



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

Investigating the effects of H3K27M on PRC1 in Diffuse Midline Glioma

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

2025

Eimear Lagan

Thesis supervisors:

Prof Adrian Bracken & Dr Gerard Brien

Chromatin Biology Laboratory
The Smurfit Institute of Genetics
Trinity College Dublin

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.

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.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

A handwritten signature in black ink, appearing to read 'Eimear Lagan', is written over a horizontal line.

Eimear Lagan, April 2025

Summary

Diffuse midline glioma (DMG) is a highly aggressive and lethal paediatric brain tumour characterised by a mutation in histone H3 (H3K27M). This mutation leads to a global reduction in the repressive histone modification H3K27me3. However, a subset of neurodevelopmental genes retains H3K27me3, keeping them aberrantly repressed. This repression is crucial for restraining differentiation and keeping DMG cells in a neural stem cell-like state, thereby promoting their proliferative potential. It is known that some forms of Polycomb Repressive Complex 1 (PRC1), known as canonical PRC1 (cPRC1), contribute to gene repression by binding to H3K27me3. Here, I demonstrate that DMG cells exhibit a specific dependence on a form of cPRC1 containing CBX4 and PCGF4 (but not their paralogs). Using a combination of CRISPR-Cas9 screens, quantitative chromatin mapping techniques, and functional perturbations, I demonstrate that cPRC1 complexes containing CBX4 are essential in DMG disease biology, acting downstream of H3K27me3. I demonstrate that the altered H3K27me3 landscape in DMG cells leads to a reorganization of cPRC1 complex distribution on chromatin, highlighting previously unknown differences between CBX paralogs. The work presented here not only advances our understanding of cPRC1 complexes, but also provides new insights into the molecular mechanisms underlying DMG disease and highlights the potential therapeutic value of targeting cPRC1 complexes in the treatment of this devastating paediatric cancer.

Publications & Manuscripts

- 1) Gerard L Brien, Raul Bardini Bressan, Craig Monger, Dáire Gannon, **Eimear Lagan**, Anthony M Doherty, Evan Healy, Hannah Neikes, Darren J Fitzpatrick, Orla Deevy, Vivien Grant, Maria-Angeles Marqués-Torrejón, Neza Alfazema, Steven M Pollard, Adrian P Bracken. Simultaneous disruption of PRC2 and enhancer function underlies histone H3.3-K27M oncogenic activity in human hindbrain neural stem cells. *Nature Genetics*, 53, 1221–1232 (2021). <https://doi.org/10.1038/s41588-021-00897-w>
- 2) Mimmi LE Lundahl, Morgane Mitermite, Dylan Gerard Ryan, Sarah Case, Niamh C Williams, Ming Yang, Roisin I Lynch, **Eimear Lagan**, Filipa M Lebre, Aoife L Gorman, Bojan Stojkovic, Adrian P Bracken, Christian Frezza, Frederick J Sheedy, Eoin M Scanlan, Luke AJ O'Neill, Stephen V Gordon, Ed C Lavelle. Macrophage innate training induced by IL-4 and IL-13 activation enhances OXPHOS driven anti-mycobacterial responses. *eLife*, 11, e74690 (2022). <https://doi.org/10.7554/eLife.74690>
- 3) **Eimear Lagan***, Daire Gannon*, Ademar Jesus Silva, Peter Bibawi, Anthony Doherty, Darragh Nimmo, Rachel McCole, Craig Monger, Giovanni Genesi, Aurelie Vanderlinden Dibekeme, James Innes, Lu Yang, Bryan Chen, Guido van Mierlo, Pascal WTC Jansen, Kieran Wynne, Francisco J Sanchez Rivera, Yadira Soto-Feliciano, Michiel Vermeulen, Giorgio Oliviero, Chun Wei Chen, Richard Phillips[†], Adrian Patrick Bracken[†], Gerard Brien[†]. *Manuscript in review. bioRxiv* (2023). <https://doi.org/10.1101/2023.12.02.569451>

* Authors contributed equally

† Co-senior authors

Awards

Irish Research Council Government of Ireland Postgraduate Scholarship Programme Awardee (GOIPG/2021/870).

Acknowledgments

First and foremost, I would like to thank my supervisors Professor Adrian Bracken and Dr Gerard Brien for their unwavering support and guidance over the years. Undertaking a research project under their supervision as an undergraduate student sparked in me what will no doubt be a life-long interest in chromatin biology and cancer epigenetics. I cannot thank them enough for seeing potential in me and providing a nurturing environment for me to grow. In particular, thank you to Dr Gerard Brien for always having the time and patience to provide training and discuss the details of experiments and ideas, big or small.

Thank you to all the members of the Bracken & Brien labs for their good humour, friendship, collaboration and support. Thank you to Dr Dáire Gannon for taking me under his wing on the DMG project, for his patience and positive attitude, and for guiding me through the genome-wide CRISPR screen. Thank you to AJ Silva for his trojan work on the DMG project, his hilarious sense of humour and for being a ray of sunshine in the lab. Thank you to Dr Anthony Doherty for providing most of the bioinformatic analyses for my experiments, and for being a rock-solid support both professionally and personally – your friendship is treasured. Thank you to Dr Ellen Tuck and Dr Francesca Matrá for your kindness, support, and friendship. Thank you to Dr Daniel Angelov for your calming influence and for sending photos of your beautiful baby boy when morale needed a boost. Thank you to Dr Andrew Malcolm and Charlie Jones for being wonderful people and colleagues in the Brien lab. You are all incredible scientists that I have been privileged to work with.

Thank you to my father Terry - your love and support means the world; I am incredibly lucky to be your daughter, and I cannot express how much I appreciate all that you do for me. Thank you to my brother Conor and my sister Aoife, who I love dearly. Huge thank you to Kejan Lim for his unwavering patience, understanding, love, and emotional support provided over the years. Thank you to Peng and Kim Lim for always treating me like part of the family and providing a safe and loving home environment whenever I needed respite. Thank you to my oldest and dearest friends; Therese Friis, Eoin Dowling, Molly Wade, Saoirse Reynolds & Aoife O'Donoghue for all the laughs, chats, and precious time spent together over the years. I am truly blessed to have you all in my life.

Finally, thank you to the Irish Research Council (GOIPG/2021/870) for making this research possible by funding this project.

Go raibh míle maith agaibh go léir.

TABLE OF CONTENTS

<i>Declaration</i>	<i>i</i>
<i>Summary</i>	<i>ii</i>
<i>Publications & Manuscripts, Awards</i>	<i>iii</i>
<i>Acknowledgements</i>	<i>iv</i>
<i>Table of Contents</i>	<i>v</i>
<i>List of Figures</i>	<i>viii</i>
<i>List of Abbreviations</i>	<i>x</i>
1. CHAPTER 1: Introduction	1
1.1. Distinct gene expression programmes give rise to diverse cellular identities.....	2
1.2. The organisation of eukaryotic DNA into chromatin regulates gene expression.....	3
1.3. Histone post-translational modifications can influence chromatin structure and gene expression.....	4
1.4. Polycomb group protein complexes play an important role in development and cellular identity maintenance.....	7
1.5. PRC1 complex structures and functions.....	8
1.6. PRC2 complex structures and functions.....	11
1.7. PRC1 and PRC2 cooperate to maintain target gene repression.....	13
1.8. Polycomb group proteins are frequently dysregulated in cancer.....	15
1.9. Diffuse Midline Glioma (H3K27-altered).....	17
1.10 H3K27M induces a distinctive chromatin landscape which restrains differentiation.....	19
1.11 PRC2 is a promising therapeutic target in DMG (H3K27-altered).....	21
2. CHAPTER 2: Materials & Methods	24
2.1 Cell culture.....	25
2.1.1. Cell line culture.....	25
2.1.2 Lentivirus production.....	25
2.2 Protein methods.....	26
2.2.1 Preparation of whole cell protein lysates & immunoblotting.....	26
2.3 CRISPR-Cas9 screening and validation.....	26
2.3.1 Genome-wide Brunello CRISPR-Cas9 knock-out screens.....	26
2.3.2 CBX4 CRISPR-Cas9 tiling screens.....	26
2.3.3 CRISPR-Cas9 screen validations – growth competition assays.....	27
2.4 Genomics & transcriptomics.....	27
2.4.1 CUT&RUN-Rx.....	27
2.4.2 CUT&RUN-Rx sequencing library preparation.....	28
2.4.3 Quant-seq.....	28
2.4.4 Quantitative PCR.....	28

2.5 Data analysis.....	29
2.5.1 Genome-wide CRISPR-Cas9 knock-out screen data analysis.....	29
2.5.2 CBX4 tiling screen data analysis.....	29
2.5.3 CUT&RUN-Rx data analysis.....	29
2.5.4 Quant-seq data analysis.....	29
3. CHAPTER 3: <i>CBX4</i> and <i>PCGF4</i> are genetic dependencies in H3K27M-DMG.....	31
3.1. Introduction.....	32
3.2. Results.....	33
3.2.1 Genome-wide CRISPR-Cas9 knock-out screens validate core PRC2 members as genetic dependencies in H3K27M-DMG.....	33
3.2.2 Genome-wide CRISPR-Cas9 knock-out screens reveal specific canonical PRC1 subunits CBX4 and PCGF4 are genetic dependencies in H3K27M-DMG.....	36
3.2.3 Growth competition assays validate CBX4 and PCGF4 as functionally essential to DMG.....	40
3.3. Discussion.....	43
4. CHAPTER 4: H3K27M redistributes cPRC1 on chromatin.....	46
4.1. Introduction.....	47
4.2. Results.....	47
4.2.1 Characterisation of isogenic H3K27M and H3K27M-KO cell lines.....	47
4.2.2 The H3K27M does not localise to Polycomb target sites.....	51
4.2.3 H3K27M drastically reorganises cPRC1 chromatin occupancy.....	54
4.3. Discussion.....	61
5. CHAPTER 5: Identification of functionally essential regions of CBX4 in H3K27M-DMG.....	64
5.1. Introduction.....	65
5.2. Results.....	65
5.2.1 CRISPR-Cas9 knock-out tiling screen identifies functionally essential domains in CBX4.....	65
5.2.2 Rescue experiments validate the CBX4 chromodomain as functionally essential for H3K27M-DMG disease biology.....	68
5.2.3 The CBX chromodomains are highly similar but are not equally sensitive to H3K27me3 loss.....	71
5.2.4 The CBX4 internal disordered region is uniquely predicted to contact core cPRC1..	80
5.3. Discussion.....	83
6. CHAPTER 6: General Discussion.....	85
6.1. Introduction.....	86

6.1.1 Understanding the molecular mechanisms driving disease biology is important for the approval of effective oncology treatments.....	86
6.2 Functional genomic screens identify CBX4 as a mechanistically-anchored potential therapeutic target for H3K27M-DMG.....	87
6.3. Accumulating evidence for non-redundant roles of cPRC1 complexes.....	90
6.4. Understanding the molecularly distinct functions of different CBX paralogs.....	92
6.5. Future directions.....	95
6.6. Conclusions.....	96
APPENDICES.....	97
Appendix I.....	98
Appendix II.....	99
Appendix III.....	99
Appendix IV.....	100
Appendix V.....	101
Appendix VI.....	102
Appendix VII.....	102
Appendix VIII.....	102
REFERENCES.....	103

LIST OF FIGURES

Chapter 1

Figure 1.1 Waddington's landscape depicting differentiation during development.

Figure 1.2 Packaging of DNA in the eukaryotic cell nucleus.

Figure 1.3 Histone "readers", "writers", and "erasers".

Figure 1.4 Mammalian canonical PRC1 complex structure.

Figure 1.5 Mammalian variant PRC1 complex structure.

Figure 1.6 Mammalian PRC2 complex structure.

Figure 1.7 PRC1/PRC2 recruitment and interplay.

Figure 1.8 Diffuse midline glioma (H3K27-altered)

Figure 1.9 Dysregulation of PRC2 activity by H3K27M restrains differentiation

Chapter 3

Figure 3.1 Genome-wide CRISPR-Cas9 knock-out screen confirms PRC2 catalytic core components are genetic dependency in H3K27M DMG.

Figure 3.2 PRC2 catalytic core components rank amongst the top hits in the genome-wide CRISPR-Cas9 screen.

Figure 3.3 Genome-wide CRISPR-Cas9 knock-out screen indicates vPRC1 components are not genetic dependencies in H3K27M DMG.

Figure 3.4 Genome-wide CRISPR-Cas9 knock-out screen indicates specific cPRC1 components CBX4 and PCGF4 are genetic dependencies in H3K27M DMG.

Figure 3.5 cPRC1 components CBX4 and PCGF4 rank amongst the top hits in the genome-wide CRISPR-Cas9 screen.

Figure 3.6 Immunoblots demonstrate Brunello sgRNA sequences have on-target efficacy.

Figure 3.7 Growth competition assays confirms *CBX4* and *PCGF4* are genetic dependencies in H3K27M DMG.

Chapter 4

Figure 4.1. Knock out of H3K27M in SU-DIPG-XIII.

Figure 4.2. Differential gene expression in H3K27M-KO cells.

Figure 4.3. Differential expression of Polycomb target genes in H3K27M-KO cells.

Figure 4.4. Capturing chromatin dynamics with quantitative CUT&RUN-Rx.

Figure 4.5. H3K27M localises mainly to actively expressed genes.

Figure 4.6 CUT&RUN-Rx shows PRC2 and PRC1 complexes accumulate on chromatin in H3K27M cells.

Figure 4.7 CUT&RUN-Rx shows differential behaviour of cPRC1 complexes on chromatin in H3K27M cells.

Figure 4.8 CUT&RUN-Rx shows vPRC1 complexes are agnostic to changes in H3K27me3 distribution.

Figure 4.9 Whole-cell cPRC1 protein subunit abundances in isogenic H3K27M and H3K27M-KO cell lines.

Chapter 5

Figure 5.1 *CBX4* CRISPR-Cas9 knock-out tiling screen sgRNA library.

Figure 5.2 *CBX4* CRISPR-Cas9 knock-out tiling screens workflow and quality control.

Figure 5.3 *CBX4* CRISPR-Cas9 knock-out tiling screens reveal the chromodomain and a section of the internal disordered region are functionally essential.

Figure 5.4 Validation of *CBX4* tiling screens: *CBX4* rescue experiment design.

Figure 5.5 *CBX4* rescue experiment confirm the chromodomain is functionally essential.

Figure 5.6 Comparison of CBX protein paralogs.

Figure 5.7 EZH2 inhibition with Tazemetostat induces transcriptional changes similar to H3K27M knock-out.

Figure 5.8 EZH2 inhibition with Tazemetostat reverses H3K27M-induced aberrant repression of Polycomb target genes.

Figure 5.9 Mapping the consequences of Tazemetostat treatment on PRC2 and cPRC1 chromatin occupancy with CUT&RUN-Rx.

Figure 5.9. CUT&RUN-Rx shows Tazemetostat treatment displaces all cPRC1 complexes from chromatin.

Figure 5.10 CUT&RUN-Rx shows displacement of cPRC1 complexes from correlated with loss of H3K27me3 and Polycomb gene derepression.

Figure 5.11 Polycomb target gene derepression correlates with cPRC1 displacement.

Figure 5.12 *CBX4* internal disordered region is predicted to have an extra contact point with core cPRC1.

Figure 5.14 Only *CBX4* is consistently predicted to contact core cPRC1.

LIST OF ABBREVIATIONS

AEBP2	adipocyte enhancer-binding protein 2
bp	base pairs
Cas9	CRISPR associated protein 9
CBX	chromobox protein homolog
CCNA2	cyclin A2
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
cPRC1	canonical Polycomb repressive complex 1
CRISPR	clustered regularly interspaced short palindromic repeats
cryo-EM	Cryogenic electron microscopy
CST	Cell Signaling Technology
CUT&RUN	Cleavage Under Targets & Release Using Nuclease
DIPG	diffuse intrinsic pontine glioma
DMG	diffuse midline glioma
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EED	embryonic ectoderm development protein
EPOP	elongin BC and Polycomb repressive complex 2-associated protein
EZH1/2	enhancer of zeste homolog 1/2
EZH1P	EZH inhibitory protein
fNSC	foetal neural stem cell
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GBM	glioblastoma multiforme
gDNA	genomic DNA
GFP	green fluorescent protein
H2A	histone H2A
H2AK119ub	histone H2A mono-ubiquitinated at lysine 119
H2B	histone H2B
H3	histone H3
H3K27M	histone H3 amino acid 27 lysine-to-methionine
H3F3A	H3.3 histone A
H3F3B	H3.3 histone B
H3K27me1/2/3	histone H3 lysine 27 mono-/di-/tri-methylation
HEK293T	Human Embryonic Kidney 293T cells (with SV40 T antigen)
Hox	homeotic
IgG	immunoglobulin G
indels	insertion or deletion mutations
IP	immunoprecipitation
JARID2	jumonji and AT-rich interaction domain containing 2
KDM2B	lysine demethylase 2B
KO	knockout
me1/2/3	mono-, di- and tri-methylation
mESC	mouse embryonic stem cell
mRNA	messenger RNA
MTF2	metal response element binding transcription factor 2
NGS	next-generation sequencing
NSC	neural stem cell
PAL1/2	PRC2 associated LCOR isoform 1/2
PAM	protospacer adjacent motif
PcG	Polycomb group
PCGF	Polycomb group ring finger
PCGF2	Polycomb group ring finger 2 (MEL18)
PCGF4	Polycomb group ring finger 4 (BMI1)
PCL	Polycomb-like protein
PCR	polymerase chain reaction
PDGFRA	platelet derived growth factor receptor alpha
PHC	polyhomeotic homolog
PRC1	Polycomb repressive complex 1

PRC2	Polycomb repressive complex 2
qPCR	quantitative PCR
RBBP4/7	histone-binding protein RBB4/7
RING	ring finger protein
RNA	ribonucleic acid
RPA3	replication protein A3
RYBP	RING1 and YY1 binding protein
SCMH1	Scm Polycomb group protein homolog
SCML1/2	Scm Polycomb group protein like 1/2
sgRNA	single-guide RNA
shRNA	short hairpin RNA
SUMO	small ubiquitin-like molecule
SUV39H1	SUV39H1 histone lysine methyltransferase
SUZ12	suppressor of zeste 12 protein homolog
Taz	Tazemetostat
TP53	tumour protein p53
UTX	lysine demethylase 6A
vPRC1	variant Polycomb repressive complex 1
WT	wild-type
YAF2	YY1 associated factor 2

CHAPTER 1

Introduction

1.1 Distinct gene expression programmes give rise to diverse cellular identities

Complex multicellular organisms are comprised of a vast array of diverse cell and tissue types. Yet, all organisms originate from a single totipotent cell - the zygote. Through a process known as development, this initial cell goes through a series of many cell divisions and sequential cell fate commitments. In this way, the initial single cell zygote gives rise to the diverse cell types that make up the organism. Notably, the genetic material, or DNA, is faithfully replicated from one generation of cells to the next, ensuring that all cells within an organism possess essentially the same DNA sequence. This concept was elegantly demonstrated in a seminal experiment in 1970 by Laskey and Gurdon in *Xenopus*, where the nuclei of adult somatic cells (such as lung, skin, and kidney cells) were transferred into enucleated eggs. The resulting zygotes developed normally, proving that the DNA is not changed to produce distinct cell types during development (Laskey & Gurdon, 1970). Instead, this is achieved through the activation and silencing of specific genes, resulting in distinct gene expression profiles that define the unique characteristics of each cell type.

Importantly, the regulatory mechanisms governing gene expression are impermanent, allowing for gene expression profiles to change over time as cells commit to specific identities. The concept of cellular differentiation during development is often illustrated by the analogy of a ball rolling down a hill (Waddington, 1957). At the top of the hill, a totipotent zygote has the potential to give rise to any cell type in the organism, with its daughter cells able to take any of the various paths down the hill (Fig 1.1). However, as development progresses, cells become increasingly restricted in their ability to differentiate due to the establishment of specific gene expression patterns. This can be thought of as the cell committing to a particular path down the hill, from which it cannot reverse. For instance, under normal conditions a neural stem cell can differentiate into lineage-specific cell types such as neurons or glial cells, but not into a haemopoietic stem cell. However, cell fate can be artificially reprogrammed across lineages or even reversed by forcing the expression of certain transcription factors. For example, in 1987 it was discovered that it is possible to convert fibroblasts to the myogenic lineage by forcing the expression of a master regulatory transcription factor, MYOD (Davis et al., 1987). Additionally, in 2006 it was shown that expression of just 4 embryonic stem cell (ESC) specific transcription factors can induce differentiated somatic cells to become pluripotent (Takahashi & Yamanaka, 2006).

Transcription factors are the master regulators of gene expression, but other mechanisms play a crucial role in maintaining correct gene expression patterns. In fact, there are many layers of regulatory mechanisms governing gene expression. The term “epigenetic” refers to heritable changes in gene expression that do not change the underlying DNA sequence. Epigenetic gene expression regulatory mechanisms include covalent modification of the DNA and/or its associated proteins (Badeaux & Shi, 2013).

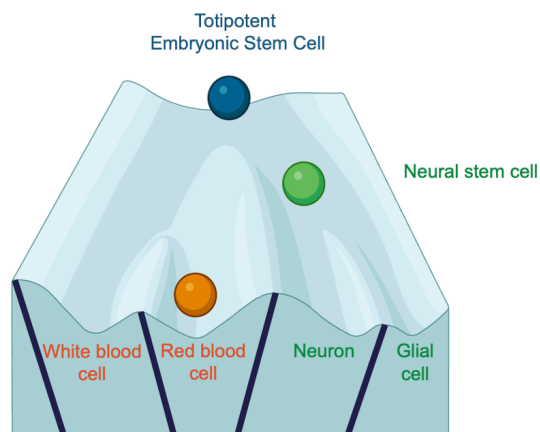


Figure 1.1 Waddington's landscape depicting differentiation during development.

A representation of the analogy used by Waddington to describe cellular differentiation during development. A differentiating cell is depicted by a ball rolling down the hill. At the top of the hill, a totipotent stem cell has the potential to roll down any path towards any cellular identity. As development progresses, cells become increasingly restricted in their ability to differentiate due to earlier lineage commitment decisions. A neural stem cell (green ball) can terminally differentiate into a cell type of the neural lineage (green), but not the haematological lineage (orange). A terminally differentiated cell such as a red blood cell (orange ball) cannot roll back up the hill to become a blood stem cell, nor can it change cell type even within its lineage, for example become a white blood cell. The support beams beneath the hills represent the underlying genetic code, as this affects how differentiation proceeds. For example, a genetic mutation in a developmental pathway may prevent a neural stem cell from being able to proceed down a differentiation pathway. Figure created with BioRender.com.

1.2 The organisation of eukaryotic DNA into chromatin regulates gene expression

The human genome, comprising approximately six billion base pairs of DNA, exists in the nucleus of the cell. If the DNA from a single diploid human cell were stretched out end-to-end, it would span two meters (Piovesan et al., 2019). However, the diameter of a typical human cell nucleus is only around 10µm, presenting a significant packaging problem. This problem is overcome through the formation of "chromatin". Chromatin is formed from the tight association between the negatively charged DNA wrapped around a core of positively charged histone proteins. The fundamental building block of chromatin is the "nucleosome", which comprises approximately 147 base pairs of DNA coiled around a histone octamer (Kornberg, 1974). This histone octamer contains dimers of the four core histone proteins (H2A, H2B, H3, and H4) (Luger et al., 1997). These core histones are among the most evolutionarily conserved proteins known. For example, histone H4 is almost perfectly conserved from yeast to mammals despite over a billion years of evolutionary divergence (Smith & Andrésson, 1983). This highlights the essential role of nucleosomes in chromatin organisation. Nucleosomes provide the first level of organisation, compacting the DNA approximately 6-fold into a fibre approximately 10nm in diameter. Nucleosomes are separated by a stretch of "linker" DNA. Histone H1 aids DNA compaction by binding the nucleosome and the linker DNA (Bednar et al., 2017). The second level of organisation involves coiling this 10nm nucleosome array into 30nm polynucleosomal fibres, which are stabilised by histone H1 and compact the DNA approximately 40-fold (Tremethick, 2007). This

30nm fibre is then further condensed and compacted into chromosomes, resulting in between 1,000- and 10,000-fold overall compaction of the DNA (Woodcock & Ghosh, 2010) (Fig. 1.2).

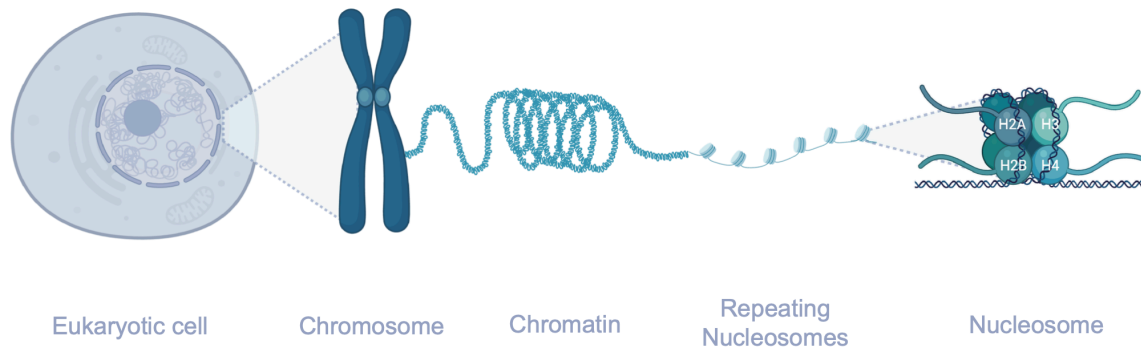


Figure 1.2 Packaging of DNA in the eukaryotic cell nucleus.

Illustration of the organisation of DNA in the nucleus. Higher order structures are on the right, beginning with the chromosome (highest order structure) and decreasing to the basic repeating subunit, the nucleosome, shown on the left. DNA double helical strands are wrapped around histone octamers, forming nucleosomes. The histone octamer contains dimers of the four core histone proteins (H2A, H2B, H3, and H4). Repeating nucleosomes are separated by linker DNA. The nucleosomal fibre (10nm diameter) is coiled and folded to create condensed polynucleosomal chromatin fibres (30nm diameter). These fibres further coil and condense, eventually forming the chromatid of a chromosome. Polynucleosomal chromatin fibres can be arranged into large, organized structures, the most compact being the metaphase chromosome which compacts DNA over 10,000 fold. Figure created with BioRender.com.

The local density of chromatin compaction influences the accessibility of the underlying DNA sequence to transcriptional machinery, thereby affecting gene expression. “Euchromatin” refers to regions of the genome that are characterized by a more open and accessible chromatin structure and is generally associated with active gene transcription. In contrast, “heterochromatin” is a more compact and densely packed form of chromatin that is typically found in transcriptionally inactive regions of the genome, where gene expression is suppressed. Heterochromatin can be further categorized into two types: constitutive and facultative (Quina et al., 2006). Constitutive heterochromatin is typically found in regions of low gene density, such as satellite, centromeric, and pericentromeric repeats, and exhibits consistent patterns of localisation across different cell types. In contrast, facultative heterochromatin is developmentally regulated, meaning its formation is highly specific to particular cell types and is shaped by cell identity. In this way, chromatin organisation functions beyond its role in compacting DNA within the nucleus, playing an additional role in regulating gene expression.

1.3 Histone post-translational modifications can influence chromatin structure and gene expression

As previously discussed, nucleosomes are composed of DNA wrapped around a core histone octamer. The globular C-terminal domains of the histones interact directly with the DNA, while the N-terminal tail domains protrude from the nucleosome. The histones and DNA can be covalently

modified by a wide range of chemical groups - acetylation and methylation of core histones H3 and H4 have been the most studied, but many other types of histone modifications exist including; ubiquitination, phosphorylation and SUMOylation among others (Strahl & Allis, 2000). These histone post-translational modifications can influence the regulation of chromatin structure and gene expression. This regulation is mediated by three key groups of proteins: chromatin "readers", "writers", and "erasers", which work together to dynamically modify chromatin structure (Allis & Jenuwein, 2016). Chromatin "writers" are responsible for catalysing histone post-translational modifications, which can alter the physical properties of nucleosomes and influence chromatin structure and gene expression. For instance, histone acetylation is generally associated with actively transcribed genes and their regulatory elements. A study published by Hebbes, Thorne, and Crane-Robinson in 1988 demonstrated the link between histone acetylation and active genes. They used an antibody to immunoprecipitate nucleosomes containing acetylated histones from chicken erythrocyte nuclei, and then probed the resulting chromatin with beta-globin and ovalbumin DNA sequences. The results showed that the beta-globin locus was enriched, but not the ovalbumin gene, indicating a correlation between histone acetylation and gene activity (Hebbes et al., 1988). The addition of acetyl groups to histones neutralizes their positive charge, weakening their interaction with negatively charged DNA and leading to a more open euchromatin conformation. This, in turn, increases accessibility for transcription factors and transcriptional machinery (Brownell & Allis, 1996).

The histone modifications established by chromatin "writers" can affect transcription not only by directly altering chromatin structure, but also by serving as binding sites for chromatin "reader" proteins. These effector proteins are recruited to chromatin through the recognition of specific modifications, which they detect using specialised binding domains such as chromo-, bromo-, and PHD-finger domains. Upon binding, reader proteins can modify the local chromatin environment by adding additional modifications or recruiting other proteins, thereby amplifying the regulatory signal. The specific amino acid residue that is modified, as well as the degree and location of the modification, provides context that influences the recruitment of reader proteins. For instance, lysine residues can be modified to varying degrees, with different readers responding to distinct patterns of methylation (Black et al., 2012). This nuanced system allows for opposing effects on transcription, as exemplified by the contrasting roles of trimethylation at H3K27 (H3K27me3) and H3K36 (H3K36me3). While H3K27me3 recruits Polycomb Repressive Complex 1 (PRC1) to chromatin, leading to chromatin compaction and gene repression, H3K36me3 does not recruit PRC1 and is instead associated with active gene expression (Schmitges et al., 2011; Yuan et al., 2011; Zheng et al., 2012). Additionally, histone tail domains can be densely modified with different marks, allowing for "cross-talk" between them, where the interpretation of one modification is influenced by another nearby modification (Jenuwein & Allis, 2001). This highlights the complex interplay between different histone modifications and their combinations, which can have distinct effects on chromatin structure and gene expression. Importantly, histone modifications are generally reversible with their removal being catalysed by chromatin "eraser" proteins. For example, H3K27me3 can be removed by histone demethylases such as UTX and JMJD3, while histone acetylation can be removed by histone

deacetylases (HDACs) (Hong et al., 2007; Seto & Yoshida, 2014) (Fig. 1.3). This allows for dynamic regulation of chromatin structure and gene expression. However, while dynamic and reversible, this epigenetic form of gene regulation can be inherited by daughter cells following cell division, ensuring the maintenance of appropriate gene expression programmes and cell identity.

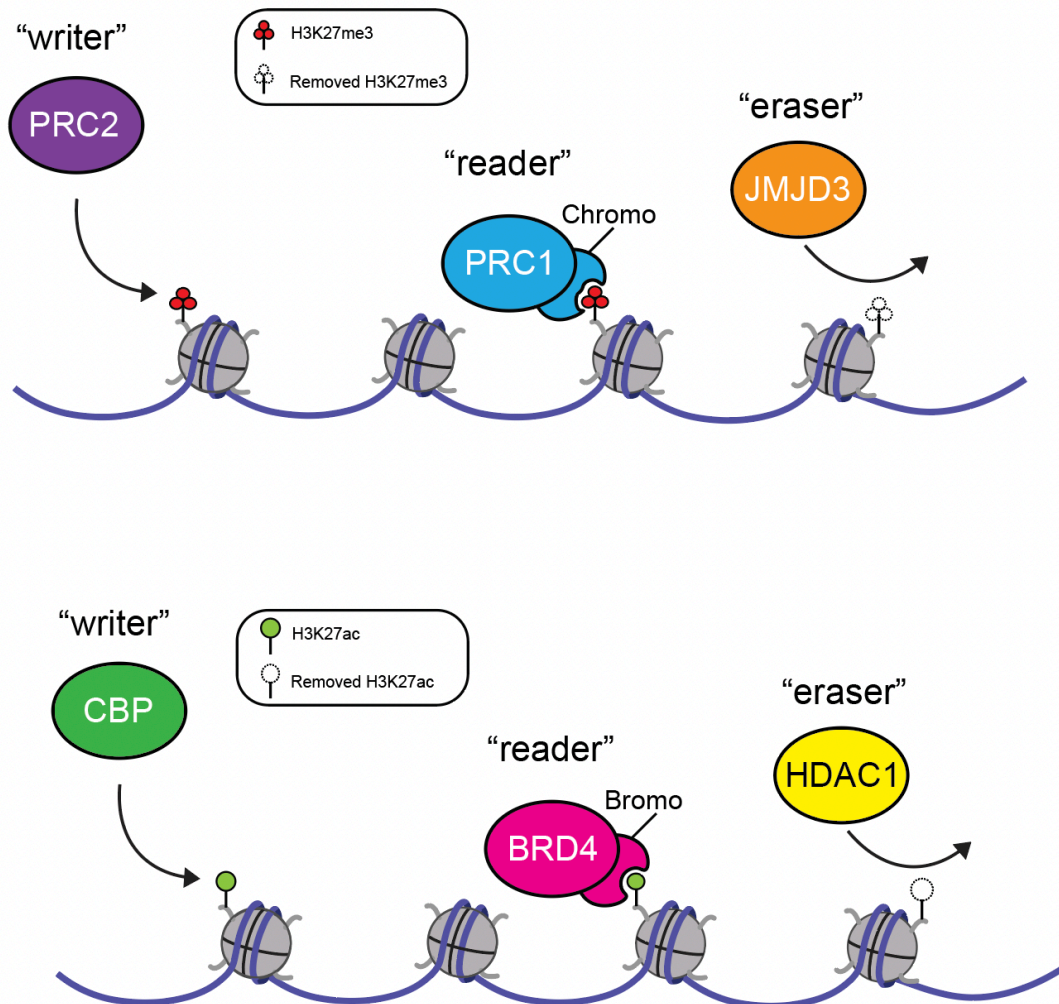


Figure 1.3 Histone “readers”, “writers”, and “erasers”.

Diagrammatic representing some examples of histone “readers”, “writers”, and “erasers” cooperating to maintain dynamic regulation of chromatin. Top: H3K27me3, associated with inactive gene promoters, is catalysed by PRC2. Certain PRC1 complexes can bind H3K27me3 via a protein subunit containing a chromodomain, and H3K27me3 can be removed by histone demethylases such as JMJD3. Bottom: H3K27ac, associated with active gene promoters, can be catalysed by histone acetyltransferases (HATs) such as CBP. BRD4 can bind H3K27ac via its bromodomain. H3K27ac can be removed by histone deacetylases (HDACs) such as HDAC1. Figure adapted from Dr Ellen Tuck.

Of particular relevance for this thesis are histone lysine modifications and associated chromatin reader and writer proteins. Most histone lysine writers, or methyl transferases, contain a conserved SET domain which is named after the three *Drosophila* proteins in which it was first identified: Suppressor of variegation (*Su(var)3-9*), Enhancer of zeste (*E(z)*), and Trithorax (*Trx*) (Jones &

Gelbart, 1993; Tschiersch et al., 1994). The human homolog of *Su(var)3-9*, *SUV39H1*, was the first characterized histone methyltransferase, shown to tri-methylate Histone H3 at lysine 9 (H3K9me3) (Rea et al., 2000). Since then, over 28 additional methyltransferases have been identified, including *EZH1* and *EZH2*, ((*E(z)*) orthologs that methylate histone H3 at lysine 27) (Kuzmichev et al., 2002; Müller et al., 2002; Shen et al., 2008), as well as *KMT2A-D* (*Trx* orthologs) (Piunti & Shilatifard, 2016) in addition to many others.

1.4 Polycomb group protein complexes play an important role in development and cellular identity maintenance

Polycomb group proteins are highly conserved chromatin regulators among multicellular organisms. Polycomb group proteins were discovered in *Drosophila melanogaster* where seminal studies revealed the key role they play in development and maintaining cellular identity. The homeotic (*Hox*) genes encode transcription factors that are essential for determining body segment identity in *Drosophila*. *Hox* mutants display dramatic homeotic transformations, where body segments or structures develop at inappropriate locations. In the initial stages of normal development, transcription factor proteins transiently establish segment identity. Segment identity must then be stabilised by adjusting chromatin conformation to 'lock' the appropriate identity in place and allow differentiation to occur (Ringrose & Paro, 2004). Trithorax group proteins maintain *Hox* gene expression in the developing embryo by maintaining an open chromatin configuration at these genes. Polycomb group proteins antagonise Trithorax group proteins and repress *Hox* gene expression, restricting them to their appropriate segments (Lewis, 1978; Schuettengruber & Cavalli, 2009). In *Drosophila*, Polycomb loss-of-function mutants display cell type transformations that recapitulate *Hox* gene gain-of-function mutants, demonstrating the importance of Polycomb group proteins in maintaining cellular identity.

Indeed, the name "Polycomb" originates from the distinctive phenotype of the mutants in which one of the first Polycomb group gene was identified. In *Drosophila*, wild-type adult males normally have sex combs only on their first pair of legs. However, the *Polycomb* (*Pc*) gene mutant exhibited additional sex combs on the second and third pairs of legs (Lewis, 1947). This abnormal phenotype was caused by the derepression of the *Hox* gene *sex combs reduced*, which is normally restricted to the first pair of legs (Kassis et al., 2017). Following the initial *Pc* discoveries, a further 15 Polycomb group genes were identified in *Drosophila* through genetic screens that produced mutants with an extra sex combs phenotype and aberrant *Hox* gene expression. Today, we know that Polycomb group proteins play a crucial role in regulating not only *Hox* genes but also many other developmental genes (Schuettengruber et al., 2017). Polycomb group proteins do not primarily determine the expression of their target genes. Instead, they play a crucial role in maintaining the repressed state of these genes, which is initially established by transcription factors and repressors. In the absence of Polycomb group proteins, target genes are initially repressed as normal, but as development progresses, they become aberrantly activated. This is because Polycomb group proteins are responsible for recognizing and maintaining the repressed state of these genes, ensuring that it is

preserved through cell division and passed on to daughter cells. It is now appreciated that Polycomb group proteins play an important role in development and maintenance of cellular identity by controlling the spatial and temporal expression of thousands of developmental genes (Bracken et al., 2006; Kim & Kingston, 2022).

Polycomb group proteins form large multi-subunit complexes that cooperate to maintain gene repression. There are two broad families of such complexes: Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2). PRC2 complexes contain *EZH1* or *EZH2*, which are the human homologs of the *Drosophila* methyltransferase *E(z)*, meaning PRC2 is responsible for catalysing methylation at H3K27 (Kuzmichev et al., 2002; Shen et al., 2008). The Polycomb group proteins have expanded and diversified in higher order species, particularly in mammals, producing a much more diverse array of complexes in mammals compared to *Drosophila* (Whitcomb et al., 2007). This enables more sophisticated regulation of gene expression to support the development and maintenance of complex tissues.

Experiments in *Drosophila* showed that H3K27 to arginine missense mutations (H3K27R) result in a failure to maintain repression of *Hox* genes and other Polycomb target genes, causing homeotic transformation phenotypes similar to those seen in PRC2 loss-of-function mutants (Pengelly et al., 2013). This demonstrates that H3K27 is the critical substrate for PRC2, and its methylation is essential for Polycomb-mediated gene repression. Consistent with this, introducing H3K27R mutations in all 28 histone H3 alleles in mouse embryonic stem cells (mESCs) results in similar transcriptional profiles and differentiation defects as PRC2-null mutants (Sankar et al., 2022). Homozygous loss-of-function mutations in any of the core PRC2 genes is embryonic lethal in mice, causing death at gastrulation (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004). This demonstrates the key role of PRC2-mediated methylation in mammalian development. Furthermore, mutations in the PRC1 core subunits have an even more severe impact, causing embryonic development to arrest at the two-cell stage (Posfai et al., 2012; Voncken et al., 2003). These findings highlight the conserved and critical role of Polycomb group proteins in development across species. Notably, the composition of PRC1 and PRC2 complexes can change throughout development and between cell types (Kim & Kingston, 2022). Given the importance of these complexes in maintaining cellular identity, and how frequently they are dysregulated in cancer (Parreno et al., 2022), furthering our understanding of these complexes is of great biological and clinical importance.

1.5 PRC1 complex structures and functions

In 2001, Saurin et al. successfully purified the first Polycomb group protein complex assembly from *Drosophila* embryonic extracts, now known as Polycomb Repressive Complex 1 (PRC1) (Saurin et al., 2001). Subsequent studies in mammalian cells led to the identification of similar complexes which also contained homologs of the *Drosophila* Polycomb group protein subunits - the mammalian canonical PRC1 (cPRC1) complexes (Vandamme et al., 2011). Further investigation into the interacting partners of the RING proteins led to the discovery of an alternative PRC1 complex with

H2AK119ub ligase activity in both mammals and *Drosophila* (Gearhart et al., 2006; Lagarou et al., 2008). This complex, referred to as the Ring Associated Factors (dRAF) complex in *Drosophila*, is known as variant PRC1 (vPRC1) in mammals.

Polycomb Repressive Complex 1 (PRC1) assembles around a core consisting of two subunits: an E3 ubiquitin ligase Really Interesting New Gene 1 (RING1) and a Polycomb Group Ring Finger (PCGF) protein. In *Drosophila* there is just one ortholog for RING1 (*Sce/dRING*) (Gorfinkiel et al., 2004) and two for the PCGF (*Psc* and *Su(z)2*) (Brunk et al., 1991; Lo et al., 2009). These have duplicated in mammals to produce multiple paralogs; mammalian PRC1 contains either RING1A or RING1B, and one of six PCGF proteins (PCGF1-6) (Gao et al., 2012). PRC1 can catalyse the addition of a ubiquitin molecule to histone H2A at lysine 119 in mammals, or lysine 118 in *Drosophila* (H2AK119ub/H2AK118ub) (Wang et al., 2004). This modification is enriched at the promoters of repressed developmental genes (Fursova et al., 2019). The RING protein forms a dimer with a PCGF protein through their shared conserved RING domains. This dimerization enables the complex to interact with an E2 conjugating enzyme, which is necessary for histone ubiquitination (Blackledge & Klose, 2021; Wang et al., 2004). Additionally, RING1A/B and PCGF proteins contain a RAWUL (Ring-finger and WD40-associated ubiquitin-like) domain, which is important for selective binding interactions with other proteins in the complex (Blackledge & Klose, 2021).

PRC1 complexes are defined by the accessory subunits present, which is determined by the PCGF paralog in the complex. The core of canonical PRC1 (cPRC1) contains RING1A/B and either PCGF2 (also known as MEL18) or PCGF4 (also known as BMI1) (Blackledge & Klose, 2021). The *Drosophila* ortholog of *PCGF2/PCGF4* is *Su(z)2* (Lo et al., 2009). The RAWUL domains of PCGF2 and PCGF4 are highly similar, which allows them to share many binding partners, including a chromobox (CBX) protein (the original *Polycomb* (*Pc*) in *Drosophila*) and a PHC protein (*Polyhomeotic-like*, *Ph* in *Drosophila*) (Gao et al., 2012; Whitcomb et al., 2007). Although cPRC1 is thought to have minimal catalytic activity, it plays a crucial role in transcriptional repression by facilitating chromatin compaction and looping (Eskeland et al., 2010; Gao et al., 2012; Grau et al., 2011). Different CBX paralogs (CBX-2/4/6/7/8) can be incorporated into cPRC1 (Fig. 1.4). The chromodomain of CBX proteins recognizes and binds to the H3K27me3 histone modification, while the C-box domain interacts with the core and facilitates assembly into the complex (Fischle et al., 2003; Min et al., 2003; Schoorlemmer et al., 1997). The PHC proteins (PHC1-3) oligomerize to facilitate cPRC1-mediated chromatin looping and compaction (Isono et al., 2013). This sequesters target developmental genes into dense, three-dimensional structures known as "Polycomb bodies", which is thought to repress their expression (Isono et al., 2013; Saurin et al., 1998). Generally, cPRC1 complexes are structurally very similar, containing the same protein subunits but varying in which paralog of these subunits is present. cPRC1 complexes are also presumed to be largely functionally redundant due to their structural similarity and a lack of evidence of unique functions (Scelfo et al., 2019; Zhu et al., 2022).

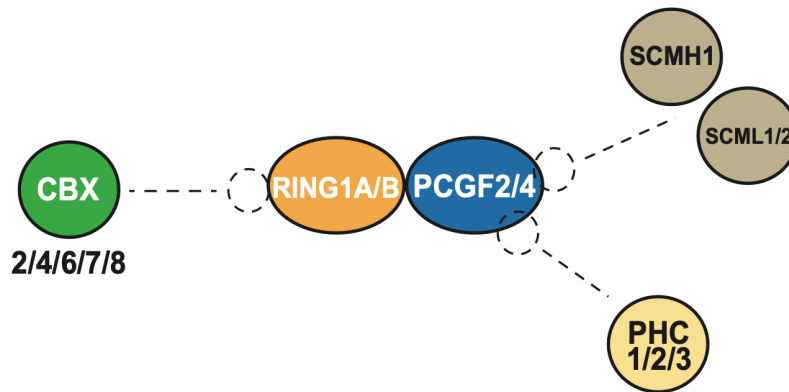


Figure 1.4 Mammalian canonical PRC1 complex structure.

Diagrammatic representation of mammalian canonical Polycomb Repressive Complex 1 (cPRC1) structure. Paralogs are indicated within the same subunit. PCGF2 and PCGF4 are also known as MEL18 and BMI1 respectively.

Any of the six mammalian PCGF paralogs can incorporate into vPRC1 complexes. vPRC1 complexes also contain either RYBP or YAF2 in place of a CBX, and an array of other unique protein subunits (Fig. 1.5A). vPRC1 complexes are thought to be the more catalytically active complexes as the RYBP subunit can stimulate catalytic activity (Ciapponi et al., 2024; Rose et al., 2016; Zhao et al., 2020). The classification of different vPRC1 complexes is based on the specific PCGF subunit they contain, with the current nomenclature denoting the PCGF subunit present (e.g., vPRC1.1, vPRC1.2, vPRC1.3, vPRC1.4, vPRC1.5, vPRC1.6) (Fig. 1.5B). The PCGF subunits interact directly with RING1A/B via their RING domain but exhibit paralogue-specific binding to partner proteins due to alterations in their RAWUL domain, which confer distinct binding specificities (Tamburri et al., 2024). Consequently, the type of PCGF subunit plays a crucial role in determining the composition of the PRC1 complex. The vPRC1 complexes exhibit a remarkable degree of structural diversity, with a wide range of accessory subunits that can bind to the PCGFs. While vPRC1 complexes do have largely overlapping functions, this structural diversity is accompanied by well-documented functional diversity, with different vPRC1 complexes exhibiting distinct functional properties (Tamburri et al., 2024). For instance, PRC1.6 complexes can be targeted to chromatin via specific accessory subunits, which possess DNA binding capacity (Farcas et al., 2012; Scelfo et al., 2019; Stielow et al., 2018). Furthermore, while PRC1.1 and PRC1.6 complexes are associated with transcriptional repression (Scelfo et al., 2019), PRC1.3 and PRC1.5 complexes are paradoxically linked to enhanced transcriptional activity (Gao et al., 2014). These findings highlight the diverse structures and functions of vPRC1 complexes. It is worth noting that PCGF2 and PCGF4 were initially identified as exclusive components of CBX-containing cPRC1 complexes. However, a systematic analysis of the human complexome using a robust affinity purification mass spectrometry approach revealed detectable interactions between PCGF2/4 and RYBP or YAF2, suggesting the existence of vPRC1.2 and

vPRC1.4 (Hauri et al., 2016). Notably, PCGF2 and PCGF4 do not interact with any additional subunits beyond RING1A/B and RYBP/YAF2, raising the possibility that these vPRC1.2 and vPRC1.4 complexes may represent transient intermediate products of PRC1 holo-complex assembly (Bajusz et al., 2018). Consequently, PCGF2 and PCGF4 are still primarily associated with cPRC1.

A



B

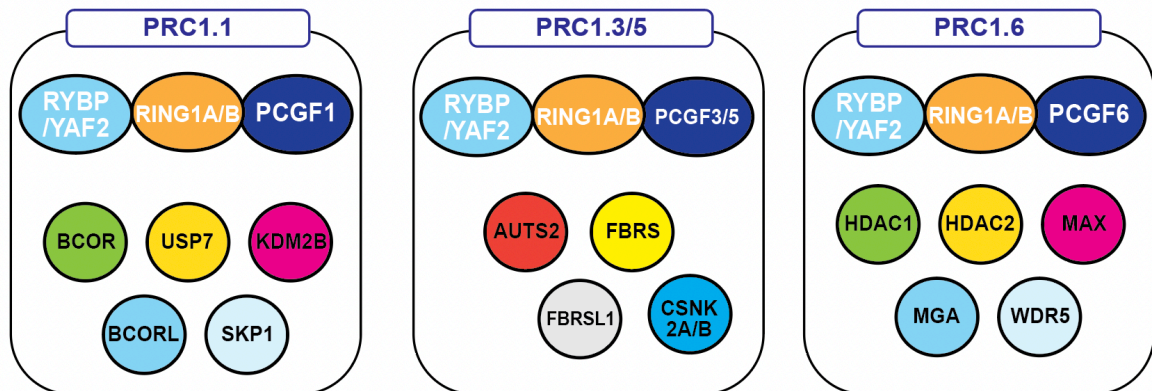


Figure 1.5 Mammalian variant PRC1 complex structure.

A) Diagrammatic representation of general mammalian variant Polycomb Repressive Complex 1 (vPRC1) complex structure. Paralogs are indicated within the same subunit.

B) Diagrammatic representation of some examples of vPRC1 complex subunit composition.

1.6 PRC2 complex structures and functions

Polycomb Repressive Complex 2 (PRC2) catalyses mono-, di- and tri-methylation of histone H3 lysine 27 (H3K27me_{1/2/3}). The exact role of each degree of H3K27 methylation is not fully understood. H3K27me₁ is lowly abundant, found on just 5-10% of all histone H3, and is enriched at active gene bodies. Contrasting this, H3K27me₂ is highly abundant, present at 50-70% of all histone H3 and is generally found at inter- and intragenic regions (Ferrari et al., 2014). The function of widespread H3K27me₂ is unknown but it has been suggested to act as a repressive 'blanket' preventing gene enhancers of alternative lineages from inappropriately promoting expression (Conway et al., 2015). H3K27me₃ is found at 5-10% of histone H3 and can be at low levels dispersed throughout the genome but is enriched at promoters where PRC2 is bound (Højfeldt et al., 2018).

The homologs of *Drosophila* E(z), EZH1 and EZH2, contain the catalytic SET domain (Czermin et al., 2002; Müller et al., 2002). Although EZH1 has some residual catalytic activity, EZH2 is the more catalytically active subunit and is predominantly responsible for the methylation of H3K27

(Margueron et al., 2008). The catalytic core of PRC2 consists of a trimer of EZH1/2, EED and SUZ12 (Fig. 1.6) (Højfeldt et al., 2018). RBBP4 and RBBP7 can both associate with core PRC2, but are not necessary for catalytic activity. EED is essential for the allosteric activation of PRC2, which is necessary for maintaining H3K27me3 levels (Margueron et al., 2009). The binding of EED to H3K27me3 triggers a conformational change in the EZH2 protein, leading to increased stabilisation and enhanced catalytic activity (Jiao & Liu, 2015). SUZ12 protein plays a critical role in stabilizing the PRC2 complex, binding to EZH1/2 and EED through its C-terminus (Ciferri et al., 2012). The SUZ12-EZH2-EED trimer is sufficient for catalytic activity. However, additional subunits are required to bind the N-terminus of SUZ12 to target PRC2 to specific sites on chromatin (Højfeldt et al., 2019; Laugesen et al., 2019). There are 2 subcomplexes of PRC2 depending on the accessory subunits present. PRC2.1 contains either EPOP or PALI1/2 and one of three PCL proteins (PCL1-3), whereas PRC2.2 contains both JARID2 and AEBP2 (Laugesen et al., 2019). The most recent PRC2 subunit to be discovered is EZHIP (also known as CATACOMB) which can associate with either PRC2.1 or PRC2.2 and has an inhibitory effect on PRC2 catalytic activity (Hübner et al., 2019; Jain et al., 2019; Pajtler et al., 2018; Piunti et al., 2019).

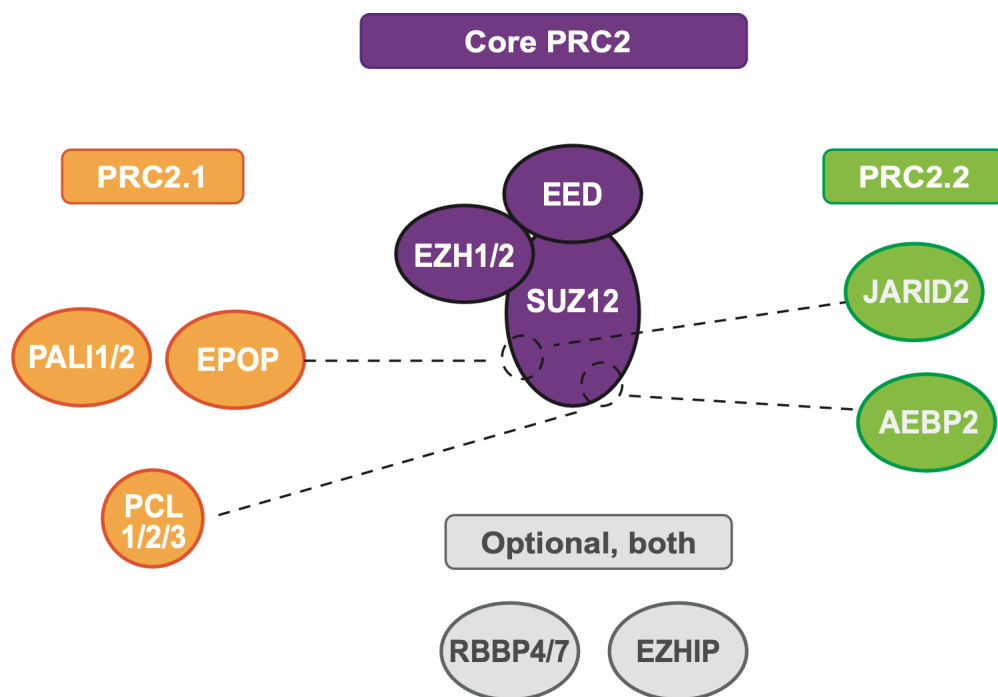


Figure 1.6 Mammalian PRC2 complex structure.

Diagrammatic representation of mammalian Polycomb Repressive Complex 2 (PRC2) complex structure. The catalytic trimeric core of PRC2 common to all complexes is indicated in purple. RBBP4/7 (grey) are found in both PRC2.1 and PRC2.2 and while they can increase methyltransferase activity, they are not essential for it. Like RBBP4/7, EZHIP (grey) has been reported to associate with PRC2.1 and PRC2.2, but is inhibitory to catalytic activity and its expression is restricted to the testes. PRC2.1 specific subunits (orange) and PRC2.2 specific subunits (green) are indicated. Paralogs are denoted within the same circle. Figure adapted from Dr Dáire Gannon.

Notably, the relative abundances of different subunits and paralogs differs by cell type and developmental stage. For example, EZH2 is the most abundant paralog in pluripotent stem cells, with EZH1 becoming more abundant in quiescent and terminally differentiated cell types (Bracken et al., 2003; Margueron et al., 2008; Shen et al., 2008). Additionally, EZHIP is normally only expressed in the gonads (Ragazzini et al., 2019). The mechanism by which PRC2 complexes are recruited to chromatin and mediate gene repression are not yet fully understood but the accessory subunits and interplay with PRC1 are important for PRC2 recruitment and activity.

1.7 PRC1 and PRC2 cooperate to maintain target gene repression

PRC1 and PRC2 are bound on chromatin at CpG-rich promoters of transcriptionally repressed genes (Bracken et al., 2006; Ku et al., 2008). CpG dinucleotides account for just 1% of the human genome (Antequera & Bird, 1993). The vast majority (70-80%) of these CpG dinucleotides are marked by DNA methylation in mammalian genomes (Lister et al., 2009). The CpG-rich promoters where PRC1/PRC2 are bound are known as CpG islands (CGIs). CGIs are short stretches of DNA (typically 1-2 kilobases in length) characterized by an unusually high density of cytosine and guanine dinucleotides (Illingworth & Bird, 2009). Interestingly, these CGIs are refractory to DNA methylation. Notably, while Polycomb-mediated repression is reversible, DNA methylation is associated with the stable repression of genes; suggesting PRC1 & PRC2 maintain a more dynamic state of transcriptional silencing in the absence of DNA methylation.

Chromatin immunoprecipitation sequencing (ChIP-seq) experiments have revealed that PRC2 and H3K27me3 are highly enriched at repressed Polycomb target CGIs, forming distinctive peaks of signal (Bracken et al., 2006). However, genome-wide analyses have also shown that H3K27me2 is present in a diffuse pattern across the genome, as is H3K27me3, albeit at much lower levels than at CGIs (Ferrari et al., 2014; Højfeldt et al., 2019). Notably, PRC2 is not co-enriched with this diffuse H3K27me2 and H3K27me3, suggesting that the complex is capable of transiently interacting with chromatin to deposit H3K27me2 and to a lesser extent, H3K27me3 (Sneeringer et al., 2010). This is consistent with the enzymatic kinetics of PRC2, which exhibits a preference for catalysing the initial mono-methylation of H3K27 (H3K27me1) and decreasing efficiency for each subsequent methylation event (H3K27me1 to H3K27me2 to H3K27me3). The longer residence time of PRC2 at Polycomb-bound target sites is therefore required for efficient H3K27me3 deposition.

PRC1 and PRC2 have complicated interplay, as both can act upstream and downstream of the other (Fig. 1.7). vPRC1 complexes can be directed to chromatin through interactions with accessory protein subunits, such as KDM2B, which is specifically associated with PCGF1-containing vPRC1 complexes (Fig. 1.7A) (Blackledge et al., 2014). KDM2B's zinc finger-CxxC (ZF-CxxC) domain enables it to bind to non-methylated CpG-rich DNA sequences, thereby recruiting the vPRC1 complex to these regions and facilitating the catalysis of H2AK119ub (Blackledge et al., 2014). Furthermore, RYBP, a subunit common to all vPRC1 complexes, can bind H2AK119ub and facilitate the propagation of this modification to adjacent nucleosomes (Zhao et al., 2020).

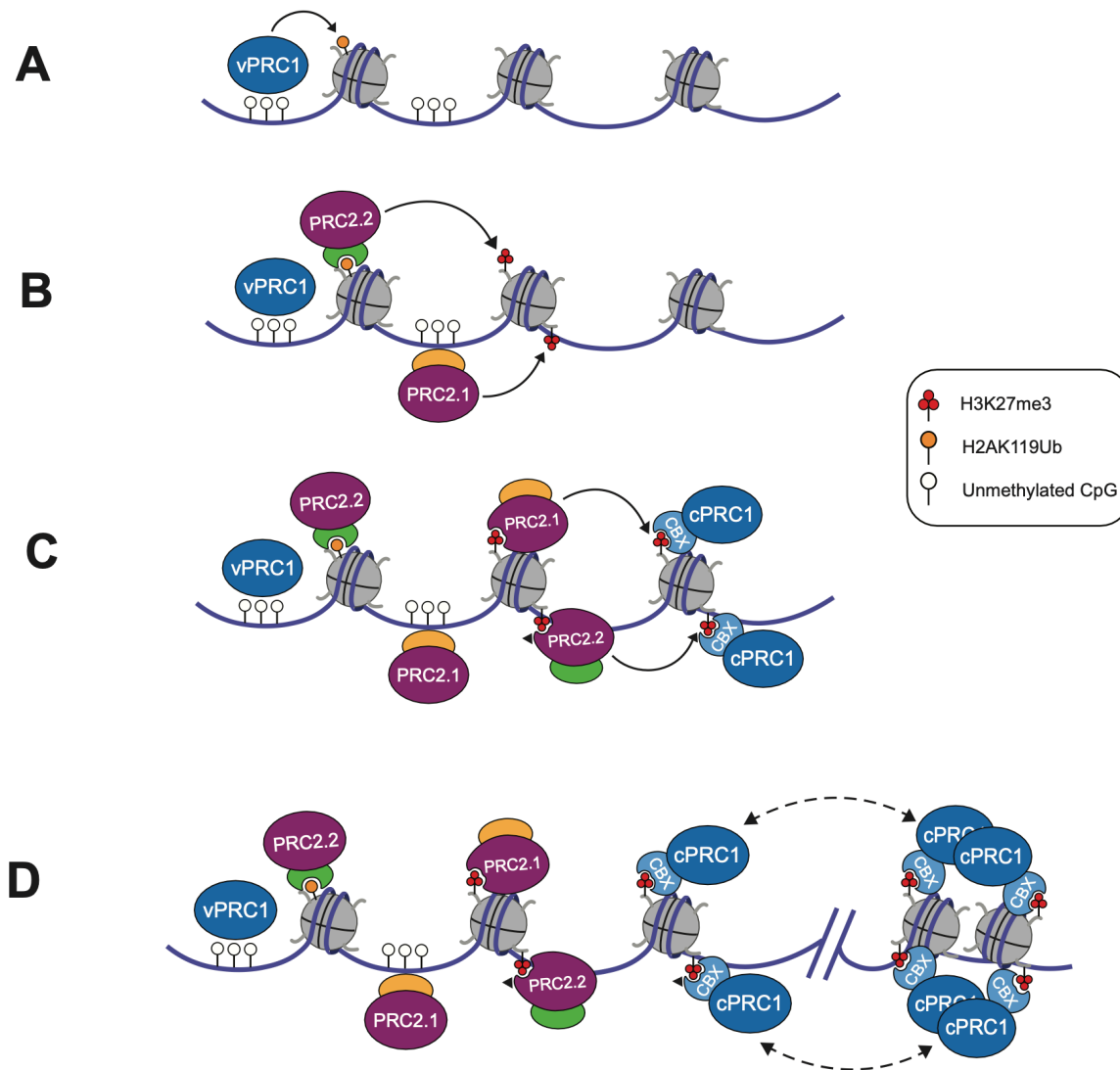


Figure 1.7 PRC1/PRC2 recruitment and interplay.

A) vPRC1 complexes can be directed to chromatin through accessory protein subunits, such as KDM2B, which enables it to bind to non-methylated CpG-rich DNA and catalyse H2AK119ub.

B) PRC2 is recruited to chromatin either via binding to unmethylated CpG islands in the case of PRC2.1, or via binding to H2AK119Ub in the case of PRC2.2. Both catalyse H3K27me3.

C) PRC2 is allosterically activated by the H3K27me3 it deposits, stimulating it to deposit more. cPRC1 binds to H3K27me3 mark via its CBX subunit.

D) cPRC1 complexes oligomerize via the PHC subunit to condense and compact chromatin, repressing gene expression.

PRC1-mediated H2AK119ub can facilitate the recruitment of some PRC2 complexes to chromatin (Fig. 1.7B). Cryo-electron microscopy (cryo-EM) structures and biochemical studies have demonstrated the PRC2.2-specific subunits AEBP2 and JARID2 can bind to H2AK119ub-modified nucleosomes (Kasinath et al., 2021). In contrast, PRC2.1 complexes can be targeted to chromatin independent of PRC1 activity through the PCL subunit, which possesses a winged helix domain that enables interaction with GC-rich DNA sequences (Fig. 1.7B) (Li et al., 2017; Perino et al., 2018). The affinity of PCL proteins for GC-rich DNA has led to the proposal that they play a role in recruiting PRC2 to CpG islands. Upon chromatin targeting, both PRC2 subcomplexes catalyse the H3K27me3,

a hallmark of Polycomb-mediated repression. Both PRC2 subcomplexes can be allosterically activated to further enhance H3K27me3 catalysis (Fig. 1.7C). Specifically, the PRC2.1 accessory subunit PALI1 can be methylated by PRC2, and the resulting methylated residue can bind to the EED aromatic cage, thereby allosterically activating PRC2 (Conway et al., 2018; Zhang et al., 2021). Similarly, the PRC2.2-specific subunit JARID2 can also be methylated by PRC2, leading to allosteric stimulation of PRC2 activity and further promoting H3K27me3 deposition (Sanulli et al., 2015).

PRC1 can also function downstream of PRC2 by interacting with H3K27me3 (Fig. 1.7C). The CBX proteins present in cPRC1 complexes can bind to H3K27me3 via the chromodomain, facilitating the recruitment of cPRC1 (Bernstein et al., 2006; Cao et al., 2002; Min et al., 2003). Consistent with this recruitment mechanism, chromatin immunoprecipitation (ChIP) has shown that the majority of PRC2-associated, H3K27me3-marked genes are co-bound by cPRC1 in multiple mammalian cell types (Boyer et al., 2006; Bracken et al., 2006). Although *in vitro* studies have shown that cPRC1 complexes exhibit reduced catalytic activity compared to vPRC1 complexes, cPRC1 has been found to oligomerize via the PHC subunits, allowing them to bridge distant chromatin loci and facilitate long-range chromatin interactions (Fig. 1.7D) (Isono et al., 2013). The ability of CBX subunits to undergo phase separation, combined with the oligomerization of cPRC1 complexes, has been proposed to contribute to the formation of "Polycomb bodies" - microscopically visible nuclear clusters of condensed chromatin thought to be refractory to transcription (Tatavosian et al., 2019).

1.8 Polycomb group proteins are frequently dysregulated in cancer

The maintenance of chromatin regulation is a crucial process that extends beyond developmental stages, playing a vital role in the preservation of tissue homeostasis and the prevention of cancer in adult somatic tissues. Recent studies have emphasized the significance of phenotypic plasticity and non-mutational epigenetic reprogramming as emerging hallmarks of cancer, building upon the foundational hallmarks of cancer first described by Hanahan and Weinberg in 2000 (Hanahan, 2022). Notably, genes encoding chromatin modification regulators are among the most frequently mutated genes in cancer, underscoring the importance of epigenetic dysregulation in tumorigenesis. Next-generation sequencing data of cancer genomes has revealed that chromatin regulators are disproportionately represented among recurrent cancer driver genes (Vogelstein et al., 2013). Mutation of chromatin regulators is particularly pervasive in paediatric cancers (Gröbner et al., 2018). Oncogenesis in adults is often a slow process requiring progressive accumulation of mutations that incrementally dysregulate cell proliferation and function over time. However, paediatric malignancies have a much lower (~14-fold) mutational burden on average than adult cancers, likely because they have had less opportunity to accumulate mutations over time through replication and exposure to carcinogens or mutagens (Gröbner et al., 2018). Consequently, the underlying disease aetiology of paediatric cancers is often very different to their adult counterparts. Tumour sequencing studies over the last decade have brought this phenomenon to the fore. Whole genome sequencing of nearly 1000 tumour samples from 24 distinct cancer types revealed that almost 60% of paediatric tumours harboured just one genetic mutation driving the disease, contrasting adult cancers which typically

require multiple driver mutations (Gröbner et al., 2018; Vogelstein et al., 2013). This striking enrichment highlights the critical role of chromatin dysregulation in the development and progression of cancer.

Mutations in PRC1 are not commonly enriched in cancer genome studies, and when they are present they are generally not considered driver mutations. However, overexpression of cPRC1 subunits, such as PCGF4 or CBXs, has been observed in various types of cancer (Bracken et al., 2019; Parreno et al., 2022; Pasini & Di Croce, 2016). Notably, PCGF4, a component of PRC1, was initially identified as an oncogene that cooperates with *Myc* to promote lymphomagenesis (Jacobs et al., 1999). Nevertheless, the observed over expression of cPRC1 components in some cancers may be a consequence, rather than a cause, of tumorigenesis. Specifically, the expression of these genes is correlated with cell proliferation and stem cell identity, suggesting that their upregulation may be a secondary effect of increased cell proliferation or cancer cells with more stem cell-like features (Bracken et al., 2019).

In contrast to PRC1, sequencing studies have revealed that PRC2 subunits are frequently mutated in a variety of cancers. Specifically, a recurrent heterozygous mutation at Y641 in the catalytic SET domain of EZH2 has been identified in certain haematological malignancies, namely diffuse large B-cell lymphomas (22%) and follicular lymphomas (7%) (Morin et al., 2010). This gain-of-function mutation enhances the enzymatic activity of EZH2, leading to an increased conversion of H3K27me2 to H3K27me3 (Sneeringer et al., 2010). Although the precise mechanism by which this mutation contributes to oncogenesis is not fully known, it is hypothesized that the resultant increase in H3K27me3 at the promoters of genes involved in cell differentiation leads to their aberrant repression, thereby blocking differentiation and facilitating cellular transformation (Béguelin et al., 2013).

While EZH2 is recognized as an oncogene in certain lymphoma contexts, PRC2 genes can also exhibit tumour suppressor functions in different cancer contexts. Notably, somatic loss-of-function mutations and deletions in *EZH2* and *SUZ12* have been identified in a significant proportion of T-cell acute lymphoblastic leukaemia cases (25%) and in the majority of sporadic malignant peripheral nerve sheath tumours (90%) (Lee et al., 2014; Ntziachristos et al., 2012). These genetic alterations result in decreased levels of H3K27me3. These studies suggest that this reduction in H3K27me3 leads to the derepression of PRC2 target genes. The observation that both activating and inactivating mutations of *EZH2* are linked to oncogenesis underscores the context-dependent role of chromatin regulation in cancer biology.

Chromatin regulators, such as PRC2, can also be dysregulated through mutations in their substrate histones. Notably, large-scale cancer genomics studies have revealed that several paediatric malignancies harbour recurrent somatic mutations in genes encoding histone proteins, a finding that was unexpected given the high degree of redundancy in histone genes (Behjati et al., 2013; Schwartzentruber et al., 2012; Wu et al., 2012). These "oncohistones" often serve as the initiating

driver of oncogenesis. To date, mutations in the N-terminal tail of histone H3 have been the most well-characterized. These mutations disrupt histone modifications at or near the mutated residue, leading to aberrant chromatin regulation. One of the first recurrent oncohistone mutations to be identified was found in a rare type of paediatric brain cancer (Schwartzentruber et al., 2012; Wu et al., 2012). This mutation affects the PRC2 substrate H3K27 and is known to dysregulate PRC2 activity. Together, these examples highlight the complex and context-dependent role of Polycomb group proteins in cancer.

1.9 Diffuse Midline Glioma (H3K27-altered)

Diffuse Midline Glioma (DMG), previously known as Diffuse Intrinsic Pontine Glioma (DIPG), is a rare but universally fatal paediatric malignancy which usually arises in the brainstem and midline structures of the CNS (Fig. 1.8A) (Louis et al., 2016; Louis et al., 2021). Peak-incidence occurs at around 5-7 years old and median survival is just 9 months from diagnosis (Jones & Baker, 2014).

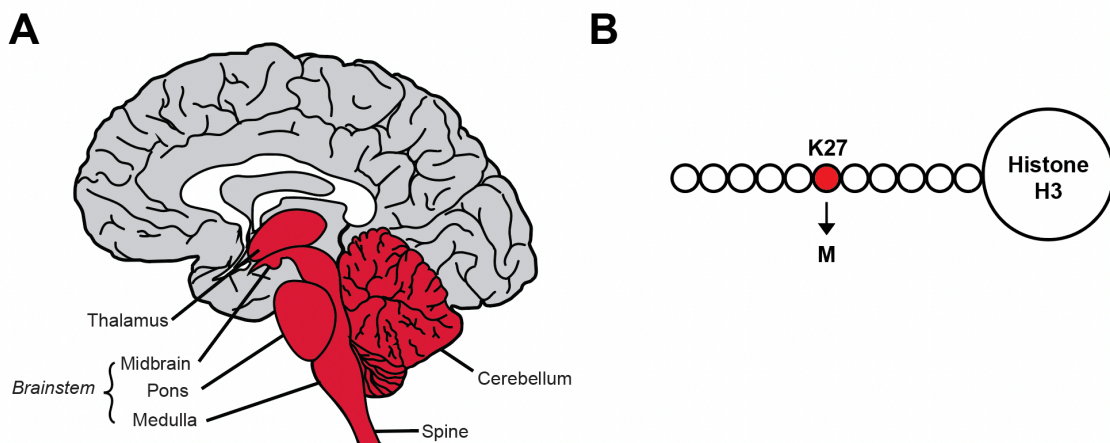


Figure 1.8 Diffuse midline glioma (H3K27-altered)

A) Diagram of the midline structures in which diffuse midline gliomas are known to arise.

B) Diagram representing the lysine-27 to methionine substitution mutation in histone H3 (H3K27M) that frequently occurs in diffuse midline glioma patients.

Initially, the histological and morphological similarities between paediatric and adult gliomas led to the assumption that the underlying disease biology was identical. However, landmark exome sequencing studies of paediatric gliomas revealed a striking difference. Approximately 80% of DIPG tumours harboured a somatic, heterozygous mutation in a gene encoding histone H3, resulting in a lysine-to-methionine substitution at position 27 (H3K27M) (Fig. 1.8B) (Schwartzentruber et al., 2012; Wu et al., 2012). Notably, this mutation is absent in adult tumours, highlighting a key distinction between adult gliomas and these paediatric gliomas. Tumour sequencing studies further demonstrated that when the H3K27M mutation is present, it is present in all tumour clones across various neuroanatomical locations, suggesting that it is an early initiating event in tumorigenesis (Nikbakht et al., 2016; Vinci et al., 2018). The presence of this mutation is also associated with a poorer prognosis in children, with H3K27M-positive tumours exhibiting significantly worse outcomes

compared to those with wild-type H3 tumours (Schwartzentruber et al., 2012). Over the past decade, molecular characterization of these tumours has led to an appreciation of their distinct molecular biology that sets them apart from counterparts, which is now recognized by the World Health Organization (WHO). In 2016, the WHO reclassified H3K27M-positive tumours as a unique disease entity – DMG (H3K27M-mutant) (Louis et al., 2016). However, it was later discovered that H3 wild-type DMG tumours aberrantly overexpress the PRC2 accessory subunit EZHIP, which acts as a H3K27M mimic (Hübner et al., 2019; Jain et al., 2020). Thus, these tumours use an alternative mechanism to achieve the same pathological molecular consequences as H3K27M. Subsequently, the nomenclature was changed in 2021 to reflect a growing understanding of the molecular drivers of the disease. This subclass of disease is now known as Diffuse Midline Glioma (H3K27-altered), acknowledging the convergent molecular consequences of H3K27M and EZHIP positive tumours (Louis et al., 2021). Due to the diffuse nature of these tumours and their anatomical location, complete surgical resection is not possible. The only current treatment is radiotherapy, which extends life by merely a few months at best (Clymer & Kieran, 2018; Vitanza & Monje, 2019). Thus, there is a pressing need to better understand the underlying disease biology and develop more effective targeted treatments.

There are 15 genes that encode histone H3, for which there are 3 variants. These H3 variants are classified into two main subtypes: replication-dependent canonical H3 (H3.1 and H3.2) and replication-independent non-canonical H3 (H3.3). Most H3-K27M alterations occur in the *H3F3A* gene (also known as *H3-3A*), which encodes the histone H3.3 variant (Mackay et al., 2017). While H3-K27M mutations can also be found in other H3-encoding genes, such as *HIST1H3B* (also known as *H3C2*) which encodes H3.1, these alterations are less common (Mohammad & Helin, 2017). Although all three histone variants are found throughout the genome, H3.3 is preferentially enriched at specific genomic regions, including transcriptional start sites, enhancers, and the gene bodies of actively transcribed genes (Mito et al., 2005).

The characteristics of the disease differ depending on which histone H3 variant harbours the K27M mutation. Specifically, tumours with H3.1-K27M mutations are restricted to the pons, whereas tumours with H3.3-K27M mutations can be more widespread, occurring in the brainstem and more rarely in other structures of the midline, such as the spinal cord (Mackay et al., 2017). Recent studies suggest that tumours with H3.1-K27M and H3.3-K27M mutations have distinct cells of origin (Jessa et al., 2022). Furthermore, the co-occurring mutations associated with H3.1-K27M and H3.3-K27M differ. H3.3-K27M mutations often co-occur with *TP53* and *PDGFRA* mutations, whereas H3.1-K27M mutations are more commonly associated with *ACVR1* mutations (Larson et al., 2019; Mackay et al., 2017). Research using an inducible mouse model has shown that H3K27M cooperates with these co-occurring mutations to accelerate glioma development (Larson et al., 2019). While H3K27M alone is not sufficient to initiate gliomagenesis, it is necessary for tumour establishment and maintenance. Studies have demonstrated that knock-down or knock-out of H3K27M in mice injected with human H3K27M glioma cells delays or prevents tumour formation, respectively (Harutyunyan et al., 2019;

Silveira et al., 2019). Notably, this is not true for all oncohistones. The H3G34R/V mutation, found in 30% of adolescent and young adult high-grade gliomas in the temporo-parietal cortex, is important for tumour formation but is dispensable for tumour maintenance once other oncogenic mutations have been co-opted (Chen et al., 2020). The fact that DMG (H3K27-altered) tumours occur in very discrete anatomical locations and generally within specific developmental windows suggest that the oncogenic effects of H3K27M/EZHIP are context dependent and linked to development. Attempts to model the disease *in-vivo* reinforce the importance of spatio-temporal context for H3K27M-driven gliomagenesis. Knock-in of H3K27M in mouse embryonic stem cells (mESCs) is embryonic lethal at the 4-cell stage (Pathania et al., 2017). Expression of H3K27M along with typical co-occurring mutations (*TP53* loss, *PDGFRA* overexpression) in human foetal neural stem cells (fNSCs) recapitulates the molecular features of the disease (Brien et al., 2021). The same three lesions cause fully penetrant tumour formation in mice when they are expressed in neural progenitor cells (NPCs) of developing embryos at E12.5 or E13.5 (Pathania et al., 2017). These tumours phenotypically and molecularly resemble DMG (H3K27-altered) tumours. Therefore, understanding the molecular mechanisms by which H3K27M contributes to gliomagenesis is clinically relevant to patients.

1.10 H3K27M induces a distinctive chromatin landscape which restrains differentiation

The precise mechanisms by which the H3K27M mutation drives DMG tumorigenesis remain unclear. However, it is known that the substitution of methionine for lysine at position 27 on the histone H3 protein disrupts post-translational modifications at this residue, as H3K27 is a substrate for both PRC2 and acetyltransferases. Quantitative mass spectrometry has revealed that H3K27M peptides comprise 3-18% of total histone H3 in DMG tumours (Lewis et al., 2013). Interestingly, immunoblots and mass spectrometry have revealed that despite the relatively low abundance, the presence of H3K27M leads to a significant increase in H3K27 acetylation (H3K27ac) levels and a profound loss of H3K27me2 and H3K27me3 levels, far exceeding an 18% change (Bender et al., 2013; Lewis et al., 2013; Venneti et al., 2014). These findings suggest that the H3K27M oncohistone exerts a dominant-negative effect on modifications at H3K27. Initially, it was hypothesized that the global loss of H3K27me3 and subsequent increase in H3K27ac, a modification associated with active promoters and enhancers, would lead to the de-repression of Polycomb target genes, which could drive disease progression. Surprisingly however, subsequent ChIP-seq studies have revealed that PRC2 and H3K27me3 are not entirely lost but are retained or even enriched at specific genomic sites (Bender et al., 2013; Brien et al., 2021; Harutyunyan et al., 2019; Mohammad et al., 2017). These are Polycomb target promoters are at neurodevelopmental genes and tumour suppressor genes, and the retention of H3K27me3 and PRC2 at these sites is associated with increased repression of these genes (Brien et al., 2021; Chan et al., 2013). These findings suggest that the H3K27M mutation induces a unique chromatin landscape characterized by global disruption of modifications at H3K27, accompanied by the aberrant repression of key Polycomb target sites.

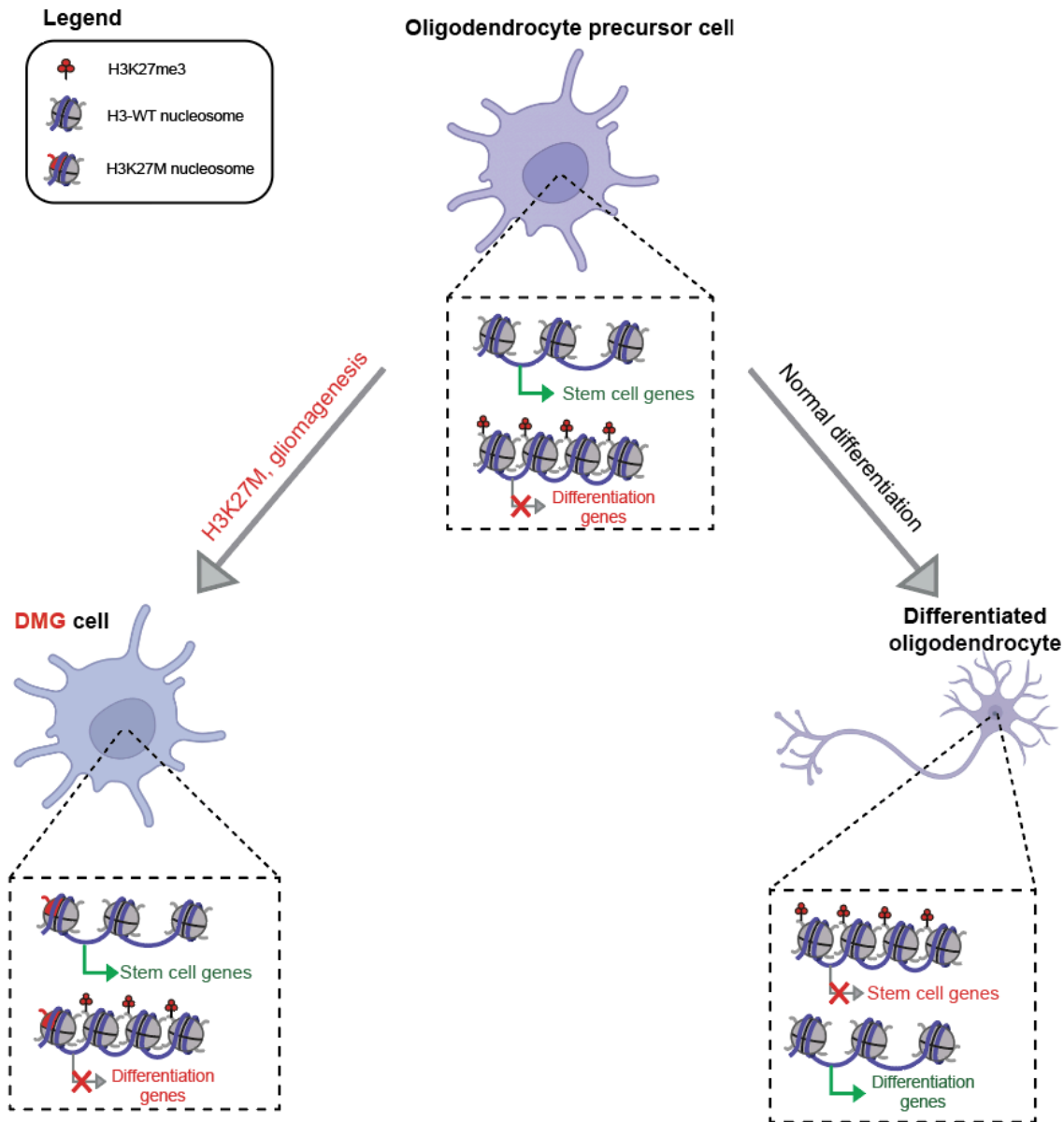


Figure 1.9 Dysregulation of PRC2 activity by H3K27M restrains differentiation

Schematic representation of normal neurodevelopment (right) and abnormal stalling of developmental programmes thought to occur in H3K27M-DMG (left). Three red circles indicate H3K27me3. Red histone indicates H3K27M.

Both *in vitro* structural studies and studies conducted in cell lines have provided evidence that PRC2 can directly interact with the H3K27M oncohistone peptide (Bender et al., 2013; Diehl et al., 2019; Jain et al., 2019). Furthermore, *in vitro* experiments have suggested that PRC2 may exhibit a higher affinity for H3K27M compared to wild-type H3 (Bender et al., 2013; Chan et al., 2013; Diehl et al., 2019). This interaction is thought to lead to the "stalling" of PRC2 at specific Polycomb target promoters, as captured by increased PRC2 ChIP-seq signal and retention of H3K27me3 levels at these sites (Brien et al., 2021; Jain et al., 2020). Notably, the PRC2 accessory component EZHIP, which is normally restricted to the gonads, can integrate into PRC2 complexes and inhibit their activity in a manner similar to H3K27M (Hübner et al., 2019; Jain et al., 2020). This is achieved

through a peptide that mimics H3K27M, thereby inducing the same characteristic modification landscape caused by H3K27M. Interestingly, EZHIP overexpression and H3K27M mutations are mutually exclusive, further suggesting that they may be functionally equivalent in terms of their impact on PRC2 activity. This aberrant chromatin landscape is thought to lock oligodendrocyte precursor cells (OPCs) – the most probable cell-type of origin - in a stem cell-like state and restrain differentiation (Fig. 1.9) (Brien et al., 2021; Jessa et al., 2022; Silveira et al., 2019). By locking OPCs in a self-renewing proliferative state, the likelihood of acquiring co-occurring oncogenic mutations is increased and this is thought to drive tumorigenesis.

While our understanding of the molecular consequences of the H3K27M oncohistone mutation have advanced greatly over the last decade, DMG (H3K27-altered) remains an incurable and effectively untreatable disease. Understanding how H3K27M affects PRC2 function remains an active area of research. Much remains to be elucidated about this characteristic chromatin landscape and exactly how it contributes to DMG gliomagenesis.

1.11 PRC2 is a promising therapeutic target in DMG (H3K27-altered)

The prevalence of recurrent heterozygous mutations in the catalytic SET domain of EZH2 in haematological malignancies led to the development of highly selective EZH2 inhibitors which competitively bind to the SET domain (Kim & Roberts, 2016). EZH2 inhibition reduces H3K27me3 levels in cells and EZH2 mutant lymphoma cells have been shown to have heightened sensitivity to EZH2 inhibition (Brien et al., 2016). This led to a number of clinical trials and ultimately FDA approval of an EZH2 inhibitor, Tazemetostat, for the treatment of certain patients with epithelioid sarcoma and patients with refractory follicular lymphoma (Straining & Eighmy, 2022).

Preclinical models of DMG (H3K27-altered) have demonstrated sensitivity to EZH2 inhibitors, with H3K27M-positive cells exhibiting greater sensitivity to treatment than H3 wild-type cells (Brien et al., 2021). The heightened sensitivity of H3K27M-positive cells to PRC2 inhibition is attributed to the local retention of H3K27me3 and PRC2 at key Polycomb target genes, including the *CDKN2A/INK4A* locus, which encodes the tumour suppressor p16^{INK4A} (Mohammad et al., 2017). ChIP experiments have shown that treatment with the EZH2 inhibitor Tazemetostat reduces H3K27me3 levels at the *CDKN2A/INK4A* promoter, leading to increased expression of p16^{INK4A} and decreased proliferation of H3K27M-positive cells (Mohammad et al., 2017). However, while Tazemetostat has shown promise, its poor blood-brain penetrability and modest response rates in preclinical models suggest that it is unlikely to be a transformative therapy for patients (Liang et al., 2022; Zhang et al., 2015). Furthermore, chronic treatment with EZH2 inhibitors poses significant challenges, including the risk of therapeutic resistance and secondary treatment-associated malignancies (Bracken et al., 2019; Brien et al., 2016). The chronic use of monotherapies is associated with a high risk of selecting for resistant cancer cells, and mutations conferring resistance to EZH2 inhibitors have been reported in other PRC2-dependent cancer types (Baker et al., 2015). Notably, deletion of the *INK4A* locus from a H3K27M-positive DMG cell line abrogated the response to Tazemetostat treatment, highlighting

the importance of this locus in mediating the antitumour effects of EZH2 inhibition and a potential resistance mechanism that could arise in patients (Mohammad et al., 2017). Additionally, PRC2 is a tumour suppressor in several malignancies, including childhood T-ALL, raising concerns that systemic inhibition of PRC2 may lead to secondary treatment-associated malignancies (Straining & Eighmy, 2022). Indeed, as more data is emerging there is already a documented risk of developing T-ALL in patients receiving long-term Tazemetostat treatment (Hsu et al., 2024). Therefore, identifying novel dependencies with fewer potential side effects is a critical and clinically important area of research.

1.12 Aims of this thesis

Although this disease is now classified as DMG (H3K27-altered), all of the work presented here has been carried out in DMG cells positive for the H3.3-K27M oncohistone. Therefore, although the underlying disease mechanisms in EZHIP-positive DMG are likely convergent, all findings and observations here will be specifically discussed in reference to H3K27M-DMG unless otherwise stated.

The primary aims of this research are to:

1. To identify genetic dependencies in H3K27M-DMG. The specific sub-aims to achieve this are:
 - Perform genome-wide CRISPR-Cas9 screens in H3K27M-DMG cells to identify genetic dependencies.
 - Validate identified dependencies using an orthogonal technique e.g. growth competition assays.
2. To understand the effects of H3K27M on Polycomb group proteins beyond PRC2. The specific sub-aims to achieve this are:
 - Generate or obtain isogenic H3K27M and H3K27M-KO cells.
 - Quantitatively map PRC1 and PRC2 subunits and associated histone modifications in H3K27M/H3K27M-KO isogenic cells to get a full and comprehensive picture of how H3K27M disrupts Polycomb repressive complexes.
3. To understand the role of an identified genetic dependency in H3K27M-DMG biology. The specific sub-aims to achieve this are:
 - Perform a gene-focused CRISPR-Cas9 tiling screen of a genetic dependency to identify functionally relevant domains for disease biology.
 - Validate these results with functional and/or chemical perturbation of the relevant domains.

CHAPTER 2

Materials & Methods

2.1 Cell culture

2.1.1. Cell line culture

HSJD-DIPG-021 and HSJD-DIPG-007 cell lines were derived at the Hospital de Sant Joan de Déu Barcelona and were gifted from Dr Ángel Montero Carcaboso. HSJD-DIPG-021 cell line was derived from a tumour biopsy of a 5.9-year-old girl. HSJD-DIPG-007 cell line was derived from the autopsy of a 9.9-year-old boy. Immunohistochemistry found both cell lines to be negative for H3K27me3. Sequencing of both cell lines found them to be positive for a H3K27M mutation in the *H3F3A* gene. Sequencing also revealed HSJD-DIPG-007 to be positive for a R206H mutation in *ACVR1*. Isogenic SU-DIPGXIII H3K27M and H3K27M-KO cells were a kind gift from Prof. Nada Jabado (McGill University). SU-DIPGXIII cell line was derived and characterised at Stanford University by Prof. Michelle Monje and colleagues (Grasso et al., 2015). The cell line was derived from the autopsy of a 6 year old girl. Sequencing found it to be positive for a H3K27M mutation in *H3F3A* (Nagaraja et al., 2017). Prof. Nada Jabado's group performed CRISPR/Cas9 editing of H3K27M and confirmed knockout using sequencing and immunoblotting (Harutyunyan et al., 2019). Mouse neural stem cells (mNSCs) were derived from ESCs following the standard differentiation protocol previously described (Conti et al., 2005).

All cell cultures were maintained in a humidified, normoxic incubator at 37°C, 5% CO₂. Cell lines were cultured as adherent monolayers in CellBind flasks and plates (Corning). Cells were initially grown in neural stem cell (NSC) medium with laminin (Bio-Techne). NSC medium consists of 1:1 DMEM-HAMS/F12 (Sigma D8437) supplemented with Glucose solution (Sigma G8644), MEM Non-Essential Amino Acids (Gibco 11140-035), Pen-Strep (Gibco 15140-122), B27 without vitamin A (Gibco 17504-044), N2 supplement (Gibco 17502-048), mouse EGF (10ng/mL; Peprotech) and Human FGF (10ng/mL; Peprotech). Accutase (Sigma A6964) was used to detach cells during passaging.

2.1.2 Lentivirus production

HEK293T cells were grown in DMEM-high glucose (Sigma) supplemented with 10% (v/v) FBS (Gibco), 100 U/mL Penicillin and 100µg/mL Streptomycin (Gibco). Trypsin-EDTA (Gibco) was used for passaging HEK293T cells.

Lentiviral supernatants were generated by co-transfecting HEK293T cells with a lentiviral transfer plasmid encoding the insert of interest, a viral packaging plasmid encoding PAX2 and a viral envelope plasmid encoding VSV-G; using polyethylenimine in Opti-MEM (Gibco). Approximately 48 and 72 hours post-transfection, viral supernatants were collected and filtered through a 0.45µm syringe filter. Filtered supernatants were then centrifuged at 20,000G for 2 hours at 4°C in a Beckman Coulter Optima L-100XP using a SW 22 Ti rotor. Pelleted lentivirus was resuspended in warm NSC media. Resuspended lentivirus was then to cell culture media with Polybrene (8.5µg/mL). Cells centrifuged at 1000rpm for 15 mins and then kept in an incubator for approximately 2 hours before

the media was removed and replaced with fresh NSC media. Cells were subject to selection 24 hours post-transfection.

2.2 Protein methods

2.2.1 Preparation of whole cell protein lysates & immunoblotting

Cells were washed twice with cold (4°C) PBS in-situ and an appropriate volume of RIPA buffer (10mM Tris-HCl pH 8, 1mM EDTA, 1% (v/v) Triton X-100, 0.1% (v/v) sodium deoxycholate, 0.1% (v/v) SDS and 140mM NaCl) supplemented with HALT Protease Inhibitor . Cell lysates were incubated on ice for 15 minutes followed by sonication and a further 15 minutes on ice before being spun down at full speed for 15 minutes at 4°C. Bradford assays were performed to determine protein concentration and samples were normalised to the same concentration. Denatured protein lysates were separated by SDS-PAGE and transferred to nitrocellulose membranes (Amersham). Membranes were blocked against nonspecific binding by rocking in Odyssey Blocking Buffer PBS (Li-Cor) for ~1 hour. Membranes were then incubated in the appropriate primary antibody (Appendix IV) overnight at 4°C. The membrane was washed 6x 5minutes in PBST (0.1% Tween) before being incubated with the appropriate secondary antibody (labelled with near-infrared dye) for an hour. Membranes were then washed 6x 5minutes in PBST (0.1% Tween) before a final wash in PBS. Membranes were imaged using a Li-Cor Odyssey Fc system. Relative proteins levels were quantified using Image Studio software (Li-Cor).

2.3 CRISPR-Cas9 screening and validation

2.3.1 Genome-wide Brunello CRISPR-Cas9 knock-out screens

Cas9-expressing HSJD-DIPG-007 and HSJD-DIPG-021 cells were transduced with the Brunello lentiviral library (Doench et al., 2016) at a target MOI ~0.3 at 350-500x library coverage. Approximately 24 hours following transduction, puromycin was added to cell cultures to select for sgRNA expressing cells. Cell pellets were harvested for genomic DNA extraction at Day 21 or Day 28 post-transduction. Genomic DNA was extracted using the NucleoSpin Blood L (Macherey-Nagel), following the manufacturer's protocol. Prior to PCR, the OneStep PCR Inhibitor Removal Kit (Zymo Research) was used, following the manufacturer's protocol. PCR using lentiGuide-Puro specific primers (from Broad GPP) and PCR purification was performed as per tiled pooled PRC2 screen above. Individually indexed libraries were pooled and sequenced using an Illumina NextSeq500.

2.3.2 CBX4 CRISPR-Cas9 tiling screens

For the *CBX4* gene tiling CRISPR-Cas9 library, a total of 385 sgRNA sequences targeting the coding exons of the *CBX4* were designed using the Genetic Perturbation Platform (Broad Institute) (Doench et al. 2016). Briefly, guide RNA oligos were synthesized by microarray (CustomArray) and cloned into the ipUSEPR lentiviral sgRNA vector (hU6-driven sgRNA co-expressed with EF-1a-driven red fluorescent protein (RFP) and puromycin-resistance gene) using the BsmBI (NEB) restriction sites.

Cas9-expressing HSJD-DIPG-007 and HSJD-DIPG-021 were transduced with the CBX4 sgRNA lentiviral library at a multiplicity of infection (MOI) of 0.3 at 1000x library coverage. Approximately 24 hours following transduction, puromycin was added to cell cultures to select for sgRNA expressing cells. Cell pellets were harvested for genomic DNA extraction at Day 4 and Day 28 post-transduction. Genomic DNA was extracted using the Qiagen QIAmp DNA Blood Mini Kit, following the manufacturer's protocol. PCR using ipUSEPR specific primers (DCF01 5'-CTTGTGGAAAGGACGAAACACCG-3' and DCR03 5'-CCTAGGAACAGCGGTTTAAAAAAGC-3') was performed: 50µL DreamTaq Hot Start Green PCR Master Mix (Thermo Scientific), 0.5µL P5 primer mix (100µM), 0.5µL P7 index primer (100µM) using the following cycling conditions: 95°C for 60s, 28 cycles of 95°C for 30s, 53°C for 30s, 72°C for 30s; 72°C for 10min. Multiple PCRs on 5 µg input gDNA were performed for each gDNA pool. PCR products were subject to agarose gel electrophoresis, gel purified, and isopropanol precipitated. Individually indexed libraries were pooled and sequenced using an Illumina NextSeq500.

2.3.3 CRISPR-Cas9 screen validations – growth competition assays

Cas9-expressing HSJD-DIPG-007 and HSJD-DIPG-021 cell lines were transduced with sgRNA-GFP expressing lentivirus at a low MOI and passaged without selection. The percentage of GFP positive cells was measured at each passage using the Guava easyCyte Flow Cytometer (Cytek). The relative proportion of GFP positive cells was monitored throughout this serial cell culture. For CBX4 rescue negative selection experiments HSJD-DIPG-007 cells were transduced with lentivirus expressing pLEX306 codon-optimised CBX4 constructs and selected with puromycin. Purified cultures were transduced with sgRNA-GFP expressing lentivirus at a low MOI and passaged with selection. The percentage of GFP positive cells was monitored at each passage using the Guava easyCyte Flow Cytometer (Cytek). Sequences of sgRNAs used in these experiments can be found in Appendix V.

2.4 Genomics & transcriptomics

2.4.1 CUT&RUN-Rx

Cells were harvested using Accutase (STEMCELL Technologies) and counted. Following counting mouse NSCs were spiked-in to human DIPGXIII cell suspensions (1:10 cell number ratio). Cells were fixed in culture media containing 0.1% formaldehyde at room temperature for 1 min. Formaldehyde was quenched with glycine at 0.125 M, followed by a 5-min incubation at room temperature. Fixed cells were washed with PBS. Fixed cells were incubated on ice for 10 min in nuclear extraction buffer (20mM HEPES pH 7.5, 10mM KCL, 0.1% Triton X-100, 20% Glycerol, 1X protease inhibitor cocktail and 0.5mM spermidine). Extracted nuclei were collected by centrifugation (4°C, 600g) and resuspended in cold nuclei extraction buffer (100µL per sample). CUT&RUN was performed following Epicpypher CUT&RUN Protocol v1.5.1. CUT&RUN DNA was purified using the Monarch PCR & DNA Cleanup Kit (New England Biolabs) in accordance with the manufacturer's instructions. Primer sequences used for quantitative PCR analysis can be found in Appendix VI and VII.

2.4.2 CUT&RUN-Rx sequencing library preparation

Purified DNA was quantified using a Qubit fluorimeter (Invitrogen). 1-10 ng of DNA was used to generate CUT&RUN-Rx libraries with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) as per the manufacturer's instructions. Library DNA was quantified using the Qubit, and size distributions were ascertained on a TapeStation (Agilent) using the D1000 ScreenTape assay reagents (Agilent; 5067- 5583). This information was used to calculate pooling ratios for multiplex library sequencing. Pooled libraries were diluted and processed for 38-bp paired-end sequencing on an Illumina NextSeq500 instrument using the NextSeq 500 High Output v2 kit (Illumina; FC-404-2005) in accordance with the manufacturer's instructions.

2.4.3 Quant-seq

Total RNA was isolated from untreated DIPGXIII-K27M/KO cells, or DIPGXIII-K27M following 3, 6 or 9 days of Tazemetostat (2 μ M) treatment. The quality of extracted RNA was confirmed using the TapeStation (Agilent) with the RNA ScreenTape assay reagents (Agilent; 5067-5576). Total RNA (500 ng) was used as input material from each sample for library preparation. Libraries were generated using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen; 015.24) in accordance with the manufacturer's instructions. Library DNA was quantified using the Qubit, and size distributions were ascertained on a TapeStation (Agilent) using the D1000 ScreenTape assay reagents (Agilent; 5067-5583). This information was used to calculate pooling ratios for multiplex library sequencing. Pooled libraries were diluted and processed for 75-bp single-end sequencing on an Illumina NextSeq instrument using the NextSeq 500 High Output v2 kit (75 cycles) (Illumina; FC-404-2005) in accordance with the manufacturer's instructions.

2.4.4 Quantitative PCR

For absolute quantification of *H3F3A* copy number, genomic DNA from HMEC primary human cells and DIPG cell lines was isolated using the QiaAMP DNA Blood Minikit. Concentrations were measured using a Nanodrop ND-1000 spectrophotometer. Primers were designed and confirmed to highly specifically amplify the endogenous *H3F3A* sequence. Dilution series of the HMEC gDNA were performed to generate standard curves using SYBR Green I detection chemistry (NEB) on a QuantStudio 3 Real-Time PCR System (Applied Biosystems). These standard curves were used to estimate the copy number of *H3F3A* from each cell line.

For CUT&RUN qPCRs, 5 μ L of purified CUT&RUN DNA was added to a PCR mix consisting of H₂O, primers and SYBR Green I detection chemistry (NEB). CT values were measured on a QuantStudio 3 Real-Time PCR System (Applied Biosystems).

For quantification of mRNA transcript abundances, total RNA was extracted from cells using the RNeasy kit (Qiagen) according to the manufacturer's protocol. RNA was used to generate cDNA by reverse transcriptase PCR using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Relative mRNA expression levels were determined using SYBR Green I detection

chemistry (Applied Biosystems) on the QuantStudio instrument. Primer sequences used in quantitative PCR experiments can be found in Appendix VII and VIII.

2.5 Data analysis

2.5.1 Genome-wide CRISPR-Cas9 knock-out screen data analysis

The MAGeCK algorithm was used to calculate CRISPR dependency beta scores comparing sgRNA abundance in the original plasmid pool with gDNA extracted at the end of a screen, using the count and mle functions (Wang et al., 2019). In the genome-wide CRISPR dependency score plots, the frequency of CRISPR dependency scores was calculated by rounding beta scores for each gene to two decimal places and counting the number of times each of the rounded scores was present. Common essential genes, as defined by depmap.org (Achilles, 19Q3), were removed and plots were generated using Prism.

2.5.2 CBX4 tiling screen data analysis

For the CBX4 tiling screen, sgRNA read counts were generated using PinAPL-Py (Spahn et al., 2017). sgRNAs with less than 100x representation at the Day 4 timepoint and sgRNAs flagged as off-target by CRISPick (Broad Institute) were removed. Log₂ fold-changes for each sgRNA were calculated by comparing the number of reads mapping to individual sgRNAs at day 4 and day 28 normalised to library sequencing depth. The Log₂ fold changes for each sgRNA were analysed using the publicly available pipeline ProTiler (He et al., 2019) to generate plots and call statistically significant hypersensitive regions.

2.5.3 CUT&RUN-Rx data analysis

The prefix mm10_ was added to the chromosome names of the mm10 genome build. Next, a human/mouse hybrid genome was generated by appending the human hg38 genome build to the mm10 genome build prior to indexing with bowtie2. Bowtie2 was used to align reads to the hybrid genome in the --very-sensitive mode. SAMtools was employed to retain reads with a mapping quality >2, separate the human and mouse reads, remove duplicate reads and convert SAM files to BAM files. The spike in normalisation factor for each target was calculated using the formula for normalised reference-adjusted reads per million (RRPM) (1 per million spike-in reads). The bamCoverage module from the deepTools suite was used to generate RRPM normalised bigwig files for genome browser visualisation with the following parameters where SCALEFACTOR is equal to the respective RRPM normalization factor. Peaks were called using macs2, BroadPeak files were converted to bed format using standard bash commands. The bedtools intersect module was used to exclude regions that overlapped with a custom hg38 blacklist, and to overlap peaks to generate consensus peak sets from replicate experiments. Peaks were annotated with HOMER.

2.5.4 Quantseq data analysis

The BBDuk module from the BBTools package was used to perform adapter trimming and quality filtering as recommended by the manufacturer. The STAR aligner was used to align filtered reads to

the hg38 human genome build. The htseq-count tool was used to calculate the number of reads that align to each gene. Differentially expressed genes were detected using a >1.5 fold change and a <0.05 Benjamini-Hochberg adjusted P value using the DESeq2 R package.

Please note that all protocols were performed at room temperature unless otherwise stated.

CHAPTER 3

CBX4 and *PCGF4* are genetic dependencies in
H3K27M-DMG

3.1 Introduction

'Genetic dependencies' are genes that are essential for the proliferation or survival of a cell. The goal of precision medicine is to provide specific, targeted therapies tailored to the disease in order to limit harmful side-effects and increase survival (Chakravarty & Solit, 2021). Identifying disease-specific genetic dependencies is an important first step in this process. As a result, discovering cancer type specific genetic dependencies has been one of the main goals of preclinical cancer research in recent years (Lin & Sheltzer, 2020). This endeavour has been greatly expedited in recent years with the advent of genetic screening approaches, particularly those utilising CRISPR-Cas9 based technologies. The goal of this research chapter is to identify novel genetic dependencies in H3K27M-DMG and to prioritise the most promising candidates for further investigation. At the time of conception of this project, no genome-wide screens had yet been published for any H3K27M-DMG cell line. Thus, these screens would be amongst the first to potentially identify novel genetic dependencies in H3K27M-DMG.

CRISPR-Cas9 knock-out screens provide an unbiased method to scan for genetic dependencies by using CRISPR-Cas9 to inactivate genes. A single guide RNA (sgRNA) targets the Cas9 nuclease to a specific site in the genome via sequence complementarity with the target locus. Here, Cas9 recognises a 3bp sequence called the protospacer adjacent motif (PAM). Cas9 cleaves the DNA 3bp upstream of the PAM, inducing a double-stranded break and triggering cellular DNA repair mechanisms (Barrangou & Doudna, 2016). Double-stranded breaks are most frequently resolved by a DNA repair mechanism known as non-homologous end-joining (NHEJ). NHEJ is an error prone process and usually results in insertions or deletions ('indels') at the site of repair (Wang & Doudna, 2023). If NHEJ occurs in a protein-coding gene, indels can have a range of consequences on the coded protein. Non-synonymous mutations can change the amino acid sequence of the protein, which can disrupt protein folding and function. Frameshift mutations can change the entire amino acid sequence of the protein downstream from the mutation, majorly disrupting the coded protein. Alternatively, frameshift mutations can introduce a premature stop codon which results in a truncated protein, or no protein if the mRNA transcript is degraded by nonsense-mediated decay.

By pooling a library of sgRNAs targeting different genes, CRISPR-Cas9 screens can be used to test thousands of genetic perturbations simultaneously (Doench, 2018). To begin, a plasmid library is introduced into a population of cells such that each cell receives only one plasmid. Each plasmid contains an sgRNA and a selectable marker. Lentivirus is the typical vehicle for this step, as lentivirus integrates into the genome of infected cells. This prevents the dilution of the sgRNAs over time as the cells divide, permanently tagging the genome of the parent and daughter cells with the sgRNA sequence. Infected cells are selected then passaged for several weeks. At the end of the screen, the sgRNA sequences in the population are retrieved using PCR followed by high-throughput sequencing. The frequency of each sgRNA in the population at the end of the screen is then compared back to the frequencies at the very beginning of the screen. The sgRNAs targeting genetic

dependencies will be significantly depleted from the population as they reduce the fitness of cells carrying those perturbations within the wider population (Doench, 2018).

3.2 Results

3.2.1 Genome-wide CRISPR-Cas9 knock-out screens validate core PRC2 members as genetic dependencies in H3K27M-DMG

The aim of this research chapter was to identify genetic dependencies in H3K27M-DMG cell lines using an unbiased CRISPR-Cas9 screening approach. The Brunello genome-wide pooled sgRNA library was chosen for this. The Brunello genome-wide pooled sgRNA library which contains 76,441 sgRNAs targeting 19,114 human genes (Doench et al., 2016). Each gene is targeted by 4 distinct sgRNAs. The sgRNAs in this library were specifically designed and optimised to minimise off-target activity. To rule out cell-line specific dependencies, the genome-wide CRISPR-Cas9 knock-out screens were performed in two H3.3K27M-positive DMG cell lines. The screens were performed in HSJD-DIPG-021 and HSJD-DIPG-007 cell lines by myself and Dr Dáire Gannon respectively (Fig 3.1A). Results are shown for HSJD-DIPG-021 unless otherwise stated. Some crucial preparation and optimisation steps were performed by senior lab members prior to the commencement of this project. A lentiviral preparation of the Brunello CRISPR-Cas9 knockout library was made by Dr Gerard Brien. HSJD-DIPG-007 and HSJD-DIPG-021 cell lines were previously engineered to express Cas9 by Dr Dáire Gannon.

To assess the technical success of the screen, the change in relative abundance of the positive and negative control sgRNAs were assessed (Fig 3.1B). Genes that function in cellular processes that are essential to many or all cell types commonly score as strong genetic dependencies in CRISPR-Cas9 knockout screens, irrespective of the cell or tissue type in which the screen is performed. Such genes are referred to as “pan-essential genes”. Examples include housekeeping genes and genes involved in processes such as DNA replication or transcription. sgRNAs targeting pan-essential genes (e.g. *CDK1*, *CDK9*, *POLR2D*, *POLR2A*, *RPA3*, *PCNA*, *RPL23A* and *RPL9*) served as positive controls and most depleted robustly in both replicates of the screen (Fig 3.1B). Non-targeting sgRNAs serve as negative controls and should not deplete over the course of the screen. The Brunello sgRNA library contains 1000 non-targeting sgRNAs. As expected, the overwhelming majority of these negative control sgRNAs did not change in abundance over the course of the screen (Fig 3.1B).

The suggested minimum sequencing depth for CRISPR-Cas9 screens is 100-200 reads per sgRNA (Bodapati et al., 2020). Both replicates were sequenced to a coverage of approximately 330 – 460 reads per sgRNA. The changes in sgRNA abundance were highly correlated (Pearson correlation coefficient; $r = 0.666$) across replicates, demonstrating that there was minimal technical noise in the experiment (Fig 3.1B). Taken together, these quality control measures confirm the screens were technically sound and noise due to technical variation was minimised. The same quality control

measures performed by Dr Dáire Gannon confirmed that the HSJD-DIPG-007 screen was similarly technically robust.

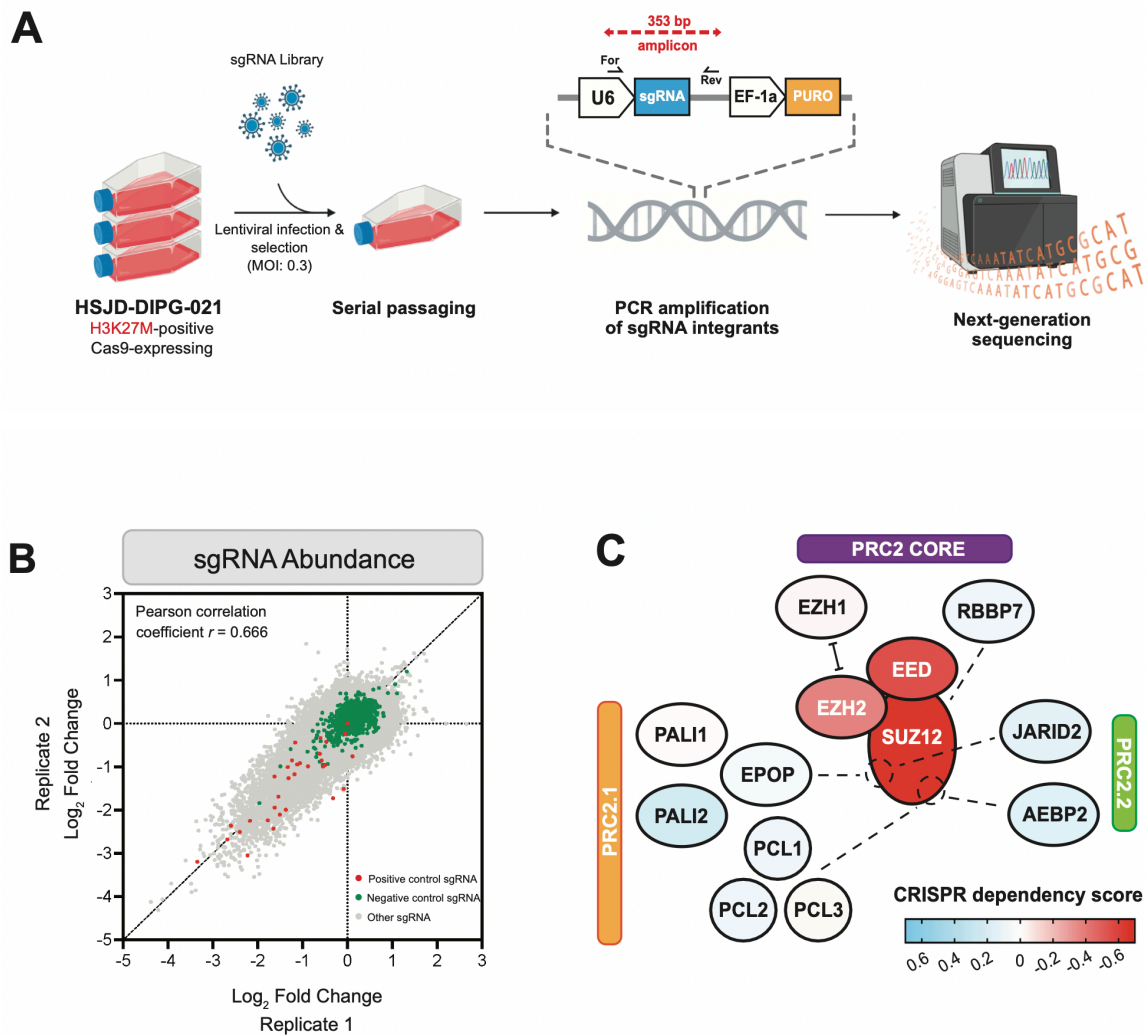


Figure 3.1 Genome-wide CRISPR-Cas9 knock-out screen confirms PRC2 catalytic core components are genetic dependency in H3K27M DMG.

A) Schematic of the genome-wide CRISPR screen experiment design (created with Biorender.com).

B) Scatterplot showing correlation $\log_2$ fold change in sgRNA abundance between 2 replicates of the genome-wide screen performed in HSJD-DIPG-021 (Pearson correlation coefficient $r = 0.666$). Green data points indicate the 1000 non-targeting negative control sgRNAs. Red data points indicate positive control sgRNAs targeting commonly known pan-essential genes (*CDK1*, *CDK9*, *POLR2D*, *POLR2A*, *RPA3*, *PCNA*, *RPL23A* and *RPL9*). Grey datapoints indicate all other sgRNAs in the Brunello sgRNA library.

C) Heatmap representation of CRISPR dependency score for PRC2 components from the genome-wide CRISPR screen. A positive score (blue) indicates positive selection, a negative score (red) indicates negative selection i.e. a genetic dependency.

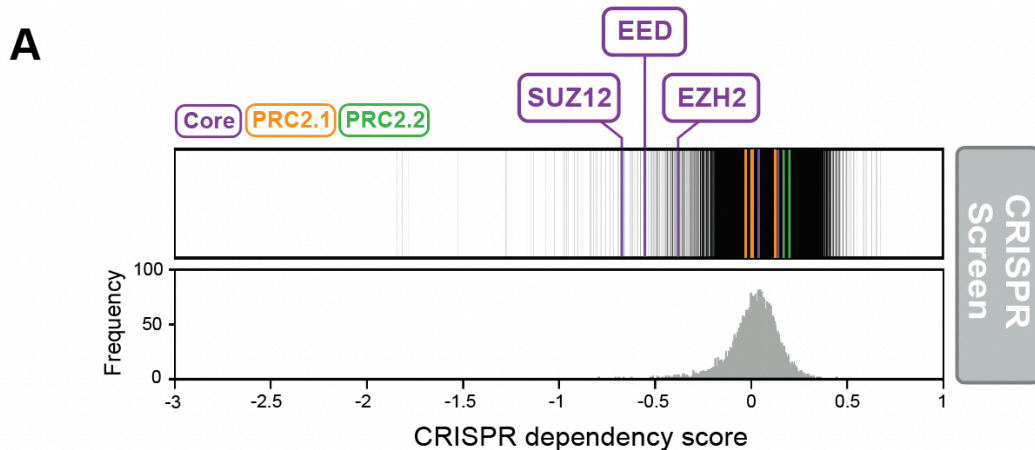
MaGeCK is a commonly used pipeline for analysing CRISPR-Cas9 knock-out screen data (Wang et al., 2019). MaGeCK integrates several statistical components (including log-fold change of sgRNA abundance and p-values) into one numerical value for each gene called a 'beta score'. Collapsing multiple statistical components into one numerical value for each gene allows for a more direct and mathematically robust method of comparing gene essentiality than log fold-change alone. Beta

scores are often referred to as 'CRISPR dependency scores' in published literature. In line with this nomenclature, beta scores will hereby be referred to as CRISPR dependency scores. The CRISPR dependency score for a gene describes the selection that is occurring when that gene is knocked out – a positive score indicates positive selection, while a negative score indicates negative selection. Dr Craig Monger used the MaGeCK algorithm to calculate CRISPR dependency scores for all 19,114 genes targeted by the Brunello library.

As discussed in Chapter 1, multiple studies have shown that H3K27M-positive cells are more sensitive to PRC2 inhibitors than H3 wild-type cells. I first compared the CRISPR dependency cores of PRC2 members. Based on this previous work, sgRNAs targeting the PRC2 core catalytic components were expected to undergo negative selection over the course of the screen. This was observed for *EED*, *SUZ12* and the gene encoding the catalytic subunit *EZH2*, but not its less catalytically active paralog *EZH1* (Fig 3.1C). This underscores the importance of PRC2 catalytic activity in H3K27M-DMG disease biology.

While the dependence of H3K27M-DMG cell on PRC2 catalytic activity was previously established (Brien et al., 2021; Harutyunyan et al., 2019; Mohammad et al., 2017), it was not known if H3K27M-DMG biology relies on any other PRC2 functions, or the activity of any specific PRC2 subcomplex. In contrast to sgRNAs targeting PRC2 catalytic components, sgRNAs targeting non-enzymatic PRC2 auxiliary components did not experience negative selection over the course of the screen (Fig. 3.1C). This suggests that H3K27M-DMG biology does not rely on any specific form of PRC2.1 or PRC1.2. However, it is important to note that there may be some functional redundancy between auxiliary subunits (e.g. PCL proteins 1-3). As only one gene is targeted per cell, it is therefore not possible to conclude that H3K27M-DMG biology does not rely on PRC2.1 or PRC2.2 complexes more generally. In order to test this, sgRNAs targeting multiple genes (e.g. all three *PCL* genes) would have to be delivered simultaneously to cells.

Known pan-essential genes as defined by the Cancer Dependency Map project were then removed from the dataset. Project Achilles is an endeavour by the Cancer Dependency Map project which continuously catalogues pan-essential genes identified in CRISPR-Cas9 knockout screens. While a more up-to-date list may currently exist, please note that the most recent list at the time (Achilles 19, Q3) was used to filter out pan-essential genes. The remaining 16,801 genes were then ranked according to CRISPR dependency score from lowest (strongest genetic dependency) to highest (no dependency). Core PRC2 genes *EED*, *SUZ12* and *EZH2* were amongst the strongest overall genetic dependencies (Fig. 3.2A), ranking in the top percentile of genes targeted, excluding pan-essential genes (Fig 3.2B).



B

	Gene	Rank	Percentile
CORE	SUZ12	55	1
	EED	86	1
	EZH2	203	2
	EZH1	8905	54
	RBBP7	14523	87
PRC2.1	PAL1	6068	37
	PAL12	16578	99
	PCL1	12987	78
	PCL2	13959	84
	PCL3	4616	28
PRC2.2	EPOP	11878	71
	JARID2	15480	93
	AEBP2	16094	99

Figure 3.2 PRC2 catalytic core components rank amongst the top hits in the genome-wide CRISPR-Cas9 screen.

A) Heatmap (top) and frequency distribution (bottom) for all genes grouped in bins according to CRISPR dependency score. CRISPR dependency scores were rounded to 3 decimal places, with all genes with the same rounded CRISPR dependency score value allocated to the same bin.

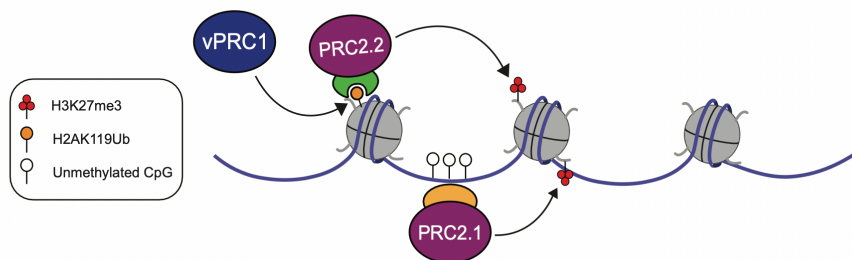
B) Table indicating the rank (middle column) and percentile rank (right column) for each PRC2 gene calculated using the MaGeCK analysis pipeline.

3.2.2 Genome-wide CRISPR-Cas9 knock-out screens reveal specific canonical PRC1 subunits CBX4 and PCGF4 are genetic dependencies in H3K27M-DMG

Given the established dependency on PRC2 catalytic activity, and the functional interplay between PRC2 and PRC1 (discussed in Chapter 1), a natural question arose – are any PRC1 components genetic dependencies in H3K27M-DMG? If so, could this give us any more insight into the mechanism by which Polycomb contributes to H3K27M-DMG disease biology?

Variant PRC1 (vPRC1) complexes are known to be structurally and functionally very diverse complexes. vPRC1 acts upstream of PRC2, depositing H2AK119Ub which can recruit PRC2.2 (Fig 3.3A). However, neither of the PRC1 catalytic core components (RING1A/B) scored as genetic dependencies (Fig 3.3B). Similarly, no PRC1 components exclusive to vPRC1 complexes scored as genetic dependencies (Fig 3.3B). Please note that not all possible vPRC1 auxiliary components are shown in Fig 3.3B as not all were represented in the Brunello sgRNA library. This is also true for *RYBP*, hence it is not included in Fig 3.3B. While vPRC1 complexes are known to have different functions depending on the incorporation of distinct unrelated auxiliary proteins, this suggests that these complexes are not functionally important to H3K27M-DMG biology. While it is clear that there is not a functional dependence on any one specific vPRC1 complex defined by specific auxiliary subunit, it is worth noting that there may be compensation between subunits for which paralogs exist – e.g. PCGFs, or RING1A/B. Paralogs of the same gene are likely to be structurally and functionally more similar to each other than unrelated proteins, which could mask a functional dependence on vPRC1. For example, vPRC1 complexes are generally considered to be more catalytically active than canonical (cPRC1) complexes. Unlike PRC2 complexes in which the catalytic subunit EZH2 is more catalytically active than its paralog EZH1, there is no obvious difference in catalytic activity between the paralogs RING1A and RING1B. Therefore, as only one paralog is targeted in the cell, it is possible that the other may compensate. As such, while neither RING1A or RING1B alone are genetic dependencies in H3K27M-DMG, it is impossible to determine the importance of H2AK119Ub catalysis in H3K27M-DMG biology from this experiment.

A



B

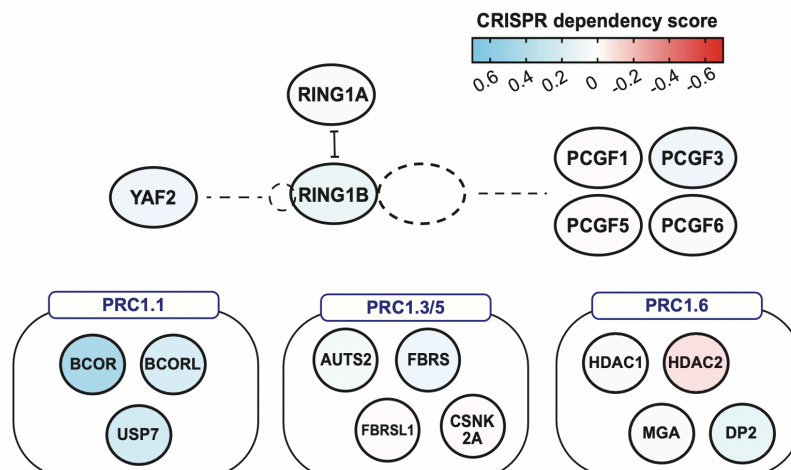


Figure 3.3 Genome-wide CRISPR-Cas9 knock-out screen indicates vPRC1 components are not genetic dependencies in H3K27M DMG.

A) Schematic of the interplay between vPRC1 and PRC2. vPRC1 acts upstream of PRC2 and can contribute to PRC2.2 recruitment to chromatin.

B) Heatmap representation of CRISPR dependency score for vPRC1-specific components from the genome-wide CRISPR screen. A positive score (blue) indicates positive selection, a negative score (red) indicates negative selection i.e. a genetic dependency. Note that there are no sgRNAs targeting *RYBP* in the Brunello library, hence only its paralog *YAF2* is represented. *PCGF2* and *PCGF4* are not shown (bottom) as they are also core components of cPRC1 and thus are not exclusive to vPRC1.

Canonical PRC1 (cPRC1) acts downstream of PRC2 and is recruited to chromatin via binding of the CBX subunit to H3K27me3 (Fig 3.4A). Two cPRC1 component genes – *CBX4* and *PCGF4* – immediately stand out as genetic dependencies (Fig 3.4B). Given the established dependency on PRC2 catalytic activity, it is perhaps unsurprising that some cPRC1 components score as genetic dependencies. Remarkably however, paralogs for these proteins exist and until now these paralogs have largely been presumed to be functionally redundant. This very specific dependence on *CBX4* and *PCGF4* suggests that these cPRC1 complexes are functionally distinct, that these functions are essential to H3K27M-DMG biology, and these functions cannot be compensated for by closely related paralogs.

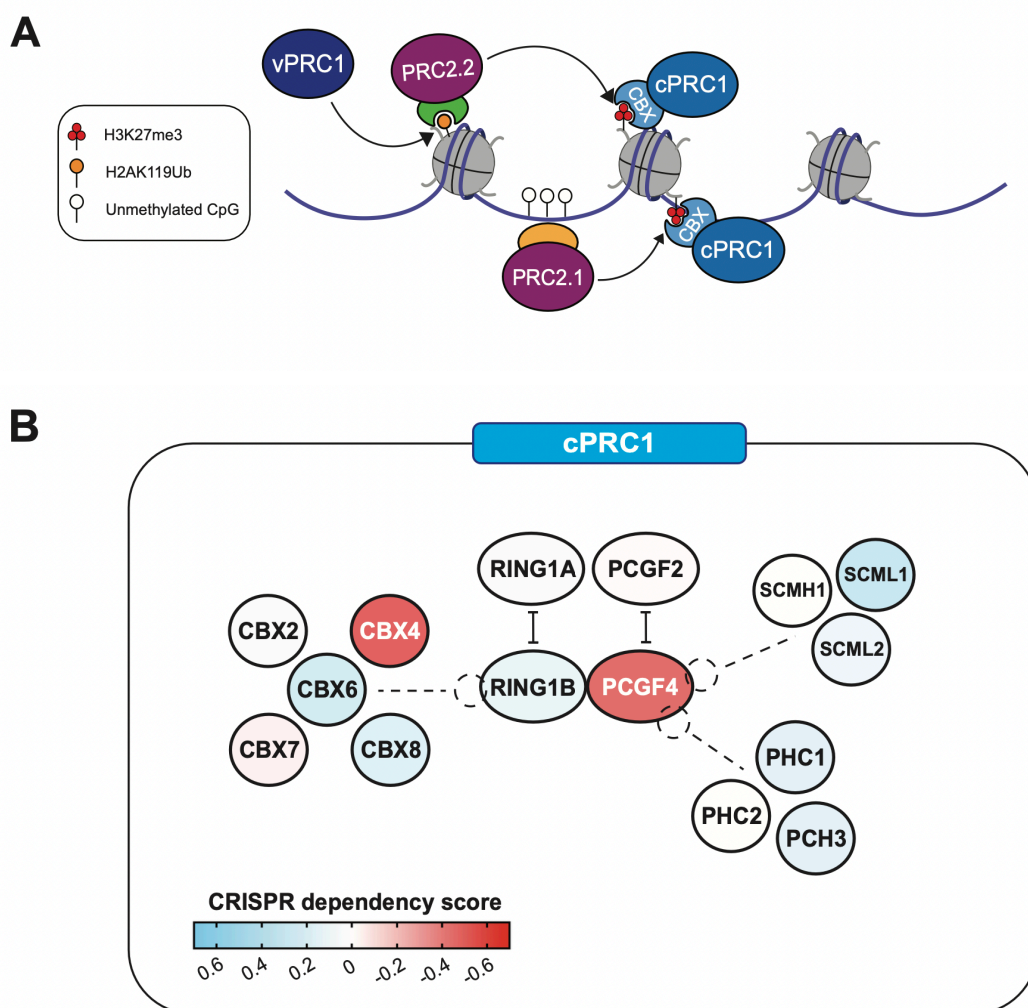


Figure 3.4 Genome-wide CRISPR-Cas9 knock-out screen indicates specific cPRC1 components CBX4 and PCGF4 are genetic dependencies in H3K27M DMG.

A) Schematic of the interplay between vPRC1, PRC2 and cPRC1. cPRC1 acts downstream of PRC2 and is recruited to chromatin via PRC2 catalytic activity.

B) Heatmap representation of CRISPR dependency score for cPRC1 components from the genome-wide CRISPR screen. A positive score (blue) indicates positive selection, a negative score (red) indicates negative selection i.e. a genetic dependency. Note that there are no sgRNAs targeting RYBP in the Brunello library, hence it is excluded (left).

In terms of CRISPR dependency score, *CBX4* and *PCGF4* rank similarly to functionally essential PRC2 core components (Fig. 5.3A). This places them in the top percentile of all genes screened, excluding pan-essential genes (Fig 3.5B). The next closest ranking CBX paralog is *CBX7* (ranked 3736 – 23rd percentile) and *PCGF2* (ranked 5225 – 32nd percentile). Importantly, these results are replicated by the screens performed by Dr Dáire Gannon in the HSJD-DIPG-007 cell line.

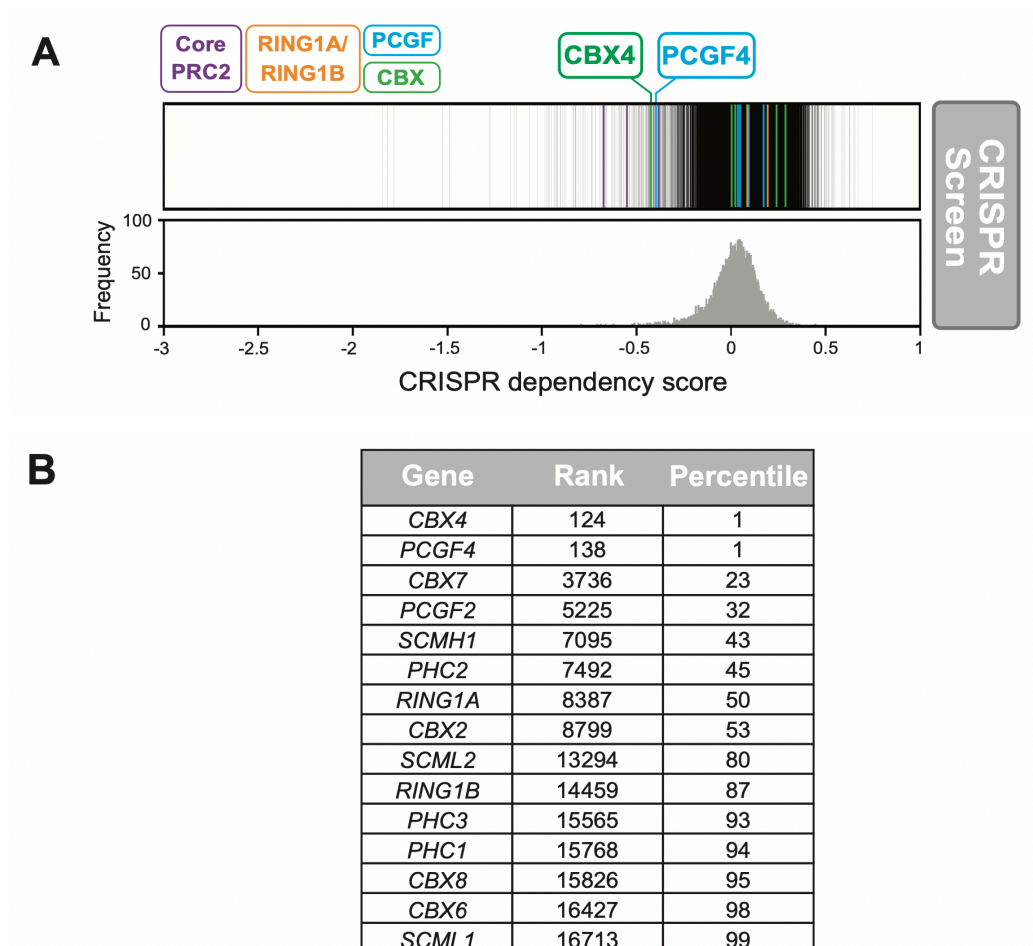


Figure 3.5 cPRC1 components CBX4 and PCGF4 rank amongst the top hits in the genome-wide CRISPR-Cas9 screen.

A) Heatmap (top) and frequency distribution (bottom) for all genes grouped in bins according to CRISPR dependency score. CRISPR dependency scores were rounded to 3 decimal places, with all genes with the same rounded CRISPR dependency score value allocated to the same bin.

B) Table in descending order according to rank (middle column) for each cPRC1 gene, calculated using the MaGeCK analysis pipeline. Percentile rank for each gene is also shown (right column).

3.2.3 Growth competition assays validate CBX4 and PCGF4 as functionally essential to DMG

Genetic screens are powerful tools for identifying functional dependencies, but they are not without limitations. It is important to validate the results using orthogonal techniques to rule out that: 1) the apparent dependence on *CBX4* and *PCGF4* is not a false-positive hit, and 2) the lack of dependence on their paralogs is not a false-negative result.

False-positive results can arise, for example when genes that are not expressed score as genetic dependencies. This can occur due to off-target effects of sgRNAs. Although the Brunello library is specifically designed to limit off-target effects, it is important to verify that *CBX4* and *PCGF4* are expressed in H3K27M-DMG and are not false positive hits. Unfortunately RNA-seq data was not available for HSJD-DIPG-021. Ultimately, information on protein abundance and PRC1 complex composition is more informative than mRNA transcript data, as it is the proteins that execute functions in the cells. Dr Dáire Gannon & AJ Silva in the lab performed endogenous RING1B IP-mass spectrometry in two independent H3K27M-DMG cell lines (HSJD-DIPG-007 and SU-DIPG-XIII), capturing the makeup of all PRC1 complexes in these cells and also the relative abundances of the CBX and PCGF paralogs. The ratio of PCGF2 to PCGF4 varied broadly between cell lines, with almost all cPRC1 containing PCGF4 in SU-DIPG-XIII (data not shown). However, in HSJD-DIPG-007 where Dr Dáire Gannon also found a specific dependence on *PCGF4*, there is an almost equal mix of the two paralogs. This suggests that the dependency on PCGF4 may not be due solely to its abundance. Dr Dáire Gannon and AJ Silva also found that CBX7 and CBX6 were not present in cPRC1 complexes. Interestingly, they also found that the ratios of CBX2, CBX4 and CBX8 in cPRC1 complexes were consistent across cell lines and remarkably, CBX4 was overwhelmingly the least abundant CBX paralog making up less than 5% of the total CBX pool (Fig 3.6A). This demonstrates that CBX4 and PCGF4 are not functional dependencies in H3K27M-DMG simply because they are the most abundant/only paralogs present in cells and again suggests that complexes containing these subunits have distinct non-redundant functions.

Next, it is important to validate that the sgRNAs used in the screen have on-target activity. Although the Brunello library is specifically designed to maximise on-target efficacy, it is important to verify that sgRNAs have an appreciable effect on target protein levels as inefficient sgRNAs could lead to false-negative results. Western blots demonstrate that all sgRNAs targeting *CBX2*, *CBX4* and *CBX8* are effective at knocking out the relevant target (Fig 3.6B). As the proteomic data revealed that CBX6 and CBX7 are not present/negligible in cPRC1 complexes in H3K27M-DMG cells (Fig3.6A), the efficacy of these sgRNAs were not verified by immunoblot. Immunoblots also demonstrated that sgRNAs targeting *PCGF2* and *PCGF4* depleted the relevant target protein, although the sgRNAs targeting *PCGF4* displayed limited efficacy (Fig 3.6B). Despite the reduced efficacy of the *PCGF4* sgRNAs, they were strongly negatively selected against in the genome-wide screen, suggesting that H3K27M-DMG cells are very sensitive to even limited perturbation PCGF4 levels.

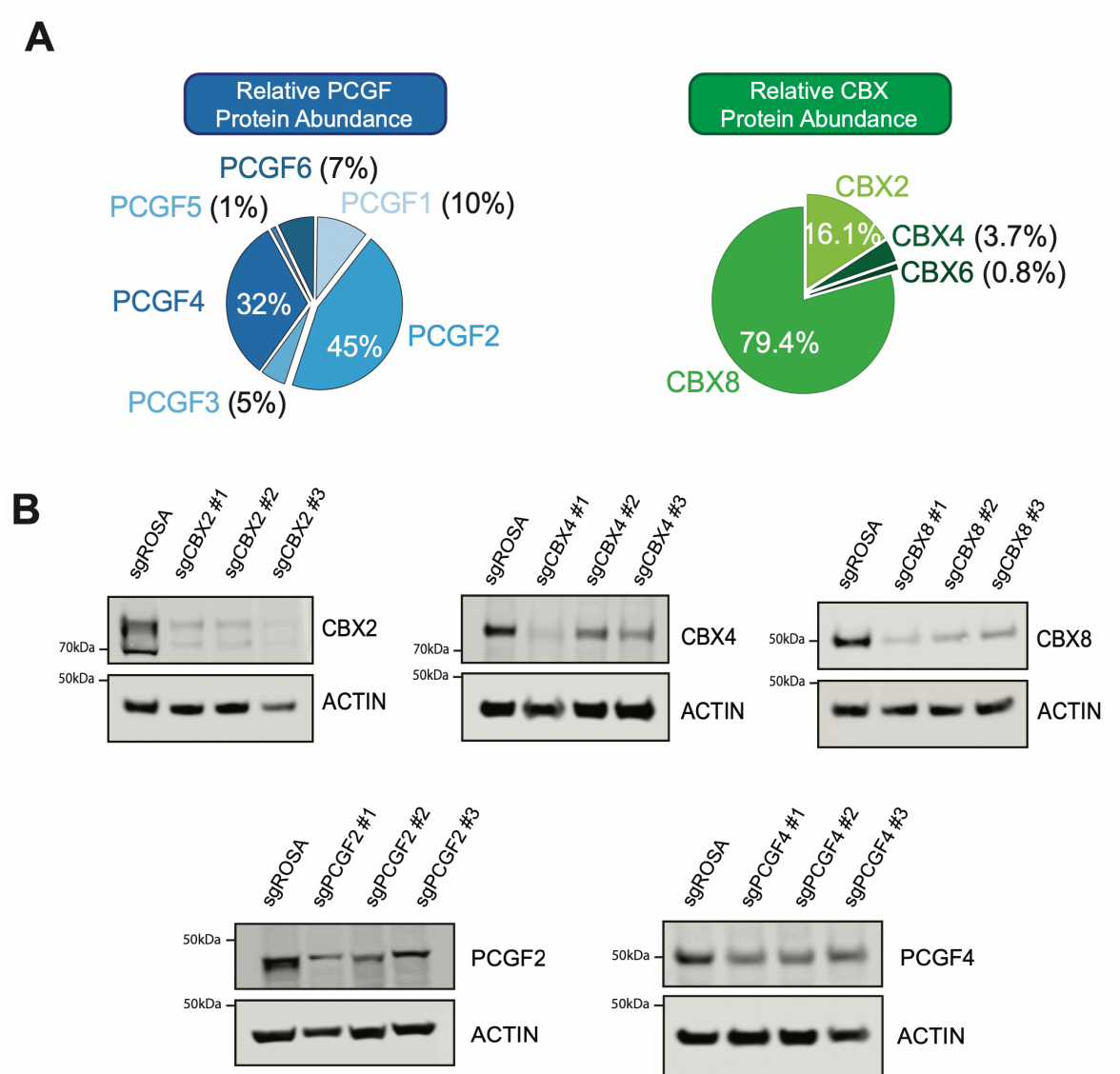


Figure 3.6 Immunoblots demonstrate Brunello sgRNA sequences have on-target efficacy.

A) Pie charts showing the relative abundances of PCGF and CBX proteins in HSJD-DIPG-007 as captured by endogenous RING1B IP-mass spectrometry. This proteomics data was generated by Dr Dáire Gannon and AJ Silva in the Bracken Lab.

B) Immunoblots using whole cell lysates from HSJD-DIPG-007-Cas9 cells following 72 hour stable expression of the indicated sgRNA.

Finally, growth competition assays were performed to validate the results of the genome-wide CRISPR-Cas9 screen (Fig 3.7A). Cas9-expressing HSJD-DIPG-021 cells were transduced with lentivirus containing a plasmid encoding GFP and an sgRNA, resulting in a heterogenous population of GFP-positive and GFP-negative cells. The percentage of GFP-positive cells in the population was measured every 4 days using flow cytometry. Measuring population dynamics over time using this GFP reporter assay allows us to infer if selection is occurring when a particular gene is targeted by CRISPR-Cas9. It is expected that sgRNAs targeting essential genes will experience negative

selection and the proportion of GFP-positive cells will decrease as they are outcompeted by uninfected cells.

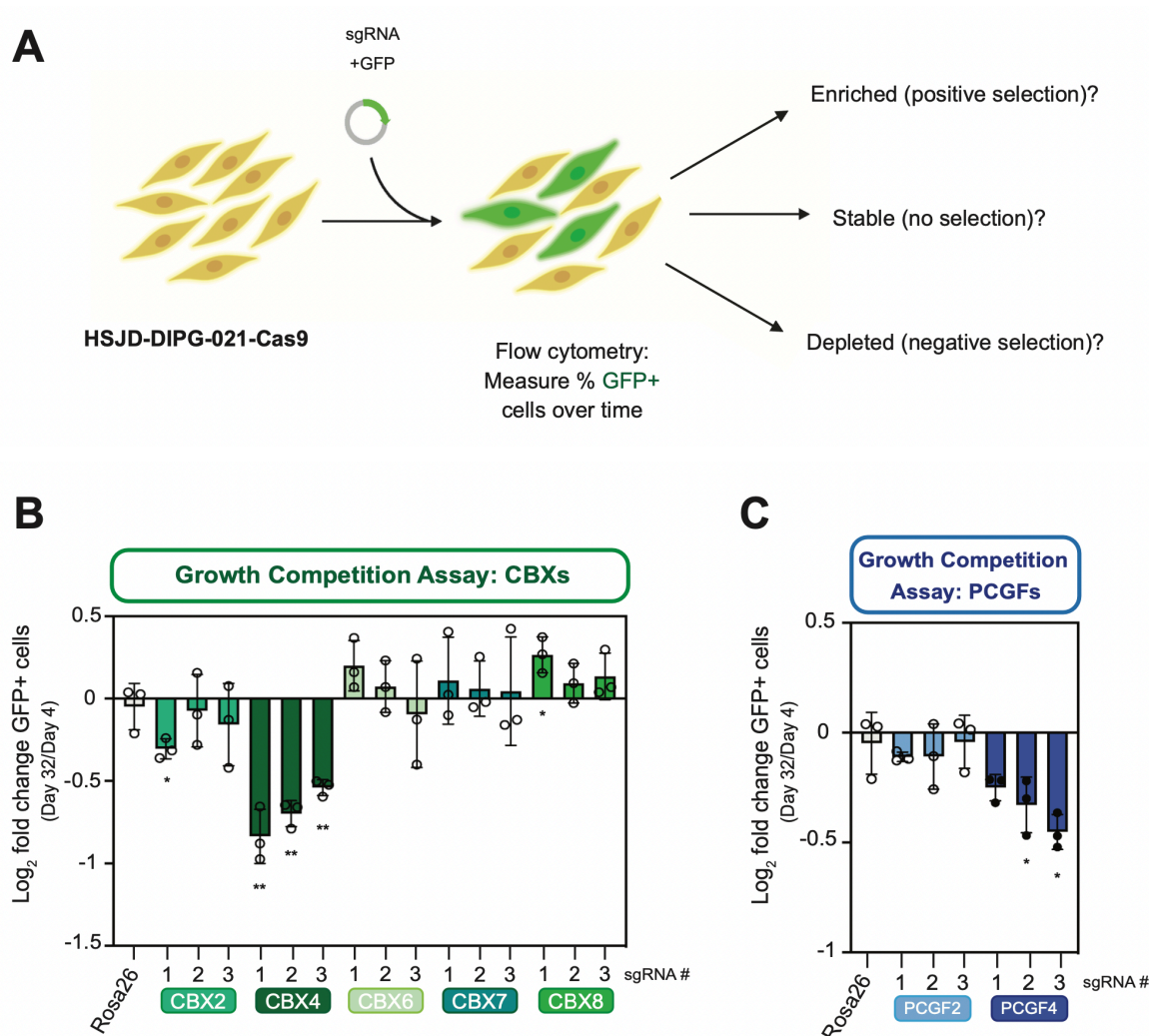


Figure 3.7 Growth competition assays confirms *CBX4* and *PCGF4* are genetic dependencies in H3K27M DMG.

A) Schematic of the growth competition assay experiment design (created with Biorender.com).

B) Growth competition assay results for sgRNAs targeting *CBX* genes and (C) cPRC1 specific *PCGF* genes in HSJD-DIPG-021. Graphs show log₂ fold change in the proportion of GFP+ cells in the population when the indicated sgRNA was delivered with lentivirus. *P*-value calculated using Student's *t*-test comparing against the negative control *Rosa26* sgRNA (* = *p* < 0.05, ** = *p* < 0.001).

Growth competition assays were performed in parallel using 3 sgRNAs targeting each *CBX* gene, as well as *PCGF2* and *PCGF4*. An sgRNA targeting the *Rosa26* locus controlled for negative cell fitness effects that could occur due to a number of technical factors. The *Rosa26* locus in the mouse genome is considered a “safe harbour” (i.e. a site in the genome that can be disrupted without adverse effects for the cell). While this sgRNA is not fully complementary to any site in the human genome, it is possible that erroneous double-stranded breaks could be induced by Cas9, especially as there is a homologous region in the human genome. This is an important negative control as the act of inducing double-stranded breaks by the Cas9 nuclease can itself cause cell death if the break is not resolved

properly. This must be accounted for and serves as a point of comparison when considering sgRNAs targeting other genes of interest. Cells transduced with the *Rosa26* sgRNA also control for factors such as silencing of GFP that could occur in a proportion of cells, and stress induced by lentiviral infection.

As expected, cells that received any sgRNAs targeting *CBX4* consistently and significantly depleted over the course of the assay (Fig 3.7B). This was not observed for the other *CBX* genes. Similarly, most sgRNAs targeting *PCGF4* depleted significantly over the course of the assay. While these did not reach the same level of significance as the *CBX4* sgRNAs, and indeed one *PCGF4* sgRNA did not meet the threshold of significance, this may be due to the relative inefficiency of the *PCGF4* sgRNAs (Fig 3.6B). Taken together with similar validations performed by Dr Dáire Gannon in HSJD-DIPG-007 cells, these growth competition assays validate the results of the genome-wide CRISPR-Cas9 screen and confirms that *CBX4* and *PCGF4* are genetic dependencies in H3K27M-DMG.

3.3 Discussion

There are many examples illustrating the clinical benefits of targeting a specific cancer genetic dependency. One of the best-known examples can be found in chronic myeloid leukaemia (CML). In 1960, a hallmark translocation known as the 'Philadelphia chromosome' was discovered in CML, but not healthy leukocytes (Nowell, 1962). Prior to this, standard therapy for CML was unselective cytotoxic chemotherapeutic agents. Specific targeting of malignant cells is not achieved with standard chemotherapy or radiation therapy, and the survival rate for CML was just 6% (Kantarjian et al., 2012). Following the discovery of this hallmark translocation, studies showed that the BCR-Abl fusion protein product of the Philadelphia chromosome is an oncogenic driver and genetic dependency in CML (Pierce et al., 2002). BCR-Abl encodes a constitutively active tyrosine kinase which drives the proliferative capacity of leukemic cells. Since then, drugs such as imatinib have been designed as inhibitors of BCR-Abl kinase activity. Use of imatinib and subsequent generation BCR-Abl inhibitors in the clinic means long-term survival of CML patients is now 87% (Lauseker et al., 2023). In short, rationally designed drugs targeting a genetic dependency have transformed CML from a fatal to treatable disease.

Small molecules and antibodies targeting specific cancer dependencies (e.g. BRAF in melanoma, HER2 in HER2-positive breast cancers, EZH2 in epithelioid sarcoma) are now being used in the clinic to treat a multitude of cancers. However, despite their clinical success, most cancer patients are still not benefiting from targeted therapeutics. This is partly due to a limited knowledge of candidate drug targets with a clear basis in underlying disease biology and a resultant lack of efficacy of cancer drugs entering clinical trials (Behan et al., 2019). Over 90% of drugs entering clinical trials to treat cancer fail to receive FDA approval (Hay et al., 2014). This failure rate is higher than that of drugs aimed to treat other diseases (e.g. infectious, autoimmune, neurological or cardiological diseases) (Hay et al., 2014). This discrepancy highlights the need to expand the range of therapeutic

targets for cancers by identifying cancer genetic dependencies, and if possible, to prioritise easily druggable targets with a clear link to disease biology.

In this chapter, genome-wide CRISPR-Cas9 screens identified two genetic dependencies in H3K27M-DMG: *CBX4* and *PCGF4*. According to publicly available CRISPR-Cas9 screening data across 1,150 cancer cell lines from the Cancer Dependency Map (depmap.org) *CBX4* is a genetic dependency in only 9 cell lines screened, and *PCGF4* is a dependency in only 16, indicating they are not common genetic dependencies and are more likely to be cell-type or disease-specific dependencies.

The results of the screen and validations performed in this chapter are supported by an epigenetic-focussed RNAi screen in a H3K27M-DMG cell line that ranked *PCGF4* as a top hit (Balakrishnan et al., 2020). Interestingly, genome-wide CRISPR-Cas9 screens performed in EZHIP-positive PFA tumour cells, another brain tumour type very similar to DMG (H3K27-altered), found *CBX4* to be the top ranked hit closely followed by core PRC2 members *SUZ12*, *EED* and *EZH2* (Michealraj et al., 2020). This suggests the dependency on *CBX4* could be linked to the characteristic pathological molecular changes to the chromatin landscape known to be induced by both H3K27M and EZHIP. Additionally, another study employed an epigenetic gene-focused CRISPR-Cas9 screen in four H3K27M DMG cell lines (including HSJD-DIPG-007) and one H3-wildtype glioblastoma (GBM) cell line, further supports the findings in this research chapter. Among the 1,350 genes assayed, *CBX4* and *PCGF4* were among the top 215 scoring genes in the H3K27M-DMG lines (Panditharatna et al., 2022). Strikingly, sgRNAs targeting *CBX4* or *PCGF4* did not deplete in the H3-WT GBM line. This suggests that these genes might be dependencies specifically in the H3K27M context. However, other factors such as differences in growth rate or cell-line specific idiosyncrasies could affect this result. Notably, evidence suggests that *PCGF4* is required for the self-renewal of neural stem cells (NSCs), meaning it could be a neural lineage-specific dependency (Molofsky et al., 2005). Therefore, before this conclusion could be drawn this would need to be tested in a panel of H3K27M-DMG and H3 wild-type cancer cell lines, for example using growth competition assays like those performed in Figure. 3.7. Ideally, these assays would also have been performed in isogenic H3K27M and H3K27M-KO DMG cell lines. Although these cell lines were not available at the time of the initial study, subsequent experiments have been performed by other lab members and collaborators to address this question. Researchers in the Phillips lab, Philadelphia, USA, conducted growth competition assays in a range of H3K27M-DMG cell lines and H3 wild-type glioblastoma cells. The results showed that sgRNA knock-out of *CBX4* had no effect on the growth of glioblastoma cells but did impact H3K27M-DMG cells (data not shown). Additionally, AJ Silva in the Bracken lab found that shRNA knock-down of *CBX4* impaired the growth of H3K27M-DMG cells but had no effect on isogenic H3K27M-KO DMG cells (data not shown). This suggests that the *CBX4* genetic dependency is specifically linked to H3K27M disease biology, rather than being a neural-lineage specific dependency. Genetic dependencies that are specific to a defining feature of the disease biology are ideal therapeutic targets, limiting toxicity. *CBX4* also presents an attractive potential therapeutic

target as it contains a conventionally drug targetable domain – the chromodomain. It has also reported that H3K27M-DMG cells are sensitive to a compound reported to inhibit PCGF4 (Balakrishnan et al., 2020). However, the clinical analogue of this inhibitor has been reported to primarily target microtubule polymerization, with potential indirect effects on PCGF4, suggesting that this inhibitor is largely non-specific and may exhibit significant toxicity in its current form (Eberle-Singh et al., 2019).

The specific dependency on *CBX4* and *PCGF4* is surprising, as Dr Dáire Gannon & AJ Silva in the lab demonstrated that paralogs for these proteins are expressed in H3K27M-DMG cells. Particularly striking is the observation that *CBX4* is a genetic dependency despite making up less than 5% of the total CBX pool in these cells. This indicates that these proteins have specific roles that are not redundant with their paralogs in H3K27M-DMG. There is currently very limited evidence of distinct molecular functions of distinct cPRC1 complex paralogs. This is mainly because most studies investigating the molecular functions of cPRC1 complex components have been conducted in mouse embryonic stem cells (mESCs), where Polycomb-mediated gene repression operates largely independent of PRC2, H3K27me3, or cPRC1 and instead relies on vPRC1 activity (Blackledge et al., 2020; Tamburri et al., 2020). For example, loss of *EED* or *PCGF2/4* does not impact Polycomb target gene repression whereas simultaneous perturbation of vPRC1 complexes leads to transcriptional reactivation (Scelfo et al., 2019). This demonstrates that vPRC1 complexes and likely the H2AK119ub modification they catalyse are necessary for maintaining repression in mESCs. However, this not entirely consistent with observations in *in vivo* systems, where the knockout of individual cPRC1 paralogs in *Drosophila* and mouse models results in a range of phenotypes and developmental defects (Bajusz et al., 2018). This discrepancy highlights a significant gap in our current knowledge of cPRC1 complexes. Clearly, mESCs do not accurately represent the functions of cPRC1 complexes in other stem cell types or adult tissues. It is likely that certain molecular features of cPRC1 complex subunits are context-dependent and may be dispensable in ESCs, but essential in other contexts. For example, PRC2 plays a crucial role in later developmental stages and different tissue contexts, and is also implicated in the biology of diseases such as H3K27M-DMG (Kim & Roberts, 2016; Tamburri et al., 2024). The findings presented in this chapter suggest that cPRC1 complexes may also execute distinct functions in specific cell contexts, which are not apparent in ESCs. The H3K27M mutation may have a unique impact on *CBX4* and *PCGF4*, either by directly altering the function of these proteins or by modifying the chromatin landscape in a manner that specifically affects their activity. This is particularly plausible, given the distinctive changes to the H3K27me3 chromatin landscape that are induced by the H3K27M mutation. The next chapter will investigate how this affects PRC1 dynamics on chromatin.

CHAPTER 4

H3K27M redistributes cPRC1 on chromatin

4.1 Introduction

As discussed in Chapter 1, the H3K27M oncohistone induces distinctive changes to the chromatin landscape that is central to the disease biology and now defines this subclass of paediatric glioma. The H3K27M oncohistone is necessary for establishment and maintenance of these tumours (Harutyunyan et al., 2019; Silveira et al., 2019). Therapeutic targeting of the oncohistone itself would therefore be ideal, and indeed a H3K27M vaccine is being trialled in patients, with one patient in complete remission 3 year following treatment (Boschert et al., 2024). Direct targeting of the H3K27M oncohistone with alternative methods is technically challenging as it does not have typically “druggable” domains and differs from wild-type histone H3 by just one amino acid. Understanding the downstream consequences of the oncohistone and how it drives disease biology could instead provide more readily targetable weak points. The H3K27M oncohistone is already known to interfere with PRC2 chromatin occupancy and function, and indeed PRC2 has been shown to be a promising potential therapeutic target in H3K27M-DMG (Brien et al., 2021; Harutyunyan et al., 2019; Mohammad et al., 2017). The aim of this research chapter is to explore the dependence on CBX4 and PCGF4 in more detail by investigating if the presence of the H3K27M oncohistone impacts the normal chromatin occupancy and function of these proteins, and if this drives disease biology.

4.2 Results

4.2.1 Characterisation of isogenic H3K27M and H3K27M-KO cell lines

Previous work in the lab had used human foetal neural stem cells (fNSCs) engineered to express H3.3-WT or H3.3-K27M to model DMG biology in an isogenic system (Brien et al., 2021). Chromatin immunoprecipitation (ChIP) experiments in this model system demonstrated that PRC2 chromatin occupancy is altered by the presence of the H3K27M oncohistone. Additionally, these experiments showed that cPRC1 component PCGF4 chromatin occupancy is also altered in H3K27M-positive human fNSCs, with PCGF4 occupancy drastically increased at Polycomb target genes. This indicates that PRC1 complex occupancy, and possibly function, are also likely to be affected by the H3K27M oncohistone. In light of the results of the genome-wide CRISPR-Cas9 screen, it was decided to study the effects of H3K27M on PRC1 and PRC2 dynamics in more detail using an isogenic system.

Human fNSCs used previously in the lab are difficult to engineer and to culture. Additionally, human fNSCs senesce after approximately a dozen passages (Brien et al., 2021). These problems could be overcome by knocking out the mutant *H3F3A* allele in H3K27M-DMG cell lines which are easier to culture and grow indefinitely. However, essential oncogenes are often amplified in cancer, particularly in cell lines that have been extensively cultured *in vitro*. The copy number of the mutant *H3F3A* allele(s) must be quantified to determine if CRISPR-Cas9 knock-out would be feasible in the cell lines available in the lab. A quantitative PCR (qPCR) of the *H3F3A* gene locus revealed that the gene is extensively amplified in both HDSJ-DIPG-021 and HSJD-DIPG-007 with approximately 10 times the number of copies compared to a diploid genome (Fig 4.1A). Engineering isogenic H3K27M-KO lines using either of these H3K27M-DMG cell lines available in the lab therefore was deemed

unfeasible within the desired timeframe. Instead, pre-established isogenic H3K27M and H3K27M-KO cell lines were gifted from Dr. Nada Jabado at McGill University in Canada. The H3K27M-KO cell line was derived from knock-out of mutant *H3F3A* allele in the SU-DIPGXIII cell line (Fig. 4.1B) (Harutyunyan et al., 2019).

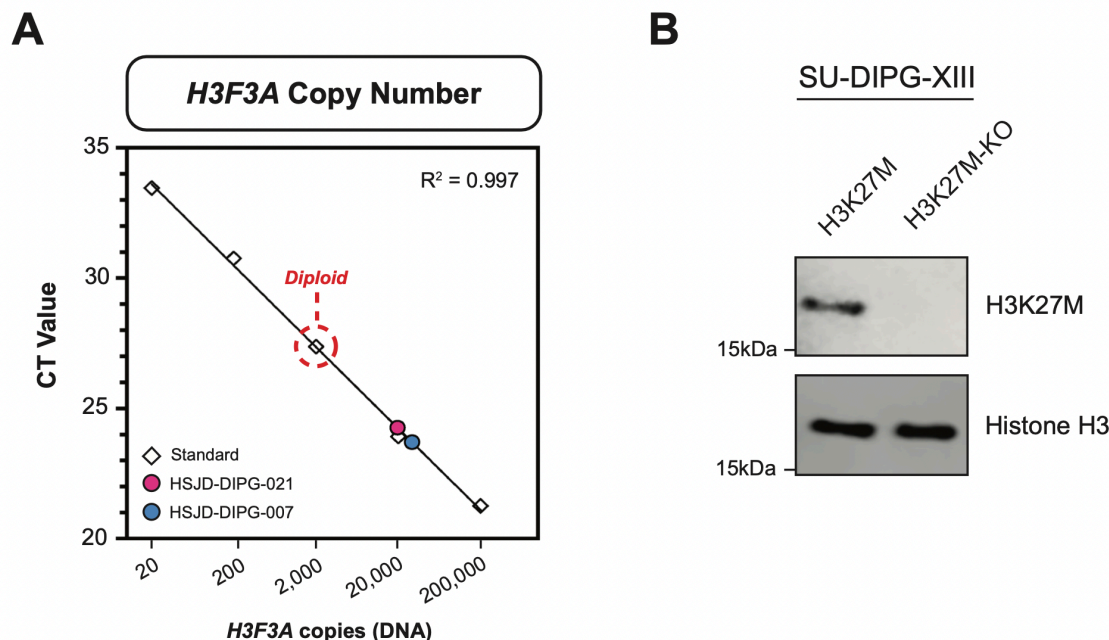


Figure 4.1. Knock out of H3K27M in SU-DIPG-XIII

A) Quantitative PCR (qPCR) showing *H3F3A* gene copy number in two H3K27M DMG cell lines; HSJD-DIPG-007 and HSJD-DIPG-021. The standard curve was generated using a diploid human genome template (HMEC). If the DMG cell lines were diploid for *H3F3A*, then the amount of DMG template gDNA used for the assay would equate to approximately 2,000 *H3F3A* copies (indicated by red dotted line). However, the actual number of *H3F3A* copies present in the DMG gDNA was approximately 10 times greater than expected. $n = 3$ technical replicates.

B) Immunoblot for H3K27M and histone H3 performed on whole-cell lysate from SU-DIPG-XIII H3K27M and H3K27M-KO isogenic cell lines. These cell lines were a kind gift from Dr Nada Jabado, McGill University (Haratunyan et. al. 2019).

Before more exploratory studies could be performed, the isogenic lines were thoroughly characterised to ensure this was an appropriate model system to study the effects of H3K27M. Importantly, knock-out of H3K27M reverses the global loss of H3K27me3 and global increase in H3K27ac levels, as expected (Fig 4.2A). Quant-seq (a bulk transcriptomics technique) followed by differential expression analysis (performed by Dr Craig Monger) revealed 2,443 significantly differentially expressed genes in the H3K27M-KO cell line compared to the parental H3K27M-positive cell line (Fig 4.2B). Gene-ontology analysis of these significantly differentially expressed genes particularly enriched for terms involved in differentiation and neurodevelopment (Fig 4.2C). This is consistent with numerous previous bodies of work that have shown that the H3K27M oncohistone disrupts neurodevelopmental pathways which is thought to be important for establishing disease mechanism.

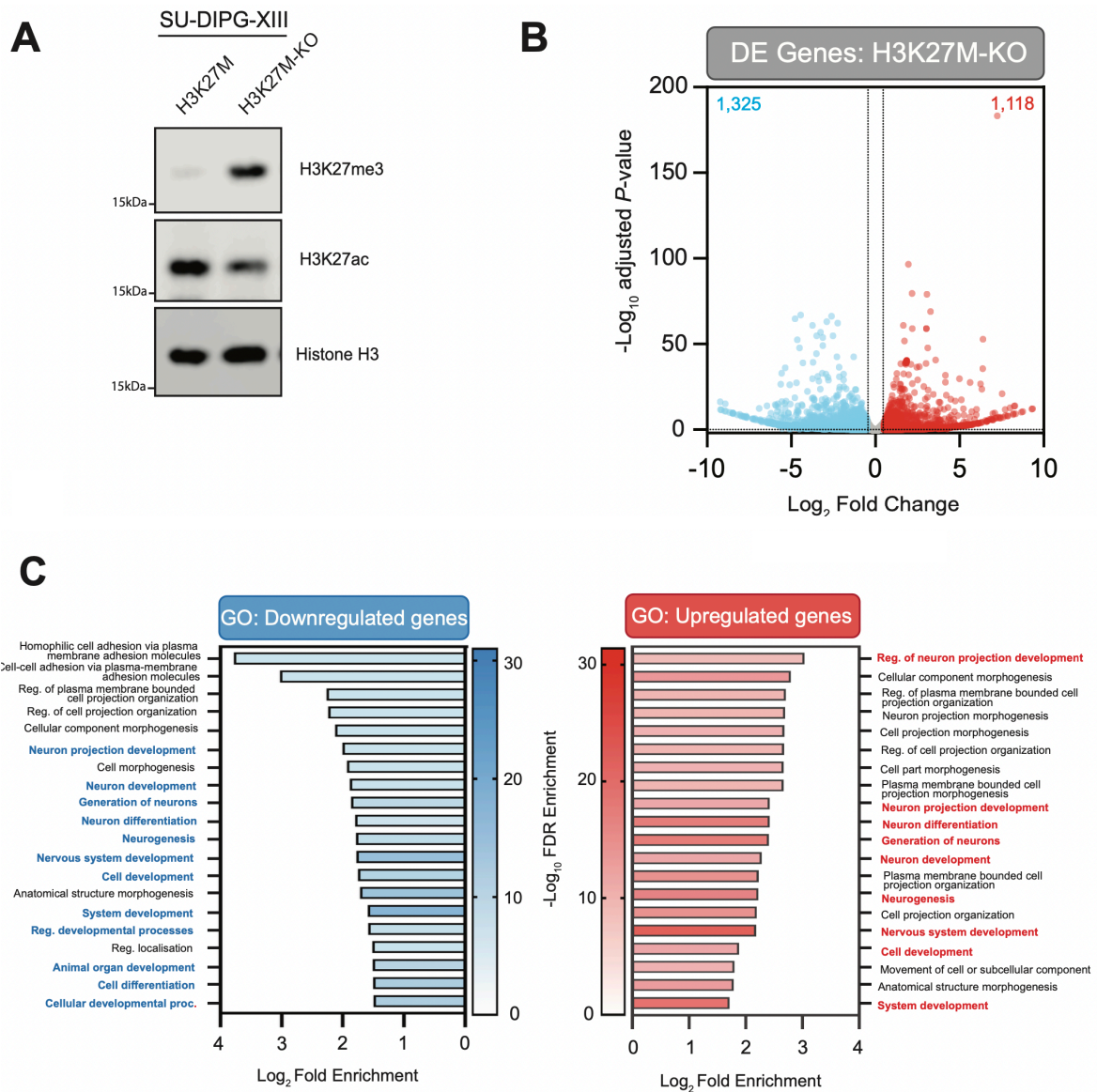


Figure 4.2. Differential gene expression in H3K27M-KO cells

A) Immunoblot performed on whole-cell lysates showing H3K27me3 and H3K27ac levels in SU-DIPG-XIII H3K27M and H3K27M-KO cell lines. Total histone H3 shown as a loading control.

B) Volcano plot showing differentially expressed genes in H3K27M-KO cells compared to H3K27M cells. $n = 3$ biological replicates. Cut off parameters: $p < 0.05$, minimum 1.5 fold change in mRNA abundance. Red indicates upregulated genes, blue indicates downregulated.

C) Gene-ontology analysis for statistically significant upregulated (red) or downregulated (blue) genes in H3K27M-KO compared to H3K27M cells. Gene ontology analysis performed using ShinyGOv0.80. Biological process terms shown.

Polycomb target genes are involved in differentiation and developmental processes and have been shown to be disrupted in H3K27M-positive cells (Brien et al., 2021; Harutyunyan et al., 2019; Silveira et al., 2019). The presence of the H3K27M oncohistone has been shown to cause increased repression of Polycomb target genes involved in neurodevelopment and this aberrant repression is thought to be an important contributor to disease mechanism (Brien et al., 2021; Silveira et al., 2019). As discussed in Chapter 1, the Polycomb target genes that are actively repressed depend on the cell type (i.e. in a neural progenitor cell, neural differentiation genes are repressed while more stem

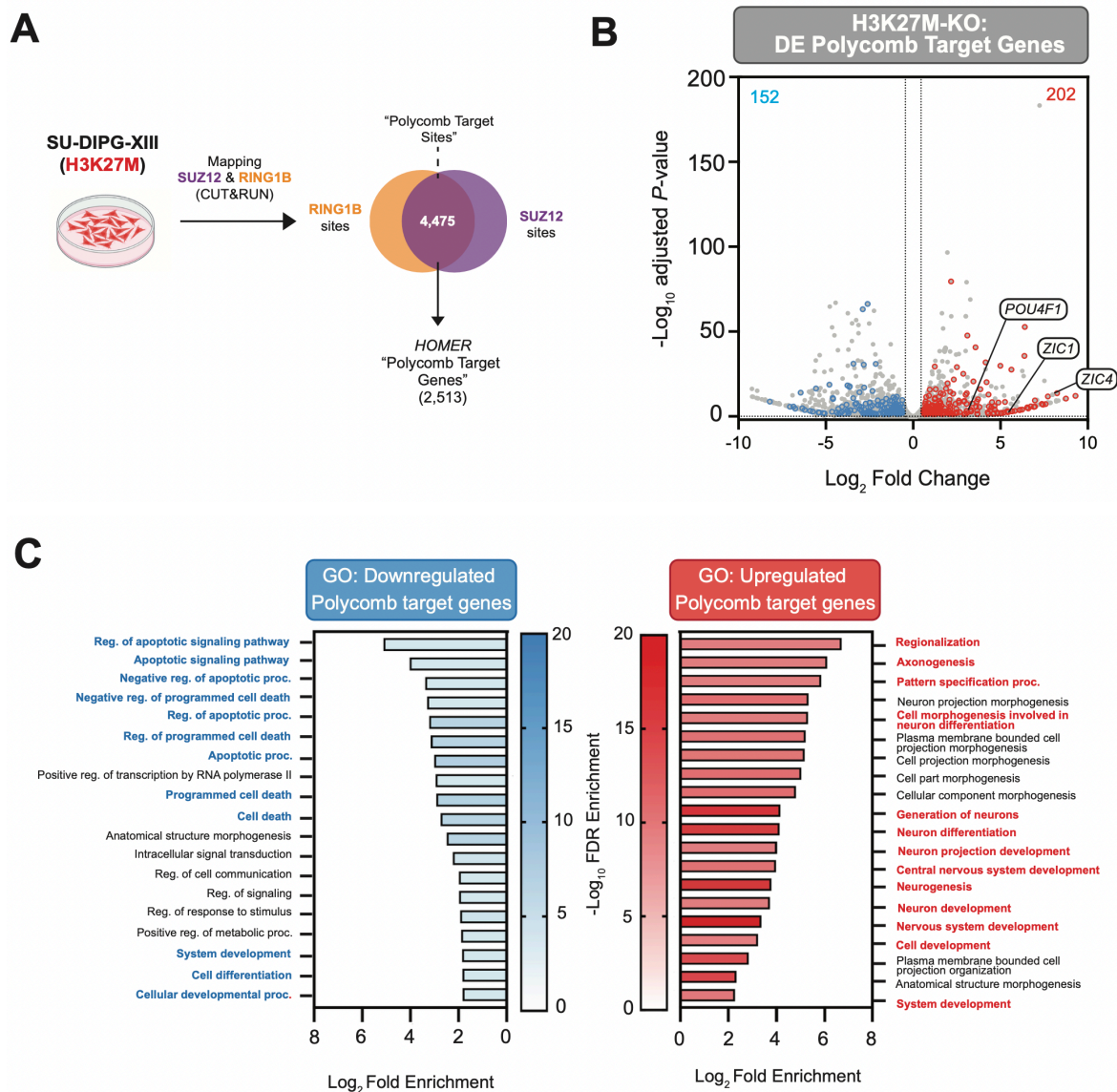


Figure 4.3. Differential expression of Polycomb target genes in H3K27M-KO cells

A) Schematic depicting the process for defining Polycomb target sites in SU-DIPG-XIII H3K27M cells. Polycomb target sites were defined as peaks common to both SUZ12 and RING1B CUT&RUN datasets in at least 2/3 replicates. $n = 3$ biological replicates. The consensus list of 4,475 Polycomb target sites were annotated with genes using HOMER.

B) Volcano plot of differentially expressed genes in H3K27M-KO vs H3K27M cells. Statistically significant differentially expressed Polycomb target genes are indicated in red (upregulated) or blue (down-regulated). Significance cut-offs: $p < 0.05$, minimum 1.5 fold change in mRNA abundance. Transcription factors involved in neural lineage differentiation, *POU4F1*, *ZIC1*, and *ZIC4* are indicated.

C) Gene-ontology analysis for statistically significant upregulated (red) or downregulated (blue) Polycomb target genes in H3K27M-KO compared to H3K27M cells. Gene ontology analysis performed using ShinyGOv0.80. Biological process terms shown.

cell-like genes involved in self-renewal are expressed). To verify if H3K27M-induced aberrant repression is accurately modelled in this isogenic system, relevant Polycomb target genes first had to be identified. This was done by mapping PRC2 and PRC1 via CUT&RUN of SUZ12 and RING1B in the parental H3K27M SU-DIPG-XIII cell line. Dr Anthony Doherty in the Brien lab overlapped these

peak sets to identify 4,475 regions common to both PRC2 and PRC1. These Polycomb target sites were annotated with HOMER to associate them with genes. The gene set associated with these 4,475 PRC1 & PRC2 sites were designated as Polycomb target genes for these cells (Fig 4.3A). 2,513 Polycomb target genes were defined this way in SU-DIPG-XIII H3K27M cells.

354 of these Polycomb target genes were significantly differentially expressed in the H3K27M-KO cells (Fig. 4.3B). Gene-ontology analysis revealed that Polycomb target genes that were significantly upregulated in H3K27M-KO cells enriched for neurodevelopmental terms, while Polycomb target genes that were significantly downregulated enriched for terms involved in cell cycle regulation and negative regulation of apoptosis (Fig. 4.3C). Indeed, expression of neural lineage transcription factor genes *POU4F1*, *ZIC1* and *ZIC4* involved in differentiation are among the upregulated Polycomb target genes in H3K27M-KO cells. This means that in the H3K27M cells, neural differentiation genes are more repressed by Polycomb while cell cycle genes and negative regulators of apoptosis are more expressed – a gene signature consistent with a more stem-cell like state. This is also consistent with the observed higher proliferation rate of the H3K27M cells (doubling time: ~24 hours) versus the H3K27M-KO cells (doubling time: ~72 hours).

Altogether, these data demonstrate that this isogenic model system replicates previously established findings on the role of H3K27M in DMG disease biology. Therefore, this is an appropriate system for further exploratory studies dissecting the roles of distinct PRC1 complexes in H3K27M disease biology.

4.2.2 The H3K27M does not localise to Polycomb target sites

As mentioned above, previous work from our lab and others has shown that the H3K27M oncohistone affects the distribution of PRC2 and H3K27-associated histone modifications. Previous work from our lab also suggested that the presence of H3K27M may also alter the chromatin occupancy of PCGF4-containing PRC1 complexes. It was therefore decided to use the SU-DIPG-XIII isogenic system to quantitatively map these perturbations and comprehensively document the effect, if any, of H3K27M on various PRC1 complexes in unprecedented detail (Fig 4.4A). Traditionally, ChIP-seq and CUT&RUN-seq reads are normalized using the “reads per million” normalization method which precludes accurate quantitative comparisons between samples – i.e. only qualitative comparisons can be made. In order to make accurate quantitative comparisons between the isogenic H3K27M and H3K27M-KO cell lines, a quantitative CUT&RUN approach was established to accurately capture dynamics. To make genome-wide quantitative comparisons across samples, reads are normalized to a reference exogenous genome (Rx). Crucially, antibodies that cross-react with their target in both human and mouse cells were identified and validated using quantitative CUT&RUN-qPCR (Appendix I). For CUT&RUN-Rx, mouse cells were added at a ratio of 1:10 as an exogenous spike-in to human glioma cells at the very beginning of the CUT&RUN protocol, which was then performed as standard (Fig 4.4B). High-quality reads mapping to the spike-

in genome were obtained (Fig. 4.4C) allowing quantitative spike-in normalisation factors to be calculated and applied to data derived from SU-DIPG-XIII cells. Additionally, three biological replicates were performed and sequenced for each condition to maximise the robustness and reproducibility of the data. The average Rx-normalised signal for the three biological replicates is represented in all plots involving CUT&RUN-Rx data. This custom CUT&RUN-Rx computational workflow was developed by Dr Anthony Doherty in the lab.

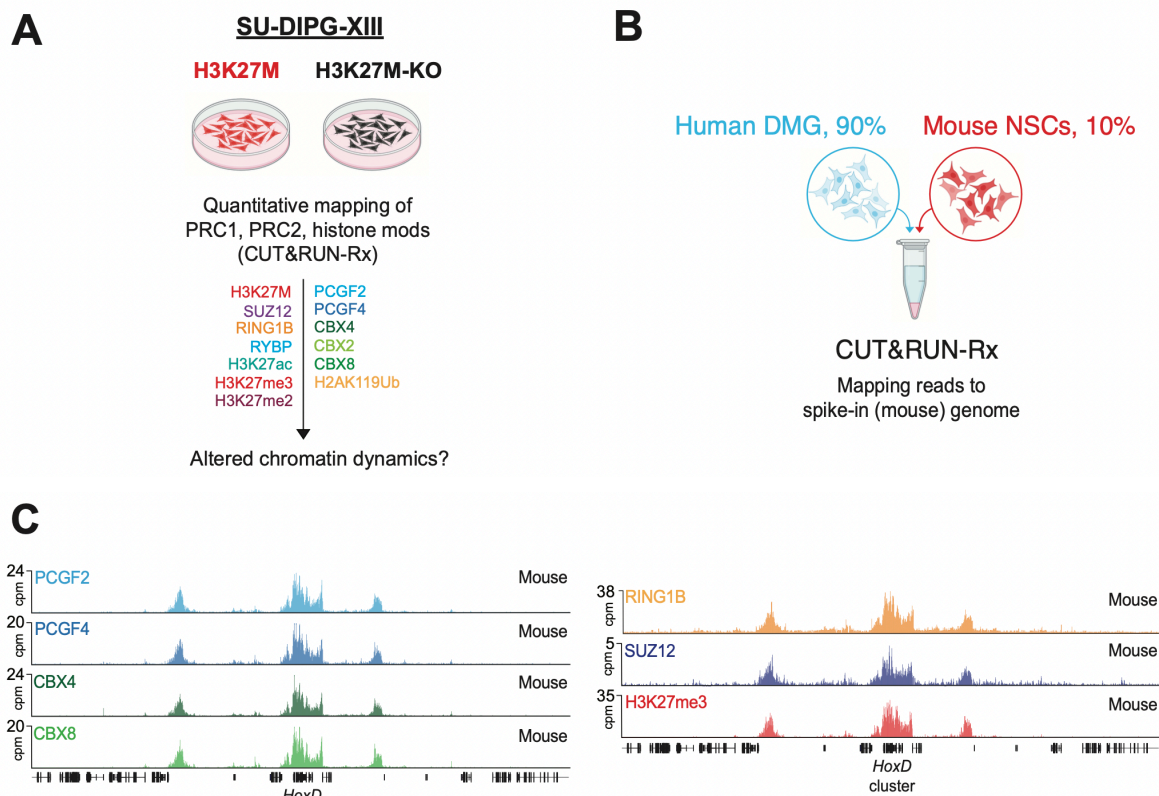


Figure 4.4. Capturing chromatin dynamics with quantitative CUT&RUN-Rx

A) Schematic representation of the quantitative CUT&RUN-Rx method mapping multiple proteins and histone modifications in SU-DIPG-XIII H3K27M and KO cells.

B) Schematic representation of the reference genome (Rx) spike-in process at the beginning of the CUT&RUN-Rx protocol.

C) Genome browser tracks for spike-in mouse reads for the indicated antibodies.

The H3K27M oncohistone is understood to have a dominant-negative inhibitory effect on PRC2, resulting in the characteristic punctate H3K27me3 distribution retained at Polycomb target sites (Brien et al., 2021; Lewis et al., 2013; Venneti et al., 2014). A previous study from the lab exogenously expressed V5-tagged H3K27M in human fNSCs (Brien et al., 2021). The exogenous V5-tagged oncohistone was mainly mapped to actively expressed genomic regions marked by H3K27ac, a histone modification associated with active gene expression. Additionally, it has been shown that the H3K27M does not need to incorporate into chromatin to induce the characteristic histone modification landscape changes (Jain et al., 2020). The inhibitory effect on PRC2 is therefore thought to be mainly an *in trans* effect rather than *in cis* at Polycomb target sites. However, it should be noted that the fNSC isogenic system did find some incorporation of the H3K27M oncohistone at Polycomb target

sites too, meaning there could also be an *in cis* inhibitory effect (Brien et al., 2021). This has yet to be independently confirmed in another system.

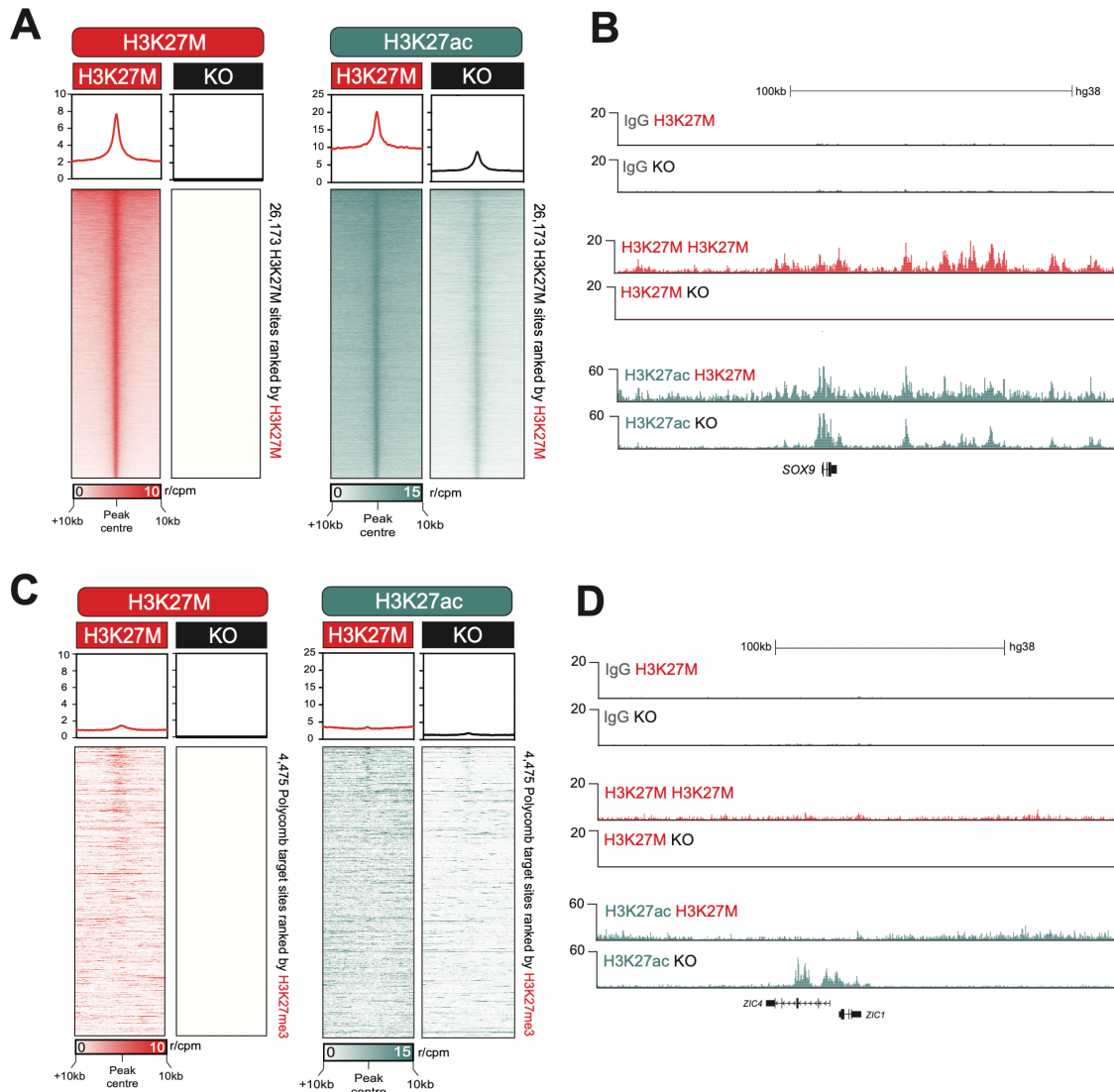


Figure 4.5. H3K27M localises mainly to actively expressed genes

A) Tornado plots and average plots showing average CUT&RUN-Rx signal (n = 3) for endogenous H3K27M oncohistone and H3K27ac in SU-DIPG-XIII H3K27M and H3K27M-KO cells at H3K27M peaks.

B) Genome browser tracks showing average CUT&RUN-Rx signal for H3K27M and H3K27ac at the indicated genomic locus in SU-DIPG-XIII H3K27M and H3K27M-KO.

C) Tornado plots and average plots showing average CUT&RUN-Rx signal (n = 3) for endogenous H3K27M oncohistone and H3K27ac in SU-DIPG-XIII H3K27M and H3K27M-KO cells at Polycomb target sites ranked by H3K27me3 CUT&RUN-Rx signal.

D) Genome browser tracks showing average CUT&RUN-Rx signal for H3K27M and H3K27ac at the indicated genomic locus in SU-DIPG-XIII H3K27M and H3K27M-KO.

Here, H3K27ac and the endogenous H3K27M oncohistone were mapped using the newly established CUT&RUN-Rx technique. In agreement with this previous study, H3K27ac was also highly enriched at the genomic loci where H3K27M localised (Fig 4.5A). These loci were generally actively expressed genes in SU-DIPG-XIII e.g. SOX9 (Fig 4.5B). Conversely, H3K27ac did not localise to Polycomb target genes in H3K27M cells (Fig. 4.5C, Fig. 4.5D). In agreement with the

fNSC isogenic system, only a very small amount of H3K27M signal could be detected at Polycomb target genes (Fig. 4.C, Fig. 4.D). Impressively, virtually no H3K27M signal was detected in the H3K27M-KO cells at any site (Fig. 4.5A, Fig. 4.5C). This reflects a combination of the cleanliness of the antibody used, the low background signal generally associated with CUT&RUN compared to traditional ChIP-methods, and the complete knock-out of H3K27M in these cells. In summary, these results support previous work and further confirm that this isogenic system accurately models the dynamics and molecular features associated with the presence or absence of the H3K27M oncohistone.

4.2.3 H3K27M drastically reorganises cPRC1 chromatin occupancy

PRC2 chromatin occupancy has previously been shown to be dramatically increased in H3K27M positive cells at Polycomb target sites. CUT&RUN-Rx of SUZ12 validated this result in this independent isogenic system (Fig. 4.6A). Here, CUT&RUN-Rx of RING1B demonstrates that PRC1 chromatin occupancy is also dramatically increased at Polycomb target sites in H3K27M cells compared to H3K27M-KO cells (Fig. 4.6A). There was a positively correlated relationship between SUZ12 (PRC2) and RING1B (PRC1) accumulation and H3K27me3 signal – i.e. the sites where SUZ12 and RING1B accumulated the most are the sites that retain the most H3K27me3 signal in SU-DIPG-XIII H3K27M cells (Fig. 4.6B). H3K27me3 signal can be thought of as a proxy for the “strength” of the Polycomb target site – the strongest Polycomb target sites have the most H3K27me3 signal. While there is a positive correlation between SUZ12 accumulation and H3K27me3 signal, SUZ12 chromatin occupancy increases at the vast majority of Polycomb target sites (Fig. 4.6B). However, RING1B accumulation in H3K27M cells appears to correlate more with the “strength” of the Polycomb target site – i.e. the amount of H3K27me3 signal. In H3K27M cells, RING1B accumulates mostly at the “strongest” Polycomb target sites with the most H3K27me3 signal, while it remains largely unchanged at the “weaker” Polycomb target sites which have less H3K27me3 signal (Fig. 4.6B). The accumulation of SUZ12 and RING1B in H3K27M cells can be observed at an example “strong” Polycomb target locus (Fig. 4.6C).

In line with the inhibitory effect of H3K27M on PRC2, H3K27me2 has also been shown to be globally reduced in H3K27M cells. The exact function of H3K27me2 is unknown., but it is thought to have a generally repressive effect and H3K27me2 catalysis is a necessary precursor for H3K27me3. Under normal cellular conditions H3K27me2 is an abundant modification diffusely present throughout the genome. Average plots of CUT&RUN-Rx signal show that H3K27me2 is largely unchanged at Polycomb target peak centres in H3K27M cells, while signal is mainly lost beyond these sites (Fig. 4.6A). In fact, while H3K27me2 is unchanged at the majority of Polycomb target sites in H3K27M cells, it is increased at the “strongest” Polycomb target sites and reduced at the “weakest” Polycomb target sites (Fig. 4.6B). The inhibitory effect of the H3K27M oncohistone on PRC2 is likely most profound at intergenic sites where PRC2 interactions are transient compared to its target sites where interactions are more stable, leading to loss of diffuse H3K27me2 and H3K27me3 signal beyond

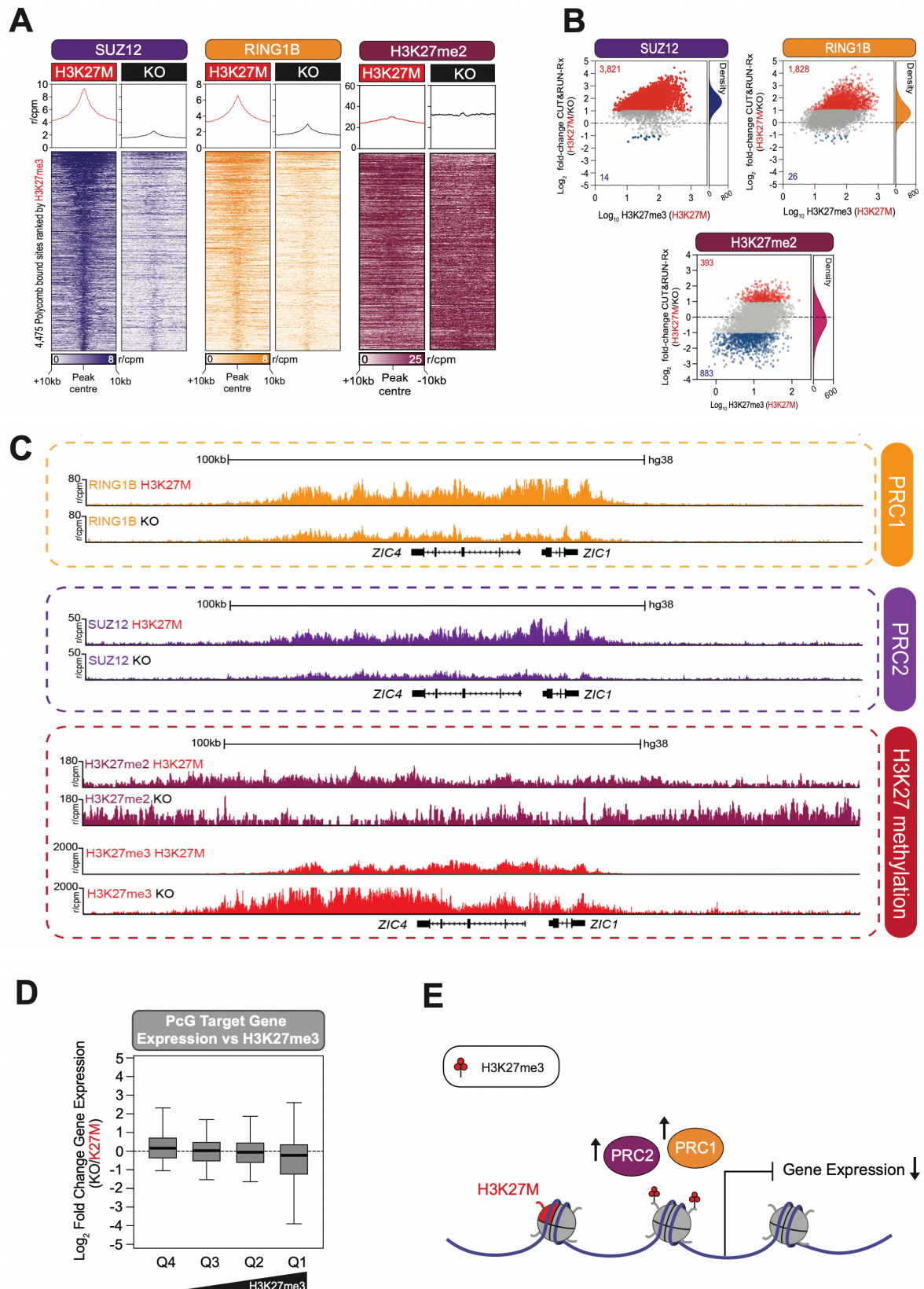


Figure 4.6 CUT&RUN-Rx shows PRC2 and PRC1 complexes accumulate on chromatin in H3K27M cells
A) Tornado plots and average plots showing average CUT&RUN-Rx signal ($n = 3$) for SUZ12, RING1B and H3K27me2 in SU-DIPG-XIII H3K27M and H3K27M-KO cells at Polycomb target sites ranked by H3K27me3 CUT&RUN-Rx signal in SU-DIPG-XIII H3K27M cells.

Figure 4.6 (continued)

B) MA-plots showing fold-change in binding for SUZ12, RING1B and H3K27me2 at Polycomb target sites in H3K27M compared to H3K27M-KO cells. Sites changing >2-fold are shown in red (increasing) or blue (decreasing) with the number of sites in these categories also indicated. The densities of chromatin binding changes are shown on the right of each plot.

C) Genome browser tracks showing average CUT&RUN-Rx signal for SUZ12, RING1B, H3K27me2 and H3K27me3 at the indicated genomic locus in SU-DIPG-XIII H3K27M and H3K27M-KO.

D) Boxplots showing log₂ fold change in RNA abundance of Polycomb target genes between H3K27M and KO cells; genes are separated into quartiles based on their local enrichment of H3K27me3 in SU-DIPG-XIII H3K27M cells.

E) Schematic representation depicting the increase in PRC2 and PRC1 complexes at Polycomb target sites where H3K27me3 is retained and the decrease in gene expression associated with this.

PRC2 target sites. At PRC2 target sites where PRC2 occupancy is increased in H3K27M cells, PRC2 can continue to catalyse H3K27me2 and to a lesser extent convert some H3K27me2 to H3K27me3. This likely leads to the observed accumulation of H3K27me2 at “strong” Polycomb target sites defined by the retention of some H3K27me3. Indeed, this has been documented previously by others, and can be observed here at an example “strong” Polycomb target locus in these cells, the *ZIC1/ZIC4* locus (Fig. 4.6C). H3K27M not only disrupts the H3K27me3 landscape, but also restricts H3K27me2 distribution. In the presence of H3K27M, this normally diffuse modification concentrates more at Polycomb target sites. This distribution is reminiscent of H3K27me3 in H3-wild type cells.

Next, the transcriptomic data (Fig. 4.2B) was revisited to investigate the relationship between the observed Polycomb related CUT&RUN-Rx dynamics and gene expression dynamics. The Polycomb target genes defined earlier (Fig. 4.3A) were ranked in descending order according to H3K27me3 signal in SU-DIPG-XIII H3K27M cells. This ranked list of genes was then divided into quartiles from Q1 (genes with the most H3K27me3 signal in SU-DIPG-XIII i.e. the “strongest” Polycomb target genes) to Q4 (genes with the least H3K27me3 signal i.e. the “weakest” Polycomb target genes). Examples of Q1 genes are *ZIC1* and *ZIC4*, which show dramatic accumulation of SUZ12 and RING1B in H3K27M cells. Q1 genes, which generally experienced the most PRC2/PRC1 accumulation in H3K27M cells, were also generally more repressed in H3K27M cells than in H3K27M-KO cells (Fig. 4.6D). Together, these results suggest that inhibition of PRC2 by the H3K27M oncohistone leads to the distinctive changes in H3K27me2 distribution and a characteristic punctate distribution of H3K27me3, with sites retaining the most H3K27me3 signal accumulating the most PRC1, leading to stronger repression of nearby Polycomb target genes (Fig. 4.6E).

Next, it was decided to investigate PRC1 dynamics in more granular detail, beginning with cPRC1 complexes. Each CBX and PCGF paralog that exists in H3K27M-DMG cells was mapped with quantitative CUT&RUN-Rx. Interestingly, there were radical differences in paralog behaviour in the presence of H3K27M. In H3K27M cells, CBX2, CBX4 and PCGF4 accumulated dramatically on chromatin at Polycomb target sites, whereas CBX8 and PCGF2 were somewhat displaced

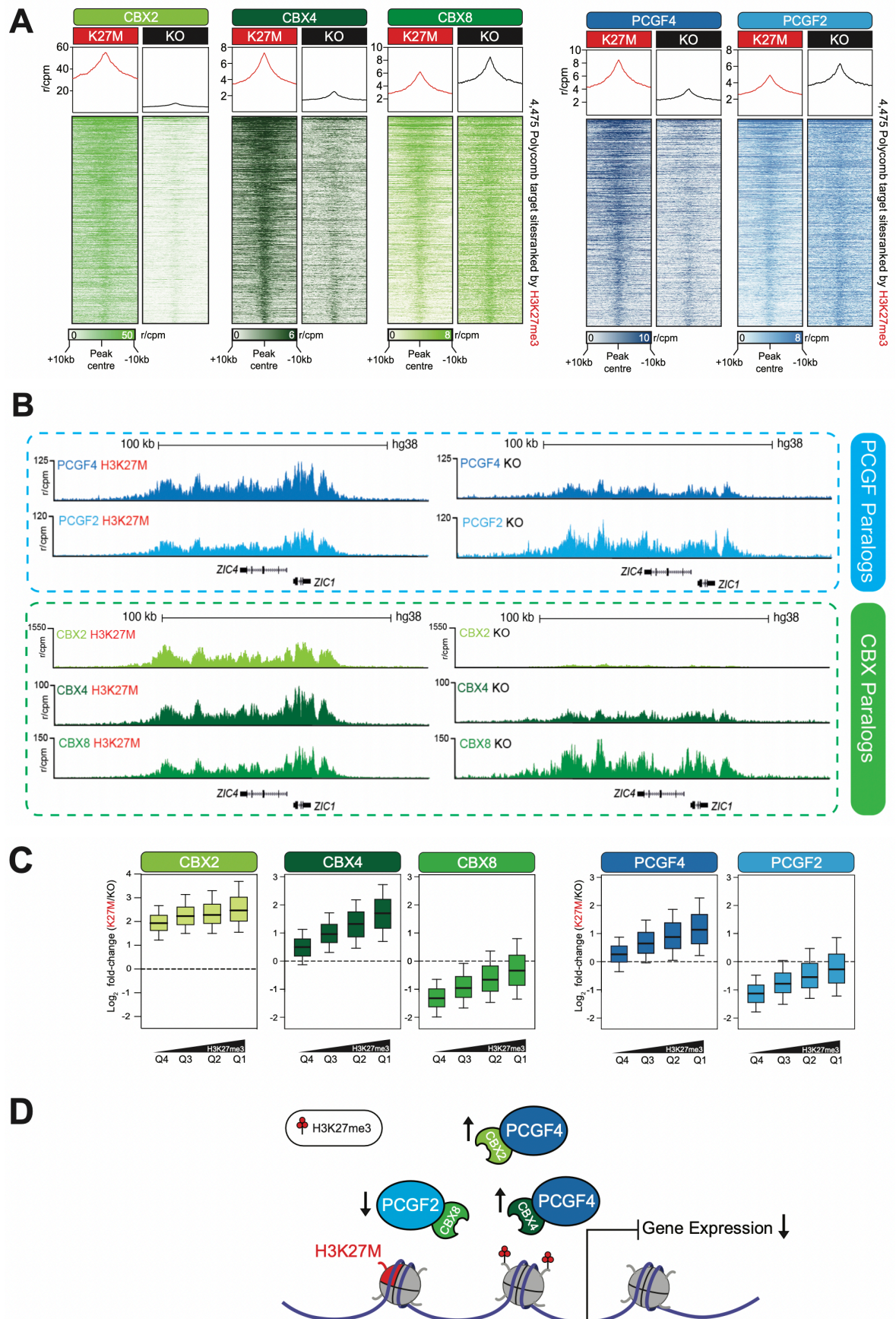


Figure 4.7 CUT&RUN-Rx shows differential behaviour of cPRC1 complexes on chromatin in H3K27M cells

A) Tornado plots and average plots showing average CUT&RUN-Rx signal ($n = 3$) for CBX2, CBX4, CBX8,

Figure 4.7 (continued)

PCGF2 and PCGF4 in H3K27M and H3K27M-KO cells at Polycomb target sites ranked by H3K27me3 CUT&RUN-Rx signal in SU-DIPG-XIII H3K27M cells.

B) Genome browser tracks showing average CUT&RUN-Rx signal for PCGF2, PCGF4, CBX2, CBX4 and CBX8 at the indicated genomic locus in SU-DIPG-XIII H3K27M and H3K27M-KO.

C) Boxplots showing log₂ fold change in CUT&RUN-Rx signal for the indicated protein at Polycomb target genes; genes are separated into quartiles based on their local enrichment of H3K27me3 in SU-DIPG-XIII H3K27M cells.

D) Schematic representation depicting the increase in CBX2, CBX4, and PCGF4 containing cPRC1 complexes and decrease in complexes containing CBX8 and PCGF2 at Polycomb target sites in H3K27M cells. Accumulation of cPRC1 complexes correlates with H3K27me3 retention and a reduction in gene expression (previous figure).

(Fig. 4.7A). To further validate these findings, I also investigated the dynamics of CBX2, CBX4, and CBX8 in an independent mouse neural stem cell (NSC) model, where either H3K27M or H3-WT were exogenously expressed (Appendix II). This suggests that the opposing dynamics of cPRC1 subcomplexes are a general property of cells harbouring the H3K27M mutation, rather than a cell-type specific phenomenon.

There was a positive correlation in the relationship between paralog abundance on chromatin and H3K27me3 signal – while CBX2, CBX4, and PCGF4 accumulated at Polycomb target genes in all quartiles, they accumulated the most at Q1 genes. Interestingly, this positive correlation also existed for the paralogs that were generally displaced in H3K27M cells – CBX8 and PCGF2 were generally retained/slightly increased at the “strongest” Q1 Polycomb target genes and displaced the most at “weaker” Polycomb target genes (Fig. 4.7C). These dynamics can be seen at the Q1 genes *ZIC1* and *ZIC4*, which show dramatic accumulation of CBX2, CBX4 and PCGF4 but not a reciprocal loss of CBX8 and PCGF2 – rather, CBX8 and PCGF2 are retained/slightly increased (Fig. 4.7B). Generally, it seems that the punctate H3K27me3 distribution induced by H3K27M leads to a re-organisation of cPRC1 complexes at Polycomb target sites; complexes containing CBX2, CBX4, and/or PCGF4 accumulate on all Polycomb target sites, while CBX8 and PCGF2 are generally displaced. The accumulation of cPRC1 complexes is associated with increased gene repression (Fig. 4.7D).

Finally, vPRC1 dynamics were captured via CUT&RUN-Rx of protein subunit RYBP. While not a constitutive member of all vPRC1 complexes – YAF2 can substitute – mapping RYBP captures a broad range of vPRC1 complexes. CUT&RUN-Rx revealed that RYBP and H2AK119ub are slightly increased at Polycomb target sites in H3K27M cells (Fig. 4.8A). While there is a slight general increase in RYBP and H2AK119ub at Polycomb target sites independent of H3K27me3 levels, at most loci there is no difference in RYBP or H2AK119ub signal between the two isogenic cell lines (Fig. 4.8C). This modest difference can be observed even at a “strong” Polycomb target locus (Fig. 4.8B). Consistent with vPRC1 acting upstream rather than downstream of PRC2, there is no correlation between RYBP or H2AK119ub signal and H3K27me3 signal (Fig. 4.8C). This is intuitive as vPRC1 lacks the H3K27me3 reader chromodomain that cPRC1 complexes have in the CBX

subunit. As H3K27me3 signal correlates with increased gene repression in H3K27M cells, there does not appear to be a correlation between RYBP or H2AK119ub signal and increased gene repression

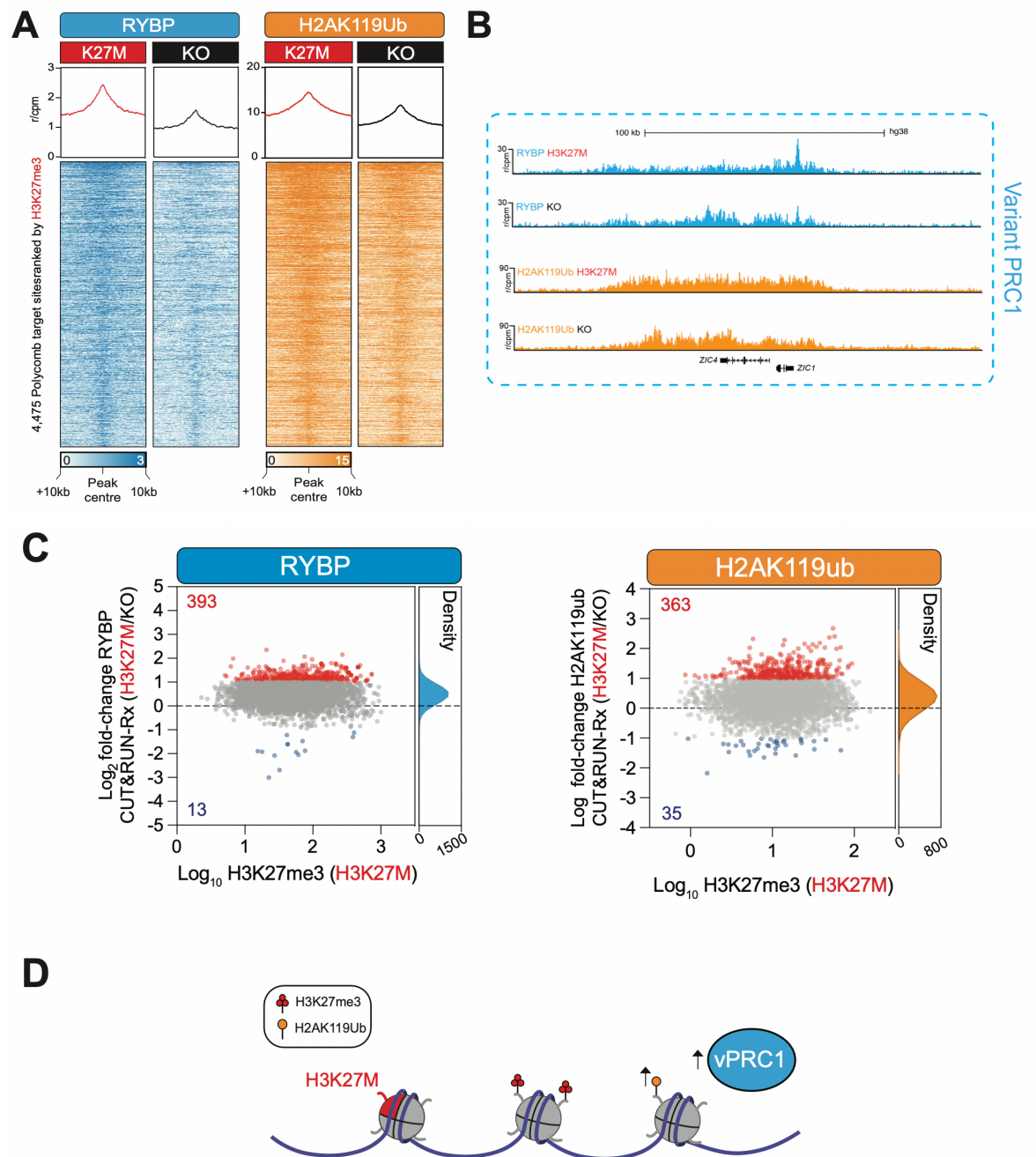


Figure 4.8 CUT&RUN-Rx shows vPRC1 complexes are agnostic to changes in H3K27me3 distribution.

A) Tornado plots and average plots showing average CUT&RUN-Rx signal (n = 3) for RYBP & H2AK119Ub in H3K27M and H3K27M-KO cells at Polycomb target sites ranked by H3K27me3 CUT&RUN-Rx signal in SU-DIPG-XIII H3K27M cells.

B) Genome browser tracks showing average CUT&RUN-Rx signal for RYBP & H2AK119ub at the indicated genomic locus in SU-DIPG-XIII H3K27M and H3K27M-KO.

C) MA-plot showing fold-change in binding for RYBP & H2AK119ub at Polycomb target sites in H3K27M compared to H3K27M-KO cells. Sites changing >2-fold are shown in red (increasing) or blue (decreasing) with the number of sites in these categories also indicated. The densities of chromatin binding changes are shown on the right of each plot.

D) Schematic depicting a slight increase in RYBP and H2AK119ub in H3K27M cells independent of changes in the H3K27me3 landscape.

in H3K27M. This suggests that cPRC1 complexes, but not vPRC1 complexes, are responsible for this increased repression of Polycomb target genes. The slight increase in H2AK119ub observed is likely due to the slight increase in vPRC1, which are the more catalytically active complexes. Overall, it appears that there is a slight increase in vPRC1 and H2AK119ub on chromatin in H3K27M cells, but this does not correlate with H3K27me3 signal or increased gene repression (Fig 4.8D).

In summary, the presence of H3K27M leads to a reorganisation of cPRC1 complexes on chromatin. H3K27M has an inhibitory effect on PRC2. PRC2 dramatically accumulates on chromatin at its target sites and concentrates H3K27me3 to a characteristic punctate pattern at Polycomb target sites. This change in H3K27 methylation pattern appears to affect how it is read by CBX chromodomains. CBX2 and CBX4 containing cPRC1 complexes dramatically accumulate at Polycomb target sites, while CBX8 chromatin occupancy is reduced. Similarly, complexes containing PCGF4 accumulate on chromatin while complexes containing PCGF2 are displaced. This redistribution of H3K27me3 and subsequent reorganisation of cPRC1 complexes correlates with the aberrant gene repression shown to be important in driving H3K27M DMG disease biology. vPRC1 complexes are largely unaffected, with a slight global increase on chromatin which does not correlate with H3K27me3 or gene repression.

It is interesting that the two PRC1 genetic dependencies identified in the CRISPR-Cas9 screen, CBX4 and PCGF4, accumulate at Polycomb target genes and their accumulation correlates with increased gene repression. Meanwhile, their non-essential paralogs are largely displaced from chromatin. A notable exception to this is CBX2, which behaves similarly to CBX4 and PCGF4 in the presence of H3K27M. There are some nuances that should be considered when interpreting these results. Firstly, proteomics performed by Dr Dáire Gannon and AJ Silva in the Bracken lab revealed that PCGF4 is vastly more abundant than PCGF2 in SU-DIPG-XIII cells. Additionally, whole-cell immunoblots show that there is more PCGF4 in SU-DIPG-XIII H3K27M cells than in the isogenic H3K27M-KO cells (Fig. 4.9A). The fact that there is overall more PCGF4 protein in the H3K27M cells could contribute to its increased abundance on chromatin in these cells, compared to the H3K27M-KO cells where there is less of it available. The same is true for CBX2, which is also generally more abundant in H3K27M cells than H3K27M-KO cells (Fig. 4.9B). The other proteins and paralogs, including CBX4, are equally abundant between the two cell lines (Fig. 4.9A, Fig. 4.9B). The accumulation of PCGF4 and CBX4 may be important to their functionally essential roles in H3K27M-DMG. However, accumulation alone is clearly not the basis of their essentiality, as CBX2 also accumulates but has been verified to be functionally redundant in H3K27M-DMG. Therefore, CBX4 and PCGF4 must have some distinct functions that are essential to H3K27M disease biology.

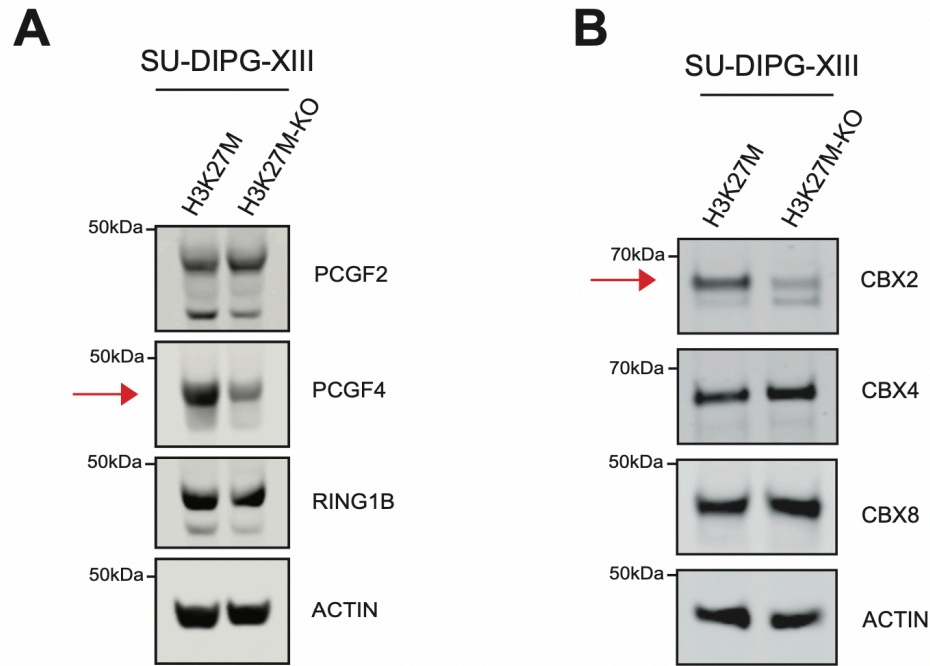


Figure 4.9. Whole-cell cPRC1 protein subunit abundances in isogenic H3K27M and H3K27M-KO cell lines

A) Immunoblot for cPRC1 PCGF paralogs and RING1B performed on whole-cell protein lysate from SU-DIPG-XIII H3K27M and H3K27M-KO cells. Abundances are equal between the two cell lines, with the exception of PCGF4 (indicated by red arrow). Beta-Actin serves as a loading control.

B) Immunoblot for CBX paralogs performed on whole-cell protein lysate from SU-DIPG-XIII H3K27M and H3K27M-KO cells. Abundances are equal between the two cell lines, with the exception of PCGF4 (indicated by red arrow). Beta-actin serves as a loading control.

4.3 Discussion

This chapter has shown that the presence of H3K27M leads to a reorganization of cPRC1 complexes on chromatin. A comparison of the binding dynamics of CBX2, CBX4 and CBX8, as well as PCGF2 and PCGF4, revealed that their binding patterns are not directly competitive, i.e. they are not anti-correlated. In fact, regions with the largest increases in CBX2, CBX4, and PCGF4 binding maintained relatively stable levels of CBX8 and PCGF2. The underlying levels of H3K27me3 appear to determine cPRC1 dynamics. Specifically, regions with the highest levels of H3K27me3 retention showed the greatest increase in CBX2, CBX4, and PCGF4 binding. Notably, quantitative mapping of the vPRC1-specific protein RYBP demonstrated that while RYBP levels exhibit a modest increase at Polycomb target sites, this increase is not correlated with underlying H3K27me3 levels. Consistent with the established direct relationship between H3K27me3 and cPRC1, but not vPRC1, quantitative mapping of the vPRC1-specific protein RYBP demonstrated that while RYBP levels exhibit a modest increase at Polycomb target sites, this increase is not correlated with underlying H3K27me3 levels. These data indicate that H3K27M induces a distinctive change to the H3K27me3 landscape, leading to a redistribution of cPRC1 complexes on chromatin. This redistribution is likely driven by the CBX subunits, which contain the H3K27me3-reader chromodomain. This implies that different CBX chromodomains have different affinities or sensitivities when it comes to sensing H3K27me3 patterning.

This H3K27me3 landscape characteristic to H3K27M-DMG results in an enrichment of complexes containing CBX2 or CBX4 and PCGF4 at Polycomb target sites. This enrichment correlates with aberrant repression of Polycomb target genes. It is interesting that aberrant repression of non-essential CBX2 is also correlates with increased gene repression. Immunoblots revealed that CBX4 levels are equal between the H3K27M isogenic cell lines. However, H3K27M-mutant cells exhibit elevated levels of CBX2 and PCGF4 compared to H3K27M-KO cells. This increase in abundance could certainly contribute to their increased abundance on chromatin in this context. In fact, overexpression of cPRC1 subunits, including PCGF4 and CBX genes, is a common feature of various cancers, but it is thought that this overexpression could be a consequence of increased cell proliferation and stem-like characteristics, rather than a contributing factor in tumorigenesis (Bracken et al., 2019). This notion is supported by the observation that the expression of these genes is correlated with cell proliferation and stem cell identity. However, very little is currently known about CBX chromodomain methylation sensing abilities (discussed further in Chapter 6). It is also entirely possible that the increase in abundance of CBX2 and CBX4 at these sites is at least partly due to these paralogs having similar H3K27me3 sensing abilities.

From the experiments presented in this chapter, any relationship between cPRC1 subcomplex dynamics and gene expression changes is purely correlative. This cannot determine to what extent different cPRC1 complexes are contributing to repression of these genes. To address this, AJ Silva in the Bracken lab used shRNAs to knock-down individual CBX paralogs in H3K27M and H3K27M-KO cells and measured Polycomb target gene mRNA abundance using RT-qPCR. He found that only knock-down of CBX4 in H3K27M cells led to derepression of Polycomb target genes – knock-down of CBX2 and CBX8 did not lead to derepression of Polycomb target genes, nor did knock-down of CBX4 in H3K27M-KO cells (data not shown). With these experiments, AJ demonstrated that cPRC1 complexes containing CBX4 are specifically responsible for aberrant repression of Polycomb target genes. Importantly, the H3K27M oncohistone does not localise to the sites where these dynamics and gene expression changes are occurring, indicating that this repression is linked to the change in chromatin modification landscape induced by H3K27M, rather than the presence of the oncohistone itself.

Finally, it is important to note that the development of a quantitative CUT&RUN-Rx approach was crucial for capturing the dynamic behaviour of distinct cPRC1 subcomplexes. It is now widely appreciated that spike-in of a reference genome is essential for quantitative chromatin mapping, as it enables the accurate assessment of global changes (Orlando et al., 2014). Without spike-in normalisation, chromatin mapping data can be skewed, leading to misrepresentation of changes. Indeed, spike-in normalisation was necessary for obtaining an accurate global view of cPRC1 dynamics in these cells. Standard read depth normalisation resulted in misleading conclusions, suggesting increased abundance of PCGF2 and CBX8 in H3K27M-mutant cells. However,

quantitative CUT&RUN-qPCRs corroborated the spike-in normalised results, instilling confidence in the global dynamics captured by CUT&RUN-Rx (Appendix III).

The next chapter will focus on identifying the features of CBX4 that are essential to H3K27M-DMG biology.

CHAPTER 5

Identification of functionally essential regions of CBX4 in H3K27M-DMG

5.1 Introduction

The main aim of this chapter is to identify functionally essential domains of CBX4 and to provide further mechanistic understanding as to what unique features functionally separate CBX4 from its paralogs.

The discovery of *CBX4* as a genetic dependency in H3K27M-DMG suggests that CBX4-containing cPRC1 complexes have unique functions distinct from other cPRC1 complexes. To begin identifying what these might be it was decided to perform a CRISPR-Cas9 knock-out tiling screen, a targeted approach that differs from traditional genome-wide screens. This method involves introducing an array of mutations across a single protein target using a library of sgRNAs that target every available *S. pyogenes* Cas9 PAM sequence (NGG) within the protein-coding region of the target gene. The tiling screen is designed to identify functionally essential domains within a protein, as even minor perturbations (such as in-frame mutations) in these critical domains are likely to be deleterious (Shi et al., 2015). Therefore, cells transduced with sgRNAs targeting these domains will experience particularly strong negative selection. Indeed, studies have consistently shown that robust depletion is correlated with the targeting of regions encoding functionally important protein domains (Brien et al., 2018; Michel et al., 2018; Shi et al., 2015). The high-density tiling library allows for the precise identification of specific protein domains that are essential for CBX4 function in H3K27M-DMG. In this way, even previously unknown functional domains within CBX4 can be identified, providing valuable insights into the unique functions of CBX4-containing complexes.

5.2 Results

5.2.1 CRISPR-Cas9 knock-out tiling screen identifies functionally essential domains in CBX4

The pooled sgRNA library targeting *CBX4* was generated by collaborators in the lab of Dr Chun-Wei David Chen (City of Hope, USA). The CBX4 library was composed of 385 sgRNAs tiling every PAM sequence along the exons of CBX4 (Fig. 5.1). The tiling screen was performed in HSJD-DIPG-007 and HSJD-DIPG-021 Cas9-positive cells in a similar manner to the genome-wide CRISPR-Cas9 knock out screen (Fig. 5.2A). The abundance of the sgRNAs in both libraries was determined by sequencing sgRNA integrants at an early (Day 4) timepoint and late (Day 28) timepoint. The early timepoint is used as a comparison to determine the change in sgRNA abundance over time, but also provides important information for quality control. Cells were transduced with lentivirus with the aim to achieve approximately 400x representation of each sgRNA. sgRNAs with less than 100x representation at the early (Day 4) timepoint were discarded from further analyses to prevent the introduction of noise. Additionally, all sgRNA sequences were checked using the CRISPick (Broad Institute) online tool. Any sgRNAs flagged as having off-target effects were also removed from further analyses to prevent false-positive results. The vast majority of the *CBX4* sgRNA library (84.7%) remained after these filtration steps (Fig. 5.2B) maintaining good coverage of the *CBX4* exons.



Figure 5.1. *CBX4* CRISPR-Cas9 knock-out tiling screen sgRNA library
Map of all 385 sgRNAs targeting *CBX4* protein coding regions.

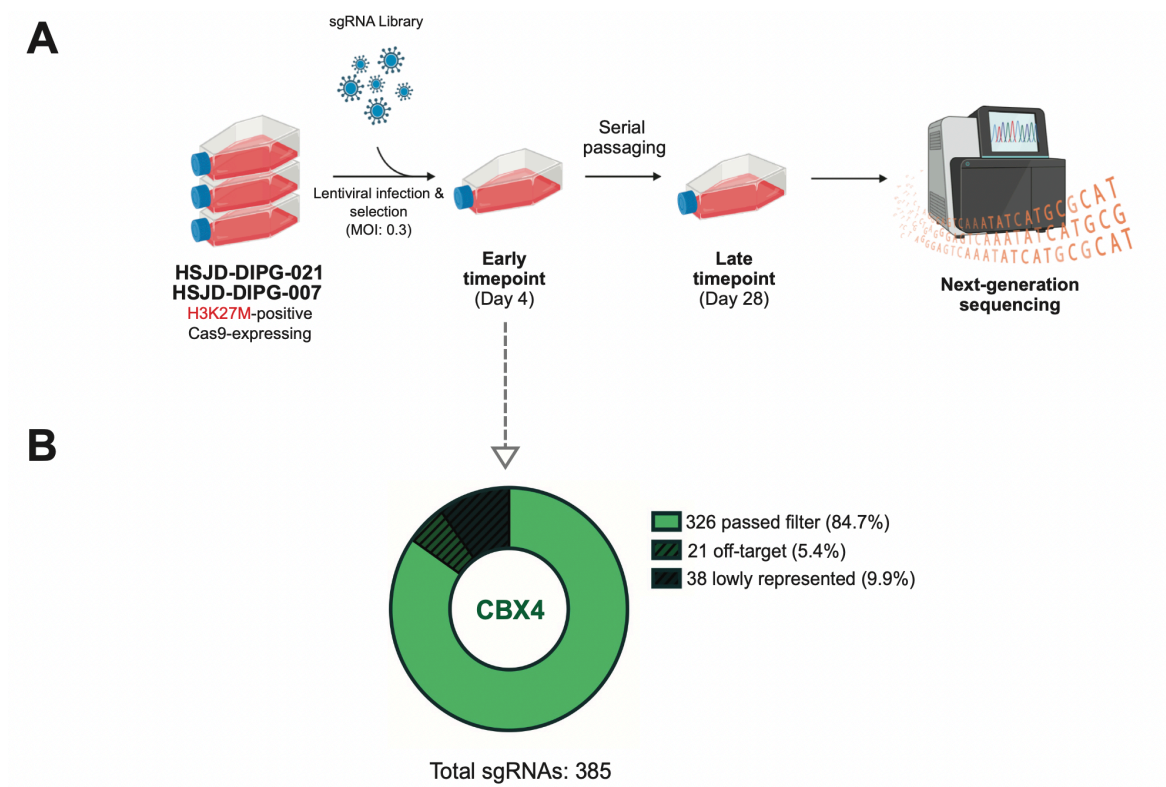


Figure 5.2. *CBX4* CRISPR-Cas9 knock-out tiling screens workflow and quality control

A) Schematic representation of the CRISPR-Cas9 tiling screen workflow.

B) Pie-chart of the *CBX4* tiling screen library composition post quality control filtration steps. sgRNAs with off-target effects according to CRISPick (Broad Institute) were removed, as were sgRNAs that were lowly represented (<100x) in the cell population at the Day 4 timepoint.

The fold change in abundance of the remaining *CBX4* sgRNAs screens was plotted and mapped to the *CBX4* protein (Fig. 5.3A). The average fold change of the positive and negative control sgRNAs are indicated by dotted lines on the graphs and are also represented by the highest and lowest values on the heat maps below each graph. Encouragingly, the positive control sgRNAs targeting depleted robustly from the population over the course of the screens, while the non-targeting negative control sgRNAs did not (Fig. 5.3A). The apparent increase in abundance of the non-targeting negative control sgRNAs is likely because all other sgRNAs in the library are targeting *CBX4*, a functionally essential gene and are therefore experiencing negative selection to some extent. In general, there was less robust depletion of sgRNAs in the HSJD-DIPG-021 screen than the HSJD-DIPG-007 screen. This is due to a flaw in experimental design, as both cell lines were screened for equal lengths of time (28 days). This design did not account for the slower growth rate of HSJD-DIPG-021, which can mask the negative selection acting on individual sgRNAs. Nevertheless, sgRNAs targeting the chromodomain were negatively selected against in both cell lines. In the HSJD-DIPG-007 screen many sgRNAs experienced robust depletion similar to those targeting pan-essential genes, notably in the internal disordered region and the C-box domain. To more objectively identify functionally essential domains, data from the HSJD-DIPG-007 screen was analysed using the publicly available computational analysis tool ProTiler (He et al., 2019) (Fig. 5.3B). ProTiler assigns Z-scores to each sgRNA by calculating the mean drop-out of sgRNAs across the gene and measuring the dropout of each sgRNA in terms of standard deviation from the mean. ProTiler identifies weak or inactive sgRNAs by comparing the dropout of each sgRNA to its neighbours (± 5 sgRNAs) and removing the sgRNA if its dropout is significantly weaker than 2/3 of left and 2/3 of right neighbours (grey data points, Fig. 5.3B). Regions that consistently deviate from the mean by at least one standard deviation are considered “hypersensitive” to perturbations and are therefore likely to be functionally essential.

ProTiler analysis of HSJD-DIPG-007 tiling screen data identified the first 108 amino acids encompassing the chromodomain to be hypersensitive to perturbation (Fig. 5.3B), further suggesting that the functional dependency on *CBX4* in H3K27M-DMG could be linked to changes in H3K27me3 distribution. Interestingly, amino acids 220 – 360 in the internal disordered region also appeared to be extra sensitive to perturbation as sgRNAs targeting this region consistently depleted more than the mean, although to a lesser extent than the chromodomain (Fig. 5.3B). As the sgRNAs targeting this central region depleted on average less than one full standard deviation from the mean, this region was designated as “sensitive” rather than hypersensitive. The function(s) of the internal disordered region are not well understood. In contrast to its paralogs, *CBX4* has been uniquely reported to function as an E3 small ubiquitin-like molecule (SUMO) ligase via its SUMO-interacting motifs (SIM1 & SIM2). SUMOylation is a reversible post-translational modification that can affect protein function, for example by altering its activity, localization, or stability. To date, only a few proteins have been reported to be SUMOylated by *CBX4*, including PCGF4 and EZH2. Notably, the SUMOylation of EZH2 has been shown to enhance its catalytic activity, stimulating the deposition of H3K27me3. As this internal sensitive region encompasses SIM1, it could be that the SUMO-ligase activity of *CBX4* is important for disease biology. However, sgRNAs targeting the SIM1 region were

filtered out prior to analysis due to off-target effects, so there was no indication of the negative selection acting on these sgRNAs.

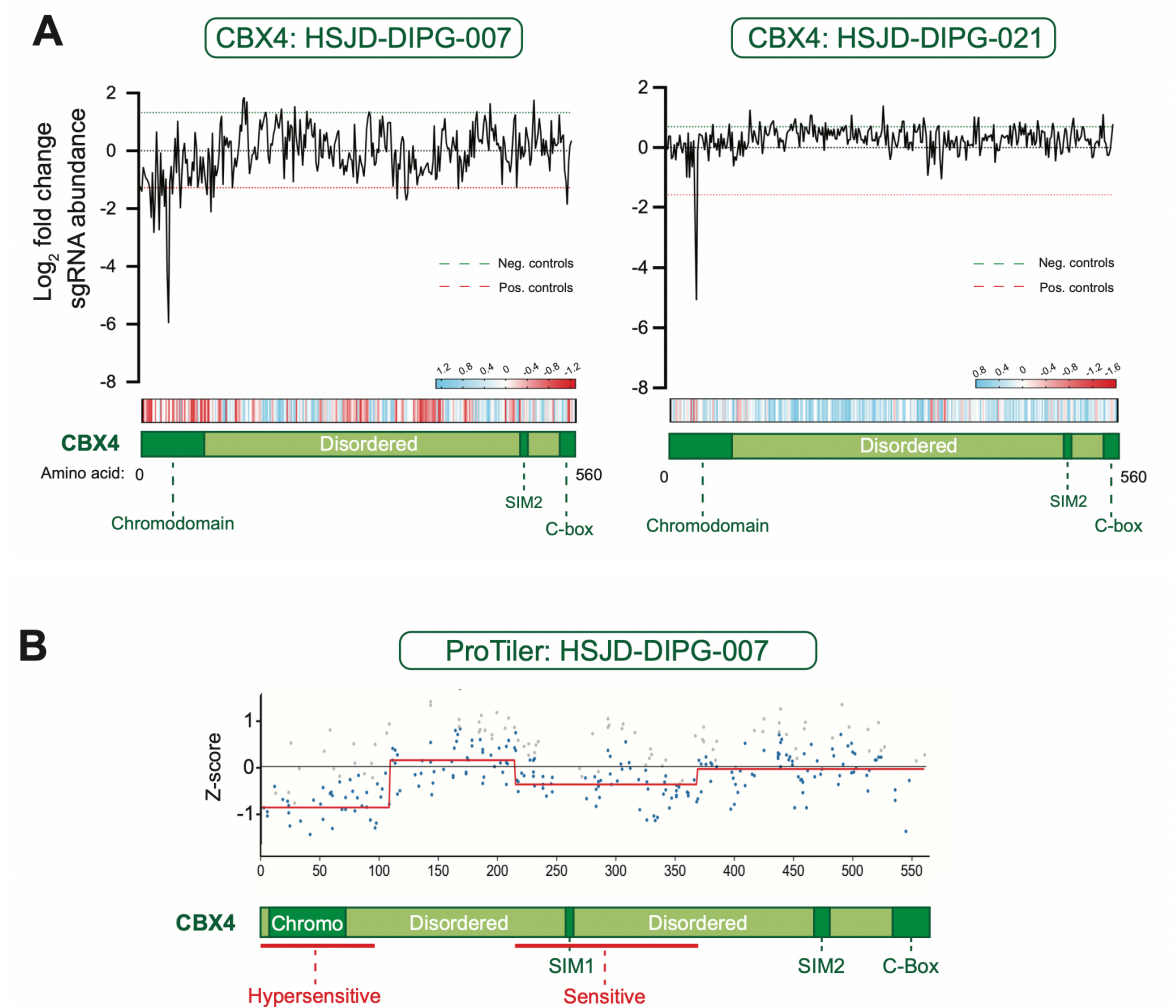


Figure 5.3. CBX4 CRISPR-Cas9 knock-out tiling screens reveal the chromodomain and a section of the internal disordered region are functionally essential

A) Log₂ fold change in sgRNA abundance over the course of the CRISPR-Cas9 tiling screen in HSJD-DIPG-007 and HSDJ-DIPG-021. Average log₂ fold change of positive (red) and negative (green) sgRNAs are indicated by dotted lines on the graph. Heatmaps also indicate log₂ fold change of each sgRNA relative to the average log₂ fold change of the positive (red) and negative (blue) control sgRNAs. Diagram of CBX4 protein is shown below the heatmap. n = 2 replicates for each cell line.

B) ProTiler analysis output of CBX4 CRISPR-Cas9 tiling screen in HSJD-DIPG-007 with a diagram of the CBX4 protein below. Blue dots indicate sgRNAs included in analysis. Grey dots indicate sgRNAs that deviate significantly from neighbouring sgRNAs (+/-5 sgRNAs) and thus are likely lowly active outlier sgRNAs and are excluded from analysis. The red line indicates the average dropout signal across different segments of the protein. Regions that are hypersensitive or sensitive to perturbation are indicated in red below the protein.

5.2.2 Rescue experiments validate the CBX4 chromodomain as functionally essential for H3K27M-DMG disease biology

The results of the CRISPR-Cas9 tiling screen strongly indicated that the CBX4 chromodomain is functionally essential to DMG biology. The tiling screen also indicated that a portion of the internal disordered domain encompassing the SIM1 region could also be functionally essential to H3K27M-

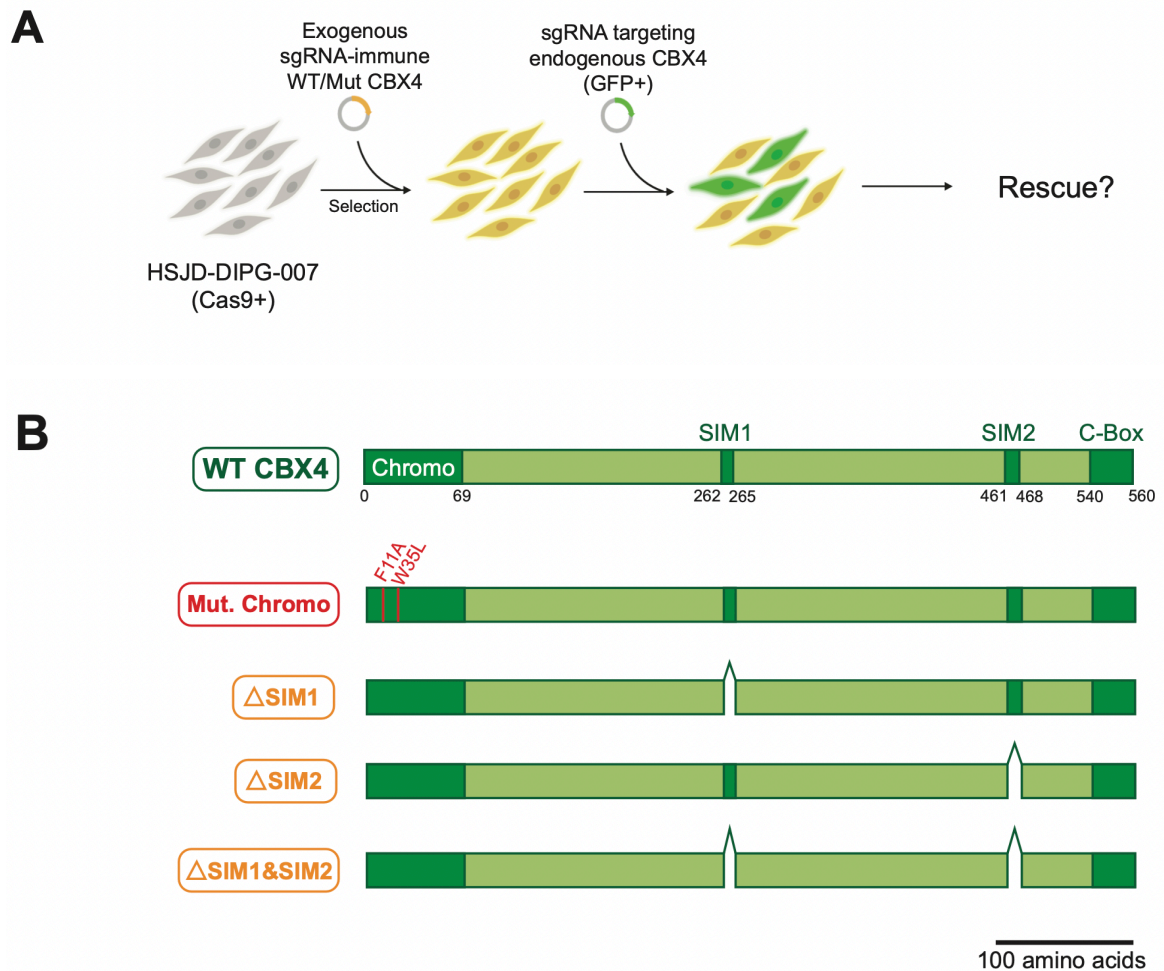


Figure 5.4. Validation of CBX4 tiling screens: CBX4 rescue experiment design

A) Schematic representation of the CBX4 rescue experiment workflow. HSJD-DIPG-007 Cas9-positive cells are lentivirally transduced with a codon-optimised CBX4 (wild-type or mutant) construct resistant to sgRNA targeting. Cells are selected with puromycin to produce a pure population. Cells are then lentivirally transduced with a vector expressing GFP and an sgRNA targeting the endogenous CBX4. The proportion of GFP-positive cells is measured over time using flow cytometry to determine if the exogenous CBX4 construct can rescue the GFP-positive cells from depleting in the population.

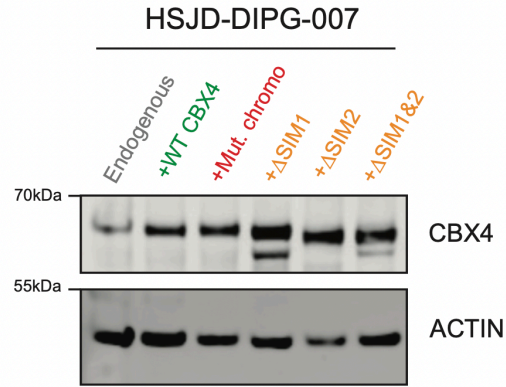
B) Diagrams of the exogenous CBX4 constructs used for rescue experiments.

DMG biology. However, the relevance of the SUMO-E3 ligase activity CBX4 is purported to have remains ambiguous as (1) sgRNAs targeting SIM2 did not experience stronger than average negative selection, and (2) the sgRNAs targeting SIM1 were removed from analysis due to off-target effects. It was therefore important to validate the functional importance of chromodomain and to further investigate if the SIM regions are functionally relevant to H3K27M-DMG biology. Validations were carried out in HSJD-DIPG-007 cells stably expressing Cas9 and an exogenous codon-optimised CBX4 construct resistant to sgRNA targeting (Fig. 5.4A). Codon optimisation of the CBX4 construct had two benefits: (1) the underlying DNA sequence was extensively altered such that there was no complementarity with the *CBX4* sgRNAs, and (2) over expression of the constructs would be boosted by utilising the most abundant amino acid codons. A proportion of cells were then lentivirally transduced with an additional vector expressing GFP and a sgRNA targeting the endogenous *CBX4*

locus, producing a heterogeneous population of GFP-positive and negative cells. The proportion of GFP-positive cells in the population was measured over time using flow cytometry. If the exogenous CBX4 construct has the functionally essential domains for H3K27M-DMG disease biology intact, it is expected to rescue the GFP-positive cells from depletion. It was decided to perform rescue experiments with: wild-type CBX4; CBX4 with a double-point mutant (F11A/W35L) rendering the chromodomain non-functional; CBX4 with SIM1, SIM2 or both SIM1 & SIM2 deleted (Fig. 5.4B).

The exogenous CBX4 constructs were successfully overexpressed in HSJD-DIPG-007 Cas9-positive cells (Fig. 5.5A). The constructs were not tagged to prevent any additional interference other than the perturbation being tested. Interestingly, it seems that deletion of SIM1, which is just 4 amino acids in length (IVIV motif) may affect protein expression and/or processing as an additional truncated CBX4 protein is detected on the immunoblot (Fig. 5.5A). All rescue experiments were performed with a negative control *Rosa26* sgRNA, a positive control *RPA3* sgRNA and 2 independent sgRNAs targeting the endogenous *CBX4* locus. As expected, in all cases cells that received the *Rosa26* sgRNA remained relatively stable in the population while cells that received the *RPA3* sgRNA depleted robustly (Fig. 5.5B). The chromodomain mutant CBX4 construct did not rescue GFP-positive cells from depletion – instead, these cells depleted from the population at the same rate as cells that only expressed endogenous CBX4 (Fig. 5.5B). This validates the functional essentiality of the CBX4 chromodomain for H3K27M-DMG biology. Contrasting this, cells that exogenously expressed wild-type CBX4 achieved partial rescue of GFP-positive cells (Fig. 5.5B). This partial rather than full rescue could be due to the strong over-expression of CBX4. CBX4 normally makes up a tiny minority of total CBX protein in H3K27M-DMG cells. Although CBX4 is functionally essential to H3K27M-DMG biology, it could be that there is an optimal level of CBX4 for maintaining H3K27M-DMG biology and that overexpression of CBX4 could also be stressful for cells. Indeed, there is evidence from experiments performed in mESCs that overexpression of wild-type CBX2 is not well tolerated. Nonetheless, cells that overexpressed exogenous CBX4 with either SIM1 or SIM2 deleted rescued GFP-positive cells to the same extent as exogenous wild-type CBX4 (Fig. 5.5B). This validates that SIM2 is functionally redundant for H3K27M-DMG biology and confirms that while SIM1 is within the hypersensitive internal disordered region, the SIM1 motif itself is functionally redundant for H3K27M-DMG biology. These results suggest that the purported SUMO E3-ligase activity of CBX4 is likely not important for supporting H3K27M-DMG disease biology. This is further reinforced by the SIM1&SIM2 double mutant rescue experiment. Unfortunately, due to a technical fault the final timepoint for *CBX4* sgRNA #1 could not be measured. However, from data at the previous timepoints appears to show a partial rescue of GFP-positive cells similar to that of exogenous wild-type CBX4. Ideally this experiment should be repeated, but taken together with the individual deletions, it appears unlikely that the reported SUMO E3-ligase of CBX4 is essential for H3K27M-DMG biology. More certainly, the functional essentiality of the internal disordered region is likely to be due to some other function that remains to be elucidated.

A



B

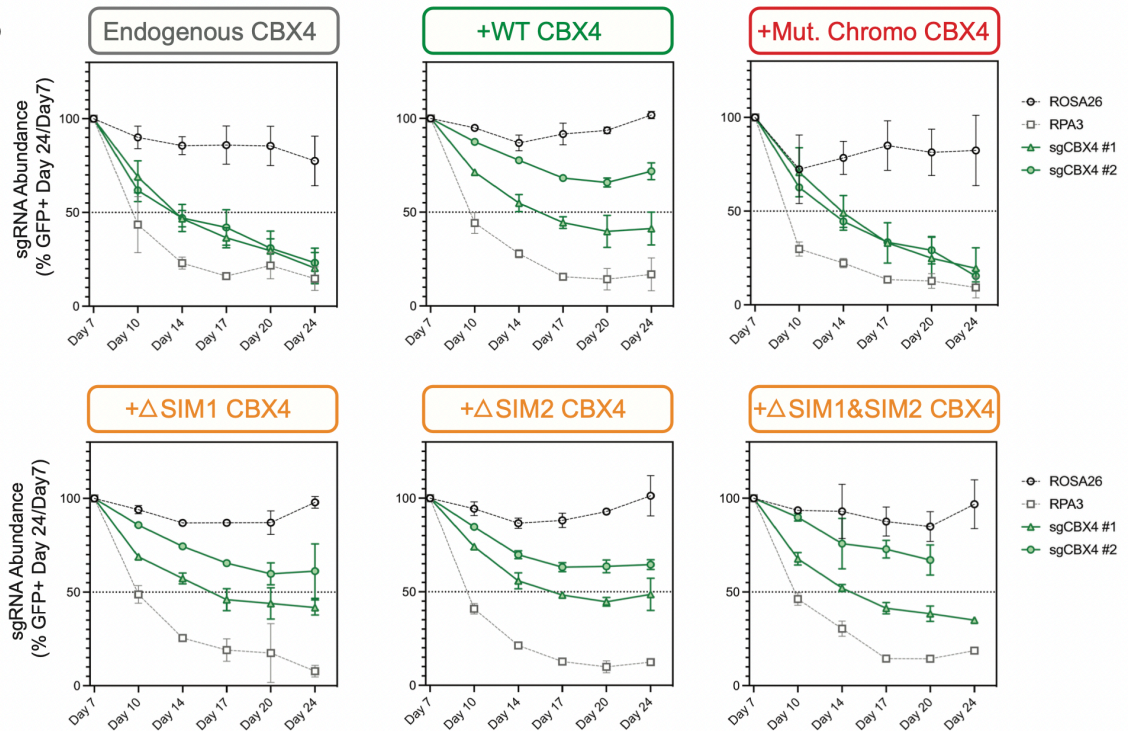


Figure 5.5. CBX4 rescue experiment confirm the chromodomain is functionally essential

A) Immunoblot demonstrating over-expression of indicated exogenous CBX4 constructs in HSJD-DIPG-007 Cas9-positive cells.

B) Rescue experiments showing %GFP-positive cells measured by flow cytometry over time. GFP- positive cells almost entirely deplete from the population when no exogenous CBX4 construct was co-delivered with the sgRNA, or when F11A/W35L double point mutant chromodomain CBX4 was co-delivered. All other CBX4 constructs achieved partial rescue. n = 3 biological replicates.

5.2.3 The CBX chromodomains are highly similar but are not equally sensitive to H3K27me3 loss

Next, it was decided to dissect the functional dependency on the CBX4 chromodomain more thoroughly. In the previous chapter, it was shown that CBX2 and CBX4 behave very similarly in the presence of the H3K27M oncohistone. Structurally, all CBXs have a chromodomain (H3K27me3 reader domain), a C-box (contact site with RING1B) and an internal disordered region (Fig. 5.6A). CBX2 is the most similar to CBX4 in terms of size (Fig. 5.6A). The overall sequence identity of CBX4

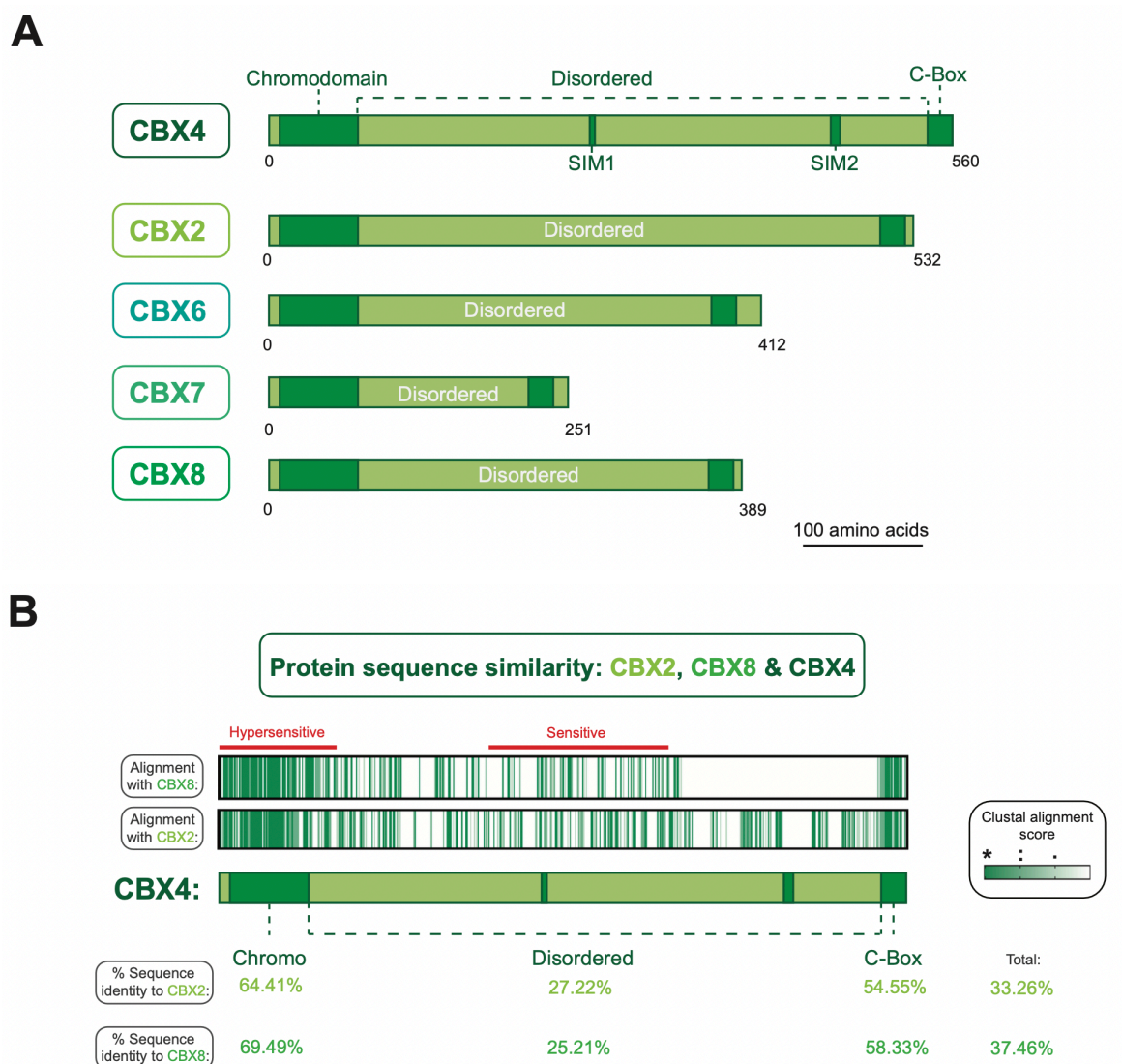


Figure 5.6. Comparison of CBX protein paralogs.

A) Diagram of the CBX paralogs shown to scale. All CBX paralogs have a chromodomain, a C-box and a generally disordered internal region.

B) Heat map (top) showing protein sequence alignment similarity between CBX2/CBX8 and CBX4. CBX4 hypersensitive and sensitive regions identified in the CRISPR-Cas9 tiling screen are indicated in red. Bottom: A diagram of CBX4 with percent sequence identity of indicated domains to CBX2/CBX8.

with CBX2 and CBX8 is quite low (33.26% and 37.46% respectively) mainly due to the large internal disordered region (Fig. 5.6B). The chromodomains are quite well conserved, sharing 64.41% amino acid sequence identity with CBX2 and 69.49% amino acid sequence identity with CBX8 (Fig 5.6B). It is interesting that the CBX4 chromodomain is more similar to the CBX8 chromodomain, yet their chromatin dynamics are opposite in H3K27M cells. In contrast, the CBX2 chromodomain is slightly less similar to the CBX4 chromodomain yet CBX4 and CBX2 exhibit very similar behaviour in the context of H3K27M-induced changes in H3K27me3 distribution. However, despite this similar behaviour and the relatively conserved chromodomain sequence identity, the CBX2 and CBX4 chromodomains differ at more than 35% of amino acid residues. Could these differences be enough to confer some essential, non-redundant function to the CBX4 chromodomain? For example, it is

known that H3K27M-DMG cells are sensitive to inhibition of PRC2 catalytic activity (i.e. EZH2 inhibitors) and subsequent derepression of Polycomb target genes. As mentioned previously, unpublished data from another member of the lab has shown that knock-down of CBX4 in H3K27M-DMG cells, but not the other CBX paralogs, phenocopies EZH2 inhibition or knockdown and leads to derepression of Polycomb target genes. It therefore seems that this aberrant repression of Polycomb target genes is somehow mediated through CBX4-containing cPRC1 complexes. Could subtle differences in the chromodomain confer a certain sensitivity to CBX4 in sensing changes in H3K27me3 that mediates this function? For example, upon EZH2 inhibition and subsequent loss of the remaining H3K27me3, is CBX4 displaced from chromatin more than the other paralogs? Could greater displacement of CBX4 be the mechanism driving the therapeutic effects of EZH2 inhibition in H3K27M-DMG cells?

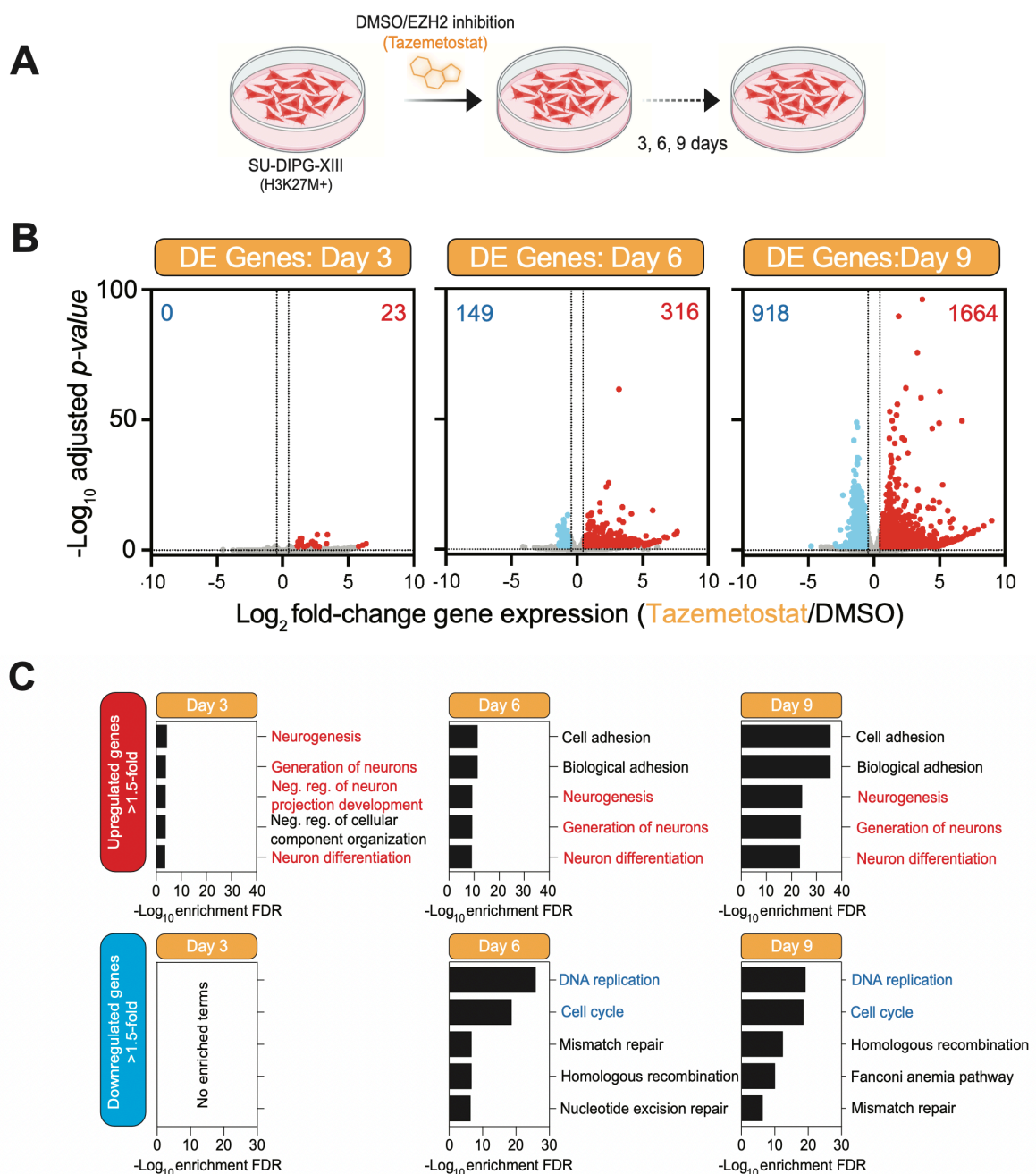


Figure 5.7. EZH2 inhibition with Tazemetostat induces transcriptional changes similar to H3K27M knock-out

A) Schematic representation of DMSO/Tazemetostat (2 μ M) treatment of SU-DIPG-XIII cells. RNA was extracted for sequencing at 3, 6 and 9 days of treatment.

B) Quant-seq data shows differentially expressed genes after 3, 6 and 9 days Tazemetostat treatment compared to DMSO treated cells. $n = 3$ biological replicates. $p < 0.05$, minimum 1.5 fold change in mRNA abundance. Red indicates upregulated genes, blue indicates downregulated.

C) Gene-ontology analyses of differentially expressed genes following 3, 6 and 9 days Tazemetostat treatment. The top 5 categories enriched among upregulated and downregulated genes are shown for each time point.

Before mapping cPRC1 complex dynamics post-EZH2 inhibition, an appropriate dose and treatment length had to be determined. EZH2 inhibition is known to take several days to have transcriptional effects and up to a week to have phenotypic effects. SU-DIPG-XIII cells were treated with DMSO or the FDA-approved EZH2 inhibitor Tazemetostat for 3, 6 and 9 days with RNA extracted for transcriptomic analysis at each timepoint (Fig 5.7A). Consistent with previous observations that EZH2 inhibition requires 5-7 days to elicit a robust phenotypic impact, at day 3 just 23 genes were differentially expressed all of which were upregulated. The number of differentially expressed genes increased sequentially across the time course, with more genes being upregulated than downregulated at all time points (Fig. 5.7B). Gene ontology analyses demonstrated a progressive enrichment of neurodevelopmental terms among upregulated genes; and DNA replication and cell cycle among those downregulated (Fig. 5.7C). This supports previous findings that PRC2 inhibition promotes both the differentiation and cell cycle arrest of H3K27M-DMG cells, and is similar to the transcriptional effects seen when the H3K27M oncohistone is knocked out. Consistent with the on-target activity of Tazemetostat, at 3 days treatment 18 (78%) of these upregulated genes were direct Polycomb target genes (Fig. 5.8A, Fig 5.8B). Following 6 and 9 days of treatment, the number of differentially expressed Polycomb target genes increased sequentially - e.g. cell cycle regulator *CDKN2A*, neurodevelopmental transcription factor *POU4F1* (Fig. 5.8B). This was independently verified with RT-qPCR for the *CDKN2A/INK4A* locus (Appendix VIII). However, the proportion of differentially expressed non-Polycomb target genes also increased sequentially, eventually greatly outnumbering Polycomb target genes at 9 days (Fig. 5.8A). This suggests that while Tazemetostat has on-target activity, secondary transcriptional changes occur with prolonged treatment (Fig. 5.8A).

It was therefore decided to quantitatively map cPRC1 dynamics following 6 day Tazemetostat treatment as robust transcriptional changes occur at this timepoint, many of which are direct Polycomb target genes (Fig. 5.9A). H3K27me3 was almost completely lost from Polycomb target sites following 6 day Tazemetostat treatment (Fig. 5.9B, Fig 5.9C). Consistent with previous studies, SUZ12 was displaced from most Polycomb target sites following Tazemetostat treatment, but was retained at the “strongest” Polycomb target sites, such as the *ZIC1/ZIC4* locus, which had the most

H3K27me3 signal pre-treatment (Fig. 5.9B, Fig 5.9C). Quantitative CUT&RUN-Rx mapping of CBX2, CBX4, CBX8, PCGF2 and PCGF4 showed that most cPRC1 members were displaced to comparable levels following PRC2 inhibition and the associated loss of H3K27me3 (Fig. 5.10A, Fig. 5.9B).

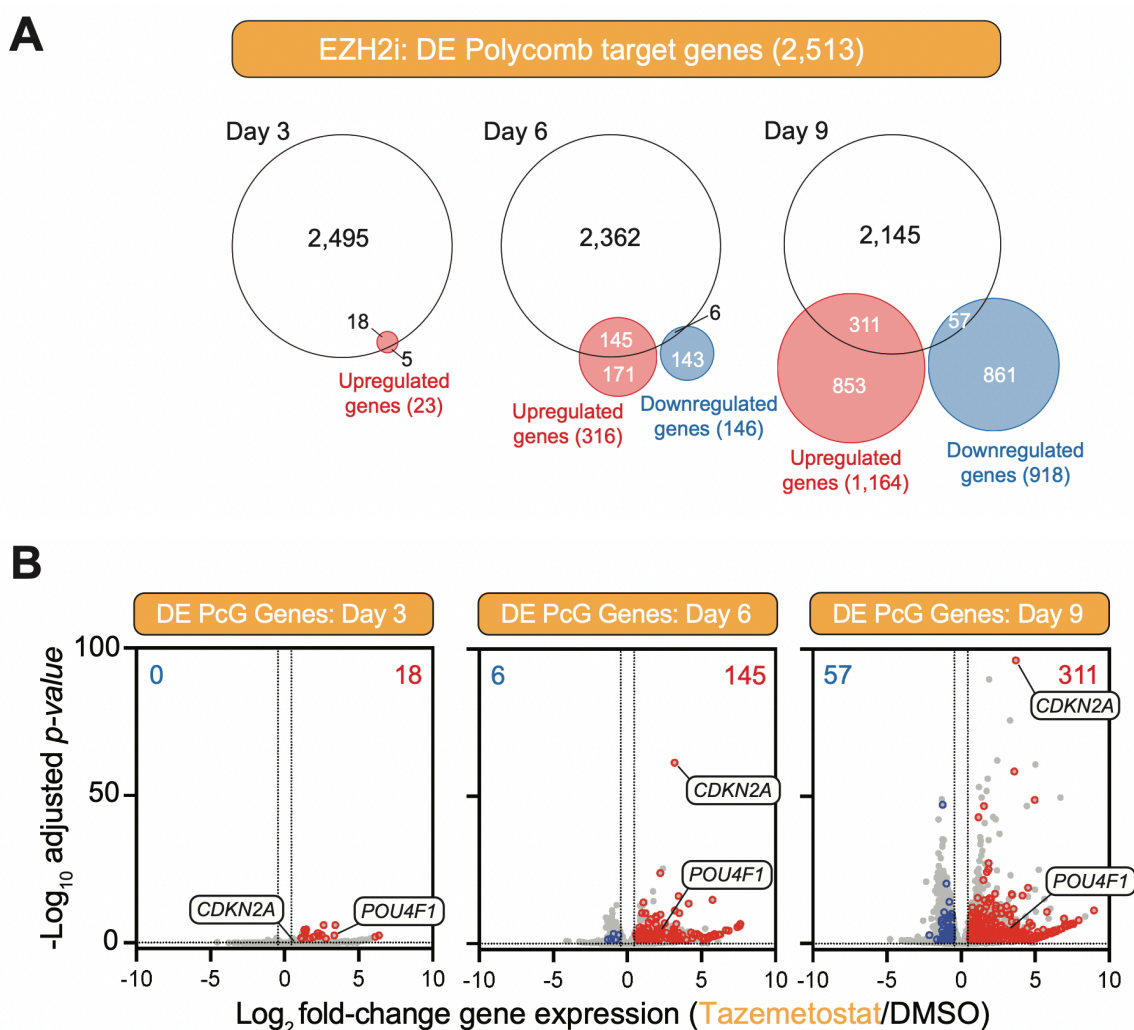


Figure 5.8. EZH2 inhibition with Tazemetostat reverses H3K27M-induced aberrant repression of Polycomb target genes.

A) Venn diagrams depicting the total number of Polycomb target genes and their overlap with genes differentially expressed following 3, 6 and 9 days of Tazemetostat treatment.

B) Volcano plots showing differentially expressed Polycomb target genes at each time point. $n = 3$ biological replicates. $p < 0.05$, minimum 1.5 fold change in mRNA abundance. Red indicates upregulated genes, blue indicates downregulated.

Comparing the average plots for the PCGF paralogs, the scale of PCGF2 displacement from chromatin appears to be slightly less than PCGF4 displacement (Fig. 5.10A). However, this could be due to combination of the fact that there is far more PCGF4 in these cells than PCGF2, and the dramatic accumulation of PCGF4 and general loss of PCGF2 on chromatin observed in these H3K27M-DMG cells – put simply, a slightly greater magnitude of PCGF4 loss could just be because there is more PCGF4 on chromatin to be displaced in these cells. Comparing the average plots for each CBX paralog however, CBX2 is noticeably displaced to a lesser extent than its paralogs (Fig. 5.10A). While there is large displacement of CBX2 at peak centres, there is hardly any loss of CBX2 signal beyond peak centres, while there is large signal loss at both peak centres and beyond for CBX4 and CBX8 upon loss of H3K27me3 (Fig. 5.10A). This could suggest that CBX2 is greatly

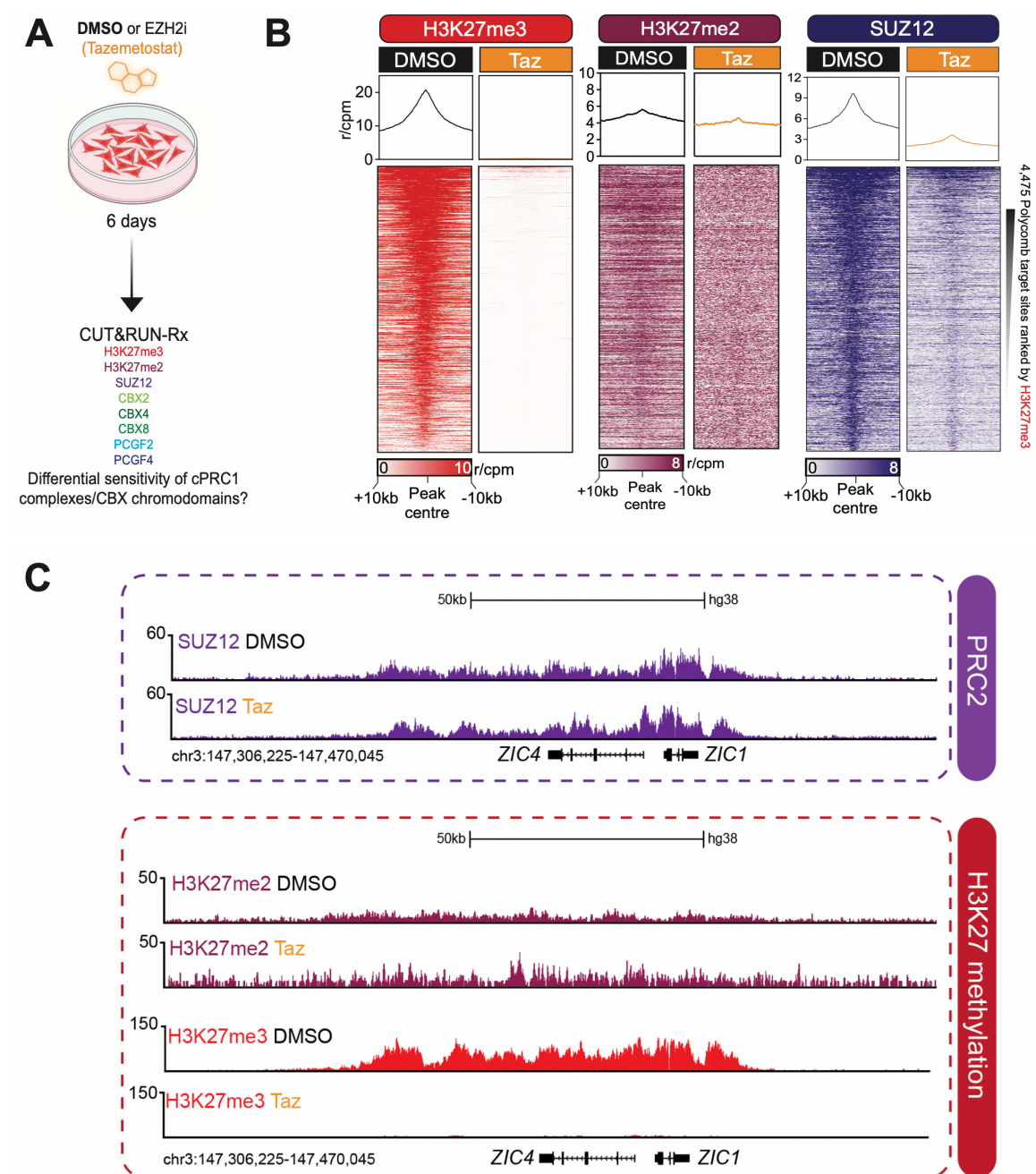


Figure 5.9. Mapping the consequences of Tazemetostat treatment on PRC2 and cPRC1 chromatin occupancy with CUT&RUN-Rx

A) Schematic representation of DMSO/Tazemetostat (2 μ M) treatment of SU-DIPG-XIII cells for 6 days followed by quantitative mapping of PRC2 and cPRC1 complex components with CUT&RUN-Rx.

B) Tornado plots and average plots showing average CUT&RUN-Rx signal (n = 3) for H3K27me3 and SUZ12 in DMSO and Tazemetostat treated SU-DIPG-XIII H3K27M cells at Polycomb target sites. Polycomb target sites are ranked by H3K27me3 CUT&RUN-Rx signal in SU-DIPG-XIII H3K27M cells.

C) Genome browser tracks showing average CUT&RUN-Rx signal for H3K27me3 and SUZ12 at the indicated genomic locus in DMSO and Tazemetostat treated cells.

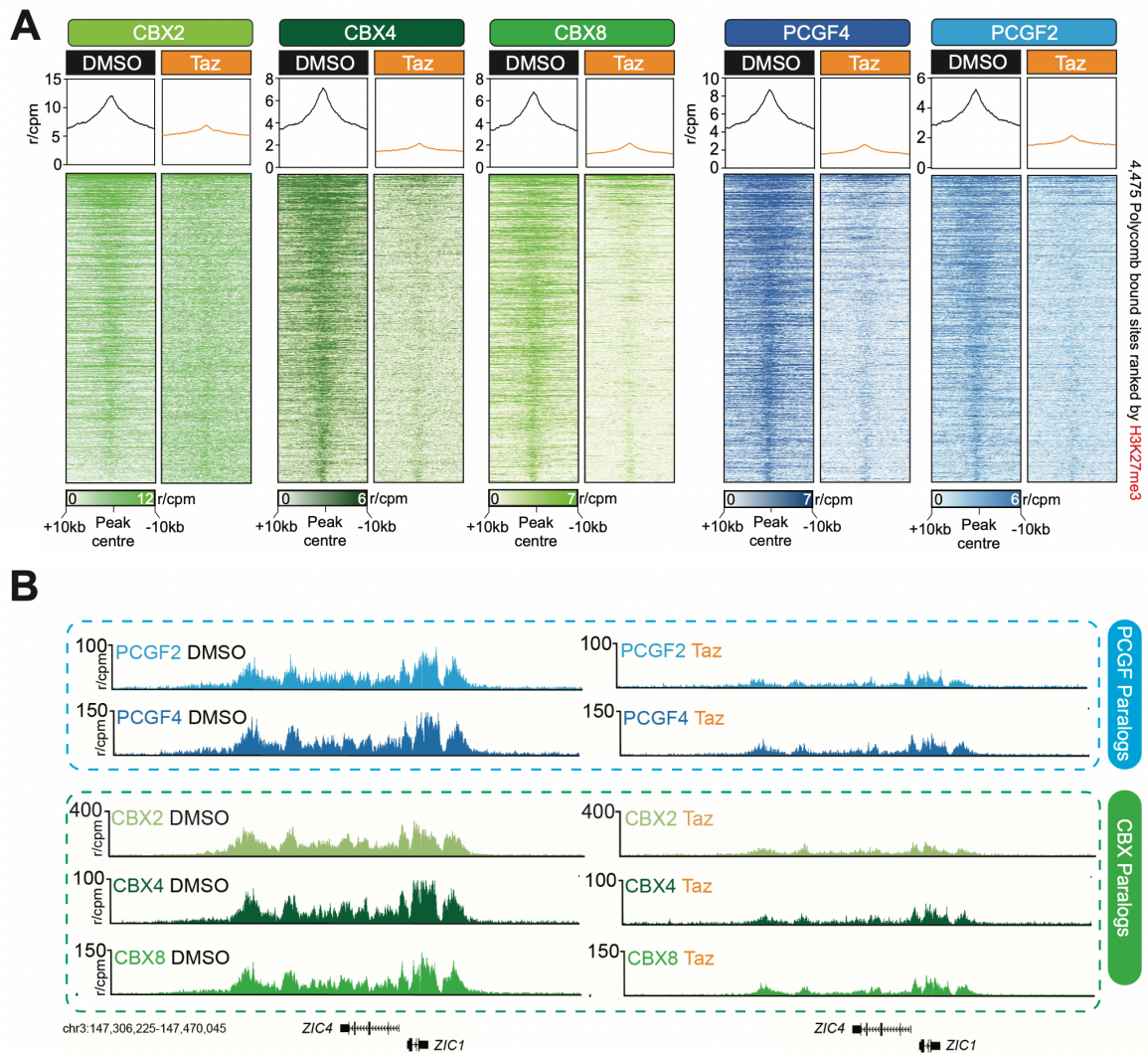


Figure 5.9. CUT&RUN-Rx shows Tazemetostat treatment displaces all cPRC1 complexes from chromatin.
A) Tornado plots and average plots showing average CUT&RUN-Rx signal ($n = 3$) for CBX2, CBX4, CBX8, PCGF2 and PCGF4 in DMSO and Tazemetostat treated SU-DIPG-XIII H3K27M cells at Polycomb target sites. Polycomb target sites are ranked by H3K27me3 CUT&RUN-Rx signal in SU-DIPG-XIII H3K27M cells.
B) Genome browser tracks showing average CUT&RUN-Rx signal for H3K27me3 and SUZ12 at the indicated genomic locus in DMSO and Tazemetostat treated cells.

displaced from the strongest Polycomb target sites like its paralogs but is possibly lost to a lesser extent or not at all at weaker sites. Indeed, dramatic loss of CBX2, CBX4 and CBX8 at strong Polycomb target site *ZIC1/ZIC4* is observed (Fig. 5.10B).

The relationship between H3K27me3 and cPRC1 component displacement was investigated further. Loss of all CBX and PCGF paralogs anti-correlated with H3K27me3 i.e. Polycomb target site “strength” (Fig. 5.11A). All paralogs were retained or lost at all Polycomb target sites, with the notable exception of CBX2 (Fig. 5.11A). While CBX2 was displaced from the strongest Polycomb target sites to similar extents as its paralogs, the vast majority of Polycomb target sites retained or even gained CBX2 occupancy (Fig. 5.11A) despite near total H3K27me3 loss (Fig. 5.9A). Interestingly, H3K27me2 was also largely retained or slightly increased at most Polycomb target sites (Fig. 5.11A).

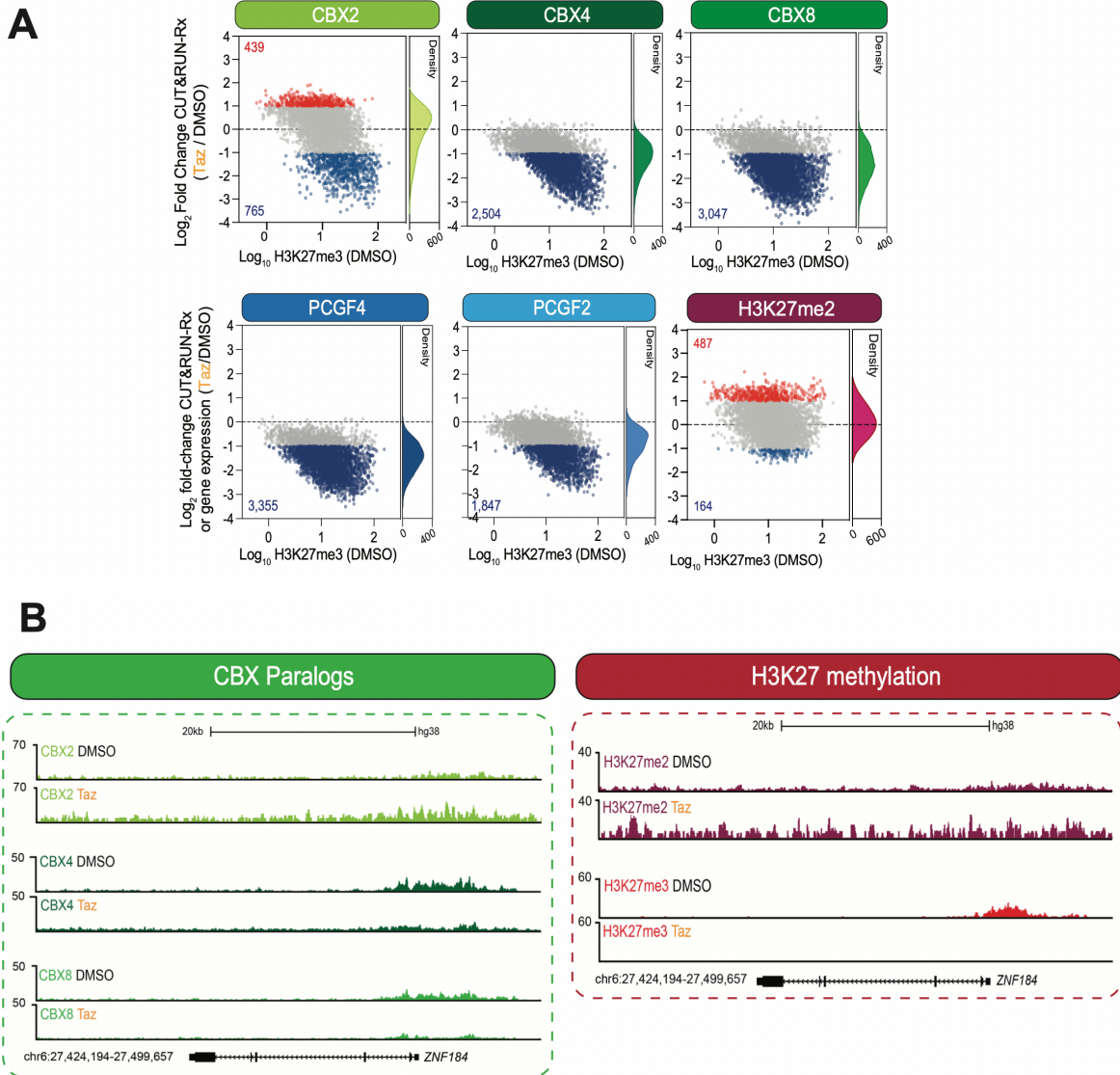


Figure 5.10. CUT&RUN-Rx shows displacement of cPRC1 complexes from correlated with loss of H3K27me3 and Polycomb gene derepression.

A) MA-plots showing $\log_2$ fold-change in binding for CBX2, CBX4, CBX8, PCGF2, PCGF4 and H3K27me3 (y-axis) at Polycomb target sites in SU-DIPG-XIII H3K27M treated with 2 μ M Tazemetostat for 6 days. Sites changing >2-fold are shown in red (increasing) or blue (decreasing) with the number of sites in these categories also indicated. The densities of chromatin binding changes are shown on the right of each plot.

B) Genome browser tracks showing average CUT&RUN-Rx signal for CBX2, CBX4, CBX8, H3K27me2 and H3K27me3 at *ZNF184*, a designated quartile 2 (Q2) Polycomb target gene, in DMSO and Tazemetostat treated cells.

The greatest retention/increase of both CBX2 and H3K27me2 appeared to occur at weaker Polycomb target sites. This perhaps suggests that CBX2 may have some affinity for H3K27me2 that the other CBX paralogs do not have. These dynamics can be seen at the *ZNF184* locus, a weaker Polycomb target Q2 gene than the Q1 *ZIC1/ZIC4* locus (Fig. 5.10C). Tazemetostat treatment results in loss of H3K27me3 and displacement of CBX4 and CBX8, but an increase in H3K27me2 and CBX2. Interestingly, Tazemetostat treatment appears to restore H3K27me2 to a more diffuse pattern typically seen in H3 wild-type cells. Taken together, these data suggest that despite the high

sequence similarity between CBX chromodomains, they are differentially sensitive to H3K27me3 loss and may even have different affinities for different degrees of H3K27 methylation.

Displacement of CBX2, CBX4, PCGF2 and PCGF4 correlates strongly with an increase in Polycomb target gene expression following Tazemetostat treatment (Fig. 5.11A). CBX2 dynamics do not follow this trend, as it is increased or retained on chromatin at Q2-4 genes which are derepressed. This is interesting, as although both CBX2 and CBX4 accumulate on chromatin in H3K27M cells, while CBX2 may be contributing somewhat to gene repression it is clearly not the main driver of repression in H3K27M cells. The CUT&RUN-Rx data presented here provide a clear picture of the dynamics that occur upon PRC2 inhibition (Fig. 5.11B). Inhibition of PRC2 in H3K27M cells leads to a decrease of PRC2, CBX4-cPRC1 and CBX8-cPRC1 on chromatin at Polycomb target sites, and displacement of these complexes correlates with Polycomb target gene derepression. On average, CBX2-cPRC1 remains on chromatin, although its distribution changes. This suggests that it is the accumulation of CBX4 that is the main driver of aberrant gene repression in H3K27M cells, and although most cPRC1 is displaced from chromatin upon PRC2 inhibition, the therapeutic effects of PRC2 are likely mediated through the displacement of CBX4-containing cPRC1. However, while the chromodomain is clearly an essential aspect of the function of CBX4 in H3K27M-DMG disease biology, it is not unique either in its response to H3K27me3 pattern redistribution in H3K27M (where CBX2 accumulates similarly) nor in its response to H3K27me3 loss (where CBX8 is displaced similarly). Therefore, it is likely that another aspect of CBX4 that contributes to its unique ability to aberrantly repress Polycomb target genes in this specific context.

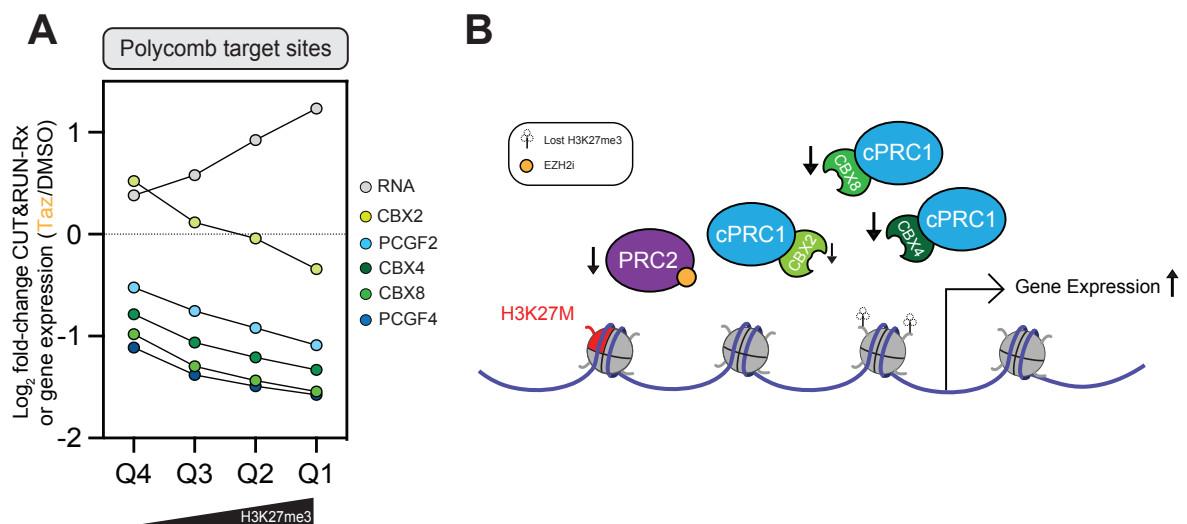


Figure 5.11 Polycomb target gene derepression correlates with cPRC1 displacement

A) Changes in gene expression and chromatin occupancy of cPRC1 complexes at Polycomb target genes, quartiles ranked by H3K27me3 signal in SU-DIPG-XIII H3K27M cells. Shown are the mean changes of the indicated cPRC1 members and RNA levels relative to control (DMSO) treated cells.

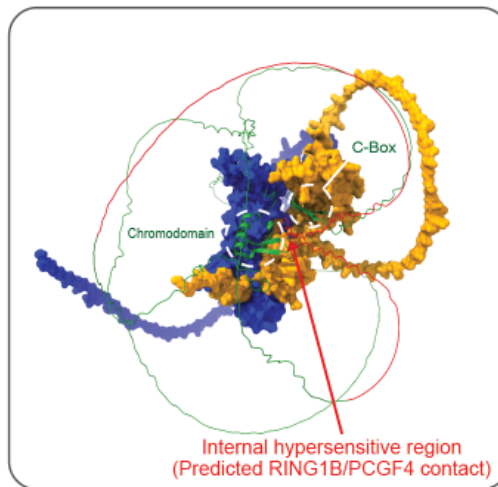
B) Schematic representation of the consequences of EZH2 inhibitor treatment in H3K27M cells. EZH2i results in an almost complete loss of H3K27me3 and a general displacement of PRC2 and cPRC1 complexes at strong (Q1) Polycomb target genes. An exception to this is CBX2, which is retained at weaker (Q2-Q4) Polycomb target genes. Displacement of cPRC1 complexes correlates with derepression of Polycomb target genes.

5.2.4 The CBX4 internal disordered region is uniquely predicted to contact core cPRC1

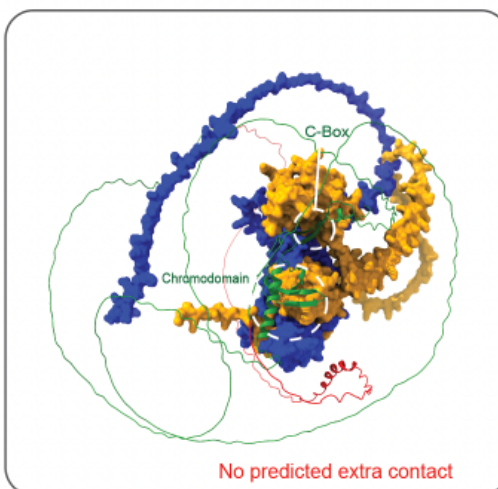
The CRISPR-Cas9 tiling screen indicated a ~140 amino acid section of the internal disordered region of CBX4 is likely functionally important for H3K27M-DMG biology. The internal disordered region is known to play an important role in the compaction and phase separation functions of CBXs (Kim & Kingston, 2020). This is particularly well-characterised for CBX2 (discussed further in Chapter 6). However, CBX2 and CBX4 have been shown to have similar phase separation and compaction properties, at least *in vitro* (Brown et al., 2023). Yet, CBX2 is functionally redundant in H3K27M-DMG biology. In general, very little is known about the function of the internal disordered regions of CBXs compared to the better characterised chromodomain and C-box domains. It is important not to exclude the possibility of other, unknown functions beyond phase separation and chromatin compaction. Disordered protein domains tend to evolve more rapidly than structured domains as they are not subject to structural constraints (Brown et al., 2011). Indeed, there is low protein sequence conservation of the internal disordered region amongst CBX4 and its paralogs (Fig. 5.6B). This raises the possibility of unique function(s) of the CBX4 internal disordered domain which could potentially be important for H3K27M-DMG biology.

To facilitate hypothesis generation and inform future experimental design, AlphaFold2-multimer was used to model the structure of core PRC1 (RING1B and PCGF4) in complex with CBX2, CBX4 or CBX8. Consistent with expectations, the predictions indicate that RING1B and PCGF4 form a dimer, with the C-box of each CBX protein interacting with this cPRC1 core dimer (Fig 5.13A, 5.13B, 5.13C). Interestingly, the 'sensitive' internal disordered region of CBX4 identified in the tiling screen was also predicted to contact RING1B & PCGF4 at their binding interface (Fig. 5.13A). This was not predicted for the equivalent region of CBX2 (Fig. 5.13B) or CBX8 (Fig. 5.13C). AlphaFold2-multimer simulates five predictions ranked by confidence. Predicted alignment error (PAE) is a measure of confidence in the relative position of two residues within the structure. A lower PAE indicates higher confidence. The structure of disordered protein domains are difficult to capture and therefore difficult to predict, and so will typically have very high PAE scores. All five predictions for CBX4 predicted a contact between the sensitive internal disordered region and core cPRC1 to a low-to-medium degree of confidence (Fig. 5.14A). Notably, no extra contact point was predicted in any of the five simulations for CBX2 (Fig. 5.14B) or CBX8 (Fig. 5.14C). While little weight should be given to predicted models with no supporting data, these models are a useful tool for generating hypotheses around what the function of this sensitive internal region of CBX4 could be and designing experiments to address these hypotheses. This will be discussed further in the following discussion section.

RING1B-PCGF4-CBX4



RING1B-PCGF4-CBX2



RING1B-PCGF4-CBX8

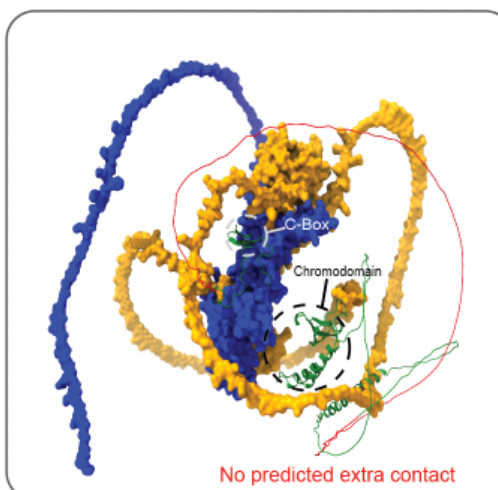


Figure 5.12. CBX4 internal disordered region is predicted to have an extra contact point with core cPRC1. AlphaFold multimer prediction of structure and folding of RING1B (yellow) and PCGF4 (blue) with CBX4 (top), CBX2 (middle), or CBX8 (bottom) in green. Chromodomain and C-box are indicated for each. The CBX4 sensitive internal disordered region identified in the CRISPR-Cas9 tiling screen is shown in red, with the equivalent regions in CBX2 and CBX8 also shown in red.

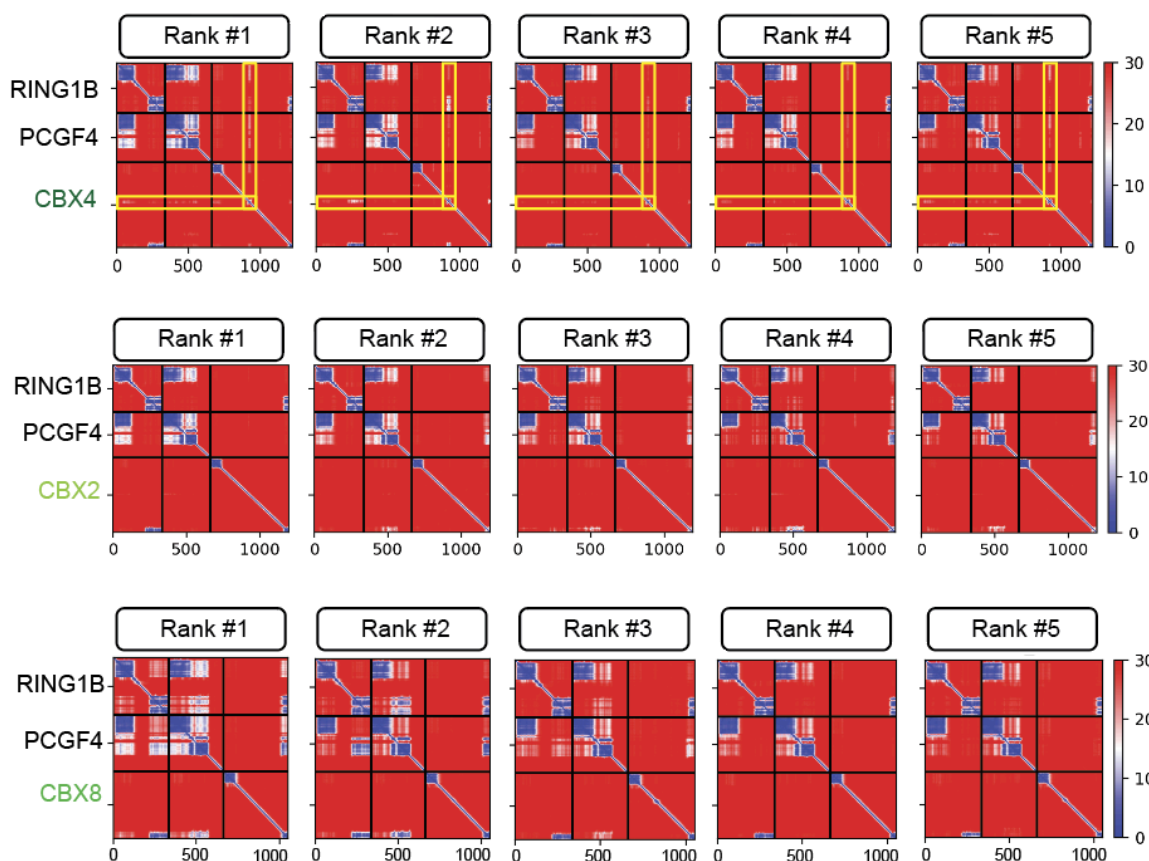


Figure 5.13. Only CBX4 is consistently predicted to contact core cPRC1

Heatmaps showing predicted alignment error (PAE). A lower score indicates higher confidence. The internal disordered region of CBX4 is consistently predicted to contact core cPRC1 with low-medium confidence (highlighted in yellow). No extra contact point is predicted for CBX2 or CBX8 in any model.

5.3 Discussion

In this chapter, functionally essential domains of CBX4 in H3K27M-DMG were identified. The functional essentiality of the chromodomain H3K27me3 reader ability was validated using an orthogonal assay, although the importance of the central disordered domain remains to be validated with rescue experiments.

It is interesting that the output of ProTiler did not seem to indicate that the C-box domain of CBX4 is sensitive to perturbation. This could indicate that some other cPRC1 independent function of CBX4 is important for H3K27M-DMG biology. Of potential relevance to this, CBX4 has been reported to have some distinct functions from other CBX proteins that could be independent of cPRC1. CBX4 has been reported to act as an E3 SUMO ligase, to be recruited to sites of DNA damage, and to bind the transcriptional co-repressor CtBP (Ismail et al., 2012; Kagey et al., 2003). However, it was validated here that the proposed SUMOylation activity of CBX4 is not an essential function in H3K27M-DMG disease biology. Additionally, although the C-box was not highlighted as particularly sensitive to perturbation by ProTiler, it is worth noting that the sgRNAs targeting the C-box do appear to deplete more than average in all plots (Fig. 5.3A & 5.3B). This indicates that the association of CBX4 likely is important for its function, rather than cPRC1-independent functions. It is likely that this domain was not highlighted by ProTiler as just 4 sgRNAs target this domain, and each sgRNA is compared to its neighbouring sgRNAs. Therefore, the increased sensitivity of this domain is likely being averaged out by its less essential neighbouring regions. It will be necessary to perform rescue experiments using exogenous CBX4 constructs with the C-box deleted to confirm this hypothesis.

Additionally, in this chapter I have demonstrated that the CBX paralogs are not equally sensitive to H3K27me3 loss. Building on the results presented in Chapter 4, this further implies differences in the methylation sensing abilities of different CBX chromodomains. CBX4 and CBX8 are displaced from chromatin with similar magnitudes following PRC2 inhibition, which correlates strongly with the loss of H3K27me3. This suggests that the chromodomains of these paralogs are capable of sensing H3K27me3, as they are both displaced from chromatin in its absence. However, the dynamics observed in Chapter 4 imply that the patterning of H3K27me3 may cause CBX4 and CBX8 to sense this modification differently. In contrast, the results presented in this chapter indicate that the CBX2 chromodomain also senses H3K27me3, as it is displaced at sites that experience significant H3K27me3 loss upon PRC2 inhibition. However, unlike CBX4 and CBX8, CBX2 is retained and even increased at most Polycomb target sites. This retention/increase of CBX2 correlates with the retention/increase of H3K27me2 at these sites. This retention/increase of H3K27me2 is likely because PRC2 has a higher catalytic efficiency for H3K27me2 than H3K27me3, and upon PRC2 inhibition H3K27me2 can still be catalysed but is no longer converted to H3K27me3. This likely leads to an accumulation of H3K27me2, which might in turn allow CBX2 to remain bound to chromatin. These findings suggest that the CBX2 chromodomain is capable of binding H3K27me2, a previously unreported observation that warrants further validation, potentially through *in vitro* binding assays (discussed further in Chapter 6).

Here I have shown that PRC2 inhibition leads to progressive derepression of Polycomb target genes and other transcriptional changes, in agreement with previous work which demonstrated that the aberrant repression of Polycomb target genes in H3K27M-mutant cells can be reversed by PRC2 inhibition (Brien et al., 2021; Mohammad et al., 2017). Gene ontology (GO) analyses demonstrated a progressive enrichment of neurodevelopmental terms among upregulated genes and DNA replication and cell cycle terms among downregulated genes. This supports previous findings that PRC2 inhibition promotes both the differentiation and cell cycle arrest of H3K27M-DMG cells. As described in Chapter 4, work performed by my colleague AJ Silva demonstrated that only knock-down of CBX4 in H3K27M-DMG cells led to derepression of Polycomb target genes, such as *POU4F1* and *ZIC1*. Although cPRC1 complexes containing CBX4 are not uniquely displaced from chromatin upon PRC2 inhibition, based on the results of this knockdown experiment and the results presented here, I hypothesise that the effects of PRC2 inhibition are likely mediated specifically via displacement of CBX4 containing cPRC1 complexes.

How exactly CBX4 mediates aberrant gene repression in H3K27M-DMG cells remains to be elucidated. Accumulation on chromatin alone cannot be sufficient, as non-essential paralog CBX2 also accumulates on chromatin in H3K27M cells, and largely remains on chromatin following PRC2 inhibition. Therefore, there must be some other unique feature pertaining to CBX4 that contributes to this aberrant repressive function. The results of the tiling screen suggest that the internal disordered region may play a role in this. The AlphaFold models predict that this region of CBX4 uniquely contacts core cPRC1. If this is validated, for example using cross-linking mass spectrometry, then this extra contact point could have a number of effects on the complex – for example, it could impact complex stability, conformation, or composition. Indeed, early preliminary experiments suggest that this functionally essential internal disordered region might influence the composition of the cPRC1 complex, defining a discrete complex make-up. Overexpression of wild-type CBX4 and CBX4-internal mutant constructs in HEK293T cells followed by IP-mass spectrometry indicated that these constructs co-precipitated very different cPRC1 complexes. Wild-type CBX4 co-precipitated only RING1A/B, PCGF4, and PHC2, while mutant CBX4 co-precipitated these and other paralogs – including PCGF2, PHC1 and PHC3 (data not shown). These results remain to be validated, but possibly gives an indication as to what could be molecularly distinct about CBX4 that is essential in H3K27M-DMG biology. Regardless, even if this result validates, this raises the question – how mechanistically does this specific cPRC1 complex maintain aberrant gene repression in H3K27M-DMG? For example, are there unique compaction or phase-separation capabilities associated with CBX4-containing cPRC1 complexes that mediate gene repression in this context? This will be discussed further in Chapter 6.

CHAPTER 6

General Discussion

6.1 Introduction

To briefly summarise, in Chapter 3, *CBX4* and *PCGF4* were identified as genetic dependencies in H3K27M-DMG using unbiased genome-wide CRISPR-Cas9 screening approaches. In Chapter 4, quantitative chromatin mapping demonstrated that H3K27M redistributes cPRC1 complexes on chromatin. In Chapter 5, genetic screening revealed functionally essential domains of *CBX4* which could be useful targets for future potential therapeutic development. Additionally, quantitative chromatin mapping indicated that different CBX chromodomains may have previously unappreciated unique methylation sensing abilities. These findings will now be discussed in more detail below.

6.1.1 Understanding the molecular mechanisms driving disease biology is important for the approval of effective oncology treatments

Prior to the mid-20th century, the treatment of cancer was primarily limited to surgical intervention or cauterization for the removal of non-invasive solid tumours (Falzone et al., 2018). However, patients with advanced or metastatic disease had very limited therapeutic options. The first chemotherapeutic agents were identified during World War II, as a result of research into the effects of chemical warfare agents. The first chemotherapy agent to receive FDA approval was mechlorethamine hydrochloride, a nitrogen mustard compound, in 1949. This approval was based on a 1946 study that demonstrated the efficacy of nitrogen mustards in the treatment of lymphomas (Goodman et al., 1946).

For the following several decades, cancer therapies were largely restricted to systemic cytotoxic agents that exhibited a preferential effect on rapidly dividing cells (Scott et al., 2023). These chemotherapeutic agents lack specificity, resulting in significant toxicity to healthy tissues. However, the advent of new targeted therapies in the late 1990s and early 2000s marked a significant paradigm shift in the treatment of cancer (Scott et al., 2023). Two key milestones in this new era of cancer treatment were the approval of an anti-HER2 antibody for the treatment of breast cancer in 1998 and the approval of imatinib, a small molecule BCR-ABL inhibitor, for the treatment of chronic myeloid leukaemia in 2001 (Scott et al., 2023). These novel therapies were designed to exploit specific genetic vulnerabilities in cancer cells, thereby enabling more selective and effective treatment. These breakthroughs mark the beginning of precision oncology, where treatments are tailored to the specific characteristics of individual cancers.

Since then, there is a growing recognition of the importance of using the molecular biology of the disease to guide treatment decisions to achieve more effective and less toxic therapies. However, despite this understanding, the majority of patients have not benefited from targeted therapies in the decades since the approval of imatinib. The development of targeted cancer drugs has been hindered by a high rate of attrition, with up to 90% of candidates failing to gain regulatory approval often due to lack of efficacy (Behan et al., 2019; Nelson et al., 2015). However, there has been a notable shift in recent years. In 2014, 40% of drugs that received FDA approval were supported by genetic evidence linking the drug target to the specific biology of the disease. By 2021 and 2022, this proportion had increased to 60% (Rusina et al., 2023). As more functional genetic and genomic data

become available, it is expected that the number of approved medicines that exploit genetically defined mechanisms will continue to rise (Scott et al., 2023).

The concept of "synthetic lethalties" refers to context-dependent essentialities, where a specific gene or pathway is essential for viability only in the presence of a particular mutation. This concept was first described in studies involving *Drosophila* and yeast, where crosses between two viable mutant cell lines resulted in non-viable offspring harbouring both mutations, indicating a lethal interaction between the two mutations (Konda et al., 2023). This principle was later applied to cancer research, with the aim of identifying cancer-selective genetic vulnerabilities that can be targeted for therapeutic intervention.

A notable example of the successful exploitation of synthetic lethalties in cancer therapy is the use of PARP inhibitors in patients with mutations in the tumour suppressor genes *BRCA1* and *BRCA2* (Farmer et al., 2005; Lord & Ashworth, 2017). *BRCA1* and *BRCA2* play a critical role in the homologous recombination (HR) repair pathway, a key mechanism for maintaining genomic stability. The hypothesis was that *BRCA*-deficient cells harbouring mutations in these genes would be differentially sensitive to inhibition of *PARP1*, a different DNA repair enzyme, due to their compromised HR repair pathway. In contrast, healthy tissues with an intact HR repair pathway would be expected to be less sensitive to PARP inhibition. Consistent with this hypothesis, *BRCA*-deficient cells have been shown to be uniquely sensitive to PARP inhibitors, which have been approved for the treatment of multiple *BRCA*-deficient cancer types, including breast, ovarian, and pancreatic cancers (Fong et al., 2009; Konda et al., 2023). This therapeutic strategy demonstrates the potential of synthetic lethalties to inform the development of targeted cancer therapies. The discovery of *BRCA1/2* and *PARP1* synthetic lethality was hypothesis-led. One major obstacle for developing new targeted cancer therapeutics has been the limited knowledge of synthetic lethalties and potential therapeutic targets (Behan et al., 2019). Unbiased screening approaches have enabled the discovery of synthetic lethalties linked to specific cancer disease biology (Huang et al., 2020). Genetic screening approaches employed in the work presented in this thesis have allowed the discovery of a synthetic lethality in H3K27M-DMG and broadened the horizon for potential therapeutic targets for this devastating paediatric cancer.

6.2 Functional genomic screens identify *CBX4* as a mechanistically-anchored potential therapeutic target for H3K27M-DMG

The discovery of RNA interference (RNAi) in *C. elegans* in the late 1990s marked a significant breakthrough in the field of cancer research (Fire et al., 1998). This is because RNAi became a crucial tool for conducting functional genetic studies in cancer models. For instance, RNAi has been used to demonstrate that the amplification of genes such as *MITF* in melanoma and *SOX2* in oesophageal cancer are associated with a dependency on these genes (Bass et al., 2009; Garraway et al., 2005). In recent years, CRISPR-Cas9 has emerged as a more robust and efficient method for

large-scale genetic manipulation in cancer research (Shalem et al., 2014). Both RNAi and CRISPR-Cas9 have been employed to systematically identify cancer dependencies in both *in vitro* and *in vivo* models. Several large-scale initiatives, such as Project Achilles and Project DRIVE, have utilized genome-scale RNAi screening and to investigate cancer dependencies across hundreds of cancer cell lines (McDonald et al., 2017; Tsherniak et al., 2017). The Cancer Dependency Map (DepMap) project, which uses CRISPR-Cas9 knockout screening to identify cancer dependencies, has expanded to include over 1,000 cell lines from over 45 lineages, with accompanying DNA and RNA sequencing data available for the screened cell lines. These datasets have validated known oncogene dependencies and revealed new selective dependencies (Konda et al., 2023). These unbiased functional genetic approaches have facilitated the systematic discovery of numerous potential cancer dependencies, which require validation using complementary perturbation techniques to confirm their relevance.

However, not all malignancies are equally represented in large-scale screening endeavours such as DepMap. Hundreds of screens have been performed in haematological cancers, which are relatively common and are easy to grow and scale-up in suspension culture. In contrast, other cancer types are less represented, possibly because they are rare or challenging to culture and scale up. Notably, at the time of this study, no genome-wide screen had been published using a H3K27M-DMG cell line. To address this gap, I performed a genome-wide knockout screen in a H3K27M-DMG cell line (Chapter 3), which identified *PCGF4* and *CBX4* as genetic dependencies. This result has since been validated in cell lines by myself and Dr Dáire Gannon using growth competition assays. The *CBX4* dependency has also been validated in other cell lines by AJ Silva in the lab and collaborators in the Phillips lab, Philadelphia, USA (data not shown). Additionally, genome-wide screens in H3K27M-DMG cell lines have now been published and independently confirmed these findings (Pal et al., 2022; Wenger et al., 2023). At the time of this study, only H3K27M-DMG cell lines were available in our lab, meaning I could not perform screens and validations in isogenic H3K27M-KO cell lines. As a result, I could not determine whether loss of *CBX4* and/or *PCGF4* are synthetic lethal with the H3K27M oncohistone mutation or if they are simply lineage-specific dependencies. Given the upregulation of *PCGF4* in H3K27M cells demonstrated by immunoblot, it seemed plausible that this dependency could be lineage specific or linked to increased growth rate. However, subsequent studies by others in our lab and our collaborators have provided insight into this question regarding *CBX4*. In the Phillips lab, it was found that *CBX4* loss is synthetic lethal with H3K27M, as knocking out *CBX4* in glioblastoma cells had no effect on cell growth or viability, but did impact H3K27M-DMG cells (data not shown). Similarly, AJ Silva in the lab demonstrated that knockdown of *CBX4* in H3K27M-DMG cells impaired growth, but had no effect on isogenic H3K27M-KO DMG cells (data not shown).

The presence of the H3K27M mutation in DMG disrupts the normal deposition of the H3K27me3 histone modification, leading to a significant global reduction in its levels (Chan et al., 2013; Lewis et al., 2013). However, this modification is retained at specific Polycomb target sites, resulting in the increased repression of genes at these loci (Brien et al., 2021; Harutyunyan et al., 2019). This

repression maintains cells in a stem-cell-like state with increased proliferative capacity (Panditharatna et al., 2022; Silveira et al., 2019). cPRC1 complexes, function downstream of PRC2 through the recognition of H3K27me3 via its CBX subunits (Bernstein et al., 2006; Morey et al., 2012). Here, I have used CUT&RUN-Rx to investigate the impact of H3K27M on the cPRC1 complex and found that the altered H3K27me3 landscape leads to a redistribution of cPRC1 complexes on chromatin. Specifically, I have shown that CBX2, CBX4 and PCGF4 accumulate at Polycomb target sites, while non-essential paralogs CBX8 and PCGF2 are displaced. I hypothesise that cPRC1 complexes containing CBX4, and potentially PCGF4, are essential mediators of oncogenic gene repression in H3K27M-DMG. The CUT&RUN-Rx data presented in Chapter 4 and Chapter 5 suggest that the accumulation of cPRC1 complexes containing CBX4 and PCGF4 may be driving this aberrant gene repression in H3K27M-DMG cells. Interestingly however, non-essential paralog CBX2 also accumulates at these sites. I hypothesise that CBX4 dynamics are closely linked to H3K27me3 while CBX2 dynamics may also be linked to H3K27me2.

PRC2 inhibitors have shown promising therapeutic potential in H3K27M-DMG as the loss of residual H3K27me3 leads to derepression of Polycomb target genes – notably upregulation of genes involved in neural differentiation and downregulation of genes involved in cell cycle and DNA repair (Brien et al., 2021; Mohammad et al., 2017). This loss of all residual H3K27me3 also leads to complete displacement of PCGF2, PCGF4, CBX4 and CBX8 from chromatin. Interestingly, CBX2 does not behave identically. I and others have shown that H3K27M not only disrupts the H3K27me3 landscape, but also restricts H3K27me2 distribution (Harutyunyan et al., 2020). In the presence of H3K27M, this normally diffuse modification concentrates at Polycomb target sites. This distribution is reminiscent of H3K27me3 in H3-wild type cells or H3K27M-KO cells. Notably, PRC2 inhibition abolishes remaining H3K27me3, and CBX4 binding, but restores H3K27me2 to a diffuse pattern similar to that observed in H3-wild type cells. I propose that CBX2 also has an affinity for H3K27me2, allowing it to remain bound to chromatin at many sites despite the loss of H3K27me3. If CBX2 does have an affinity for H3K27me2, this may partially account for its dramatic accumulation on chromatin in H3K27M cells, as H3K27me2 also accumulates at certain Polycomb target sites in these cells. In contrast, I hypothesize that CBX4 dynamics are tightly linked to H3K27me3 in H3K27M-DMG, and that the displacement of this paralog upon PRC2 inhibitor treatment is a key contributor to the therapeutic response to PRC2 inhibitors.

While PRC2 inhibitors have therapeutic potential for H3K27M-DMG, and work to increase blood-brain-barrier penetrability of inhibitors is ongoing (Liang et al., 2022), directly targeting cPRC1 complexes may be preferable to targeting PRC2. Systemic inhibition of PRC2 may increase the risk of secondary treatment-associated malignancies as PRC2 can act as a tumour suppressor in certain cancers (Straining & Eighmy, 2022). Indeed, clinical evidence for this is emerging (Hsu et al., 2024). Notably, CBX4 and PCGF4 have not been identified as tumour suppressors in cancer genome sequencing studies, suggesting that their inhibition may be less likely to cause secondary malignancies. Directly targeting cPRC1 complexes containing CBX4, which appear to mediate

repression downstream of PRC2 in H3K27M-DMG, is a distinct and potentially preferable therapeutic strategy. This approach may offer a safer and more targeted alternative to PRC2 inhibition, as it could avoid systemic effects on tumour suppressor functions. However, more research is needed to better understand the long-term effects of targeting specific cPRC1 complexes.

The chromodomain is the reader domain for H3K27me3 and is essential for CBX4 function in H3K27M-DMG biology (Chapter 5), suggesting that inhibiting this domain could be an effective potential treatment strategy. Although some CBX chromodomain inhibitors exist in preclinical development, they lack specificity for individual CBX paralogs (Stuckey et al., 2016). Developing paralog-specific inhibitors may be challenging due to the high sequence conservation of the chromodomain across CBX proteins. However, this conservation could also be beneficial, as it may prevent other CBX paralogs from compensating for CBX4 inhibition over time, thereby reducing the risk of resistance developing. Another significant hurdle is the blood-brain barrier, which must be overcome for CBX chromodomain inhibitors to be effective in treating H3K27M-DMG. Further research is needed in preclinical models to assess the potential of CBX chromodomain inhibitors for H3K27M-DMG treatment and to investigate potential resistance mechanisms following long-term perturbation of CBX4, such as compensation by other CBX paralogs.

6.3 Accumulating evidence for non-redundant roles of cPRC1 complexes

PCGF2 and PCGF4 exhibit high reciprocal sequence homology and engage with the same sets of accessory subunits to form what are considered to be largely functionally overlapping ensembles, contrasting the functional diversity of vPRC1 complexes (Tamburri et al., 2024). However, there is some evidence that cPRC1 complexes also have non-overlapping functions. In mice, *Pcgf2/Pcgf4* double mutant embryos exhibit severe growth defects, accelerated apoptosis, and impaired maintenance of Hox gene repression (Akasaka et al., 1996; van der Lugt et al., 1996). *Pcgf2* null mutant mice exhibit growth defects, defects in lymphocyte development, defects in the smooth muscle of the intestine, and posterior transformations of the axial skeleton (Akasaka et al., 1996). *Pcgf4* null mutant mice are perinatal lethal, but also display posterior transformations of the axial skeleton, growth defects, defects in haematopoiesis, and neurological abnormalities (van der Lugt et al., 1996). Notably, PCGF4 is required for the self-renewal of neural stem cells (NSCs) in both the peripheral and central nervous systems, highlighting its critical role in neural development and function (Molofsky et al., 2005). This may explain why the genome-wide CRISPR screen in Chapter 3 identified PCGF4 as a dependency in H3K27M-DMG cells. However, it is possible that PCGF4 has other unique functions that contribute to the disease biology of H3K27M-DMG, but this would require further investigation before it could be proposed. To explore this, it would be necessary to test and validate the effects of PCGF4 loss in isogenic cell lines, both with and without the H3K27M mutation. However, this is challenging in the currently available SU-DIPG-XIII isogenic system, as PCGF4 is the dominant paralog in cPRC1 complexes in these cells, making it difficult to distinguish between the consequences of PCGF4 loss and the effects of perturbing the majority of cPRC1 complexes.

Studies in *Drosophila* have shown that heterozygous loss-of-function mutations in the *Polycomb* (*Pc*) gene lead to destabilization of the repressed state and posterior segment identity transformations (Denell & Frederick, 1983). Homozygous *Pc* mutant embryos are not viable, as all segments are transformed towards posterior segment identity (Lewis, 1978). In mammals, knockout of individual *Cbx* paralogs produces far less severe phenotypes than *Pc* null *Drosophila*, suggesting a level of functional redundancy between the paralogs. However, there are unique phenotypes which suggests that compensation is only partial. For example, *Cbx2* knock-out mice display posterior transformation of the axial skeleton, male-to-female sex reversal, and defects in the spleen and bone development (Kato-Fukui et al., 2019; Kato-Fukui et al., 2012). *Cbx4* knock-out mice die shortly after birth, but heterozygotes are viable and exhibit defects in the thymus, epithelial cells, and epidermal stem cells (Liu et al., 2013; Mardaryev et al., 2016). *Cbx7* knock-out mice are viable but develop lung and liver cancer in adulthood (Forzati et al., 2012). To my knowledge, constitutive *Cbx8* knock-out mice have not been generated. The distinct phenotypes observed in *Cbx* knock-out mice suggest that these proteins have some non-overlapping, distinct functions.

Overexpression of CBX proteins can also have significant consequences, suggesting that the relative stoichiometry of CBX paralogs in a given cell or developmental context is crucial. For instance, overexpression of CBX7, but not CBX2, CBX4, or CBX8 in haematopoietic stem or progenitor cells promotes proliferation and stem-cell identity (Klauke et al., 2013). Additionally, overexpression of CBX4, but not CBX6, CBX7, or CBX8, in human primary keratinocytes locked cells a quiescent and undifferentiated state (Luis et al., 2011).

The expansion of the Polycomb group gene family in mammals has resulted in a system with built-in redundancy, where paralogous proteins retain common functions, thereby preserving the mechanism of Polycomb-mediated gene silencing. However, loss-of-function and overexpression studies have revealed that individual paralogs possess unique molecular functions that are relevant in specific cell types or developmental stages. This functional diversity among Polycomb group complexes and subtle changes in stoichiometry enables the regulation of complex tissue-specific gene expression programs, but the molecular properties that distinguish between these paralogous proteins remain to be elucidated (Kim & Kingston, 2020). The different CBX paralog chromatin binding dynamics observed in Chapter 4 and Chapter 5, along with the gene expression data and the identification of a functionally essential region within CBX4, have provided clues as to what these distinct properties might be.

6.4 Understanding the molecularly distinct functions of different CBX paralogs

Although CBX4 has been suggested to have functions beyond its role in the cPRC1 complex, my findings suggest that these activities are unlikely to contribute significantly to H3K27M-DMG disease biology. For example, previous studies have reported that CBX4 has SUMOylation activity (Ismail et al., 2012; Kagey et al., 2003), but rescue experiments indicate that this activity is not relevant to CBX4's role in H3K27M-DMG disease biology (Chapter 5). Additionally, CBX4 has been reported to interact with the transcriptional co-repressor CtBP, which could contribute to the aberrant repression mediated by CBX4 in H3K27M-DMG (Ismail et al., 2012). However, the genome-wide CRISPR screen (Chapter 3) did not reveal any significant depletion of *CtBP1* or *CtBP2*, suggesting that these proteins are not crucial for the function of CBX4 in H3K27M-DMG disease biology. This suggests that the essential role of CBX4 in H3K27M-DMG biology is unlikely to be independent from cPRC1.

The binding specificities and affinities of mammalian CBX homologs remain poorly characterized. Early *in vitro* studies conducted nearly two decades ago examined the binding properties of individual CBX chromodomains to histone H3 tail peptides (Bernstein et al., 2006). These initial studies revealed limited insights into the binding specificities and affinities of individual CBX chromodomains. The CBX8 chromodomain exhibited a weak affinity for H3K27me₃, but this finding has not been independently validated. In contrast, CBX2, CBX4 and CBX7 chromodomains demonstrated no selectivity for H3K27me₃, binding to both H3K27me₃ and H3K9me₃ (Bernstein et al., 2006). The CBX6 displayed no affinity for histone H3 tails (Bernstein et al., 2006). However, despite these *in vitro* findings, mammalian CBX proteins are known to associate broadly with Polycomb repressed domains enriched in H3K27me₃ *in vivo* (Morey et al., 2012; Pemberton et al., 2014). This disparity suggests that the minimalist *in vitro* approach previously used may not accurately reflect the complexities of chromodomain binding in cells. Specifically, the results presented in this thesis imply that CBX paralogs may respond differently to H3K27me₃ patterning. Furthermore, it is possible that H3K27me₂ patterning and sensing could play a previously underappreciated role in regulating CBX binding dynamics. I propose that CBX2 may have a higher affinity for H3K27me₂ compared to CBX4 and CBX8, which could contribute to differences in their chromatin binding profiles. However, this hypothesis requires further testing *in vitro* to confirm. The specificity and affinity of CBX chromodomains for H3K27me₃ or H3K27me₂ may require the presence of whole nucleosome substrates and/or additional features beyond the chromodomain itself, such as whole proteins or protein complexes. Supporting this notion, it has been shown that CBX7 recruitment to chromatin is dependent on the presence of JARID2 (Glancy et al., 2023). To elucidate the differences in chromodomain binding specificities and affinities, it is essential to revisit these experiments using more physiologically relevant systems, including whole nucleosomes and purified CBX proteins, or recombinant PRC1 complexes.

The region between the chromodomain and the C-box in the CBXs is characterized by central disordered domain that varies in length and lacks homology with other CBX paralogs (Kim & Kingston, 2020). The central domains of CBX2, CBX4, and CBX8 exhibit predicted disordered structures, low

sequence complexity, and are generally positively charged (Kim & Kingston, 2020). The charged and disordered nature of the central domain is essential for two documented functions of cPRC1: chromatin compaction and phase separation.

The central disordered domains of CBX2, CBX4, and CBX8 share similar characteristics, with the CBX2 domain being the most extensively studied to date. Electron microscopy has shown that the central domain of these CBXs can compact nucleosomal arrays *in vitro* (Grau et al., 2011). Point mutations at the positively charged residues within this disordered region abolish the nucleosome compaction capabilities of the CBX, highlighting the critical role of this domain in chromatin compaction (Grau et al., 2011). Furthermore, the central disordered domains of CBX2, CBX4, and CBX8 have been found to mediate liquid-liquid phase separation (LLPS) *in vitro*, and are important for the formation of nuclear aggregates in cells (Plys et al., 2019; Tatavosian et al., 2019). This phenomenon was first inferred through fluorescent microscopy studies in the late 1990s, which identified dynamic nuclear aggregates containing PRC1 and PRC2 proteins, as well as megabases of highly compacted DNA (Satijn et al., 1997; Saurin et al., 1998). These nuclear aggregates are termed "Polycomb bodies". Polycomb bodies exhibit dynamic properties and are thought to create specialized compartments for target gene repression (Murphy & Boettiger, 2024; Tamburri et al., 2024). However, the role of Polycomb bodies in gene repression, if they are formed through LLPS in cells, and if LLPS is important for gene repression remain topics of ongoing debate (Murphy & Boettiger, 2024; Tamburri et al., 2024).

LLPS has been proposed as a mechanism for maintaining epigenetic memory (Kim & Kingston, 2022). A plausible hypothesis is that LLPS could help to increase the local concentration of repressive Polycomb group complexes in Polycomb domains, thereby enhancing the likelihood of the Polycomb domain re-forming on the template following DNA replication (Kim & Kingston, 2022). Supporting this idea, studies have shown that *Drosophila* cPRC1 remains bound to replicated chromatin *in vitro*, and that the mammalian cPRC1 component CBX2 associates with mitotic chromosomes in cells (Francis et al., 2009; Zhen et al., 2014). However, it is unknown if LLPS actually assists in the reformation of Polycomb bodies during cell division and maintenance of gene repression. Theoretically, sequestering repressed Polycomb domains within LLPS structures that concentrate Polycomb complexes and exclude activating proteins could enhance the memory of repression (Kim & Kingston, 2022). However, experimentally testing this hypothesis is challenging due to the inherent coupling between the compaction and LLPS activities of cPRC1 complexes. Specifically, mutations that disrupt the nucleosome compaction activity of CBX2 also abolish its ability to undergo phase separation *in vitro*, making it difficult to dissociate the mechanisms of direct gene repression from those involved in maintaining the repressed state (Grau et al., 2011; Kim & Kingston, 2022). As a result, teasing apart the distinct contributions of these mechanisms remains a significant experimental challenge.

Perturbations that disrupt direct repression mechanisms are likely to result in widespread and immediate derepression of target genes. In contrast, mutations that impair epigenetic memory maintenance may lead to stochastic or delayed derepression, as the loss of memory function would result in a gradual erosion of repression over time (Kim & Kingston, 2022). It has been suggested that a subtle reduction in the efficacy of epigenetic memory maintenance, even as small as 10%, can have a profound impact on gene regulation over multiple cell divisions. For example, a 10% decrease in memory function after 6-7 cell divisions could result in a 50% probability of derepression in a significant proportion of cells, potentially disrupting normal developmental processes (Kim & Kingston, 2022). In contrast, perturbations in components required for direct repression are likely to result in immediate and widespread derepression in cells. This might suggest that loss of CBX4 disrupts a memory function as it takes multiple days to observe significant gene expression changes with bulk transcriptomics following PRC2 inhibition (Chapter 5) or specific knockdown of CBX4 (data not shown). However, this remains to be experimentally tested. To further investigate this hypothesis, single-cell transcriptomics could be employed to examine the effects of PRC2 inhibition or CBX4 loss on gene expression at the single-cell level. This approach would allow for a more detailed understanding of the dynamics of gene expression changes and potentially shed light on the role of CBX4 in epigenetic memory.

Additionally, beyond the debated relevance of LLPS in the function and maintenance of Polycomb bodies, very little is known about the behaviour of Polycomb bodies themselves in their native context, particularly in developing organisms (Kim & Kingston, 2022). Key questions remain unanswered, including: which target genes are localized within these bodies, and do the segregated structures contribute to a self-reinforcing mechanism that consolidates epigenetic memory? To address these questions, it would be beneficial to employ acute depletion approaches, such as inducible genetic deletion or degron-based methods, to rapidly remove target proteins and assess the immediate responses. This would provide valuable insights into the dynamics of Polycomb bodies and their role in regulating gene expression *in vivo*.

There is a growing recognition of the importance of unstructured proteins in cellular function. Proteins containing intrinsically disordered regions are highly abundant in higher eukaryotes and play crucial roles in various biological processes (Tsang et al., 2020; Wright & Dyson, 2015). *Cbx2* null mutant mice exhibit axial patterning defects, characterised by anterior-to-posterior transformations in several vertebrae (Coré et al., 1997). Notably, the same point mutations in the central disordered domain of *Cbx2* that disrupt its nucleosome compaction and LLPS capabilities phenocopy the null mutant, resulting in similar axial patterning defects (Lau et al., 2017). This demonstrates that the central disordered domain is essential for the proper function of CBX2 in repressing gene expression and maintaining cell type during development. The CRISPR-Cas9 tiling screen in Chapter 5 revealed that the central domain of CBX4 is crucial for its function in H3K27M-DMG biology. However, the exact role of this disordered region remains unclear. It is uncertain whether it is directly involved in repressive functions or if it plays a structural role, such as stabilizing the cPRC1 complex or

influencing its conformation and composition. Preliminary data suggest that the central region of CBX4 may affect the composition of the cPRC1 complex. Specifically, immunoprecipitation of wild-type CBX4 co-precipitates a distinct complex consisting of RING1A/B, PCGF4, and PHC2 (data not shown). In contrast, immunoprecipitation of CBX4 with deletions in this essential disordered domain results in the co-precipitation of a more diverse array of cPRC1 complex components, including all three PHCs and PCGF2. However, this remains to be validated.

Additionally, it remains to be determined if this central disordered region of CBX4 plays a role in functions such as nucleosome compaction and Polycomb body formation/LLPS. It is also uncertain whether these functions contribute to the aberrant gene repression mediated by CBX4 in H3K27M-DMG. To address these questions, domain swap experiments could be performed, where the equivalent region from CBX2 or CBX8 is swapped with the essential region in CBX4. Rescue experiments using these domain swap constructs could help determine if the essential properties of the internal disordered domain are unique to CBX4. The sequence homology of the internal disordered region is low, increasing the likelihood of neofunctionalization. However, low sequence homology is also a common characteristic of internally disordered regions as they are not subject to structural constraints. As such, they may ultimately possess largely the same properties. Based on AlphaFold prediction models, I hypothesize that the CBX4 internal disordered domain has a unique function that cannot be compensated for by other CBX paralogs. However, this hypothesis requires experimental validation to be proven. Domain swap experiments can also provide valuable insights into the roles of Polycomb domains and LLPS in gene repression, particularly in H3K27M-DMG cells. One potential approach would be to create a CBX4 construct with the internal disordered domain of CBX2 and test its ability to compact nucleosomes and undergo LLPS. Additionally, it would be interesting to see if this modified CBX4 construct can still aberrantly repress Polycomb target genes. If the construct is able to perform certain functions, such as forming Polycomb bodies, but fails to maintain gene repression, this would suggest that these functions can be uncoupled. Furthermore, this could indicate that different CBX proteins contribute differently to various cPRC1 functions. However, these hypotheses remain to be experimentally tested.

6.5 Future directions

Fundamental questions remain regarding the CBX paralogs CBX2, CBX4 and CBX8 which are expressed in distinct combinations in differentiated cell lineages. Future research directions should focus on re-examining the binding specificities and affinities of CBX chromodomains, using whole or near-whole cPRC1 complexes and nucleosome substrates to gain a more comprehensive understanding. Additionally, it is essential to investigate the long-term consequences of depleting or inhibiting CBX4 to determine its potential as a therapeutic target for H3K27M-DMG. This will also help identify potential mechanisms of resistance that may arise, which is crucial for developing effective treatments.

It is becoming increasingly clear that the precise regulation of Polycomb protein levels is essential for maintaining proper spatio-temporal gene expression patterns. This notion is supported by the observation that both mutations in Polycomb genes and overexpression of these genes result in dosage-sensitive phenotypes. Notably, attempts to overexpress wild-type CBX2 in mESCs have been challenging, whereas a point mutant of CBX2 deficient in compaction and phase separation can be overexpressed (Bel et al., 1998; Kim & Kingston, 2022). Furthermore, the consistent ratios of different CBX paralogs across various H3K27M-DMG cell lines, as well as the difficulties encountered in overexpressing CBX4 in these cells for rescue experiments, suggest that the stoichiometry of CBX proteins is tightly regulated to prevent aberrant activity. The H3K27M oncohistone only promotes oncogenesis within a specific cellular context and developmental window (Deshmukh et al., 2022; Jessa et al., 2022). The reasons for this are currently unknown. Speculating, the composition of cPRC1 complexes may play a role in this process. It is possible that a specific stoichiometry of cPRC1 subunits, a type of "goldilocks zone", could exist in oligodendrocyte precursor cells (OPCs) that allows for the correct balance of Polycomb-mediated gene repression that restrains differentiation but supports stem cell maintenance. However, further investigation is warranted to determine the true molecular mechanisms underlying this phenomenon.

It is unknown whether CBX paralogs possess distinct abilities to compact chromatin or form Polycomb bodies or undergo LLPS, and if this is critical for their function in specific cell lineages or diseases. To address this, domain-swap experiments will provide valuable insights into the unique functions of the internal disordered regions among CBX proteins. These experiments may also help clarify the relationship between Polycomb bodies and gene repression. Furthermore, the role of this region in CBX4 in determining complex composition requires further validation and investigation. If confirmed, this finding could lead to new therapeutic approaches for H3K27M-DMG. Specifically, targeting protein-protein interactions with inhibitors or using targeted protein degraders may offer a more precise way to disrupt oncogenic multiprotein complexes on chromatin (Gao & Vakoc, 2024), potentially leading to alternative effective clinical treatments for H3K27M-DMG.

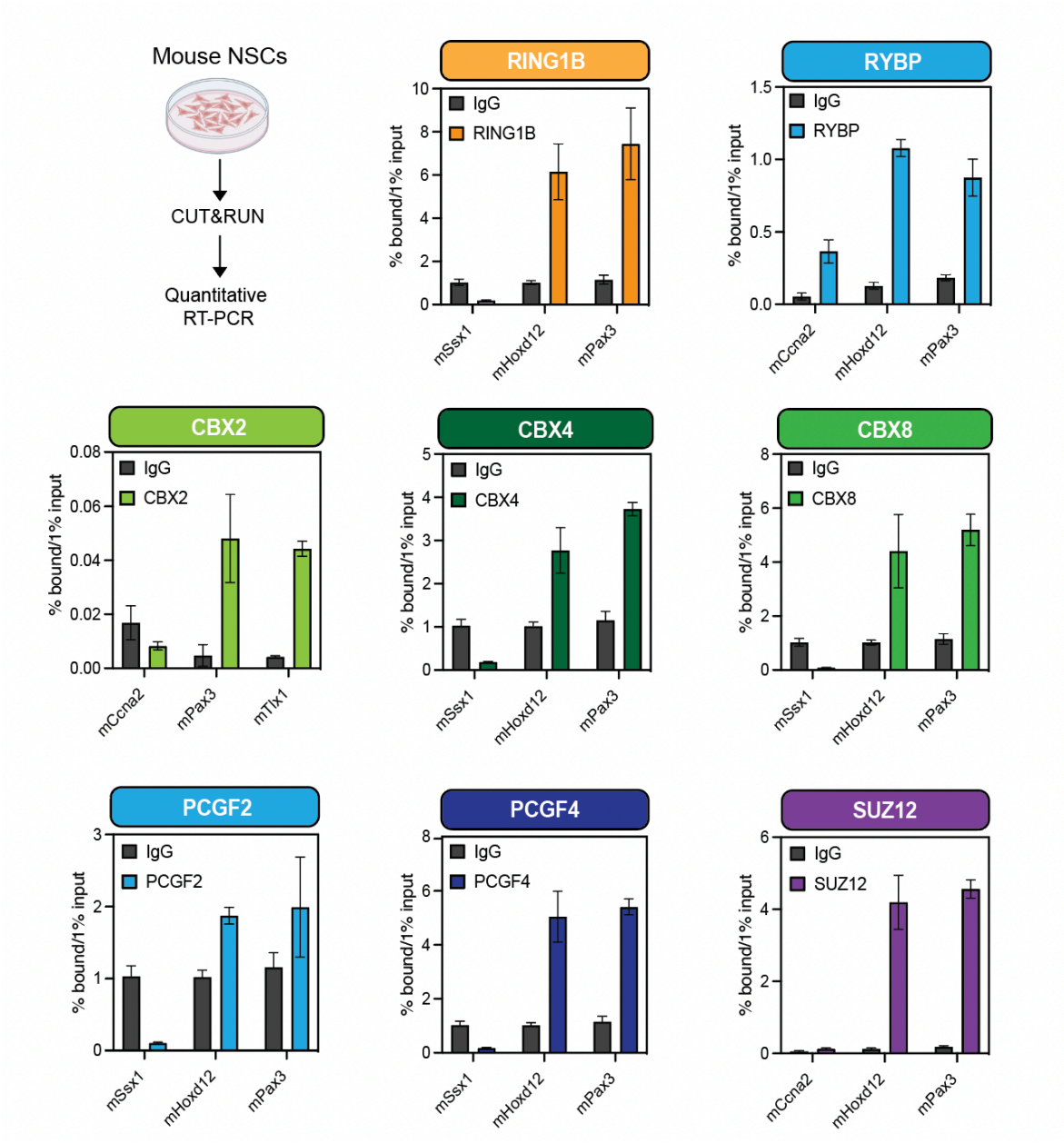
6.6 Conclusions

The data presented in this thesis supports the following conclusions:

- *CBX4* and *PCGF4* are genetic dependencies in H3K27M-DMG.
- The presence of H3K27M leads to changes in H3K27me2 and H3K27me3 patterning, along with redistribution of cPRC1 complexes on chromatin.
- The chromodomain of CBX4 is functionally essential in H3K27M-DMG disease biology.
- PRC2 inhibition in H3K27M-DMG cells leads to loss of H3K27me3, and a restoration of H3K27me2 patterning similar to that in non-H3K27M cells. The CBX2 chromodomain may have some affinity for H3K27me2.
- The purported E3 SUMO-ligase activity of CBX4 is not essential in H3K27M-DMG biology.
- The internal disordered region of CBX4 is likely functionally important in H3K27M-DMG, although this remains to be validated.

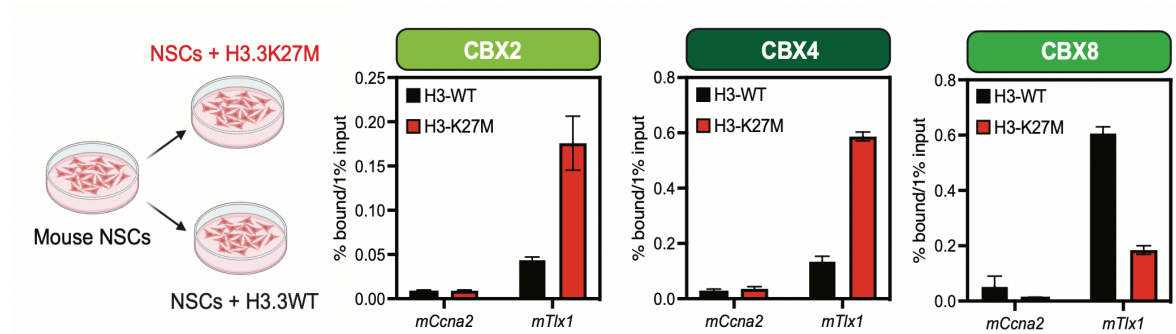
APPENDIX

APPENDIX I



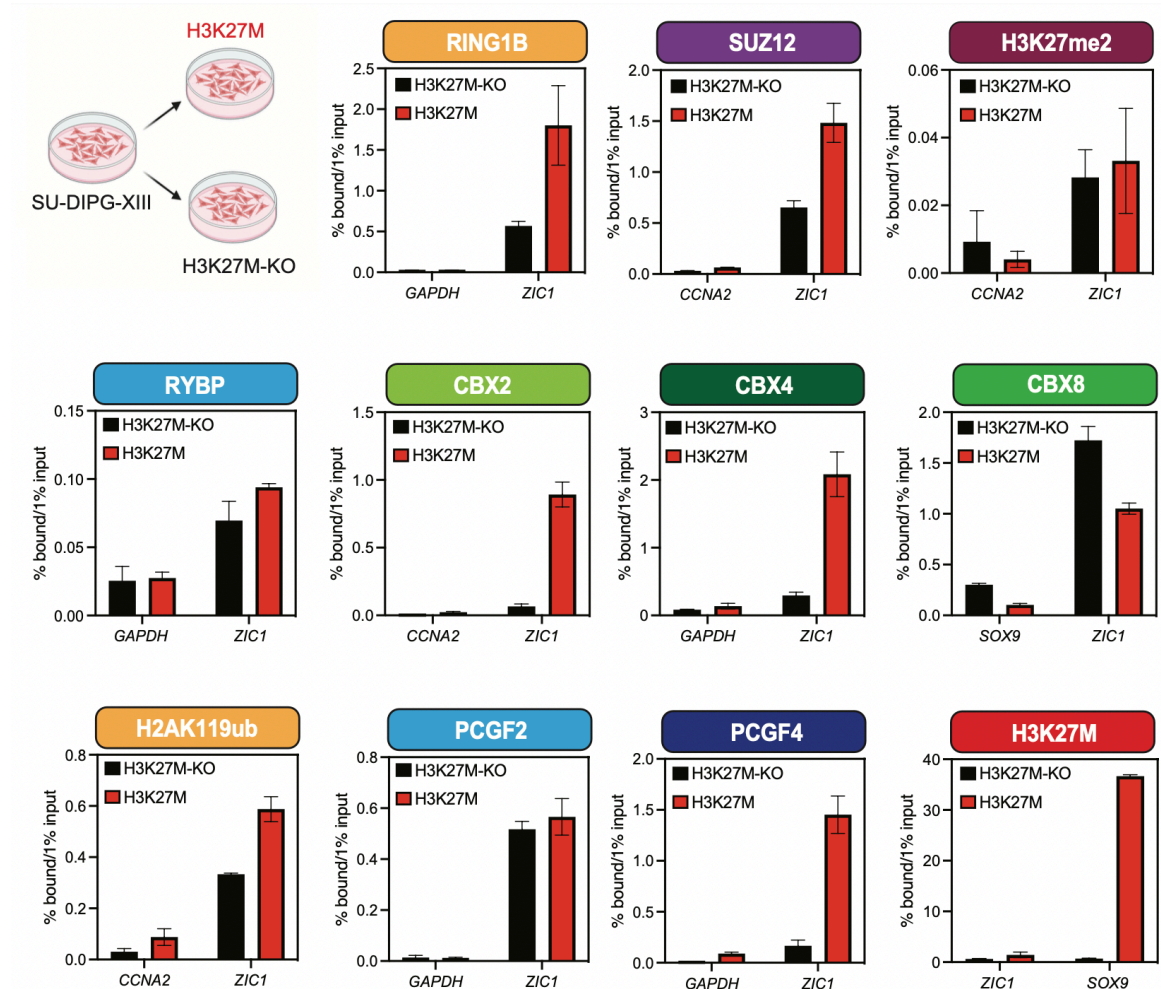
mNSC CUT&RUN-qPCRs. Negative control sites: *mSsx1* and *mCcn2*. Positive control sites: *mHoxd12* and *mPax3*. n = 3 technical replicates shown.

APPENDIX II



CUT&RUN-qPCRs in mNSCs expressing exogenous H3 wild-type or H3K27M. Negative control site: *mCcn2*. Positive control site: *mTlx1*. n = 3 technical replicates shown.

APPENDIX III



CUT&RUN-qPCRs in SU-DIPG-XIII H3K27M and H3K27M-KO cells. Negative control sites for each antibody are plotted on the left, positive control sites for each antibody are plotted on the right. n = 3 technical replicates shown.

APPENDIX IV

Antibodies used in this study.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Normal Rabbit IgG (CUT&RUN)	Cell Signaling	Cat#2729S; RRID: AB_1031062
Normal Rabbit IgG (CUT&RUN)	Millipore	Cat#12-370; RRID: AB_145841
SUZ12 (CUT&RUN)	Cell Signaling	Cat#3737S; RRID: AB_2196850
RING1B (CUT&RUN, WB)	Cell Signaling	Cat#5694S; RRID: AB_10705604
PCGF2 (CUT&RUN, WB)	Diego Pasini Lab (Scelfo et al.)	N/A
PCGF4 (CUT&RUN, WB)	Cell Signaling	Cat#5856S; RRID: AB_10838137
CBX2 (CUT&RUN, WB)	Cell Signaling	Cat#18687S; RRID: AB_2923356
CBX4 (CUT&RUN, WB)	Cell Signaling	Cat#30559S; RRID: AB_2798991
CBX8 (CUT&RUN, WB)	Cell Signaling	Cat#14696S; RRID: AB_2687589
H3K27me3 (CUT&RUN)	Cell Signaling	Cat#9733T; RRID: AB_2616029
H3K27me3 (WB)	Cell Signaling	Cat# 9733BF; RRID:
H3K27me2 (CUT&RUN)	Cell Signaling	Cat#9728; RRID: AB_1281338
H2AK119ub (CUT&RUN, WB)	Cell Signaling	Cat#8240S; RRID: AB_10891618
DEDAF/RYPB (CUT&RUN)	Millipore	Cat#AB3637; RRID: AB_2285466
Histone H3 (WB)	Active Motif	Cat# 39763; RRID: AB_2650522
Goat anti-mouse (red) IRDye 680RD (WB)	Licor	Cat#926-68070; RRID: AB_10956588
Goat anti-rabbit (green) IRDye 800CW (WB)	Licor	Cat# 926-32211; RRID: AB_621843

APPENDIX V

sgRNA sequences used in this study.

Oligonucleotide	Targeting Sequence	Source	IDENTIFIER
ROSA26	GAAGATGGGCGGGAGTCTTC		
CBX2 sgRNA 1	AGTGACTTAGATGCTAAGAG	Doench et al.	Addgene Catalog #73178
CBX2 sgRNA 2	GATCCCATCCGGAAGAAGCG	Doench et al.	Addgene Catalog #73178
CBX2 sgRNA 3	GCTTTGCAGACATAACAGCT	Doench et al.	Addgene Catalog #73178
CBX4 sgRNA 1	CCTTGCAAGATATAACACGT	Doench et al.	Addgene Catalog #73178
CBX4 sgRNA 2	CGGTCAGGACATTGGAACGA	Doench et al.	Addgene Catalog #73178
CBX4 sgRNA 3	GAGCCACCCGCCGACAAGT	Doench et al.	Addgene Catalog #73178
CBX6 sgRNA 1	AACCTGAAGGTGATCGACAA	Doench et al.	Addgene Catalog #73178
CBX6 sgRNA 2	ACTCGGCAGGTACAGCACTT	Doench et al.	Addgene Catalog #73178
CBX6 sgRNA 3	GGGCGGCCAGAAAGTAGCCCA	Doench et al.	Addgene Catalog #73178
CBX7 sgRNA 1	AGGAGAGAGACCGAGCATCG	Doench et al.	Addgene Catalog #73178
CBX7 sgRNA 2	AGGTACTTGTGGGCCTTTTCG	Doench et al.	Addgene Catalog #73178
CBX7 sgRNA 3	GCATCCGGAAGAAGCGCGTG	Doench et al.	Addgene Catalog #73178
CBX8 sgRNA 1	CTCTATGGCCCCAAAAGCG	Doench et al.	Addgene Catalog #73178
CBX8 sgRNA 2	GGGACAAACCCATGTTTCGA	Doench et al.	Addgene Catalog #73178
CBX8 sgRNA 3	GGGGACAGCTCGAAGAAGCG	Doench et al.	Addgene Catalog #73178
PCGF2 sgRNA 1	AACGGCTCCAATGAGGACCG	Doench et al.	Addgene Catalog #73178
PCGF2 sgRNA 2	CATCGACGCCACCACTATCG	Doench et al.	Addgene Catalog #73178
PCGF2 sgRNA 3	CTCTCCATCGAATTCTACGA	Doench et al.	Addgene Catalog #73178
PCGF4 sgRNA 1	AATGGCTCTAATGAAGATAG	Doench et al.	Addgene Catalog #73178
PCGF4 sgRNA 2	CTCCACAAAGCACACACATC	Doench et al.	Addgene Catalog #73178
PCGF4 sgRNA 3	GATTGATGTCATGTATGAGG	Doench et al.	Addgene Catalog #73178

APPENDIX VI

Mouse qPCR primer sequences

Target	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
<i>mSsx1</i> promoter		
<i>mCcn2</i> promoter	CTGCCGCCGACAGAACCG	CTGAGCCCGCATCGCGAC
<i>mHoxd12</i> promoter	CCTGAGAGCCGCCCAATTGG	GTGTTGTGGGAGCCCGCG
<i>mPax3</i> promoter	AGAAAGGCGGAAAGAGATTAGGAC	GAGCCAGTGAGGATGGAAAAAG
<i>mTlx1</i> promoter	CATACACCTCGGCCTTCCTC	GCACGGAGCTCAGGGAATAA

APPENDIX VII

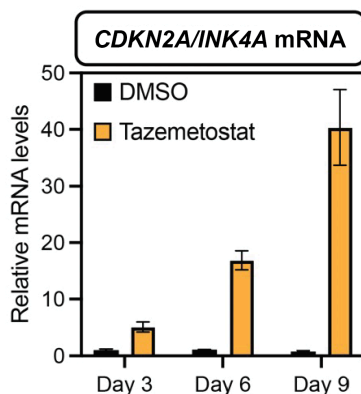
Human qPCR primer sequences

Target	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
<i>ZIC1</i> promoter	CGCTCACATTCTCTATGCTAC	GAAAGGCATGGAGCATCTCAG
<i>CCNA2</i> promoter	TGACGTCATTCAAGGCGA	GCTCAGTTTCCTTTGGTTTAC
<i>GAPDH</i> TSS	CCACATCGCTCAGACACCAT	GCGAACTCACCCGTTGACT
<i>SOX9</i> enhancer	AACCCTCTTTTCACCTCTCCAC	TCAGATGCTCAGGCCAGTAAAG
<i>H3F3A</i>	ATGGCTCGTACAAAGCAGAC	AAGGTCTTGGTCCACATTGG

APPENDIX VIII

Gene expression RT-qPCR primers

Target	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
<i>CDKN2A/INK4A</i>	AGAGGATTTGAGGGACAGGGTC	CCTCTTTCTTCCTCCGGTGC



One representative biological replicate. n = 3 technical replicates shown.

REFERENCES

- Akasaka, T., Kanno, M., Balling, R., Mieza, M. A., Taniguchi, M., & Koseki, H. (1996). A role for mel-18, a Polycomb group-related vertebrate gene, during the anteroposterior specification of the axial skeleton. *Development*, 122(5), 1513-1522. <https://doi.org/10.1242/dev.122.5.1513>
- Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, 17(8), 487-500. <https://doi.org/10.1038/nrg.2016.59>
- Antequera, F., & Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A*, 90(24), 11995-11999. <https://doi.org/10.1073/pnas.90.24.11995>
- Badeaux, A. I., & Shi, Y. (2013). Emerging roles for chromatin as a signal integration and storage platform. *Nature Reviews Molecular Cell Biology*, 14(4), 211-224. <https://doi.org/10.1038/nrm3545>
- Bajusz, I., Kovács, G., & Pirity, M. K. (2018). From Flies to Mice: The Emerging Role of Non-Canonical PRC1 Members in Mammalian Development. *Epigenomes*, 2(1), 4. <https://www.mdpi.com/2075-4655/2/1/4>
- Baker, T., Nerle, S., Pritchard, J., Zhao, B., Rivera, V. M., Garner, A., & Gonzalez, F. (2015). Acquisition of a single EZH2 D1 domain mutation confers acquired resistance to EZH2-targeted inhibitors. *Oncotarget*, 6(32), 32646-32655. <https://doi.org/10.18632/oncotarget.5066>
- Balakrishnan, I., Danis, E., Pierce, A., Madhavan, K., Wang, D., Dahl, N., Sanford, B., Birks, D. K., Davidson, N., Metselaar, D. S., Meel, M. H., Lemma, R., Donson, A., Vijmasi, T., Katagi, H., Sola, I., Fosmire, S., Alimova, I., Steiner, J., . . . Venkataraman, S. (2020). Senescence Induced by BMI1 Inhibition Is a Therapeutic Vulnerability in H3K27M-Mutant DIPG. *Cell Reports*, 33(3). <https://doi.org/10.1016/j.celrep.2020.108286>
- Barrangou, R., & Doudna, J. A. (2016). Applications of CRISPR technologies in research and beyond. *Nature Biotechnology*, 34(9), 933-941. <https://doi.org/10.1038/nbt.3659>
- Bass, A. J., Watanabe, H., Mermel, C. H., Yu, S., Perner, S., Verhaak, R. G., Kim, S. Y., Wardwell, L., Tamayo, P., Gat-Viks, I., Ramos, A. H., Woo, M. S., Weir, B. A., Getz, G., Beroukhi, R., O'Kelly, M., Dutt, A., Rozenblatt-Rosen, O., Dziunycz, P., . . . Meyerson, M. (2009). SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nature Genetics*, 41(11), 1238-1242. <https://doi.org/10.1038/ng.465>
- Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A. R., Papai, G., Reymer, A., Syed, S. H., Lone, I. N., Tonchev, O., Crucifix, C., Menoni, H., Papin, C., Skoufias, D. A., Kurumizaka, H., Lavery, R., Hamiche, A., Hayes, J. J., Schultz, P., Angelov, D., . . . Dimitrov, S. (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Molecular Cell*, 66(3), 384-397.e388. <https://doi.org/10.1016/j.molcel.2017.04.012>
- Béguelin, W., Popovic, R., Teater, M., Jiang, Y., Bunting, K. L., Rosen, M., Shen, H., Yang, S. N., Wang, L., Ezponda, T., Martinez-Garcia, E., Zhang, H., Zheng, Y., Verma, S. K., McCabe, M. T., Ott, H. M., Van Aller, G. S., Kruger, R. G., Liu, Y., . . . Melnick, A. M. (2013). EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid

- transformation. *Cancer Cell*, 23(5), 677-692.
<https://doi.org/10.1016/j.ccr.2013.04.011>
- Behan, F. M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C. M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., Ansari, R., Harper, S., Jackson, D. A., McRae, R., Pooley, R., Wilkinson, P., van der Meer, D., Dow, D., Buser-Doepner, C., . . . Garnett, M. J. (2019). Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*, 568(7753), 511-516. <https://doi.org/10.1038/s41586-019-1103-9>
- Behjati, S., Tarpey, P. S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D. C., Cooke, S. L., Gundem, G., Davies, H., Nik-Zainal, S., Martin, S., McLaren, S., Goody, V., Robinson, B., Butler, A., Teague, J. W., Halai, D., Khatri, B., . . . Flanagan, A. M. (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nature Genetics*, 45(12), 1479-1482. <https://doi.org/10.1038/ng.2814>
- Bel, S., Coré, N., Djabali, M., Kieboom, K., Van der Lugt, N., Alkema, M. J., & Van Lohuizen, M. (1998). Genetic interactions and dosage effects of Polycomb group genes in mice. *Development*, 125(18), 3543-3551.
<https://doi.org/10.1242/dev.125.18.3543>
- Bender, S., Tang, Y., Lindroth, A. M., Hovestadt, V., Jones, D. T., Kool, M., Zapatka, M., Northcott, P. A., Sturm, D., Wang, W., Radlwimmer, B., Højfeldt, J. W., Truffaux, N., Castel, D., Schubert, S., Ryzhova, M., Seker-Cin, H., Gronych, J., Johann, P. D., . . . Pfister, S. M. (2013). Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. *Cancer Cell*, 24(5), 660-672.
<https://doi.org/10.1016/j.ccr.2013.10.006>
- Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., & Allis, C. D. (2006). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol*, 26(7), 2560-2569. <https://doi.org/10.1128/mcb.26.7.2560-2569.2006>
- Black, Joshua C., Van Rechem, C., & Whetstone, Johnathan R. (2012). Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Molecular Cell*, 48(4), 491-507.
<https://doi.org/10.1016/j.molcel.2012.11.006>
- Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L. P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., Ishikura, T., Long, H. K., Sheahan, T. W., Brockdorff, N., Kessler, B. M., Koseki, H., & Klose, R. J. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, 157(6), 1445-1459. <https://doi.org/10.1016/j.cell.2014.05.004>
- Blackledge, N. P., Fursova, N. A., Kelley, J. R., Huseyin, M. K., Feldmann, A., & Klose, R. J. (2020). PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol Cell*, 77(4), 857-874.e859.
<https://doi.org/10.1016/j.molcel.2019.12.001>
- Blackledge, N. P., & Klose, R. J. (2021). The molecular principles of gene regulation by Polycomb repressive complexes. *Nature Reviews Molecular Cell Biology*, 22(12), 815-833. <https://doi.org/10.1038/s41580-021-00398-y>
- Bodapati, S., Daley, T. P., Lin, X., Zou, J., & Qi, L. S. (2020). A benchmark of algorithms for the analysis of pooled CRISPR screens. *Genome Biology*, 21(1), 62. <https://doi.org/10.1186/s13059-020-01972-x>

- Boschert, T., Kromer, K., Lerner, T., Lindner, K., Haltenhof, G., Tan, C. L., Jähne, K., Poschke, I., Bunse, L., Eisele, P., Grassl, N., Mildenerberger, I., Sahm, K., Platten, M., Lindner, J. M., & Green, E. W. (2024). H3K27M neoepitope vaccination in diffuse midline glioma induces B and T cell responses across diverse HLA loci of a recovered patient. *Science Advances*, 10(5), eadi9091. <https://doi.org/doi:10.1126/sciadv.adi9091>
- Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., & Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, 441(7091), 349-353. <https://doi.org/10.1038/nature04733>
- Bracken, A. P., Brien, G. L., & Verrijzer, C. P. (2019). Dangerous liaisons: interplay between SWI/SNF, NuRD, and Polycomb in chromatin regulation and cancer. *Genes Dev*, 33(15-16), 936-959. <https://doi.org/10.1101/gad.326066.119>
- Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H., & Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & Dev*, 20, 1123-1136. <https://doi.org/10.1101/gad.381706>
- Bracken, A. P., 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(20), 5323-5335. <https://doi.org/10.1093/emboj/cdg542>
- Brien, G. L., Bressan, R. B., Monger, C., Gannon, D., Lagan, E., Doherty, A. M., Healy, E., Neikes, H., Fitzpatrick, D. J., Deevy, O., Grant, V., Marqués-Torrejón, M. A., Alfazema, N., Pollard, S. M., & Bracken, A. P. (2021). Simultaneous disruption of PRC2 and enhancer function underlies histone H3.3-K27M oncogenic activity in human hindbrain neural stem cells. *Nat Genet*, 53(8), 1221-1232. <https://doi.org/10.1038/s41588-021-00897-w>
- Brien, G. L., Remillard, D., Shi, J., Hemming, M. L., Chabon, J., Wynne, K., Dillon, E. T., Cagney, G., Van Mierlo, G., Baltissen, M. P., Vermeulen, M., Qi, J., Fröhling, S., Gray, N. S., Bradner, J. E., Vakoc, C. R., & Armstrong, S. A. (2018). Targeted degradation of BRD9 reverses oncogenic gene expression in synovial sarcoma. *eLife*, 7. <https://doi.org/10.7554/eLife.41305>
- Brien, G. L., Valerio, D. G., & Armstrong, S. A. (2016). Exploiting the Epigenome to Control Cancer-Promoting Gene-Expression Programs. *Cancer Cell*, 29(4), 464-476. <https://doi.org/10.1016/j.ccell.2016.03.007>
- Brown, C. J., Johnson, A. K., Dunker, A. K., & Daughdrill, G. W. (2011). Evolution and disorder. *Current Opinion in Structural Biology*, 21(3), 441-446. <https://doi.org/https://doi.org/10.1016/j.sbi.2011.02.005>
- Brown, K., Chew, P. Y., Ingersoll, S., Espinosa, J. R., Aguirre, A., Espinoza, A., Wen, J., Astatike, K., Kutateladze, T. G., Colleparado-Guevara, R., & Ren, X. (2023). Principles of assembly and regulation of condensates of Polycomb repressive complex 1 through phase separation. *Cell Reports*, 42(10). <https://doi.org/10.1016/j.celrep.2023.113136>
- Brownell, J. E., & Allis, C. D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Current Opinion in Genetics & Development*, 6(2), 176-184. [https://doi.org/https://doi.org/10.1016/S0959-437X\(96\)80048-7](https://doi.org/https://doi.org/10.1016/S0959-437X(96)80048-7)

- Brunk, B. P., Martin, E. C., & Adler, P. N. (1991). *Drosophila* genes Posterior Sex Combs and Suppressor two of zeste encode proteins with homology to the murine bmi-1 oncogene. *Nature*, 353(6342), 351-353.
<https://doi.org/10.1038/353351a0>
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., & Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298(5595), 1039-1043.
<https://doi.org/10.1126/science.1076997>
- Chakravarty, D., & Solit, D. B. (2021). Clinical cancer genomic profiling. *Nat Rev Genet*, 22(8), 483-501. <https://doi.org/10.1038/s41576-021-00338-8>
- Chan, K. M., Fang, D., Gan, H., Hashizume, R., Yu, C., Schroeder, M., Gupta, N., Mueller, S., James, C. D., Jenkins, R., Sarkaria, J., & Zhang, Z. (2013). The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes Dev*, 27(9), 985-990.
<https://doi.org/10.1101/gad.217778.113>
- Chen, C. C. L., Deshmukh, S., Jessa, S., Hadjadj, D., Lisi, V., Andrade, A. F., Faury, D., Jawhar, W., Dali, R., Suzuki, H., Pathania, M., A, D., Dubois, F., Woodward, E., Hébert, S., Coutelier, M., Karamchandani, J., Albrecht, S., Brandner, S., . . . Jabado, N. (2020). Histone H3.3G34-Mutant Interneuron Progenitors Co-opt PDGFRA for Gliomagenesis. *Cell*, 183(6), 1617-1633.e1622. <https://doi.org/10.1016/j.cell.2020.11.012>
- Ciapponi, M., Karlukova, E., Schkölziger, S., Benda, C., & Müller, J. (2024). Structural basis of the histone ubiquitination read–write mechanism of RYBP–PRC1. *Nature Structural & Molecular Biology*, 31(7), 1023-1027.
<https://doi.org/10.1038/s41594-024-01258-x>
- Ciferri, C., Lander, G. C., Maiolica, A., Herzog, F., Aebersold, R., & Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. *eLife*, 1, e00005. <https://doi.org/10.7554/eLife.00005>
- Clymer, J., & Kieran, M. W. (2018). The Integration of Biology Into the Treatment of Diffuse Intrinsic Pontine Glioma: A Review of the North American Clinical Trial Perspective. *Front Oncol*, 8, 169.
<https://doi.org/10.3389/fonc.2018.00169>
- Conti, L., Pollard, S. M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q. L., Cattaneo, E., & Smith, A. (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol*, 3(9), e283. <https://doi.org/10.1371/journal.pbio.0030283>
- Conway, E., Healy, E., & Bracken, A. P. (2015). PRC2 mediated H3K27 methylations in cellular identity and cancer. *Curr Opin Cell Biol*, 37, 42-48.
<https://doi.org/10.1016/j.ceb.2015.10.003>
- Conway, E., Jerman, E., Healy, E., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D. J., Mucha, M., Watson, A., Rice, A. M., Chammas, P., Huang, C., Pratt-Kelly, I., Koseki, Y., Nakayama, M., Ishikura, T., Streubel, G., . . . Bracken, A. P. (2018). A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. *Mol Cell*, 70(3), 408-421.e408.
<https://doi.org/10.1016/j.molcel.2018.03.005>
- Coré, N., Bel, S., Gaunt, S. J., Aurrand-Lions, M., Pearce, J., Fisher, A., & Djabali, M. (1997). Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. *Development*, 124(3), 721-729.

- Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., & Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, 111(2), 185-196. [https://doi.org/10.1016/s0092-8674\(02\)00975-3](https://doi.org/10.1016/s0092-8674(02)00975-3)
- Davis, R. L., Weintraub, H., & Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, 51(6), 987-1000. [https://doi.org/10.1016/0092-8674\(87\)90585-X](https://doi.org/10.1016/0092-8674(87)90585-X)
- Denell, R. E., & Frederick, R. D. (1983). Homoeosis in Drosophila: A description of the polycomb lethal syndrome. *Developmental Biology*, 97(1), 34-47. [https://doi.org/https://doi.org/10.1016/0012-1606\(83\)90061-1](https://doi.org/https://doi.org/10.1016/0012-1606(83)90061-1)
- Deshmukh, S., Ptack, A., Krug, B., & Jabado, N. (2022). Oncohistones: a roadmap to stalled development. *Febs j*, 289(5), 1315-1328. <https://doi.org/10.1111/febs.15963>
- Diehl, K. L., Ge, E. J., Weinberg, D. N., Jani, K. S., Allis, C. D., & Muir, T. W. (2019). PRC2 engages a bivalent H3K27M-H3K27me3 dinucleosome inhibitor. *Proceedings of the National Academy of Sciences*, 116(44), 22152-22157. <https://doi.org/doi:10.1073/pnas.1911775116>
- Doench, J. G. (2018). Am I ready for CRISPR? A user's guide to genetic screens. *Nature Reviews Genetics*, 19(2), 67-80. <https://doi.org/10.1038/nrg.2017.97>
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2), 184-191. <https://doi.org/10.1038/nbt.3437>
- Eberle-Singh, J. A., Sagalovskiy, I., Maurer, H. C., Sastra, S. A., Palermo, C. F., Decker, A. R., Kim, M. J., Sheedy, J., Mollin, A., Cao, L., Hu, J., Branstrom, A., Weetall, M., & Olive, K. P. (2019). Effective Delivery of a Microtubule Polymerization Inhibitor Synergizes with Standard Regimens in Models of Pancreatic Ductal Adenocarcinoma. *Clin Cancer Res*, 25(18), 5548-5560. <https://doi.org/10.1158/1078-0432.Ccr-18-3281>
- Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A. I., Wutz, A., & Bickmore, W. A. (2010). Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Molecular Cell*, 38(3), 452-464. <https://doi.org/10.1016/j.molcel.2010.02.032>
- Falzone, L., Salomone, S., & Libra, M. (2018). Evolution of Cancer Pharmacological Treatments at the Turn of the Third Millennium. *Front Pharmacol*, 9, 1300. <https://doi.org/10.3389/fphar.2018.01300>
- Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W., Koseki, H., Brockdorff, N., Ponting, C. P., Kessler, B. M., & Klose, R. J. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife*, 1, e00205. <https://doi.org/10.7554/eLife.00205>
- Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N. J., Johnson, D. A., Richardson, T. B., Santarosa, M., Dillon, K. J., Hickson, I., Knights, C., Martin, N. M. B., Jackson, S. P., Smith, G. C. M., & Ashworth, A. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, 434(7035), 917-921. <https://doi.org/10.1038/nature03445>

- Faust, C., Schumacher, A., Holdener, B., & Magnuson, T. (1995). The *eed* mutation disrupts anterior mesoderm production in mice. *Development*, 121(2), 273-285. <https://doi.org/10.1242/dev.121.2.273>
- Ferrari, K. J., 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(1), 49-62. <https://doi.org/10.1016/j.molcel.2013.10.030>
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806-811. <https://doi.org/10.1038/35888>
- Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., & Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev*, 17(15), 1870-1881. <https://doi.org/10.1101/gad.1110503>
- Fong, P. C., Boss, D. S., Yap, T. A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M. J., Ashworth, A., Carmichael, J., Kaye, S. B., Schellens, J. H. M., & Bono, J. S. d. (2009). Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from *BRCA* Mutation Carriers. *New England Journal of Medicine*, 361(2), 123-134. <https://doi.org/doi:10.1056/NEJMoa0900212>
- Forzati, F., Federico, A., Pallante, P., Abbate, A., Esposito, F., Malapelle, U., Sepe, R., Palma, G., Troncone, G., Scarfò, M., Arra, C., Fedele, M., & Fusco, A. (2012). CBX7 is a tumor suppressor in mice and humans. *J Clin Invest*, 122(2), 612-623. <https://doi.org/10.1172/jci58620>
- Francis, N. J., Follmer, N. E., Simon, M. D., Aghia, G., & Butler, J. D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell*, 137(1), 110-122. <https://doi.org/10.1016/j.cell.2009.02.017>
- Fursova, N. A., Blackledge, N. P., Nakayama, M., Ito, S., Koseki, Y., Farcas, A. M., King, H. W., Koseki, H., & Klose, R. J. (2019). Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol Cell*, 74(5), 1020-1036.e1028. <https://doi.org/10.1016/j.molcel.2019.03.024>
- Gao, Y., & Vakoc, C. R. (2024). Therapeutic index of targeting select chromatin complexes in human cancer patients. *Current Opinion in Genetics & Development*, 85, 102162. <https://doi.org/https://doi.org/10.1016/j.gde.2024.102162>
- Gao, Z., Lee, P., Stafford, J. M., von Schimmelmänn, M., Schaefer, A., & Reinberg, D. (2014). An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature*, 516(7531), 349-354. <https://doi.org/10.1038/nature13921>
- Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., & Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Molecular Cell*, 45(3), 344-356. <https://doi.org/10.1016/j.molcel.2012.01.002>
- Garraway, L. A., Widlund, H. R., Rubin, M. A., Getz, G., Berger, A. J., Ramaswamy, S., Beroukhi, R., Milner, D. A., Granter, S. R., Du, J., Lee, C., Wagner, S. N., Li, C., Golub, T. R., Rimm, D. L., Meyerson, M. L., Fisher, D. E., & Sellers, W. R. (2005). Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature*, 436(7047), 117-122. <https://doi.org/10.1038/nature03664>

- Gearhart, M. D., Corcoran, C. M., Wamstad, J. A., & Bardwell, V. J. (2006). Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol Cell Biol*, 26(18), 6880-6889. <https://doi.org/10.1128/mcb.00630-06>
- Glancy, E., Wang, C., Tuck, E., Healy, E., Amato, S., Neikes, H. K., Mariani, A., Mucha, M., Vermeulen, M., Pasini, D., & Bracken, A. P. (2023). PRC2.1- and PRC2.2-specific accessory proteins drive recruitment of different forms of canonical PRC1. *Mol Cell*, 83(9), 1393-1411.e1397. <https://doi.org/10.1016/j.molcel.2023.03.018>
- Goodman, L. S., Wintrobe, M. M., & et al. (1946). Nitrogen mustard therapy; use of methyl-bis (beta-chloroethyl) amine hydrochloride and tris (beta-chloroethyl) amine hydrochloride for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *J Am Med Assoc*, 132, 126-132. <https://doi.org/10.1001/jama.1946.02870380008004>
- Gorfinkiel, N., Fanti, L., Melgar, T., García, E., Pimpinelli, S., Guerrero, I., & Vidal, M. (2004). The Drosophila Polycomb group gene Sex combs extra encodes the ortholog of mammalian Ring1 proteins. *Mechanisms of Development*, 121(5), 449-462. <https://doi.org/https://doi.org/10.1016/j.mod.2004.03.019>
- Grasso, C. S., Tang, Y., Truffaux, N., Berlow, N. E., Liu, L., Debily, M. A., Quist, M. J., Davis, L. E., Huang, E. C., Woo, P. J., Ponnuswami, A., Chen, S., Johung, T. B., Sun, W., Kogiso, M., Du, Y., Qi, L., Huang, Y., Hütt-Cabezas, M., . . . Monje, M. (2015). Functionally defined therapeutic targets in diffuse intrinsic pontine glioma. *Nat Med*, 21(6), 555-559. <https://doi.org/10.1038/nm.3855>
- Grau, D. J., Chapman, B. A., Garlick, J. D., Borowsky, M., Francis, N. J., & Kingston, R. E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev*, 25(20), 2210-2221. <https://doi.org/10.1101/gad.17288211>
- Gröbner, S. N., Worst, B. C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V. A., Johann, P. D., Balasubramanian, G. P., Segura-Wang, M., Brabetz, S., Bender, S., Hutter, B., Sturm, D., Pfaff, E., Hübschmann, D., Zipprich, G., Heinold, M., Eils, J., Lawerenz, C., . . . Project, I. M.-S. (2018). The landscape of genomic alterations across childhood cancers. *Nature*, 555(7696), 321-327. <https://doi.org/10.1038/nature25480>
- Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discov*, 12(1), 31-46. <https://doi.org/10.1158/2159-8290.Cd-21-1059>
- Harutyunyan, A. S., Chen, H., Lu, T., Horth, C., Nikbakht, H., Krug, B., Russo, C., Bareke, E., Marchione, D. M., Coradin, M., Garcia, B. A., Jabado, N., & Majewski, J. (2020). H3K27M in Gliomas Causes a One-Step Decrease in H3K27 Methylation and Reduced Spreading within the Constraints of H3K36 Methylation. *Cell Rep*, 33(7), 108390. <https://doi.org/10.1016/j.celrep.2020.108390>
- Harutyunyan, A. S., Krug, B., Chen, H., Papillon-Cavanagh, S., Zeinieh, M., De Jay, N., Deshmukh, S., Chen, C. C. L., Belle, J., Mikael, L. G., Marchione, D. M., Li, R., Nikbakht, H., Hu, B., Cagnone, G., Cheung, W. A., Mohammadnia, A., Bechet, D., Faury, D., . . . Majewski, J. (2019). H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. *Nature Communications*, 10(1), 1262. <https://doi.org/10.1038/s41467-019-09140-x>

- 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(2), 583-595. <https://doi.org/10.1016/j.celrep.2016.08.096>
- Hay, M., Thomas, D. W., Craighead, J. L., Economides, C., & Rosenthal, J. (2014). Clinical development success rates for investigational drugs. *Nat Biotechnol*, 32(1), 40-51. <https://doi.org/10.1038/nbt.2786>
- He, W., Zhang, L., Villarreal, O. D., Fu, R., Bedford, E., Dou, J., Patel, A. Y., Bedford, M. T., Shi, X., Chen, T., Bartholomew, B., & Xu, H. (2019). De novo identification of essential protein domains from CRISPR-Cas9 tiling-sgRNA knockout screens. *Nature Communications*, 10(1), 4541. <https://doi.org/10.1038/s41467-019-12489-8>
- Hebbes, T. R., Thorne, A. W., & Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. *The EMBO Journal*, 7(5), 1395-1402-1402. <https://doi.org/https://doi.org/10.1002/j.1460-2075.1988.tb02956.x>
- Højfeldt, J. W., 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(3), 423-436.e423. <https://doi.org/10.1016/j.molcel.2019.07.031>
- Højfeldt, J. W., Laugesen, A., Willumsen, B. M., Damhofer, H., Hedehus, L., Tvardovskiy, A., Mohammad, F., Jensen, O. N., & Helin, K. (2018). Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nat Struct Mol Biol*, 25(3), 225-232. <https://doi.org/10.1038/s41594-018-0036-6>
- Hong, S., Cho, Y.-W., Yu, L.-R., Yu, H., Veenstra, T. D., & Ge, K. (2007). Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proceedings of the National Academy of Sciences*, 104(47), 18439-18444. <https://doi.org/doi:10.1073/pnas.0707292104>
- Hsu, C., Konner, J. A., & Gounder, M. M. (2024). Epigenetic Therapy in a Rare Ovarian Cancer - A Double-Edged Sword. *N Engl J Med*, 391(8), 770-772. <https://doi.org/10.1056/NEJMc2407971>
- Huang, A., Garraway, L. A., Ashworth, A., & Weber, B. (2020). Synthetic lethality as an engine for cancer drug target discovery. *Nature Reviews Drug Discovery*, 19(1), 23-38. <https://doi.org/10.1038/s41573-019-0046-z>
- Hübner, J. M., Müller, T., Papageorgiou, D. N., Mauermann, M., Krijgsveld, J., Russell, R. B., Ellison, D. W., Pfister, S. M., Pajtler, K. W., & 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(7), 878-889. <https://doi.org/10.1093/neuonc/noz058>
- Illingworth, R. S., & Bird, A. P. (2009). CpG islands--'a rough guide'. *FEBS Lett*, 583(11), 1713-1720. <https://doi.org/10.1016/j.febslet.2009.04.012>
- Ismail, I. H., Gagné, J.-P., Caron, M.-C., McDonald, D., Xu, Z., Masson, J.-Y., Poirier, G. G., & Hendzel, M. J. (2012). CBX4-mediated SUMO modification regulates BMI1 recruitment at sites of DNA damage. *Nucleic Acids Research*, 40(12), 5497-5510. <https://doi.org/10.1093/nar/gks222>
- Isono, K., Endo, T. A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B. E., & Koseki, H. (2013). SAM domain

- polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell*, 26(6), 565-577. <https://doi.org/10.1016/j.devcel.2013.08.016>
- Jacobs, J. J., Scheijen, B., Voncken, J. W., Kieboom, K., Berns, A., & van Lohuizen, M. (1999). Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. *Genes Dev*, 13(20), 2678-2690. <https://doi.org/10.1101/gad.13.20.2678>
- Jain, S. U., Do, T. J., Lund, P. J., Rashoff, A. Q., Diehl, K. L., Cieslik, M., Bajic, A., Juretic, N., Deshmukh, S., Venneti, S., Muir, T. W., Garcia, B. A., Jabado, N., & Lewis, P. W. (2019). PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism. *Nature Communications*, 10(1), 2146. <https://doi.org/10.1038/s41467-019-09981-6>
- Jain, S. U., Rashoff, A. Q., Krabbenhoft, S. D., Hoelper, D., Do, T. J., Gibson, T. J., Lundgren, S. M., Bondra, E. R., Deshmukh, S., Harutyunyan, A. S., Juretic, N., Jabado, N., Harrison, M. M., & Lewis, P. W. (2020). H3 K27M and EZHIP Impede H3K27-Methylation Spreading by Inhibiting Allosterically Stimulated PRC2. *Mol Cell*, 80(4), 726-735.e727. <https://doi.org/10.1016/j.molcel.2020.09.028>
- Jenuwein, T., & Allis, C. D. (2001). Translating the Histone Code. *Science*, 293(5532), 1074-1080. <https://doi.org/doi:10.1126/science.1063127>
- Jessa, S., Mohammadnia, A., Harutyunyan, A. S., Hulswit, M., Varadharajan, S., Lakkis, H., Kabir, N., Bashardanesh, Z., Hébert, S., Faury, D., Vladiou, M. C., Worme, S., Coutelier, M., Krug, B., Faria Andrade, A., Pathania, M., Bajic, A., Weil, A. G., Ellezam, B., . . . Kleinman, C. L. (2022). K27M in canonical and noncanonical H3 variants occurs in distinct oligodendroglial cell lineages in brain midline gliomas. *Nature Genetics*, 54(12), 1865-1880. <https://doi.org/10.1038/s41588-022-01205-w>
- Jiao, L., & Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science*, 350(6258), aac4383. <https://doi.org/10.1126/science.aac4383>
- Jones, C., & Baker, S. J. (2014). Unique genetic and epigenetic mechanisms driving paediatric diffuse high-grade glioma. *Nat Rev Cancer*, 14(10). <https://doi.org/10.1038/nrc3811>
- Jones, R. S., & Gelbart, W. M. (1993). The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. *Mol Cell Biol*, 13(10), 6357-6366. <https://doi.org/10.1128/mcb.13.10.6357-6366.1993>
- Kagey, M. H., Melhuish, T. A., & Wotton, D. (2003). The Polycomb Protein Pc2 Is a SUMO E3. *Cell*, 113(1), 127-137. [https://doi.org/10.1016/S0092-8674\(03\)00159-4](https://doi.org/10.1016/S0092-8674(03)00159-4)
- Kantarjian, H., O'Brien, S., Jabbour, E., Garcia-Manero, G., Quintas-Cardama, A., Shan, J., Rios, M. B., Ravandi, F., Faderl, S., Kadia, T., Borthakur, G., Huang, X., Champlin, R., Talpaz, M., & Cortes, J. (2012). Improved survival in chronic myeloid leukemia since the introduction of imatinib therapy: a single-institution historical experience. *Blood*, 119(9), 1981-1987. <https://doi.org/10.1182/blood-2011-08-358135>
- 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(6527). <https://doi.org/10.1126/science.abc3393>

- Kassis, J. A., Kennison, J. A., & Tamkun, J. W. (2017). Polycomb and Trithorax Group Genes in *Drosophila*. *Genetics*, 206(4), 1699-1725. <https://doi.org/10.1534/genetics.115.185116>
- Katoh-Fukui, Y., Baba, T., Sato, T., Otake, H., Nagakui-Noguchi, Y., Shindo, M., Suyama, M., Ohkawa, Y., Tsumura, H., Morohashi, K.-i., & Fukami, M. (2019). Mouse polycomb group gene *Cbx2* promotes osteoblastic but suppresses adipogenic differentiation in postnatal long bones. *Bone*, 120, 219-231. <https://doi.org/10.1016/j.bone.2018.10.021>
- Katoh-Fukui, Y., Miyabayashi, K., Komatsu, T., Owaki, A., Baba, T., Shima, Y., Kidokoro, T., Kanai, Y., Schedl, A., Wilhelm, D., Koopman, P., Okuno, Y., & Morohashi, K.-i. (2012). *Cbx2*, a Polycomb Group Gene, Is Required for *Sry* Gene Expression in Mice. *Endocrinology*, 153(2), 913-924. <https://doi.org/10.1210/en.2011-1055>
- Kim, J., & Kingston, R. E. (2020). The CBX family of proteins in transcriptional repression and memory. *J Biosci*, 45.
- Kim, J. J., & Kingston, R. E. (2022). Context-specific Polycomb mechanisms in development. *Nature Reviews Genetics*, 23(11), 680-695. <https://doi.org/10.1038/s41576-022-00499-0>
- Kim, K. H., & Roberts, C. W. (2016). Targeting EZH2 in cancer. *Nat Med*, 22(2), 128-134. <https://doi.org/10.1038/nm.4036>
- Klauke, K., Radulović, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K., Bystrykh, L., & de Haan, G. (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat Cell Biol*, 15(4), 353-362. <https://doi.org/10.1038/ncb2701>
- Konda, P., Garinet, S., Van Allen, E. M., & Viswanathan, S. R. (2023). Genome-guided discovery of cancer therapeutic targets. *Cell Reports*, 42(8). <https://doi.org/10.1016/j.celrep.2023.112978>
- Kornberg, R. D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. *Science*, 184(4139), 868-871. <https://doi.org/10.1126/science.184.4139.868>
- Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., Adli, M., Kasif, S., Ptaszek, L. M., Cowan, C. A., Lander, E. S., Koseki, H., & Bernstein, B. E. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet*, 4(10), e1000242. <https://doi.org/10.1371/journal.pgen.1000242>
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., & Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev*, 16(22), 2893-2905. <https://doi.org/10.1101/gad.1035902>
- Lagarou, A., Mohd-Sarip, A., Moshkin, Y. M., Chalkley, G. E., Bezstarosti, K., Demmers, J. A., & Verrijzer, C. P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev*, 22(20), 2799-2810. <https://doi.org/10.1101/gad.484208>
- Larson, J. D., Kasper, L. H., Paugh, B. S., Jin, H., Wu, G., Kwon, C. H., Fan, Y., Shaw, T. I., Silveira, A. B., Qu, C., Xu, R., Zhu, X., Zhang, J., Russell, H. R., Peters, J. L., Finkelstein, D., Xu, B., Lin, T., Tinkle, C. L., . . . Baker, S. J. (2019). Histone H3.3 K27M Accelerates Spontaneous Brainstem Glioma

- and Drives Restricted Changes in Bivalent Gene Expression. *Cancer Cell*, 35(1), 140-155.e147. <https://doi.org/10.1016/j.ccell.2018.11.015>
- Laskey, R. A., & Gurdon, J. B. (1970). Genetic Content of Adult Somatic Cells tested by Nuclear Transplantation from Cultured Cells. *Nature*, 228(5278), 1332-1334. <https://doi.org/10.1038/2281332a0>
- Lau, M. S., Schwartz, M. G., Kundu, S., Savol, A. J., Wang, P. I., Marr, S. K., Grau, D. J., Schorderet, P., Sadreyev, R. I., Tabin, C. J., & Kingston, R. E. (2017). Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. *Science*, 355(6329), 1081-1084. <https://doi.org/10.1126/science.aah5403>
- Laugesen, A., Højfeldt, J. W., & Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Molecular Cell*, 74(1), 8-18. <https://doi.org/https://doi.org/10.1016/j.molcel.2019.03.011>
- Lauseker, M., Hehlmann, R., Hochhaus, A., & Sauße, S. (2023). Survival with chronic myeloid leukaemia after failing milestones. *Leukemia*, 37(11), 2231-2236. <https://doi.org/10.1038/s41375-023-02028-2>
- Lee, W., Teckie, S., Wiesner, T., Ran, L., Prieto Granada, C. N., Lin, M., Zhu, S., Cao, Z., Liang, Y., Sboner, A., Tap, W. D., Fletcher, J. A., Huberman, K. H., Qin, L. X., Viale, A., Singer, S., Zheng, D., Berger, M. F., Chen, Y., . . . Chi, P. (2014). PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. *Nat Genet*, 46(11), 1227-1232. <https://doi.org/10.1038/ng.3095>
- Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature*, 276(5688), 565-570. <https://doi.org/10.1038/276565a0>
- Lewis, P. (1947). New mutants report. *Drosoph. Inf. Serv*, 21, 69.
- Lewis, P. W., Müller, M. M., Koletsky, M. S., Cordero, F., Lin, S., Banaszynski, L. A., Garcia, B. A., Muir, T. W., Becher, O. J., & Allis, C. D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science*, 340(6134), 857-861. <https://doi.org/10.1126/science.1232245>
- Li, H., Liefke, R., Jiang, J., Kurland, J. V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D. J., Bulky, M. L., Shi, Y., & Wang, Z. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature*, 549(7671), 287-291. <https://doi.org/10.1038/nature23881>
- Liang, R., Tomita, D., Sasaki, Y., Ginn, J., Michino, M., Huggins, D. J., Baxt, L., Kargman, S., Shahid, M., Aso, K., Duggan, M., Stamford, A. W., DeStanchina, E., Liverton, N., Meinke, P. T., Foley, M. A., & Phillips, R. E. (2022). A Chemical Strategy toward Novel Brain-Penetrant EZH2 Inhibitors. *ACS Med Chem Lett*, 13(3), 377-387. <https://doi.org/10.1021/acsmedchemlett.1c00448>
- Lin, A., & Sheltzer, J. M. (2020). Discovering and validating cancer genetic dependencies: approaches and pitfalls. *Nat Rev Genet*, 21(11), 671-682. <https://doi.org/10.1038/s41576-020-0247-7>
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315-322. <https://doi.org/10.1038/nature08514>

- Liu, B., Liu, Y. F., Du, Y. R., Mardaryev, A. N., Yang, W., Chen, H., Xu, Z. M., Xu, C. Q., Zhang, X. R., Botchkarev, V. A., Zhang, Y., & Xu, G. L. (2013). Cbx4 regulates the proliferation of thymic epithelial cells and thymus function. *Development*, 140(4), 780-788. <https://doi.org/10.1242/dev.085035>
- Lo, S. M., Ahuja, N. K., & Francis, N. J. (2009). Polycomb group protein Suppressor 2 of zeste is a functional homolog of Posterior Sex Combs. *Mol Cell Biol*, 29(2), 515-525. <https://doi.org/10.1128/mcb.01044-08>
- Lord, C. J., & Ashworth, A. (2017). PARP inhibitors: Synthetic lethality in the clinic. *Science*, 355(6330), 1152-1158. <https://doi.org/doi:10.1126/science.aam7344>
- Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., Ohgaki, H., Wiestler, O. D., Kleihues, P., & Ellison, D. W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta neuropathologica*, 131(6), 803-820. <https://doi.org/10.1007/s00401-016-1545-1>
- Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H. K., Pfister, S. M., Reifenberger, G., Soffietti, R., von Deimling, A., & Ellison, D. W. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol*, 23(8), 1231-1251. <https://doi.org/10.1093/neuonc/noab106>
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), 251-260. <https://doi.org/10.1038/38444>
- Luis, N. M., Morey, L., Mejetta, S., Pascual, G., Janich, P., Kuebler, B., Cozutto, L., Roma, G., Nascimento, E., Frye, M., Di Croce, L., & Benitah, S. A. (2011). Regulation of human epidermal stem cell proliferation and senescence requires polycomb- dependent and -independent functions of Cbx4. *Cell Stem Cell*, 9(3), 233-246. <https://doi.org/10.1016/j.stem.2011.07.013>
- Mackay, A., Burford, A., Carvalho, D., Izquierdo, E., Fazal-Salom, J., Taylor, K. R., Bjerke, L., Clarke, M., Vinci, M., Nandhabalan, M., Temelso, S., Popov, S., Molinari, V., Raman, P., Waanders, A. J., Han, H. J., Gupta, S., Marshall, L., Zacharoulis, S., . . . Jones, C. (2017). Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma. *Cancer Cell*, 32(4), 520-537.e525. <https://doi.org/10.1016/j.ccell.2017.08.017>
- Mardaryev, A. N., Liu, B., Rapisarda, V., Poterlowicz, K., Malashchuk, I., Rudolf, J., Sharov, A. A., Jahoda, C. A., Fessing, M. Y., Benitah, S. A., Xu, G. L., & Botchkarev, V. A. (2016). Cbx4 maintains the epithelial lineage identity and cell proliferation in the developing stratified epithelium. *J Cell Biol*, 212(1), 77-89. <https://doi.org/10.1083/jcb.201506065>
- Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury Iii, W. J., Voigt, P., Martin, S. R., Taylor, W. R., De Marco, V., Pirrotta, V., Reinberg, D., & Gamblin, S. J. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, 461(7265), 762-767. <https://doi.org/10.1038/nature08398>
- Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C. L., Dynlacht, B. D., & Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell*, 32(4), 503-518. <https://doi.org/10.1016/j.molcel.2008.11.004>
- McDonald, E. R., III, de Weck, A., Schlabach, M. R., Billy, E., Mavrikakis, K. J., Hoffman, G. R., Belur, D., Castelletti, D., Frias, E., Gampa, K., Golji, J.,

- Kao, I., Li, L., Megel, P., Perkins, T. A., Ramadan, N., Ruddy, D. A., Silver, S. J., Sovath, S., . . . Sellers, W. R. (2017). Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell*, 170(3), 577-592.e510. <https://doi.org/10.1016/j.cell.2017.07.005>
- Michealraj, K. A., Kumar, S. A., Kim, L. J. Y., Cavalli, F. M. G., Przelicki, D., Wojcik, J. B., Delaidelli, A., Bajic, A., Saulnier, O., MacLeod, G., Vellanki, R. N., Vladoiu, M. C., Guilhamon, P., Ong, W., Lee, J. J. Y., Jiang, Y., Holgado, B. L., Rasnitsyn, A., Malik, A. A., . . . Taylor, M. D. (2020). Metabolic Regulation of the Epigenome Drives Lethal Infantile Ependymoma. *Cell*, 181(6), 1329-1345.e1324. <https://doi.org/10.1016/j.cell.2020.04.047>
- Michel, B. C., D'Avino, A. R., Cassel, S. H., Mashtalir, N., McKenzie, Z. M., McBride, M. J., Valencia, A. M., Zhou, Q., Bocker, M., Soares, L. M. M., Pan, J., Remillard, D. I., Lareau, C. A., Zullo, H. J., Fortoul, N., Gray, N. S., Bradner, J. E., Chan, H. M., & Kadoch, C. (2018). A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol*, 20(12), 1410-1420. <https://doi.org/10.1038/s41556-018-0221-1>
- Min, J., Zhang, Y., & Xu, R. M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev*, 17(15), 1823-1828. <https://doi.org/10.1101/gad.269603>
- Mito, Y., Henikoff, J. G., & Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. *Nature Genetics*, 37(10), 1090-1097. <https://doi.org/10.1038/ng1637>
- Mohammad, F., & Helin, K. (2017). Oncohistones: drivers of pediatric cancers. *Genes Dev*, 31(23-24), 2313-2324. <https://doi.org/10.1101/gad.309013.117>
- Mohammad, F., Weissmann, S., Leblanc, B., Pandey, D. P., Højfeldt, J. W., Comet, I., Zheng, C., Johansen, J. V., Rapin, N., Porse, B. T., Tvardovskiy, A., Jensen, O. N., Olaciregui, N. G., Lavarino, C., Suñol, M., de Torres, C., Mora, J., Carcaboso, A. M., & Helin, K. (2017). EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nature Medicine*, 23(4), 483-492. <https://doi.org/10.1038/nm.4293>
- Molofsky, A. V., He, S., Bydon, M., Morrison, S. J., & Pandal, R. (2005). Bmi-1 promotes neural stem cell self-renewal and neural development but not mouse growth and survival by repressing the p16Ink4a and p19Arf senescence pathways. *Genes Dev*, 19(12), 1432-1437. <https://doi.org/10.1101/gad.1299505>
- Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A., & Di Croce, L. (2012). Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell*, 10(1), 47-62. <https://doi.org/10.1016/j.stem.2011.12.006>
- Morin, R. D., Johnson, N. A., Severson, T. M., Mungall, A. J., An, J., Goya, R., Paul, J. E., Boyle, M., Woolcock, B. W., Kuchenbauer, F., Yap, D., Humphries, R. K., Griffith, O. L., Shah, S., Zhu, H., Kimbara, M., Shashkin, P., Charlot, J. F., Tcherpakov, M., . . . Marra, M. A. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet*, 42(2), 181-185. <https://doi.org/10.1038/ng.518>
- Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E., & Simon, J. A. (2002). Histone

methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*, 111(2), 197-208. [https://doi.org/10.1016/s0092-8674\(02\)00976-5](https://doi.org/10.1016/s0092-8674(02)00976-5)

- Murphy, S. E., & Boettiger, A. N. (2024). Polycomb repression of Hox genes involves spatial feedback but not domain compaction or phase transition. *Nat Genet*, 56(3), 493-504. <https://doi.org/10.1038/s41588-024-01661-6>
- Nagaraja, S., Vitanza, N. A., Woo, P. J., Taylor, K. R., Liu, F., Zhang, L., Li, M., Meng, W., Ponnuswami, A., Sun, W., Ma, J., Hulleman, E., Swigut, T., Wysocka, J., Tang, Y., & Monje, M. (2017). Transcriptional Dependencies in Diffuse Intrinsic Pontine Glioma. *Cancer Cell*, 31(5), 635-652.e636. <https://doi.org/10.1016/j.ccell.2017.03.011>
- Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P. C., Li, M. J., Wang, J., Cardon, L. R., Whittaker, J. C., & Sanseau, P. (2015). The support of human genetic evidence for approved drug indications. *Nature Genetics*, 47(8), 856-860. <https://doi.org/10.1038/ng.3314>
- Nikbakht, H., Panditharatna, E., Mikael, L. G., Li, R., Gayden, T., Osmond, M., Ho, C.-Y., Kambhampati, M., Hwang, E. I., Faury, D., Siu, A., Papillon-Cavanagh, S., Bechet, D., Ligon, K. L., Ellezam, B., Ingram, W. J., Stinson, C., Moore, A. S., Warren, K. E., . . . Nazarian, J. (2016). Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma. *Nature Communications*, 7(1), 11185. <https://doi.org/10.1038/ncomms11185>
- Nowell, P. C. (1962). The minute chromosome (Ph1) in chronic granulocytic leukemia. *Blut*, 8, 65-66. <https://doi.org/10.1007/bf01630378>
- Ntziachristos, P., Tsirigos, A., Van Vlierberghe, P., Nedjic, J., Trimarchi, T., Flaherty, M. S., Ferres-Marco, D., da Ros, V., Tang, Z., Siegle, J., Asp, P., Hadler, M., Rigo, I., De Keersmaecker, K., Patel, J., Huynh, T., Utro, F., Poglio, S., Samon, J. B., . . . Aifantis, I. (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat Med*, 18(2), 298-301. <https://doi.org/10.1038/nm.2651>
- O'Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A., & Jenuwein, T. (2001). The polycomb-group gene *Ezh2* is required for early mouse development. *Mol Cell Biol*, 21(13), 4330-4336. <https://doi.org/10.1128/mcb.21.13.4330-4336.2001>
- Orlando, D. A., Chen, M. W., Brown, V. E., Solanki, S., Choi, Y. J., Olson, E. R., Fritz, C. C., Bradner, J. E., & Guenther, M. G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep*, 9(3), 1163-1170. <https://doi.org/10.1016/j.celrep.2014.10.018>
- Pajtler, K. W., Wen, J., Sill, M., Lin, T., Orisme, W., Tang, B., Hübner, J.-M., Ramaswamy, V., Jia, S., & Dalton, J. D. (2018). Molecular heterogeneity and CXorf67 alterations in posterior fossa group A (PFA) ependymomas. *Acta neuropathologica*, 136, 211-226.
- Pal, S., Kaplan, J. P., Nguyen, H., Stopka, S. A., Savani, M. R., Regan, M. S., Nguyen, Q.-D., Jones, K. L., Moreau, L. A., Peng, J., Dipiazza, M. G., Perciaccante, A. J., Zhu, X., Hunsel, B. R., Liu, K. X., Alexandrescu, S., Drissi, R., Filbin, M. G., McBrayer, S. K., . . . Haas-Kogan, D. A. (2022). A druggable addiction to de novo pyrimidine biosynthesis in diffuse midline glioma. *Cancer Cell*, 40(9), 957-972.e910. <https://doi.org/https://doi.org/10.1016/j.ccell.2022.07.012>

- Panditharatna, E., Marques, J. G., Wang, T., Trissal, M. C., Liu, I., Jiang, L., Beck, A., Groves, A., Dharia, N. V., Li, D., Hoffman, S. E., Kugener, G., Shaw, M. L., Mire, H. M., Hack, O. A., Dempster, J. M., Lareau, C., Dai, L., Sigua, L. H., . . . Filbin, M. G. (2022). BAF Complex Maintains Glioma Stem Cells in Pediatric H3K27M Glioma. *Cancer Discov*, 12(12), 2880-2905. <https://doi.org/10.1158/2159-8290.Cd-21-1491>
- Parreno, V., Martinez, A.-M., & Cavalli, G. (2022). Mechanisms of Polycomb group protein function in cancer. *Cell Research*, 32(3), 231-253. <https://doi.org/10.1038/s41422-021-00606-6>
- Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E., & Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *Embo j*, 23(20), 4061-4071. <https://doi.org/10.1038/sj.emboj.7600402>
- Pasini, D., & Di Croce, L. (2016). Emerging roles for Polycomb proteins in cancer. *Curr Opin Genet Dev*, 36, 50-58. <https://doi.org/10.1016/j.gde.2016.03.013>
- Pathania, M., De Jay, N., Maestro, N., Harutyunyan, A. S., Nitarska, J., Pahlavan, P., Henderson, S., Mikael, L. G., Richard-Londt, A., Zhang, Y., Costa, J. R., Hébert, S., Khazaei, S., Ibrahim, N. S., Herrero, J., Riccio, A., Albrecht, S., Ketteler, R., Brandner, S., . . . Salomoni, P. (2017). H3.3(K27M) Cooperates with Trp53 Loss and PDGFRA Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas. *Cancer Cell*, 32(5), 684-700.e689. <https://doi.org/10.1016/j.ccell.2017.09.014>
- Pemberton, H., Anderton, E., Patel, H., Brookes, S., Chandler, H., Palermo, R., Stock, J., Rodriguez-Niedenführ, M., Racek, T., de Breed, L., Stewart, A., Matthews, N., & Peters, G. (2014). Genome-wide co-localization of Polycomb orthologs and their effects on gene expression in human fibroblasts. *Genome Biology*, 15(2), R23. <https://doi.org/10.1186/gb-2014-15-2-r23>
- Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A., & Müller, J. (2013). A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb. *Science*, 339(6120), 698-699. <https://doi.org/doi:10.1126/science.1231382>
- Perino, M., van Mierlo, G., Karemaker, I. D., van Genesen, S., Vermeulen, M., Marks, H., van Heeringen, S. J., & Veenstra, G. J. C. (2018). MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat Genet*, 50(7), 1002-1010. <https://doi.org/10.1038/s41588-018-0134-8>
- Pierce, A., Spooncer, E., Ainsworth, S., & Whetton, A. D. (2002). BCR-ABL alters the proliferation and differentiation response of multipotent hematopoietic cells to stem cell factor. *Oncogene*, 21(19), 3068-3075. <https://doi.org/10.1038/sj.onc.1205424>
- Piovesan, A., Pelleri, M. C., Antonaros, F., Strippoli, P., Caracausi, M., & Vitale, L. (2019). On the length, weight and GC content of the human genome. *BMC Research Notes*, 12(1), 106. <https://doi.org/10.1186/s13104-019-4137-z>
- Piunti, A., & Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science*, 352(6290), aad9780. <https://doi.org/doi:10.1126/science.aad9780>
- Piunti, A., Smith, E. R., Morgan, M. A. J., Ugarenko, M., Khaltan, N., Helmin, K. A., Ryan, C. A., Murray, D. C., Rickels, R. A., Yilmaz, B. D., Rendleman, E. J., Savas, J. N., Singer, B. D., Bulun, S. E., & Shilatifard, A. (2019). CATACOMB: An endogenous inducible gene that antagonizes H3K27

- methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. *Science Advances*, 5(7), eaax2887.
<https://doi.org/doi:10.1126/sciadv.aax2887>
- Plys, A. J., Davis, C. P., Kim, J., Rizki, G., Keenen, M. M., Marr, S. K., & Kingston, R. E. (2019). Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev*, 33(13-14), 799-813. <https://doi.org/10.1101/gad.326488.119>
- Posfai, E., Kunzmann, R., Brochard, V., Salvaing, J., Cabuy, E., Roloff, T. C., Liu, Z., Tardat, M., van Lohuizen, M., Vidal, M., Beaujean, N., & Peters, A. H. (2012). Polycomb function during oogenesis is required for mouse embryonic development. *Genes Dev*, 26(9), 920-932.
<https://doi.org/10.1101/gad.188094.112>
- Quina, A. S., Buschbeck, M., & Di Croce, L. (2006). Chromatin structure and epigenetics. *Biochemical Pharmacology*, 72(11), 1563-1569.
<https://doi.org/https://doi.org/10.1016/j.bcp.2006.06.016>
- Ragazzini, R., Pérez-Palacios, R., Baymaz, I. H., Diop, S., Ancelin, K., Zielinski, D., Michaud, A., Givélet, M., Borsos, M., Aflaki, S., Legoix, P., Jansen, P., Servant, N., Torres-Padilla, M. E., Bourc'his, D., Fouchet, P., Vermeulen, M., & Margueron, R. (2019). EZHIP constrains Polycomb Repressive Complex 2 activity in germ cells. *Nat Commun*, 10(1), 3858.
<https://doi.org/10.1038/s41467-019-11800-x>
- Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., & Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796), 593-599.
<https://doi.org/10.1038/35020506>
- Ringrose, L., & Paro, R. (2004). Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annual Review of Genetics*, 38(Volume 38, 2004), 413-443.
<https://doi.org/https://doi.org/10.1146/annurev.genet.38.072902.091907>
- Rose, N. R., King, H. W., Blackledge, N. P., Fursova, N. A., Ember, K. J. I., Fischer, R., Kessler, B. M., & Klose, R. J. (2016). RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. *eLife*, 5, e18591. <https://doi.org/10.7554/eLife.18591>
- Rusina, P. V., Falaguera, M. J., Romero, J. M. R., McDonagh, E. M., Dunham, I., & Ochoa, D. (2023). Genetic support for FDA-approved drugs over the past decade. *Nat Rev Drug Discov*, 22(11), 864. <https://doi.org/10.1038/d41573-023-00158-x>
- Sankar, A., Mohammad, F., Sundaramurthy, A. K., Wang, H., Lerdrup, M., Tatar, T., & Helin, K. (2022). Histone editing elucidates the functional roles of H3K27 methylation and acetylation in mammals. *Nature Genetics*, 54(6), 754-760.
<https://doi.org/10.1038/s41588-022-01091-2>
- Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S. T., Offer, J., Loew, D., Servant, N., Wassef, M., Burlina, F., Gamblin, S. J., Heard, E., & Margueron, R. (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol Cell*, 57(5), 769-783.
<https://doi.org/10.1016/j.molcel.2014.12.020>
- Satijn, D. P. E., Gunster, M. J., van der Vlag, J., Hamer, K. M., Schul, W., Alkema, M. J., Saurin, A. J., Freemont, P. S., van Driel, R., & Otte, A. P. (1997).

- RING1 Is Associated with the Polycomb Group Protein Complex and Acts as a Transcriptional Repressor. *Molecular and Cellular Biology*, 17(7), 4105-4113. <https://doi.org/10.1128/MCB.17.7.4105>
- Saurin, A. J., Shao, Z., Erdjument-Bromage, H., Tempst, P., & Kingston, R. E. (2001). A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature*, 412(6847), 655-660. <https://doi.org/10.1038/35088096>
- Saurin, A. J., Shiels, C., Williamson, J., Satijn, D. P., Otte, A. P., Sheer, D., & Freemont, P. S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J Cell Biol*, 142(4), 887-898. <https://doi.org/10.1083/jcb.142.4.887>
- Scelfo, A., Fernández-Pérez, D., Tamburri, S., Zanotti, M., Lavarone, E., Soldi, M., Bonaldi, T., Ferrari, K. J., & Pasini, D. (2019). Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities. *Molecular Cell*, 74(5), 1037-1052.e1037. <https://doi.org/10.1016/j.molcel.2019.04.002>
- Schmitges, Frank W., Prusty, Archana B., Faty, M., Stützer, A., Lingaraju, Gondichatnahalli M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., Bunker, Richard D., Wirth, U., Bouwmeester, T., Bauer, A., Ly-Hartig, N., Zhao, K., Chan, H., Gu, J., Gut, H., . . . Thomä, Nicolas H. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Molecular Cell*, 42(3), 330-341. <https://doi.org/https://doi.org/10.1016/j.molcel.2011.03.025>
- Schoorlemmer, J., Marcos-Gutiérrez, C., Were, F., Martínez, R., García, E., Satijn, D. P., Otte, A. P., & Vidal, M. (1997). Ring1A is a transcriptional repressor that interacts with the Polycomb-M33 protein and is expressed at rhombomere boundaries in the mouse hindbrain. *Embo j*, 16(19), 5930-5942. <https://doi.org/10.1093/emboj/16.19.5930>
- Schuettengruber, B., Bourbon, H.-M., Di Croce, L., & Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell*, 171(1), 34-57. <https://doi.org/10.1016/j.cell.2017.08.002>
- Schuettengruber, B., & Cavalli, G. (2009). Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development*, 136(21), 3531-3542. <https://doi.org/10.1242/dev.033902>
- Schwartzentruber, J., Korshunov, A., Liu, X.-Y., Jones, D. T. W., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A. M., Quang, D.-A. K., Tönjes, M., Hovestadt, V., Albrecht, S., Kool, M., Nantel, A., Konermann, C., Lindroth, A., Jäger, N., Rausch, T., Ryzhova, M., . . . Jabado, N. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*, 482(7384), 226-231. <https://doi.org/10.1038/nature10833>
- Scott, E. C., Baines, A. C., Gong, Y., Moore, R., Jr., Pamuk, G. E., Saber, H., Subedee, A., Thompson, M. D., Xiao, W., Pazdur, R., Rao, V. A., Schneider, J., & Beaver, J. A. (2023). Trends in the approval of cancer therapies by the FDA in the twenty-first century. *Nat Rev Drug Discov*, 22(8), 625-640. <https://doi.org/10.1038/s41573-023-00723-4>
- Seto, E., & Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb Perspect Biol*, 6(4), a018713. <https://doi.org/10.1101/cshperspect.a018713>
- Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., & Zhang, F. (2014).

- Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*, 343(6166), 84-87. <https://doi.org/doi:10.1126/science.1247005>
- Shen, X., Liu, Y., Hsu, Y. J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G. C., & Orkin, S. H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell*, 32(4), 491-502. <https://doi.org/10.1016/j.molcel.2008.10.016>
- Shi, J., Wang, E., Milazzo, J. P., Wang, Z., Kinney, J. B., & Vakoc, C. R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol*, 33(6), 661-667. <https://doi.org/10.1038/nbt.3235>
- Silveira, A. B., Kasper, L. H., Fan, Y., Jin, H., Wu, G., Shaw, T. I., Zhu, X., Larson, J. D., Easton, J., Shao, Y., Yergeau, D. A., Rosencrance, C., Boggs, K., Rusch, M. C., Ding, L., Zhang, J., Finkelstein, D., Noyes, R. M., Russell, B. L., . . . Baker, S. J. (2019). H3.3 K27M depletion increases differentiation and extends latency of diffuse intrinsic pontine glioma growth in vivo. *Acta Neuropathol*, 137(4), 637-655. <https://doi.org/10.1007/s00401-019-01975-4>
- Smith, M. M., & Andr sson,  . S. (1983). DNA sequences of yeast H3 and H4 histone genes from two non-allelic gene sets encode identical H3 and H4 proteins. *Journal of Molecular Biology*, 169(3), 663-690. [https://doi.org/https://doi.org/10.1016/S0022-2836\(83\)80164-8](https://doi.org/https://doi.org/10.1016/S0022-2836(83)80164-8)
- Sneeringer, C. J., Scott, M. P., Kuntz, K. W., Knutson, S. K., Pollock, R. M., Richon, V. M., & Copeland, R. A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci U S A*, 107(49), 20980-20985. <https://doi.org/10.1073/pnas.1012525107>
- Spahn, P. N., Bath, T., Weiss, R. J., Kim, J., Esko, J. D., Lewis, N. E., & Harismendy, O. (2017). PinAPL-Py: A comprehensive web-application for the analysis of CRISPR/Cas9 screens. *Scientific Reports*, 7(1), 15854. <https://doi.org/10.1038/s41598-017-16193-9>
- Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., & Suske, G. (2018). MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet*, 14(1), e1007193. <https://doi.org/10.1371/journal.pgen.1007193>
- Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765), 41-45. <https://doi.org/10.1038/47412>
- Straining, R., & Eighmy, W. (2022). Tazemetostat: EZH2 Inhibitor. *J Adv Pract Oncol*, 13(2), 158-163. <https://doi.org/10.6004/jadpro.2022.13.2.7>
- Stuckey, J. I., Dickson, B. M., Cheng, N., Liu, Y., Norris, J. L., Cholensky, S. H., Tempel, W., Qin, S., Huber, K. G., Sagum, C., Black, K., Li, F., Huang, X. P., Roth, B. L., Baughman, B. M., Senisterra, G., Pattenden, S. G., Vedadi, M., Brown, P. J., . . . Frye, S. V. (2016). A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1. *Nat Chem Biol*, 12(3), 180-187. <https://doi.org/10.1038/nchembio.2007>
- Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), 663-676. <https://doi.org/10.1016/j.cell.2006.07.024>
- Tamburri, S., Lavarone, E., Fern ndez-P rez, D., Conway, E., Zanotti, M., Manganaro, D., & Pasini, D. (2020). Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Mol Cell*, 77(4), 840-856.e845. <https://doi.org/10.1016/j.molcel.2019.11.021>

- Tamburri, S., Rustichelli, S., Amato, S., & Pasini, D. (2024). Navigating the complexity of Polycomb repression: Enzymatic cores and regulatory modules. *Molecular Cell*, 84(18), 3381-3405. <https://doi.org/10.1016/j.molcel.2024.07.030>
- Tatavosian, R., Kent, S., Brown, K., Yao, T., Duc, H. N., Huynh, T. N., Zhen, C. Y., Ma, B., Wang, H., & Ren, X. (2019). Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *J Biol Chem*, 294(5), 1451-1463. <https://doi.org/10.1074/jbc.RA118.006620>
- Tremethick, D. J. (2007). Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell*, 128(4), 651-654. <https://doi.org/10.1016/j.cell.2007.02.008>
- Tsang, B., Pritisanac, I., Scherer, S. W., Moses, A. M., & Forman-Kay, J. D. (2020). Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell*, 183(7), 1742-1756. <https://doi.org/10.1016/j.cell.2020.11.050>
- Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., & Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *Embo j*, 13(16), 3822-3831. <https://doi.org/10.1002/j.1460-2075.1994.tb06693.x>
- Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., Gill, S., Harrington, W. F., Pantel, S., Krill-Burger, J. M., Meyers, R. M., Ali, L., Goodale, A., Lee, Y., Jiang, G., Hsiao, J., Gerath, W. F. J., Howell, S., Merkel, E., . . . Hahn, W. C. (2017). Defining a Cancer Dependency Map. *Cell*, 170(3), 564-576.e516. <https://doi.org/10.1016/j.cell.2017.06.010>
- van der Lugt, N. M. T., Alkema, M., Berns, A., & Deschamps, J. (1996). The Polycomb-group homolog Bmi-1 is a regulator of murine Hox gene expression. *Mechanisms of Development*, 58(1), 153-164. [https://doi.org/https://doi.org/10.1016/S0925-4773\(96\)00570-9](https://doi.org/https://doi.org/10.1016/S0925-4773(96)00570-9)
- Vandamme, J., Völkel, P., Rosnoblet, C., Le Faou, P., & Angrand, P. O. (2011). Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Mol Cell Proteomics*, 10(4), M110.002642. <https://doi.org/10.1074/mcp.M110.002642>
- Venneti, S., Santi, M., Felicella, M. M., Yarinlin, D., Phillips, J. J., Sullivan, L. M., Martinez, D., Perry, A., Lewis, P. W., Thompson, C. B., & Judkins, A. R. (2014). A sensitive and specific histopathologic prognostic marker for H3F3A K27M mutant pediatric glioblastomas. *Acta neuropathologica*, 128(5), 743-753. <https://doi.org/10.1007/s00401-014-1338-3>
- Vinci, M., Burford, A., Molinari, V., Kessler, K., Popov, S., Clarke, M., Taylor, K. R., Pemberton, H. N., Lord, C. J., Gutteridge, A., Forshe, T., Carvalho, D., Marshall, L. V., Qin, E. Y., Ingram, W. J., Moore, A. S., Ng, H. K., Trabelsi, S., H'Mida-Ben Brahim, D., . . . Jones, C. (2018). Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. *Nat Med*, 24(8), 1204-1215. <https://doi.org/10.1038/s41591-018-0086-7>
- Vitanza, N. A., & Monje, M. (2019). Diffuse Intrinsic Pontine Glioma: From Diagnosis to Next-Generation Clinical Trials. *Curr Treat Options Neurol*, 21(8), 37. <https://doi.org/10.1007/s11940-019-0577-y>
- Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*, 339(6127), 1546-1558. <https://doi.org/10.1126/science.1235122>

- Voncken, J. W., Roelen, B. A. J., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J., & van Lohuizen, M. (2003). *Rnf2* (*Ring1b*) deficiency causes gastrulation arrest and cell cycle inhibition. *Proceedings of the National Academy of Sciences*, 100(5), 2468-2473. <https://doi.org/doi:10.1073/pnas.0434312100>
- Waddington, C. H. (1957). *The Strategy of the Genes* (1 ed.). Routledge. <https://doi.org/https://doi.org/10.4324/9781315765471>
- Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C. H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z., Mei, S., Cui, Y., Shi, S., Lipp, J. J., Hinterndorfer, M., Zuber, J., Brown, M., Li, W., & Liu, X. S. (2019). Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc*, 14(3), 756-780. <https://doi.org/10.1038/s41596-018-0113-7>
- Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, 431(7010), 873-878. <https://doi.org/10.1038/nature02985>
- Wang, J. Y., & Doudna, J. A. (2023). CRISPR technology: A decade of genome editing is only the beginning. *Science*, 379(6629), eadd8643. <https://doi.org/10.1126/science.add8643>
- Wenger, A., Karlsson, I., Kling, T., & Carén, H. (2023). CRISPR-Cas9 knockout screen identifies novel treatment targets in childhood high-grade glioma. *Clinical Epigenetics*, 15(1), 80. <https://doi.org/10.1186/s13148-023-01498-6>
- Whitcomb, S. J., Basu, A., Allis, C. D., & Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. *Trends in Genetics*, 23(10), 494-502. <https://doi.org/10.1016/j.tig.2007.08.006>
- Woodcock, C. L., & Ghosh, R. P. (2010). Chromatin Higher-order Structure and Dynamics. *Cold Spring Harb Perspect Biol*. <https://doi.org/10.1101/cshperspect.a000596>
- Wright, P. E., & Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol*, 16(1), 18-29. <https://doi.org/10.1038/nrm3920>
- Wu, G., Broniscer, A., McEachron, T. A., Lu, C., Paugh, B. S., Becksfort, J., Qu, C., Ding, L., Huether, R., Parker, M., Zhang, J., Gajjar, A., Dyer, M. A., Mullighan, C. G., Gilbertson, R. J., Mardis, E. R., Wilson, R. K., Downing, J. R., Ellison, D. W., . . . Baker, S. J. (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet*, 44(3), 251-253. <https://doi.org/10.1038/ng.1102>
- Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., & Zhu, B. (2011). H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation ^{*}. *Journal of Biological Chemistry*, 286(10), 7983-7989. <https://doi.org/10.1074/jbc.M110.194027>
- Zhang, P., de Gooijer, M. C., Buil, L. C., Beijnen, J. H., Li, G., & van Tellingen, O. (2015). ABCB1 and ABCG2 restrict the brain penetration of a panel of novel EZH2-Inhibitors. *Int J Cancer*, 137(8), 2007-2018. <https://doi.org/10.1002/ijc.29566>
- Zhang, Q., Agius, S. C., Flanigan, S. F., Uckelmann, M., Levina, V., Owen, B. M., & Davidovich, C. (2021). PAL1 facilitates DNA and nucleosome binding by PRC2 and triggers an allosteric activation of catalysis. *Nature Communications*, 12(1), 4592. <https://doi.org/10.1038/s41467-021-24866-3>
- Zhao, J., Wang, M., Chang, L., Yu, J., Song, A., Liu, C., Huang, W., Zhang, T., Wu, X., Shen, X., Zhu, B., & Li, G. (2020). RYBP/YAF2-PRC1 complexes and

- histone H1-dependent chromatin compaction mediate propagation of H2AK119ub1 during cell division. *Nat Cell Biol*, 22(4), 439-452.
<https://doi.org/10.1038/s41556-020-0484-1>
- Zhen, C. Y., Duc, H. N., Kokotovic, M., Phiel, C. J., & Ren, X. (2014). Cbx2 stably associates with mitotic chromosomes via a PRC2- or PRC1-independent mechanism and is needed for recruiting PRC1 complex to mitotic chromosomes. *Mol Biol Cell*, 25(23), 3726-3739.
<https://doi.org/10.1091/mbc.E14-06-1109>
- Zheng, Y., Sweet, S. M. M., Popovic, R., Martinez-Garcia, E., Tipton, J. D., Thomas, P. M., Licht, J. D., & Kelleher, N. L. (2012). Total kinetic analysis reveals how combinatorial methylation patterns are established on lysines 27 and 36 of histone H3. *Proceedings of the National Academy of Sciences*, 109(34), 13549-13554.
<https://doi.org/doi:10.1073/pnas.1205707109>
- Zhu, Y., Dong, L., Wang, C., Hao, K., Wang, J., Zhao, L., Xu, L., Xia, Y., Jiang, Q., & Qin, J. (2022). Functional redundancy among Polycomb complexes in maintaining the pluripotent state of embryonic stem cells. *Stem Cell Reports*, 17(5), 1198-1214. <https://doi.org/10.1016/j.stemcr.2022.02.020>