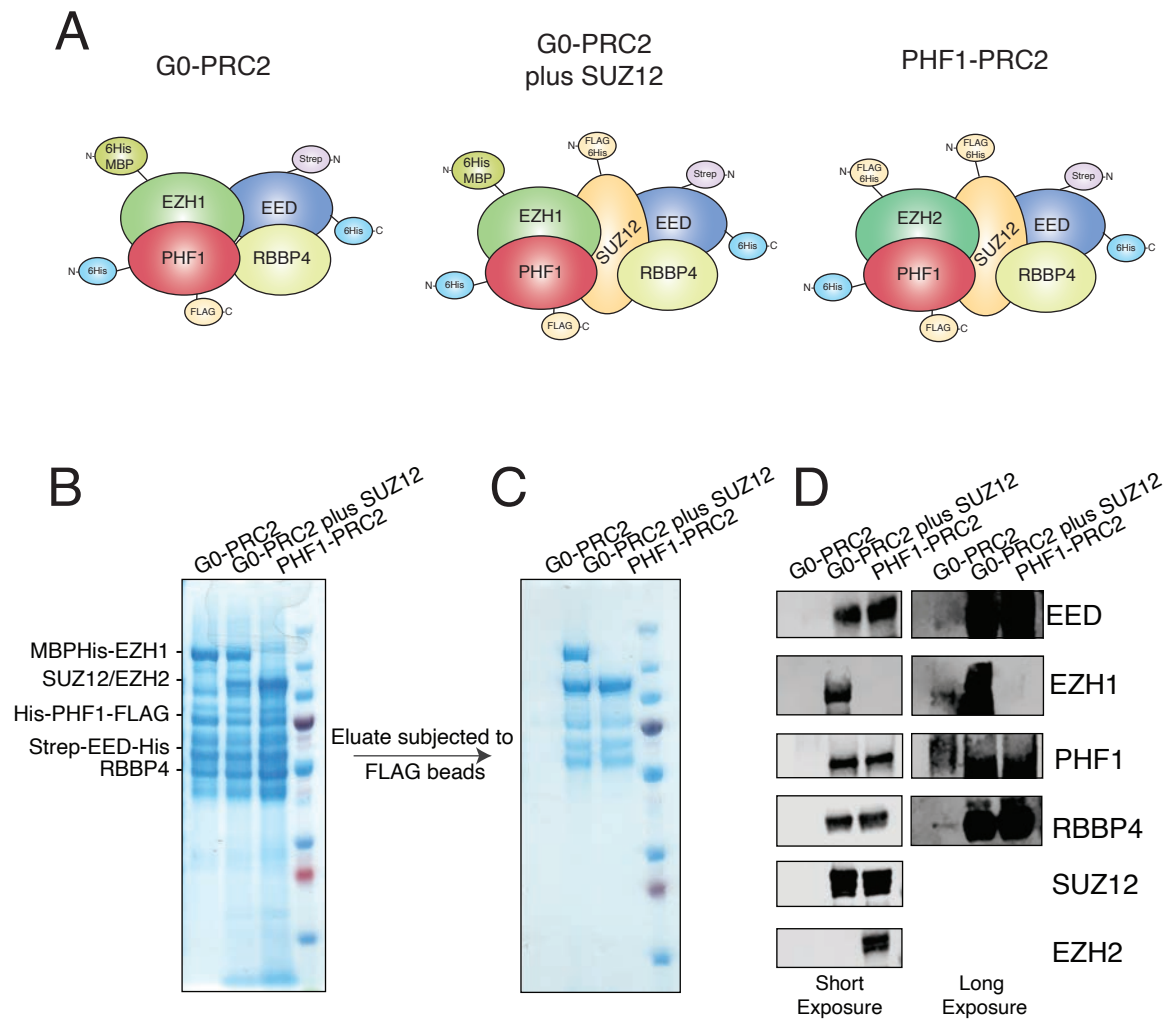


Figure 4.1 Purification of recombinant human G0-PRC2.

(A) Schematic of G0-PRC2, G0-PRC2 plus SUZ12 and PHF1-PRC2 complexes.
 (B) Coomassie-stained gel after nickel purification of PRC2 complexes.
 (C) Coomassie-stained gel after FLAG purification of PRC2 complexes from (B).
 (D) Western blot of PRC2 complex after FLAG purification in (C). Short and long exposures have been included.

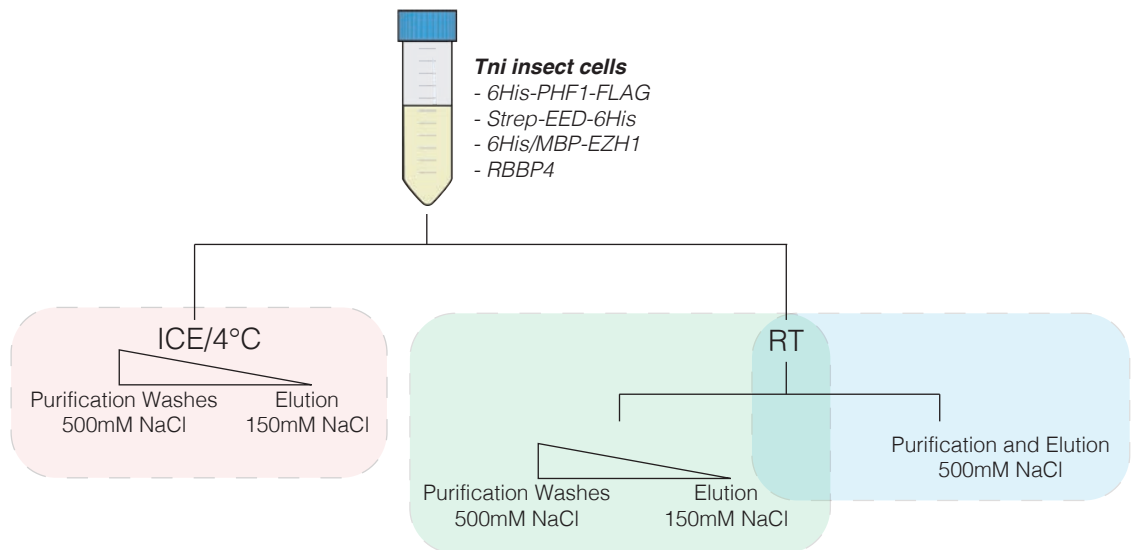
containing complexes (Figure 4.1D). Taken together, this experiment suggests that SUZ12 is required for the stability of the EZH1-PRC2 complex.

4.2.2 High salt buffer and low temperature improve the purification of G0-PRC2 subunits

Due to the low yield and recovery of G0-PRC2, I decided to focus on optimising the purification conditions for this complex. Structural prediction tools have revealed PHF1 as a highly disordered protein, particularly towards its C-terminus. It has been suggested that increasing salt concentration and temperature during the purification of disordered proteins may alleviate protein insolubility and instability (Alberti *et al*, 2018). Similarly, recent PRC2 structural papers have employed a higher salt concentration (i.e. minimum of 250 mM NaCl) throughout the affinity tag purification protocols and occasionally in the final gel filtration steps (Li *et al.*, 2017; Zhang *et al.*, 2019). Taken together, I hypothesised that maintaining G0-PRC2 (without SUZ12) in a high salt buffer could potentially improve complex stability.

The standard protocol for purification of PRC2 complexes within the Bracken lab is to perform several rounds of high salt washes (500 mM NaCl) followed by 1-2 washes in a lower salt (150 mM) with the final elutions performed in 150 mM NaCl. As noted above, temperature has been reported to improve the purification of disordered proteins and with this in mind I chose to modify the experiment by performing the standard protocol at room temperature (RT). Additionally, the purification was also performed with a high salt for both washes and elution steps at room temperature (Figure 4.2A). A significant improvement was observed in the Nickel-NTA purification of G0-PRC2 members when all washes and elutions were maintained in 500 mM NaCl with a notably better yield of PHF1. However, performing the purification at room temperature did not improve recovery of the PRC2 components (Figure 4.2B). I then decided to assess the stability of the G0-PRC2 subunits by snap-freezing the eluates in liquid nitrogen. Samples were thawed and centrifuged to remove any precipitated material prior to loading on the SDS-PAGE gel (Represented by the lanes “+” on within Figure 4.2B). Overall, EZH1,

A



B

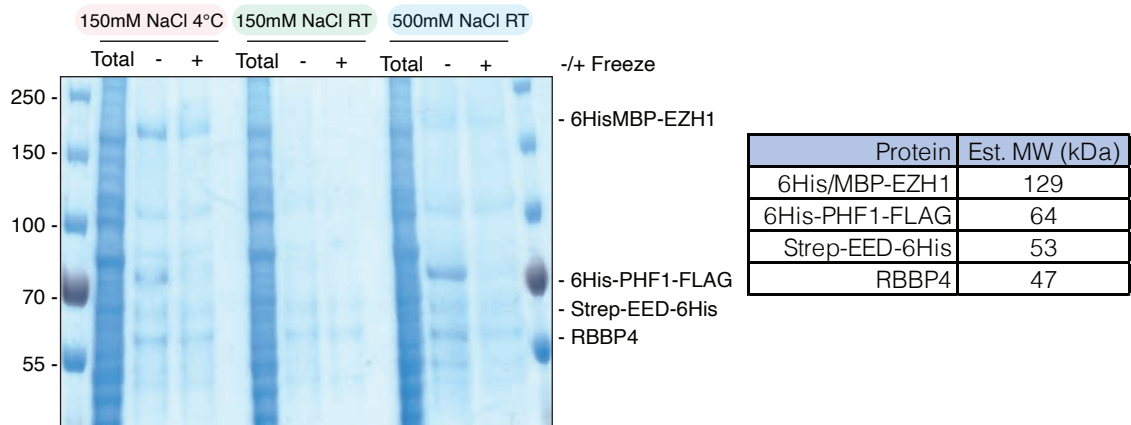


Figure 4.2 The stability of PHF1 in G0-PRC2 increases in high salt.

(A) Schematic for the purification of G0-PRC2 (6His-PHF1-FLAG, Strep-EED-6His, 6His/MBP-EZH1 and RBBP4) from *Tni* cells.

(B) Coomassie stained gel following Nickel purification of PRC2 complexes. Insect cell lysate prior to purification is indicated in the lane "Total". 10% glycerol was added to samples which were then frozen in liquid nitrogen (indicated by +)

EED, PHF1 and RBBP4 were all found to largely precipitate after a round of freeze-thaw (Figure 4.2B)

Building on the observation that PHF1 favours a higher salt buffer, I repeated another round of purifications with the aim of further improving the protocol by purification of G0-PRC2 subunits in a high salt buffer at 4°C (Figure 4.3A). The results, in agreement with the preceding purifications, show that purification of PRC2 components was significantly improved in the presence of a high salt wash/elution buffer (Figure 4.3B). Additionally, purification of the G0-PRC2 subunits in a lower temperature (i.e. 4°C) recovered more protein (Figure 4.3B). As described previously, I assessed the stability of G0-PRC2 components after a round of freeze-thaw and found the subunits to be largely unstable (Represented by the lanes “+” on within Figure 4.3B). It is evident, from all the purifications conducted, that the proteins in G0-PRC2 do not freeze particularly well which is not optimal if needed for subsequent biochemical assays.

I then decided to subject the G0-PRC2 complex to gel filtration to establish the size of the complex and to further polish the complex following affinity purification. The members of G0-PRC2 did not co-elute from the column within the same peak. EED-EZH1 appeared to elute as a dimer whilst PCL1 and RBBP4 eluted as monomers suggesting a highly instable complex (Data not shown). Taken together with the earlier observations, these efforts support the idea that the presence of SUZ12 is essential and a recombinant PRC2 complex cannot be produced in its absence.

4.2.3 The composition of PRC2 in quiescent and cycling primary human fibroblasts

Based on the preliminary endogenous immunoprecipitation data in the Bracken laboratory, the G0-PRC2 complex was hypothesised to consist of PCL1, EED, EZH1 and RBBP4. However, as described above, the recombinant purification of these proteins followed by gel filtration resulted in complex dissociation and instability. To

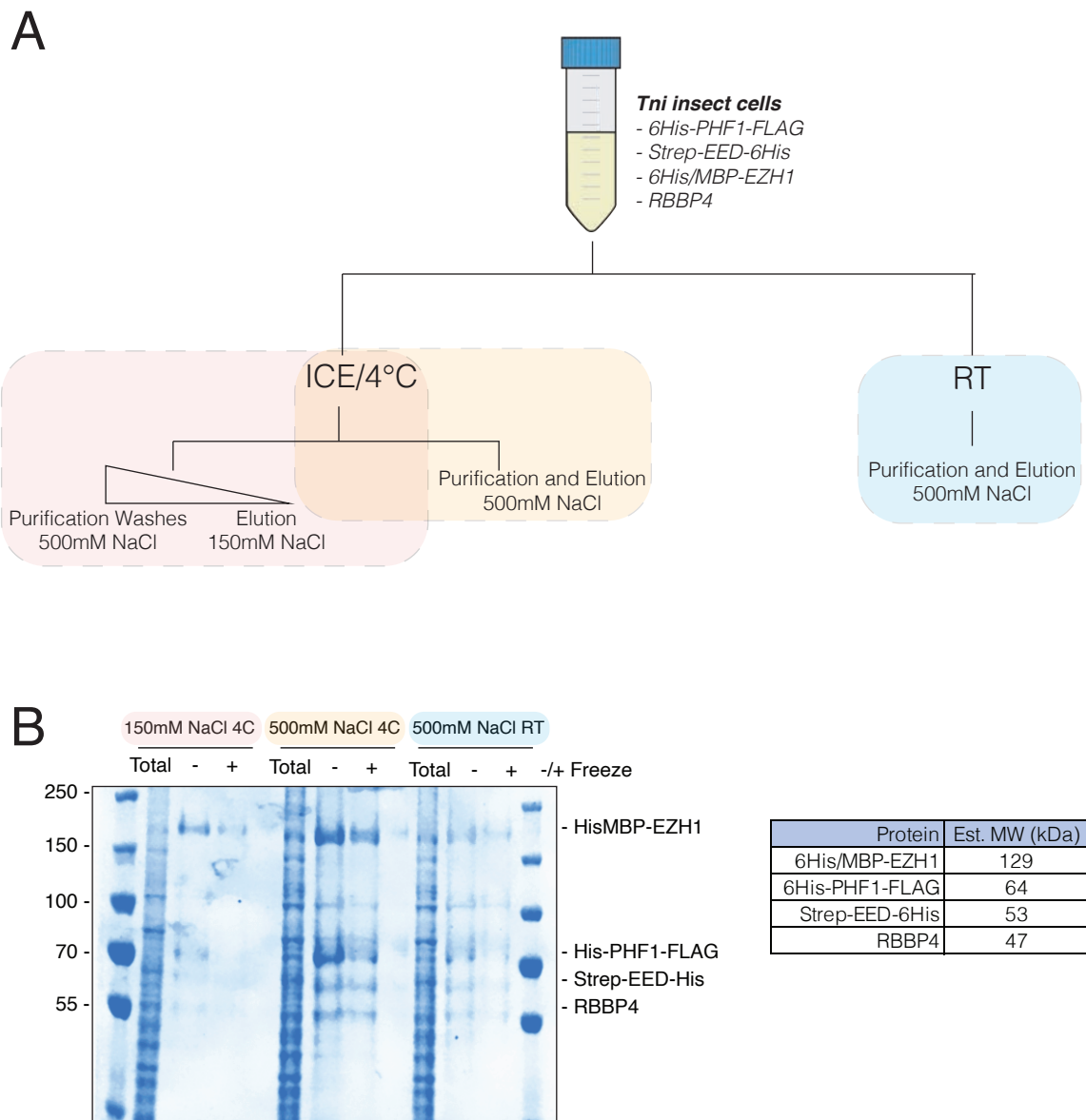


Figure 4.3 High salt and temperature are critical for G0-PRC2 purification.
 (A) Schematic for the purification of G0-PRC2 (6His-PHF1-FLAG, Strep-EED-6His, 6His/MBP-EZH1 and RBBP4) from *Tni* cells.
 (B) Coomassie stained gel following Nickel purification of PRC2 complexes. Insect cell lysate prior to purification is indicated in the lane "Total". 10% glycerol was added to samples which were then frozen in liquid nitrogen (indicated by +)

assess the various assemblies of PRC2 within quiescent and cycling cells, I performed endogenous EED, EZH1 and EZH2 immunoprecipitations (IP) coupled with mass spectrometry on nuclear lysates of serum-starved (quiescent) and serum re-stimulated (cycling) immortalised human TIG3-T fibroblasts. To induce quiescence, serum was removed from primary human fibroblasts for a total of 96 hours and cells were stimulated to re-enter the cell cycle with serum for 24 hours. The nuclear lysate from serum-starved and cycling fibroblasts were subjected to endogenous IP with antibodies against EED, EZH1, EZH2 and IgG. Immunoprecipitated material was subjected to trypsin digestion followed by LC-MS/MS and Perseus analysis (Figure 4.4A). RT-qPCR analysis was performed prior to endogenous IP to confirm cells had been sufficiently serum-starved with high levels of EZH1 and CDKN1A expression which correlated with quiescent, serum-starved cells whilst EZH2 and CCNA2 expression levels were elevated in cycling fibroblasts (Figure 4.4B).

Prior to digestion of the immunoprecipitated material and mass spectrometry, I performed Western blot analysis to confirm the presence of known PRC2-interacting proteins. Whilst the Ips performed in quiescent cells appeared to have low or undetectable levels of proteins for the indicated antibodies, the Ips performed in cycling cells did co-IP with some of the indicated PRC2-associated proteins (Figure 4.4C). An additional EZH2 antibody from Cell Signalling Technology (CST) was included but due to the low levels of PRC2 components analysed by Western blot compared to the EZH2 (AC22) antibody, mass spectrometry was only performed using the latter antibody (Figure 4.4C). Analysis of the peptide abundance confirmed that all three antibodies, (EED, EZH1 and EZH2) successfully co-immunoprecipitated PRC2-associated proteins while the IgG control did not (Figure 4.4D). Similar amounts of peptide were enriched for core PRC2 components whilst the abundance of accessory proteins varied slightly across the different antibodies in quiescent and cycling cells, however, PHF1 peptides were amongst the most abundant of all accessory proteins in both contexts suggesting that PHF1 is highly expressed in human fibroblast cells (Figure 4.4D).

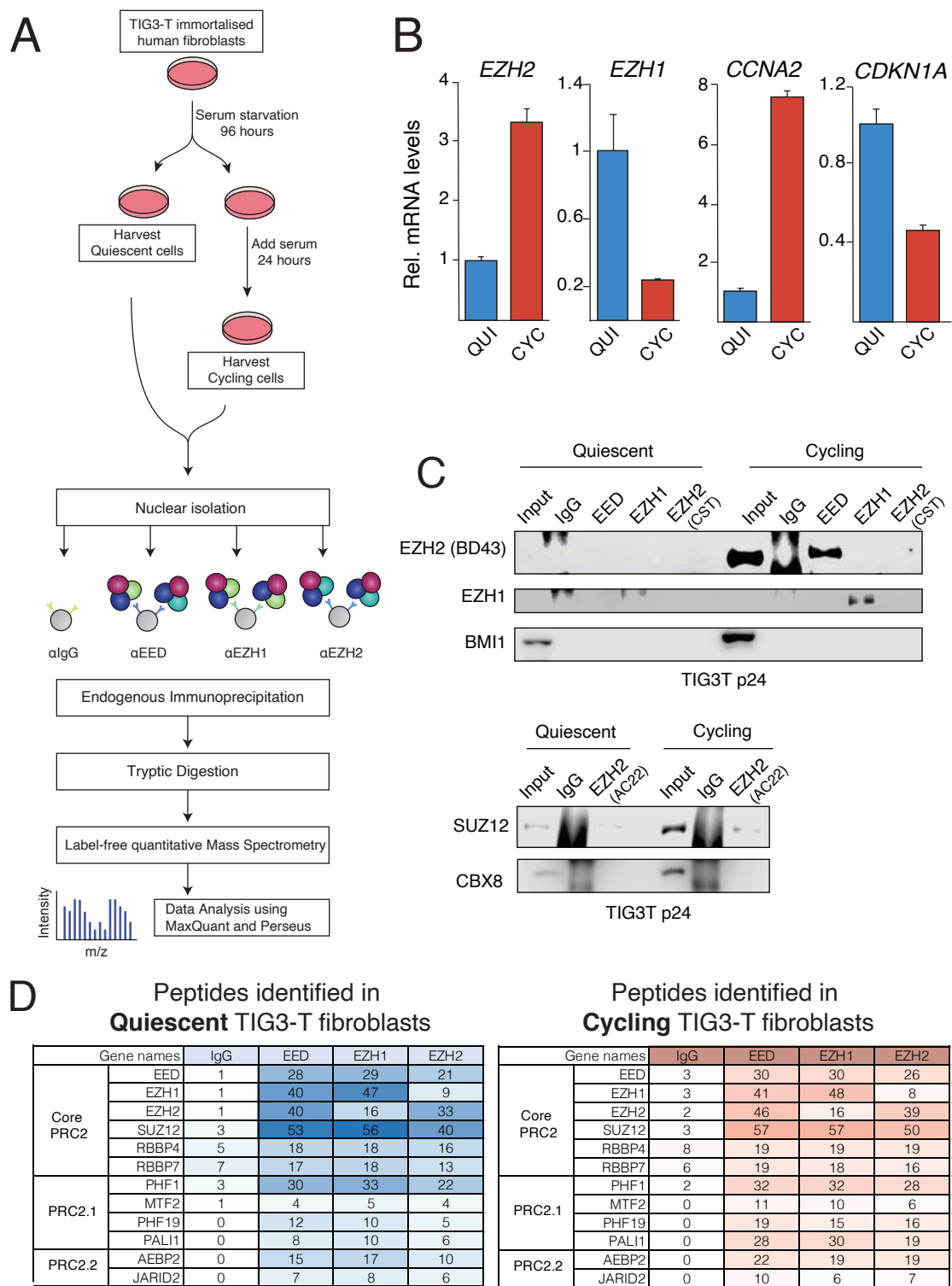


Figure 4.4 Overview of EED, EZH1 and EZH2 IP-MS in quiescent and cycling TIG3-T fibroblasts.

(A) Schematic of EED, EZH1 and EZH2 immunoprecipitation in TIG3-T nuclear lysates followed by in solution tryptic digestion methods.

(B) RT-qPCR analysis of serum-starved and serum-stimulated TIG3-T cells.

(C) Western blot analysis prior to subjecting IP eluates to LC-MS/MS.

(D) Table showing the average peptides identified across three replicates following EED, EZH1 and EZH2 endogenous co-IPs.

4.2.4 Evaluation of EED IP-MS for identification of PRC2 accessory proteins in quiescent and cycling cells

In quiescent and cycling TIG3-T fibroblasts, core PRC2 members (SUZ12, EZH1/2 and EED (bait)) were all immunoprecipitated along with members of PRC2.1 (PHF1/MTF2/PHF19 and PALI1) as well as PRC2.2 (AEBP2 and JARID2), highlighting the many variations of PRC2 that can exist in fibroblasts (Fig 4.5A&B). Interestingly, in the quiescent context, whilst all PRC2 accessory proteins immunoprecipitated with EED, IBAQ stoichiometry analysis highlighted that PHF1 was the most abundant of the Polycomb-like accessory proteins, while AEBP2 was also found to be enriched in quiescent cells. Contrastingly, in cycling cells, in addition to AEBP2 and PHF19, PHF1-containing complexes were observed suggesting that the majority of PRC2 within cycling cells contains these accessory proteins. The majority of PRC2 within quiescent cells was found to contain EZH1 with an IBAQ score of 0.55 relative to the EED bait, whilst EZH2 was the more abundant of the methyltransferases in cycling cells with an IBAQ score of 0.58 (Figure 4.5C). There were no peptides detected for EPOP in any context suggesting it is not expressed in these cells (Figure 4.5C). In line with these observations on a protein level, integration with RNA-seq analysis performed in Chapter 3, confirmed negligible expression levels of EPOP, suggesting that it is either lowly or not expressed in TIG3-T fibroblasts (Figure 3.9C).

Although JARID2 levels were marginally elevated in quiescent over cycling, the overall abundance was insignificant relative to the bait. Similarly, IBAQ scores for MTF2 and PALI1 in both quiescent and cycling cells were all less than 0.0125 (i.e. occur in less than 1.25% of all complexes) suggesting that these components are lowly expressed and do not make up a large portion of PRC2 within TIG3-T cells (Figure 4.5 C). Overall, EZH1 and PHF1 were enriched in quiescent TIG3-T whilst in cycling cells elevated levels of EZH2, PHF19 and AEBP2 were noted (Figure 4.5D).

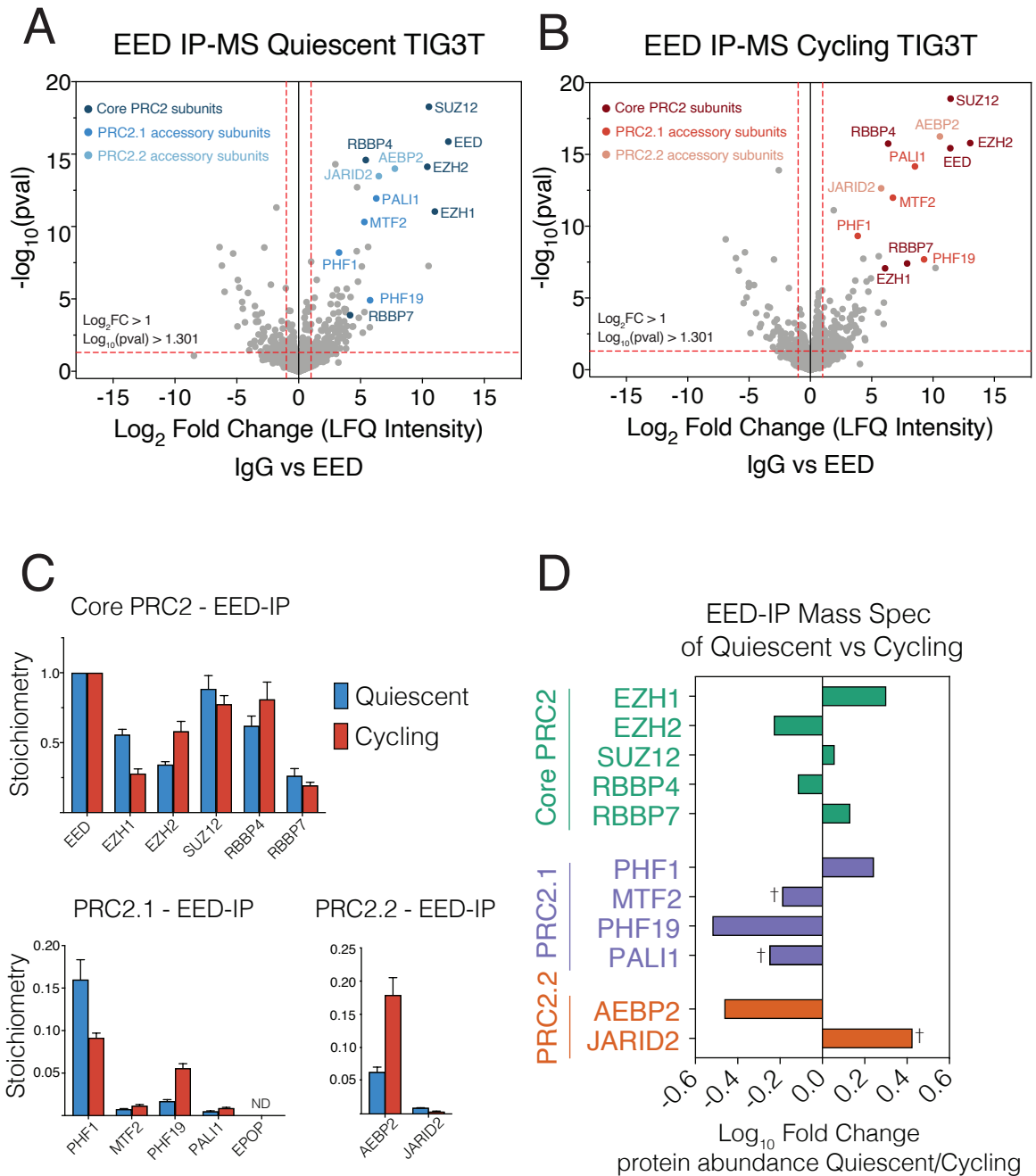


Figure 4.5 Identification of EED-interacting proteins in quiescent and cycling TIG3-T fibroblasts.

(A) Volcano plot representation of EED-interacting proteins in quiescent TIG3-T cells. The label free quantification (LFQ) intensity of the bait (EED) over IgG control is plotted against the $-\log_{10}$ (p value).

(B) Volcano plot representation of EED-interacting proteins in cycling TIG3-T cells. The label free quantification (LFQ) intensity of the bait (EED) over IgG control is plotted against the $-\log_{10}$ (p value).

(C) Intensity-based absolute quantification (IBAQ) scores for PRC2 components following EED IP-MS in quiescent and cycling TIG3-T.

(D) $\log_{10}$ of the fold change of IBAQ scores for PRC2 subunits following EED IP-MS in quiescent and cycling TIG3-T. † denotes IBAQ scores of <0.0125 .

4.2.5 Evaluation of EZH1/2 IP-MS for identification of PRC2 accessory proteins in quiescent and cycling cells

Previous work within the lab and within the RNA-seq analysis in Chapter 3 of this thesis highlighted a distinct expression profile between EZH1 and EZH2 in serum-starved and serum-stimulated human fibroblasts. Therefore, I next performed EZH1 and EZH2 endogenous immunoprecipitation, coupled with mass spectrometry, to compare the proteins associated with these histone methyltransferases in quiescent and cycling fibroblasts.

Initial EZH1 IP-MS analysis confirmed the presence of all known PRC2 accessory proteins (with the exception of EPOP) with EZH1 in the quiescent and cycling context (Figure 4.6A&B). Similar to the results observed in the previous EED IP-MS, PHF1 was the most abundant of the accessory protein in quiescent cells followed by AEBP2 (Figure 4.6C). The remaining accessory proteins were all represented in less than 5% of PRC2 complexes in quiescent cells. Similar to the protein abundance in quiescent cells, AEBP2 and PHF1 were the predominant accessory proteins, however, there is more of an enrichment of PHF19 proteins compared to quiescent cells. Interestingly, negligible levels of EZH2 were observed in the EZH1 IP-MS (IBAQ score of 0.003 and 0.006 in G0 and cycling cells) indicating that these two histone-methyltransferase likely do not co-occur in a dimeric complex (Figure 4.6C). I wanted to evaluate if there were any significant differences in protein abundance between quiescent and cycling cells by comparing the stoichiometry of each PRC2 protein. While there was no significant enrichment for any protein in quiescent cells, PHF19 and AEBP2 were enriched in the cycling context (Figure 4.6D). Analogous to the EED IP results, whilst JARID2 appeared enriched in quiescence (and PALI1/MTF2 enriched in cycling) these observations are not significant due to the overall low abundance of these proteins in both cellular contexts (Figure 4.6C&D).

I next performed an immunoprecipitation of EZH2 in TIG3-T fibroblasts in order to further assess the variance in PRC2 associated proteins in the two cellular contexts

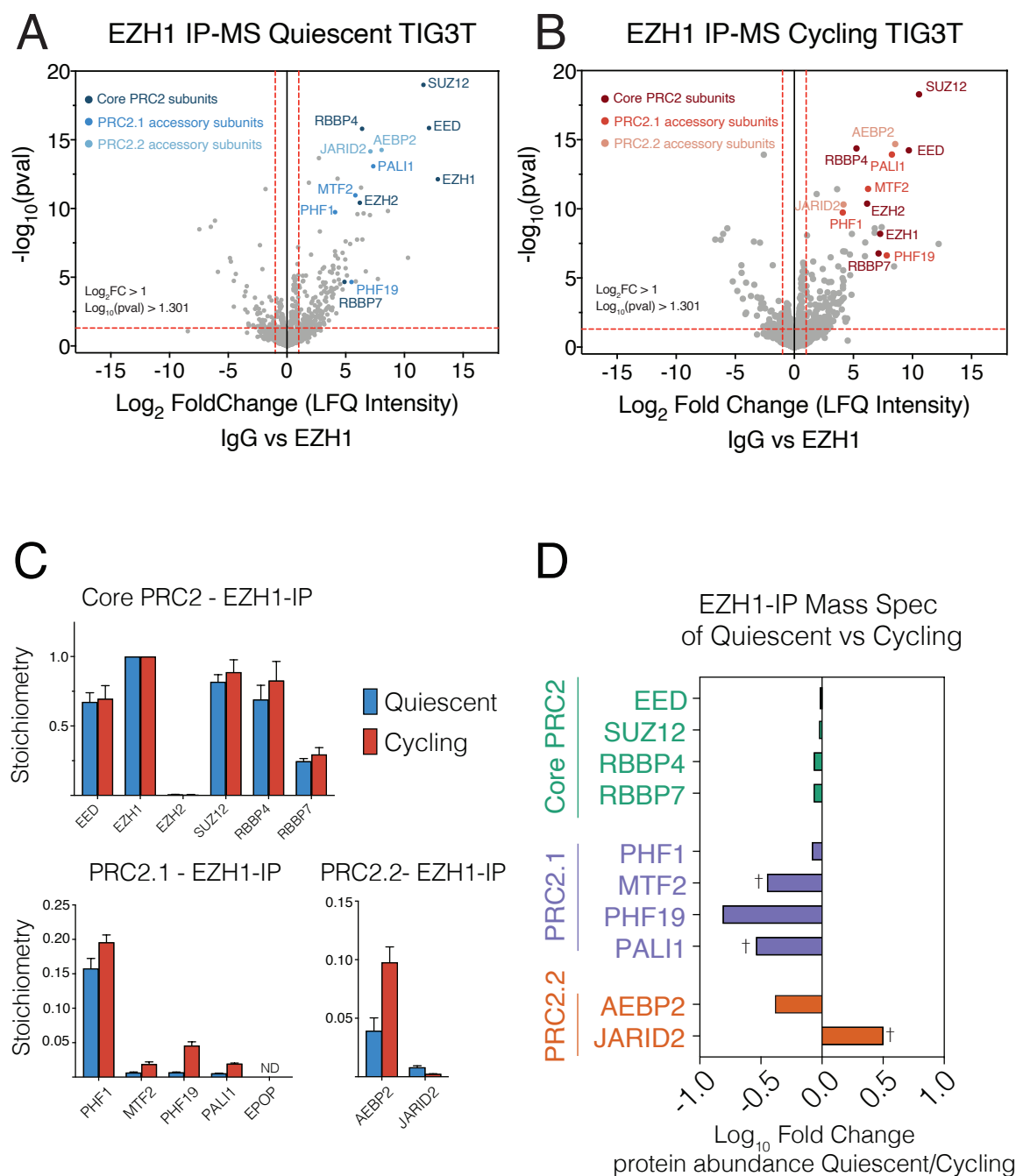


Figure 4.6 Identification of EZH1-interacting proteins in quiescent and cycling TIG3-T fibroblasts.

(A) Volcano plot representation of EZH1-interacting proteins in quiescent TIG3-T cells. The label free quantification (LFQ) intensity of the bait (EZH1) over IgG control is plotted against the $-\log_{10}$ (p value).

(B) Volcano plot representation of EZH1-interacting proteins in cycling TIG3-T cells. The label free quantification (LFQ) intensity of the bait (EZH1) over IgG control is plotted against the $-\log_{10}$ (p value).

(C) Intensity-based absolute quantification (IBAQ) scores for PRC2 components following EZH1 IP-MS in quiescent and cycling TIG3-T.

(D) $\log_{10}$ of the fold change of IBAQ scores for PRC2 subunits following EZH1 IP-MS in quiescent and cycling TIG3-T. † denotes IBAQ scores of <0.02 .

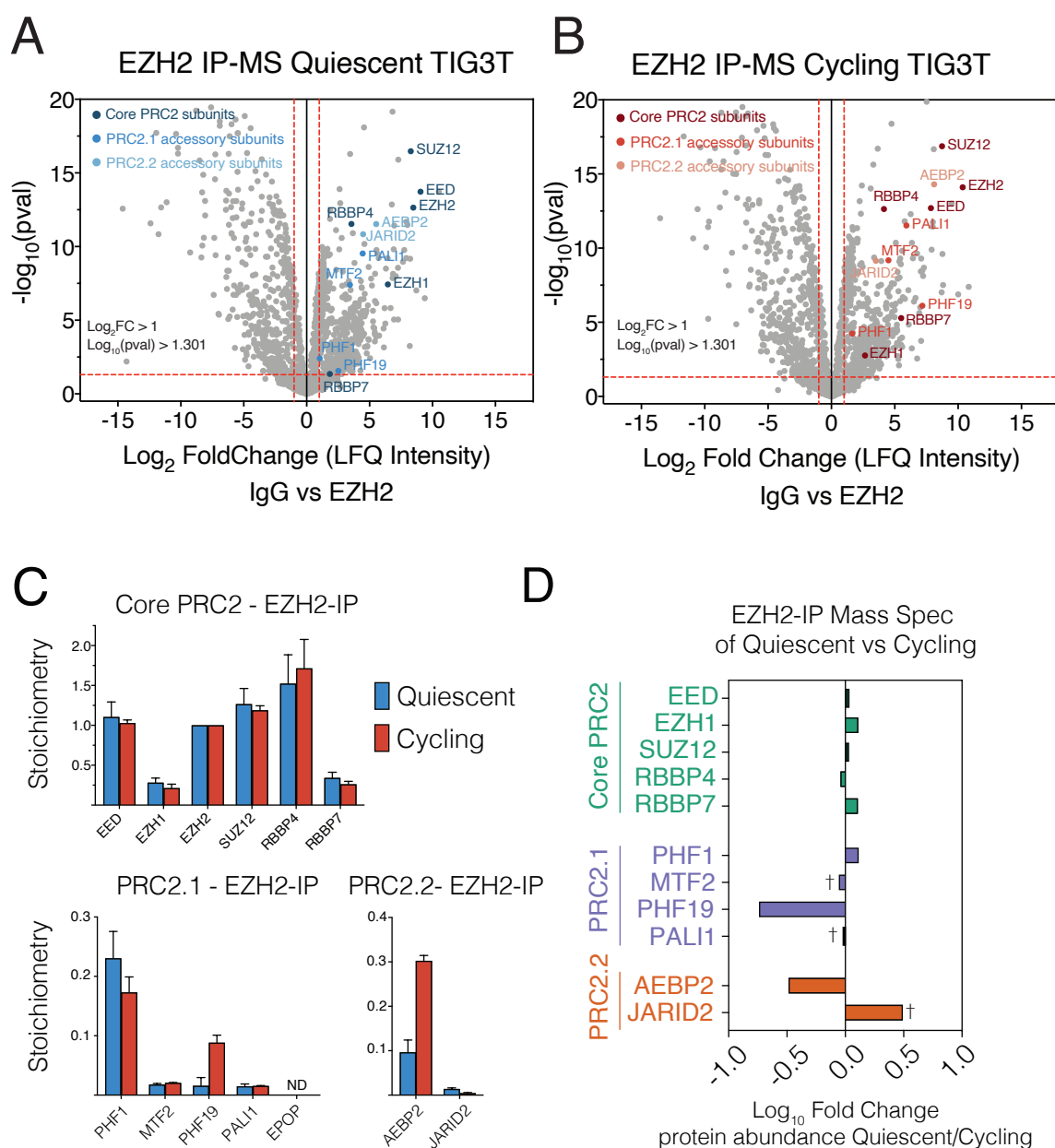


Figure 4.7 Identification of EZH2-interacting proteins in quiescent and cycling TIG3-T fibroblasts.

(A) Volcano plot representation of EZH2-interacting proteins in quiescent TIG3-T cells. The label free quantification (LFQ) intensity of the bait (EZH2) over IgG control is plotted against the $-\log_{10}$ (p value).

(B) Volcano plot representation of EZH2-interacting proteins in cycling TIG3-T cells. The label free quantification (LFQ) intensity of the bait (EZH2) over IgG control is plotted against the $-\log_{10}$ (p value).

(C) Intensity-based absolute quantification (IBAQ) scores for PRC2 components following EZH2 IP-MS in quiescent and cycling TIG3-T.

(D) $\log_{10}$ of the fold change of IBAQ scores for PRC2 subunits following EZH2 IP-MS in quiescent and cycling TIG3-T. † denotes IBAQ scores of <0.02.

and as a comparison to the EZH1 IP. It was evident, based on the label-free quantification and p-values of IgG versus EZH2, that the EZH2 antibody immunoprecipitated more background proteins compared to EZH1 or EED antibodies (Figure 4.7A&B). Regardless of the abundance of background proteins, PRC2 associated proteins were amongst the most enriched and similar to EZH1 and EED immunoprecipitations, EPOP was not detected in the EZH2 Ips (Figure 4.7C). This confirms that EPOP is either lowly or not expressed in these cells. Mirroring the previous results from EZH1 and EED immunoprecipitations, PHF1 and AEBP2 were the most abundant of the accessory proteins in quiescent cells whereas PHF19, PHF1 and AEBP2 protein levels were enriched in cycling cells (Figure 4.7C). The fold change of PRC2 protein stoichiometry for quiescent versus cycling cells highlighted PHF1 and EZH1 as marginally enriched in quiescent over cycling cells while AEBP2 and PHF19 are significantly higher in cycling cells (Figure 4.7D). As discussed previously, the enrichment of JARID2 in quiescent cells and MTF2/PAL11 in cycling cells is likely a consequence of the negligible protein levels in both cell cycle contexts.

4.2.6 EZH1/PHF1-PRC2 is likely the most abundant PRC2 within quiescent TIG3-T

Next, I sought to correlate the results from RNA-sequencing of quiescent and cycling TIG3-T in Chapter 3 with the IP-MS results already established in this chapter. I chose to use the IBAQ scores relative to EED as EED is consistent in both quiescent and cycling cells whilst EZH1 and EZH2 expression fluctuates depending on cell cycle. Taken together, there was a broad correlation in RNA-sequencing and the EED IP-MS data with EZH2 and PHF19 expression elevated on an RNA and protein level in cycling cells while the inverse was observed for EZH1 and PHF1 in quiescent cells (Figure 4.8A). These results suggest that the predominant complex found within quiescent cells most likely consists of PHF1, EZH1, EED, SUZ12 and RBBP4 while EZH2, EED, SUZ12, RBBP4 and either PHF19 or AEBP2 likely make up the majority of PRC2 found in cycling cells (Figure 4.8B).

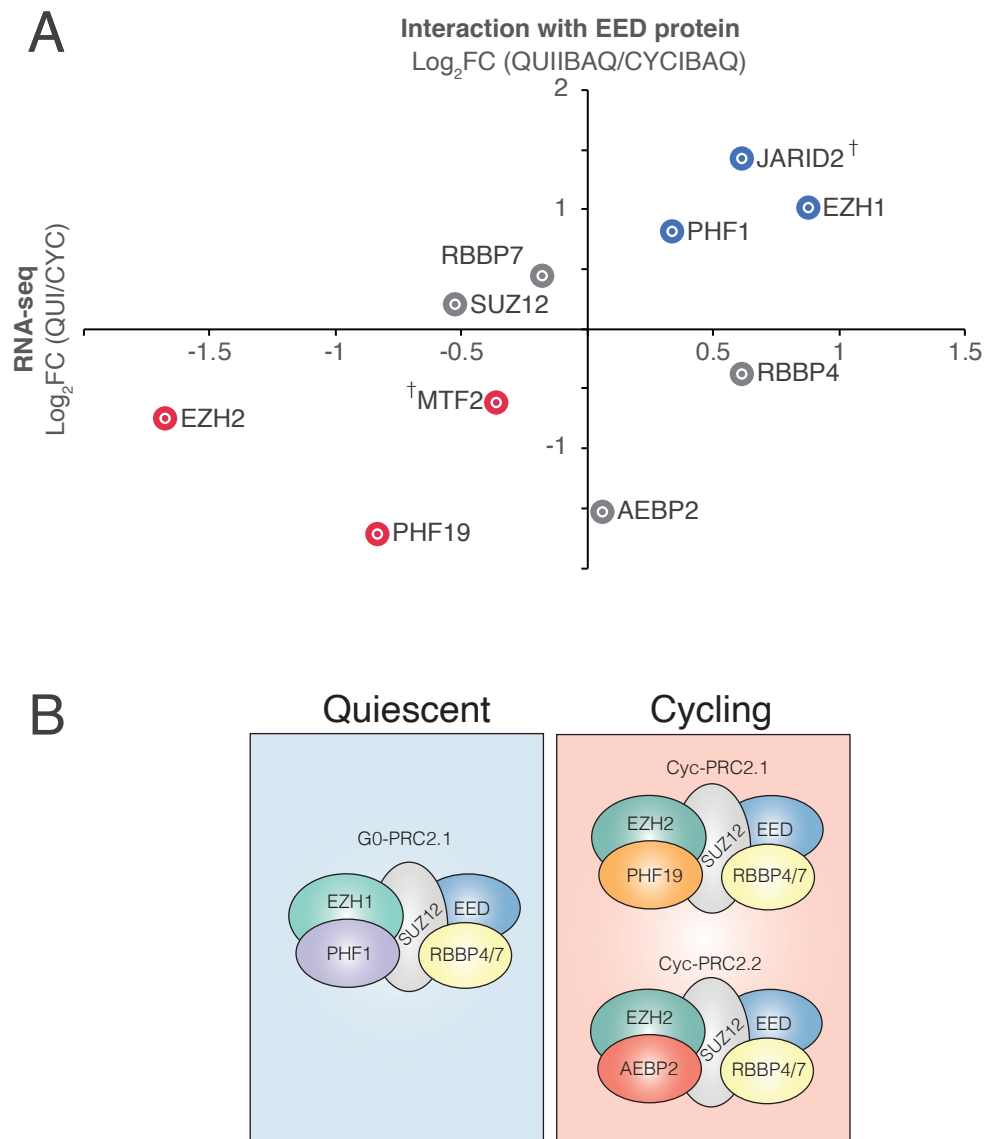


Figure 4.8 PRC2 interactions in quiescent and cycling cells are broadly correlated between RNA and protein expression.

(A) Log₂ IBAQ scores for proteins interacting with EED in quiescent and cycling fibroblasts were plotted against the Log₂ normalised counts in quiescent and cycling cells. † denotes IBAQ scores of <0.0125 (Relative to EED bait) in EED IP-MS.

(B) Schematic of the predicted PRC2 assemblies in quiescent and cycling cells.

4.2.7 Degradation of EED destabilises G0-PRC2 in human fibroblasts

Subsequently, I endeavoured to assess the effect of EED degraders on PRC2 stability and function in quiescent and cycling TIG3-T immortalised human fibroblasts. EED degraders were kindly provided for the experiments within this thesis as part of a collaborative project with the lab of Dr. Lindsey I. James (Potjewyd *et al*, 2020). UNC6852 (EED degrader/dEED) is composed of two distinct moieties, an EED ligand comprised of an analogue of EED226 which is a known interactor of EED and a VHL ligand consisting of a VH032-amine. The negative control compound (UNC7043) is similar in structure to UNC6852 with the exception of an opposite enantiomer at the hydroxyproline moiety on the VHL ligand (Figure 4.9A) (Potjewyd *et al.*, 2020). These components are connected with a linker which is sufficient to degrade PRC2 by tethering the complex to a ubiquitin ligase complex which subsequently tags PRC2 for proteasomal degradation (Figure 4.9B). TIG3-T cells were induced to quiesce by removal of serum from the cells for 48 hours. Following serum removal for 48 hours, cells were treated with DMSO, the inactive control drug (UNC7043) and the EED degrader (UNC6852) for 72 hours. Quiescent cells were maintained in a serum-free media for the duration of the drug treatment while asynchronously (AS) growing cells were treated with the same compounds in a serum-containing media (Figure 4.9C).

Western blot analysis of core PRC2 proteins (EED, EZH1, EZH2 and SUZ12) and CCNA2 confirmed sufficient serum starvation of TIG3-T cells as evidenced by the elevated levels of CCNA2 and EZH2 in AS cells and treatment with dEED resulted in degradation of core PRC2 proteins in both quiescent and cycling TIG3-T cells (Figure 4.9D). H3K27 di- and tri-methylation were largely reduced in cycling cells where EED was degraded, they were not reduced after degradation of EZH1-PRC2 in quiescent cells (Figure 4.9D). Interestingly, H3K27me1 remained largely unchanged in both quiescent and cycling cells treated with EED degrader. Based on quantification of Western blots using Licor imaging software, H3K27 tri- and di-methylation levels were reduced by 70-90%, respectively, in cycling cells yet remain unchanged in quiescent cells (Figure 4.9E).

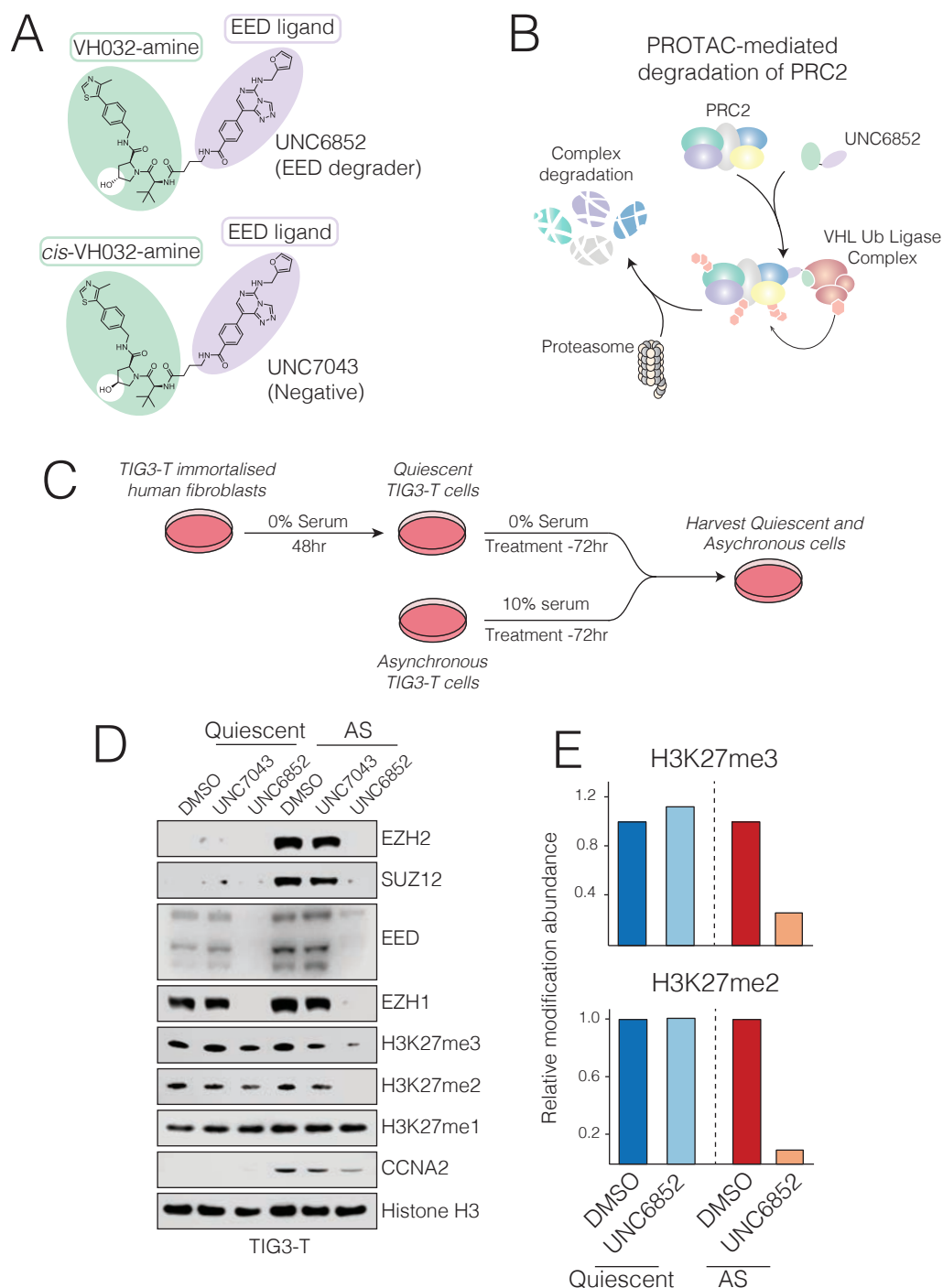


Figure 4.9 Degradation of PRC2 destabilises the complex but does not reduce H3K27 methylations in quiescent cells.

(A) Structure of EED degrader (UNC6852) and the negative control compound (UNC7043).

(B) Mechanism of PROTAC-mediated degradation of PRC2.

(C) Schematic of dEED treatment of quiescent and asynchronous TIG3-T cells.

(D) Western blot analysis of quiescent and asynchronous TIG3-T cells treated with DMSO, UNC7043 and UNC6852.

(E) Quantification of H3K27me2/3 band intensity from western blots in panel (D) using the Odessey LICOR.

4.2.8 Quiescent human fibroblasts can re-enter the cell cycle upon inhibition of PRC2

To establish the long-term effects of PRC2 degradation in TIG3-T cells, I next performed a growth assay following treatment with an EED degrader. EED inhibitors were included as a control as previous, unpublished data (Healy et al, 2021 in preparation) had shown that the proliferation rate of quiescent human primary cells, which had been stimulated to re-enter the cell cycle by addition of serum, was not affected by PRC2 inhibition. Therefore, I decided to examine if degradation of PRC2 affected cell growth and proliferation and also to compare the effects of the EED inhibitor and degrader on PRC2 stability and on histone modifications. Serum was removed for 48 hours prior to treatment of the cells with DMSO, EED inhibitor (A395/EEDi) and EED degrader (UNC6852) for 72 hours. Quiescent cells were counted, and 1 million cells were re-seeded in serum-containing media for each condition. Asynchronous cells were treated and re-seeded in parallel (Figure 4.10A). The EED inhibitor and EED degrader were sufficient to reduce H3K27me3 and H3K27me2 in cycling cells, whereas only the EED degrader destabilised the complex. PRC2 was degraded in quiescent cells following treatment with EED degrader (UNC6852) but not EED inhibitor (A395) while H3K27 di- and trimethylation marks remained unchanged in after treatment of quiescent cells with either EEDi or dEED in quiescent cells (Figure 4.10B).

I hypothesised that the effect of PRC2 inhibitors and degrader drugs on quiescent cells may not be observed within the short 3-day drug treatment therefore I stimulated the treated cells with serum to re-enter the cell cycle and cultured the cells for a further 18 days to observe the potential growth defects of PRC2 inhibition/degradation. Interestingly, quiescent fibroblasts treated with either PRC2 inhibitor or degrader re-entered the cell cycle following serum addition and continued to proliferate for the duration of the assay (Figure 4.10C). In contrast, proliferation of asynchronous TIG3-T cells treated with EED inhibitors was slower than the DMSO control although this growth defect was not fully observed until around Day 9 (Figure 4.10D). Surprisingly, degradation of PRC2 did not perturb the growth in neither quiescent nor asynchronous TIG3-T (Figure 4.10C&D). This was

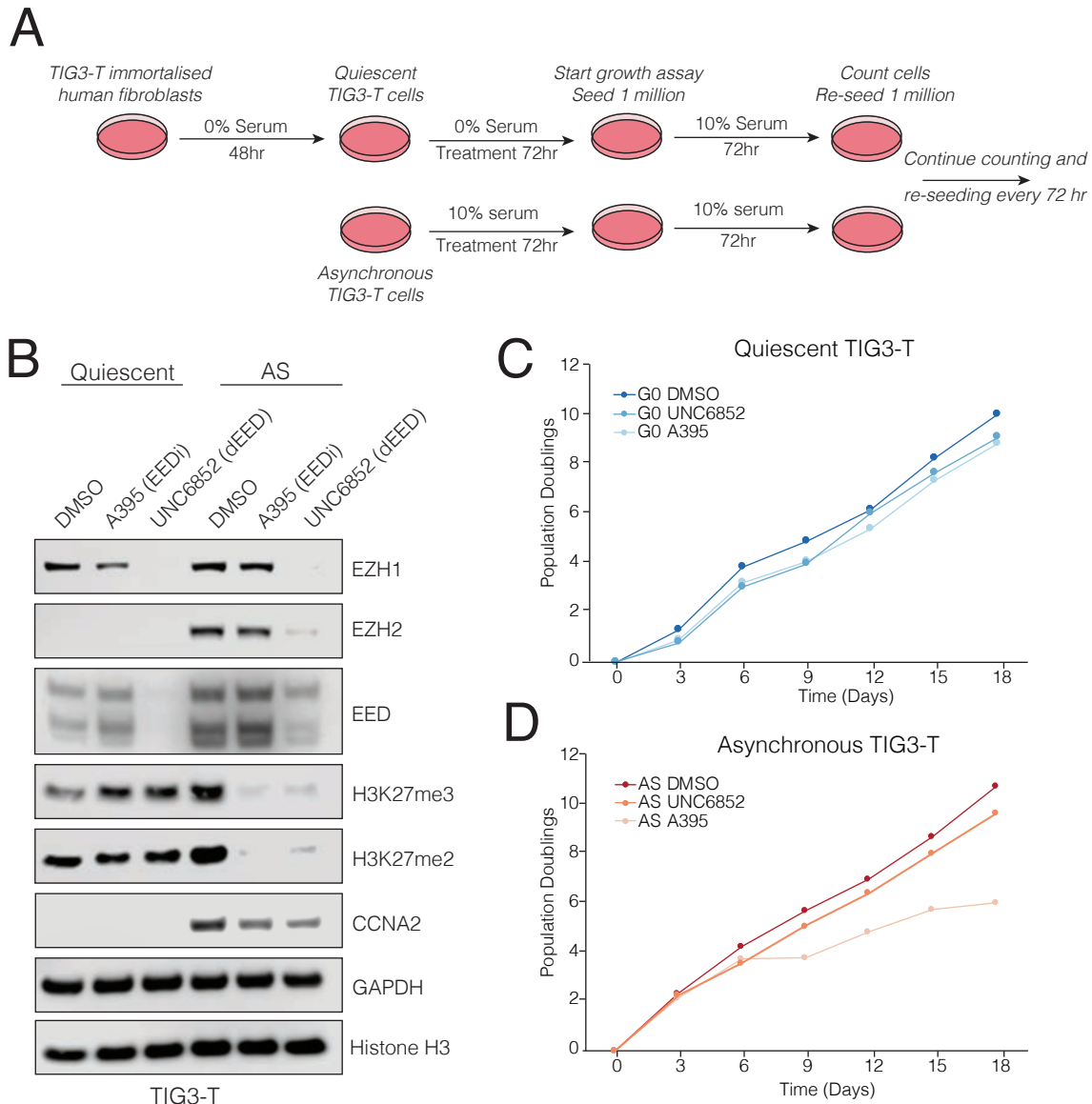


Figure 4.10 Inhibition of PRC2 in quiescent cells does not affect their ability to re-enter the cell cycle.

(A) Schematic of experimental design for the treatment of quiescent and cycling TIG3-T fibroblasts with EED inhibitor (A395) and degrader (UNC6852).

(B) Western blot analysis of quiescent and cycling TIG3-T fibroblasts treated with EED inhibitor and degrader.

(C) Growth assay on quiescent and cycling TIG3-T fibroblasts treated with EED inhibitor and degrader. Assays were initiated following 72 hour treatment and continued for 18 days.

surprising, given that previous data had shown that primary human cells were sensitive to EED and EZH2 inhibition and fail to efficiently re-enter the cell cycle and proliferate.

4.2.9 Serum starvation of certain cancer cell lines is sufficient to induce quiescence

The results so far have revealed that quiescent human fibroblasts are insensitive to degradation (or inhibition) of PRC2 by small molecule drugs. It has been hypothesized that within tumour microenvironments, cancer stem cells remain in a dormant state until activation by an outside factor. To assess whether quiescent cancer cells are rendered insensitive to PRC2 degradation/inhibition, I performed serum-starvation of multiple cancer cell lines and RT-qPCR for known quiescent and cycling markers to check whether cells had entered quiescence. Similar to the serum starvation of immortalised TIG3-T fibroblasts, BCPAP, a thyroid papillary carcinoma cell line, was cultured for 5 days in the absence of serum (Quiescent). Cells re-entered the cell cycle following addition of serum for 24 hours (Cycling). Serum was not removed from asynchronous (AS) cells. EZH1, PHF1 and CDKN1A mRNA expression levels were all elevated after removal of serum in BCPAP and in contrast, PHF19, EZH2 and CCNA2 expression increased upon addition of serum which would suggest re-entry to the cell cycle (Figure 4.11A). Publicly available RNA-sequencing data of cancer cell lines reported high expression of PHF1 and EZH1 within BCPAP cells while expression of the remaining PRC2 accessory proteins was relatively lower (<https://www.ebi.ac.uk/gxa/experiments/E-MTAB-2770>). This expression profile was similar to the RNA-sequencing results observed in Chapter 3 for TIG3-T cells. RNA-seq data for Pfeiffer was included as it has been previously established that EZH2 is expressed in these lymphoma cell lines (McCabe *et al.*, 2012) (Figure 4.11B)

Supporting the hypothesis that BCPAP cells have successfully undergone quiescence, light microscopy images of cancer cells illustrated that removal of serum impeded growth of the cancer cell lines (Quiescent Day 2 and Day 5), while

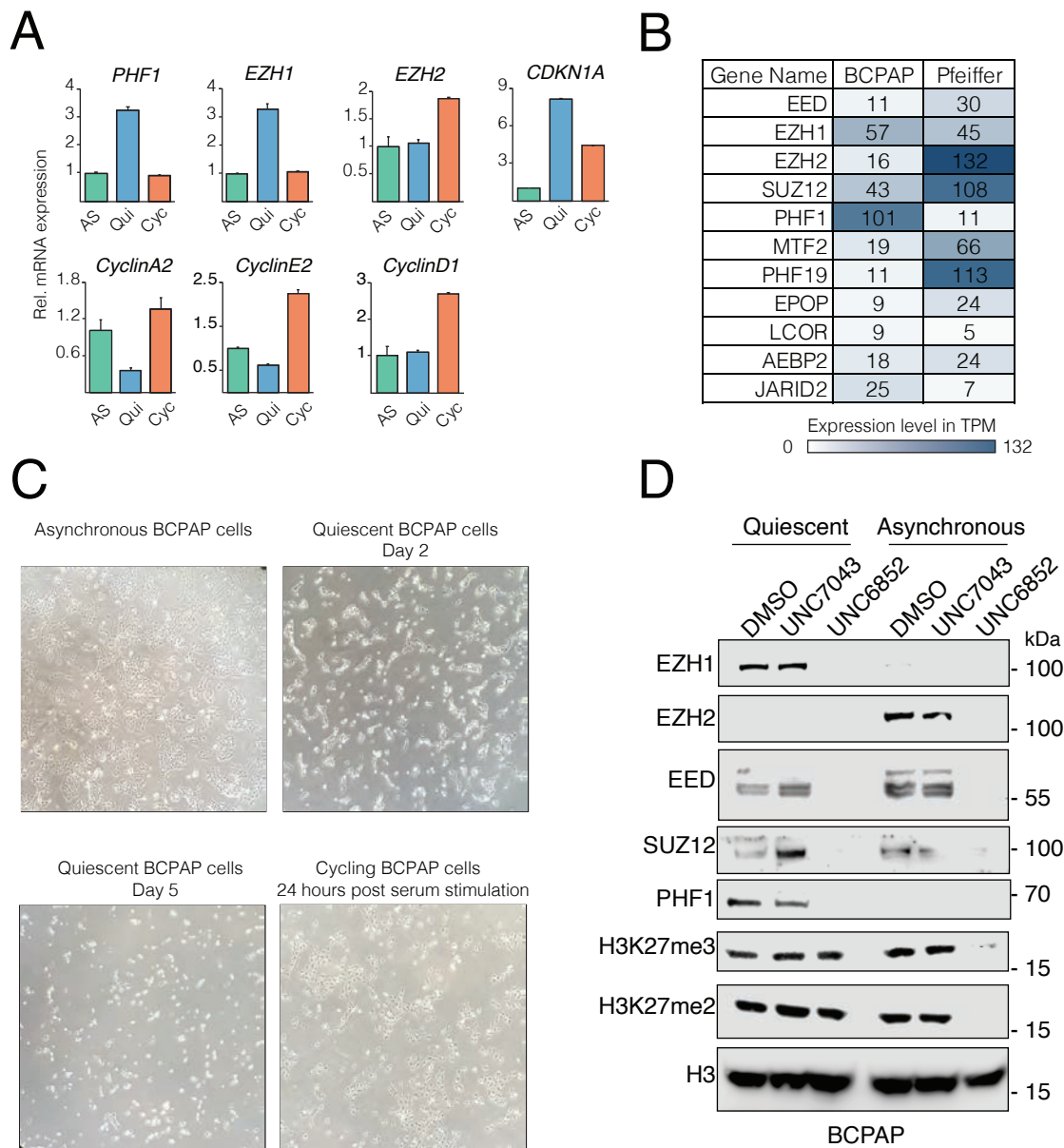


Figure 4.11 Serum starvation of BCPAP is sufficient to induce quiescence.

(A) RT-qPCR analysis of asynchronous, serum-starved (quiescent) and serum-stimulated (Cycling) BCPAP cells. Cells were serum starved for 5 days in total and subsequently released into the cell cycle for 24 hours.

(B) RNA-sequencing data taken from the Cancer Cell Line Encyclopedia (CCLE) include PRC2-associated genes specifically expressed above the expression level cutoff: 0.5 TPM in experiment.

(C) Representative images of BCPAP cells prior to removal of serum (asynchronous), Day 2, 5 after serum starvation (Quiescent) and after addition of serum to stimulate cells to re-enter the cell cycle (Cycling).

(D) Western blot analysis of BCPAP treated with EED degrader for 72 hours. Quiescent BCPAP were serum-starved for 48 hour prior to treatment with drugs.

cells stimulated with serum efficiently re-enter the cell cycle and proliferate (Figure 4.11C). Taken together, I decided that BCPAP would be a valuable model for studying quiescence in a cancer cell context.

4.2.10 Degradation of PRC2 in thyroid cancer cell lines destabilise the complex but does not reduce H3K27me3 levels

Subsequently, I explored whether degradation of PRC2 in a quiescent cancer model would have a similar cellular response as observed previously in TIG3-T fibroblasts. Cells were treated as described in Figure 4.9C. Briefly, BCPAP cells were cultured in a serum-free media for 48 hours prior to treatment with dEED for 72 hours while maintained in a serum-free media. Asynchronous BCPAP cells were cultured in media containing 10% serum. Serum starvation of BCPAP resulted in increased EZH1 expression while EZH2 was elevated in asynchronously growing BCPAP. Core PRC2 was largely destabilised and H3K27me_{2/3} was reduced following treatment of asynchronous BCPAP with EED degrader. Core PRC2 was degraded in quiescent cells after treatment with EED degrader, however, H3K27 di- and tri-methylation marks were unchanged. These preliminary results would suggest a cautionary note in the treatment of cancers with PRC2 degraders as quiescent cancer cells may be refractory to degradation of PRC2. Combinatorial drug therapies could be considered as a potential therapy for the treatment of quiescent cancer cells.

4.3 Discussion

The primary aim of this chapter was to characterise the composition of PRC2 within quiescent and cycling TIG3-T cells. Using endogenous immunoprecipitation of EED, EZH1 and EZH2 coupled with mass spectrometry, I successfully identified a form of PRC2, termed G0-PRC2.1 and consisting of PHF1, EZH1, EED, SUZ12 and RBBP4, that exists primarily in serum-starved, quiescent TIG3-T. I predict that G0-PRC2.1 would have considerably less methyltransferase activity than Cycling-PRC2 based previous observations in the literature. It has been well established that EZH1 is less catalytically active than EZH2 (Grau *et al.*, 2021; Margueron *et al.*, 2009; Son *et al.*, 2013) and recent studies have highlighted that PHF1-PRC2 has slightly reduced methyltransferase activity on nucleosome substrates over PHF19-PRC2 (Zhang *et al.*, 2019). The initial attempts at purifying G0-PRC2 complexes demonstrated that the inclusion of SUZ12 in PRC2 greatly improved the stability and purification of PRC2 components (Figure 4.1C, middle lane) and further characterisation of the biochemical properties of this complex is required. It has been reported that EZH1-containing PRC2 complexes contribute to the chromatin compaction through a dimerization of two complexes via the localisation lobe (i.e. the N-terminal region of SUZ12) (Grau *et al.*, 2021). Quiescent cells generally have a more condensed chromatin make-up than proliferating cells which subsequently correlates with the overall reduced transcription levels in quiescent yeast models (McKnight *et al.*, 2015; Swygert *et al.*, 2019). While we do yet not fully understand the function or role of G0-PRC2.1 within quiescent mammalian cells, I speculate that it has minimal catalytic activity and likely instead facilitates the compaction of chromatin. The primary aim of the IP-MS experiments was to identify the accessory proteins associated with PRC2 during quiescence and the identification of these accessory proteins could provide insight on the function of the complex in non-dividing, quiescent cells. In addition to this, I have also identified potentially novel interactors of PRC2 in quiescent cells. For example, HDAC2, a histone deacetylase, was observed as an interactor in the EZH1/EZH2 IPs and was found to be marginally elevated in quiescent compared to cycling cells. It has been reported that Rpd3, the yeast equivalent of HDAC2, is important for driving transcriptional quiescence in yeast (McKnight *et al.*, 2015). While the validity of these interactions has not been explored in this thesis, the implication of these quiescent-PRC2 interactors could be an interesting in future experiments.

The second aim of this chapter was to evaluate the sensitivity of quiescent and proliferating cells to PRC2 inhibition and degradation. Unpublished data from the Bracken lab has highlighted the lack of sensitivity of quiescent primary cells to catalytic and allosteric PRC2 inhibitors (Healy et al, 2021 in preparation) and the results within this chapter have supported these observations in an immortalised cell line. Treatment of TIG3-T with EED degrader (UNC6852) did not affect the proliferation status of quiescent or asynchronously growing cells. The variance in the affect these drugs have on cells may be due to the differences in potency of the two drugs with the EED degrader molecule reporting a half-maximal inhibitory concentration (IC_{50}) of 247 ± 2.9 nM and the capacity to degrade approximately 80% of EED in cells (Potjewyd *et al.*, 2020) and EED inhibitor (A395) has an IC_{50} of 18 ± 2 nM. While the IC_{50} of two individual drugs cannot be directly compared, it is informative regarding the drug's efficacy. Treatment of asynchronous TIG3-T with EED degrader was sufficient to degrade the complex and reduce H3K27me2/3 levels, however, the levels of H3K27me1 remained unchanged (Figure 4.9D). Given that PRC2 catalyses H3K27me1 more readily over dimethylation and subsequently trimethylation, it could be hypothesized that residual PRC2 may remain in the cells and may be capable of catalysing the mark. A longer treatment or increased dose of the EED degrader drug may be sufficient to remove H3K27me1 completely. While it is surprising that H3K27me1/2/3 did not decrease in quiescent cells treated with EED degraders, it could be speculated that these methylation marks are not altered due to the cells being in a dormant, non-dividing state and thus the marks are not diluted out following cell division.

Within the tumour microenvironment are cancer stems cells which are thought to exist in a non-dividing quiescent state and their resistance to chemotherapeutic drugs is believed to drive the re-population and recurrence of tumours (Batlle & Clevers, 2017). Due to the slow-growing and low numbers of quiescent cells *in vivo*, it is challenging to study these dormant cells. The BCPAP “quiescent cancer model” established in this chapter has highlighted that quiescent cancer cells are potentially resistant to PRC2 targeting small molecules and further strategies are required to target these cells. Quiescent TIG3-T were unaffected by PRC2 inhibition and proliferate as normal following re-entry to cell cycle so therefore a combination drug

therapy may be a better approach for sensitising these non-dividing cells to chemotherapy drugs. An example of such a strategy may require the use of treatments to stimulate cells out of quiescence and back into the cell cycle alongside PRC2 inhibitors. A combinatorial approach would hopefully result in increased efficacy in targeting quiescent cancer cells.

Chapter 5: Analysing the dependency of a panel of cancer cell lines on PRC2 subunits

5.1 Introduction

Genomic analyses of human cancer cell lines have revealed a multitude of mutations in genes which can promote cell growth and remarkably, can confer drug resistance (Baker *et al.*, 2015; Gibaja *et al.*, 2016). Dependencies in cancer cells have been identified with a number of different methods including computational approaches based on available datasets, chemical approaches with small drug molecules to target particular proteins of interest and genome-wide RNAi and CRISPR screening approaches which aim to discover new therapeutic targets. Computational analysis of previously published datasets can be used to predict vulnerabilities and as more data becomes available, computational predictions are a valuable tool in the early stages of experimental design when selecting cancer cell lines, however, all predicted dependencies require further validation in cells (Meyers *et al.*, 2017; Tsherniak *et al.*, 2017). Chemical screening approaches have been used to identify dependencies in various cancer cell lines (Murashima *et al.*, 2019; Torrance *et al.*, 2001; Turcotte *et al.*, 2008) and the main advantage of this approach is that the results can be more easily translated to a clinical setting. However, the limitations of drug-based screens is the potential for off-target hits with certain compounds and the requirement of a druggable target. RNAi have been implemented in several studies to identify novel vulnerabilities (Cowley *et al.*, 2014; McDonald *et al.*, 2017) and can be combined with small molecule screening to identify synergistic gene and drug combinations (Williams *et al.*, 2017). As RNAi functions at a post-transcriptional level, it can be difficult to achieve complete loss of the target of interest and often, only partial knockdown is achieved. Conversely though, this partial loss using RNAi may better represent the effects of using small molecules in cell to inhibit or disrupt the target of interest (Boettcher & McManus, 2015) (Table 5.1).

Initial genome-wide screens generally employed RNAi to disrupt gene expression and function, however, recent advances in genome-editing have shifted large scale screening towards CRISPR-based technologies (Shalem *et al.*, 2014; Wang *et al.*, 2014). The origins of CRISPR were first described in 1987 when a DNA sequence of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) was accidentally discovered in *E. coli* (Ishino *et al.*, 1987). However, it wasn't until nearly

20 years later that the true potential of gene-editing using CRISPR-Cas9 technologies became apparent (Cong *et al*, 2013; Jinek *et al*, 2012). CRISPR-Cas9 has since been widely adapted by scientists across the globe in many different fields for genome editing of various mammalian cell lines as well as plant and yeast models. CRISPR-Cas9 screening involves the stable expression of Cas9 in the cell line of interest coupled with a library of single guide RNAs (sgRNAs) targeting either a certain gene or a selection of genes. This approach results in the introduction of double-stranded breaks at the targeted site which are then repaired through nonhomologous end-joining. This repair pathway is thought to be error-prone resulting in the introduction of insertion-deletions (INDELs) of unpredictable length at the sgRNA-targeted sites which can often lead to frameshift mutations and gene knockout (Cong *et al*, 2013; Jinek *et al*, 2012; Shalem *et al*, 2014; Wang *et al*, 2014). These CRISPR-based screens not only identify potential new weaknesses in cancer cells, but the precise targeting mechanism of CRISPR-Cas9 can be exploited to pinpoint exact regions/domains within proteins which are critical for function (Brien *et al*, 2018; Doench, 2018; Michel *et al*, 2018; Shi *et al*, 2015).

The primary aim of this chapter is to perform a PRC2-focused, high-density, CRISPR-Cas9 screen in cancer cell lines which were selected based on previous screening data suggesting a dependency on the Polycomb-like proteins (PHF1, MTF2 and PHF19). Using a PRC2-centric sgRNA library of approximately 4000 sgRNA to tile across the exonic sequences of 12 PRC2 subunits, I intend to identify regions or domains of functional importance within subunit of PRC2. The secondary aim of this chapter is to validate the dropout of amino acids 19-27 and 125-133 of PHF1 which were identified from the CRISPR screen by performing a rescue experimental strategy in order to rule out potential off-target, false-positive hits.

Table 5.1 Types of screening methods to identify novel dependencies

Screen type	Advantages	Disadvantages
Computational prediction	Easy to compile data from multiple dataset (i.e. NGS results, previous screens) Quicker than setting up physical screens in cells	Predicted dependencies require experimental validation
Drug/small molecule	Easy to screen thousands of compounds Not limited by technical issues with cell lines (i.e. poor Cas9 cutting efficiency) Direct clinical relevance	Off-target effects of drug Target must be druggable Toxicity in cells could mask results
RNAi (RNA interference)	No manipulation of target cell lines required (i.e. Cas9 expression) Can be combined with drug screening RNAi may better reflect the effects of drug inhibition over complete KO	Off target effects of RNAi RNAi not as effective as complete KO of gene Functions at a post-transcriptional level
CRISPR-Cas9	Can screen all genes as well as non-coding regions of genome Useful to see effects of complete loss of gene Can be combined with drug screening	Off target effects of sgRNA Can be susceptible to false positives/negatives KO of gene may not have same outcome as drug inhibition Technical issues (i.e. Large scale of cells may be required to maintain sgRNA coverage and ensure low MOI. Poor Cas9 cutting efficiency)

5.2 Results

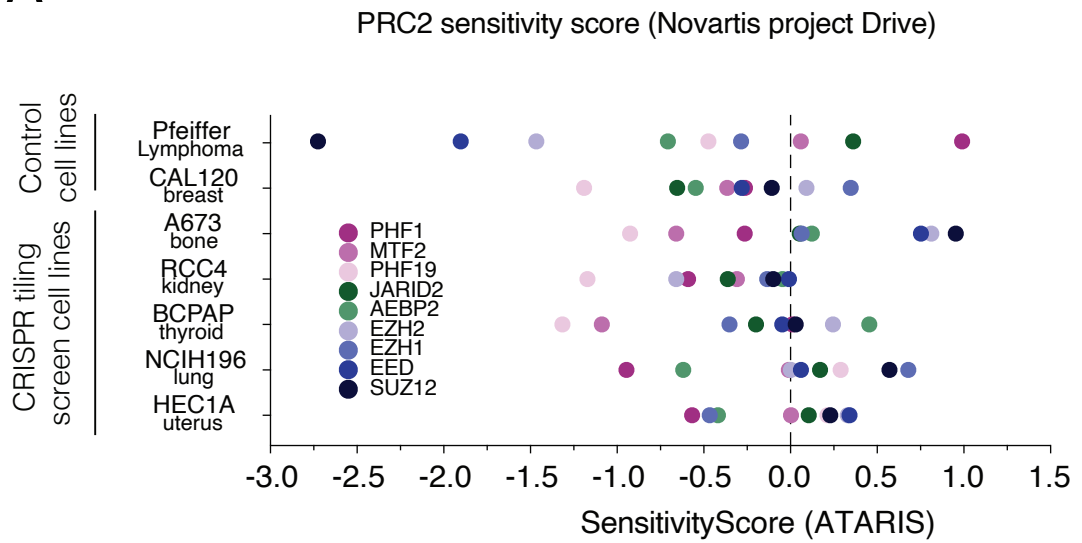
5.2.1 Selection of PHF1/MTF2/PHF19 -dependent cancer cell lines

Project DRIVE is a genome-wide cancer screen examining ~7400 genes in ~400 cancer cell lines. The screen uses on average 20 shRNAs per gene and ranks the dependency of a particular gene in cancer cell lines (McDonald *et al.*, 2017). Cell lines which were sensitive for depletion of PHF1, MTF2 and PHF19 were selected for PRC2 library CRISPR screen based on the sensitivity score in Project DRIVE. Cell lines were then chosen from multiple different cancer types to evaluate the effect of PHF1, MTF2 and PHF19 depletion in a broader context as opposed to focusing on a specific cancer type. The cell lines RCC4_kidney, BCPAP_thyroid, HEC1A_endometrium, NCIH196_lung, A673_bone were selected based on their sensitivity to depletion of PHF1, MTF2 and PHF19. CRISPR screens in CAL120 and Pfeiffer cell lines were performed by Marlena Mucha, Dr. Eric Conway and Dr. James Nolan and have been included to serve as negative controls due to their lack of dependency on PHF1/MTF2/PHF19 (Figure 5.1A). I next decided to check publicly available RNA-sequencing data (<https://www.ebi.ac.uk/gxa/experiments/E-MTAB-2770/Results>) from the Cancer Cell Line Encyclopedia (CCLE) of 934 cancer cells lines to confirm the expression of PHF1/MTF2/PHF19 in the cell lines selected for the CRISPR tiling screen. PHF1/MTF2/PHF19 were all sufficiently expressed in the chosen cell lines based on transcripts per million (TPM) while PRC2.2 members (AEBP2 and JARID2) were slightly less (Figure 5.1B).

5.2.2 Generation of Cas9-expressing cancer cell lines and testing Cas9 cutting efficiency

I next wanted to assess the effect of depletion of PHF1/MTF2/PHF19 in cancer cell lines to identify essential and/or novel domains within these proteins using CRISPR-Cas9 genome editing technology. A PRC2 sgRNA library containing nearly 4000 guides against 12 PRC2 coding sequences (EED, EZH1/2, SUZ12, PALI1/2, PHF1, MTF2, PHF19, EPOP, JARID2 and AEBP2) was used to generate a detailed map

A



B

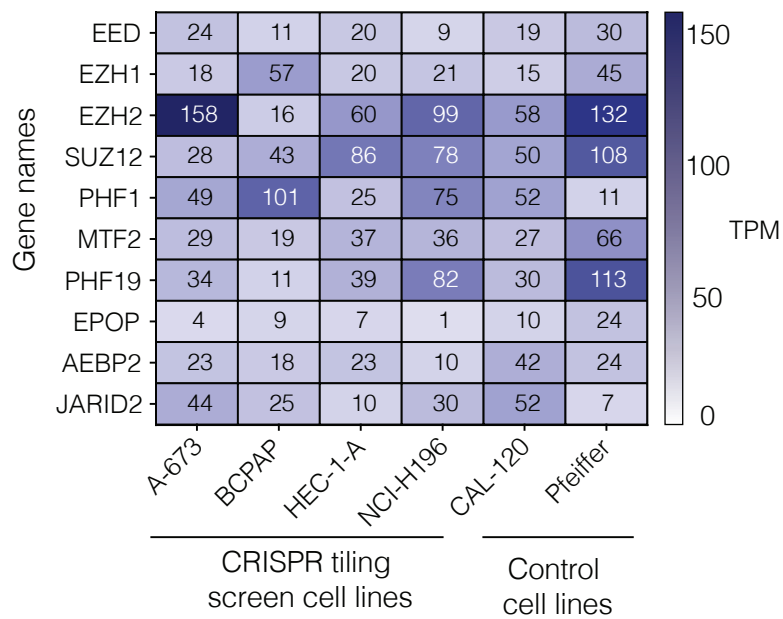


Figure 5.1 Selection of cancer cell lines which are sensitive to PRC2 depletion.

(A) Summary of cell lines and sensitivity to PRC2 component depletion by RNAi
 (B) Heatmap of the TPM values for indicated genes from the RNA-seq studies of 934 human cancer cell lines. RNA-sequencing data was extracted the Cancer Cell line Encyclopedia.

of PRC2 sensitivities in multiple cancer cell lines. Briefly, the general workflow of the PRC2 CRISPR tiling screen involves stably expressing Cas9 within the chosen cancer cell lines. Cancer cells then undergo rigorous quality control checks to confirm expression of Cas9, to test Cas9 cutting efficiency and to test the optimal concentration of PRC2 library virus to ensure cells only receive a single sgRNA on average. Cell lines which pass the quality control checks are then subjected to transduction with the PRC2 sgRNA library lentivirus at an multiplicity of infection (MOI) of approximately 0.1-0.3 in order to reduce the odds of two guides targeting a cell. Cells undergo antibiotic selection and cell pellets are recovered at days 3 and 18. DNA is subsequently extracted, PCR amplified and sent for next generation sequencing (NGS) (Figure 5.2A).

The cell lines RCC4_kidney, BCPAP_thyroid, HEC1A_endometrium, NCIH196_lung, A673_bone were selected previously based on their sensitivity to loss of PHF1, MTF2 and PHF19 and Cas9-expressing stable cell lines were generated using a lentiviral system. The A673 parental and Cas9-expressing cell lines were kindly provided by Dr. Gerard Brien. Western blot analysis confirmed expression of FLAG-tagged Cas9 in 4 out of the 5 cancer cell lines, RCC4, BCPAP, HEC1A and A673 (Figure 5.2B). Western blot analysis was not sufficient to detect Cas9 expression in NCIH196 and this cell line also showed a slower proliferation rate compared with other cell lines. The CRISPR screen protocol requires counting and seeding the sgRNA-infected cells every 3 days to maintain sgRNA abundance. Based on the slow growth and lowly detectable Cas9 expression, NCIH196 was dropped from subsequent future experiments. To test Cas9 cutting efficiency, Cas9-expressing cancer cell lines were infected with pXPR_011 lentivirus containing a plasmid that encodes for GFP protein in addition to an sgRNA targeting against GFP (Figure 5.2C). Cancer cell lines underwent antibiotic selection to remove non-transduced cells from the population and following 14 days in culture, the cell lines were analysed by flow cytometry using the GFP channel on the BD Accuri C6. Cell lines scoring >70% Cas9 cutting efficiency were carried forward for the PRC2 library screen. RCC4, BCPAP and A673 cell lines scored ~90% Cas9 efficiency at 14 days following pXPR infection suggesting that these cell lines were suitable for the PRC2 library screen (Figure 5.2D). Cell lines falling below the 70% Cas9 efficiency

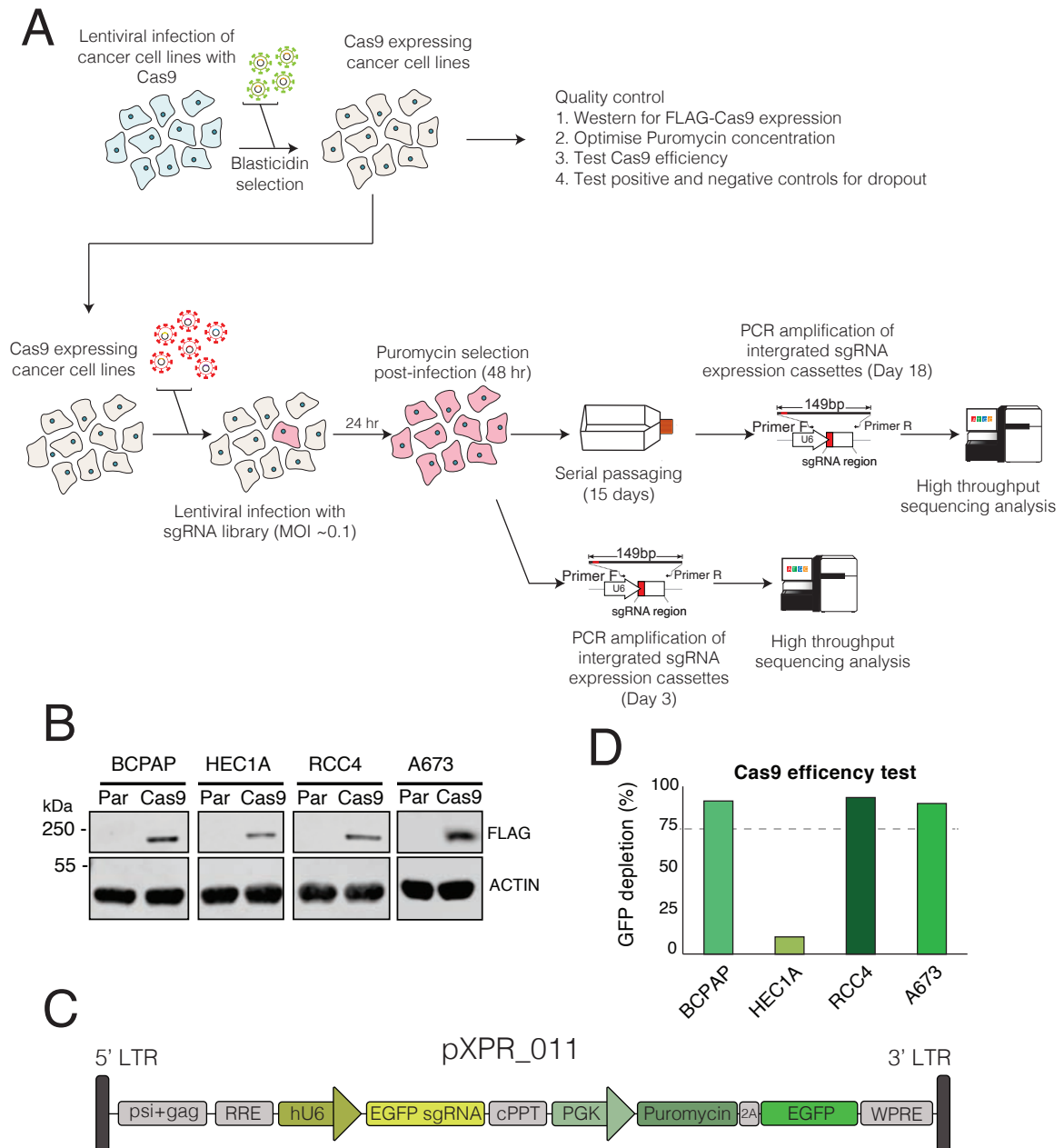


Figure 5.2 Schematic of PRC2 CRISPR tiling screen.

(A) Cancer cell lines are infected with Cas9 lentivirus particles and selected with blasticidin to generate stable cell lines. Cas9-expressing cancer cell lines are infected with PRC2 sgRNA library lentivirus at an MOI of ~0.1. DNA is extracted from cells at day 3 and day 18, used in PCR amplification and finally analysed by high throughput sequencing.

(B) Western blots of cancer cell lines expressing Cas9.

(C) Schematic of pXPR_011 plasmid.

(D) Cas9 efficiency test in the cell lines BCPAP, HEC1A, RCC4 and A673.

threshold, i.e. HEC1A, were cultured for an additional 20-30 days, as it has been speculated that Cas9 efficiency can improve following a longer culture incubation period. Repetition of the pXPR_011 transduction and GFP analysis by flow cytometry revealed insufficient Cas9 activity in HEC1A and thus, this endometrial cancer cell line was not continued to the PRC2 library screen. A673 (Ewing sarcoma), BCPAP (thyroid carcinoma) and RCC4 (renal clear cell carcinoma) all had sufficient Cas9 activity for screening purposes, however, due to the scale of tissue culture required to implement the PRC2 sgRNA screen on RCC4, the screen was performed on A673 and BCPAP.

The PRC2 sgRNA library consists of approximately 4000 sgRNAs targeting 12 PRC2-associated coding sequences and each sgRNA was on a ipUSEPR backbone containing an RFP cassette. The cancer cells were transduced with the PRC2 sgRNA lentivirus at an MOI of 0.1-0.3 to reduce the chance of two guides targeting each cell. Importantly, sgRNAs were maintained at 1000x in order to ensure sufficient guide coverage for downstream next generation sequencing (NGS) analysis. Cells were harvested on days 3 and 18 after transduction with the PRC2 sgRNA library virus. Extraction of genomic DNA from the pellets and subsequent NGS was performed by the lab of Dr. Chun-Wei (David) Chen (City of Hope, Los Angeles). Briefly, genomic DNA was extracted, amplified using primers containing NGS indexes and sequenced. Following NGS, PRC2 sgRNAs were assessed for their off-target efficiency and based on the Broad CRISPR sgRNA tool, guides which were Tier I Match Bin I Matches (i.e. has off-target effects in another coding region) were removed from the analysis. Additionally, sgRNAs with less than fifty reads at day 3 were removed. The total number of sgRNAs which were included in downstream analysis decreased from circa 4000 to 2800. In order to calculate the dependency (or dropout) of a particular sgRNA, the Log2 reads from day 18 were divided by day 3. These values were then plotted against the predicted cut site position on the protein. A value of -1 indicated that an sgRNA depleted by approximately 50% over the course of the screen.

5.2.3 Highlighting dependencies on core PRC2 in Ewing Sarcoma and thyroid papillary carcinoma cancer cell lines

CRISPR tiling screen analysis revealed dependencies on both the core and accessory proteins of PRC2. Surprisingly, sgRNAs targeting EZH2 did not deplete in A673 Ewing sarcoma nor BCPAP thyroid cancer cell lines. This result is quite stark in comparison to other cell lines, for example CAL120, which have highlighted a strong dependence on EZH2 (Figure 5.3). RNA-sequencing data, available from the CCLE, suggested that EZH2 may not be as highly expressed as EZH1 in the BCPAP cell line and may provide an explanation as to why EZH2 guides are not dropping out (Figure 5.1B). While EZH1 sgRNAs did appear to deplete slightly across the entire gene in the BCPAP cell line, overall, there were no significant clusters of guides depleting within functional domains (SANT1/2, SET, etc.) over the course over 18 days. Although sgRNAs targeting the remaining core members, EED and SUZ12, did appear to marginally drop out over the course of 18 days across the entire gene, the guides did not deplete considerably within any of the functionally important domains (i.e NR/C2/VEFS domain of SUZ12). Contrastingly, a large proportion of EED and SUZ12 sgRNAs depleted in both CAL120 (Figure 5.3) suggesting a dependency on core PRC2 in this cancer cell line. Taken together, these results suggest that A673 and BCPAP were not particularly sensitive to loss of core PRC2.

5.2.4 PHF1 is a predicted dependency in Ewing Sarcoma and thyroid carcinoma cancer cell lines

In addition to highlighting the dependency (or lack of in some instances) of core PRC2, dropout of accessory protein sgRNAs was also observed in A673 and BCPAP. Both cell lines were sensitive to depletion of PHF1, however, this dependency was not observed in the CAL120 (Figure 5.3). Surprisingly, a dependency on MTF2 or PHF19 was not observed and PHF1 was the only Polycomb-like protein to dropout in A673 or BCPAP cell lines (Figure 5.3) sgRNA targeting PRC2.2 components, AEBP2 and JARID2, did not score in A673 and BCPAP, however, given that these cell lines were selected for CRISPR screen

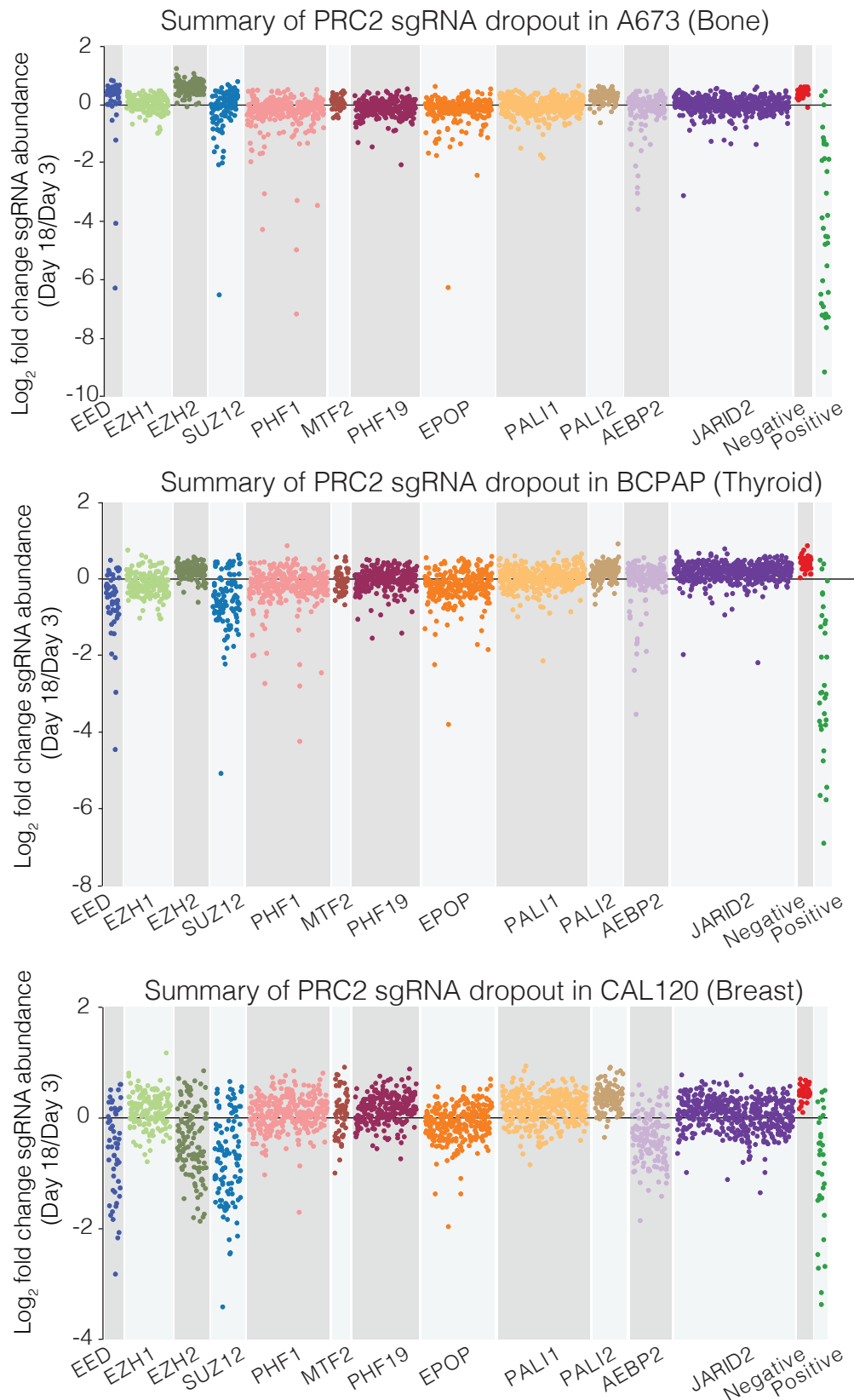


Figure 5.3 PHF1 is a dependency in A673 and BCPAP cell lines but not CAL120.

Tiled CRISPR screening for PRC2 components in (A) A673, (B) BCPAP and (C) CAL120. The Log_2 Fold change was calculated based on the average reads from Days 18 over Day 3. The PRC2 CRISPR tiling screen in the cell line CAL120 was performed by Marlena Mucha.

analysis based on publicly available data suggesting sensitivity to the Polycomb-like proteins, these results are not surprising. Initial analysis of AEBP2 highlighted potential sensitivities within the N-terminus, however, due to the long stretches of GC repeating regions within the N-terminus of AEBP2, a large proportion of the sgRNA were found to be off-target and I therefore decided to not pursue any of these hits observed in AEBP2.

sgRNAs targeting EPOP, PALI1 and PALI2 were also examined during the screen, however, no significant clusters of sgRNAs for PALI1/2 dropped out. DepMap data has predicted EPOP as an essential gene in approximately 75% of the cancer cell lines screened to date. The structure of EPOP is poorly understood and the functional domains have been loosely annotated as the N-terminal Elongin-BC-interacting, the central region and the C-terminal PRC2-interacting domains (Liefke *et al.*, 2016; Liefke & Shi, 2015). A region of EPOP corresponding to amino acids 59-63 was dropping out by approximately 50% in both A673 and BCPAP (Figure 5.3). A BC-box motif (P/S/T-L-x-x-x-C/S/A-x-x-x-L/I/A/V) has been predicted within this region and so depletion of sgRNA within this region may be involved in the interaction with ElonginBC, however, I elected to focus on the sensitivities identified in PHF1 as these guides were the most significant in A673 and BCPAP.

5.2.5 CRISPR tiling sgRNA targeting amino acids 19-27 and 125-133 of PHF1 deplete in A673 and BCPAP but not CAL120 or Pfeiffer

CRISPR tiling screens have revealed a potential dependency on PHF1 in A673 and BCPAP in two separate regions of the proteins; amino acids 19-27 located immediately adjacent to the Tudor domain and amino acids 125-133 which fall within the PHD domain (Figure 5.4A). The majority of sgRNAs targeting amino acids 19-27 of PHF1 reduced by 50% in both cell lines over the course of the 18 day screen. Surprisingly, despite the overall structural and sequence similarity between the three Polycomb-like proteins, depletion of these amino acids was only specific to PHF1 (Figure 5.4B). In addition to sgRNAs depleting at amino acids 19-27, guides targeting the region 125-133 within the PHD1 domain also dropped out by at least

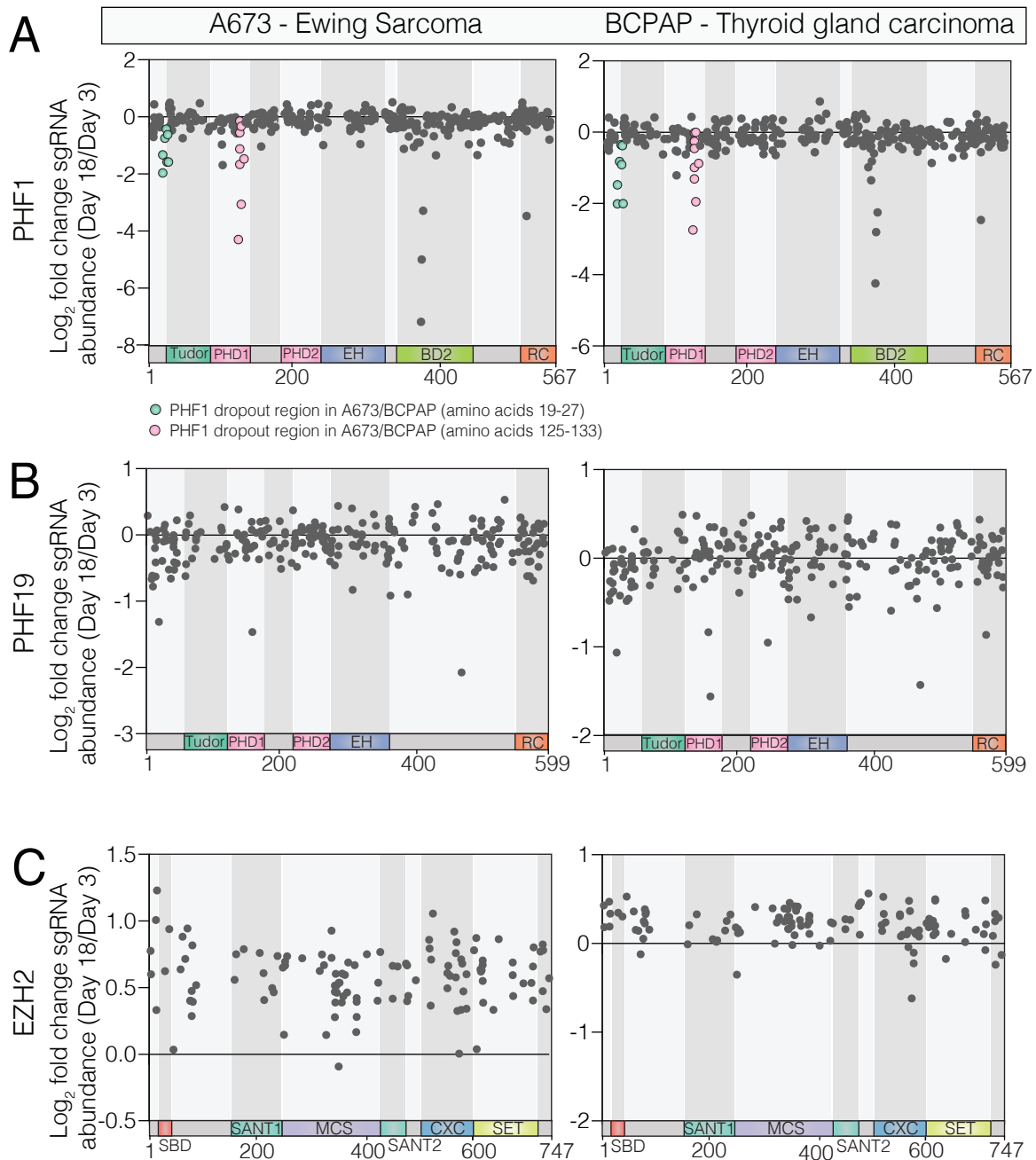


Figure 5.4 CRISPR tiling sgRNA targeting codons corresponding to amino acids 19-27 and 125-133 of PHF1 deplete in A673 and BCPAP cancer cell lines.

CRISPR tiling screen of (A) PHF1, (B) PHF19 and (C) EZH2 in A673 and BCPAP cell lines. Log₂ fold change in sgRNA abundance was calculated from the reads on Day 18 over Day 3. Off-target sgRNA and sgRNA with reads of less than 50 on Day 3 were removed from the analysis.

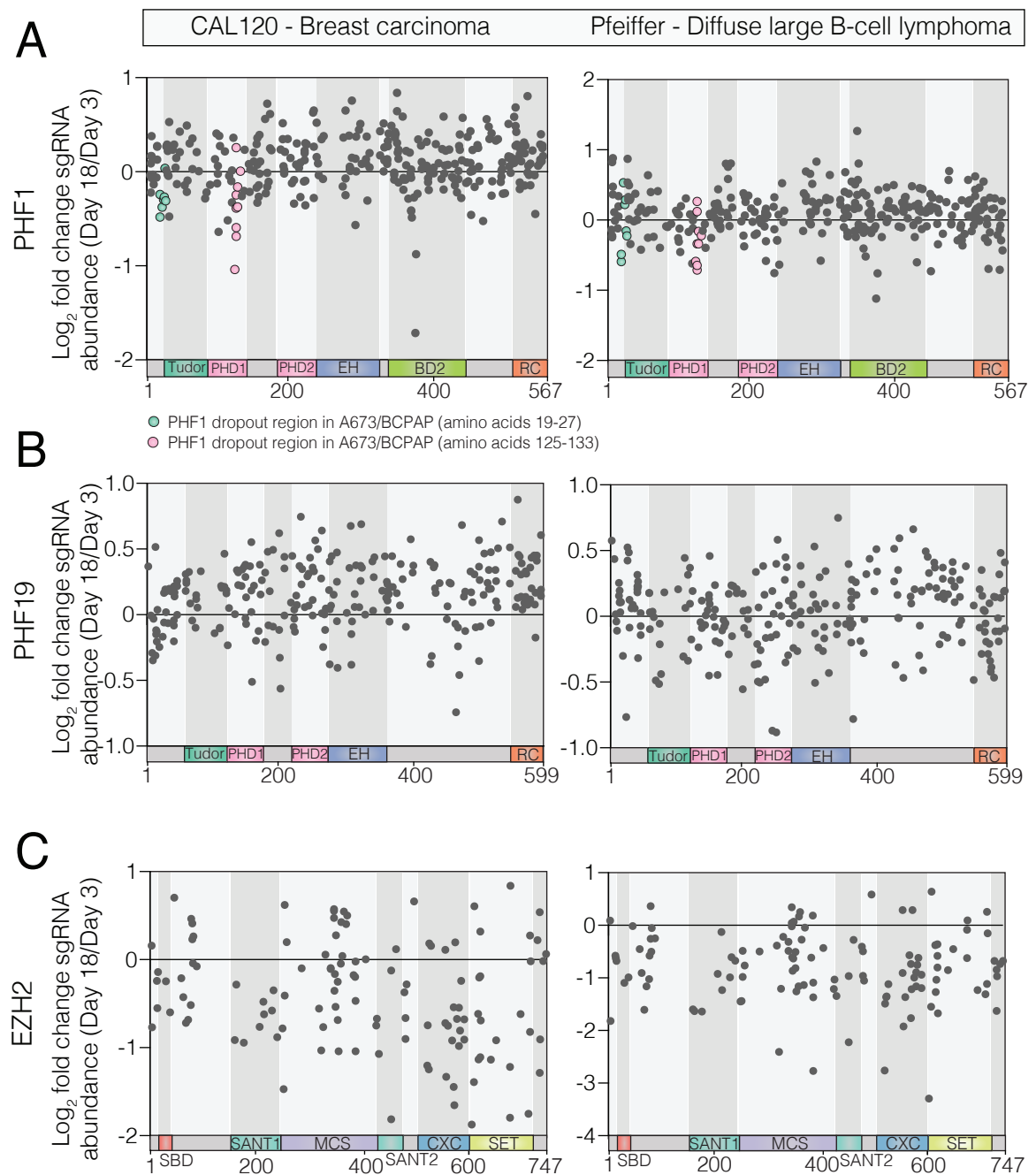


Figure 5.5 EZH2 but not PHF1 is a dependency in CAL120 and Pfeiffer cancer cell lines.

CRISPR tiling screen of (A) PHF1, (B) PHF19 and (C) EZH2 in CAL120 and Pfeiffer cell lines. Log₂ fold change in sgRNA abundance was calculated from the reads on Day 18 over Day 3. Off-target sgRNA and sgRNA with reads of less than 50 on Day 3 were removed from the analysis. The CRISPR tiling screens for the cell lines CAL120 and Pfeiffer were performed by Marlena Mucha, Dr. Eric Conway and Dr. James Nolan.

50% over the course of the screen (Figure 5.4A). As established previously, the depletion of these guides was specific to PHF1 as the region corresponding to 125-133 did not drop-out within PHF19 (Figure 5.4B). sgRNA targeting the amino acids 19-27 and 125-133 of PHF1 did not deplete in the cell lines CAL120 and Pfeiffer, however, these cell lines were sensitive to EZH2 depletion confirming there is sufficient Cas9 activity within these cells but the cells are not dependent on PHF1 (Figure 5.5C).

5.2.6 Validation of dependency on amino acids 19-27 and 125-133 in A673, BCPAP but not CAL120 cells

One potential fault in CRISPR tiling screens is that the predicted dependencies on PHF1 in A673 and BCPAP could be as a result of two or more guides entering each cell. Cells were transduced with the PRC2 sgRNA library viruses at an MOI of 0.1-0.3 to account for this inconsistency however I next decided to validate the results of the PRC2 CRISPR tiling screen by performing CRISPR using each individual guide targeting codons corresponding to amino acids 19-27 and 125-133. To ensure sufficient coverage, I used between 7-9 sgRNA for each of the predicted novel regions (Figure 5.6A&C). Additionally, I included several positive and negative control sgRNAs such as sgLuciferase, sgRPA3 (RNA Polymerase 3) (Figure 5.6B) and sgRNAs targeting the C-terminus of PHF1 (Figure 5.6D). I performed the validation in the A673, BCPAP and as a negative control, CAL120 cell lines. As each sgRNA was cloned into a plasmid containing an RFP cassette, I measured RFP intensity using the Guava easyCyte HT system Flow cytometer at days 6, 12, 18 and 24 to assess the abundance of each guide. The depletion of cells transduced with the sgRNA targeting RPA3 confirmed sufficient Cas9 activity within these three cell lines. sgEZH2 guides were included to serve as a negative control for A673 and a positive control for the CAL120 cell lines. Mirroring the results from the CRISPR screen, sgRNA targeting EZH2 did not deplete in A673, however, EZH2 was identified as a sensitivity in CAL120 (Figure 5.6B). sgRNA targeting the C-terminus of PHF1 did not substantially deplete over the course of the experiments in any of the cell lines confirming that lack of dependency on the C-terminus. The majority of sgRNA targeting codons corresponding to amino acids 19-27 and 125-133 were

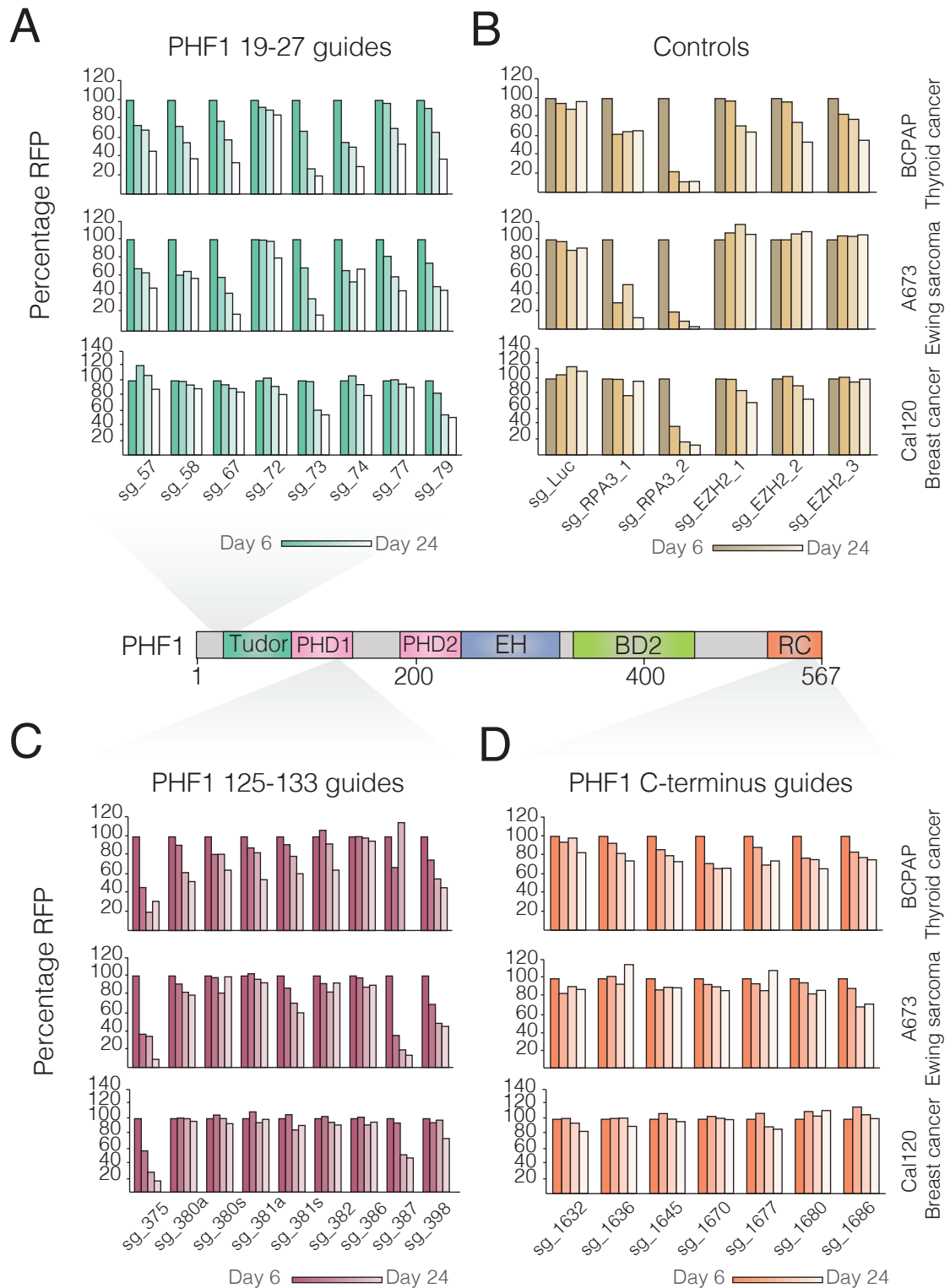


Figure 5.6 Validation of dependency on regions 19-27 and 125-133 of PHF1 in A673 and BCPAP but not CAL120 cell lines.

Individual guides targeting amino acids within the regions 19-27, 125-133 were transduced into the indicated cell lines. The abundance of RFP was measured in the cells at 6-day intervals. sgRNA targeting the C-terminus of PHF1 were included as negative controls.

dropping out thus confirming a sensitivity to PHF1 depletion in A673 and BCPAP but not CAL120. There were some sgRNA dropout in CAL120, for example sg79 and sg375 and sg387, which perhaps indicates that some of these sgRNA have off-target effects (Figure 5.6A&C).

5.2.7 The regions 19-27 and 125-133 of PHF1 are highly conserved in mammalian PHF1

Evolutionary analysis of these two clusters of amino acids in PHF1 revealed a strong sequence conservation across many mammalian species with a loss of conservation outside of mammals (Figure 5.7A). Additionally, overlaying the crystal structures of PHF1 and MTF2 did not highlight any structural similarities between the amino acids 125-133 of PHF1 and the equivalent 140-148 of MTF2 (Figure 5.7B). Interestingly, sgRNAs targeting the amino acids 19-27 and 125-133 of PHF1 did not score in CAL120, indicating that the dependency on PHF1 occurs in certain cancer cell types but not all. Surprisingly, sgRNA targeting the C-terminus of PHF1 did not drop out during the screen as the C-terminus of PHF1 has been implicated in binding to SUZ12. I hypothesised that the sgRNA dropping out towards the N-terminus of PHF1 were involved in a PRC2-independent function. A PRC2-independent role of PHF1 was characterised by a number of groups in recent years whereby PHF1 is bound to the C-terminal tail of p53 subsequently blocking MDM2-mediated ubiquitination and degradation (Brien *et al.*, 2015; Yang *et al.*, 2013). In addition to binding to p53, PHF1 has been speculated to recognise H4R3me2s via the PHD1 domain (Liu *et al.*, 2018). I hypothesised that the regions 19-27 and 125-133 were contributing to a PRC2-independent role of PHF1 in these cancer cell lines.

5.2.8 Expression of mouse PHF1 does not rescue PHF1 sgRNA depletion in A673

To confirm the PHF1 sensitivity and rule out false positive hits, I elected to stably express mouse PHF1 in A673 cells to determine whether exogenous expression of mouse Phf1 was sufficient to rescue depletion of endogenous human PHF1 (Figure

5.8A). The mouse Phf1, Phf19 and empty vector constructs were transfected into cells and selected for a number of days to ensure integration of the plasmid into the genome of the cells. Due to difficulties maintaining A673 at single cell density, cells were continued as a polyclonal cell line. Western blot analysis confirmed the expression of FLAG-tagged PHF1 and PHF19 constructs in A673 (Figure 5.8B). Individual sgRNA targeting the regions 19-27, 125-133 and the C-term of PHF1 were transduced into A673 stably expressing mouse PHF1 and PHF19. Five sgRNA targeting codons corresponding to amino acids 19-27 (sg57, sg58, sg67, sg74, sg77), three sgRNA targeting codons corresponding to amino acids 125-133 (sg380, sg381 and sg398) and a single sgRNA (sg1632) targeting the C-terminus of PHF1 were included in the rescue experiment. Two independent sgRNAs targeting RPA3 depleted rapidly in the three cell lines, A673-EV, A673-mPhf1 and A673-mPhf19, confirming sufficient Cas9 cutting activity. While sgRNA targeting amino acid regions 19-27 and 125-133 dropped out in the A673-EV cell line, they also comparably depleted in the A673-mPhf1 cell line suggesting that expression of mouse Phf1 was not sufficient to rescue the effect of knocking of PHF1 in A673 cells (Figure 5.8C).

5.2.9 Strategy for improved rescue of PHF1 depletion in A673 cells

I next examined the protein levels of exogenously expressed FLAG-tagged mPhf1 and mPhf19 in A673 cells after finishing the rescue experiment as described in Figure 5.8. Western blot analysis suggested that expression of mouse Phf1 was at a low level compared to Phf19 (Figure 5.9A). I hypothesised that the poor expression of mouse Phf1 coupled with potential targeting of mouse Phf1 may have resulted in the inconclusive rescue experiment results.

Subsequently, I designed a new codon-optimised human PHF1 construct which is more varied than mouse Phf1 and therefore should not be targeted by the PHF1 sgRNAs (Figure 5.9B). The new, improved strategy for the rescue of PHF1 depletion in A673 cells involves lentiviral transduction of the construct into cells as opposed to transient transfection followed by antibiotic selection. This can potentially improve

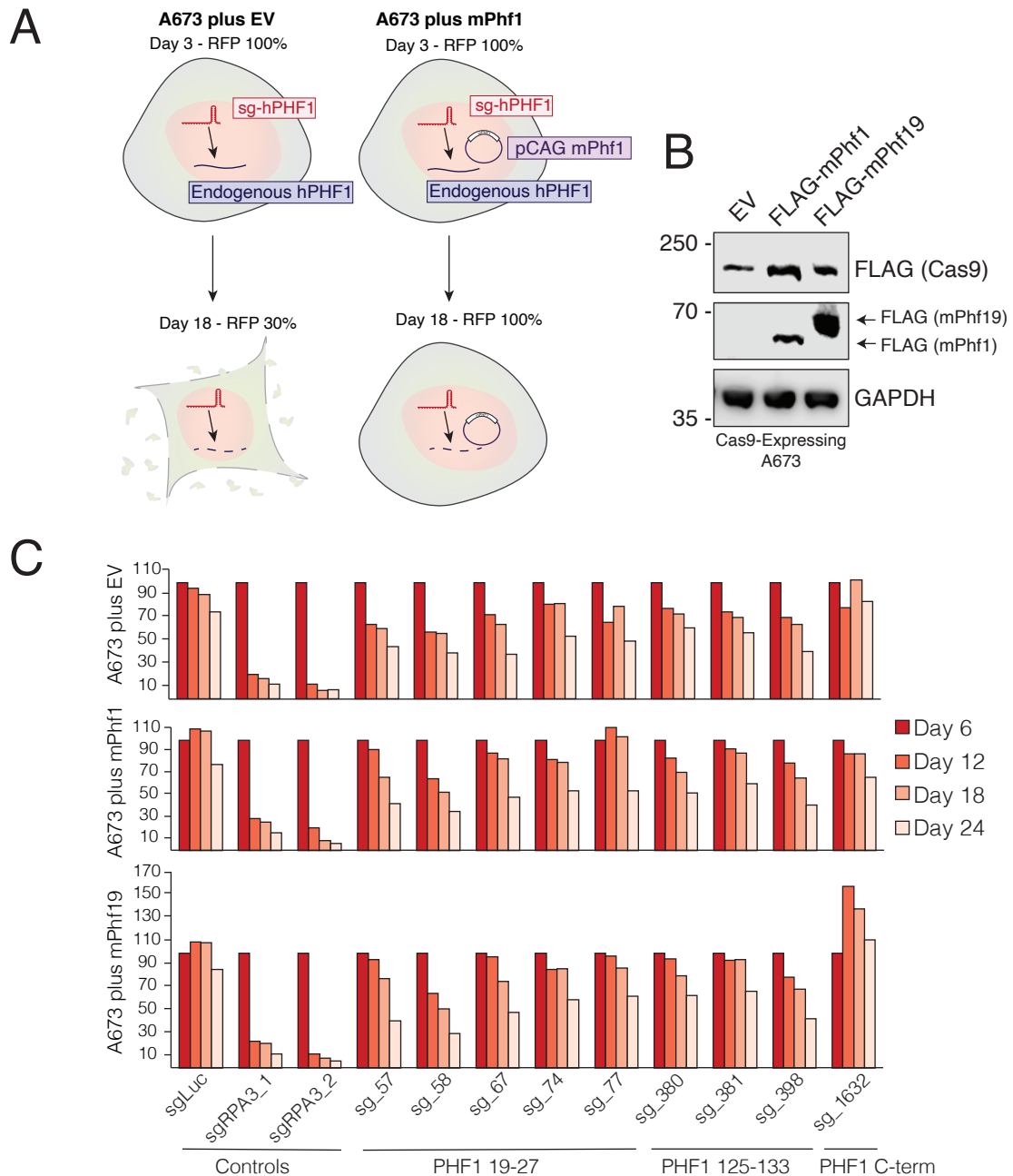


Figure 5.8 Expression of mouse PHF1 does not rescue depletion of human PHF1 in A673 cells.

(A) Schematic of PHF1 rescue experiment in A673 cells. Cas9-Expressing A673 cells were transfected with FLAG-tagged mouse Phf1 (mPhf1), followed by transduction with PHF1 sgRNA. sgRNA abundance was based on RFP abundance measured using Guava easyCyte flow cytometer.

(B) Western blot analysis of A673 cell lines expressing empty vector (EV), mouse PHF1 and mouse PHF19 (negative control).

(C) Depletion of sgRNA in EV, mPhf1 and mPhf19-expressing A673. sgRNA abundance on Days 6, 12, 18 and 24 was based on RFP abundance measured using Guava easyCyte flow cytometer.

expression levels of exogenous PHF1. Whilst the mouse Phf1 construct should not have been recognised by the human sgRNA, there is the potential that some PHF1 sgRNA did. The codon-optimised construct varies from human PHF1 at every third base and cannot be targeted by the sgRNA. Taken together, the codon-optimised PHF1 construct coupled with lentiviral transduction will provide a better cell model for validation of the CRISPR screen hits.

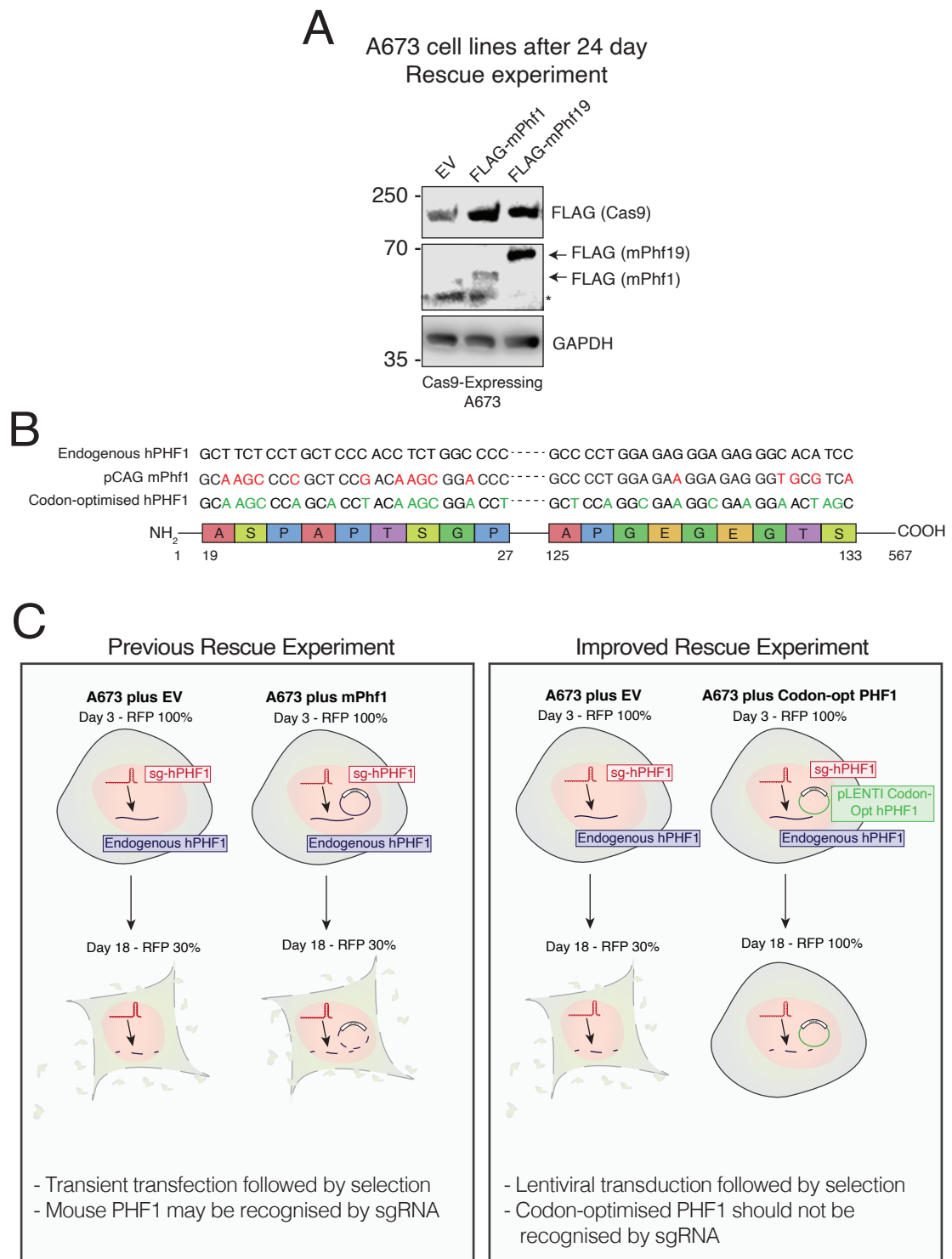


Figure 5.9 Strategy for improved rescue of PHF1 depletion in A673 cells

(A) Western blot analysis of A673 cell lines after completion of the rescue experiment in Figure 5.8 (C). * denotes a non-specific band recognised by the FLAG antibody.

(B) Comparison of endogenous human PHF1 (hPHF1), mouse PHF1 and codon-optimised hPHF1 constructs.

(C) Schematic of initial and improved rescue experiments in A673 cells.

5.3 Discussion

CRISPR-Cas9 screens are valuable tools for identifying novel therapeutic targets in cancer cell lines. In this chapter, I have performed a PRC2-focused CRISPR-Cas9 knockout screen in a selection of cancer cell lines with the aim of identifying novel dependencies on PRC2 proteins. Focusing on PRC2 proteins, I employed an sgRNA library of approximately 4000 guides to tile across each of the PRC2 associated coding sequences in two biological replicates in the cancer cell lines A673 and BCPAP. The advantage of a detailed, PRC2-focused screen over the traditional genome-wide screens allows us to use many sgRNAs to tile across the entirety of each PRC2-associated gene to pinpoint regions or domains of functional importance in each gene. Additionally, targeting with multiple sgRNA within each gene is important as it assists to rule out false-positive hits and provide confidence in the hits observed.

Analysis of the CRISPR screen results from A673, BCPAP, CAL120 and Pfeiffer have identified potential dependencies on PRC2 proteins in these cancer cell lines. Initially, it was surprising that neither A673 nor BCPAP were dependent on EZH2 as this core PRC2 subunit has been reported to promote tumour proliferation. Specifically, downregulation of EZH2 in A673 has been suggested to impair cell proliferation *in vitro* and tumour growth *in vivo* (Richter *et al*, 2009) and EZH2 has also been identified as a potential therapeutic target in Ewing sarcoma (Pandya *et al*, 2018). These discrepancies may be attributable to the variance between CRISPR knockout compared to short hairpin targeting strategies. Additionally, Pandya *et al*, have reported that while tumour growth was impaired 5 days following treatment, tumour growth has recovered by day 10. Alternatively, it has been reported that EZH1 can compensate for loss of EZH2 in cells and this may, in part, explain why EZH2 was not a dependency in these cells. On a similar note, given the sequence similarity between PHF1, MTF2 and PHF19, it could be speculated that Polycomb-like proteins are capable of compensating after knockout of one of the PCLs. This may explain why no regions of MTF2 and PHF19 and only two notable regions of PHF1 dropped out in the screen. In future, it could be interesting to assess the effects of Polycomb-like sgRNA in a KO background where only one of the Polycomb-like proteins exist.

I identified two potential candidate regions within PHF1, amino acids 19-27 and 125-133, which specifically depleted in A673 (Ewing sarcoma) and BCPAP (thyroid papillary carcinoma) cancer cell lines. The depletion of guides targeting amino acids 19-27 was intriguing because although this region is predicted to be largely unstable with no characterised role, it has been speculated that the amino acids preceding the Tudor domain of PHF1 give rise to more stable histone engagement (Gibson *et al.*, 2017). Similarly, the PHD1 domain has been reported to interact with the C-terminal tail of p53 (Brien *et al.*, 2015; Yang *et al.*, 2013) and recently, H4R3me2s modification (Liu *et al.*, 2018). I speculated that disruption of this region within A673 and BCPAP is inhibiting the interaction with other proteins in a similar manner. I attempted to validate the depletion of PHF1 in A673 cells by expression of a mouse Phf1 construct to “rescue” the loss of PHF1, however, I was unable to validate the candidate PHF1 hits in the CRISPR tiling screen. Potentially, this may occur due to the unpredicted off-target effects of sgRNA highlighting false-positive hits within the cells or alternatively, may be due to technical issues with the experiment. For example, it could be conceivable that the sgPHF1 were still capable of recognising and binding to the exogenous mouse construct. An alternative solution is to engineer a codon-optimised human construct. Additionally, generation of a stable cell line using a lentiviral backbone plasmid would potentially improve expression of the construct. While the mouse PHF1 and PHF19 constructs were detectable by Western Blot analysis prior to beginning the rescue experiment, expression decreased after approximately 24 days in culture. Generating a stable cell line by lentiviral transduction may result in prolonged, stable expression of the rescue constructs. Further validation of the PHF1 hits is necessary in order to confirm the dependency on amino acids 19-27 and 125-133, however, the results from CRISPR knockout screen have also highlighted dependencies on the remaining PRC2-subunits. EPOP has been predicted in a dependency in approximately 75% of all cancer cell lines analysed on DepMap.org and reported to drive the proliferation of cancer cell lines via the ElonginBC interaction (Liefke & Shi, 2015). EPOP does present as a possible dependency in A673, BCPAP and CAL120 and while these results have not been validated within this thesis, further experiments could be performed to test these dependencies.

Chapter 6: **General Discussion**

6.1 Introduction

6.1.1 The role of Polycomb-like proteins in the PRC2 complex

PcG proteins are essential regulators during the development, differentiation and maintenance of cell lineages (Bracken *et al.*, 2006; Conway *et al.*, 2015) and the importance of these chromatin regulating complexes is particularly striking following the knockout of core components. Notably, loss of core PRC2 members, Ezh2, Eed or Suz12, results in embryonic lethality in mice underlining the critical function of PRC2 during early development (Faust *et al.*, 1995; O'Carroll *et al.*, 2001; Pasini *et al.*, 2004). Polycomb-like proteins (PHF1, MTF2 and PHF19) are key PRC2 accessory proteins within PRC2 and define a unique PRC2 subcomplex, termed PRC2.1 (Alekseyenko *et al.*, 2014; Hauri *et al.*, 2016). Polycomb-like proteins likely mediate recruitment of the complex to chromatin via an N-terminal Tudor domain which recognises and binds to H3K36me3 (Ballaré *et al.*, 2012; Brien *et al.*, 2012; Cai *et al.*, 2013) as well as through the centrally-located DNA-binding extended homologous (or winged helix-like) domain (Choi *et al.*, 2017; Li *et al.*, 2017; Perino *et al.*, 2018). Loss of Polycomb-like proteins in mouse ES cells evicts PRC2.1 from chromatin and results in genome-wide re-distribution of H3K27 methylations and consequently results in deregulation of Polycomb target genes during ESC differentiation (Healy *et al.*, 2019). In addition to the targeting role of Polycomb-like proteins, they also stimulate the in vitro and are key PRC2 accessory proteins methyltransferase activity of the complex (Cao *et al.*, 2008; Healy *et al.*, 2019; Højfeldt *et al.*, 2019; Sarma *et al.*, 2008; Zhang *et al.*, 2019).

6.1.2 The PRC2-independent role of PHF1

Interestingly, while the three Polycomb-like proteins (PHF1, MTF2 and PHF19) share a high level of sequence similarity and serve similar functions within PRC2, several groups have reported a PRC2-independent role of PHF1 whereby it can bind to and interact with other proteins outside of PRC2. The first report of a PRC2-independent function of PHF1 was highlighted by Hong *et al.* 2013 where PHF1 localised at sites of DNA damage and interacted with DNA damage response

proteins including p53 (Hong *et al.*, 2008). This apparent relationship between PHF1 and DNA damage was interrogated further by Yang *et al.* and Brien *et al.* which reported that PHF1 was capable of interacting with p53 through divergent domains within both proteins (Brien *et al.*, 2015; Yang *et al.*, 2013). Brien *et al.* showed that this interaction stabilises p53 protein levels and promotes cellular quiescence (Brien *et al.*, 2015). A recent study has also proposed PHF1 as a bridge between PRMT5–WDR77, and CRL4B complex subunits and also as a reader of symmetric demethylation on H4R3 (H4R3me₂s), a histone mark deposited by PRMT5 (Liu *et al.*, 2018).

6.2 General Discussion

6.2.1 Summary of findings

The overall main goal of the research presented in this thesis was to characterise the PRC2-dependent and independent roles of PHF1. In chapter 3, I successfully recovered the PHF1-p53 complex from human cells and provided evidence that PHF1 interacts more stably with members of the PRC2 complex than p53. Supporting the hypothesis of a PHF1-p53 assembly, I purified these proteins from insect cells and have estimated the size to be approximately 200 kDa. I hypothesise that the complex contains a dimer of p53 coupled with a single PHF1 subunit, however, the complex requires further characterisation. I next assessed the binding profile of p53 by performing ChIP-seq in 4 cell conditions, DMSO-treated, Etoposide-treated, Quiescent and Cycling in two separate cell lines. The majority of p53 peaks largely overlapped in Etoposide and quiescent cells, however, there was a small subset of p53 targets that were enriched in one condition over the other. In parallel, RNA-seq was performed on these cellular conditions to assess mRNA expression with a particular focus on PRC2 component expression in quiescent and cycling cells. PHF1 and EZH1 were found to be highly expressed in quiescent cells while EZH2 and PHF19 were elevated in proliferating cells.

In chapter 4, I next characterised PRC2 protein assembly in quiescent and cycling cells. Previously unpublished data generated within the Bracken lab had suggested an absence of SUZ12 from quiescent cells, however the absence of SUZ12

rendered the complex highly unstable following recombinant purification from insect cells in Chapter 4. Towards characterising the complex composition, I next performed mass spectrometry in quiescent and cycling cells with EED, EZH1 and EZH2 antibodies, Mirroring the RNA-seq results from chapter 3, PHF1 and EZH1 were highly expressed and formed the predominant complex within quiescent cells which subsequently switches to EZH2, PHF19 and AEBP2 in proliferating TIG3-T. Treatment of quiescent TIG3-T or BCPAP (thyroid cancer cell line) with an EED degrader did destabilise the complex but did not result in a reduction of H3K27me3. A growth assay of TIG3-T suggested that quiescent cells treated with an EED degrader or inhibitor re-entered the cell cycle as normal.

In chapter 5, I performed a CRISPR-Cas9 tiling screen using an sgRNA library of approximately 4000 guides to target 12 PRC2 subunits. I identified potential dependencies in two regions of PHF1, amino acids 19-27 and 125-133, in two cell lines, A673 and BCPAP. I have attempted to validate of these predicted dependencies by expressing mouse Phf1 in A673 cells to rescue of the effect of sgRNA targeting the predicted dependent regions. Unfortunately, these validations were unsuccessful, however, I have re-designed an improved rescue experiment using a codon-optimised human PHF1 construct.

6.2.2 PHF1 as a p53 target

A study by Brien et al has reported that while the expression of MTF2 and PHF19 are regulated by E2F transcription factors, PHF1 is a p53-regulated gene (Brien *et al.*, 2015). These observations were based on ChIP-qPCR analysis performed in primary human diploid fibroblasts, however, the results from the research presented in Chapter 3 of this thesis do not support this hypothesis.

In Chapter 3, I performed p53 ChIP-seq in two human fibroblast cell lines, IMR90 and TIG3-T, in four cell conditions, DMSO-treated, Etoposide-treated, Quiescent and Cycling. Although IMR90 and TIG3-T cells are of the same lineage to the HDFs used in the 2015 study, it could be possible that PHF1 is not a p53 target in all quiescent fibroblast models. A recent review of p53 targets has reported that of all p53 target genes identified from 16 high-throughput datasets, CDKN1A and RRM2B

were the only two genes present in all datasets (Fischer, 2017). It is possible that PHF1 may only be a target in the HDF cell line used in the 2015 paper. Additionally, as described in the Discussion section of Chapter 3, it is possible that technical issues within the ChIP protocol may account for the absence of p53 on the PHF1 promoter. The p53 antibody used in this thesis was derived from an in house monoclonal supernatant which potentially has reduced in potency and affinity for antigen over the course of several years (Johnson, 2012). Likewise, another technical variance is the crosslinking of quiescent cells with formaldehyde prior to performing ChIP. Given that PHF1 is not expected to be a strong p53 target, the interaction between p53 and the PHF1 promoter may be more transient compared to typical p53 targets, for example CDKN1A, and so it would be valuable to repeat the p53 ChIPs using a stronger cross-linker and with a new stock of p53 antibody to rule out these technical discrepancies. While the p53 ChIP results did not reveal PHF1 as a quiescent p53 target, the datasets generated are still a valuable resource for comparison of p53 target genes in 4 separate cell conditions two cell lines.

6.2.3 PRC2 complex assemblies in quiescent and cycling cells

The RNA-seq and mass spectrometry results presented in Chapters 3 and 4 have highlighted PHF1 as the most abundantly expressed PRC2 subunit in quiescent TIG3-T cells. These data are consistent with previous observations of a switch from PHF1 to MTF2/PHF19 expression in quiescent and cycling cells (Brien *et al.*, 2015). The role of Pcl proteins in the recruitment of the complex to polycomb target sites has been well characterised in proliferating mouse ES cells (Brien *et al.*, 2012; Casanova *et al.*, 2011; Healy *et al.*, 2019; Højfeldt *et al.*, 2019; Walker *et al.*, 2010) however, the majority of studies have focused on Mtf2, given that it is the most expressed of the Pcls in mouse ES cells. The Pcl proteins are hypothesised to serve a similar function within PRC2, however, characterisation of PHF1, MTF2 and PHF19 binding within the same cellular context would be valuable to compare differences and similarities in the targeting abilities of these three structurally similar proteins. Furthermore, it would be interesting to contrast against the binding profile of PHF1 which is elevated during quiescence versus PHF19 in cycling cells. Additionally, the data presented within this thesis provided evidence to suggest that a smaller portion of PRC2 within quiescent cells contained the PRC2.2 component,

AEBP2 and it will be important to determine the role (if any) of PRC2.2 during quiescence.

6.2.4 The role of EZH1-PRC2 in facilitating chromatin compaction

In Chapter 3 and 4 of this thesis, I have presented data suggesting the elevated expression of EZH1 in serum-starved TIG3-T cells. EZH1-PRC2 has a higher affinity for nucleosomes than EZH2-PRC2 and has been speculated to contribute to the condensation of chromatin, an activity speculated to be mediated by dimerization of the complex (Grau *et al.*, 2021; Margueron *et al.*, 2008; Son *et al.*, 2013). On a similar note, EZH1 is among the most upregulated genes expressed during the quiescent-like state of diapause in African Turquoise Killifish, Diapause is an event that occurs during the dry seasons for a period of 4-6 month during which cells are maintained in a dormant, suspended state (Hu & Brunet, 2018; Hu *et al.*, 2020). Cell quiescence has typically been associated with an overall reduction in transcription and hypothesised to correlate with a more compacted, condensed chromatin structure (McKnight *et al.*, 2015; Swygert *et al.*, 2019). A similar phenomena of high levels of chromatin compaction is observed in quiescent B-cells prior to activation. B-cell activation results in an increase in global acetylation coupled with decompaction of chromatin (Kieffer-Kwon *et al.*, 2017). The precise mechanism in which EZH1-PRC2 facilitates chromatin compaction remains unclear for now but it is potentially related to the dimerisation of the complex (Grau *et al.*, 2021). RNA-seq analysis in quiescent and cycling TIG3-T have also highlighted elevated expression levels of several cPRC1 subunits, for example CBX7, CBX8, PHC1 (Data not shown) which I hypothesise may facilitate compaction of chromatin *in vivo*. While these results have not been explored in this thesis, it is interesting to speculate that the increased presence of cPRC1 may also contribute another layer of chromatin compaction in quiescent cells complementing the predicted EZH1-mediated compaction. Taken together, I hypothesise that EZH1-PRC2 together with cPRC1 are facilitating the compaction of chromatin during quiescence. Treatment of quiescent cells with PRC2 inhibiting/degrading drugs in combination with CBX7/8 inhibitors may provide a potential strategy in targeting these chemotherapeutic-resistant quiescent cells (Lamb *et al.*, 2019; Suh *et al.*, 2021).

6.2.5 PHF1 as a potential target in cancer therapies

The validation of PHF1 hits identified from the CRISPR-Cas9 PRC2 tiling screen in Chapter 5 remains an ongoing experiment, however, it would be interesting to perform PRC2 tiling screen in a context where only one Pcl protein is present. The cell lines selected for the screen (A673 and BCPAP) expressed all three Polycomb-like proteins to varying degrees and therefore a compensatory mechanism between the proteins is likely (Jain *et al.*, 2020).

The data presented in this thesis has provided strong evidence for the assembly of PHF1 within two distinct complexes, namely PHF1-PRC2 and PHF1-p53. PHF1-PRC2 has been suggested as the predominate complex within quiescent TIG3-T cells and similarly, PHF1-p53 has been reported as a complex that only occurs in quiescent cells (Brien *et al.*, 2015; Yang *et al.*, 2013). Loss of PHF1 using RNAi results in a reduction in the number of quiescent cells and therefore, inhibition or degradation of PHF1 could be a potential strategy in targeting both PHF1-containing complexes in quiescent cells. It has been established that when grown under selective pressure, lymphoma cancer cells develop mutations rendering them insensitive to PRC2 inhibition (Baker *et al.*, 2015; Gibaja *et al.*, 2016). These mutations are of particular concern given that EZH2 inhibitors have been approved by the FDA for the treatment of follicular lymphomas and represent the urgency in finding alternative strategies for disruption of the complex.

6.2.6 Further perspectives

In the research chapters of this thesis, I have presented strong data characterising of PRC2 complex assemblies in quiescent and proliferating cells and the data is consistent with published data as well as unpublished data within the Bracken lab. Further validation of these observations is needed to determine the impact of the various subcomplex assemblies during quiescence and proliferation of TIG3-T cells. Additional validation in another quiescence model, for example, BCPAP the

quiescent cancer model highlighted in Chapter 5, would further solidify these results. The absence of PHF1 as a p53 target was surprising however as discussed in section 6.2.2, there are several technical issues that may explain the discrepancies in results.

Further experiments are required to validate the CRISPR screen results and provide more confidence in predicted dependencies. I performed a PRC2 CRISPR tiling screen using a library of approximately 4000 sgRNA but ultimately lost around one-third of guides due to predicted off-target effects or low reads on the starting day of the screen. The loss of a huge number of guides can make it difficult to interpret the findings from this screen and going forward I would propose a re-design of the PRC2 sgRNA pooled library to remove the off-target guides.

6.3 Conclusions

The data presented in this thesis support the following conclusions:

- PHF1 and p53 stably interact as a complex and can be purified from human and insect cells
- A large proportion of p53 targets are shared in Etoposide and Quiescent TIG3-T fibroblasts
- RNA expression of EZH1 and PHF1 is elevated in quiescent TIG3-T cells with a switch in expression to EZH2 and PHF19 in cycling cells
- G0-PRC2.1, consisting of PHF1, EZH1, EED, SUZ12 and RBBP4, is the predominant complex found quiescent TIG3-T
- Quiescent TIG3-T are refractory to PRC2 inhibition or degradation

References

- Agger K, Cloos PA, Christensen J, Pasini D, Rose S, Rappsilber J, Issaeva I, Canaani E, Salcini AE, Helin K (2007) UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 449: 731-734
- Alabert C, Barth TK, Reveron-Gomez N, Sidoli S, Schmidt A, Jensen ON, Imhof A, Groth A (2015) Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes Dev* 29: 585-590
- Alberti S, Saha S, Woodruff JB, Franzmann TM, Wang J, Hyman AA (2018) A User's Guide for Phase Separation Assays with Purified Proteins. *J Mol Biol* 430: 4806-4820
- Alekseyenko AA, Gorchakov AA, Kharchenko PV, Kuroda MI (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
- Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M *et al* (2006) Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* 8: 532-538
- Baker T, Nerle S, Pritchard J, Zhao B, Rivera VM, Garner A, Gonzalez F (2015) Acquisition of a single EZH2 D1 domain mutation confers acquired resistance to EZH2-targeted inhibitors. *Oncotarget* 6: 32646-32655
- Ballaré C, Lange M, Lapinaite A, Martin GM, Morey L, Pascual G, Liefke R, Simon B, Shi Y, Gozani O *et al* (2012) Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat Struct Mol Biol* 19: 1257-1265

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410: 120-124

Batlle E, Clevers H (2017) Cancer stem cells revisited. *Nat Med* 23: 1124-1134

Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS (2018) Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death Differ* 25: 154-160

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

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

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

Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LLP, 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

Blackledge NP, Fursova NA, Kelley JR, Huseyin MK, Feldmann A, Klose RJ (2020) PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol Cell* 77: 857-874.e859

Blighe K, Rana S, Lewis M (2019) EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. *R package version 1*

Boettcher M, McManus MT (2015) Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Mol Cell* 58: 575-585

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

Boucas J, Fritz C, Schmitt A, Riabinska A, Thelen L, Peifer M, Leeser U, Nuernberg P, Altmueller J, Gaestel M *et al* (2015) Label-Free Protein-RNA Interactome Analysis Identifies Khgrp Signaling Downstream of the p38/Mk2 Kinase Complex as a Critical Modulator of Cell Cycle Progression. *PLoS One* 10: e0125745

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

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

Bratkowski M, Yang X, Liu X (2018) An Evolutionarily Conserved Structural Platform for PRC2 Inhibition by a Class of Ezh2 Inhibitors. *Sci Rep* 8: 9092

Brien GL, Bracken AP (2016) The PCL1-p53 axis promotes cellular quiescence. *Cell Cycle* 15: 305-306

Brien GL, Gambero G, O'Connell DJ, Jerman E, Turner SA, Egan CM, Dunne EJ, Jurgens MC, Wynne K, Piao L *et al* (2012) Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat Struct Mol Biol* 19: 1273-1281

Brien GL, Healy E, Jerman E, Conway E, Fadda E, O'Donovan D, Krivtsov AV, Rice AM, Kearney CJ, Flaus A *et al* (2015) A chromatin-independent role of polycomb-like 1 to stabilize p53 and promote cellular quiescence. *Genes and Development* 29: 2231-2243

Brien GL, Remillard D, Shi J, Hemming ML, Chabon J, Wynne K, Dillon ET, Cagney G, Van Mierlo G, Baltissen MP *et al* (2018) Targeted degradation of BRD9 reverses oncogenic gene expression in synovial sarcoma. *Elife* 7

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

Calo E, Wysocka J (2013) Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49: 825-837

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

Cao R, Wang H, He J, Erdjument-Bromage H, Tempst P, Zhang Y (2008) Role of hPHF1 in H3K27 methylation and Hox gene silencing. *Mol Cell Biol* 28: 1862-1872

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

Cao R, Zhang Y (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

Chammas P, Mocavini I, Di Croce L (2020) Engaging chromatin: PRC2 structure meets function. *Br J Cancer* 122: 315-328

Chen S, Jiao L, Liu X, Yang X (2020) A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin. *Mol Cell* 77: 1265-1278.e1267

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

Cheung TH, Rando TA (2013) Molecular regulation of stem cell quiescence. *Nat Rev Mol Cell Biol* 14: 329-340

Choi J, Bachmann AL, Tauscher K, Benda C, Fierz B, Müller 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 GC, Maiolica A, Herzog F, Aebersold R, Nogales E (2012) Molecular architecture of human polycomb repressive complex 2. *eLife* 2012: 1-22

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA *et al* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819-823

Conway E, Healy E, Bracken AP (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 DJ, 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.e408

Cooper S, Grijzenhout A, Underwood E, Ancelin K, Zhang T, Nesterova TB, Anilkirmizitas B, Bassett A, Kooistra SM, Agger K *et al* (2016) Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nature Communications* 7: 1-8

Cowley GS, Weir BA, Vazquez F, Tamayo P, Scott JA, Rusin S, East-Seletsky A, Ali LD, Gerath WF, Pantel SE *et al* (2014) Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci Data* 1: 140035

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA *et al* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* 107: 21931-21936

Cruz-Molina S, Respuela P, Tebartz C, Kolovos P, Nikolic M, Fueyo R, van Ijcken WFJ, Grosveld F, Frommolt P, Bazzi H *et al* (2017) PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* 20: 689-705.e689

Dey A, Chitsaz F, Abbasi A, Misteli T, Ozato K (2003) The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proc Natl Acad Sci U S A* 100: 8758-8763

Doench JG (2018) Am I ready for CRISPR? A user's guide to genetic screens. *Nat Rev Genet* 19: 67-80

Dong C, Nakagawa R, Oyama K, Yamamoto Y, Zhang W, Dong A, Li Y, Yoshimura Y, Kamiya H, Nakayama JI *et al* (2020) Structural basis for histone variant H3tK27me3 recognition by PHF1 and PHF19. *Elife* 9

Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW *et al* (2012) KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* 1: e00205

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

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

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

Finogenova K, Bonnet J, Poepsel S, Schäfer IB, Finkl K, Schmid K, Litz C, Strauss M, Benda C, Müller J (2020) Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. *Elife* 9

Fischer M (2017) Census and evaluation of p53 target genes. *Oncogene* 36: 3943-3956

- Fischer M, Steiner L, Engeland K (2014) The transcription factor p53: not a repressor, solely an activator. *Cell Cycle* 13: 3037-3058
- Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* 17: 1870-1881
- Flaus A (2011) Principles and practice of nucleosome positioning in vitro. *Frontiers in Life Science* 5: 5-27
- Friberg A, Oddone A, Klymenko T, Müller J, Sattler M (2010) Structure of an atypical Tudor domain in the Drosophila Polycomb-like protein. *Protein Sci* 19: 1906-1916
- Fursova NA, Blackledge NP, Nakayama M, Ito S, Koseki Y, Farcas AM, King HW, Koseki H, Klose RJ (2019) Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol Cell* 74: 1020-1036.e1028
- Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, Reinberg D (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* 45: 344-356
- Gatchalian J, Kingsley MC, Moslet SD, Rosas Ospina RD, Kutateladze TG (2015) An aromatic cage is required but not sufficient for binding of Tudor domains of the Polycomb-like protein family to H3K36me3. *Epigenetics* 10: 467-473
- Gibaja V, Shen F, Harari J, Korn J, Ruddy D, Saenz-Vash V, Zhai H, Rejtar T, Paris CG, 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
- Gibson MD, Gatchalian J, Slater A, Kutateladze TG, Poirier MG (2017) PHF1 Tudor and N-terminal domains synergistically target partially unwrapped nucleosomes to increase DNA accessibility. *Nucleic Acids Res* 45: 3767-3776

Giuffrida R, Adamo L, Iannolo G, Vicari L, Giuffrida D, Eramo A, Gulisano M, Memeo L, Conticello C (2016) Resistance of papillary thyroid cancer stem cells to chemotherapy. *Oncol Lett* 12: 687-691

Glancy E, Ciferri C, Bracken AP (2020) Structural basis for PRC2 engagement with chromatin. *Curr Opin Struct Biol* 67: 135-144

Grau D, Zhang Y, Lee CH, Valencia-Sánchez M, Zhang J, Wang M, Holder M, Svetlov V, Tan D, Nudler E *et al* (2021) Structures of monomeric and dimeric PRC2:EZH1 reveal flexible modules involved in chromatin compaction. *Nat Commun* 12: 714

Grijzenhout A, Godwin J, Koseki H, Gdula MR, Szumska D, McGouran JF, Bhattacharya S, Kessler BM, Brockdorff N, 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

Harris SL, Levine AJ (2005) The p53 pathway: positive and negative feedback loops. *Oncogene* 24: 2899-2908

Haupt Y, Maya R, Kazaz A, Oren M (1997) Mdm2 promotes the rapid degradation of p53. *Nature* 387: 296-299

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

He GP, Kim S, Ro HS (1999) Cloning and characterization of a novel zinc finger transcriptional repressor. A direct role of the zinc finger motif in repression. *J Biol Chem* 274: 14678-14684

He J, Kallin EM, Tsukada Y, Zhang Y (2008) The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). *Nat Struct Mol Biol* 15: 1169-1175

Healy E, Mucha M, Glancy E, Fitzpatrick DJ, Conway E, Neikes HK, Monger C, Van Mierlo G, Baltissen MP, Koseki Y *et al* (2019) PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol Cell* 76: 437-452.e436

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576-589

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

Hong Z, Jiang J, Lan L, Nakajima S, Kanno S, Koseki H, Yasui A (2008) A polycomb group protein, PHF1, is involved in the response to DNA double-strand breaks in human cell. *Nucleic Acids Res* 36: 2939-2947

Hrzenjak A (2016) JAZF1/SUZ12 gene fusion in endometrial stromal sarcomas. *Orphanet J Rare Dis* 11: 15

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

Hu CK, Wang W, Brind'Amour J, Singh PP, Reeves GA, Lorincz MC, Alvarado AS, Brunet A (2020) Vertebrate diapause preserves organisms long term through Polycomb complex members. *Science* 367: 870-874

Højfeldt JW, Hedehus L, Laugesen A, Tatar T, Wiehle L, Helin K (2019) Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. *Mol Cell* 76: 423-436.e423

Højfeldt JW, Laugesen A, Willumsen BM, Damhofer H, Hedehus L, Tvardovskiy A, Mohammad F, Jensen ON, Helin K (2018) Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nat Struct Mol Biol* 25: 225-232

Hübner JM, Müller T, Papageorgiou DN, Mauermann M, Krijgsveld J, Russell RB, Ellison DW, Pfister SM, Pajtler KW, Kool M (2019) EZHIP/CXorf67 mimics K27M mutated oncohistones and functions as an intrinsic inhibitor of PRC2 function in aggressive posterior fossa ependymoma. *Neuro Oncol* 21: 878-889

Idogawa M, Ohashi T, Sasaki Y, Maruyama R, Kashima L, Suzuki H, Tokino T (2014) Identification and analysis of large intergenic non-coding RNAs regulated by p53 family members through a genome-wide analysis of p53-binding sites. *Hum Mol Genet* 23: 2847-2857

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* 169: 5429-5433

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

Jain P, Ballare C, Blanco E, Vizán P, Di Croce L (2020) PHF19 mediated regulation of proliferation and invasiveness in prostate cancer cells. *Elife* 9

Jain SU, Do TJ, Lund PJ, Rashoff AQ, Diehl KL, Cieslik M, Bajic A, Juretic N, Deshmukh S, Venneti S *et al* (2019) PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism. *Nat Commun* 10: 2146

Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293: 1074-1080

Jiao L, Liu X (2015) Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science* 350: aac4383

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816-821

Johnson M (2012) Antibody Shelf Life/How to Store Antibodies. *Materials and Methods* 2

Jung M, Philpott M, Müller S, Schulze J, Badock V, Eberspächer U, Moosmayer D, Bader B, Schmees N, Fernández-Montalván A *et al* (2014) Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J Biol Chem* 289: 9304-9319

Kadoch C, Crabtree GR (2013) Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell* 153: 71-85

Kalb R, Latwiel S, Baymaz HI, Jansen PW, Müller CW, Vermeulen M, Müller J (2014) Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* 21: 569-571

Kasinath V, Beck C, Sauer P, Poepsel S, Kosmatka J, Faini M, Toso D, Aebersold R, Nogales E (2021) JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. *Science* 371

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

Katz C, Low-Calle AM, Choe JH, Laptenko O, Tong D, Joseph-Chowdhury JN, Garofalo F, Zhu Y, Friedler A, Prives C (2018) Wild-type and cancer-related p53 proteins are preferentially degraded by MDM2 as dimers rather than tetramers. *Genes Dev* 32: 430-447

Kieffer-Kwon KR, Nimura K, Rao SSP, Xu J, Jung S, Pekowska A, Dose M, Stevens E, Mathe E, Dong P *et al* (2017) Myc Regulates Chromatin Decompaction and Nuclear Architecture during B Cell Activation. *Mol Cell* 67: 566-578.e510

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

Kim TG, Chen J, Sadoshima J, Lee Y (2004) Jumonji represses atrial natriuretic factor gene expression by inhibiting transcriptional activities of cardiac transcription factors. *Mol Cell Biol* 24: 10151-10160

Kim TW, Kang BH, Jang H, Kwak S, Shin J, Kim H, Lee SE, Lee SM, Lee JH, Kim JH *et al* (2015) Ctbp2 Modulates NuRD-Mediated Deacetylation of H3K27 and Facilitates PRC2-Mediated H3K27me3 in Active Embryonic Stem Cell Genes During Exit from Pluripotency. *Stem Cells* 33: 2442-2455

Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, Ghosh D, Sewalt RG, Otte AP, Hayes DF *et al* (2003) EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* 100: 11606-11611

Kloet SL, Makowski MM, Baymaz HI, van Voorthuijsen L, Karemaker ID, Santanach A, Jansen PWTC, Di Croce L, Vermeulen M (2016) The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol* 23: 682-690

Knutson SK, Warholic NM, Wigle TJ, Klaus CR, Allain CJ, Raimondi A, Porter Scott M, Chesworth R, Moyer MP, Copeland RA *et al* (2013) Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc Natl Acad Sci U S A* 110: 7922-7927

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

Kruse JP, Gu W (2009) Modes of p53 regulation. *Cell* 137: 609-622

Kycia I, Kudithipudi S, Tamas R, Kungulovski G, Dhayalan A, Jeltsch A (2014) The Tudor domain of the PHD finger protein 1 is a dual reader of lysine trimethylation at lysine 36 of histone H3 and lysine 27 of histone variant H3t. *J Mol Biol* 426: 1651-1660

Lamb KN, Bsteh D, Dishman SN, Moussa HF, Fan H, Stuckey JI, Norris JL, Cholensky SH, Li D, Wang J *et al* (2019) Discovery and Characterization of a

Cellular Potent Positive Allosteric Modulator of the Polycomb Repressive Complex 1 Chromodomain, CBX7. *Cell Chem Biol* 26: 1365-1379.e1322

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

Lavarone E, Barbieri CM, Pasini D (2019) Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. *Nat Commun* 10: 1679

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

Lee HG, Kahn TG, Simcox A, Schwartz YB, Pirrotta V (2015) Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome Res* 25: 1170-1181

Levine AJ, Oren M (2009) The first 30 years of p53: growing ever more complex. *Nat Rev Cancer* 9: 749-758

Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276: 565-570

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

Liao Y, Smyth GK, 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, Shi Y (2016) EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape. *Molecular Cell* 64: 659-672

Liefke R, Shi Y (2015) The PRC2-associated factor C17orf96 is a novel CpG island regulator in mouse ES cells. *Cell Discov* 1: 15008

Liu R, Gao J, Yang Y, Qiu R, Zheng Y, Huang W, Zeng Y, Hou Y, Wang S, Leng S *et al* (2018) PHD finger protein 1 (PHF1) is a novel reader for histone H4R3 symmetric dimethylation and coordinates with PRMT5-WDR77/CRL4B complex to promote tumorigenesis. *Nucleic Acids Res* 46: 6608-6626

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

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

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389: 251-260

Ma X, Wang J, Ma CX, Gao X, Patriub V, Sklar JL (2017) The JAZF1-SUZ12 fusion protein disrupts PRC2 complexes and impairs chromatin repression during human endometrial stromal tumorigenesis. *Oncotarget* 8: 4062-4078

Maier VK, Feeney CM, Taylor JE, Creech AL, Qiao JW, Szanto A, Das PP, Chevrier N, Cifuentes-Rojas C, Orkin SH *et al* (2015) Functional Proteomic Analysis of Repressive Histone Methyltransferase Complexes Reveals ZNF518B as a G9A Regulator. *Mol Cell Proteomics* 14: 1435-1446

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

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

Margueron R, Reinberg D (2010) Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* 11: 285-296

Masudo K, Suganuma N, Nakayama H, Oshima T, Rino Y, Iwasaki H, Matsuzu K, Sugino K, Ito K, Kondo T *et al* (2018) EZH2 Overexpression as a Useful Prognostic Marker for Aggressive Behaviour in Thyroid Cancer. *In Vivo* 32: 25-31

McCabe MT, Graves AP, Ganji G, Diaz E, Halsey WS, Jiang Y, Smitheman KN, Ott HM, Pappalardi MB, Allen KE *et al* (2012) 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

McConnell AM, Yao C, Yeckes AR, Wang Y, Selvaggio AS, Tang J, Kirsch DG, Stripp BR (2016) p53 Regulates Progenitor Cell Quiescence and Differentiation in the Airway. *Cell Rep* 17: 2173-2182

McDonald ER, de Weck A, Schlabach MR, Billy E, Mavrakis KJ, Hoffman GR, Belur D, Castelletti D, Frias E, Gampa K *et al* (2017) Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* 170: 577-592.e510

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

McKnight JN, Boerma JW, Breeden LL, Tsukiyama T (2015) Global Promoter Targeting of a Conserved Lysine Deacetylase for Transcriptional Shutoff during Quiescence Entry. *Mol Cell* 59: 732-743

Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia NV, Montgomery PG, Cowley GS, Pantel S *et al* (2017) Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* 49: 1779-1784

Michel BC, D'Avino AR, Cassel SH, Mashtalir N, McKenzie ZM, McBride MJ, Valencia AM, Zhou Q, Bocker M, Soares LMM *et al* (2018) A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol* 20: 1410-1420

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP *et al* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553-560

Milner J, Medcalf EA (1991) Cotranslation of activated mutant p53 with wild type drives the wild-type p53 protein into the mutant conformation. *Cell* 65: 765-774

Min J, Zhang Y, Xu RM (2003) Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* 17: 1823-1828

Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schübeler D (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* 30: 755-766

Moll UM, Petrenko O (2003) The MDM2-p53 Interaction. *Molecular Cancer Research* 1: 1001-1008

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

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

Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, Kuchenbauer F *et al* (2010) Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* 42: 181-185

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

Murashima A, Shinjo K, Katsushima K, Onuki T, Kondoh Y, Osada H, Kagaya N, Shin-Ya K, Kimura H, Yoshida M *et al* (2019) Identification of a chemical modulator

of EZH2-mediated silencing by cell-based high-throughput screening assay. *J Biochem* 166: 41-50

Musselman CA, Avvakumov N, Watanabe R, Abraham CG, Lalonde M-E, Hong Z, Allen C, Roy S, Nuñez JK, Nickoloff J *et al* (2012) Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nature Structural & Molecular Biology* 19: 1266-1272

Musselman CA, Mansfield RE, Garske AL, Davrazou F, Kwan AH, Oliver SS, O'Leary H, Denu JM, Mackay JP, Kutateladze TG (2009) Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem J* 423: 179-187

Müller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA (2002) Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* 111: 197-208

Nguyen TT, Grimm SA, Bushel PR, Li J, Li Y, Bennett BD, Lavender CA, Ward JM, Fargo DC, Anderson CW *et al* (2018) Revealing a human p53 universe. *Nucleic Acids Res* 46: 8153-8167

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

O'Connell S, Wang L, Robert S, Jones CA, Saint R, Jones RS (2001) Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein. *J Biol Chem* 276: 43065-43073

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

Olins AL, Olins DE (1974) Spheroid chromatin units (v bodies). *Science* 183: 330-332

Oliviero G, Brien GL, Waston A, Streubel G, Jerman E, Andrews D, Doyle B, Munawar N, Wynne K, Crean J *et al* (2016) Dynamic Protein Interactions of the Polycomb Repressive Complex 2 during Differentiation of Pluripotent Cells. *Molecular & cellular proteomics : MCP* 15: 3450-3460

Pajtler KW, Wen J, Sill M, Lin T, Orisme W, Tang B, Hübner JM, Ramaswamy V, Jia S, Dalton JD *et al* (2018) Molecular heterogeneity and CXorf67 alterations in posterior fossa group A (PFA) ependymomas. *Acta Neuropathol* 136: 211-226

Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R, Thomson JA (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1: 299-312

Pandya P, Bailey B, Elmi AE, Bates H, Hemenway CN, Sinn A, Bijangi-Vishehsaraei K, Saadatzadeh M, Shannon H, Ding J *et al* (2018) Abstract 3180: Preclinical validation of EZH2 as a therapeutic target in pediatric Ewing's sarcoma. *Cancer Research* 78: 3180-3180

Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K (2007) The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* 27: 3769-3779

Pasini D, Bracken AP, Jensen MR, Lazzerini Denchi E, Helin K (2004) Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J* 23: 4061-4071

Perino M, van Mierlo G, Karemaker ID, van Genesen S, Vermeulen M, Marks H, van Heeringen SJ, Veenstra GJC (2018) MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat Genet* 50: 1002-1010

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11: 1650-1667

Perucca P, Cazzalini O, Madine M, Savio M, Laskey RA, Vannini V, Prosperi E, Stivala LA (2009) Loss of p21 CDKN1A impairs entry to quiescence and activates

a DNA damage response in normal fibroblasts induced to quiescence. *Cell Cycle* 8: 105-114

Pitolli C, Wang Y, Mancini M, Shi Y, Melino G, Amelio I (2019) Do Mutations Turn p53 into an Oncogene? *Int J Mol Sci* 20

Piunti A, Smith ER, Morgan MAJ, Ugarenko M, Khaltayan N, Helmin KA, Ryan CA, Murray DC, Rickels RA, Yilmaz BD *et al* (2019) CATACOMB: An endogenous inducible gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. *Sci Adv* 5: eaax2887

Plass C, Pfister SM, Lindroth AM, Bogatyrova O, Claus R, Lichter P (2013) Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat Rev Genet* 14: 765-780

Potjewyd F, Turner AW, Beri J, Rectenwald JM, Norris-Drouin JL, Cholensky SH, Margolis DM, Pearce KH, Herring LE, James LI (2020) Degradation of Polycomb Repressive Complex 2 with an EED-Targeted Bivalent Chemical Degradar. *Cell Chem Biol* 27: 47-56.e15

Qiu Y, Liu L, Zhao C, Han C, Li F, Zhang J, Wang Y, Li G, Mei Y, Wu M *et al* (2012) Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription. *Genes Dev* 26: 1376-1391

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-842

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470: 279-283

Ragazzini R, Pérez-Palacios R, Baymaz IH, Diop S, Ancelin K, Zielinski D, Michaud A, Givelet M, Borsos M, Aflaki S *et al* (2019) EZHIP constrains Polycomb Repressive Complex 2 activity in germ cells. *Nat Commun* 10: 3858

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42: W187-191

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44: W160-165

Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2: 1896-1906

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

Richter GH, Plehm S, Fasan A, Rössler S, Unland R, Bennani-Baiti IM, Hotfilder M, Löwel D, von Lüttichau I, Mossbrugger I *et al* (2009) EZH2 is a mediator of EWS/FLI1 driven tumor growth and metastasis blocking endothelial and neuro-ectodermal differentiation. *Proc Natl Acad Sci U S A* 106: 5324-5329

Sammons MA, Zhu J, Drake AM, Berger SL (2015) TP53 engagement with the genome occurs in distinct local chromatin environments via pioneer factor activity. *Genome Res* 25: 179-188

Sarma K, Margueron R, Ivanov A, Pirrotta V, Reinberg D (2008) Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27 Trimethylation In Vivo. *Molecular and Cellular Biology* 28: 2718-2731

Scelfo A, Fernández-Pérez D, Tamburri S, Zanotti M, Lavarone E, Soldi M, Bonaldi T, Ferrari KJ, Pasini D (2019) Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities. *Mol Cell* 74: 1037-1052.e1037

Schafer KA (1998) The cell cycle: a review. *Vet Pathol* 35: 461-478

- Schuettengruber B, Bourbon HM, Di Croce L, Cavalli G (2017) Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* 171: 34-57
- Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, Heckl D, Ebert BL, Root DE, Doench JG *et al* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343: 84-87
- Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE (1999) Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98: 37-46
- Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR (2015) Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* 33: 661-667
- Simon J, Chiang A, Bender W, Shimell MJ, O'Connor M (1993) Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products. *Dev Biol* 158: 131-144
- Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. *Nat Rev Genet* 14: 204-220
- Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* 41: e28
- Sneeringer CJ, Scott MP, Kuntz KW, Knutson SK, Pollock RM, Richon VM, Copeland RA (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 SS, Margueron R, Reinberg D (2013) Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev* 27: 2663-2677

Suh JL, Bsteh D, Si Y, Hart B, Weaver TM, Pribitzer C, Lau R, Soni S, Ogana H, Rectenwald JM *et al* (2021) REPROGRAMMING CBX8-PRC1 FUNCTION WITH A POSITIVE ALLOSTERIC MODULATOR. *bioRxiv*: 2021.2002.2023.432388

Swygert SG, Kim S, Wu X, Fu T, Hsieh TH, Rando OJ, Eisenman RN, Shendure J, McKnight JN, Tsukiyama T (2019) Condensin-Dependent Chromatin Compaction Represses Transcription Globally during Quiescence. *Mol Cell* 73: 533-546.e534

Taherbhoy AM, Huang OW, Cochran AG (2015) BMI1-RING1B is an autoinhibited RING E3 ubiquitin ligase. *Nat Commun* 6: 7621

Tamburri S, Lavarone E, Fernández-Pérez D, Conway E, Zanotti M, Manganaro D, Pasini D (2020) Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Mol Cell* 77: 840-856.e845

Tasdemir N, Banito A, Roe JS, Alonso-Curbelo D, Camiolo M, Tschaharganeh DF, Huang CH, Aksoy O, Bolden JE, Chen CC *et al* (2016) BRD4 Connects Enhancer Remodeling to Senescence Immune Surveillance. *Cancer Discov* 6: 612-629

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

Torrance CJ, Agrawal V, Vogelstein B, Kinzler KW (2001) Use of isogenic human cancer cells for high-throughput screening and drug discovery. *Nat Biotechnol* 19: 940-945

Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, Gill S, Harrington WF, Pantel S, Krill-Burger JM *et al* (2017) Defining a Cancer Dependency Map. *Cell* 170: 564-576.e516

Turcotte S, Chan DA, Sutphin PD, Hay MP, Denny WA, Giaccia AJ (2008) A molecule targeting VHL-deficient renal cell carcinoma that induces autophagy. *Cancer Cell* 14: 90-102

- Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP *et al* (2002) The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 419: 624-629
- Vella P, Barozzi I, Cuomo A, Bonaldi T, Pasini D (2012) Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res* 40: 3403-3418
- Vermeulen K, Van Bockstaele DR, Berneman ZN (2003) The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif* 36: 131-149
- Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW (2013) Cancer genome landscapes. *Science* 339: 1546-1558
- Voigt P, Tee WW, Reinberg D (2013) A double take on bivalent promoters. *Genes Dev* 27: 1318-1338
- Völkel P, Bary A, Raby L, Chapart A, Dupret B, Le Bourhis X, Angrand PO (2019) Ezh1 arises from Ezh2 gene duplication but its function is not required for zebrafish development. *Sci Rep* 9: 4319
- Walker E, Chang WY, Hunkapiller J, Cagney G, Garcha K, Torchia J, Krogan NJ, Reiter JF, Stanford WL (2010) Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* 6: 153-166
- Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y (2004a) Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431: 873-878
- Wang L, Brown JL, Cao R, Zhang Y, Kassiss JA, Jones RS (2004b) Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* 14: 637-646
- Wang T, Wei JJ, Sabatini DM, Lander ES (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343: 80-84

Wang X, Paucek RD, Gooding AR, Brown ZZ, Ge EJ, Muir TW, Cech TR (2017) Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat Struct Mol Biol* 24: 1028-1038

Wang Z, Dai J, Yan J, Zhang Y, Yin Z (2019) Targeting EZH2 as a novel therapeutic strategy for sorafenib-resistant thyroid carcinoma. *J Cell Mol Med* 23: 4770-4778

Wei F, Hao P, Zhang X, Hu H, Jiang D, Yin A, Wen L, Zheng L, He JZ, Mei W *et al* (2018) Etoposide-induced DNA damage affects multiple cellular pathways in addition to DNA damage response. *Oncotarget* 9: 24122-24139

Weinberg RL, Veprintsev DB, Fersht AR (2004) Cooperative binding of tetrameric p53 to DNA. *J Mol Biol* 341: 1145-1159

Willcockson MA, Healton SE, Weiss CN, Bartholdy BA, Botbol Y, Mishra LN, Sidhwani DS, Wilson TJ, Pinto HB, Maron MI *et al* (2021) H1 histones control the epigenetic landscape by local chromatin compaction. *Nature* 589: 293-298

Williams SP, Barthorpe AS, Lightfoot H, Garnett MJ, McDermott U (2017) High-throughput RNAi screen for essential genes and drug synergistic combinations in colorectal cancer. *Sci Data* 4: 170139

Wu X, Johansen JV, 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 JH, Nguyen M, Fujiwara Y, Peng C, Orkin SH (2014) Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* 14: 68-80

Xu C, Bian C, Yang W, Galka M, Ouyang H, Chen C, Qiu W, Liu H, Jones AE, MacKenzie F *et al* (2010) Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc Natl Acad Sci U S A* 107: 19266-19271

- Xu J, Shao Z, Li D, Xie H, Kim W, Huang J, Taylor JE, Pinello L, Glass K, Jaffe JD *et al* (2015) Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol Cell* 57: 304-316
- Xue L, Yan H, Chen Y, Zhang Q, Xie X, Ding X, Wang X, Qian Z, Xiao F, Song Z *et al* (2019) EZH2 upregulation by ER α induces proliferation and migration of papillary thyroid carcinoma. *BMC Cancer* 19: 1094
- Yang XJ, Seto E (2007) HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene* 26: 5310-5318
- 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
- Yap DB, Chu J, Berg T, Schapira M, Cheng SW, Moradian A, Morin RD, Mungall AJ, Meissner B, Boyle M *et al* (2011) Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* 117: 2451-2459
- Youmans DT, Schmidt JC, Cech TR (2018) Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits. *Genes Dev* 32: 794-805
- Young CP, Hillyer C, Hokamp K, Fitzpatrick DJ, Konstantinov NK, Welty JS, Ness SA, Werner-Washburne M, Fleming AB, Osley MA (2017) Distinct histone methylation and transcription profiles are established during the development of cellular quiescence in yeast. *BMC Genomics* 18: 107
- Yusufova N, Klotgen A, Teater M, Osunsade A, Camarillo JM, Chin CR, Doane AS, Venters BJ, Portillo-Ledesma S, Conway J *et al* (2021) Histone H1 loss drives lymphoma by disrupting 3D chromatin architecture. *Nature* 589: 299-305
- Zeng L, Zhang Q, Li S, Plotnikov AN, Walsh MJ, Zhou MM (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* 466: 258-262

Zhang Q, Agius SC, Flanigan SF, Levina V, Owen BM, Davidovich C (2020) Convergent evolution between PALI1 and JARID2 for the allosteric activation of PRC2. *bioRxiv*: 2020.2005.2028.122556

Zhang Q, McKenzie NJ, Warneford-Thomson R, Gail EH, Flanigan SF, Owen BM, Lauman R, Levina V, Garcia BA, Schittenhelm RB *et al* (2019) RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nat Struct Mol Biol* 26: 237-247

Zhang Z, Jones A, Sun CW, Li C, Chang CW, Joo HY, Dai Q, Mysliwiec MR, Wu LC, 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 XD, Han X, Chew JL, Liu J, Chiu KP, Choo A, Orlov YL, Sung WK, Shahab A, Kuznetsov VA *et al* (2007) Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* 1: 286-298

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10: 1523