

Exploring the functions of PRC2 in development and congenital disorders

Orla Deevy

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

2022

Thesis Supervisor: Prof. Adrian Bracken

Cancer Epigenetics Laboratory,

Department of Genetics,

Trinity College Dublin, The University of Dublin



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of 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, except where otherwise acknowledged.

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

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish.

A handwritten signature in black ink, reading 'Orla Deevy', written over a horizontal line.

Orla Deevy, September 2021

SUMMARY

Polycomb group (PcG) proteins are key chromatin-based repressors of gene expression that play important roles in lineage-specification and cellular memory. PcG proteins assemble into multi-subunit complexes, termed Polycomb Repressive Complexes (PRCs), which elicit transcriptional silencing through chromatin condensation. PRC2 consists of three core PcG proteins (SUZ12, EED and EZH1/2) that together catalyse the addition of up to three methyl groups on histone H3 at lysine 27 (H3K27). This catalytic core interacts with accessory proteins that function to modulate and refine PRC2 activity. Both core and accessory PRC2 components are essential for normal development and their aberrant function thus results in a range of developmental phenotypes. The aim of this PhD thesis was to generate new insight into the function of PRC2 in development and, specifically, in PRC2-related overgrowth. Firstly, I explore the function of the relatively poorly characterised PALI-family of PRC2 accessory proteins using mouse and stem cell models of development. Both PALI1 and PALI2 were found to promote PRC2 methyltransferase activity *in vivo*, though *Pali1* knockout mice exhibited a perinatal lethal phenotype while *Pali2* knockout mice were viable and fertile. PALI1 was additionally found to be dispensable for the *de novo* recruitment of PRC2 to target gene promoters. Secondly, using mouse embryonic stem cells (mESCs), I investigate the molecular effects of *EZH2* mutations found in Weaver syndrome – a developmental overgrowth and intellectual disability (OGID) disorder. I show that Weaver-syndrome-associated *EZH2* mutations are predominantly partial loss-of-function but with dominant negative effects on PRC2 methyltransferase activity. Finally, through the comparative analysis of isogenic mESC models of distinct congenital growth disorders, I identify key chromatin and transcriptional changes that are shared across genetically distinct OGID syndromes. Taken together, these data contribute to an improved understanding of the biological functions of the PRC2-associated PALI proteins, and provide important molecular insight into the aetiology of *EZH2*-mutant Weaver syndrome and related congenital growth disorders.



Dedicated to Bidly and Terry.



ACKNOWLEDGEMENTS

I would first like to thank Professor Adrian Bracken for giving me the opportunity to pursue a PhD under his supervision. His mentorship and support allowed me to develop into the scientist I am today. Thank you for providing me with such great training.

I wish to extend thanks to all my colleagues from my placement at the RIKEN Center for Integrative Medical Sciences: いろいろありがとうございました! I am grateful to Haruhiko Koseki for the opportunity, and for collaborating with us to generate invaluable cell lines and mouse strains used in this project. Also, to Takashi and Kaori Kondo, from whom I learned so much both in and outside of the lab.

I would like to thank my colleagues at the Transgenics Facility and Comparative Medicine Unit at Trinity College Dublin for their assistance with the mouse work related to this PhD. I would also like to thank Craig Monger (TCD) for his bioinformatic expertise, and without whom much of the work in this thesis would not be possible. I am grateful to Professors Kristian Helin (ICR) and Danny Reinberg (NYU) for sharing cell lines that contributed to the work in this thesis. I am also very thankful for helpful discussions had with Professor Kate Tatton Brown relating to the work in this thesis (St George's, University of London). This PhD work would not have been possible without the funding provided by the Irish Research Council (GOIPG/2017/2009) and Science Foundation Ireland.

I owe so much of who I am today to my time spent in the Bracken lab – thank you to all members past and present. I would like to personally thank Eric Conway for giving me such a good start at the bench under his supervision, for sharing his work on the PALI project with me, and for ongoing valued collaboration. Thank you also to Ellen Tuck for providing invaluable assistance and feedback on various projects – not to mention encouragement! Of course, a special shout-out goes to Evan Healy, Eleanor Glancy, Rachel McCole and Marlena Mucha for their help and friendship since day 1 of this PhD.

In addition, I would like to thank my parents, Anne and Ken, for their unwavering love and support through everything. To my sister, Mairéad, I am so appreciative of your advice and guidance, and I thank you for helping me to grow as a person over the last few years.

To Alfie, Evie, Zen and Boo: thanks for the emotional support! 🍪🍪

Finally, I truly could not have completed this PhD without my rock, Jeffrey O'Callaghan. Thank you, for everything.

PUBLICATIONS

Directly related to this work

Deevy O. and Bracken A. P. PRC2 functions in development and congenital disorders. *Development*. 146, dev181354. (2019)

Conway E., Jerman E., Healy E., Ito S., Holoch D., Oliviero G., Deevy O., [...] Koseki H. and Bracken A.P. A Family of Vertebrate-Specific Polycombs Encoded by the LCOR / LCORL Genes Balance PRC2 Subtype Activities, *Molecular Cell*. 70(3), 408-421. (2018)

Other

Brien G.L., Bardini Bressan R., Monger C., Gannon D., Lagan E., Doherty A.M., Healy E., Neikes H., Fitzpatrick D.J., Deevy O., [...] Pollard S.M., Bracken A.P. Simultaneous disruption of PRC2 and enhancer function underlies H3.3-K27M oncogenic activity in human hindbrain neural stem cells. *Nature Genetics* (2021)

Qadeer Z. A., Valle-Garcia D., Hasson D., Sun, Z., Cook A., Nguyen C., Soriano A., Ma A., Griffiths L. M., Zeineldin M., Filipescu D., Jubierre L., Chowdhury A., Deevy O., [...] Dyer M.A., Bernstein E. ATRX In-Frame Fusion Neuroblastoma Is Sensitive to EZH2 Inhibition via Modulation of Neuronal Gene Signatures. *Cancer Cell*. 36, 512-527.e9. (2019)

TABLE OF CONTENTS

Declaration	i
Summary	iii
Acknowledgements	vii
Publications	ix
Table of Contents	xi
List of Figures	xvi
List of Tables	xix
List of Abbreviations	xx
Chapter 1: Introduction	1
1.1 Cellular identity and gene expression	3
1.2 Covalent modifications of DNA.....	5
1.3 Chromatin structure.....	6
1.4 Histone post-translational modifications.....	11
1.5 Histone reader, writer, and eraser proteins	13
1.6 Polycomb group (PcG) proteins	14
1.7 PcG proteins form large Polycomb Repressive Complexes (PRCs)	15
1.7.1 <i>Polycomb Repressive Complex 2 (PRC2)</i>	15
1.7.2 <i>Polycomb Repressive Complex 1 (PRC1)</i>	18
1.8 The molecular activities of PRC2	19
1.8.1 <i>Recruitment to chromatin</i>	19
1.8.2 <i>Enzymatic activity</i>	20
1.9 PRC2 functions in mouse development	21
1.10 A role for imbalanced crosstalk between PRC2, NSD1 and DNMT3A in human developmental disorders	26
1.10.1 <i>Crosstalk between PRC2, NSD1 and DNMT3A</i>	26
1.10.2 <i>Mutations in core PRC2 components cause overgrowth and intellectual disability</i>	29
1.10.3 <i>Monoallelic loss-of-function mutations in NSD1 cause Sotos syndrome</i>	31

1.10.4 <i>De novo DNMT3A mutations are implicated in Tatton-Brown-Rahman syndrome</i>	31
1.10.5 <i>Imbalanced regulation of PRC2 at intergenic chromatin as a common feature of overgrowth syndromes</i>	32
1.11 Aims of thesis	34
Chapter 2: Materials and Methods	35
2.1 Cell culture methods.....	37
2.1.1 <i>Basic cell culture conditions</i>	37
2.1.2 <i>Generation of knockout cell lines</i>	38
2.1.3 <i>Lentivirus production</i>	39
2.1.4 <i>Generation of stable cell lines by viral transduction</i>	39
2.1.5 <i>Embryoid body differentiation</i>	40
2.1.6 <i>Cell size and competitive cell growth assay</i>	40
2.2 Mouse methods.....	41
2.2.1 <i>Animals</i>	41
2.2.2 <i>Mouse dissection and tissue collection</i>	42
2.2.4 <i>Tissue fixation</i>	42
2.2.4 <i>Paraffin embedding and tissue sectioning</i>	42
2.3 DNA methods.....	43
2.3.1 <i>High-throughput genotyping of genome-edited cell lines</i>	43
2.3.2 <i>Expression vectors</i>	43
2.3.3 <i>Chromatin immunoprecipitation (ChIP)</i>	44
2.3.4 <i>Quantitative chromatin immunoprecipitation relative to a reference exogenous genome (ChIP-Rx)</i>	45
2.4 RNA methods.....	46
2.4.1 <i>RNA preparation and qRT-PCR analysis</i>	46
2.4.2 <i>Next generation sequencing of total RNA</i>	46
2.4.3 <i>Preparation of DIG-labelled ssRNA probes</i>	47
2.4.4 <i>RNA in situ hybridization</i>	47
2.5 Protein methods	48
2.5.1 <i>Preparation of whole cell protein lysates and western blotting</i>	48
2.5.2 <i>Crude cellular fractionation</i>	48
2.5.3 <i>FLAG co-immunoprecipitation</i>	49

2.6 Bioinformatics methods.....	49
2.6.1 Evolutionary conservation analysis.....	49
2.6.2 ChIP-Rx data analysis.....	50
2.6.3 RNA-seq data analysis.....	51
2.6.4 Pathway and Process Enrichment Analysis.....	51
Chapter 3: Analysis of the PRC2-Associated PALI Proteins	53
3.1 Introduction	55
3.2 Results	61
3.2.1 The genomic localization of <i>PAL1</i> could not be determined by standard chromatin immunoprecipitation methods	61
3.2.2 Interrogating the role of <i>PAL1</i> in de novo PRC2 recruitment and activity	67
3.2.3 Analyses of <i>Pali</i> gene expression during mouse development.....	76
3.2.4 <i>Pali1/2</i> double-knockout mice phenocopy the <i>Pali1</i> single-knockout mouse perinatal lethal phenotype, but <i>Pali2</i> single-knockout mice do not...	77
3.2.5 H3K27 methylation states are altered in <i>Pali</i> knockout mice	79
3.3 Discussion.....	81
Chapter 4: Molecular Characterisation of EZH2-Mutant Weaver Syndrome	85
4.1 Introduction	87
4.2 Results	92
4.2.1 Selection of Weaver syndrome-associated <i>hEZH2</i> mutations for cellular disease modelling	92
4.2.2 No Weaver syndrome mutations are total loss of function.....	92
4.2.3 Generation of an <i>Ezh2</i> heterozygous knockout cell line for improved cellular modelling of Weaver syndrome.....	96
4.2.4 Isogenic mESC models of Weaver syndrome suggest varying degrees of dominant negative effects on PRC2 catalytic activity	97
4.2.5 Isogenic mESC models of Weaver syndrome exhibit global H3K27 hyperacetylation	100
4.2.6 Weaver syndrome-associated mutant <i>hEZH2</i> assembles into PRC2 and alters its genome-wide binding profile	101
4.2.7 Weaver syndrome-associated <i>hEZH2</i> mutations can cause divergent aberrations to H3K27me2/3 deposition at Polycomb target promoters.....	103

4.2.8 Cellular models of Weaver syndrome share broad reductions in H3K27me2 and H3K27me3 with concomitant global increases in H3K27ac	105
4.2.9 Global H3K27 hyperacetylation does not correlate with an increased abundance of non-genic RNA transcripts in cellular models of Weaver syndrome.....	109
4.2.10 Isogenic mESC models of Weaver syndrome share a small cohort of differentially expressed genes	112
4.3 Discussion	114

Chapter 5: Dysregulated Histone H3K27 Acetylation Is a Recurrent Feature of Congenital Growth Disorders.....117

5.1 Introduction	119
5.2 Results	123
5.2.1 An isogenic mESC model of EZH2-associated growth restriction displays global H3K27 hyper-trimethylation and hypo-acetylation.....	123
5.2.2 Growth-restriction-associated mutant hEZH2 assembles into PRC2.....	123
5.2.3 Growth-restriction-associated mutant EZH2 drives reciprocal molecular changes compared to Weaver-syndrome-associated mutant EZH2.....	125
5.2.4 Assaying the relative cell size and proliferation rates of wild-type, Weaver syndrome and growth restriction mESC models.....	127
5.2.5 Generation and validation of an isogenic mESC model of Nsd1 deletion mutant Sotos syndrome	134
5.2.6 Cellular models of Sotos syndrome phenocopy the H3K27 hyperacetylation signature observed in Weaver syndrome models.....	135
5.2.7 Conserved and divergent gene expression changes in Weaver syndrome, Sotos syndrome and growth restriction mESC models.....	139
5.2.8 Preliminary characterization of human lymphoblastoid cell lines derived from Weaver and Sotos syndrome-affected individuals.....	141
5.3 Discussion	143

Chapter 6: General Discussion.....147

6.1 Discussion	149
6.1.1 Summary of findings.....	149
6.1.2 PALI1-PRC2	150

6.1.3 Dysregulated H3K27ac is a recurrent feature of congenital growth disorders.....	151
6.1.4 Potential clinical relevance of findings in cellular models of disease	153
6.1.5 Comments on experimental design related to chapters four and five	155
6.1.6 Future directions.....	156
6.2 Conclusions	159
References	161
Appendices	183
Appendix I: Primers and Oligonucleotides	183
I.I Genotyping primers	183
I.II CRISPR gRNAs.....	183
III Primers for amplifying ISH probe template DNA.....	184
I.IV RT-qPCR mRNA primers	184
I.V ChIP-qPCR primers	185
Appendix II: Antibodies	186
II.I Primary antibodies.....	186
II.II Secondary antibodies.....	187
Appendix III: Wild type hEZH2 ORF	188
Appendix IV: Sanger sequencing of mutant hEZH2 constructs.....	189

LIST OF FIGURES

Chapter 1

- Figure 1.1** Cellular identity
- Figure 1.2** DNA methylation
- Figure 1.3** Chromatin structure
- Figure 1.4** Histone post-translational modifications
- Figure 1.5** Readers, writers, and erasers of histone post-translational modifications
- Figure 1.6** The variable compositions of Polycomb repressive complexes
- Figure 1.7** The crosstalk between PRC2, NSD1 and DNMT3A on chromatin

Chapter 3

- Figure 3.1** The relationship between LCOR, PALI1, LCORL and PALI2
- Figure 3.2** PALI1 target genes could not be determined by standard chromatin immunoprecipitation methods
- Figure 3.3** The combined use of long and short crosslinkers did not improve our ability to ChIP HA-tagged PALI1
- Figure 3.4** The PALI1 interaction with PRC2 (PIP) domain could not be detected at PRC2 target sites by ChIP
- Figure 3.5** Overview of experimental design utilizing the inducible wild-type EED rescue (iWT-EEDr) system in mouse embryonic stem cells
- Figure 3.6** Generation and validation of single-cell derived Pali1 homozygous knockout clones in the iWT-EEDr cellular context
- Figure 3.7** Global PRC2-mediated H3K27 methylation is re-established to a lesser degree in the absence of PALI1
- Figure 3.8** The initial nucleation and spreading of PRC2-mediated H3K27me3 is not impaired in the absence of PALI1
- Figure 3.9** Specific loci show evidence of reduced *de novo* PRC2 activity in the absence of PALI1
- Figure 3.10** *De novo* established intergenic H3K27me3 is increased in the absence of PALI1, with concomitant reductions in H3K27me2
- Figure 3.11** Analysis of *Pali* gene expression during mouse development
- Figure 3.12** Preliminary characterization of *Pali* knockout mice

Chapter 4

- Figure 4.1** Selected Weaver syndrome mutations
- Figure 4.2** Weaver-syndrome-associated EZH2 variants are not total loss of function
- Figure 4.3** Generation and validation of single-cell-derived *Ezh2* heterozygous knockout clones
- Figure 4.4** Weaver-syndrome-associated EZH2 variants can exert dominant negative effects
- Figure 4.5** Isogenic mESC models expressing Weaver-syndrome-associated EZH2 variants exhibit global H3K27 hyper-acetylation
- Figure 4.6** Weaver-syndrome-associated EZH2 variants assemble into PRC2 and alter its genome-wide binding profile
- Figure 4.7** Weaver-syndrome-associated EZH2 variants can cause divergent aberrations to H3K27me3 and H3K27me2 deposition at Polycomb target promoters
- Figure 4.8** Cellular models of Weaver syndrome share broad reductions in H3K27me3 and H3K27me2 outside of Polycomb target promoters
- Figure 4.9** Cellular models of Weaver syndrome share broad increases in H3K27ac
- Figure 4.10** In cellular models of Weaver syndrome, H3K27 acetylation increases where H3K27 methylation is lost
- Figure 4.11** Global H3K27 hyperacetylation does not correlate with an increased abundance of non-genic RNA transcripts in cellular models of Weaver syndrome
- Figure 4.12** Cellular models of Weaver syndrome share a small cohort of differentially expressed genes

Chapter 5

- Figure 5.1** An isogenic mESC model of *EZH2*-associated growth restriction displays global H3K27 hyper-trimethylation and hypo-acetylation
- Figure 5.2** The growth-restriction-associated A733T-EZH2 variant assembles into PRC2 and binds its target sites
- Figure 5.3** The growth-restriction-associated A733T-EZH2 variant causes a diffuse genome-wide increase in H3K27me3 with concomitant global reductions in H3K27me2
- Figure 5.4** Diffuse H3K27ac is lost from the cellular model of EZH2-associated growth restriction
- Figure 5.5** Reciprocal H3K27 modification changes in cellular models of *EZH2*-associated growth restriction versus Weaver syndrome
- Figure 5.6** Flow-cytometry-based method for assessing relative cell size and proliferation rate
- Figure 5.7** Analysis of relative cell size
- Figure 5.8** Analysis of relative cell proliferation rates
- Figure 5.9** Generation and validation of single-cell-derived *Nsd1* heterozygous knockout clones
- Figure 5.10** Isogenic models of *Nsd1* +/- Sotos syndrome exhibit a specific global increase in H3K27ac
- Figure 5.11** Conserved and divergent gene expression changes in isogenic cellular models of Weaver syndrome, Sotos syndrome and EZH2-associated growth restriction
- Figure 5.12** Preliminary characterisation of human lymphoblastoid cell lines derived from Weaver and Sotos syndrome-affected individuals

LIST OF TABLES

Table 1.1	Constitutive loss-of-function phenotypes of PRC2 and related histone/DNA methyltransferase mouse mutants
Table 1.2	Constitutive heterozygous mutations in PRC2 and related histone/DNA methyltransferases implicated in human congenital growth disorders
Table 2.1	A/B cell combinations used in flow-cytometry-based assay of relative cell size and proliferation rate
Table 2.2	Mouse RNA-seq datasets obtained from ENCODE
Table 4.1	Weaver-syndrome-associated EZH2 mutations included in this study
Table 4.2	Differential transcript abundance analysis

LIST OF ABBREVIATIONS

AEBP2 – AE binding protein

ALL – Acute Lymphoblastic Leukaemia

BCIP – 5-bromo-4-chloro-3'-indolylphosphate

BMI1 – B Lymphoma Mo-MLV Insertion Region 1 Homolog

BSA – Bovine Serum Albumin

C10ORF12 – Chromosome 10 open reading frame 12

CBX – Chromobox

cDNA – complementary DNA

CGI – CpG island

ChIP – Chromatin Immunoprecipitation

COGIS – Cohen-Gibson syndrome

CpG – Cytosine-phosphate-Guanine

cPRC1 – canonical Polycomb Repressive Complex 1

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

CtBP – C-terminal-binding protein

d.p.c. – Days Post Coitum

d.p.p. – Days Post-Partum

DIG – Digoxigenin

dKO – double knockout

DMEM – Dulbecco's Modified Eagle Media

DNA – Deoxyribonucleic Acid

DNMT – DNA (cytosine-5)- methyltransferase

DSG – Disuccinimidyl Glutarate

DSS – Disuccinimidyl Suberate

DTT – Dithiothreitol

E – embryonic day

E3 ligase – Ubiquitin ligase

EB – Embryoid Body

EED – Embryonic Ectoderm Development

EGS – Ethylene Glycol bis(Succinimidyl Succinate)

EPOP – Elongin BC and Polycomb Repressive Complex 2 associated protein

ES / ESC – Embryonic Stem / Embryonic Stem Cell

EtOH – Ethanol

EZH1/2 – Enhancer of Zeste 1/2

FBS – Foetal Bovine Serum

G9a – Euchromatic histone-lysine N-methyltransferase 2 (a.k.a. EHMT2)

gDNA – genomic DNA

GFP – Green Fluorescent Protein

GMEM – Glasgow Minimum Essential Medium

gRNA – guide RNA

GSK – Glycogen Synthase Kinase

H2A – Histone H2A

H2B – Histone H2B

H3 – Histone H3

H4 – Histone H4

HAT – Histone Acetyltransferase

HDAC – Histone Deacetylase

HEK293T – Human Embryonic Kidney 293T cells (with SV40 T antigen)

Het – Heterozygote

HMT – Histone methyltransferase

Hox genes – Homeobox genes

HP1 – Heterochromatin Protein 1

HTH – Helix-Turn-Helix domain

IHC – Immunohistochemistry

IMMAS – Imagawa-Matsumoto syndrome

int – intergenic

IP – immunoprecipitate / immunoprecipitation

iPSC – induced Pluripotent Stem Cell

ISH – *In Situ* Hybridization

iWT-EEDr – inducible wild-type EED rescue

JARID2 – Jumonji and AT-Rich Interaction Domain Containing 2

JMJD2A – Jumonji Domain-Containing Protein 2A

KO – Knockout (homozygote)
LCL – Lymphoblastoid Cell Line
LCOR – Ligand-dependent Co-Repressor
LIF – Leukaemia Inhibitory Factor
me(1/2/3) – (mono-/di-/tri-) methyl
MEIS2 – Meis homeobox 2
MEK – MAPK/ERK Kinase
MeOH – Methanol
mESC – mouse Embryonic Stem Cell
MNase – Micrococcal Nuclease
mRNA – messenger RNA
mTOR – mammalian Target Of Rapamycin
NBT – Nitro blue tetrazolium chloride
NHL – non-Hodgkin lymphoma
NSD – nuclear receptor-binding SET domain
OGID – Overgrowth and Intellectual Disability
ORF – Open Reading Frame
P – Postnatal day
PAL1/2 – Polycomb Associated Lcor Isoform 1/2
PBS – Phosphate Buffered Saline
PcG – Polycomb Group
PCGF – Polycomb Group Ring Finger
PCL – Polycomb-Like
PCR – Polymerase Chain Reaction
PHD – Plant Homeodomain (zinc-finger-like motif)
PHF1 – PHD Finger Protein 1
PHF19 – PHD Finger Protein 19
PI3K – Phosphoinositide 3-kinase
PIP – PAL1 interaction with PRC2
PRC1/2 – Polycomb Repressive Complex 1/2
prom – promoter
PTM – Post-Translational Modifications

PWWP – domain with conserved Pro-Trp-Trp-Pro motif

qPCR – Real-Time Quantitative PCR

qRT-PCR - Quantitative Reverse Transcription PCR

RCF – Relative Centrifugal Force

RFP – Red Fluorescent Protein

RING – RING finger domain; Really Interesting New Gene

RNA – Ribonucleic Acid

RPM – Revolutions Per Minute

rRNA – ribosomal RNA

RSTS – Rubinstein-Taybi syndrome

RT-PCR - Reverse Transcription PCR

RYBP – RING1 and YY1 binding protein

SANT – SWI3, ADA2, N-CoR, and TFIIIB 1

SDS-PAGE – Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SET – Su(var)3-9, Enhancer-of- zeste and Trithorax

sKO – single knockout

SS – Sotos syndrome

SSC – Saline-Sodium Citrate (hybridization buffer)

ssRNA – single-stranded RNA

STORM – Stochastic Optical Reconstruction Microscopy

T – *Brachyury*

TBRS – Tatton-Brown-Rahman syndrome

TBS – Tris-buffered Saline

tRNA – transfer RNA

TSS – Transcription Start Site

WS – Weaver syndrome

WT – Wild Type

2i – two small molecular inhibitors (of MEK1 and GSK3)

5-mC – 5-methylcytosine

CHAPTER 1: INTRODUCTION

1.1 Cellular identity and gene expression

Every multicellular organism begins life as a single ‘totipotent’ cell, so-called because it has the potential to create each of the many functionally diverse cell types of the developing and adult organism (Fig. 1.1A). Somatic cell nuclear transfer experiments taught us that the DNA content of a fully-developed adult cell is the same as that of a totipotent cell (Gurdon et al., 1975; Wilmut et al., 1997), so how can such cellular specification be achieved from an individual genetic code? The answer lies in differential gene expression patterns; different genes are turned on (expressed) or off (repressed) in different cell types, and each lineage-specific combination of gene expression (transcriptional programme) confers a unique cellular identity. In a totipotent cell all genes have the potential to be transcriptionally expressed or repressed at a future point in time and thus all cell fates remain possible. During development, specific cell types arise through a process known as differentiation in which the once totipotent cell becomes progressively restricted in its cellular potential; genes too become limited in their future capacity to be expressed or repressed. This process was famously depicted by Conrad Waddington as a pebble rolling down a hill, following a set of existing paths in the landscape to eventually settle in one of several possible valleys representing discrete cell fates (Moris et al., 2016; Waddington, 1942). Using this metaphor, cells lose ‘potency’ as they move down the hill: pluripotent cells can be directed towards most but not all cell types and multipotent cells can be directed towards several related cell types, whereas terminally differentiated cells are stuck at the bottom of the hill and thereby fixed in their cellular identity (Fig. 1.1B). Stem cells are considered those that are either totipotent, pluripotent, or multipotent. Stem cells additionally have the ability to self-renew, i.e., to undergo cellular division without differentiation (Fig 1.1A).

The coordination of cellular differentiation and gene expression is a highly complex and robust – albeit inherently stochastic – process shaped by both cell-extrinsic and cell-intrinsic factors (Moris et al., 2016). The fidelity of this process is crucial for the correct development of any multicellular organism, but also for the maintenance of a healthy adult state; cancers are essentially the result of this process gone awry. For accurate cell fate specification to occur, molecular mechanisms of ‘transcriptional memory’ are required to facilitate the propagation of lineage-specific transcriptional programmes through successive cell divisions (Stewart-Morgan et al., 2020). These molecular mechanisms are generally termed ‘epigenetic’, and examples include the covalent modification of DNA, the regulation of chromatin structure, and the post-translational modification of histone proteins.

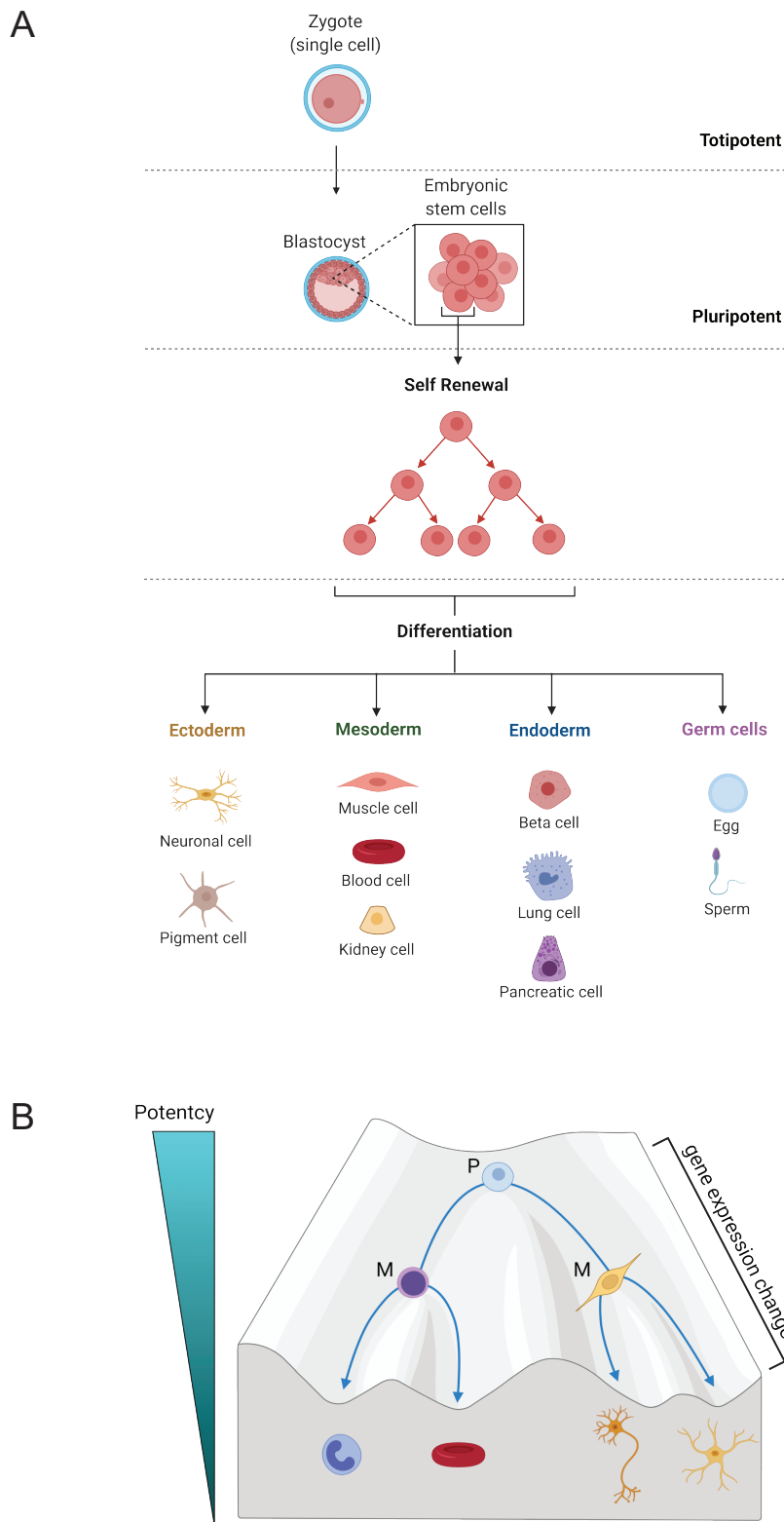


Figure 1.1 Cellular identity

(A) A schematic illustrating the diversity of cell types that can arise from a single cell. In this example, the totipotent zygote develops into the blastocyst, from which all embryonic and extra-embryonic cell types are derived. The embryonic stem cells are pluripotent, capable of both self-renewal and differentiation into the many different cell types of the developing and adult organism. Adapted from “Human Embryonic Stem Cell Differentiation”, by BioRender.com (2021). Retrieved from <https://app.biorender.com/biorender-templates>.

(B) A schematic representation of cellular differentiation based on Waddington’s landscape. In this example, the pluripotent (P) stem cell can be directed toward alternative multipotent (M) cell fates: haematopoietic (left) or neural (right) stem cells. These multipotent cells can in turn be directed towards lineage-specific mature cell types. The changes in cellular identity are driven by changes in gene expression. Created with BioRender.com.

1.2 Covalent modifications of DNA

DNA can be covalently modified by the addition of a methyl group to the fifth carbon of cytosine nucleotides (5-methylcytosine (5-mC)) (Fig. 1.2). In vertebrates, 5-mC is usually found in the context of CpG dinucleotides that are almost always symmetrically modified on both strands of the DNA (Doskočil and Šorm, 1962; Feng et al., 2010; Zemach et al., 2010). The symmetric modification of CpG dinucleotides is an exemplar epigenetic mechanism: after DNA replication, which precedes cell division, the hemimethylated parental DNA strand acts as a template for DNA methyltransferases (DNMTs) to faithfully copy the modification onto the newly synthesised daughter strand (Li and Zhang, 2014).

DNA methylation is widespread throughout mammalian genomes, with 5mC found on up to 80% of CpG dinucleotides. Having said that, the mammalian genome is generally CpG poor; 5-mC is inherently mutagenic (leading to C-to-T transitions) and as a result, CpG dinucleotides are globally depleted from the genome (Bird and Taggart, 1980; Cooper and Krawczak, 1989; Holliday and Grigg, 1993). CpG islands (CGIs) are the exception to this rule: CGIs are short (~1-2 kb) regions of the genome that are enriched for CpG dinucleotides but almost never methylated. Approximately two-thirds of mammalian promoters are associated with CGIs, which maintain their unmethylated state regardless of the transcriptional status of the associated gene (Hughes et al., 2020). The CGI promoters of tumour suppressor genes, however, often become aberrantly methylated in cancer (Baylin and Jones, 2016).

DNA methylation is generally considered to promote transcriptional repression. Accordingly, it is associated with heterochromatin formation, X-chromosome inactivation, and the repression of transposons and germline-specific genes (Greenberg and Bourc'his, 2019). However, 5-mC is also enriched along the bodies of actively transcribed genes (Feng et al., 2010; Zemach et al., 2010). Thus, DNA methylation likely performs distinct context-dependent functions that rely on its interplay with other chromatin regulatory processes.

5mC is deposited by two different classes of DNMT, divided according to whether they establish or maintain DNA methylation. In mammals, there are two major *de novo* DNA methylation enzymes which may catalyse 5-mC in any sequence context – DNMT3A and DNMT3B (Fig. 1.2). A catalytically inactive DNMT, DNMT3L, also exists and appears to function as a positive regulator of DNMT3A and DNMT3B specifically in the germline (Bourc'his et al., 2001; Ooi et al., 2007). DNMT1 is the enzyme responsible for the maintenance of symmetric CpG methylation after cell division (Fig. 1.2), though it requires the E3 ligase UHRF1 to perform this function (Sharif et al., 2007). The process of DNA

methylation is essential, as exemplified by the severe developmental abnormalities and invariably lethal phenotypes of DNMT-deficient mice (Greenberg and Bourc'h, 2019).

DNA demethylation occurs through both passive and active mechanisms (Fig 1.2). Passive DNA demethylation occurs through the dilution of 5-mC during successive rounds of replication in the absence of DNMT1 activity (Li and Zhang, 2014). Active DNA demethylation is carried out by the ten-eleven translocation (TET) family of methylcytosine dioxygenases (TET1, TET2 and TET3). TET enzymes progressively oxidise 5mC to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) (Kohli and Zhang, 2013). None of these oxidised forms stimulate the maintenance methyltransferase activity of DNMT1, therefore they all contribute to DNA demethylation during replication (Greenberg and Bourc'h, 2019). Additionally, thymine DNA glycosylase (TDG) can directly act on 5-fC and 5-caC to promote demethylation (Kohli and Zhang, 2013).

Mammals undergo two striking waves of genome-wide DNA methylation erasure and establishment during development – first following fertilization and then after germline cell specification (Greenberg and Bourc'h, 2019). Intriguingly, approximately twenty genomic regions exist in humans and mice that withstand this global reprogramming and force mono-allelic expression of the ~150 neighbouring genes, many of which are essential for foetal and placental growth and development (Greenberg and Bourc'h, 2019; Hanna and Kelsey, 2021). These special 'imprinting control regions' exhibit allele specific methylation because they were differentially methylated in the parental germlines; the maternally inherited allele is most often methylated. It is hypothesised that maternal imprints function to suppress foetal growth and thereby conserve the dam's resources during and immediately after pregnancy, whereas paternal imprints are thought to promote growth (Moore and Haig, 1991). Indeed, several imprinted genes have been identified that fit this model (Hanna and Kelsey, 2021).

1.3 Chromatin structure

In humans, each diploid cell contains about six billion base pairs of DNA which, if extended, would measure over two metres (Piovesan et al., 2019). To put that another way, the DNA from a single cell is taller than the average person (made up of trillions of such cells). The nuclei of human cells are approximately 10µm in diameter – how is it possible to pack two metres of DNA into this microscopic space? Eukaryotes do it by wrapping their DNA around special proteins called histones. Histone proteins are positively charged while DNA is negatively charged, so electrostatic attractions keep them tightly bound together in a

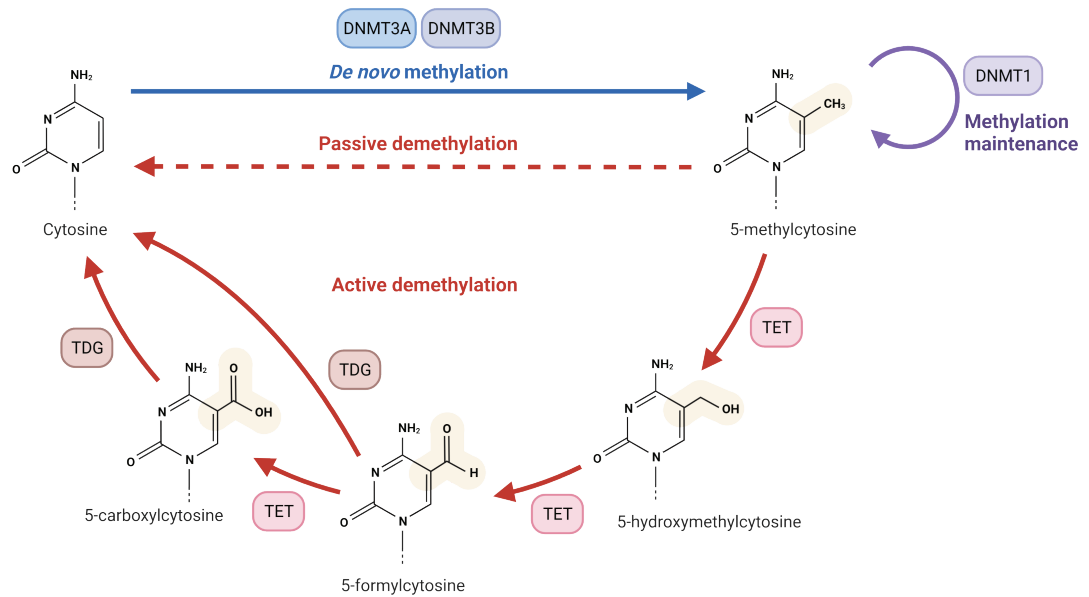


Figure 1.2 DNA Methylation

Schematic outlining the processes of DNA methylation and demethylation in mammals. DNMT = DNA methyltransferase. TET = ten-eleven translocation methylcytosine dioxygenase. TDG = thymine DNA glycosylase. Reprinted from "DNA Methylation", by BioRender.com (2021). Retrieved from <https://app.biorender.com/biorender-templates>.

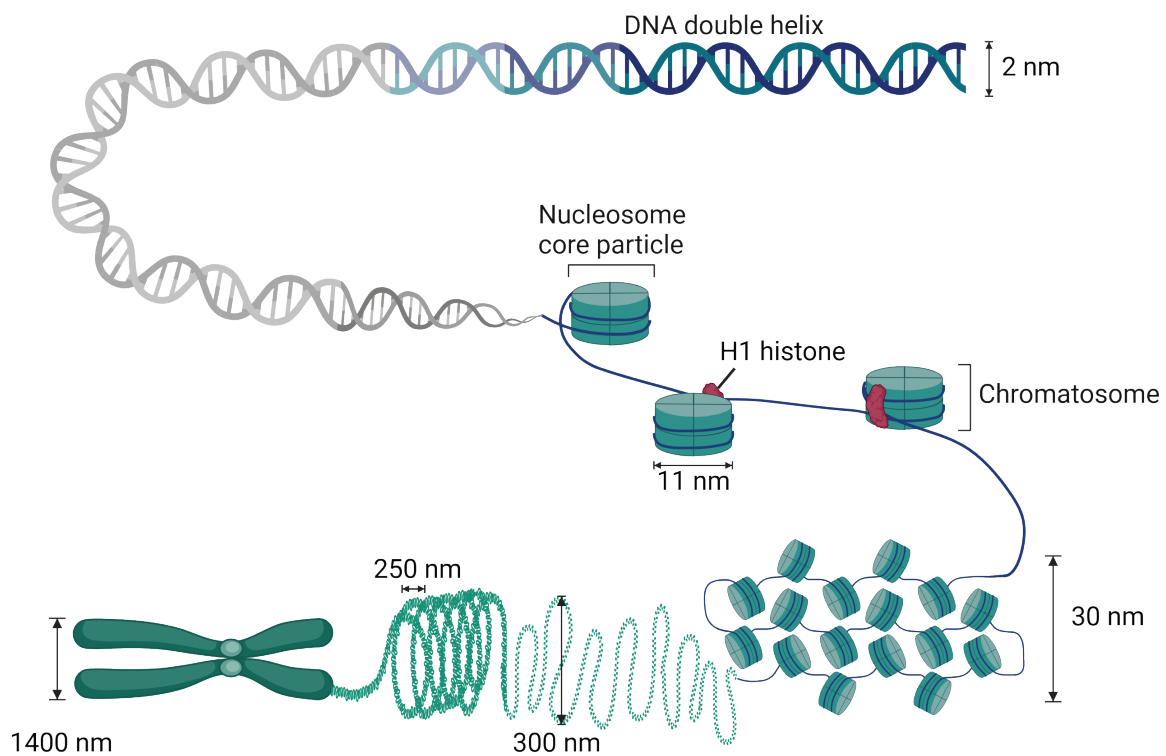


Figure 1.3 Chromatin structure

Schematic representation of chromatin structure. Negatively charged DNA is wrapped around positively charged core histones to form the nucleosome core particle. A chromatosome consists of a nucleosome plus the H1 linker histone. Nucleosomes fold together to produce 30 nm fibres that themselves coil and form loops averaging 300 nm in length. These loops are further compressed and folded to produce 250-nm-wide fibres which tightly coil to make up the chromatid of a chromosome. Created with BioRender.com.

complex known as chromatin. The basic subunit of chromatin is the nucleosome, also known as the nucleosome core particle (Fig 1.3). Nucleosomes consist of ~146 base pairs of DNA wrapped ~1.7 times around an octamer containing two copies of each of the core histone proteins – histone 2A (H2A), histone 2B (H2B), histone 3 (H3) and histone 4 (H4). The N-terminals of the core histone proteins protrude from the nucleosome as histone ‘tails’ (Luger et al., 1997). The C-terminal tail of histone H2A also protrudes from the nucleosome core particle (Vogler et al., 2010). The addition of the linker histone, H1, allows ~166 base pairs of DNA wrap twice around the core octamer to form a more compact and rigid structure called a chromatosome (Fig. 1.3) (Bednar et al., 2017). H1 interacts with DNA where it enters and exits the nucleosome core particle, protects the internucleosomal ‘linker’ DNA (~20bp), and plays important roles in the regulation of chromatin condensation (Brockers and Schneider, 2019). Chromosomes are made up of hundreds of thousands of nucleosomes linked together and folded tightly to produce ~30nm fibres which themselves loop and coil to generate higher order compacted structures and, eventually, a chromatid (Fig. 1.3) (Woodcock, 2005).

Despite this high level of compaction, chromatin remains a dynamic macromolecule as DNA needs to be accessible to transcriptional and other cellular machinery. Chromatin dynamics are regulated by ATP-dependent chromatin-remodelling complexes that can change nucleosome composition, packaging and positioning to facilitate DNA accessibility (Clapier et al., 2017). Four subfamilies of such remodeller complexes exist: imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), switch/sucrose non-fermentable (SWI/SNF) and INO80. Orphan remodellers also exist, so-called because they do not belong to any particular subfamily – examples include α -thalassaemia/mental retardation syndrome X-linked (ATRX) and Cockayne syndrome group B (CSB).

Chromatin structure is closely linked to gene expression. Chromatin that is less compact or ‘open’ (called euchromatin) is readily accessible to transcriptional machinery and so is typically associated with active gene expression. Conversely, tightly condensed chromatin (called heterochromatin) is less transcriptionally permissive and therefore associated with gene repression. An excellent example of this can be seen in X-chromosome inactivation. In female cells dosage compensation requires that all but one copy of the X-chromosome be transcriptionally silenced, and the inactive X-chromosomes are silenced through a process of heterochromatinization (Brockdorff et al., 2020). Chromatin structure is thus naturally also linked to cellular identity – consider for example that, for more than a century, abnormal nuclear morphology has been used as an essential diagnostic feature of cancerous cells (Fischer, 2020). Accordingly, mutations in histone proteins and their remodellers are

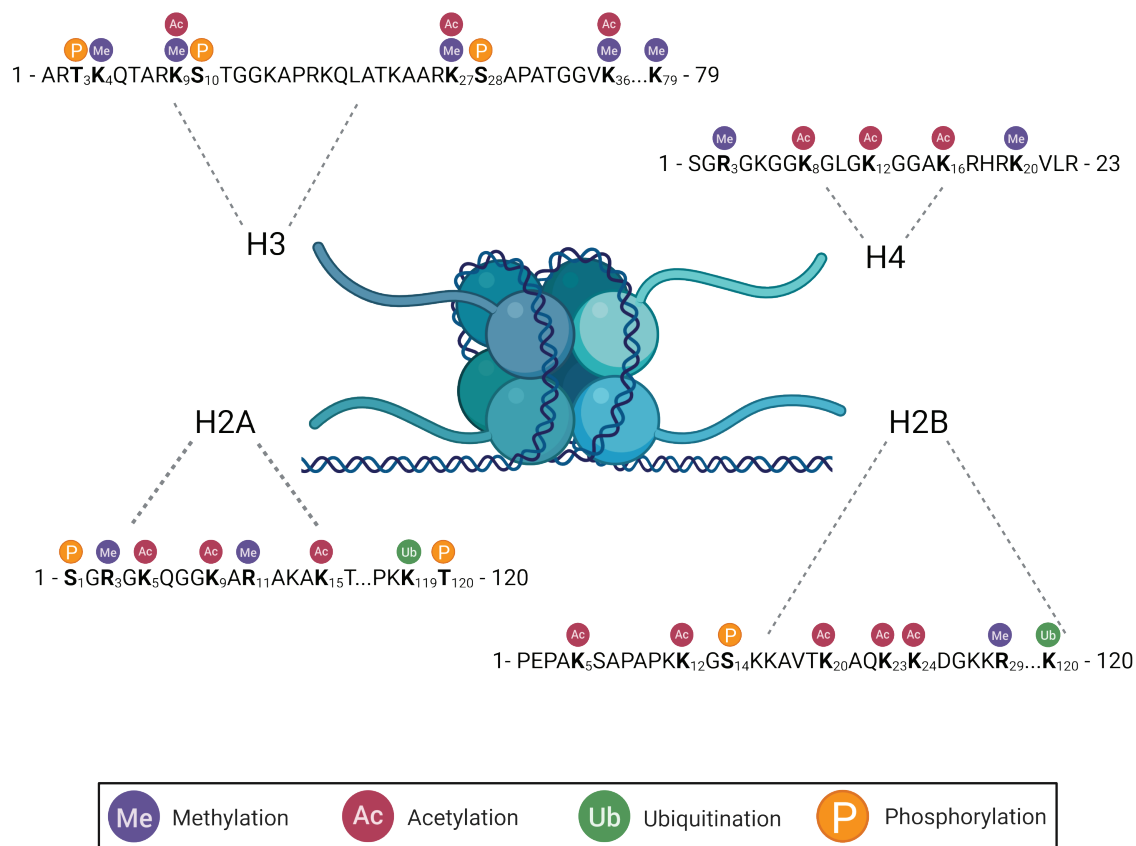


Figure 1.4 Histone post-translational modifications

Schematic of the nucleosome core particle (centre). The nucleosome core particle consists of ~146 base pairs of DNA wrapped around an octamer containing two copies of each of the core histone proteins: histone 2A (H2A), histone 2B (H2B), histone 3 (H3) and histone 4 (H4). The N-terminals of the core histone proteins protrude from the nucleosome as so-called 'tails'. The residues of these histone tails are subject to a wide variety of post-translational modifications, including methylation, acetylation, ubiquitination and phosphorylation. These modifications can exert direct or indirect effects on chromatin accessibility and transcription. Some example histone PTMs are depicted here. Created with BioRender.com.

known drivers of malignant transformation (Amatori et al., 2021; Bagert et al., 2021; Vogelstein et al., 2013).

1.4 Histone post-translational modifications

Changes to chromatin structure can be mediated by covalent modifications made to histone proteins. Histones are subject to a great variety of post-translational modifications (PTMs) including, but not limited to, methylation, acetylation, phosphorylation and ubiquitination (Fig 1.4) (Soshnev et al., 2016). These PTMs are enriched and have been best studied on the protruding ‘tails’ of the four core histones, but the H1 linker histone is also extensively post-translationally modified (Prendergast and Reinberg, 2021).

At the turn of the millennium, it was proposed that histone PTMs function in concert together as fundamental regulators of chromatin-structure and gene expression – the so-called “histone code hypothesis” (Jenuwein and Allis, 2001). Since then, much work has been focussed on elucidating the molecular mechanisms through which this may be achieved. In certain cases, the relationship between a histone PTM and chromatin structure/gene expression is direct. For example, the acetylation of histone lysine (K) residues neutralises their positive charge and thereby loosens the local interactions between histones and DNA. Accordingly, acetylated histone residues are enriched at the promoters of transcriptionally active genes – though, interestingly, a recent study suggests that active transcription often precedes histone acetylation (Martin et al., 2021). More often, histone PTMs influence chromatin structure and gene expression through less direct mechanisms. For example, the methylation of histone residues can either positively or negatively regulate transcription; mono- or trimethylation of histone H3 lysine 9 (H3K9me1) or 36 (H3K36me3), respectively, is associated with active transcription whereas trimethylation of histone H3 lysine 9 (H3K9me3) or 27 (H3K27me3) is considered repressive (Barski et al., 2007). In these cases, histone modifications elicit their regulatory effects via crosstalk with chromatin regulatory proteins, or ‘effector’ proteins – as will be discussed in section 1.5.

As with DNA methylation, cellular mechanisms are in place to facilitate the mitotic inheritance of various histone PTMs (Escobar et al., 2021; Stewart-Morgan et al., 2020). For example, analogous to the maintenance of DNA methylation by UHRF1 and DNMT1, ‘read-write’ mechanisms have been defined for H3K9me3 and H3K27me3 to propagate their associated chromatin states post-replication (Coleman and Struhl, 2017; Gaydos et al., 2014; Laprell et al., 2017; Margueron et al., 2009; Müller et al., 2016; Zhang et al., 2008a).

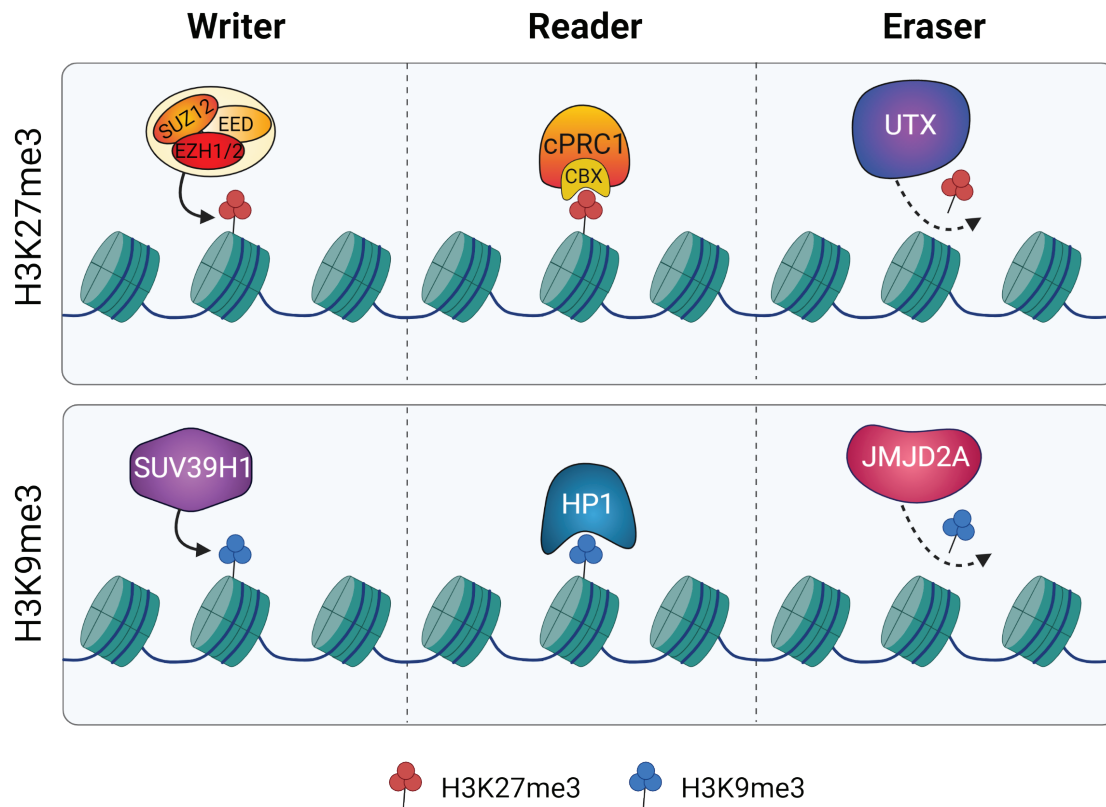


Figure 1.5 Readers, writers and erasers of histone post-translational modifications

Histone post-translational modifications are deposited by catalytic 'writer' enzymes, recognised and bound by 'reader' proteins, and removed by 'eraser' enzymes. Example writer, reader and eraser proteins for H3K27me3 (top) and H3K9me3 (bottom) are shown for illustrative purpose. Created with BioRender.com.

These inheritance mechanisms depend on the actions of so-called histone ‘reader’ and ‘writer’ enzymes, discussed below.

1.5 Histone reader, writer, and eraser proteins

Histone PTMs are deposited by ‘writer’ proteins, recognised and bound by ‘reader’ proteins, and removed by ‘eraser’ proteins (Fig. 1.5). The first discovery of such an enzyme was made in yeast – Kleff et al. (1995) identified a **writer** enzyme, specifically a histone acetyltransferase (HAT). However, the lab of David Allis is widely credited with the first major breakthrough in the field; they identified a HAT in *Tetrahymena* and showed that it was homologous to a known transcriptional regulator in yeast, Gcn5, thereby providing the first direct evidence that histone acetylation was connected to gene expression (Brownell et al., 1996). Importantly, that same year saw the complementary discovery of a mammalian histone acetylation **eraser** enzyme (a histone deacetylase (HDAC)) that was homologous to a known yeast transcriptional repressor, Rpd3p (Taunton et al., 1996). Together, these crucial studies functionally linked histone modifications to transcriptional regulation and marked the beginning of an explosion of research into histone modifications and their associated regulatory proteins.

Histone methyltransferases (HMTs) are another important class of writer enzyme discovered shortly thereafter, and highly relevant to this thesis. The first HMT, SUV39H1, was identified by Rea et al. (2000) as a H3K9 methyltransferase (Fig. 1.5). Soon after, the first histone **reader** protein identified – heterochromatin protein 1 (HP1) – was shown to bind to SUV39H1-mediated H3K9me via a methyl-lysine recognition domain known as a ‘chromodomain’ (Fig. 1.5) (Bannister et al., 2001; Lachner et al., 2001). As its name suggests, HP1 promotes heterochromatin formation (Allshire and Madhani, 2018). In other words, the H3K9me₃ modification elicits its repressive effects via HP1; HP1 is both a reader and effector protein for H3K9me₃. Incidentally, HP1 can dimerize to directly compact chromatin but it can also recruit DNMTs as additional heterochromatin effectors – demonstrating crosstalk between histone modifications and DNA methylation (Fuks et al., 2003; Smallwood et al., 2007). Finally, the first eraser for methylated histone residues, a histone demethylase (HDM), was identified by Shi et al. in 2004: LSD1 (KDM1) selectively demethylates H3K4. It is worth noting that trimethylated H3K4 is enriched at CpG islands (which are characterised by a lack of DNA methylation) and, consistent with this, it repels DNMT3A and DNMT3B – again demonstrating crosstalk between histone modifications and DNA methylation (Greenberg and Bourc’his, 2019; Hughes et al., 2020).

A multitude of reader, writer and eraser enzymes have since been identified and shown to function together to regulate the chromatin landscape (Allis et al., 2015). Multi-protein chromatin regulatory complexes are often collections of different readers, writers and erasers which work together to promote a particular chromatin or transcriptional state. For example, Polycomb Repressive Complex 2 (PRC2), which will be discussed in detail in the coming sections, both writes and reads H3K27me3 to propagate repressive chromatin domains (Margueron et al., 2009). PRC2 can also recruit the NO66 eraser enzyme to demethylate H3K36me3, a marker of active chromatin (Brien et al., 2012).

1.6 Polycomb group (PcG) proteins

Polycomb group (PcG) proteins are key chromatin-based regulators of gene expression that include both histone readers and writers (Blackledge and Klose, 2021; Laugesen et al., 2016). The first PcG gene, *Polycomb* (*Pc*), was discovered in *Drosophila* by Pamela Lewis (Lewis, 1947). Pamela's husband, Ed Lewis, subsequently determined that PcG genes function as negative regulators of Homeobox (*Hox*) gene expression (Lewis, 1978). *Hox* genes are master regulators for the specification of body plan and thus play a crucial role during embryonic development (Pearson et al., 2005). The mutations in PcG genes in *Drosophila* caused ectopic expression of *Hox* genes that resulted in posterior homeotic transformations, in which anterior body structures displayed features of more posterior structures (Lewis, 1947; Slifer, 1942). This was interpreted as a failure of the PcG proteins to appropriately repress *Hox* genes, leading to dramatic alterations of cell fate (Ringrose and Paro, 2004). Orthologous PcG genes were later identified in mammals whose products also appeared to repress *Hox* gene expression during development (Alkema et al. 1995; Akasaka et al. 1996). PcG proteins are now known to be widely conserved across eukaryotes and to play a broad role in the repression of genes that regulate cellular identity and lineage-specification, in pluripotent and differentiated cells alike (Boyer et al., 2006; Bracken, 2006; Gahan et al., 2020; Lee et al., 2006; Schuettengruber et al., 2017).

It is pertinent to mention that *Trithorax* group (TrxG) proteins – also discovered in *Drosophila* – antagonise PcG function; TrxG proteins maintain active gene expression states (Piunti and Shilatifard, 2016; Schuettengruber et al., 2017). Accordingly, mutations in TrxG genes cause anterior homeotic transformation phenotypes (Ingham, 1983, 1985; Kennison and Tamkun, 1988; Struhl and Akam, 1985). Together, PcG and TrxG proteins function to maintain appropriate gene expression profiles and thereby safeguard cellular identity during development and differentiation (Cenik and Shilatifard, 2021; Piunti and Shilatifard, 2021; Schuettengruber et al., 2017).

1.7 PcG proteins form large Polycomb Repressive Complexes (PRCs)

PcG proteins function as part of multi-subunit complexes termed Polycomb repressive complexes (PRCs) (Di Croce & Helin 2013; Blackledge et al. 2015). Two major PRCs have been identified in mammals, PRC1 and PRC2, as discussed below (Fig. 1.6).

1.7.1 Polycomb Repressive Complex 2 (PRC2)

Mammalian PRC2 consists of 3 core subunits (Fig. 1.6): SUZ12, EED and either the EZH2 or EZH1 histone methyltransferase (HMT) (Laugesen et al., 2019; Yu et al., 2019). These core PRC2 proteins associate in a 1:1:1 stoichiometry and catalyse all mono-, di- and tri-methylation of histone H3 lysine 27 (H3K27) through the SET domain of the EZH1/2 subunit (Højfeldt et al., 2018; Smits et al., 2013). This trimeric core of constitutive PRC2 components can associate with an expanding list of facultative or ‘accessory’ PRC2 components, including AEBP2, JARID2, PCL1/2/3 (PHF1/MTF2/PHF19), RBBP4/7 and the more recently identified EPOP, PALI and EZHIP proteins (Fig. 1.6) (Conway et al., 2018; Holoch and Margueron, 2017; Hübner et al., 2019; Jain et al., 2019; Ragazzini et al., 2019). Whilst these accessory components are not strictly essential for the formation of core PRC2, we are continuing to discover how they function to modulate its recruitment and enzymatic activity (Bracken et al., 2019; Laugesen et al., 2019).

Mammalian PRC2 exists in two mutually exclusive subtype assemblies – PRC2.1 and PRC2.2 – that are defined by their associations with specific accessory subunits (Fig. 1.6) (Alekseyenko et al., 2014; Grijsenhout et al., 2016; Hauri et al., 2016; Oliviero et al., 2016; Streubel et al., 2018). PRC2.1 contains one of three paralogous Polycomb-like proteins (PHF1/PCL1, MTF2/PCL2 or PHF19/PCL3) as well as either the EPOP or PALI1 proteins (Beringer et al., 2016; Conway et al., 2018; Liefke et al., 2016), whereas PRC2.2 is defined by the presence of the JARID2 and AEBP2 accessory proteins (Holoch and Margueron, 2017). *PALI2* is a paralogue of *PALI1* whose protein product is predicted to interact with PRC2 through a conserved “PIP” domain (Conway et al., 2018). Like RBBP4/7, the most recently described accessory subunit, EZHIP, has been reported to interact with both PRC2.1 and PRC2.2, although it is not as widely expressed (Hübner et al., 2019; Jain et al., 2019; Piunti et al., 2019; Ragazzini et al., 2019). The discovery of these variant PRC2 assemblies poses a question that remains to be answered – what are their distinct functions, if any? Towards this, a recent study suggests that PRC2.1 and PRC2.2

regulate distinct gene sets during mouse embryonic stem cell differentiation (Petracovici and Bonasio, 2021).

It is interesting to note that the subdivision of PRC2 into two main forms, containing either Pcl or Jarid2, is conserved in *Drosophila*, and that a homologue of AEBP2 (Jing) also exists in flies (Bracken et al., 2019; Herz et al., 2012; Nekrasov et al., 2007). It is therefore likely that any functional divergence between the two main PRC2 subtypes precedes the most recent common ancestor of flies and mammals but has been expanded upon in the vertebrate lineage with the emergence of the PALI, EPOP and EZHIP accessory proteins.

The increase in the number of PRC2 proteins in mammals when compared against the fly is due to the occurrence of two whole-genome duplication events at the base of the vertebrate lineage (Dehal and Boore, 2005; Schuettengruber et al., 2017). For example, while *Drosophila* possess a single Polycomb-like protein (Pcl), mammalian genomes encode three such homologues (Brien et al., 2015). Furthermore, the EPOP, EZHIP and PALI1/2 PRC2.1 components do not exist in flies, but are instead specific to therians, eutherians and vertebrates, respectively (Beringer et al., 2016; Conway et al., 2018; Ragazzini et al., 2019). It will be interesting to assess whether the emergence of these additional PRC2 accessory proteins correlates with certain anatomical innovations during vertebrate evolution and/or certain tissue-specific functions not shared with our most recent common ancestor with *Drosophila*.

Should differential functions for PRC2.1 and PRC2.2 be identified, there is evidence to support the convergent evolution of functionally distinct PRC2 subcomplexes in plants. While plants do not possess any known orthologues of the abovementioned PRC2 accessory proteins, they have multiple homologues of the core PRC2 components (Schuettengruber et al., 2017). *Arabidopsis*, for example, possesses three *SUZ12* homologues (*EMF2*, *VRN2*, *FIS2*) which define at least three different PRC2 subcomplexes that are known exhibit to divergent tissue- and/or developmental stage-specific functions (Derkacheva and Hennig, 2014). Interestingly, in mammals SUZ12 is the PRC2 subunit that mediates interactions between core PRC2 and accessory components, and therefore may be considered an independently evolved but analogous determinant of functional PRC2 subcomplex specificity (Laugesen et al., 2019).

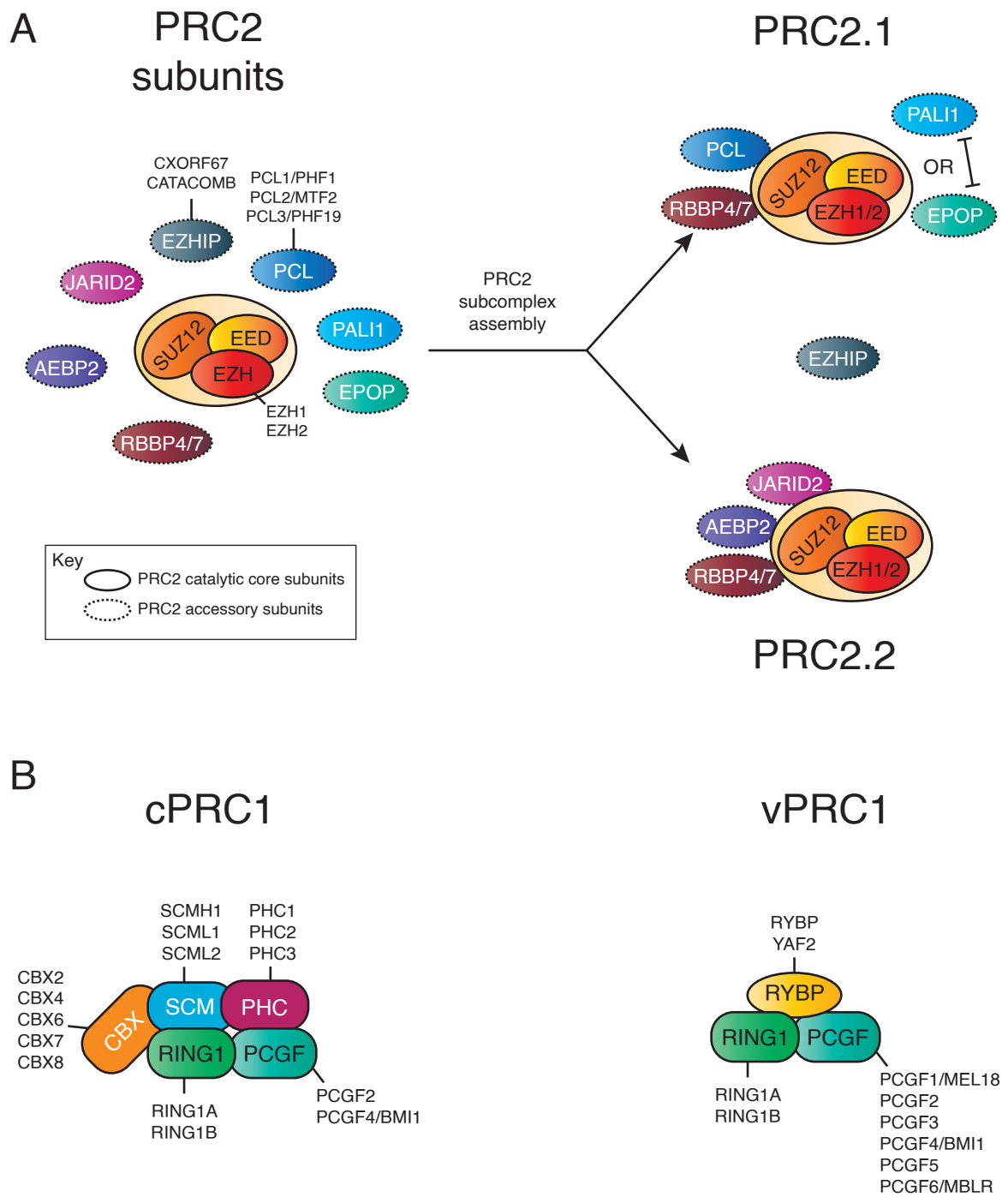


Figure 1.6 The variable compositions of Polycomb repressive complexes

(A) Schematic representation of mammalian Polycomb repressive complex 2 (PRC2), including both core and accessory subunits. The composition of the PRC2.1 and PRC2.2 subcomplexes are also depicted. Note that PALI1 and EPOP are mutually exclusive members of PRC2.1, whereas RBBP4/7 associates with both PRC2.1 and PRC2.2. EZHIP has also been reported to interact with both PRC2.1 and PRC2.2, but is expressed in a limited range of cell types.

(B) Schematic representation of mammalian Polycomb repressive complex 1 (PRC1), distinguishing between canonical PRC1 (cPRC1) versus variant PRC1 (vPRC1).

Paralogous subunits and alternative protein names are listed where applicable. Contacts shown between subunits are merely schematic and are not intended to reflect actual physical interactions.

1.7.2 Polycomb Repressive Complex 1 (PRC1)

Like PRC2, mammalian PRC1 can be subdivided into two main subtypes: canonical or variant PRC1, respectively cPRC1 or vPRC1 (Fig 1.6) (Blackledge and Klose, 2021). All PRC1 assemblies contain an E3 ubiquitin ligase (RING1A or RING1B) and one of six PCGF proteins (PCGF1-6). Canonical PRC1 was the first to be isolated, and is defined by the additional presence a CBX protein (CBX2/4/6/7/8), a PHC protein (PHC1-3) and a SCM protein (SCMH1, SCML1/2) (Blackledge and Klose, 2021). cPRC1 specifically assembles with PCGF2 or PCGF4. Variant PRC1, however, can assemble around any of the six PCGF proteins and lacks the CBX, PHC and SCM subunits but instead contains a RYBP or YAF2 protein (Blackledge et al., 2015; Gao et al., 2012). Different PRC1 assemblies can associate with additional accessory components depending on their PCGF subunit, and differential assembly-specific functions have been identified in some cases (Blackledge and Klose, 2021). For example, PCGF3/5-vPRC1 is involved in X-chromosome inactivation (Almeida et al., 2017; Fursova et al., 2019).

PRC1 functions both up- and downstream of PRC2 activity. Accordingly, PRC1 and PRC2 enrichment on chromatin is largely overlapping (Boyer et al., 2006; Bracken, 2006; Fursova et al., 2019; Schwartz et al., 2006; Tolhuis et al., 2006). vPRC1 functions upstream of PRC2 via recruitment mechanisms that are incompletely understood but include targeting by sequence-specific DNA-binding factors (e.g. MGA-MAX), targeting by long non-coding RNAs (e.g. *XIST*) and targeting by proteins that bind CpG islands (e.g. KDM2B) (Blackledge and Klose, 2021). vPRC1 catalyses the addition of a single ubiquitin moiety on lysine 119 of histone H2A (H2AK119ub) through the RING/PCGF heterodimer (Wang et al., 2004a). H2AK119ub is enriched on the promoters of developmentally repressed genes but found at low levels genome wide (Conway et al., 2021; Fursova et al., 2019; Wang et al., 2004a). vPRC1-mediated H2AK119ub recruits PRC2.2 via interactions with JARID2 and potentially also AEBP2 (Blackledge et al., 2020; Fursova et al., 2019; Healy et al., 2019; Kasinath et al., 2021; Sugishita et al., 2021; Tamburri et al., 2020). cPRC1, on the other hand, is thought to function downstream of PRC2. cPRC1 recognises and binds to PRC2-mediated H3K27me3 via a chromodomain present in its CBX subunit (Fischle et al., 2003; Min et al., 2003; Wang et al., 2004b). cPRC1 does not significantly contribute to H2AK119ub deposition, but it does directly mediate chromatin compaction via oligomerisation of the PHC subunits (Fursova et al., 2019; Isono et al., 2013).

1.8 The molecular activities of PRC2

1.8.1 Recruitment to chromatin

To date, the mechanisms by which PRC2 is recruited to chromatin have been less well defined in mammals compared to *Drosophila*. However, clues as to how mammalian PRC2 is recruited to its target genes on chromatin are emerging through the study of its accessory proteins (Glancy et al., 2021; Laugesen et al., 2019). PRC2 binds to CpG islands at the promoters of inactive developmental genes in mammalian cells (Deaton and Bird, 2011; Ku et al., 2008; Lee et al., 2006; Riising et al., 2014). Its recruitment is facilitated, at least in part, by the Polycomb-like and JARID2 accessory components of the PRC2.1 and PRC2.2 complexes, respectively (Healy et al., 2019; Højfeldt et al., 2019; Oksuz et al., 2018; Youmans et al., 2018). The Polycomb-like proteins contain a winged-helix domain that allows them to bind to unmethylated GC-rich DNA, while JARID2 is believed to recognise vPRC1-mediated H2AK119ub at CpG islands via its ubiquitin interaction motif (UIM) (Blackledge et al., 2014; Choi et al., 2017; Cooper et al., 2016; Farcas et al., 2012; Li et al., 2017; Perino et al., 2018; Wu et al., 2013). Supporting this model, vPRC1-mediated H2AK119ub1 is central to the role of JARID2 in promoting H3K27me3 during X-chromosome inactivation (Almeida et al., 2017; da Rocha et al., 2014). Furthermore, the loss of H2AK119ub1 in mouse embryonic stem cells lacking vPRC1 function leads to a partial reduction in core PRC2 members and H3K27me3 levels on Polycomb target genes (Blackledge et al., 2020; Fursova et al., 2019; Scelfo et al., 2019). Taken together, it appears that PRC2 recruitment is predominantly the product of targeting to unmethylated CpG islands both directly by Polycomb-like proteins and indirectly by JARID2 via its association with ncPRC1-mediated H2AK119ub.

It remains to be fully elucidated what, if any, contributions AEBP2, PALI1, EPOP or EZHIP make to PRC2 recruitment mechanisms. AEBP2 is reportedly capable of targeting PRC2 to methylated CpGs *in vitro* (Wang et al., 2017) – a finding incongruent with several studies reporting an anti-correlation between PRC2 and methyl-CpG DNA genome wide (Bartke et al., 2010; King et al., 2016; Lynch et al., 2011). As such, it remains to be determined whether this methyl-CpG binding preference holds true *in vivo* and, if so, what functional contribution this might make to PRC2 action in cells. In addition, a recent structural study indicates that AEBP2, like JARID2, can bind to monoubiquitinated histone H2A *in vitro* (Kasinath et al., 2021). Thus, future studies dissecting the functional relationship between AEBP2 and H2AK119ub in mouse embryonic stem cells, for example, may shed light on potential new AEBP2-dependent PRC2 recruitment mechanisms.

Similarly, PALI1 has recently been shown to facilitate DNA and nucleosome binding by PRC2 *in vitro* but it is not yet known how this may contribute to PRC2 recruitment in cellular contexts (Zhang et al., 2021a).

1.8.2 Enzymatic activity

PRC2-mediated H3K27 methylation can be catalysed by either EZH1 or EZH2. EZH2 is the more potent H3K27 methyltransferase and can fully compensate for EZH1; *Ezh2*-null cells show global reduction of H3K27me3 and H3K27me2, whereas *Ezh1*-null cells do not exhibit any reduction in H3K27 methylation levels (Table 1.1). Nevertheless, EZH1 can maintain normal levels of H3K27me1 in the absence of EZH2, and ablation of both EZH1 and EZH2 is required to completely abolish all H3K27 methylations (Højfeldt et al., 2018; Lavarone et al., 2019; Shen et al., 2008). PRC2-mediated H3K27me3 has long been known to play a role in gene silencing (Laugesen et al., 2019; Margueron and Reinberg, 2011; Pengelly et al., 2013; Yu et al., 2019). It is broadly deposited across the gene bodies and flanking regions of transcriptionally silent developmental genes and leads to recruitment of cPRC1, which initiates chromatin compaction and transcriptional repression (Bracken et al., 2019). The PRC2-mediated H3K27me2 modification is less well characterised, despite being ubiquitously distributed across intergenic sites of the genome (Conway et al., 2015; Streubel et al., 2018). It has been proposed to function as a repressive “blanket”, possibly preventing the inappropriate activation of enhancers of alternative lineages (Conway et al., 2015; Ferrari et al., 2014; Lee et al., 2015). In this sense, H3K27me2 could be considered to represent the “default” setting on chromatin. Contrastingly, the H3K27me1 modification is located along the bodies of actively transcribed genes, and correlates with the promotion of gene expression (Ferrari et al. 2014; Lee et al. 2015).

Almost all PRC2 accessory subunits enhance the histone methyltransferase (HMT) activity of the core complex *in vitro* (Laugesen et al., 2019). The one exception identified so far is the so-called Enhancer of Zeste Inhibitory Protein (EZHIP), which inhibits H3K27 methylation by allosterically activating PRC2 through binding the SET domain of EZH2 and preventing its methyltransferase activity (Hübner et al., 2019; Jain et al., 2019, 2020; Piunti et al., 2019; Ragazzini et al., 2019). Interestingly, whilst the loss of PALI1, JARID2 or Polycomb-like proteins in cells leads to reduced levels of H3K27me3, the loss or depletion of AEBP2 or EPOP results in an increase of H3K27me3 (Table 1.1) (Beringer et al., 2016; Conway et al., 2018; Grijzenhout et al., 2016; Højfeldt et al., 2018; Oksuz et al., 2018). Moreover, *Aebp2*^{-/-} mice exhibit anterior transformation of the skeleton, as opposed

to the classic Polycomb phenotype, i.e. posterior homeotic transformation (Grijzenhout et al., 2016). These paradoxical mutant phenotypes highlight the work that remains to be done to understand the distinct and potentially subtle or context-dependent functions of different PRC2 accessory components. Some clues are emerging from studies of *Aebp2*-null mouse ESCs (Conway et al., 2018; Grijzenhout et al., 2016). These suggest that although AEBP2 likely does stimulate the HMT activity of PRC2.2 *in vivo*, its loss skews the balance of PRC2 subtypes towards the more catalytically active PRC2.1.

1.9 PRC2 functions in mouse development

The requirement for PRC2 activity during mammalian embryogenesis is best exemplified by studies of germline loss-of-function mutations in *Ezh2*, *Eed* and *Suz12* in mice (Table 1.1). Loss-of-function mutants of these core PRC2 components invariably exhibit gastrulation defects and lethality around E7.5-8.5, during early post implantation stages (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004). Incidentally, loss-of-function mutations in the gene encoding the H3K36 histone methyltransferase NSD1 (discussed in detail later) also result in gastrulation defects and embryonic lethality at E7.5-8.5 (Rayasam et al., 2003). Death at this critical developmental stage reflects the essential functions of these chromatin regulators during early embryogenesis. Interestingly, although EZH1 is a core PRC2 subunit, *Ezh1*-knockout mice have been reported as “viable, fertile, and healthy” (Ezhkova et al., 2011), suggesting that EZH2 can compensate for its loss during development.

Loss-of-function mutants for accessory PRC2 subunits tend to exhibit more variable phenotypes, with lethality occurring in later embryonic, peri- or early post-natal stages of development (Table 1.1)(Conway et al., 2018; Grijzenhout et al., 2016; Rothberg et al., 2018). However, it is important to bear in mind that this does not necessarily translate to a lesser importance for the PRC2 accessory components in development. Rather, it may indicate that a level of functional redundancy exists between PRC2.1 and PRC2.2. Consider that the loss of a core PRC2 component will render both forms of the complex completely non-functional, whereas the loss of an accessory PRC2 component should affect one subcomplex only and leave the other functionally intact. It is likely that the more subtle accessory mutant phenotypes are the result of skewing the normal balance of PRC2 subtypes, as opposed to the outright loss of PRC2 activity.

Jarid2 mutant mice exhibit a range of phenotypes, the severity and age of onset of which appear to depend on the genetic background of the mice (Lee et al., 2000; Motoyama et

al., 1997; Takeuchi et al., 1995, 1999). To date, all reported constitutive *Jarid2* loss-of-function mutations have been generated by gene trapping, and cause pre-natal lethality between E10.5–E18.5 in mice (Table 1.1). Similarly, the phenotypes of different reported *Mtf2/Pcl2*-null mice range from lethal at E15.5 to viable (Li et al., 2014; Rothberg et al., 2018; Wang et al., 2007). Potential sources of phenotypic variation for a given knock-out lie not only in the genetic background of the mice, but also on the methods used to generate the null allele. For example, gene-trapping can cause different phenotypes depending on the vector insertion site (McClive et al., 1998; Olson et al., 1996). Additionally, gene-trap mutants can express tissue-specific, alternatively spliced forms of the gene in question, giving rise to hypomorphs rather than full knockouts. This phenomenon is exemplified by one strain of *Jarid2* gene-trap mutant mice, whose homozygotes retain leaky *Jarid2* expression in the nervous system (Lee et al., 2000). Alternative splicing has also been problematic with the *Mtf2/Pcl2* gene, which is expressed as multiple different isoforms (Li et al., 2014; Stanford et al., 2001). A more recent approach may have circumvented this issue by generating a *Mtf2/Pcl2* knockout mouse through a combination gene-trap and gene-targeting strategy (Rothberg et al., 2018). Accordingly, the resultant mice display the most severe and early-onset developmental phenotype, with none surviving past E15.5. Given the parallel roles of Polycomb-like proteins and JARID2 in directing the recruitment of PRC2.1 and PRC2.2 across the genome, it is entirely credible that *Mtf2/Pcl2* knockout mice should display a comparable knockout phenotype to *Jarid2*^{-/-} counterparts. Furthermore, it is reassuring to note that all prior reported *Mtf2/Pcl2* mutant mice did display some degree of the classic Polycomb defect (posterior skeletal transformation) and/or postnatal lethality (Li et al., 2014; Wang et al., 2007).

A similar disparity exists between the two published *Aebp2* loss-of-function mouse mutant phenotypes, with lethality reported to occur either postnatally or at E10.5 (Grijzenhout et al., 2016; Kim et al., 2011). Both groups employed a gene-trap strategy to target *Aebp2*, suggesting the phenotypic variation in this instance relates to the genetic background of the mice. What is most interesting to note is that one group observed an unexpected anterior homeotic transformation, or Trithorax (Schuettengruber et al., 2017), phenotype in their *Aebp2*-gene-trapped mice (Grijzenhout et al., 2016). This has since been interpreted as resulting from a skew in the balance of PRC2 subtypes towards the more catalytically active PRC2.1 (Conway et al., 2018).

Setting aside the potential for functional redundancy between PRC2.1 and PRC2.2, the relatively delayed phenotypes of PRC2 accessory subunit mutants could indicate that their key functions are executed during organogenesis (E10–E14) and/or the foetal growth and

development stages (E14 onwards), rather than during early embryogenesis. This is consistent with a model in which the core part of PRC2 is essential for the correct execution of cell fate specification during early embryogenesis, whilst the accessory PRC2 components come to the fore during later developmental stages to refine the activity of PRC2 and thereby give rise to progressively more differentiated and/or specialised cell types. An alternative interpretation could be that the consequences of the loss of PRC2 accessory proteins can be compensated for by the embryo, or the mother, up to that point in development. For example, despite the presumed persistence of some PRC2 activity via PRC2.2, PALI1-deficient embryos already show a reduction in global levels of H3K27me2/3 at E11.5, but do not exhibit lethality until the perinatal period (Conway et al., 2018). The functional consequences of these early molecular changes are tolerated by the embryos until birth, suggesting that: a) compensatory mechanisms exist to overcome this reduction in PRC2 enzymatic activity; or b) the physical manifestation of this reduction in H3K27me2/3 is not of functional consequence *in utero*. Another interpretation for the milder loss-of-function phenotypes of accessory PRC2 proteins could simply be that different accessory subunits are expressed in different tissues and at different stages during development, and that the penetrance and/or severity of the phenotype is rather a reflection of how essential the tissue(s) are for normal developmental progression. For example, a PRC2 accessory protein required for heart development would be expected to cause an earlier lethal phenotype than a subunit required for lung development. This is because a working circulatory system is essential for the embryo *in utero*, whereas independent breathing is not required until after birth. Therefore, to better understand the potential differential functions of PRC2 accessory proteins, it would be useful to generate a catalogue of their spatiotemporal expression patterns during embryonic development. Conditional loss-of-function mutant mice could then be generated to functionally validate their importance in particular lineages and/or at specific developmental stages.

Since none of the PRC2 accessory subunit loss-of-function mutants phenocopy the loss of a core PRC2 protein, it can be inferred that no single PRC2 accessory subunit is required for the initial recruitment and function of core PRC2 during mouse embryogenesis. As such, it would be interesting to generate mice with combined loss of JARID2 and Polycomb-like proteins to evaluate how this might compare to loss of core PRC2 activity.

Table 1.1 Constitutive loss-of-function phenotypes of PRC2 and related histone/DNA methyltransferase mouse mutants

	Gene	Gross developmental phenotype	Molecular mouse phenotype	Molecular mESC phenotype		References	
CORE PRC2	<i>Eed</i>	Lethal at E7.5-E8.5. Gastrulation defects.	ND	Complete loss of H3K27me1/2/3.		(Faust et al., 1995, 1998; Højfeldt et al., 2018; Montgomery et al., 2005; Schoeffner et al., 2006; Schumacher et al., 1996)	
	<i>Suz12</i>	Lethal at E7.5-E8.5. Gastrulation defects.	Loss of H3K27me2/3 at E7.5 and E8.5.	Complete loss of H3K27me1/2/3		(Højfeldt et al., 2018; Pasini et al., 2004)	
	<i>Ezh2</i>	Lethal at E7.5-E8.5. Gastrulation defects.	Complete loss of H3K27me2/3*	Global reduction of H3K27me2/3.	Complete loss of H3K27me1/2/3	(Erhardt et al., 2003; O'Carroll et al., 2001)	(Højfeldt et al., 2018; Lavarone et al., 2019; Shen et al., 2008)
	<i>Ezh1</i>	Viable, fertile and healthy.	ND	No change in H3K27me1/2/3		(Ezhkova et al., 2011)	
PRC2.2	<i>Jarid2</i>	Lethal at E10.5-E18.5. Developmental defects depending on strain.	ND	?		(Landeira and Fisher, 2011; Lee et al., 2000; Motoyama et al., 1997; Takeuchi et al., 1995, 1999)	
	<i>Aebp2</i>	Perinatal lethal. Anterior skeletal transformations.	ND	Slight increase in H3K27me3 at PRC2 target sites.		(Grijzenhout et al., 2016)	
		Lethal before E10.5.	ND			(Kim et al., 2011)	
PRC2.1	<i>Mtf2/ Pcl2</i>	Lethal at E15.5 Growth retardation and severe anaemia.	ND	Genome-wide reduction in H3K27me3.		(Perino et al., 2018; Rothberg et al., 2018)	
	<i>Pali1/ Gm340</i>	Perinatal lethal. No observable skeletal transformations.	Global reduction in H3K27me2/3 at E11.5.	Moderate loss of H3K27me3 at all genes.		(Conway et al., 2018)	
H3K36 HMT	<i>Nsd1</i>	Lethal at E7.5-E8.5. Gastrulation defects.	ND	Reduced H3K36me2. Increased H3K27me3. Reduced H3K27me2.		(Rayasam et al., 2003; Streubel et al., 2018)	
	<i>Setd2</i>	Lethal at E10.5-E11.5. Embryonic growth retardation.	Mutant embryos exhibit global reduction in H3K36me3, but no change in H3K36me1/2	ND		(Hu et al., 2010; Xu et al., 2019)	
DNMT	<i>Dnmt3a</i>	Postnatal lethality at ~4 weeks, incomplete penetrance Growth restriction.	ND	Reduced CpG methylation. Decreased SUZ12 binding and H3K27me3 at promoters of bivalent genes.		(Gu et al., 2018; Okano et al., 1999)	

ND, not determined; ?, lack of consensus to date; HMT, histone methyltransferase; DNMT, DNA methyltransferase; mESC, mouse embryonic stem cell; *Experiment performed on parthenogenetic *Ezh2*-null embryos.

Table 1.2 Constitutive heterozygous mutations in PRC2 and related histone/DNA methyltransferases implicated in human congenital growth disorders

	Gene	Mutation	Developmental Disorder	Key clinical features	Mutant molecular phenotype					References
					H3K27me3	H3K27me2	Methyl-CpG	H3K36me3	H3K36me2	
CORE PRC2	<i>EED</i>	Missense mutations in the WD40 domains of EED.	Cohen-Gibson syndrome (MIM 617561)	Overgrowth. Mild-to-severe intellectual disability. Advanced bone age. Macrocephaly.	Decreased levels in patient cells. Impaired deposition by EED-mutant-PRC2.	ND	ND	ND	ND	(Cohen and Gibson, 2016; Cohen et al., 2015; Cooney et al., 2017; Hetzelt et al., 2021; Smigiel et al., 2018; Spellicy et al., 2019; Tatton-Brown et al., 2017)
	<i>SUZ12</i>	Stop-gain, splice-site and missense mutation. Missense mutations enriched in the VEFS domain of SUZ12.	Imagawa–Matsumoto syndrome (MIM 618786)	Overgrowth. Mild-to-moderate intellectual disability. Advanced bone age. Macrocephaly.	Decreased levels in patient cells. Impaired deposition by SUZ12-mutant-PRC2.	ND	ND	ND	ND	(Cyrus et al., 2019b; Imagawa et al., 2017, 2018)
	<i>EZH2</i>	Missense mutations affecting predominantly (not exclusively) the SET domain of EZH2. Rare truncation mutations in last exon.	Weaver syndrome (MIM 277590)	Overgrowth. Mild-to-severe intellectual disability. Advanced bone age.	Decreased levels in patient cells. Decreased HMT activity <i>in vitro</i> and in a Weaver syndrome mouse model.	ND	Altered DNA methylation pattern in patient cells.	Increased H3K27 HMT activity on H3K36 _{me3} -containing substrates	ND	(Choufani et al., 2020; Cohen et al., 2016; Gibson et al., 2012; Imagawa et al., 2017; Jani et al., 2019; Lui et al., 2018; Polonis et al., 2018; Tatton-Brown et al., 2011, 2013)
H3K36 HMT	<i>NSD1</i>	Various loss of function mutations causing haploinsufficiency.	Sotos syndrome (MIM 117550)	Overgrowth. Mild-to-severe intellectual disability. Advanced bone age.	ND	ND	Altered DNA methylation pattern in patient cells.	ND	Reduced HMT activity of missense mutant NSD1 <i>in vitro</i> .	(Choufani et al., 2015; Kurotaki et al., 2002; Qiao et al., 2011; Tatton-Brown et al., 2005)
		Duplications including <i>NSD1</i> .	ND	Growth retardation. Mild-to-moderate intellectual disability. Delayed bone age. Microcephaly.	ND	ND	ND	ND	ND	(Dikow et al., 2013; Rosenfeld et al., 2013)
	<i>SETD2</i>	Loss-of-function substitution mutations.	Sotos-like syndrome	Exhibiting similar but fewer features of Sotos syndrome	ND	ND	ND	ND	ND	(Luscan et al., 2014; Tlemsani et al., 2016)
DNMT	<i>DNMT3A</i>	Various mutations affecting functional domains in DNMT3A. Deletions causing haploinsufficiency.	Tatton-Brown-Rahman syndrome (MIM 615879)	Overgrowth. Mild-to-moderate intellectual disability. Macrocephaly.	ND	ND	Altered DNA methylation pattern in patient cells.	ND	ND	(Kosaki et al., 2017; Okamoto et al., 2016; Shen et al., 2017; Tatton-Brown et al., 2014; Tlemsani et al., 2016)
		Change-of-function missense mutations affecting the PWWP domain of DNMT3A.	Microcephalic dwarfism	Growth restriction. Microcephaly.	Depleted at sites of aberrant DNA hypermethylation in patient cells.	ND	Increased at PcG-target genes in patient cells.	Binding to H3K36me3 impaired.	Binding to H3K36me2 impaired.	(Heyn et al., 2019)
		Duplication including <i>DNMT3A</i> .	ND	Growth restriction. Developmental delay.	ND	ND	ND	ND	ND	(Polonis et al., 2018)

ND, not determined; HMT, histone methyltransferase; DNMT, DNA methyltransferase

1.10 A role for imbalanced crosstalk between PRC2, NSD1 and DNMT3A in human developmental disorders

In humans, heterozygous mutations in the *EZH2*, *EED* and *SUZ12* genes cause congenital overgrowth, often marked by features of Weaver syndrome (Table 1.2). For simplicity, I may collectively refer to this group of disorders as being Weaver syndrome, a condition characterised by tall stature, a distinctive facial appearance and variable intellectual disability (Tatton-Brown et al., 2013). Two remarkably similar human overgrowth conditions, known as Sotos syndrome and Tatton-Brown-Rahman syndrome, are caused by heterozygous mutations in the genes encoding two other chromatin regulators, Nuclear receptor-binding SET Domain 1 (NSD1) and DNA (cytosine-5)-methyltransferase 3A (DNMT3A), respectively (Table 1.2). Similar to Weaver syndrome, these conditions are marked by childhood overgrowth, dysmorphic facial features and learning disabilities (Okamoto et al., 2016; Tatton-Brown et al., 2005). NSD1 is a histone methyltransferase that catalyses the addition of up to two methyl groups on histone H3 lysine residue 36, whilst DNMT3A is one of two mammalian *de novo* DNA methyltransferases, as introduced earlier (Bennett et al., 2017; Rose and Klose, 2014). At the molecular level, PRC2, NSD1 and DNMT3A exhibit context-dependent functional interplay or ‘crosstalk’ on chromatin (Fig. 1.7). Deciphering this molecular crosstalk may prove key to understanding the molecular aetiology of the Weaver, Sotos and Tatton-Brown-Rahman syndromes. For this reason, below I review recent insights into this crosstalk and summarise the phenotypes and genotypes of each of the related disorders in turn. Then I will propose a model in which the overlapping phenotypes of these genetically distinct overgrowth disorders could be due to imbalances in the landscape of chromatin modifications at intergenic regions.

1.10.1 Crosstalk between PRC2, NSD1 and DNMT3A

The activity of PRC2 on chromatin is subject to crosstalk from NSD1 and DNMT3A, which methylate H3K36 and CpG dinucleotides, respectively. Like H3K27, H3K36 can be modified by the addition of up to three methyl groups (Wagner and Carpenter, 2012). The NSD family of histone methyltransferases contains three members (NSD1-3) that catalyse the addition of up to two methyl groups on H3K36 (Wagner and Carpenter, 2012). In mammals, the histone methyltransferase SETD2 can then convert H3K36me2 to H3K36me3 (Edmunds et al., 2008). The ability of PRC2 to methylate H3K27 is influenced by the methylation state of H3K36 on the same histone such that its reaction rate is decreased on nucleosomes containing pre-existing di- or tri-methyl H3K36 modifications

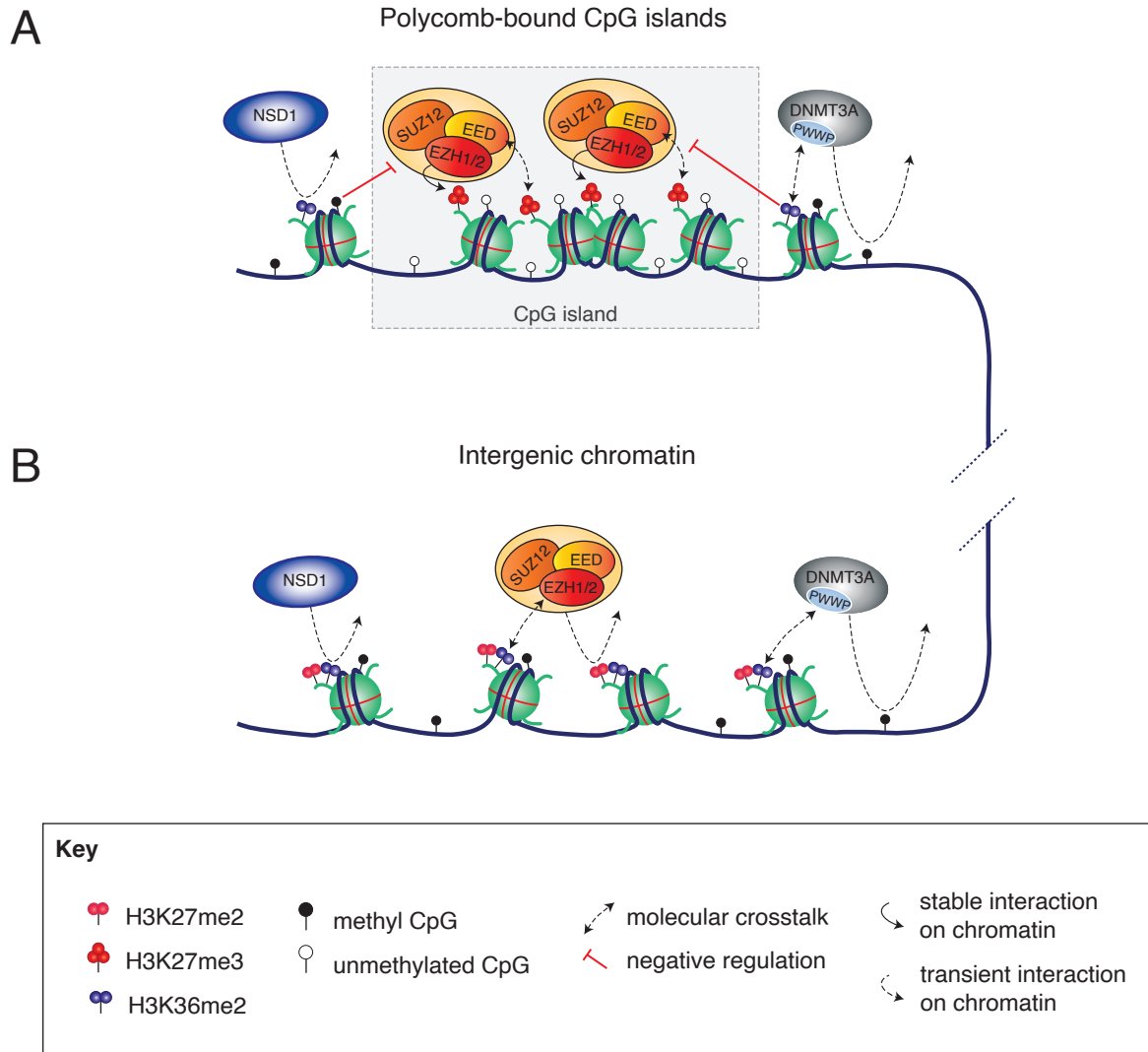


Figure 1.7 The crosstalk between PRC2, NSD1 and DNMT3A on chromatin

(A) Antagonistic interplay between NSD1-mediated H3K36me2, DNMT3A-mediated DNA methylation and PRC2-mediated H3K27me3 around Polycomb-bound CpG islands. Both DNA methylation and H3K36me2 are reported to antagonise the accumulation of PRC2-mediated H3K27me3. H3K27me3 allosterically activates PRC2.

(B) Co-existence of PRC2-mediated H3K27me2, NSD1-mediated H3K36me2 and DNMT3A-mediated DNA methylation at intergenic chromatin. The crosstalk between these different chromatin regulators is illustrated by the ability of both DNMT3A and EZH2 to read the methylation status of H3K36.

(Schmitges et al., 2011; Yuan et al., 2011; Zheng et al., 2012). In other words, the modification of H3K36 creates a less permissive environment for PRC2-mediated methylation at H3K27 (Finogenova et al., 2020). Important insight into how this crosstalk could be mediated came from the recent identification of a binding pocket on EZH2 that reportedly acts as a molecular sensor to detect the methylation state at H3K36 (Fig. 1.7) (Jani et al., 2019). The data suggest that PRC2 is activated *in cis* by unmodified H3K36, but this activation effect diminishes in the presence of an increasingly more methylated H3K36 residue (Jani et al., 2019). Reassuringly, a recent structural study of PRC2 on dinucleosomes came to the same conclusion (Finogenova et al., 2020). This would be predicted to lead to an inverse correlation between the methylation states at H3K36 and H3K27 in cells. Supporting this, H3K27me1 colocalises with H3K36me3 along the bodies of actively transcribed genes, while H3K27me2 colocalises with H3K36me2, both being broadly deposited across the genome (Ferrari et al., 2014; Streubel et al., 2018). Interestingly, despite the fact that 30–50% of all histone H3 is dimethylated at H3K36, the function of this modification still lacks thorough characterization (Rose and Klose, 2014). Some important clues came from the discovery that reduced H3K36me2 in NSD1-depleted cells leads to a quantitative increase and qualitative expansion in H3K27me3 deposition, with concomitant decreases in H3K27me2 (Shirane et al., 2020; Streubel et al., 2018). Furthermore, dimethylation of either H3K27 or H3K36 drastically reduces the rate of trimethylation occurring on the alternate residue, suggesting that an equilibrium exists between these two modifications once established (Zheng et al., 2012). One could theorise that blankets of H3K36me2 and H3K27me2 exist across broad regions of the genome (particularly at intergenic sites, which lack H3K36me1/3 and H3K27me1/3) to together function as the “default” setting on chromatin, limiting the potential for aberrant deposition of either activating H3K36me3 or repressive H3K27me3 (Fig. 1.7). In this model, intergenic decreases in the level of H3K36me2 shift the balance of PRC2-mediated methylations towards H3K27me3, at the expense of H3K27me2 (Streubel et al., 2018).

Molecular crosstalk also exists between H3K36 methylation and DNA methylation. Both di- and tri-methylation of H3K36 are known to recruit the *de novo* DNA methylases DNMT3A and DNMT3B to CpG-rich DNA via their PWWP (Pro-Trp-Trp-Pro) domains (Fig. 1.7) (Chen et al., 2004; Dhayalan et al., 2010; Ge et al., 2004). DNMT3B preferentially co-localises with H3K36me3 and methylates DNA along active gene bodies, a preference not shared with DNMT3A (Baubec et al., 2015). Like H3K36 di- and tri-methylation, DNA methylation is considered antagonistic to the deposition of PRC2-mediated H3K27me3 (Fig. 1.7) (Bartke et al., 2010; Reddington et al., 2013; Wu et al.,

2010). Consistent with this, H3K27me3 levels are reported to increase globally in mouse ESCs completely lacking DNA methyltransferase activity (Hagarman et al., 2013). In other words, decreased DNA methylation causes a global shift in the balance of PRC2-mediated methylation towards H3K27me3. Most, if not all studies to date on the relationship between PRC2 and DNA methylation have focussed on H3K27me3 alone as the read-out of PRC2 function. However, it is interesting to note that DNMT3A-mediated DNA methylation at intergenic chromatin co-exists with the H3K27me2 modification, as well as H3K36me2 (Fig. 1.7) (Wu et al., 2010). Therefore, to support the proposed model of molecular interplay, future studies are required to explore whether loss of DNMT3A function impairs deposition of intergenic H3K36me2 or H3K27me2. Likewise, DNA methylation profiles should be carefully examined under conditions of altered H3K36me2 and/or H3K27me2 deposition. Supporting this, it has been reported that *NSD1*^{+/-} Sotos syndrome patients exhibit a specific and reproducible DNA methylation signature, similar to that of patients with Tatton-Brown-Rahman syndrome (Choufani et al., 2015; Weinberg et al., 2019). Weaver syndrome patients, however, exhibit a distinct but equally specific DNA methylation signature (Choufani et al., 2020). Remarkably, the specificity of these DNA methylation signatures was such that a molecular distinction could be drawn between Sotos syndrome patients and Weaver syndrome patients, despite their highly overlapping clinical phenotypes. Importantly, this finding indicates that a germline mutation in a histone methyltransferase can have specific consequences on DNA methylation patterns that persist and are detectable in adult tissues.

1.10.2 Mutations in core PRC2 components cause overgrowth and intellectual disability

As mentioned above, heterozygous germline mutations in the genes encoding each of the core PRC2 subunits, *EZH2*, *EED* and *SUZ12*, are known to cause only recently distinguished congenital overgrowth and intellectual disability (OGID) syndromes – respectively Weaver, Cohen-Gibson and Imagawa-Matsumoto syndrome. Weaver syndrome is by far the most frequent and best studied condition of the three, to date.

Weaver syndrome (MIM #277590) is an autosomal dominant condition known to be caused by germline monoallelic mutations affecting the catalytic PRC2 subunit, *EZH2* (Gibson et al., 2012; Tatton-Brown et al., 2011). The clinical presentation of Weaver syndrome is characterised chiefly by pre- and post-natal overgrowth, mild-to-severe intellectual disability, advanced bone age and characteristic craniofacial features including

macrocephaly, round face, broad forehead, hypertelorism, large ears, prominent chin, long philtrum, low nasal bridge and retrognathia (Tatton-Brown and Rahman, 2013). The overgrowth-associated pathogenic variants of *EZH2* have, so far, been predicted to be predominantly loss-of-function (Cohen et al., 2016; Imagawa et al., 2017; Lui et al., 2018). However, alternative mechanisms cannot be ruled out in the absence of more thorough biochemical characterization – this shall be discussed in detail in chapter four.

Cohen-Gibson syndrome (MIM #617561; COGIS) is phenotypically very similar to Weaver syndrome but caused by germline heterozygous variants in *EED* (Cohen and Gibson, 2016; Cohen et al., 2015; Cooney et al., 2017; Hetzelt et al., 2021; Smigiel et al., 2018; Spellicy et al., 2019; Tatton-Brown et al., 2017). Cohen-Gibson syndrome shares the core defining features of Weaver syndrome, including its distinctive facial gestalt, overgrowth, advanced bone age and mild-to-severe intellectual disability and macrocephaly (Table 1.2). To date, twelve individuals have been identified with Cohen-Gibson Syndrome and at least nine of these individuals presented with *de novo* pathogenic variants in *EED* – intriguingly, no inherited pathogenic mutations in *EED* have yet been reported (Hetzelt et al., 2021). *EED* is recurrently affected by nonsynonymous mutations at a few key ‘hotspot’ residues within its WD-40 repeats, a key structural feature of the *EED* protein required for recognition of H3K27me3 on chromatin and allosteric activation of PRC2 (Hetzelt et al., 2021; Oksuz et al., 2018; Spellicy et al., 2019). As such, the pathogenic variants are predicted to be hypomorphic but, so far, lack comprehensive characterization (Cohen et al., 2015; Imagawa et al., 2017).

Finally, Imagawa–Matsumoto syndrome (MIM #618786; IMMAS) results from pathogenic germline heterozygous mutations in *SUZ12*. A total of thirteen individuals with suspected Imagawa–Matsumoto syndrome have been described in the literature thus far, and four of these have displayed autosomal dominant inheritance (Cyrus et al., 2019b; Imagawa et al., 2017, 2018). As with Cohen-Gibson syndrome, Imagawa–Matsumoto syndrome shares many of the characteristic features of Weaver syndrome (Table 1.2) though, interestingly, intellectual disability appears less frequently and less profound (Cyrus et al., 2019b; Imagawa et al., 2018). The reported pathogenic *SUZ12* variants include stop-gain, splice-site and missense mutations (Cyrus et al., 2019b). Notably, of the five reported missense variants, four occur within the *SUZ12* VEFS domain, a key functional domain for its association with *EZH2* and *EED* which, when deleted, abolishes the enzymatic activity of PRC2 (Cyrus et al., 2019b; Højfeldt et al., 2018; Imagawa et al., 2018). IMMAS-associated *SUZ12* mutations are reported to be partial loss-of-function but, like COGIS-associated *EED* mutations, require detailed biochemical assessment.

1.10.3 Monoallelic loss-of-function mutations in NSD1 cause Sotos syndrome

Sotos syndrome (MIM #117550) is an autosomal dominant developmental disorder resulting from either germline haploinsufficiency of, or intragenic loss-of-function mutations in, *NSD1* (Tatton-Brown and Rahman, 2013). At the phenotypic level, Sotos syndrome is remarkably similar to Weaver syndrome, characterised primarily by pre- and postnatal overgrowth, mild-to-severe intellectual disability, advanced bone age and characteristic craniofacial features that include downward slanting palpebral fissures, a long and thin face, and a prominent chin and broad forehead. The overlap between Weaver and Sotos syndrome will be discussed in more detail in chapter five.

The mutational spectrum of Sotos syndrome is diverse, including nonsense and missense mutations, partial and whole-gene deletions, intragenic indels and splice-site mutations. However it appears clear that the pathogenic variants always abrogate *NSD1* function (Tatton-Brown et al., 2005). For example, the pathogenic missense mutations in *NSD1* occur exclusively within functional domains of the protein that are implicated in chromatin regulation (Tatton-Brown et al., 2005). Furthermore, *in vitro* assays have determined that the H36 methyltransferase activity of *NSD1* is impaired in Sotos syndrome-mutant versions of the enzyme (Qiao et al., 2011). Interestingly, loss-of-function mutations in the related H3K36 histone methyltransferase *SETD2* cause a similar but less severe “Sotos-like” syndrome. This observation is resonant with what is seen in homozygous loss-of-function mouse models, with *Setd2*-null embryos exhibiting a less severe developmental mutant phenotype than *Nsd1*-null embryos (Table 1.1) (Hu et al., 2010; Rayasam et al., 2003).

1.10.4 De novo DNMT3A mutations are implicated in Tatton-Brown-Rahman syndrome

Tatton-Brown-Rahman syndrome (MIM #615879; TBRS) is an autosomal dominant genetic condition resulting from germline heterozygous mutations in the *DNMT3A* gene (Okamoto et al., 2016). The key clinical features of Tatton-Brown-Rahman syndrome are reminiscent of those observed in Weaver and Sotos syndromes, and include tall stature, mild-to-moderate intellectual disability and distinctive craniofacial characteristics including macrocephaly, a round face, heavy horizontal eyebrows and narrow palpebral fissures (Okamoto et al., 2016). Tatton-Brown-Rahman syndrome has only relatively recently been defined at the clinical level and so its molecular characterization is yet lacking.

The spectrum of reported TBRS-associated pathogenic mutations in *DNMT3A* includes microdeletions, in-frame deletions, frameshift insertions and missense mutations (Tatton-

Brown et al., 2014). Although the biochemical and functional characterization of these mutations remain to be determined, they are predicted to interfere with the intra- and inter-molecular protein-protein interactions of DNMT3A and ultimately disrupt its ability to accurately methylate DNA (Tatton-Brown et al., 2014). Given that haploinsufficiency of *DNMT3A* has been ruled in as a mutational mechanism for Tatton-Brown-Rahman syndrome, it seems likely that all pathogenic variants will prove to be loss-of-function (Okamoto et al., 2016).

1.10.5 Imbalanced regulation of PRC2 at intergenic chromatin as a common feature of overgrowth syndromes

Considering the convergence of PRC2-mediated H3K27me₂, NSD1-mediated H3K36me₂ and DNMT3A-mediated DNA methylation at intergenic chromatin, I propose that an equilibrium exists between these modifications and that a shared molecular feature of the abovenamed developmental disorders may be disruptions to this balance, which shift the landscape of PRC2-mediated methylation (Fig. 1.7). Supporting this, there is evidence for gene dosage effects of *NSD1* and *DNMT3A* on human growth (Table 1.2). In other words, an increase in the copy number of either gene appears to have the opposite effect on growth when compared to having one loss-of-function mutation in that same gene. For example, whilst *NSD1* haploinsufficiency is associated with overgrowth, reciprocal duplications involving *NSD1* correlate with opposing clinical features including growth retardation, delayed bone age and microcephaly (Dikow et al., 2013; Rosenfeld et al., 2013). Similarly, an individual with a maternally inherited duplication encompassing *DNMT3A* exhibits a growth failure phenotype marked by developmental delay, despite also possessing a paternally inherited Weaver syndrome-related point mutation in *EZH2* (Polonis et al., 2018). This phenotype suggests that the increased gene dosage of *DNMT3A* exerts dominant effects that serve to restrict growth, masking any contribution from the missense *EZH2* mutation, which normally causes overgrowth. It should be noted, however, that for both of the above cases, a functional increase in the activity of the duplicated gene product has yet to be confirmed.

Further, more complex, change-of-function mutations affecting the PWWP domain of *DNMT3A* which abrogate its ability to bind to H3K36me₂ and H3K36me₃ *in vitro*, but result in DNA hypermethylation at sites marked by H3K27me₃, also result in growth restriction in both mice and humans (Heyn et al., 2019; Sendžikaitė et al., 2019). Although only *DNMT3A* has been mutated in this context, the mutation has functional consequences

on the normal crosstalk with NSD1-mediated H3K36me₂, and an imbalance in the landscape of DNA methylation specifically at PRC2 target sites is observed. This result is consistent with a model in which a carefully balanced equilibrium exists between PRC2, NSD1 and DNMT3A that, when skewed, can affect growth regulation.

A Weaver syndrome-associated mutation affecting the H3K36-sensing pocket in EZH2 is also reported to disrupt the crosstalk between PRC2 and NSD1 on chromatin (Jani et al., 2019). When compared to wild-type PRC2 in *in vitro* assays, disease-associated mutant EZH2-containing PRC2 exhibits an increased enzymatic activity on H3K36-trimethylated nucleosomes, despite a decreased overall HMT activity on unmodified nucleosomes (Jani et al., 2019). This change would be predicted to shift the balance between H3K27me₂ and H3K36me₂ at intergenic chromatin, as well as potentially disrupting the profiles of H3K27 and H3K36 methylation genome wide. Again, this result is consistent with a model of disturbed crosstalk between PRC2 and NSD1 as a feature of Weaver syndrome.

Clearly, changes in the crosstalk between PRC2, NSD1 and DNMT3A that affect the balance of their associated modifications can have consequences on human growth. I propose that imbalances at intergenic chromatin are particularly relevant to the aetiology of the Weaver, Sotos and Tatton-Brown-Rahman syndromes. Supporting this, while the function of DNMT3A and DNMT3B largely overlap, DNMT3B differs in that it preferentially methylates genic DNA (Baubec et al., 2015). Mutations in the *DNMT3B* gene do not cause human overgrowth, indicating that disruptions to the non-overlapping functions of DNMT3A (i.e., non-genic DNA methylation) are what contribute to the overgrowth phenotype. Similarly, SETD2 tri-methylates H3K36 at active gene bodies, but heterozygous loss-of-function mutations in *SETD2* causes a less severe “Sotos-like” phenotype (Table 1.2). This might suggest a less important role for genic H3K36 methylation in the regulation of growth and development (Luscan et al., 2014; Tlemsani et al., 2016). Furthermore, genes encoding members of cPRC1 and vPRC1, which co-localise with PRC2-mediated H3K27me₃ at unmethylated CpG islands, have so far not been reported to be mutated in human overgrowth syndromes. In fact, a mutation in their core component, *RING1A*, causes a dissimilar, neurodevelopmental disorder (Pierce et al., 2018). Taken together, these findings suggest that imbalances in chromatin modifications at gene bodies and CpG islands are not foremost in the pathogenesis of the human overgrowth syndromes, leaving intergenic chromatin as the key candidate.

1.11 Aims of thesis

The overarching aim of this thesis was to explore the functions of both accessory (PALI) and core (EZH2) PRC2 components in development, using a combination of *in vivo* and *in vitro* models. Specifically, I aimed to:

- 1) Generate new insight into the biochemical and biological functions of the PALI protein family, with a focus on PALI1. This comprised the following sub-aims:
 - Determine the localisation of PALI1 on chromatin using chromatin immunoprecipitation methods in mouse embryonic stem cells.
 - Examine whether PALI1 contributes to the recruitment or targeting of PRC2 to chromatin in mouse embryonic stem cells.
 - Analyse mouse models to provide new insight into the potential biological functions of PALI1 and PALI2 *in vivo*, including the preliminary characterisation of knockout mice lacking one or both PALI proteins.
- 2) Create an isogenic panel of cellular models of *EZH2*-mutant Weaver syndrome to facilitate the following sub-aims:
 - Provide a comprehensive biochemical characterisation of the enzymatic activity of a wide variety of Weaver-syndrome-associated *EZH2* variants, and thereby define the mutational mechanism(s) of *EZH2*-mutant Weaver syndrome.
 - Investigate the consequences of any observed perturbations to PRC2 function on the wider chromatin and transcriptional landscape.
- 3) Identify common molecular signatures that may contribute to the phenotypic overlap of the genetically distinct Weaver and Sotos syndromes. This involved the following sub-aims:
 - Generate an isogenic cellular model of *Nsd1*-mutant Sotos syndrome, and for comparison, *EZH2*-associated growth restriction.
 - Identify molecular changes that are shared between isogenic cellular models of Weaver and Sotos syndrome, but divergent between overgrowth versus growth restriction models.

CHAPTER 2:

MATERIALS AND METHODS

Experimental procedures were carried out at room temperature unless otherwise stated.

2.1 Cell culture methods

2.1.1 Basic cell culture conditions

HEK293T and NTERA-2 cells were cultured using high glucose Dulbecco's Modified Eagle's Medium (DMEM; Sigma D6429) supplemented with 10% (v/v) FBS (Gibco), 100 U/mL Penicillin and 100 µg/mL Streptomycin (Gibco). These cells were maintained on TC-treated culture dishes, passaged every two or three days using 0.25% Trypsin-EDTA (Gibco), and cultured at 37 °C in a 5% CO₂ (v/v) incubator.

Lymphoblastoid cell lines (LCLs) obtained from Coriell Biorepositories were cultured in RPMI 1640 Medium supplemented with 2mM GlutaMAX (Gibco 61870036) and 15% (v/v) FBS (Gibco). These cells were grown in suspension using tissue culture flasks in an upright position, split every three to five days, and cultured at 37 °C in a 5% CO₂ (v/v) incubator.

Mouse embryonic stem cells (mESCs) were maintained on 0.1% gelatin-coated culture dishes, passaged every two days using 0.25% Trypsin-EDTA (Gibco) and cultured at 37 °C in a 5% CO₂ (v/v) incubator. The basic (Serum/LIF) mESC culture medium consisted of Glasgow Minimum Essential Medium (GMEM; Sigma) supplemented with 10% (v/v) heat-inactivated ES-qualified FBS (Millipore) or 20% (v/v) heat-inactivated FBS (Gibco), 100 U/mL Penicillin and 100 µg/mL Streptomycin (Gibco), 50 µM β-mercaptoethanol, 1:100 GlutaMAX (Gibco), 1:100 non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco) and 1:500 homemade leukaemia inhibitory factor (LIF). Under Serum/LIF/2i growth conditions, the GSK inhibitor CHIR99021 (Cayman Chemical Company) and MEK inhibitor PD0325901 (Cayman Chemical Company) were also added at final concentrations of 3 µM and 1 µM, respectively. The iWT-EEDr cell line (a gift from Prof Danny Reinberg) and all its derivatives were cultured using high glucose Dulbecco's Modified Eagle's Medium (DMEM; Sigma D6429) supplemented with 15% (v/v) heat-inactivated ES-qualified FBS (Millipore), 100 U/mL Penicillin and 100 µg/mL Streptomycin (Gibco), 50 µM β-mercaptoethanol, 1:100 GlutaMAX (Gibco), 1:100 non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), 1:500 homemade leukaemia inhibitory factor (LIF), 3 µM CHIR99021 (Cayman Chemical Company) and 1 µM PD0325901 (Cayman Chemical Company).

2.1.2 Generation of knockout cell lines

The genetic manipulation of the *Ezh2*^{fl/fl}; *Rosa26*^{CreERT2} mouse embryonic stem cell line – kindly sent to us by Prof. Kristian Helin – lay the foundation for much of the research presented in this thesis. This cell line was created to derive OHT-inducible homozygous *Ezh2*-knockout (KO) mESCs (Højfeldt et al., 2018). I modified the protocol to preferentially induce heterozygous *Ezh2*-knockouts (*Ezh2*^{fl/Δ}); *Ezh2*^{fl/fl}; *Rosa26*^{CreERT2} mESCs were seeded at a density of 1.5×10⁶ cells per 10cm dish and two days later either treated with (a) 0.5 μM 4-OHT for 12 hours or (b) 0.25 μM 4-OHT for 24 hours. Single cells were sorted into 96-well culture plates using a FACSAria Fusion Flow Cytometer operated by Dr Barry Moran at the Trinity Biomedical Sciences Institute Flow Cytometry Facility. Single cell clones were screened at the DNA level for the desired heterozygous deletion in *Ezh2* (see appendix I.I for genotyping primers) and positive clonal populations were expanded for further validation at the mRNA and protein expression level.

The *Ezh2*^{fl/fl}; *Rosa26*^{CreERT2} cell line was also used to generate *Nsd1* heterozygous knockout mESC lines via CRISPR/Cas9 genome editing. Guide RNAs (gRNAs; see appendix I.II) were designed in pairs (using the Broad Institute GPP sgRNA design tool) to target the 5'-end of the mouse *Nsd1* gene with the aim of modelling the partial *NSD1* deletions observed in human Sotos syndrome patients and to cease *Nsd1* expression (Douglas et al., 2005). Each gRNA was cloned into the pSpCas9(BB)-2A-GFP (PX458) plasmid backbone – a gift from Feng Zhang (Addgene plasmid #48138; <http://n2t.net/addgene:48138>; RRID: Addgene_48138) – following the Zhang lab general cloning protocol, and positive clones were confirmed by Sanger sequencing. *Ezh2*^{fl/fl}; *Rosa26*^{CreERT2} mESCs were seeded at a density of 1.5×10⁶ cells per 10cm dish to achieve ≥70% confluency for transfection the following day. Lipofectamine 2000 (Invitrogen) was used as per the manufacturer's protocol to either (a) co-transfect cells with 2.5 μg of each pSpCas9(BB)-2A-GFP-gRNA plasmid of a pair (i.e., 5 μg total plasmid DNA), or (b) to transfect cells with 5 μg of control plasmid DNA: pSpCas9(BB)-2A-GFP “empty vector” or the GFP-expressing pGIPZ plasmid. Although the efficiency of this transfection reaction is typically low (~5%), transfection success can be inferred from observable GFP expression on the pGIPZ-transfected control plate after 24 hours. Approximately 48-hours post-transfection, GFP-expressing (GFP+) single cells were sorted into 96-well culture plates using a FACSAria Fusion Flow Cytometer operated by Dr Barry Moran at the Trinity Biomedical Sciences Institute Flow Cytometry Facility. Single cell clones were screened at the DNA level for the desired

deletions in *Nsd1* (see appendix I.I for genotyping primers) and positive clonal populations were expanded for further validation at the mRNA and protein expression level.

The iWT-EEDr cell line was used to generate *Pali1* homozygous knockout mESC lines via CRISPR/Cas9 genome editing as described above. gRNAs were designed to target the *Pali1*-specific exon of the *Lcor* gene locus (see appendices I.I-II).

2.1.3 Lentivirus production

A second-generation lentiviral system was used to produce lentivirus particles. HEK293T cells were seeded in complete DMEM at a density of $\sim 2 \times 10^6$ cells per T75 flask to achieve $\geq 70\%$ confluency for transfection the following day. A standard polyethylenimine (PEI) transfection protocol was followed to co-transfect 3.5 μg of lentiviral transfer plasmid (i.e., the desired expression vector) with 3 μg PAX8 packaging plasmid and 2 μg VSV-G envelope plasmid. An additional flask of HEK293T cells was transfected with 8.5 μg of the GFP-expressing pGIPZ plasmid alone; high transfection efficiency was inferred from the level of GFP-expression observed on this control flask. Approximately 18 hours post-transfection, the DMEM was washed from HEK293T cells with Dulbecco's Phosphate Buffered Saline (DPBS) and replaced with appropriate mESC medium. Cell culture supernatants containing lentivirus particles were harvested at 48- and 72-hours post transfection. Viral supernatants were pooled and passed through a 0.45 μm polyethersulfone (PES) filter.

2.1.4 Generation of stable cell lines by viral transduction

Mouse embryonic stem cells were seeded at a density of $\sim 2 \times 10^4$ cells per cm^2 . The following morning viral supernatant – undiluted or diluted with complete mESC medium, containing appropriate final concentrations of LIF/2i and 5 $\mu\text{g}/\text{mL}$ polybrene – was applied to cells. This process was repeated ~ 24 hours later. Approximately 48 hours post-transduction all viral supernatant was removed, cells were washed with DPBS, and complete mESC medium containing 1 $\mu\text{g}/\text{mL}$ puromycin was applied. Antibiotic selection was maintained for an empirically defined length of time (typically four to six days) to obtain putative stable cell lines, which were subsequently screened for expression of the intended transgene.

2.1.5 Embryoid body differentiation

For embryoid body (EB) differentiation, mouse embryonic stem cells were washed with DPBS twice then detached using 0.25% Trypsin-EDTA (Gibco). Trypsinization was stopped by the addition of 10% (v/v) FBS in DPBS, cells were washed once more with DPBS and then seeded onto non-adherent petri dishes at a density of 3×10^5 cells/mL in basic mESC culture medium without LIF.

2.1.6 Cell size and competitive cell growth assay

I designed this assay to compare mouse embryonic stem cell types A and B in terms of their relative size and proliferation rate (see Fig. 5.6). Cell types A and B were first engineered by viral transduction to stably express green or red fluorescent proteins, respectively; Cell type A was transduced with the EF.PGK.GFP plasmid while cell type B was transduced with the EF.PGK.DsRed plasmid (see Table 2.1, below). Cell types A and B were then bulk sorted to obtain pure GFP- or DsRed-expressing populations using a FACS Aria Fusion Flow Cytometer operated by Dr Barry Moran at the Trinity Biomedical Sciences Institute Flow Cytometry Facility. On day 0, pure A-GFP+ and B-DsRed+ populations were separately counted (LUNA Automated Cell Counter) then mixed at an expected 1:1 ratio and seeded in technical triplicate onto a 96-well plate at a density of 1.6×10^4 total cells per well. The Guava easyCyte 8HT bench cytometer (Luminex) was used to determine the profile and relative abundance of GFP+/DsRed+ (sub)populations for each cell line individually and in A/B combination. Every 2 days, cells were split at the required ratio and the relative abundance of GFP+/DsRed+ subpopulations re-assessed. For each well of the 96-well plate processed by the flow cytometer ≥ 5000 total cells (events) were assessed. Flow cytometry data was processed using FlowJo v10 software with the assistance of Dr Barry Moran, and results were further analysed and plotted using GraphPad Prism 9.

To assess relative cell size, first the median forward scatter area (FSC-ALin) value was determined for each A and B subpopulation for each replicate and at each time point. Next, within a given technical triplicate dataset, all median FSC-ALin values were expressed relative to the average median FSC-ALin value of the DsRed-expressing cell line (*Ezh2*^{fl/Δ} + hEZH2 WT or *Ezh2*^{fl/Δ} + hEZH2 A733T) for that A/B combination. This resulted in ten (n=10) relative cell size values per cell line, per A/B combination. For a given A/B combination, these relative cell size values were visualised as violin plots and analysed using an unpaired t-test with Welch's correction.

Relative cellular proliferation rates were determined by tracking the proportion of GFP+/DsRed+ cells within a given A/B combination over time, expressed as a percentage of the total number of cells analysed.

Table 2.1 A/B cell combinations used in flow-cytometry-based assay of relative cell size and proliferation rate

A – GFP expressing	B – DsRed expressing
<i>Ezh2^{fl/Δ}</i> + hEZH2 R679C	<i>Ezh2^{fl/Δ}</i> + hEZH2 WT
	<i>Ezh2^{fl/Δ}</i> + hEZH2 A733T
<i>Ezh2^{fl/Δ}</i> + hEZH2 K563E	<i>Ezh2^{fl/Δ}</i> + hEZH2 WT
	<i>Ezh2^{fl/Δ}</i> + hEZH2 A733T
<i>Ezh2^{fl/Δ}</i> + hEZH2 Y728*	<i>Ezh2^{fl/Δ}</i> + hEZH2 WT
	<i>Ezh2^{fl/Δ}</i> + hEZH2 A733T
<i>Ezh2^{fl/Δ}</i> + hEZH2 A733T	<i>Ezh2^{fl/Δ}</i> + hEZH2 WT

2.2 Mouse methods

2.2.1 Animals

The Health Products Regulatory Authority (HPRA) is responsible for the correct implementation of EU directive 2010/63/EU in Ireland. All animals and procedures used in this study were approved under HPRA project authorisation number AE19136/P124 or used according to the in-house guidelines for the care and use of laboratory animals of the RIKEN Center for Integrative Medical Sciences, Yokohama, Japan [Approval number: Kei-27-001(7)]. The *Pali1* sKO, *Pali2* sKO and *Pali1/2* sKO mice used in this study were originally generated by the Koseki lab (RIKEN, Japan). The (129×C57BL/6)F1-*Gm340^{tm2.1Hko}* × C57BL/6NTac-*Gt(ROSA)26Sor^{tm9(cre/ESR1)Arte}* mouse strain, from which *Pali1* sKO mice can be derived, is now bred and housed in specific-pathogen-free environments at the Comparative Medicine Unit, Trinity Biomedical Sciences Institute, Trinity College Dublin.

2.2.2 Mouse dissection and tissue collection

For embryo collection, pregnant females were euthanized by cervical dislocation. The uterus containing the embryos was dissected from the dam and transferred to a petri dish of ice-cold (4°C) PBS. Individual embryos were dissected out of the uterus under a stereomicroscope in a petri dish with fresh ice-cold PBS. Care was taken to remove all extra-embryonic membranes a small sample of the yolk sac was taken for gDNA isolation and subsequent genotyping. Postnatal mice (P0) and embryos at E14.5 or later were euthanized by decapitation. P0 pups were cut along the ventral midline and terminal dissection was performed for tissue collection. Tissues of interest were isolated by dissection in ice-cold PBS under a stereomicroscope. If not for use in histochemical analyses, freshly dissected tissues were submerged in an excess of RNAlater solution (Sigma R0901) and incubated at 4°C overnight before transfer to -20°C storage.

2.2.4 Tissue fixation

Freshly dissected mouse embryos and P0 mouse tissues were transferred to an appropriate volume of ice-cold 4% formaldehyde (Sigma) and fixed at 4°C with rotation overnight. After fixation, embryos/tissues were washed thrice for 10 minutes each in PBS and then dehydrated stepwise through a methanol series (25% MeOH in PBS, 50% MeOH in PBS and 75% MeOH in H₂O) into 100% methanol. Samples could be stored long-term at -20°C in 100% methanol.

2.2.4 Paraffin embedding and tissue sectioning

Samples were transferred from 100% methanol to a 50% methanol 50% ethanol mix, and then transferred to 70% ethanol in H₂O. A Leica TP1020 benchtop tissue processor was used to process samples stepwise through an ethanol dehydration series, xylene clearance and paraffin infiltration. Samples were embedded in paraffin with the aid of a Leica EG1150 H Heated Paraffin Embedding Module. A rotary microtome was used with Leica DB80 LS blades to section paraffin-embedded specimens at a 5 µm thickness. Sections were transferred onto Superfrost Plus adhesion microscope slides (Thermo Fisher) and baked on at 40°C overnight.

2.3 DNA methods

2.3.1 High-throughput genotyping of genome-edited cell lines

Single colonies growing in 96-well plates were detached and split onto identical wells of two separate 96-well plates. One of these 96-well plates was to facilitate the continued cell culture of desired clones, while the other was reserved for high-throughput gDNA extraction. The cells on the genotyping plate were cultured without special supplements (i.e., LIF/2i) but grown to complete confluence. Once confluent, cells were washed with DPBS then scraped and resuspended in 20 μ L QuickExtract DNA Extraction Solution 1.0 (Lucigen). Cell lysates were directly transferred to PCR strip tubes using a multichannel pipette, then heated in a thermocycler at 65°C for 15 minutes, followed by 68°C for 15 minutes and finally 98°C for 10 minutes. While this lysis cycle was running, genotyping PCR reactions were pre-loaded – minus template DNA – into PCR strip tubes using OneTaq Quick-Load 2X Master Mix with Standard Buffer (NEB M0486) as per the manufacturer's protocol. Next, 3 μ L of the so-called 'quick extracted' gDNA was directly transferred by multichannel pipette to the pre-loaded PCR reaction mix and subject to appropriate PCR cycling. Finally, PCR products were directly loaded onto an agarose gel for electrophoretic analysis, thanks to the tracking dyes already present in the PCR reaction mix.

2.3.2 Expression vectors

The ectopic expression of the 3xHA-tagged wild-type or double mutant extPIP domain of PALI1 was performed using an in-house lentiviral expression vector derived from pLEX_305 (Addgene plasmid #41390; <http://n2t.net/addgene:41390>; RRID: Addgene_41390).

The ectopic expression of hEZH2 ORFs was performed using the in-house pLENTI EFIA FLAG/HA lentiviral expression vector (Conway et al., 2018). Mutant hEZH2 ORFs were generated by Q5 Site-Directed Mutagenesis (NEB E0554S) as per the manufacturer's protocol. See appendix III for the wild-type ORF sequence, and appendix IV for Sanger sequence confirmation of the mutant constructs.

The EF.PGK.GFP lentiviral expression vector was used to drive ectopic expression of green fluorescent protein in mouse embryonic stem cells. EF.PGK.GFP was a gift from Linzhao Cheng (Addgene plasmid #17618; <http://n2t.net/addgene:17618>; RRID: Addgene_17618). The DsRed2 ORF from the EF.CMV.RFP plasmid was cloned into the

EF.PGK.GFP backbone, in place of the GFP ORF, to generate the EF.PGK.DsRed lentiviral expression vector used to drive ectopic expression of red fluorescent protein in mouse embryonic stem cells. EF.CMV.RFP was a gift from Linzhao Cheng (Addgene plasmid #17619; <http://n2t.net/addgene:17619>; RRID: Addgene_17619)

2.3.3 Chromatin immunoprecipitation (ChIP)

Cells were counted and normalized cell pellets (each of equal cell number, up to 50×10^6) were washed twice with DPBS before crosslinking for 10 minutes with 1% formaldehyde (Sigma) in DBPS. Crosslinking was quenched with 125 mM glycine for 5 minutes before two PBS washes were performed. Crosslinked cells were lysed in an appropriate volume of SDS-lysis buffer (100 mM NaCl, 50 mM Tris pH 8.1, 5 mM EDTA pH 8.0, 0.5% SDS, 2 μ g/mL aprotinin, 1 μ g/mL leupeptin, 1 mM PMSF). Chromatin was pelleted by centrifugation at 350 RCF for 5 minutes at room temperature. The supernatants were discarded, and chromatin resuspended in 1 mL of ChIP buffer (2:1 dilution of SDS-lysis buffer/Triton dilution buffer [100 mM Tris pH 8.6, 100 mM NaCl, 5 mM EDTA pH 8.0, 5% Triton X-100, 2 μ g/mL aprotinin, 1 μ g/mL leupeptin, 1 mM PMSF]). Chromatin was sheared to approximately 200bp–800bp fragments by successive 1 second pulses of sonication at 50% amplitude, for a total of 4 minutes sonication time (Branson Sonifier SFX 150). Efficient sonication was verified by analysing a small sample of the fragmented DNA on an agarose gel. Input samples were taken from well-sonicated chromatin before being divided according to the desired cell number (equivalent) per antibody to be used. These chromatin aliquots were incubated overnight with the desired antibody while rotating at 4°C. To ChIP FLAG-tagged proteins, chromatin aliquots were instead incubated overnight with 40 μ L equilibrated anti-FLAG beads (gift from Dr Claudio Ciferri).

The following morning, chromatin-antibody mixes were spiked with 40 μ L equilibrated Protein G/A Dynabeads (Invitrogen) and then left to rotate for a further two to four hours at 4°C. After this, all beads were washed three times in mixed micelle buffer (150 mM NaCl, 20 mM Tris pH 8.1, 5 mM EDTA pH 8.0, 5.2% Sucrose, 1% Triton X-100, 0.2% SDS), twice with buffer 500 (0.1% Sodium Deoxycholate, 1 mM EDTA pH 8.0, 50 mM HEPES pH7.5, 1% Triton X-100), twice with LiCl detergent wash (0.5% Sodium Deoxycholate, 1 mM EDTA pH 8.0, 250 mM LiCl, 0.5% NP-40, 10 mM Tris pH 8.0) and finally one wash with TE (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0). Immunoprecipitated material was eluted from the beads with elution buffer (0.1 M

NaHCO₃, 1% SDS) while shaking for 1 hour at 65°C. The supernatant was retained and incubated overnight at 65°C while shaking to reverse the crosslinks.

The eluted complexes were then subject to RNase (Thermo Fisher) and Proteinase K (Sigma) treatment prior to column-based DNA fragment purification (Qiagen or Thermo Fisher). ChIP enrichment was analysed by qPCR using the SYBR Green I detection chemistry (NEB M3003E) on an Applied Biosystems Quant Studio 3 platform. See appendix I.V for a list of primers used in ChIP-qPCR analyses.

Frozen embryos were prepared for ChIP (chapter three) by thawing on ice then dissociating in ice-cold DPBS using a gentleMACS Dissociator with C-tubes (Miltenyi Biotec). Dissociated cells were pooled according to genotype and washed twice with DPBS before crosslinking and proceeding with the protocol as described above.

In attempts to optimize our ChIP protocol for PALI1 (chapter three), different crosslinkers were trialled for their ability to stabilize the predicted PALI1-DNA interaction. These were ethylene glycol bis(succinimidyl succinate), disuccinimidyl glutarate and disuccinimidyl suberate – also respectively known as EGS, DSG and DSS. I tried double crosslinking with EGS (Thermo Fisher 21565) and formaldehyde, and then quadruple crosslinking with EGS, DSG (Thermo Fisher 20593), DSS (Thermo Fisher 21655) and formaldehyde. In my modified protocol, cells were first crosslinked for up to 1 hour at room temperature with 2 mM DSS and 2 mM DSG and/or 2 mM EGS, and then fixed with 1% formaldehyde for 10 minutes before quenching with 125 mM glycine as per the standard protocol. Double or quadruple fixed chromatin required longer sonication times to achieve the desired 200–800 bp fragments.

2.3.4 Quantitative chromatin immunoprecipitation relative to a reference exogenous genome (ChIP-Rx)

Quantitative chromatin immunoprecipitation relative to a reference exogenous genome (ChIP-Rx) and coupled with massively parallel DNA sequencing was performed for the genome-wide mapping of histone modifications, SUZ12 and epitope-tagged EZH2, as described previously by Orlando et al., (2014). Human chromatin lysate (from NTERA-2 cells) was spiked into mouse chromatin lysate at an expected ratio of 1:9, respectively. This spike-in chromatin was added at the beginning of the workflow (before sonication) and the combined human-mouse chromatin sample was treated as a single ChIP-seq experiment until completion of DNA sequencing. Following completion of the ChIP protocol described above (section 2.3.3), the precipitated DNA was quantified using the Qubit

dsDNA High Sensitivity Assay Kit (ThermoFisher Q32854). ChIP-Rx sequencing libraries were prepared using the NEBNext Ultra II DNA Library Kit for Illumina (NEB E7645) and NEBNext Multiplex Oligos for Illumina (Index Primers Set 1 and 2; NEB E7335 and E7500) as per the manufacturer's protocol. Input DNA amounts ranged from 5–50 ng, but stayed consistent within each set of ChIPs (e.g., 50ng for all H3K27me2 ChIPs but 5ng for all H3K27me3 ChIPs). Following adaptor ligation, DNA was PCR amplified for 4–7 cycles, depending on the input DNA amount. PCR-amplified DNA was purified (without size selection) using AMPure XP beads (Beckman Coulter A63881) warmed to room temperature. The quality of DNA libraries was analysed on a High Sensitivity D1000 Screen Tape (Agilent). The resulting libraries were then used for cluster generation and sequencing using an Illumina NextSeq 500 (ID: NB501524), with 75bp read length.

2.4 RNA methods

2.4.1 RNA preparation and qRT-PCR analysis

Total RNA was extracted from cells using the RNeasy Mini kit (Qiagen) as per the manufacturer's protocol. Oligo-dT₍₁₈₎ primers were used to generate cDNA by reverse transcription PCR using the High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems 4374966). Relative RNA abundance was determined by SYBR Green I detection chemistry (NEB M3003E) on the Applied Biosystems QuantStudio 3 Real-Time PCR System. The level of *18S* rRNA or *Rplp0* mRNA was used as a normaliser. See appendix I.IV for a list of primers used in qRT-PCR analyses.

2.4.2 Next generation sequencing of total RNA

RNA-seq libraries were prepared using the NEBNext rRNA Depletion Kit v2 (Human/Mouse/Rat) (NEB E7405) and the NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB E7770) as per the manufacturer's protocol. Prior to starting library prep, RNA concentrations were measured on a Qubit 3.0 and RNA integrity numbers (RIN) calculated using an RNA ScreenTape (Agilent). A total 1 µg of high-quality RNA (RIN ≥ 8.9) was used for library preparation. Following adaptor ligation, DNA was PCR amplified for 8 cycles. DNA purification was then performed using AMPure XP beads (Beckman Coulter A63881). The quality of cDNA libraries was analysed on a High Sensitivity D1000

Screen Tape (Agilent). The resulting libraries were then used for cluster generation and sequencing using an Illumina NextSeq 500 (ID: NB501524), with 2 x 35bp read length.

2.4.3 Preparation of DIG-labelled ssRNA probes

Single-stranded DIG-labelled RNA probes were *in vitro* transcribed from 1 µg of template DNA using T7 RNA Polymerase (Sigma 10881767001) and DIG RNA labelling mix (Sigma 11277073910) in the presence of an RNase inhibitor (Sigma 3335399001). The RNA probes were purified by gel filtration on a column based on Sephadex G-50 Fine (GE Healthcare 17004201). See appendix I.III for the primers used to amplify the DNA templates.

2.4.4 RNA in situ hybridization

Tissue slides were deparaffinized and then rehydrated through an ethanol series (100% EtOH x2, 90% EtOH in H₂O). Sections were treated with 1% hydrogen peroxide for 30 minutes to remove pigmentation and block endogenous peroxidase activity. Antigen retrieval was performed by soaking sections in HistoVT One solution (Nakalai Tesque 06380-05) in a glass vessel, which was then placed in a boiling water bath for 20 minutes. Immediately after this treatment, the sections were washed with distilled water to remove detergents. Slides were next treated with acetic anhydride (Wako chemicals) to acetylate amino residues and prevent nonspecific binding of the RNA probe. Sections were incubated with pre-hybridisation buffer (50% formamide, 5X SSC (750 mM NaCl and 75 mM sodium citrate) and 40 µg/mL salmon sperm DNA (Invitrogen)) at 65°C for 2 hours. Hybridization of RNA probes to target mRNA in tissue sections was achieved by incubation with the appropriate hybridization mix (50% formamide, 5X SCC, 500 µg/mL salmon sperm DNA (Invitrogen), 5X Denhardt's solution (0.1% Ficoll (type 400), 0.1% polyvinylpyrrolidone and 0.1% BSA), 1mM DTT, 250 µg/mL tRNA (Roche 10109495001) and the desired dilution of RNA probe preparation) at 65°C overnight. Following hybridization, slides were washed extensively with SSC at 65°C and then incubated for 30-60 minutes with blocking buffer (0.5% (w/v) blocking reagent (GE Healthcare 10106424) in TBS). From this point on, slides were protected from light. Sections were incubated for 30-60 minutes with 1/1000 alkaline phosphatase-coupled Anti-DIG antibody (Roche 11093274910) in blocking buffer. Slides were washed thrice with TBS and once with Alphos buffer (100 mM Tris pH9.5, 100 mM NaCl, 5 mM MgCl₂).

Detection of Anti-DIG-bound hybridized probe was achieved by incubation with detection solution (100 mg/mL NBT and 50 mg/mL BCIP in Alphos buffer) overnight, or until colour development was deemed sufficient. Slides were washed once with Alphos buffer, thrice with TBS and once with 70% ethanol. Slides were air-dried before mounting with Immu-Mount (Thermo Fisher 9990402).

2.5 Protein methods

2.5.1 Preparation of whole cell protein lysates and western blotting

Cells were harvested by scraping, washed three times in PBS and then lysed by resuspension in ice-cold High Salt buffer (50 mM Tris-HCl, pH 7.2, 300 mM NaCl, 0.5% (v/v) NP-40, 1mM EDTA pH7.4, 2 µg/mL aprotinin, 1 µg/mL leupeptin, 1mM PMSF). Cell suspensions were sonicated and then rotated at 4°C for 20 minutes to ensure more complete lysis. Lysates were next clarified by centrifugation at $\geq 20,000$ RCF at 4°C for 20 minutes. Bradford Assays were performed to determine the concentration of protein lysates, and samples were then normalised to the same concentration. Denatured protein lysates were separated on SDS-PAGE (in-house), Bolt Bis-Tris (Invitrogen), NuPAGE Bis-Tris (Invitrogen) or NuPAGE Tris-Acetate (Invitrogen) gels and transferred to nitrocellulose (0.2 µM, Amersham) or PVDF (0.45 µM, Merck Immobilon-FL) membranes; immunoblotting for NSD1 specifically required NuPAGE gel chemistry. Membranes were blocked against nonspecific binding and then incubated with the relevant primary (overnight at 4°C) and secondary (1 hour at room temperature) antibodies (see appendix II). Relative protein levels were determined by chemiluminescence or infrared fluorescence detection on a LI-COR Odyssey Fc.

2.5.2 Crude cellular fractionation

Cells were resuspended in 400µL (two volumes) pre-extraction buffer (20 mM HEPES pH 7.2, 0.5% Triton X-100, 50 mM NaCl, 3 mM MgCl₂, 300 mM Sucrose, 2 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 mM PMSF) and incubated on ice for 30 minutes. One volume of cell suspension was removed and labelled “total extract”. The remaining lysate was clarified at $\geq 20,000$ RCF at 4°C for 10 minutes. The supernatant (one volume) was kept and labelled “soluble”. The insoluble pellet was washed once in 1 mL of pre-extraction buffer before resuspension in one volume of pre-extraction buffer – this suspension was labelled “insoluble”. All samples were diluted with one volume of 2X SDS PAGE sample

buffer (150 mM Tris pH 6.8, 1% SDS, 3% (v/v) glycerol, 15% (v/v) β -mercaptoethanol, 1.8% (w/v) Bromophenol blue) and boiled at 100°C for 7 minutes before being subject to four ten-second rounds of sonication at 40% amplitude (Branson Sonifier SFX 150).

2.5.3 FLAG co-immunoprecipitation

Cell pellets were resuspended in an appropriate volume of high salt buffer (50 mM Tris-HCl pH 7.2, 300 mM NaCl, 0.5% (v/v) NP-40, 1 mM EDTA pH7.4, 2 μ g/mL aprotinin, 1 μ g/mL leupeptin, 1mM PMSF) and sonicated. Lysates were rotated at 4°C for 20 minutes before dilution with one volume of no salt buffer (50 mM Tris-HCl pH 7.2, 0.5% (v/v) NP-40, 1 mM EDTA pH7.4, 2 μ g/mL aprotinin, 1 μ g/mL leupeptin, 1 mM PMSF). The lysates were then clarified at $\geq 20,000$ RCF at 4°C for 20 minutes. Protein lysates were pre-cleared by rotating with equilibrated mouse IgG-coupled beads (Sigma) for one hour at 4°C. Input samples were taken from precleared lysates, and the remaining lysate was incubated with equilibrated anti-FLAG beads (gift from Dr Claudio Ciferri) and ≥ 250 units of Benzonase nuclease (Sigma), rotating at 4°C overnight. The beads were then washed five times with wash buffer (50 mM Tris-HCl pH 7.2, 150 mM NaCl, 0.5% (v/v) NP-40, 1 mM EDTA pH7.4, 2 μ g/mL aprotinin, 1 μ g/mL leupeptin, 1mM PMSF). FLAG-tagged immunoprecipitated peptides were then eluted from the beads by shaking incubation at 25°C with elution buffer (wash buffer with 250 μ g/mL 3xFLAG peptide (Sigma)). Samples were centrifuged at $\geq 20,000$ RCF and the eluate supernatant was transferred to a new tube and denatured in an appropriate sample buffer for western blot.

2.6 Bioinformatics methods

2.6.1 Evolutionary conservation analysis

EZH2 protein sequences from the named species (Fig. 4.1) were collected from the UniProt and NCBI protein databases (Bateman et al., 2021; Sayers et al., 2021). Protein sequences were aligned using the EMBL-EBI high speed multiple sequence alignment programme MAFFT (Katoh and Standley, 2013; Madeira et al., 2019). The protein alignment data was analysed using the Jalview Desktop application (version 2.11.1.4; Waterhouse et al., 2009).

2.6.2 ChIP-Rx data analysis

ChIP-Rx data analysis was performed by Dr Craig Monger. Chromosome names for the human genome were modified with the prefix hg38_ and a metagenome was created by concatenating the mouse and human reference genomes (mm10 and hg38, respectively) before indexing with Bowtie 2 (version 2.3.5; Langmead and Salzberg, 2012) as described (Orlando et al., 2014). Reads were aligned to the metagenome using Bowtie 2 with default parameters, with the exception of `--no-unal`. Non-unique read alignments were filtered out using SAMtools (version 1.10; Li et al., 2009) to exclude those with an alignment quality of <2 , and the hg38_ prefix appended to chromosome names was used to separate reads as aligned to the reference mouse or spike-in human genomes. SAMtools was used to convert SAM files to BAM files and to remove duplicate aligned reads. Spike-in normalization factors were calculated for each ChIP using the formula for normalized reference-adjusted reads per million (RRPM) as described (Orlando et al., 2014; 1 per million spike-in reads). BigWig files were generated using the bamCoverage tool from the deepTools suite (version 3.5.0; Ramírez et al., 2016) with a bin size of 10 and scaled using the RRPM normalization factor derived for each sample. BAM files were converted to BED format using the bamToBed function in BEDTools (version 2.27.1; Quinlan and Hall, 2010) and used for peak calling with HOMER (version 4.10; Heinz et al., 2010), using BED files from input ChIP-Rx as a control. The computeMatrix function in deepTools was used to score the coverage of BigWigs in genomic regions of interest prior to generating average and tornado plots using the plotProfile and plotHeatmap functions respectively.

The mm10 reference genome was split into 10-kb windows using the makewindows function in BEDTools, and regions in the ENCODE mm10 blacklist were removed using bedtools intersect with the `-v` and `-wa` flags (Amemiya et al., 2019). BEDTools coverage was used to summarize overlapping read alignments in BAM files within these bins and scaled using RRPM normalization. Resulting counts were used for the generation of chromosome-wide-coverage line graphs and XY density scatter plots.

Promoter regions were identified by taking $-1,000$ bases upstream and $+100$ bases downstream of annotated transcription start sites in the Ensembl 98 mm10 genome annotation. Active promoters were defined by overlapping H3K27ac peak sets with gene promoter regions defined within the mm10 build of the mouse genome. Active enhancers were defined as H3K27ac ChIP-Rx peaks from macs2 that were distal ($\geq 5\text{kb}$) to a gene body. PRC2 target promoters were defined by overlapping SUZ12 peak sets with gene

promoter regions defined within the mm10 build of the mouse genome. Gene body regions were defined by taking the gene annotations as defined in the Ensembl 98 mm10 genome annotation and converted to the bed format using a custom script. The BEDTools complement tool was used to generate a BED file of intergenic regions by supplying the gene body BED file a FASTA index of the mm10 genome as input (generated using SAMtools faidx).

2.6.3 RNA-seq data analysis

Total RNA-seq data analysis was performed by Dr Craig Monger. RNA-seq reads were pre-processed to remove the universal Illumina sequencing adapter ‘AGATCGGAAGAG’ using cutadapt (version 3.2; Martin, 2011). The STAR (version 2.7.3a; Dobin et al., 2013) algorithm was used to align reads to the mouse reference genome (mm10) in two-pass mode, enabling the identification of novel splice junctions during alignment. SAM files were converted to sorted BAM files using SAMtools (version 1.10). Read alignments were assigned to gene identifier features annotated in the Ensembl 98 mm10 genome annotation using featureCounts (version 1.6.4; Liao et al., 2014). The DESeq2 R package (version 1.24.0; Love et al., 2014) was utilized to identify differentially abundant transcripts at a fold change of >1.5 with a baseMean expression of >20 and a Benjamini–Hochberg-adjusted P value of <0.05. Batch effects were reduced by the inclusion of the formula “~ batch + mutation status” to the design of the linear model.

Lcor/Pali1 and *Lcorl/Pali2* RNA-seq data analysis was performed by undergraduate student Tom Payne. StringTie (Pertea *et al.*, 2015) was used for the assembly and quantification of transcripts from mouse RNA-seq data obtained from the ENCODE database (see Table 2.2; Davis et al., 2018; Dunham et al., 2012). Ballgown (Frazee *et al.*, 2015) was used for differential transcript expression analysis and the formatting and export of gene and transcript expression data, represented as a table of FPKM values. These data were transformed to log2(FPKM+1) values and utilised to compare the expression of LCOR with PALI1/PALI2 throughout embryonic development represented with scatterplots and Pearson correlations in Figure 3.11.

2.6.4 Pathway and Process Enrichment Analysis

The gene ontologies were obtained using the Metascape database (Zhou et al., 2019) through the Metascape web interface and were visualized in Figures 3.12 and 4.12 as follows. The terms with the strongest enrichments are shown in the Figure, with p values

represented using a colour scheme ranging from red (lowest p values) to blue (highest p values). For each term the number of genes in the different categories is written in brackets. Circle size represents the percentage of the total number of genes in each GO category. Gene ontologies obtained using the Metascape database [<http://metascape.org>].

Table 2.2 Mouse RNA-seq datasets obtained from ENCODE

Each ENCODE RNA-seq dataset features two isogenic replicates. The datasets available and used in this thesis (Fig. 3.11B) are presented in the table below, with the tissue and developmental stage of origin indicated along the top and side, respectively. A tick indicates a dataset available on ENCODE and used in this thesis, while unticked boxes indicate a developmental stage/tissue of origin combination not represented on ENCODE.

	Brain	EFP	Neural Tube	Liver	Kidney	Lung	Placenta
E10.5	✓	✓					
E11.5	✓	✓	✓	✓			
E12.5	✓	✓	✓	✓			
E13.5	✓	✓	✓	✓			
E14.5	✓	✓	✓	✓	✓	✓	
E15.5	✓	✓	✓	✓	✓	✓	
E16.5	✓			✓	✓	✓	
p0	✓			✓	✓	✓	
A8w				✓	✓	✓	✓

The brain includes separate fore-, mid- and hindbrain datasets; EFP = embryonic facial prominence; E = embryonic day; p0 = postnatal day 0; A8w = 8-week-old adult.

CHAPTER 3:
ANALYSIS OF THE PRC2-ASSOCIATED
PAL1 PROTEINS

3.1 Introduction

C10ORF12 (the uncharacterised protein product of the human chromosome 10 open reading frame 12, *C10ORF12*) was identified as a novel interactor of core PRC2 over the last several years (Alekseyenko et al., 2014; Smits et al., 2013). C10ORF12 was also observed to specifically co-immunoprecipitate with the PRC2 accessory protein PHF1 (Alekseyenko et al., 2014; Conway et al., 2018). The ligand-dependent co-repressor (LCOR) protein was similarly demonstrated to interact with core PRC2 and PHF1 (Conway et al., 2018; Smits et al., 2013). Like *C10ORF12*, *LCOR* is encoded on human chromosome 10. LCOR contains an N-terminal nuclear receptor binding box and functions to repress ligand-dependent transactivation by nuclear receptors; this repressive effect was shown to be histone deacetylase and C-terminal binding protein (CtBP) dependent (Fernandes et al., 2003). A study conducted by the Bracken lab identified a new protein, now called 'PALI1' (PRC2 associated LCOR isoform 1), which encompasses C10ORF12 and the N-terminal portion of LCOR (Conway et al., 2018). Hauri et al. (2016) also published on the discovery of this new *LCOR* gene product under the name of 'LCOR-CRA_b'.

Within the Bracken lab, the discovery of PALI1 followed numerous observations. Firstly, as mentioned above, both LCOR and C10ORF12 peptides were identified to interact with PHF1 by IP mass spectrometric analysis. It was subsequently noted that the *C10ORF12* gene locus is located just ~10kb directly downstream of *LCOR*, a gene known to produce alternative splice variants. It was hypothesised that the LCOR and C10ORF12 peptides detected to interact with PHF1 could arise from an unannotated alternative splicing event from the *LCOR* gene locus, whereby *C10ORF12* is incorporated into the *LCOR* mRNA transcript. This hypothesis was supported by the observation that PCL1-co-immunoprecipitated peptides representing LCOR and C10ORF12 were detected at molecular weights much higher than expected for the individual proteins (Conway et al., 2018). The existence of the predicted splice form was validated by cDNA sequence analysis showing that *PALI1* contains the first two coding exons of *LCOR* followed by the *C10ORF12* open reading frame, separated by a region of *PALI1* unique sequence (see Conway et al. (2018) Fig. 1). It was further determined that *C10ORF12* is only expressed as a part of the *PALI1* mRNA transcript, and that the C10ORF12 protein is unlikely to exist independently of PALI1. The existence of the PALI1 protein was biochemically validated and it was confirmed that it was indeed PALI1 – not LCOR or C10ORF12 peptides – that immunoprecipitated with PHF1 in the initial study that had led to its discovery (Conway et al., 2018).

PHF1 is a member of the Polycomb-like (PCL) family of proteins. PCL proteins are the defining members of PRC2.1 subtype assemblies, crucial for the faithful targeting of core PRC2 to CpG islands (Ballaré et al., 2012; Brien et al., 2012; Hauri et al., 2016; Healy et al., 2019; Højfeldt et al., 2019; Li et al., 2017; Oksuz et al., 2018). PALI1 interacts with these PRC2.1-specific proteins, but not the PRC2.2-specific JARID2 or AEBP2 proteins (Alekseyenko et al., 2014; Conway et al., 2018; Hauri et al., 2016). In other words, PALI1 is a PRC2.1 specific accessory protein. I generated some of the data supporting this conclusion published in Conway et al. (2018) – specifically Figure 2 panel B. PALI1 (PRC2.1) and AEBP2 (PRC2.2) exhibit functional antagonism likely underpinned by competitive binding to core PRC2, similar to recent reports of competition between EPOP (PRC2.1) and JARID2 (PRC2.2) for binding to SUZ12 (Chen et al., 2018; Conway et al., 2018). Interestingly, PALI1 associates with PRC2.1 exclusively of the EPOP accessory protein (Fig. 1.6) suggesting further biochemical complexity within this sub-complex (Alekseyenko et al., 2014; Hauri et al., 2016).

The Bracken lab comprehensively demonstrated that the PALI1 protein associates with PRC2 via the C-terminal portion of the protein that is not shared with LCOR (Conway et al., 2018). A protein homology search of the PALI1 interaction with PRC2 (PIP) domain identified only one homologous human protein which maps to a predicted alternative splice form from the human *LCORL* gene (located on chromosome 4; XM_011513822.2). *LCORL* is a paralogue of *LCOR*. The predicted protein product of this *LCORL* isoform was thus named PALI2, and accordingly its homologous PIP domain was shown to similarly associate with PRC2 (see Conway et al. (2018) Fig. 3B). It is worth noting that PALI1 was also observed to interact with Euchromatic Histone-Lysine N-Methyltransferase 2 (EMHT2, a.k.a. G9A) via a more N-terminal region which is not thought to be shared with PALI2 (Fig. 3.1B). Figure 3.1A presents a schematic of the relationship between LCOR, PALI1, LCORL and PALI2. Figure 3.1B shows the actual evolutionary relationship between LCOR, PALI1, LCORL and PALI2 and highlights the high degree of amino acid sequence conservation within the PIP domains of PALI1 and PALI2. This high degree of conservation supports the hypothesis that the peptide sequences are under functional constraint, which in turn implies that they may be engaged in essential biochemical functions. Moreover, PALI protein orthologues are not found in invertebrates, thus their conserved functions may be vertebrate-specific (Fig. 3.1B).

Potentially essential biochemical functions are beginning to be elucidated for the PALI proteins. We have shown that PALI1 stimulates the histone methyltransferase (HMT) activity of core PRC2 *in vitro* and *in vivo* (Conway et al., 2018): recombinant PALI1

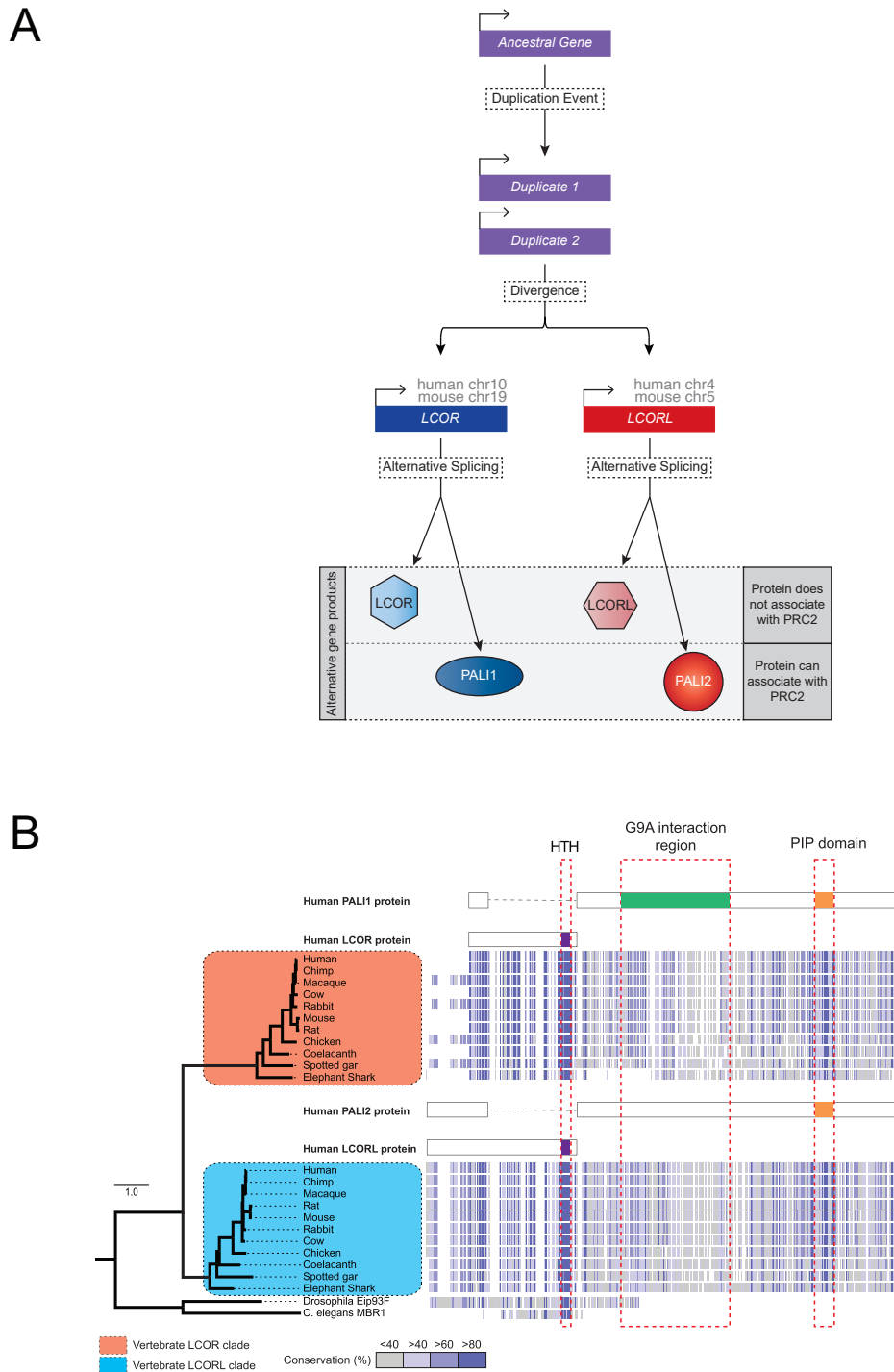


Figure 3.1 The relationship between LCOR, PALI1, LCORL and PALI2

(A) A schematic representation of the relationship between LCOR, PALI1, LCORL and PALI2. *LCOR* and *LCORL* are paralogous genes whose gene products are known to undergo alternative splicing. PALI1 and PALI2 are proteins encoded by the *LCOR* and *LCORL* loci, respectively. The PALI proteins differ from the LCOR and LCORL proteins in their ability to associate with PRC2.

(B) A maximum-likelihood phylogenetic tree (left) inferred from aligned amino acid sequences of LCOR and LCORL exons (right). The percentage of the residues that agree with the consensus sequence is indicated by colour. Positions corresponding to either the human LCOR/LCORL and PALI1/2 proteins are indicated above the alignments. Figure taken from Conway et al. (2018).

protein robustly stimulates core PRC2 HMT activity *in vitro*, whereas mouse embryonic stem cells lacking *Pali1* (or *Pali1* and *Pali2*) show reduced H3K27me3 at PRC2 target genes. The Davidovich lab has since provided insight into the molecular mechanisms through which PALI1 promotes the enzymatic activity of PRC2 (Zhang et al., 2021a). Firstly, they determined that PALI1 is methylated by PRC2 at lysine residue (K) 1241 (and possibly K1219). They additionally showed that this methylated form of PALI1 allosterically activates PRC2 through binding to the aromatic cage of EED in a manner analogous to the allosteric activation of PRC2 by methylated JARID2 (JARID2-K116me3) (Sanulli et al., 2015). Remarkably, they also provided evidence to suggest that PALI2 may be similarly methylated by PRC2 to trigger allosteric activation of the complex. Finally, they showed that PALI1 facilitates PRC2 binding to DNA and nucleosomes independently of its allosteric activation function, which could promote the HMT activity of PRC2 simply by increasing its residency time on chromatin. This finding was consistent with the prior observation that PALI1 is exclusively localised to the chromatin fraction of mouse embryonic stem cells (Conway et al., 2018; Figure 6F).

It is possible that the activation of PRC2 by PALI1/2 constitutes the essential biochemical function which keeps the PIP domains under functional constraint. Consistent with this idea, the methylated residues identified by Zhang et al. (2021a) are amongst those that were 100% conserved in the evolutionary conservation analysis by Conway et al. (2018). Furthermore, the genetic ablation of *Pali1* causes a perinatal lethal phenotype in mice which suggests that PALI1 is essential for mouse prenatal development (Conway et al., 2018). This lethal phenotype supports the hypothesis that the PALI1 protein sequence is under functional constraint. We speculate that the interplay between PALI1 and PRC2 may be a key factor in this functional constraint and in the essentiality of PALI1; as discussed in chapter one, the loss of related PRC2 accessory components have reported similar late embryonic/perinatal lethal phenotypes (Table 1.1; Deevy and Bracken, 2019). It is worth noting that an independent research group has generated a *Lcor*-null mouse model which exhibits placental defects, foetal growth restriction and perinatal lethality (Shalom-Barak et al., 2018). However, these mice were generated from mES cells carrying a *lacZ* trap at the sixth intron of *Lcor*, a modification that would also result in loss of PALI1 function. In other words, there is independent validation that the loss of PALI1 is perinatal lethal.

This chapter presents a follow up study to Conway et al. (2018) with the aim of generating new insight into the biochemical and biological functions of the PALI protein family, with a focus on PALI1. There are three specific aims, the first of which is to determine the localization of PALI1 on chromatin using chromatin immunoprecipitation methods in

mouse embryonic stem cells. Secondly, I aimed to determine whether PALI1 contributes to the recruitment or targeting of PRC2 to chromatin, also using mouse embryonic stem cells, as has been investigated for other PRC2 accessory proteins (Glancy et al., 2021; Laugesen et al., 2019). The third and final aim of this project was to utilize mouse models to provide new insight into the biological functions of PALI1 and PALI2 *in vivo*. This comprised of preliminary characterization of knockout mice lacking one or both PALI proteins.

In this chapter, I provide comprehensive evidence that the PALI1 protein is not amenable to analysis by chromatin immunoprecipitation using mouse embryonic stem cells; I show that neither full-length PALI1 nor its PIP domain are detected on chromatin by ChIP, despite employing an epitope-tagging strategy and testing various combinations of protein-DNA crosslinking agents. I also use new *Pali1* knockout mouse embryonic stem cell models to confirm that PRC2 enzymatic activity is impaired in the absence of PALI1, despite appearing to function as normal at its sites of initial recruitment. I then go on to show that *Pali1/Lcor* and *Pali2/Lcorl* mRNA transcripts are both broadly and largely co-expressed throughout the developing mouse embryo. Finally, I provide the first descriptions of the *Pali2* single knockout and *Pali1/2* double knockout mouse phenotypes and show that the loss of either *Pali1* or *Pali2* is sufficient to impair the histone methyltransferase activity of PRC2, while their combined loss appears to impair H3K27me3 but not H3K27me2 deposition.

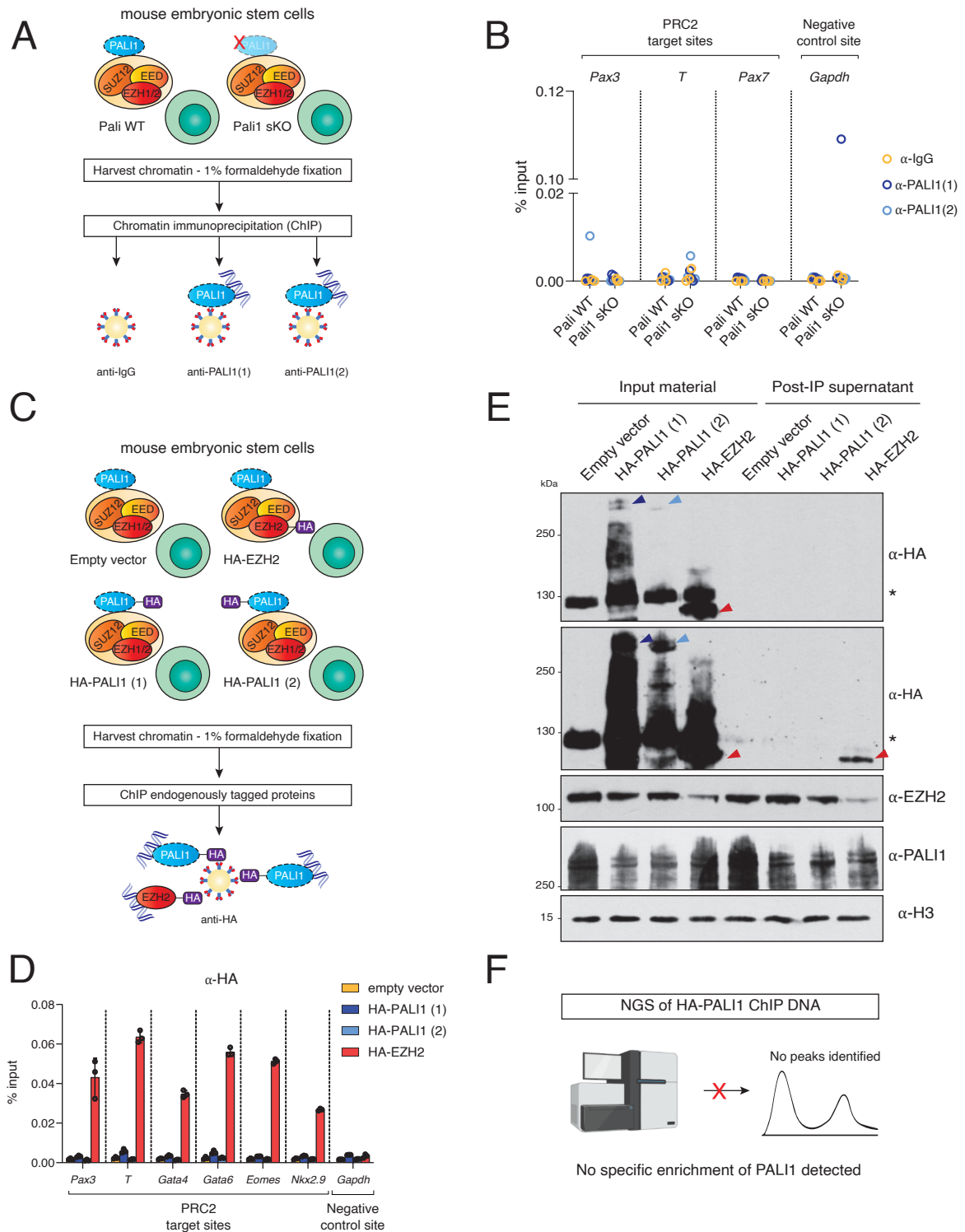


Figure 3.2 PALI1 target genes could not be determined by standard chromatin immunoprecipitation methods

(A) Schematic of the attempted endogenous PALI1 chromatin immunoprecipitation (ChIP) experiment whose data is presented in panel B.

(B) ChIP-qPCR signals for each of two custom α -PALI1 antibodies (dark/light blue) and immunoglobulin G (IgG) control (yellow) in *Pali* wild-type (WT) and single knockout (sKO) cell lines at PRC2 target sites and a control region. Data values represent technical replicates.

(C) Schematic of the HA-ChIP experiment whose data is presented in panels D-E.

(D) ChIP-qPCR signals for α -HA in the indicated cell lines at PRC2 target sites and a control region. Data values represent technical replicates, and their median value is plotted with the range.

(E) Western blot for HA-tagged and endogenous PALI1 and EZH2, probing extracts prepared from the indicated cell lines both before (input material) and after (post-IP supernatant) immunoprecipitation. Arrows indicate the HA-tagged proteins. The asterisks denotes a nonspecific band.

(F) Schematic of the attempted HA-PALI1 ChIP-sequencing experiment that failed to identify enriched peaks.

3.2 Results

3.2.1 The genomic localization of PALI1 could not be determined by standard chromatin immunoprecipitation methods

Ever since the discovery of PALI1 as a PRC2 accessory protein, the Bracken lab have endeavoured to determine the binding profile of PALI1 on chromatin. Conway et al. (2018) showed that PALI1 is both tightly bound to chromatin and largely destabilized in the absence of stable PRC2. Furthermore, it was shown that H3K27me3 is decreased at PRC2 target genes in *Pali1* knockout mESCs. Taken together, we hypothesised that PALI1 co-localises with PRC2 at a subset of its target genes.

We first attempted to investigate this by performing chromatin immunoprecipitation (ChIP) with custom antibodies generated against PALI1-specific antigens, followed by quantitative real-time PCR (qPCR) to search for enrichment at known PRC2 target sites. (Fig. 3.2A). Representative results are presented in Figure 3.2B, whose raw data was generated by Dr Eric Conway. In this particular experiment, two different α -PALI1 antibodies were tested for their ability to ChIP endogenous PALI1 in both wild-type and *Pali1*-KO mouse embryonic stem cells. Unfortunately, the signal for PALI1 at representative PRC2 target genes (*Pax3*, *T* and *Pax7*) was not enriched over background (α -IgG) values, nor was it specifically depleted in the *Pali1*-knockout context. The cumulative data generated by the lab failed to identify an α -PALI1 antibody that could successfully immunoprecipitate PALI1 at PRC2 target genes.

To overcome the technical limitations of our α -PALI1 antibodies, we next engineered mouse embryonic stem cell lines in which the endogenous *Pali1*-specific exon was labelled with a 3xHA-tag. Dr Eric Conway generated two such tagged cell lines, with PALI1 tagged at two different positions: at the C-terminal (HA-PALI1 (1)) or at an internal site (HA-PALI1 (2)). An additional control cell line was generated in which *Ezh2* was also 3xHA-tagged (Fig. 3.2C) – I assisted Dr Conway with the technical validation of these tagged cell lines (data not shown). ChIP-qPCR with α -HA antibodies confirmed that HA-EZH2 was specifically enriched at PRC2 target promoters (*Pax3*, *T*, *Gata4*, *Gata6*, *Eomes*, *Nkx2.9*) but failed to reliably detect either form of HA-PALI1 at the same loci (Fig. 3.2D). However, western blot analysis of the material that went into this ChIP experiment (input) versus what failed to immunoprecipitate with α -HA (post-IP supernatant) shows that HA-PALI1 was indeed depleted which in turn suggests the ChIP was successful (Fig. 3.2E). Taken together, we reasoned that perhaps HA-PALI1 had been ‘ChIP-ed’ successfully and that we were simply looking at the wrong sites in our qPCR analysis. For this reason, we performed

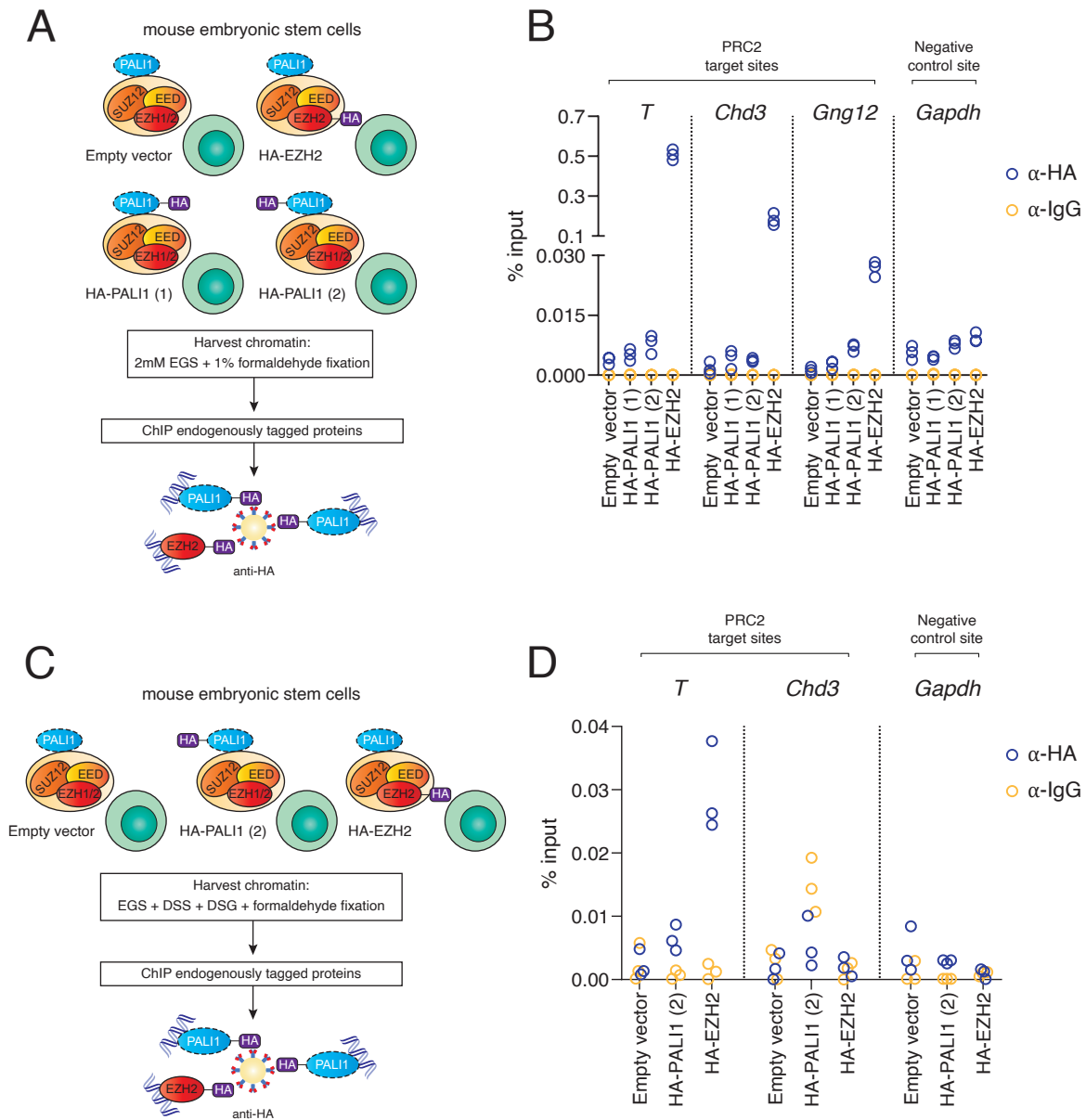


Figure 3.3 The combined use of long and short crosslinkers did not improve our ability to ChIP HA-tagged PAL11

(A) Schematic of the attempted HA-ChIP experiment whose data is presented in panels B.

(B) ChIP-qPCR signals for HA (blue) and immunoglobulin G (IgG) control (yellow) in the indicated cell lines at PRC2 target sites (CGI promoters of *T*, *Chd3* and *Gng12*) and a control region (CGI promoter of *Gapdh*). Data values represent technical replicates.

(C) Schematic of the attempted HA-ChIP experiment whose data is presented in panel D.

(D) ChIP-qPCR signals for HA (blue) and immunoglobulin G (IgG) control (yellow) in the indicated cell lines at PRC2 target sites (CGI promoters of *T* and *Chd3*) and a control region (CGI promoter of *Gapdh*). Data values represent technical replicates.

unbiased next generation sequencing on the HA-PALI1 ChIP DNA. Model-based Analysis of ChIP-Seq (MACS2; Zhang et al., 2008) failed to identify any peaks representing potential PALI1 binding sites from the NGS data (Fig. 3.2F). Due to the successful depletion of HA-PALI1 post-ChIP, we interpreted this negative result as a potential failure to co-immunoprecipitate enough DNA associated with PALI1. In other words, we hypothesised that our chromatin fixation protocol was not effectively crosslinking associated DNA to PALI1.

To this end, I performed a HA-ChIP in one of the HA-PALI1 and the HA-EZH2 tagged cell lines using combined EGS and formaldehyde crosslinking (Fig. 3.3A). EGS is a protein-protein crosslinking agent with a 16.1 angstrom spacer arm. This dual crosslinking strategy has previously successfully aided the chromatin immunoprecipitation of difficult targets including PCGF1 and BAP1 (Conway et al., 2021; Scelfo et al., 2019). The crosslinking was performed as described in section 2.3.3. ChIP-qPCR analysis revealed that the modified protocol could effectively detect HA-tagged EZH2 specifically at PRC2 target genes (*T*, *Chd3*, *Gng12*) but again, there was no specific enrichment for HA-PALI1 at these same loci (Fig. 3.3B). To further increase the crosslinker capture length I next used a comprehensive crosslinking cocktail of EGS, DSS (11.4 angstrom spacer arm) and DSG (7.7 angstrom spacer arm) followed by formaldehyde (Fig. 3.3C). The crosslinking was performed as described in section 2.3.3. ChIP-qPCR analysis indicated that I had once again successfully ChIP-ed HA-EZH2 at a PRC2 target site (*T*) – albeit with a lower signal to noise ratio – with no evidence for the specific enrichment of HA-PALI1 at this same target (Fig. 3.3D). The technical difficulty in capturing PALI1 on chromatin was not resolved with the use of longer crosslinkers.

At this point I felt that perhaps the large (~260 kDa), disordered nature of PALI1 itself was contributing to the technical difficulties I was encountering. I adjusted the approach and attempted to ChIP just a portion of PALI1 rather than the whole protein, to simply interrogate whether PALI1 had the biochemical potential to co-localise with PRC2 at its target sites. As mentioned in the introduction to this chapter, I had previously contributed to mapping and characterising the minimal region of interaction between PALI1 and PRC2 (Fig. 3.4A). My efforts helped to identify a highly conserved region that we termed the PALI1 interaction with PRC2 domain, the so-called ‘PIP’ domain, which was sufficient to co-immunoprecipitate EZH2 in a series of FLAG-IP experiments (Fig. 3.4B). We followed up on this by mutating the two most conserved tryptophan residues within this PIP domain (to alanine – a nonconservative substitution), and I performed the FLAG-IP experiments which showed that this double mutant PIP domain lost almost all ability to associate with

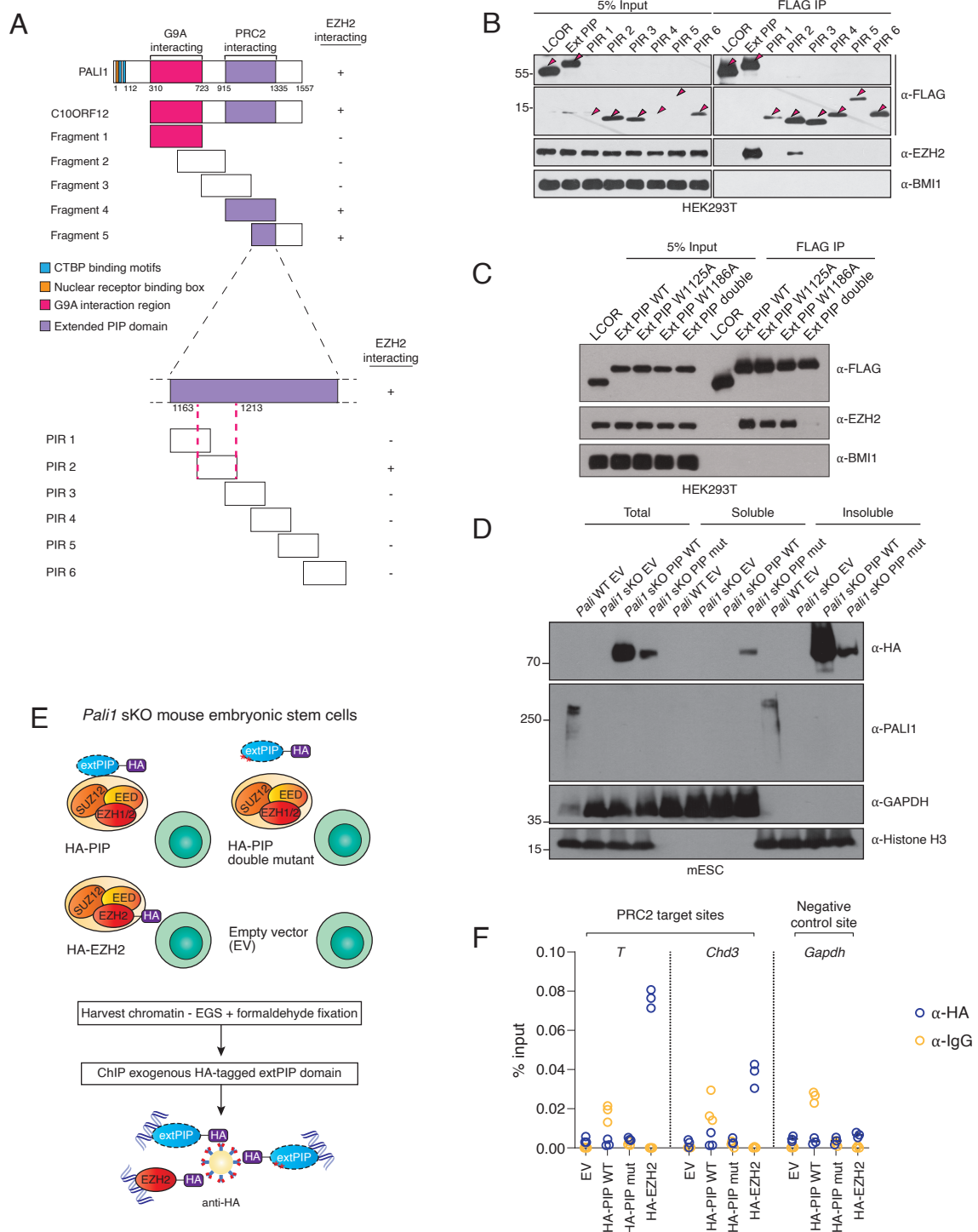


Figure 3.4 The PALI1 interaction with PRC2 (PIP) domain could not be detected at PRC2 target sites by ChIP

Panels A-C are adapted with permission from Conway et al. (2018).

(A) Left: Map of PALI1 and the 5 fragments of PALI1 cloned individually for use in FLAG-IPs in Figure 2B of Conway et al. (2018). An additional 6 smaller PALI1 PRC2 interaction regions (PIR) were subsequently cloned from the extended PIP domain for use in FLAG-IPs. Right: Table indicating the status of interaction of each fragment with EZH2.

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

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

(D) Western blot analysis using the indicated antibodies, probing extracts prepared from the total, soluble and insoluble (≈ chromatin-bound) fractions of the indicated cell lines. PIP WT = wild-type extPIP; PIP mut = double mutant extPIP.

(E) Schematic of the HA-ChIP experiment whose data is presented in panel F.

(F) ChIP-qPCR signals for HA (blue) and immunoglobulin G (IgG) control (yellow) in the indicated cell lines at PRC2 target sites (CGI promoters of *T* and *Chd3*) and a control region (CGI promoter of *Gapdh*). Data values represent technical replicates.

PRC2. (Fig. 3.4C). This highlights the importance of these tryptophan residues as potentially directly interacting with PRC2, or at least being required to maintain the structure of the PIP domain to mediate the interaction.

I reasoned that it might be possible to ChIP the extended PIP (extPIP) domain of PALI1 at PRC2 target sites, since the extPIP domain is a lot smaller and more structured than the full-length PALI1 protein. My experimental strategy was to stably express a HA-tagged version of the wild-type and double mutant extPIP domain in *Pali1*-knockout mESCs, and then to perform a HA-ChIP. I predicted that the wild-type extPIP domain (PIP WT) would associate with PRC2 on chromatin while the double mutant extPIP domain (PIP mut) would not. I chose to perform the experiment in *Pali1*-knockout mESCs to avoid a scenario in which the full-length PALI1 protein out competes the extPIP domain for binding sites on PRC2. I performed a crude fractionation followed by western blot to confirm that my constructs were expressing and predominantly localizing to chromatin (~ the insoluble fraction) (Fig. 3.4D). The results indicated that my constructs were expressed, though the wild-type extPIP expressed to a higher level than the double mutant extPIP. All of the wild-type extPIP protein was indeed found in the insoluble fraction whereas the double mutant extPIP appeared to be divided between the soluble and insoluble fractions. This result was consistent with our prior observation that the double mutant extPIP domain exhibits dramatically reduced affinity for PRC2 (Fig. 3.4C). Taken together, I interpret that the unbound mutant extPIP has lost its association with chromatin and is accordingly lost from the insoluble fraction.

To try ChIP the extPIP I followed a standard HA-ChIP protocol with combined EGS and formaldehyde crosslinking, and I included the abovementioned HA-*Ezh2* mESCs as positive control for the HA-ChIPs themselves (Fig. 3.4E). As before, ChIP-qPCR analysis successfully detected HA-EZH2 at known PRC2 target sites (*T* and *Chd3*), but there was no specific enrichment for the wild-type PIP at these same loci (Fig. 3.4F). I thus concluded that the PIP domain of PALI1 could not be ChIP-ed with this approach.

In summary, neither endogenous α -PALI1 antibodies nor an epitope-tagging strategy, nor the combined use of long and short crosslinkers facilitated the detection of full-length PALI1 by ChIP analysis. A small fragment of PALI1 known to associate with PRC2 also evaded detection on chromatin at PRC2 target sites. Potential technical reasons for this are discussed in section 3.3, but include the use of sonication to fragment DNA, limited antibody availability and protein solubility issues.

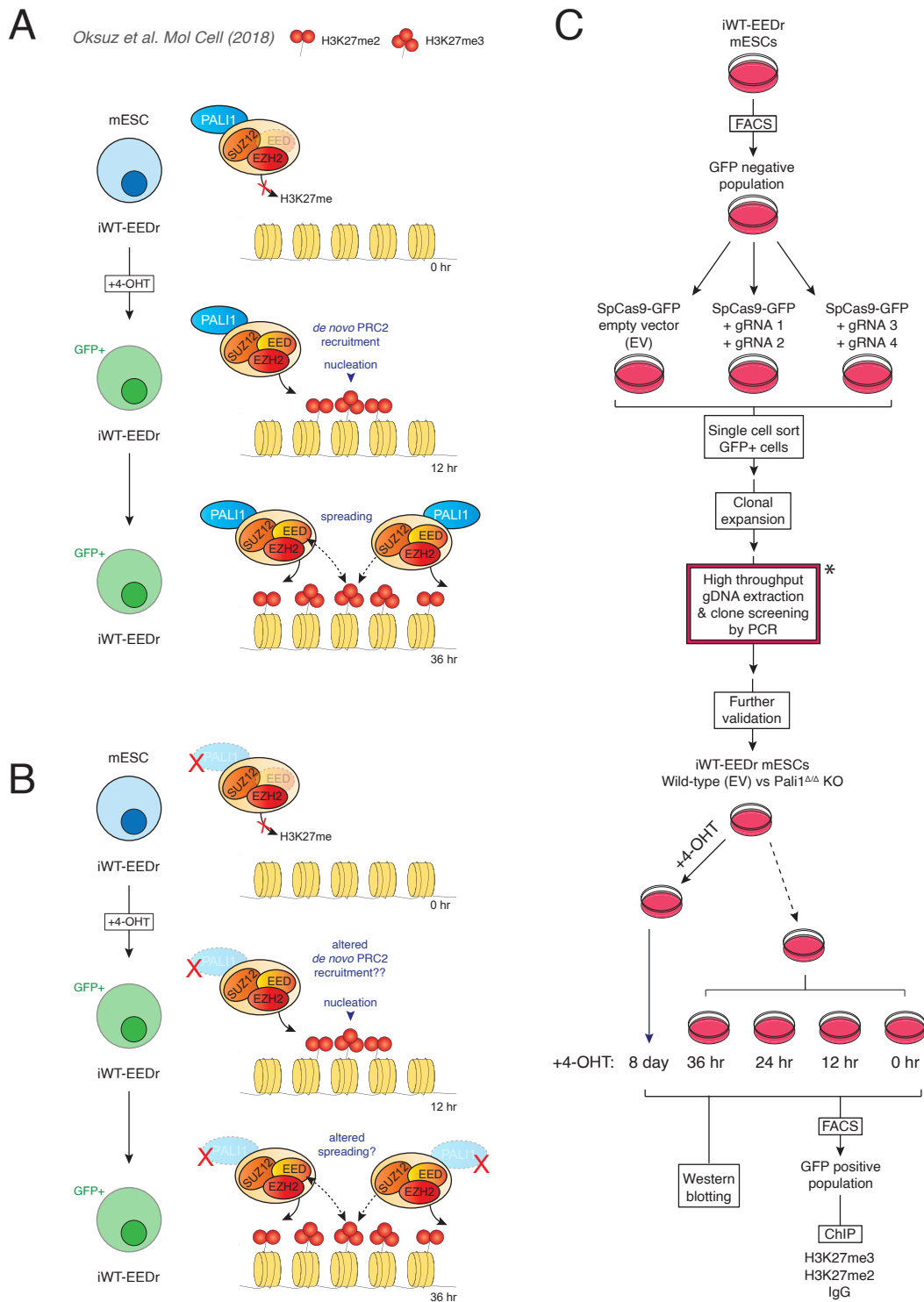


Figure 3.5 Overview of experimental design utilizing the inducible wild-type EED rescue (iWT-EEDr) system in mouse embryonic stem cells

(A) Schematic representation of the inducible wild-type EED rescue (iWT-EEDr) system created by Oksuz et al. (2018). The core PRC2 component EED is reversibly knocked-out in this Cre-lox based mESC model, and cells are propagated until devoid of H3K27me2/3. Wild-type EED is then reintroduced to the system - along with a GFP marker - by tamoxifen-induced expression of Cre-recombinase, and the subsequent re-establishment of PRC2-mediated H3K27 methylation can be tracked over time. See Oksuz et al. (2018) figure 2 for more details.

(B) Schematic representation of our plans to utilize the inducible wild-type EED rescue (iWT-EEDr) system to generate insight into the relative contribution of PALI1 to PRC2 recruitment and spreading. We wanted to test whether (a) PRC2 can be *de novo* recruited to chromatin as normal in the absence of PALI1 and (b) if so, are the spatiotemporal patterns of H3K27me2/3 deposition changed?

(C) Schematic outlining the complete experimental workflow employed to generate (via CRISPR/Cas9 genome-editing), validate and functionally characterize *Pali1* homozygous knockout (*Pali1*^{Δ/Δ} KO) mESCs in the iWT-EEDr cellular context.

*The protocol for high-throughput gDNA extraction and clone screening by PCR was optimized for this workflow.

3.2.2 Interrogating the role of *PAL1* in *de novo* PRC2 recruitment and activity

I next probed another open question regarding the function of *PAL1* in PRC2. In 2018, the Reinberg lab published a study that interrogated the role of the *JARID2* and *MTF2* accessory proteins in *de novo* PRC2 recruitment (Oksuz et al., 2018). These researchers had devised an elegant Cre-lox-based mESC model system in which the core PRC2 component EED was reversibly knocked-out (see Oksuz et al. 2018 Fig. 2). These cells were propagated in the knockout state until they were completely devoid of H3K27me_{2/3}. However, epitope-tagged wild type EED could then be restored by tamoxifen-induced expression of Cre-recombinase, and the subsequent re-establishment of PRC2-mediated H3K27 methylation could be tracked over time (Fig. 3.5A). The authors constitutively knocked-out *JARID2* and/or *MTF2* against this so-called ‘iWT-EEDr’ (inducible wild-type EED rescue) genetic background and then studied the re-establishment of H3K27 methylation in their absence. From this, they could draw conclusions about the relative contributions of *MTF2* and *JARID2* to the *de novo* recruitment and spreading of PRC2 throughout the genome. Professor Reinberg kindly sent us the iWT-EEDr mES cell line so we could ask the same questions of *PAL1* (Fig. 3.5B).

I used CRISPR/Cas9-mediated genome editing to knock-out *PAL1* from the iWT-EEDr cell line. Due to the unique genetic makeup of the iWT-EEDr cells, it was necessary to FACS purify the starting iWT-EEDr population to remove any GFP⁺ cells immediately prior to the genome-editing process (Fig. 3.5C). Figure 3.6A presents the mouse *Lcor* gene locus, highlighting the coding exon specific to the *Pali1* isoform. My knock-out strategy was to homozygously delete the entire *Pali1*-specific coding exon (~5kb) and thereby remove the possibility for any PRC2-interacting protein to be produced by alternative splicing of *Lcor*. This gene targeting approach was designed to leave canonical *Lcor* isoforms intact. A three-primer genotyping PCR reaction was used to screen for positive knock-out clones; this multiplex approach allowed for the simultaneous detection of both the wild-type and deletion allele from the same template gDNA (Fig. 3.6B). I selected two clones with homozygous deletions (*Pali1* Δ/Δ) to carry forward for further screening. RT-qPCR analysis using primers corresponding to either the *Pali1*- or *Lcor*-specific coding exons confirmed that *Pali1*-specific transcripts were absent from the homozygous deletion clones while *Lcor*-specific transcripts remained unchanged or marginally increased in relative abundance (Fig. 3.6C). Finally, western blotting further validated the complete knockout of *PAL1* in both clones (Fig. 3.6D).

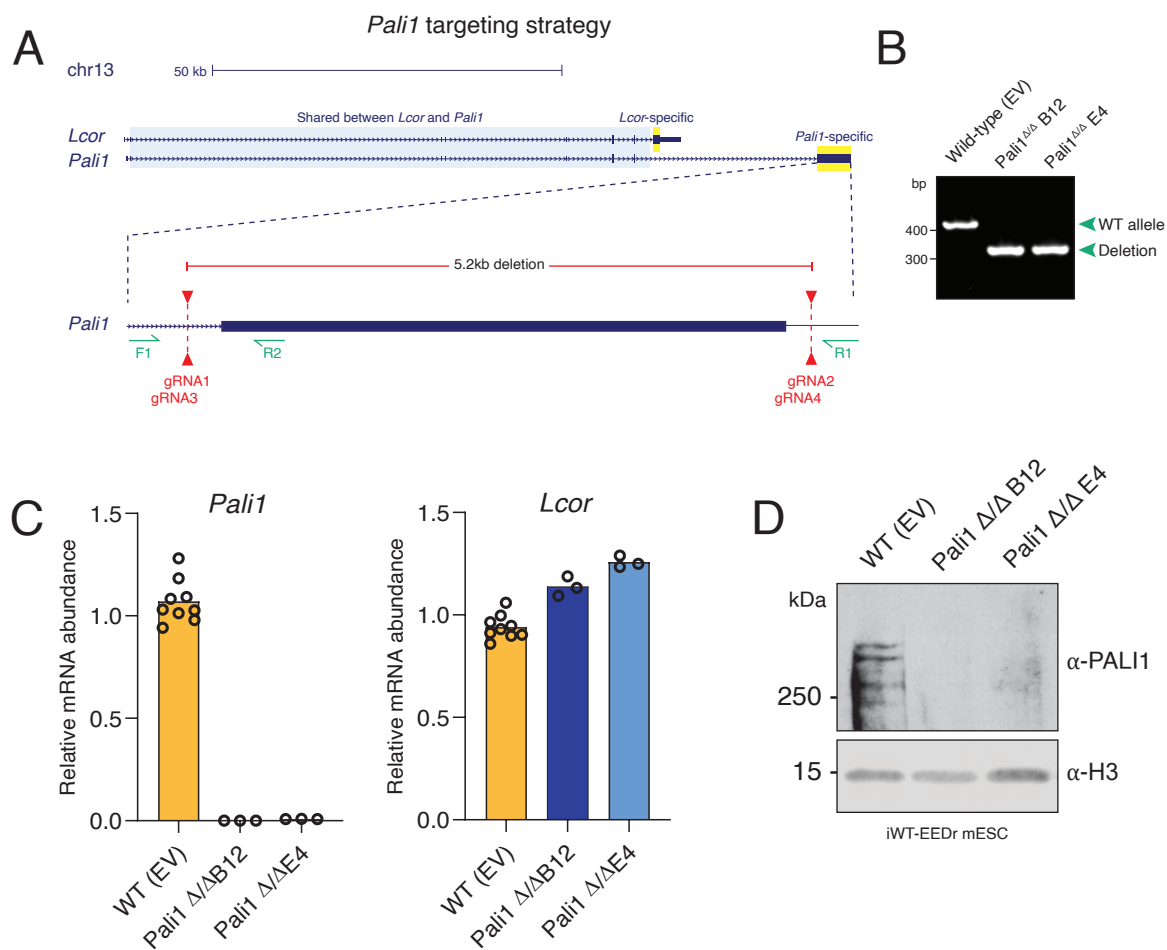


Figure 3.6 Generation and validation of single-cell-derived *Pali1* homozygous knockout clones in the iWT-EEDr cellular context

(A) Top: A graphical depiction of the mouse *Lcor* gene locus showing alternative splicing variants for canonical *Lcor* versus *Pali1*. The exons (coding and non-coding) shared between *Lcor* and *Pali1* isoforms are highlighted in blue, while coding exons specific to either *Lcor* or *Pali1* are labeled and highlighted in yellow. Bottom: A schematic representation of the *Pali1* targeting strategy; paired gRNAs flank the *Pali1*-specific coding exon to induce a 5.2kb CRISPR/Cas9-mediated deletion.

(B) A representative genotyping PCR using the primers indicated in panel A, confirming homozygous knockout of the *Pali1*-specific exon in two independent single-cell-derived clones.

(C) RT-qPCR analysis using primers targeted to either the *Pali1*- or *Lcor*-specific coding exons in wild-type empty vector (WT EV) or *Pali1* homozygous knockout (*Pali1*^{Δ/Δ}) clones. WT (EV) data values come from three biologically independent clones (n=3) with three technical replicate values shown for each. Data values for each biologically independent *Pali1*^{Δ/Δ} clone (n=2) represent three technical replicates. Bars represent the mean value.

(D) Western blot for PALI1, probing whole cell lysates from the indicated cell lines.

I next set up a time-course experiment to re-introduce EED to the system and follow the re-establishment of PRC2-mediated methylations (Fig. 3.7A). I did this in wild-type control (EV) cells and two biologically independent *Pali1* Δ/Δ clones. The time points were at 0 hours, 12 hours, 24 hours, 36 hours, and 8 days post tamoxifen treatment. I tracked the induction of EED expression over time by quantifying the percentage of GFP+ cells in the population at these time points using a Guava InCyte benchtop cytometer (Fig. 3.7B). At 0 hours post tamoxifen treatment, the *Pali1* wild-type control (empty vector, EV) cell line exhibited less than one percent GFP+ cells while the *Pali1* Δ/Δ clones exhibited approximately five to ten percent GFP+ cells. This could be reflective of leaky Cre expression in the genome-edited cell lines, or perhaps due to residual detection of the pSpCas9-GFP plasmid. The proportion of GFP+ cells peaked at 36 hours post tamoxifen treatment in the wild-type cell lines but appeared to peak earlier (between 24–36 hours) in the *Pali1* Δ/Δ clones. Across all conditions, the percentage of GFP+ cells had decreased by 8 days post tamoxifen induction. In parallel, I confirmed the induction of EED expression at each time point by western blot analysis of HA: one of the epitope tags fused to EED (Fig. 3.6D). The western blots data were consistent with the analysis of GFP expression. After validating the efficacy of the system, I turned my attention to the global re-establishment of PRC2-mediated H3K27 methylations (Fig. 3.7C). Western blot analyses indicated that H3K27 methylation was re-established as expected following the rescue with wild-type EED. The average global signal for H3K27me1, H3K27me2 and H3K27me3 was lower in the *Pali1* Δ/Δ clones compared to the wild-type (EV) control at each time point tested (Fig. 3.7E). These data suggest that H3K27 methylation is re-established to a lesser degree the absence of PALI1, and that this deficit is still not overcome by the eighth day of this time-course of this experiment.

Figure 3.8A presents the model of PRC2 nucleation and spreading as proposed by Oksuz et al. (2018); PRC2 is initially recruited to specific CGI-dense “nucleation hubs” where it deposits H3K27me3, before spreading to proximal and then distal genomic locations. I wanted to assess whether PRC2 was changed in its ‘nucleation’ or ‘spreading’ ability in the absence of PALI1. To address this question, I performed ChIP-qPCR of H3K27me3 and H3K27me2, for both the wild-type (EV) control and two independent *Pali1* Δ/Δ clones; a bulk sort (by FACS) was first required to obtain pure populations of GFP+ cells before, to ensure that the population was enriched for cells in which EED had been successfully re-introduced. (Fig. 3.8B). For these ChIP-qPCR analyses, I chose primers localised within two representative “nucleation” sites and two representative “spreading” (proximal) sites (Fig. 3.8 C and D, respectively). Despite the global deficit in *de novo* H3K27me3 observed

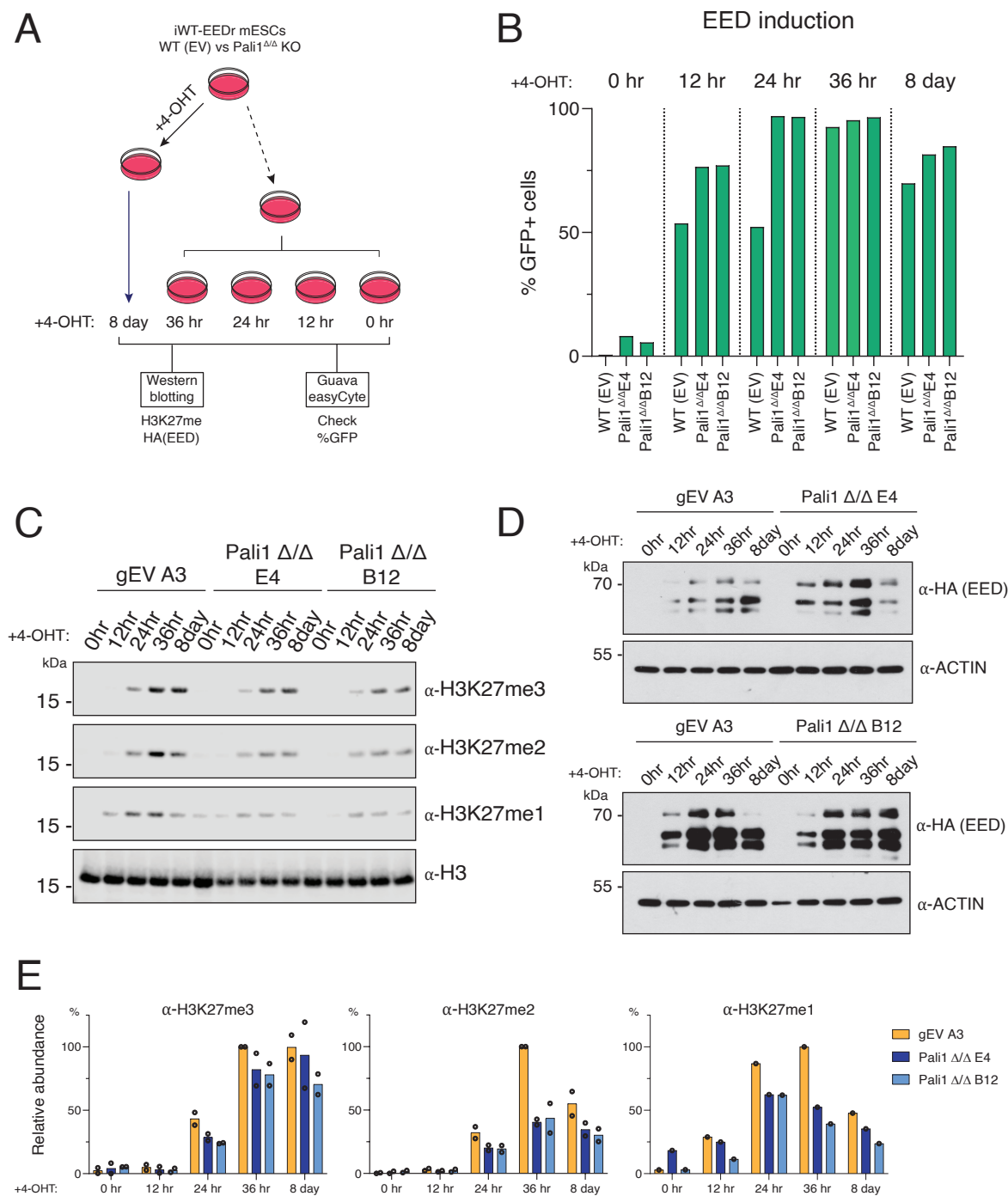


Figure 3.7 Global PRC2-mediated H3K27 methylation is re-established to a lesser degree in the absence of PALI1

(A) Schematic outlining the experimental workflow for the data presented in this figure; wild-type (EV) and *Pali1*^{Δ/Δ} homozygous knockout clones were treated with tamoxifen (+4-OHT) and analysed by flow cytometry and Western blot at 0 hrs, 12 hrs, 24 hrs, 36 hrs and 8 days post-treatment.

(B) A graphical representation of flow cytometry data obtained using the Guava easyCyte 8HT bench cytometer. At each time point post-tamoxifen treatment, the number of detected GFP-positive cells is plotted as a percentage of the total cell population analysed. In this cellular context, GFP is a marker of wild-type EED expression.

(C) Western blotting for H3K27me3, H3K27me2 and H3K27me1, probing whole cell lysates from the indicated wild-type (EV) and *Pali1*^{Δ/Δ} homozygous knockout clones.

(D) Western blotting for epitope-tagged EED, probing whole cell lysates from the indicated wild-type (EV) and *Pali1*^{Δ/Δ} homozygous knockout clones.

(E) Relative quantification of western blot analyses of H3K27 methylation abundance. Within each time-point, signal intensity values were normalized to a loading control and then expressed relative to the wild-type sample (gEV A3). The data were then scaled according to their relative abundance over time, with the 36-hour time point set as 100% intensity. Data values represent technical replicates. Bars represent the mean value.

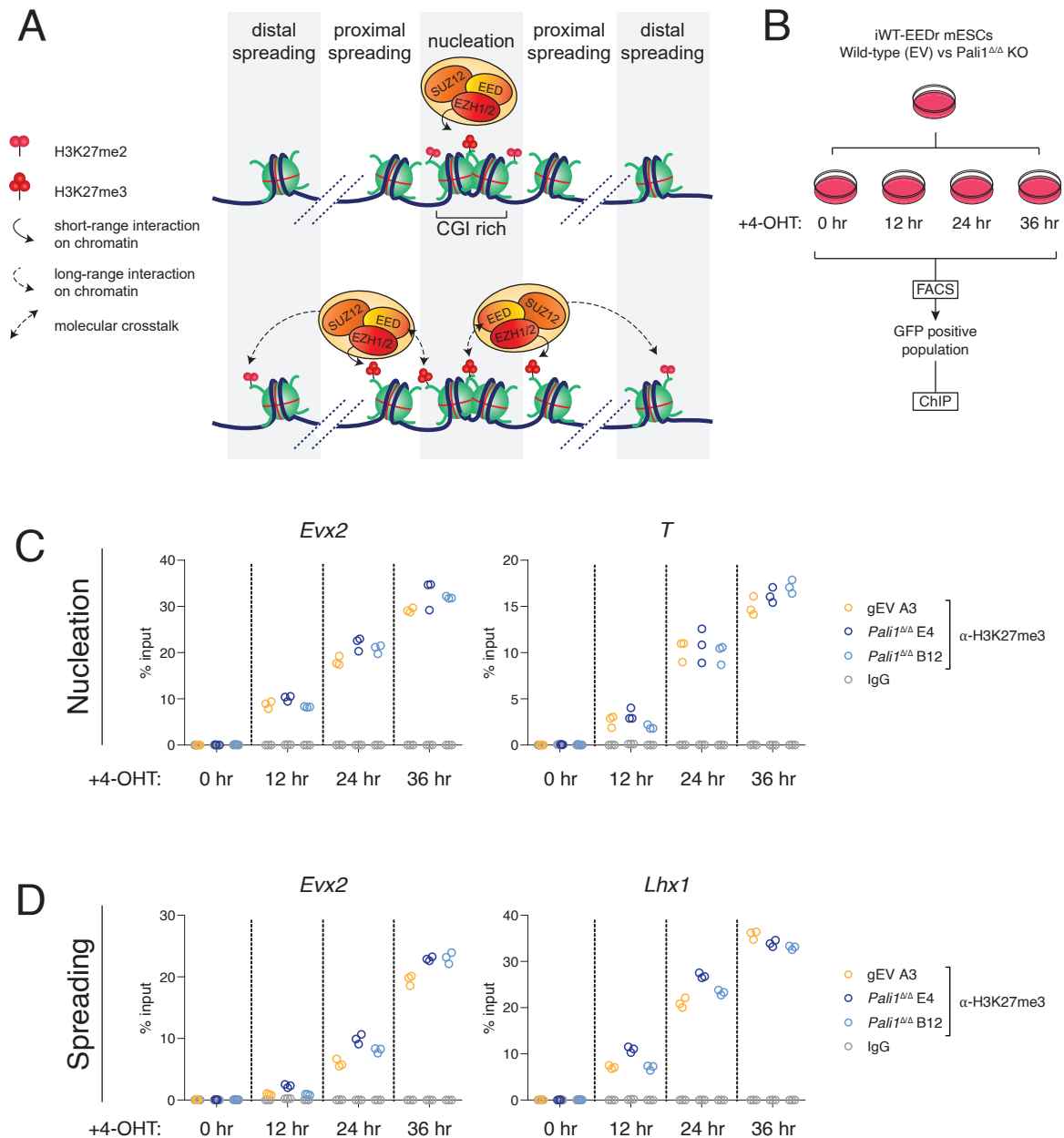


Figure 3.8 The initial nucleation and spreading of PRC2-mediated H3K27me3 is not impaired in the absence of PALI1

(A) Schematic representation of the nucleation and spreading actions of PRC2, based off findings published by *Oksuz et al. (2018)*. In brief, PRC2 is initially recruited to sites with high CGI density (nucleation sites) but then rapidly spreads H3K27me3 and H3K27me2 proximal and distal to the nucleation sites, respectively. CGI = CpG Island.

(B) A schematic outlining the experimental workflow for the data presented in this figure; wild-type (EV) and *Pali1*^{Δ/Δ} homozygous knockout clones were treated with tamoxifen (+4-OHT) and analysed by ChIP at 0, 12, 24 and 36 hours post-treatment.

(C) ChIP-qPCR signals for H3K27me3 and immunoglobulin G (IgG) control in the wild-type (gEV A3) and *Pali1*^{Δ/Δ} homozygous knockout clones at the *Evx2* 'nucleation' site (*Oksuz et al. figure S2E, position C*) and a strong PRC2 target (CGI promoter of *T*). Data values represent technical replicates.

(D) ChIP-qPCR signals for H3K27me3 and immunoglobulin G (IgG) control in the wild-type (gEV A3) and *Pali1*^{Δ/Δ} homozygous knockout clones at PRC2 proximal 'spreading' sites (*Oksuz et al. figure S2E Evx2 position A* and *figure S2G Lhx1 position A*). Data values represent technical replicates.

in *Pali1* Δ/Δ clones by western blot, the ChIP-qPCR data presented in Figure 3.8C indicate that *de novo* PRC2 HMT activity at nucleation sites is not impaired in the absence of PALI1; if anything, H3K27me3 levels appear slightly elevated at the chosen nucleation sites in both *Pali1* Δ/Δ clones compared to wild type. Similarly, there was no evidence for impaired H3K27me3 “spreading” in the absence of PALI1 (Fig. 3.8D). Taken together, it appears unlikely that PALI1 plays a critical role in the initial recruitment of PRC2 to chromatin. Nevertheless, I identified at least two gene loci that showed evidence of reduced *de novo* PRC2 activity in the absence of PALI1 – *Chd3* and *Gng12* (Fig. 3.9). I selected these genes for ChIP-qPCR analysis because they had previously demonstrated reductions in steady-state H3K27me3 in *Pali1* sKO versus wild-type mouse embryonic stem cells (Fig. 3.9A). At 24- and 36-hours post tamoxifen exposure both *Pali1* Δ/Δ clones showed a significant ($p < 0.0001$) reduction in H3K27me3 at *Chd3* compared to the wild-type control (Fig. 3.9B). Similarly, although the absolute signal for H3K27me2 was low, both *Pali1* Δ/Δ clones showed a significant ($p < 0.003$) reduction in H3K27me2 at *Gng12* compared to the wild-type control at the 24- and 36-hour timepoints (Fig. 3.9C). It is noteworthy that these loci once again displayed evidence of impaired PRC2 HMT activity, considering that this experiment was independently conducted against a different genetic background using alternative gene-targeting methodologies to ablate *Pali1*. These data suggest that the previously observed loss of H3K27me3 at these sites in steady-state *Pali1* sKO mESCs is likely a direct consequence of PALI1 inactivation.

Western blot analysis of the *Pali1* Δ/Δ clones had indicated a stark global reduction in the H3K27me2 modification which is broadly deposited at intergenic sites. Thus, I next turned my attention to intergenic DNA regions to assess whether PRC2 activity may be altered at these sites in the absence of PALI1. Of two intergenic sites chosen for analysis (Fig. 3.10A), one displayed significantly ($p < 0.0001$) decreased H3K27me2 levels in both *Pali1* Δ/Δ clones compared to wild type at 36 hours post tamoxifen exposure (Fig. 3.10B, left). At the other site (Fig. 3.10B, right), one clone (B12, light blue) showed significantly ($p = 0.0102$) reduced H3K27me2 signal compared to wild-type at the 36-hour timepoint. Conversely, H3K27me3 signal appeared significantly ($p < 0.0001$) increased relative to wild type in both *Pali1* Δ/Δ clones at both intergenic sites, at the 24- and 36-hour timepoints (Fig. 3.10C). One interpretation of these data is that loss of PALI1 may cause aberrant conversion of intergenic H3K27me2 to H3K27me3.

Taken together, these data independently confirm that the global HMT activity of PRC2 is decreased in the absence of PALI1 but argue against an essential role for PALI1 in the *de novo* recruitment and spreading of PRC2.

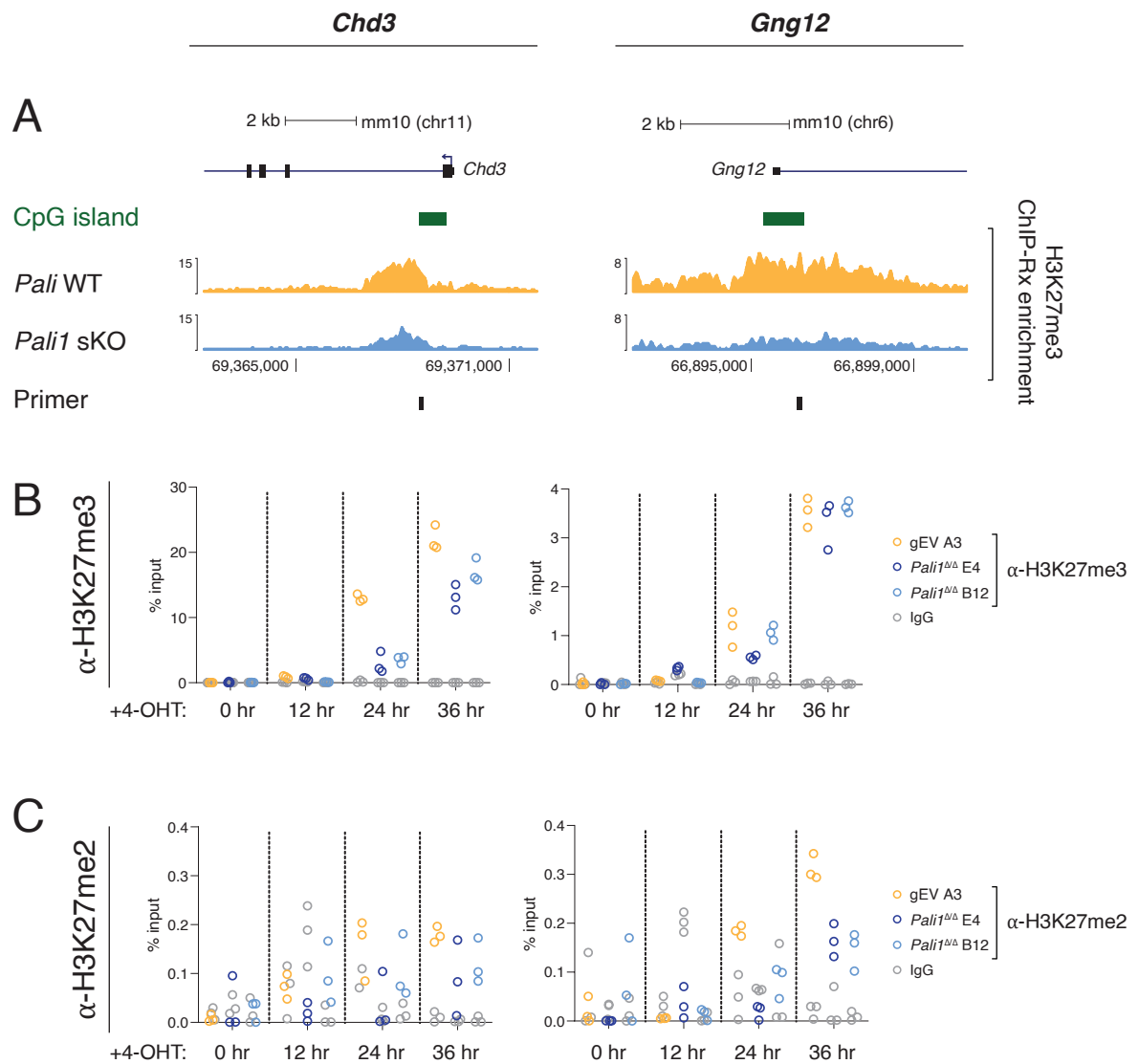


Figure 3.9 Specific loci show evidence of reduced *de novo* PRC2 activity in the absence of PALI1

(A) Genome browser representations of ChIP-Rx normalized reads for H3K27me3 in matched wild-type and *Pali* single knockout mouse embryonic stem cells, showing the locations of the primer pairs utilized in panels B and C. The ChIP-Rx data has been published in Conway *et al.* (2018).

(B) ChIP-qPCR signals for H3K27me3 and immunoglobulin G (IgG) control (grey) in the wild-type (gEV A3) and *Pali*^{Δ/Δ} homozygous knockout clones at two sites previously identified as affected by the loss of PALI1 in steady-state mouse embryonic stem cells. Data values represent technical replicates.

(C) ChIP-qPCR signals for H3K27me2 and immunoglobulin G (IgG) control (grey) in the wild-type (gEV A3) and *Pali*^{Δ/Δ} homozygous knockout clones at two sites previously identified as affected by the loss of PALI1 in steady-state mouse embryonic stem cells. Data values represent technical replicates.

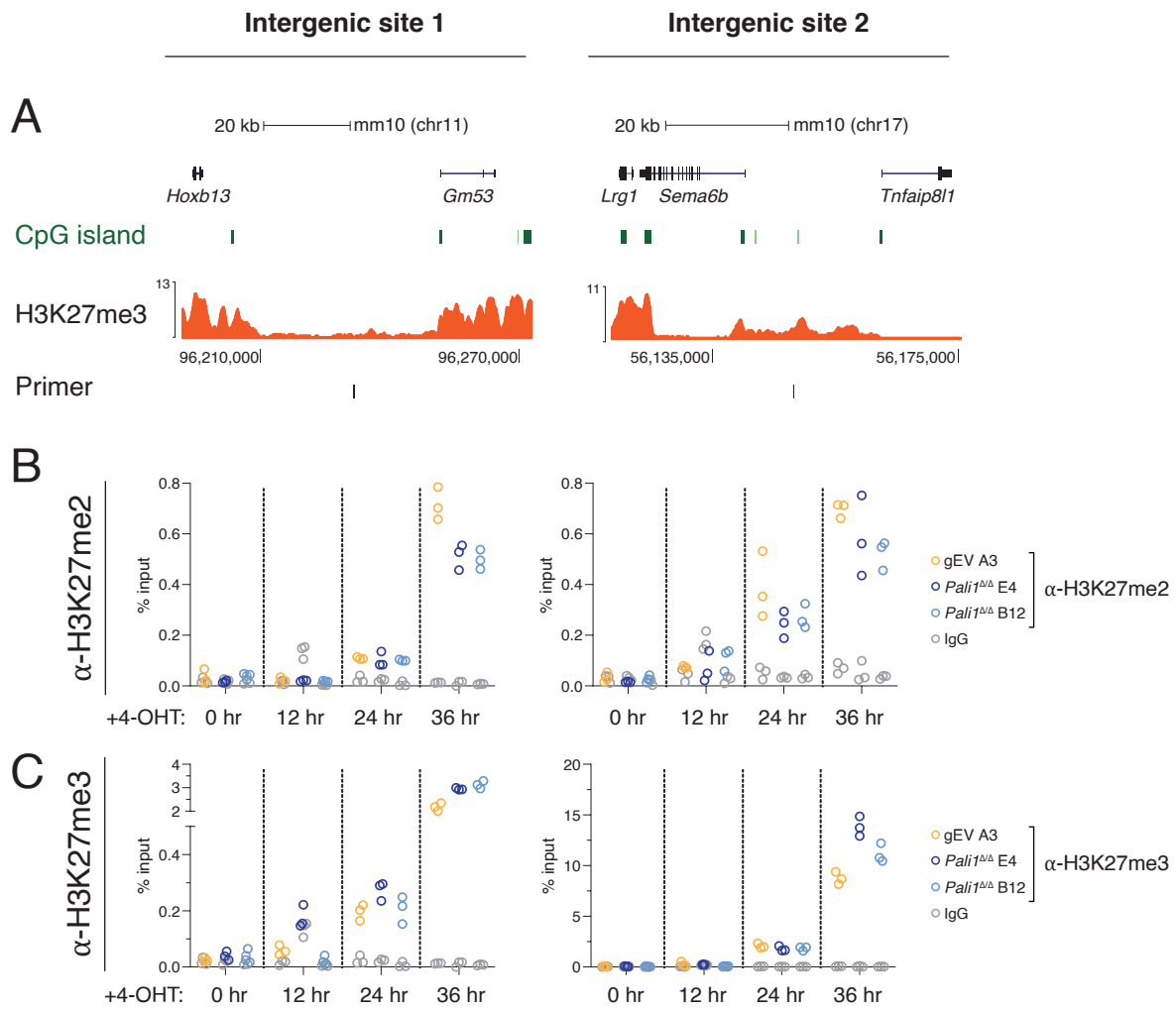


Figure 3.10 *De novo* established intergenic H3K27me3 is increased in the absence of PALI1, with concomitant reductions in H3K27me2

(A) Genome browser representations of ChIP-Rx normalized reads for H3K27me3 in wild-type mouse embryonic stem cells, showing the locations of the intergenic primer pairs utilized in panels B and C. The ChIP-Rx data has been published in Conway *et al.* (2018).

(B) ChIP-qPCR signals for H3K27me3 and immunoglobulin G (IgG) control in the wild-type (gEV A3) and *Pali1*^{Δ/Δ} homozygous knockout clones at PRC2 two intergenic sites. Data values represent technical replicates.

(C) ChIP-qPCR signals for H3K27me2 and immunoglobulin G (IgG) control in the wild-type (gEV A3) and *Pali1*^{Δ/Δ} homozygous knockout clones at PRC2 the same two intergenic sites. Data values represent technical replicates.

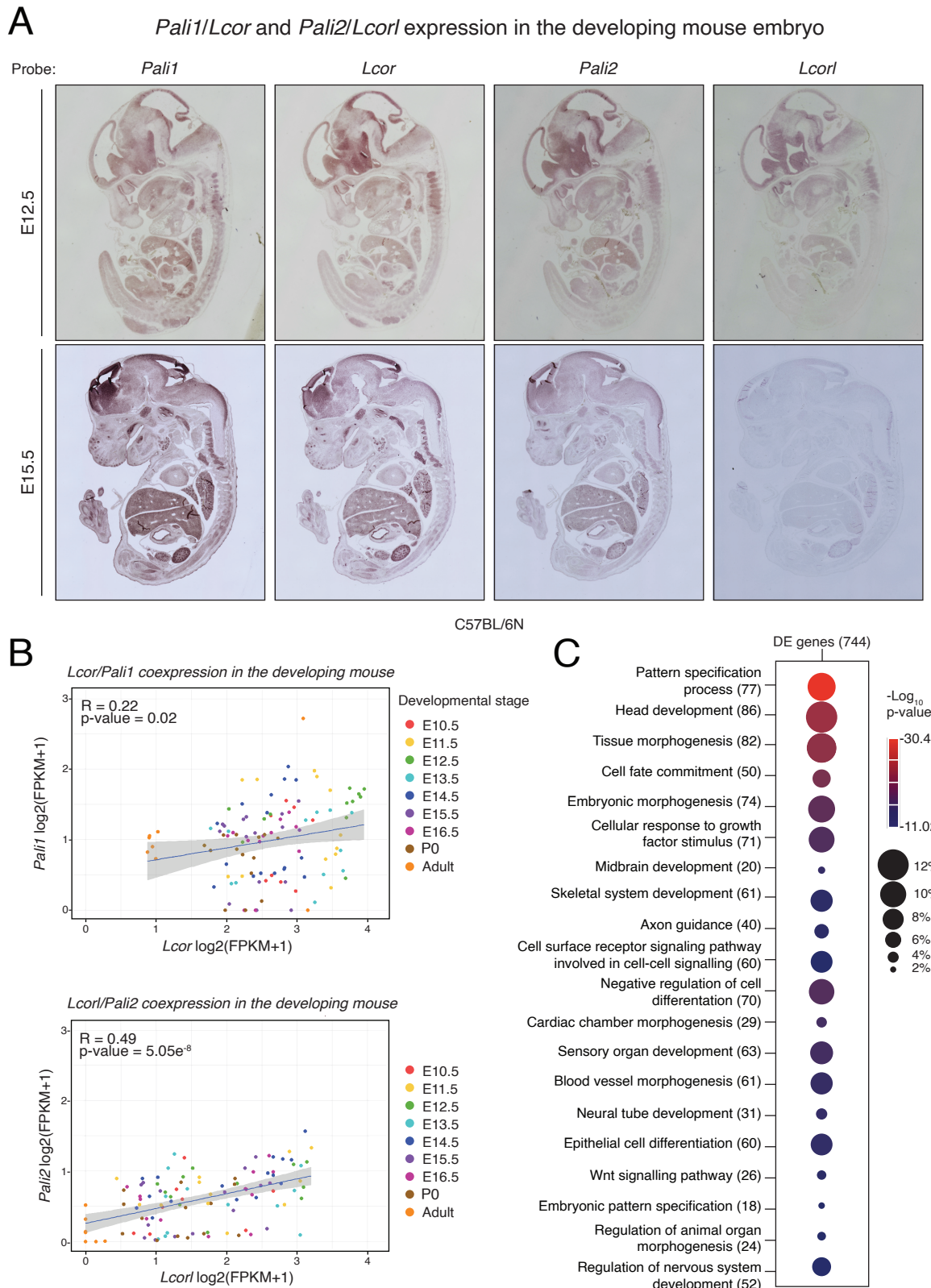


Figure 3.11 Analysis of *Pali* gene expression during mouse development

(A) Qualitative analysis of global *Pali1*, *Lcor*, *Pali2* and *Lcorl* expression as determined by RNA *in situ* hybridization performed on paraffin embedded sagittal sections of the developing mouse embryo at 12.5 (top) and 15.5 (bottom) dpc.

(B) Quantitative bioinformatic analysis of global *Pali1/Lcor* and *Pali2/Lcorl* co-expression in the developing mouse embryo (E10.5-16.5), neonate (P0) and adult (8 weeks old), using publicly available RNA-seq data from the ENCODE database. Each data point is coloured by the developmental stage of the tissue of origin; various tissues are represented within each developmental stage (see section 2.6.3 for details). Pearson correlation and associated p-value indicated in top left of each plot. Grey-shaded area around blue trend-line represents its 95% confidence interval.

(C) Pathway and process enrichment analysis of genes that are differentially expressed in *Pali1* knockout day 8 embryoid bodies compared to wild-type control embryoid bodies at the same stage. See section 2.6.4 for further detail.

3.2.3 Analyses of *Pali* gene expression during mouse development

The PALI1 and LCOR proteins are expressed from the same genetic locus, as are the PALI2 and LCORL proteins (Fig 3.1A). It is not yet known how the expression of these alternative splice forms are regulated during development. This information could potentially provide some insight into their differential biological functions. To address these questions, we employed both bioinformatic and molecular analyses to study the mRNA expression patterns of *Pali1*, *Lcor*, *Pali2* and *Lcorl* during development.

The global mRNA expression patterns of *Pali1*, *Lcor*, *Pali2* and *Lcorl* were qualitatively assessed via RNA *in situ* hybridization (ISH). I performed these RNA ISH experiments at the RIKEN Centre for Integrative Medical Sciences (Yokohama, Japan) in collaboration with the Laboratory for Developmental Genetics (Team Leader: Haruhiko Koseki). I designed single-stranded RNA probes to specifically detect and differentiate between each expressed mRNA transcript (see appendix I.III for details). I performed RNA ISH on paraffin sections taken from wild-type mouse embryos; representative results are presented in Figure 3.11A. *Pali1* appears globally co-expressed with *Lcor* at both E12.5 and E15.5, while *Pali2* appears similarly globally co-expressed with *Lcorl* at both E12.5 and E15.5 – it must be noted, however, that the *Lcorl* probe employed was either very weak or of poor quality. *Pali1*, *Lcor*, *Pali2* and *Lcorl* appear to be ubiquitously expressed at these developmental timepoints. As a control, I also designed and tested probes against the mRNA transcripts of *Pax2*, *Lhx1* and *Six2* – three lineage marker genes whose developmental expression patterns are already well established. Their staining was specific and in the expected patterns (data not shown).

Bioinformatic analysis was performed by Tom Payne (an undergraduate genetics student of at Trinity College Dublin) to examine *Pali1*, *Lcor*, *Pali2* and *Lcorl* mRNA expression across different tissues and timepoints during mouse development, using publicly available RNA-seq data from the Encyclopaedia Of DNA Elements (ENCODE) database (Davis et al., 2018; Dunham et al., 2012). The graphical output of one of his analyses is presented in Figure 3.11B – this is a scatter plot of *Pali1* versus *Lcor* (top) or *Pali2* versus *Lcorl* (bottom) transcript abundance across all tissues and developmental stages analysed, and samples were grouped according to the developmental stage of their tissue of origin. In agreement with my *in vivo* RNA ISH data, this global analysis shows that *Pali1/Lcor* and *Pali2/Lcor* are significantly co-expressed (p value = 0.02 and $5.05e^{-8}$, respectively) during mouse development. The correlation between *Pali2* and *Lcorl* transcript abundance is of moderate strength ($R = 0.49$) whilst the correlation between *Pali1* and *Lcor* transcript abundance is weaker ($R = 0.22$). It is pertinent to note that, for both *Pali1/Lcor* and *Pali2/Lcorl*,

transcript abundances are most tightly clustered and at their lowest in the 8-week-old adult tissues, whilst peak transcript abundances appear to occur at E11.5-E14.5.

In 2018, Conway et al. differentiated wild-type and *Pali1* single knockout (sKO) mESCs into embryoid bodies, which are three-dimensional aggregates of pluripotent stem cells that have been allowed to differentiate in a non-directed manner to give rise to cells representing all three germ layers. RNA sequencing was performed on mRNA isolated from day 8 embryoid bodies (EB8) to reveal what genes were significantly differentially expressed in the *Pali1* sKO EB8 compared to wild type. Figure 3.11C presents a pathway and process enrichment analysis of this list of differentially expressed genes. The most significantly enriched terms relate to important whole-organism early developmental processes such as embryonic morphogenesis, pattern specification and organ development. These results correspond well with the observed ubiquitous expression pattern of *Pali1* during mouse embryonic development (Fig. 3.11A), and the fact that *Pali1* transcript abundance appears higher during embryonic versus adult developmental stages (Fig. 3.11B).

In conclusion, a combined *in vivo* and *in silico* approach show that *Pali1/Lcor* and *Pali2/Lcorl* mRNA transcripts are ubiquitously distributed and spatiotemporally co-expressed throughout the developing mouse embryo.

3.2.4 *Pali1/2* double-knockout mice phenocopy the *Pali1* single-knockout mouse perinatal lethal phenotype, but *Pali2* single-knockout mice do not

To probe the biological function of the PALI proteins through a reverse genetics approach, the Bracken lab collaborated with the Koseki lab to generate *Pali1* single knockout (*Pali1*^{-/-}; *Pali1* sKO), *Pali2* single knockout (*Pali2*^{-/-}; *Pali2* sKO) and *Pali1/2* double knockout (*Pali1*^{-/-}*Pali2*^{-/-}; *Pali1/2* dKO) mouse models (Fig. 3.12A). To date, the Bracken lab have published preliminary data relating to the *Pali1* sKO mouse phenotype only (Conway et al., 2018). Starting with a 3-month placement at the Koseki lab, I sought to investigate the phenotype of the *Pali1* sKO mice further and to perform preliminary characterization of the additional, unpublished knockout mouse strains. Towards this, I successfully pioneered the establishment of a colony of *Pali1* sKO mice at Trinity College Dublin, thereby enabling the Bracken lab to independently continue to study the role of PALI1 *in vivo*.

We previously published data showing that *Pali1*^{-/-} mice exhibit a homozygous recessive perinatal lethal phenotype (Conway et al., 2018). We showed that offspring from a *Pali1*^{+/-} heterozygous cross are found in the expected Mendelian ratios at embryonic day 11.5 (E11.5), but that these ratios are skewed by postnatal day 0 (P0) such that no live *Pali1*

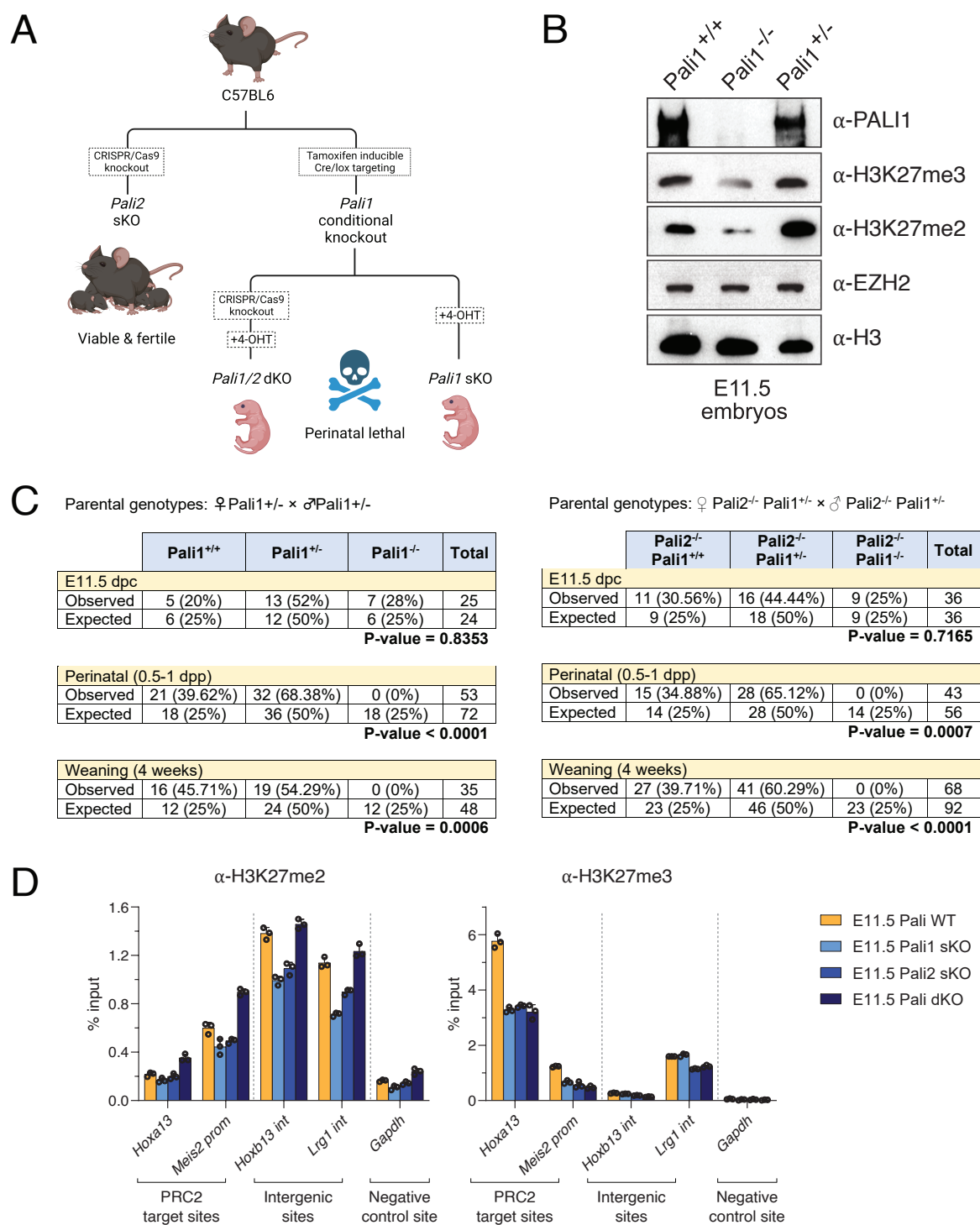


Figure 3.12 Preliminary characterization of *Pali* knockout mice

(A) Schematic outlining the various *Pali* knockout mice generated by the Koseki lab and their associated viability. Created with BioRender.com.

(B) Western blot analysis using the indicated antibodies in wild-type, *Pali1* homozygous knockout and *Pali1* heterozygous knockout embryos at embryonic day 11.5 (E11.5). Reproduced from Conway et al. (2018) with permission.

(C) Left: Tables showing the number of mice of the indicated genotypes obtained at various stages of development following a cross of *Pali1*^{+/-} mice. Right: Tables showing the number of mice of the indicated genotypes obtained at various stages of development following a cross of *Pali2*^{-/-} *Pali1*^{+/-} mice. P-values obtained in Chi-square test for goodness of fit. Significant (p<0.5) values indicate a significant deviation from the expected Mendelian ratios.

(D) ChIP-qPCR signals for H3K27me2 (left) and H3K27me3 (right) in E11.5 embryos of the indicated genotypes at PRC2 target sites, intergenic sites and a negative control region. Data values represent technical triplicates and their mean value is plotted. The significance of the difference between two means was determined by two-way ANOVA with Dunnett's multiple comparisons test. Error bars represent the standard deviation from the mean.

homozygous mutants are recovered. I have reanalysed this data and performed a chi-square test to numerically determine the significance of this skew. The results are presented in Figure 3.12C (left panel). As expected, the deviation of the observed genotypes recovered after birth versus the expected Mendelian ratios was computed as highly statistically significant.

To add to this, I also compiled and analysed new data concerning the genotypic ratios of offspring resulting from a *Pali2*^{-/-}*Pali1*^{+/-} cross. This data was kindly provided by Yoko Koseki and Miho Urabe (RIKEN IMS, Yokohama, Japan). The results of this analysis are presented in Figure 3.12C (right panel). As with the *Pali1*^{-/-} mice, *Pali1/2* double-knockout (dKO; *Pali2*^{-/-}*Pali1*^{-/-}) mice were present in the expected Mendelian ratio at E11.5 but not recovered live from P0 onwards. The discrepancy between the observed and expected genotypic ratios recovered after birth was again deemed highly statistically significant. *Pali2* single-knockout mice (*Pali2*^{-/-}*Pali1*^{+/+}; *Pali2*^{-/-}*Pali1*^{+/-}) are recovered at all stages; *Pali2* single-knockout mice are fertile and appear ostensibly normal and healthy.

In summary, these data establish that *Pali1/2* double-knockout mice phenocopy the *Pali1* single-knockout mouse perinatal lethal phenotype whilst the *Pali2* single-knockout mice are contrastingly viable and fertile.

3.2.5 H3K27 methylation states are altered in *Pali* knockout mice

Global H3K27me2 and H3K27me3 levels were previously observed to be reduced at the protein level in E11.5 *Pali1*^{-/-} embryos versus wild-type and heterozygous knockout controls (Fig. 3.12B). Furthermore, chromatin immunoprecipitation (ChIP) experiments have shown that H3K27me3 is reduced at PRC2 target loci when comparing *Pali1*^{-/-} and *Pali1/2* dKO mouse embryonic stem cells (mESCs) against matched wild-type mESCs (Conway et al., 2018). I decided to extend our analyses of PALI function *in vivo* by checking whether H3K27me3 levels are also specifically decreased at PRC2 targets in E11.5 *Pali* knockout versus wild-type embryos, and testing whether H3K27me2 is concomitantly increased or decreased. I performed a ChIP-qPCR with a protocol optimised for use in embryo tissue. This experiment used three frozen E11.5 embryos from each of the following genotypes: *Pali* WT, *Pali1*^{-/-}, *Pali2*^{-/-} and *Pali1/2* dKO. These embryos were harvested, genotyped, and kindly sent to us by Yoko Koseki (RIKEN IMS Yokohama, Japan).

The results of the ChIP-qPCR analyses are presented in Figure 3.12D. I investigated the levels of H3K27me2 and H3K27me3 at two established PRC2 target loci (CGI promoters of *Hoxa13* and *Meis2*), two intergenic loci (*Hoxb13* int, *Lrg1* int; respectively intergenic

site 1 and 2 from Fig. 3.10), and the transcriptional start site of one active gene (*Gapdh*). Compared to the wild-type embryo, H3K27me3 (right panel) was significantly reduced at the PRC2 target sites by approximately 50% across each of the three *Pali*-mutant genotypes ($p < 0.0001$). H3K27me3 was also reduced at one of two intergenic regions (*Lrg* int) in the *Pali2* sKO and *Pali1/2* dKO mutant embryos ($p < 0.0001$). *Pali1/2* dKO embryos showed significant increases in H3K27me2 at both PRC2 target sites ($p < 0.0001$), along with slight but significant increases in H3K27me2 at both intergenic sites ($p < 0.0264$). Contrastingly, H3K27me2 was significantly decreased at both intergenic sites in *Pali1*^{-/-} and *Pali2*^{-/-} embryos ($p < 0.0001$), and at one of two PRC2 target sites (the *Meis2* promoter) ($p < 0.0016$). Overall, these ChIP-qPCR analyses confirm that H3K27me3 levels are indeed specifically decreased at PRC2 targets in E11.5 *Pali* knockout embryos but also reveal an unexpected divergence of the single and double knockout phenotypes with respect to H3K27me2: the loss of either *Pali1* or *Pali2* reduces H3K27me2 levels relative to wild type, but their combined loss produces the opposite effect.

3.3 Discussion

In this chapter, I have thoroughly demonstrated the inefficacy of chromatin immunoprecipitation methods to detect PALI1 enrichment on chromatin in mouse embryonic stem cells. This is consistent with a recent statement from the Davidovich lab on their inability to identify direct target genes of PALI1 in the K-562 human leukaemia cell line by ChIP (Zhang et al., 2021; [peer review file](#)). These researchers did, however, successfully detect ectopically overexpressed PALI1 at some PRC2 target genes by ChIP-qPCR in the human HEK293T cell line (Zhang et al., 2021; supplementary Figure 10) – demonstrating that PALI1 is at least biochemically capable of co-localizing with PRC2 at its target sites on chromatin. I hypothesise that current ChIP protocols recover insufficient material to reliably detect the enrichment of endogenous PALI1 on chromatin. I speculate that protein solubility issues may factor into this; PALI1 is biochemically a difficult protein to work with due for reasons not yet completely understood. For example, in our hands it is only possible to immunoblot PALI1 from whole cell lysates and using high salt extraction buffers. I made several attempts to overcome perceived technical incompatibilities between PALI1 and the standard ChIP protocol, but to no avail. One potential technical limitation to my methodology was the use of sonication to fragment DNA – sonication is a harsh technique which may damage antibody target epitopes or disrupt PALI1 binding to chromatin. MNase digestion could be used as an alternative, gentler DNA fragmentation method. Another known limitation is the lack of specific antibodies targeting PALI1. Although I can conceive alternative experimental designs to try detect PALI1 on chromatin by ChIP analysis, I believe future studies should utilize alternative methodologies to try identify the target genes of PALI1 on chromatin; examples include DNA adenine methyltransferase identification (DamID; Steensel & Henikoff, 2000; Vogel et al., 2007), cleavage under targets & release using nuclease (CUT&RUN; Meers et al., 2019; Skene et al., 2018) and cleavage under targets & tagmentation (CUT&Tag; Kaya-Okur et al., 2019, 2020). A technical advantage of these methods is that they each circumvent the use of sonication and chromatin crosslinking altogether.

Having said that, it gives one pause to consider that Zhang et al. (2021) reported that CUT&Tag also failed to identify direct target genes of PALI1 in the K-562 human leukaemia cell line. The Davidovich lab showed that PALI1 facilitates both DNA and nucleosome binding by PRC2 (though without apparent DNA-sequence specificity), and they also demonstrated *in vitro* that the PIP domain of PALI1 can bind DNA in the absence of PRC2 (Zhang et al., 2021; supplementary Figure 6C). Mounting evidence suggests that PALI1 will associate with DNA on chromatin. So why then is it so difficult to detect? I

speculate that epigenomic profiling methods are not well suited to the detection of PALI1 on chromatin because it may bind diffusely throughout the genome, without strong peaks of enrichment. I expand on this hypothesis in chapter six (section 6.1.2).

The two *Pali1* knockout mouse embryonic stem cell lines I generated for this project provide independent validation of published findings that PALI1 stimulates the histone methyltransferase activity of PRC2 (Conway et al., 2018; Zhang et al., 2021a). These *Pali1* knockout cell lines differ from those previously reported in several ways: (1) they have been generated using a different mouse embryonic stem cell line, (2) deletion of the *Pali1* specific exon was achieved using CRISPR/Cas9-mediated genome editing rather than Cre-lox-mediated recombination, and (3) the iWT-EEDr genetic background makes it possible to investigate the effects of the constitutive loss of PALI1 on *de novo* recruitment and activity of PRC2. The aim for this project was to utilize these cell lines to determine whether PALI1 contributes to the recruitment or targeting of PRC2 to chromatin *in vivo*. My data does not support a model in which PALI1 is essential for the initial targeting or proximal spreading of PRC2 at CGI-dense nucleation hubs, based off the observation by ChIP-qPCR that normal levels of H3K27me3 accumulate at these sites in the absence of PALI1. This is likely due in large part to functional redundancy between PRC2.1 and PRC2.2, as had been previously shown in mESCs (Healy et al., 2019; Højfeldt et al., 2019). Given the established role of PRC2.1-associated PCL proteins in complex targeting, it is likely that PALI1 functions moreso as a positive regulator of PRC2.1 activity than a specific recruitment factor (Ballaré et al., 2012; Brien et al., 2012; Healy et al., 2019; Højfeldt et al., 2019; Li et al., 2017; Oksuz et al., 2018). Furthermore, with respect to the allosteric activation of PRC2 by methylated PALI1, I hypothesise that JARID2-K116me3 may partially compensates for the loss of PALI1 (Sanulli et al., 2015; Zhang et al., 2021a).

A limitation of this project is the absence of direct analyses of core PRC2 components on chromatin – without these, the *de novo* recruitment of PRC2 can only be inferred from the observed deposition of H3K27 methylation. However, the histone methyltransferase activity of PRC2 is known to be diminished in the absence of PALI1. As such, it is flawed to assume that the degree of PRC2 recruitment can be correlated to H3K27 methylation levels in the absence of PALI1; it is only possible to infer the recruitment or non-recruitment of PRC2 in this way. My data indicates that PRC2 is recruited to chromatin in the absence of PALI1 but does not rule out a function for PALI1 in targeting or tethering PRC2 to chromatin. An unbiased genome-wide quantitative ChIP-sequencing experiment of core PRC2 along with H3K27 methylations is required to properly address the question set out at the start of this project. Unfortunately, due to practical time constraints faced in the

pursuit of this doctoral research, the project was discontinued. The relevant cell lines have all been cryopreserved should future researchers wish to revisit this question.

Nevertheless, my data provide new insight into the consequence of PALI1 depletion in mouse ES cells. I have identified two candidate target genes of PALI1 (*Chd3* and *Gng12*) whose transcription start sites show specifically decreased H3K27 methylation in both mine and the previously published *Pali1* knockout cells (Conway et al., 2018). I have additionally provided evidence to suggest a role for PALI1 in establishing the correct balance of H3K27 methylations at intergenic regions in mouse embryonic stem cells. Finally, it is worth stating that in undertaking this project I optimized our existing lab protocol for the high throughput screening of clonal cell lines, and the streamlined protocol has since benefited several projects within our research group (including those discussed in chapters four and five).

To the best of my knowledge, the mRNA expression analyses presented in Figure 3.11 provide the first investigation of *Pali1* versus *Lcor* and *Pali2* versus *Lcorl* mRNA expression patterns throughout the developing mouse. We found these transcripts to be largely co-expressed in terms of both physical location and developmental stage. The data suggest that *Pali1/Lcor* and *Pali2/Lcorl* are broadly expressed throughout the developing embryo and predominantly during embryonic stages (particularly during organogenesis, ~10-14 dpc). These findings are consistent with the known perinatal lethal phenotypes of *Pali1/Lcor* knockout mice and the presented (Fig. 3.11C) pathway and process enrichment analysis of genes that are differentially expressed in *Pali1* knockout embryoid bodies compared to wild type (Conway et al., 2018; Shalom-Barak et al., 2018). In addition, Fernandes et al. (2003) similarly found *Lcor* to be widely expressed in foetal tissues. The RNA *in situ* hybridization analyses could be improved by redesigning the probes to differentiate between nascent and mature mRNA transcripts: this could be achieved by designing them to cross exon-exon boundaries. It would also be beneficial to test these probes on *Pali* knockout mice for a gold-standard verification of their specificity. Additional analyses that would complement and strengthen this dataset include RT-qPCR amplification of *Pali1*, *Lcor*, *Pali2* and *Lcorl* transcripts from mRNA obtained from embryos of a wider sample of developmental timepoints, and matched immunohistochemical (IHC) staining for the translated proteins; the former is possible as a follow-up experiment while the latter is not yet possible due to a lack of specific IHC-grade antibodies. These additional analyses would add to my work to solidify the observation that *Pali1/Lcor* and *Pali2/Lcorl* are broadly expressed throughout the developing embryo.

The preliminary characterization of the various *Pali* knockout mice has generated valuable new insight. I have provided evidence that the *Pali2* single knockout mice are viable and fertile whereas both the *Pali1* single knockout and the *Pali1/2* double knockout mice exhibit completely penetrant perinatal lethality. This suggests that PALI1 can functionally compensate for the loss of PALI2, but that PALI2 cannot functionally compensate for the loss of PALI1 (like EZH2 and EZH1 (Table 1.1)). One potential reason for this could simply be that *Pali2* is more lowly expressed compared to *Pali1* (Fig. 3.11B). Intriguingly, I show that the histone methyltransferase activity of PRC2 on chromatin is approximately equally impaired in both the *Pali1* and *Pali2* single knockout mice. Therefore, an alternative interpretation is that the essential functions of PALI1 in mouse prenatal development are not shared with PALI2. If this were the case, then the essential functions of PALI1 might not rely solely on its interaction with PRC2. Considering that the G9a-interacting region of PALI1 is not thought to be shared with PALI2, I hypothesise that the loss of the interaction between PALI1 and the G9a histone methyltransferase (Fig. 3.1B) may be a key factor in the perinatal lethal phenotype observed in *Pali1* single knockout and *Pali1/2* double knockout mice (Conway et al., 2018). Specifically, I hypothesise that PALI1 links PRC2 and G9a activity in the context of non-canonical genomic imprinting (Hanna and Kelsey, 2021; Raas et al., 2021). Given that I pioneered the establishment of a colony of *Pali1* single knockout mice at my university, another PhD student from the Bracken lab is now actively pursuing this hypothesis.

In addition, my ChIP-qPCR data from the knockout mice models provide the first direct evidence that PALI2 functions as a positive regulator of PRC2 activity *in vivo*, consistent with recent *in vitro* analyses that show PALI2 has the biochemical potential to function as an allosteric activator of PRC2 (Zhang et al., 2021a). The finding in E11.5 *Pali1* single knockout and *Pali1/2* double knockout embryos that H3K27me3 is reduced at PRC2 target genes corresponds well with previous observations made in mouse embryonic stem cells (Conway et al., 2018). I also show that intergenic H3K27 methylation levels are altered in *Pali1* single knockout mice, which aligns with my data from iWT-EEDr *Pali1* knockout stem cells (Fig. 3.10).

Overall, the pursuit of the research presented herein provided me with excellent scientific training and sparked my interest in the study of PRC2 in development, which in turn lead me to pursue the research projects presented in chapters 4 and 5.

CHAPTER 4:
MOLECULAR CHARACTERISATION OF
EZH2-MUTANT WEAVER SYNDROME

4.1 Introduction

As introduced in chapter one, Weaver syndrome (WS) is a rare congenital overgrowth and intellectual disability (OGID) syndrome, so-called as it was first described by Weaver et al. in 1974. Almost four decades later, *EZH2* was identified as the gene mutated in Weaver syndrome (Gibson et al., 2012; Tatton-Brown et al., 2011). Since then, approximately fifty individuals with *EZH2* mutations and Weaver syndrome have been reported in the literature (Tatton-Brown et al., 2013).

The study of these individuals has revealed that Weaver-syndrome-associated *EZH2* mutations are predominantly missense variants; the majority of these occur in the catalytic SET domain of EZH2, but some fall within the CXC and SANT domains (Tatton-Brown et al., 2013). Rare truncating (stop-gain) mutations in *EZH2* have also been reported, but they all fall within its last exon and therefore are likely to escape nonsense mediated decay at the mRNA level (Hentze and Kulozik, 1999). Indeed, such a case was recently reported and it was shown that the global EZH2 protein level in the Weaver syndrome patient sample was comparable to control samples (Imagawa et al., 2017). The dearth of reported stop-gain mutations in more N-terminal exons suggests that simple haploinsufficiency is unlikely to be the major mutational mechanism of Weaver syndrome. Having said that, there is a report of an individual who is haploinsufficient for *EZH2* and exhibits some of the features of Weaver syndrome, including overgrowth and intellectual disability (Suri and Dixit, 2017).

OGID-associated pathogenic variants of *EZH2* have been predicted to be predominantly loss-of-function (Cohen et al., 2016; Imagawa et al., 2017; Lui et al., 2018; Tatton-Brown et al., 2013). However, it is my opinion that alternative mutational mechanisms cannot be ruled out in the absence of more thorough biochemical characterization. To date, three published studies have explored the functions of Weaver syndrome-associated EZH2 mutant proteins (Cohen et al., 2016; Imagawa et al., 2017; Lui et al., 2018). Collectively, they determined that Weaver syndrome mutant EZH2 variants exhibit impaired global histone methyltransferase activity both *in vitro* and *in vivo*. The two initial studies did not delineate between the effects on PRC2-mediated H3K27me1, H3K27me2 or H3K27me3 (Cohen et al., 2016; Imagawa et al., 2017). The third and most recent study involves the first described mouse model of Weaver syndrome (Lui et al., 2018). CRISPR-mediated genome editing was used to generate mice harbouring a Weaver syndrome patient-specific missense mutation targeting the SET domain of EZH2 (p.Val626Met). Mice homozygous for this mutation exhibit perinatal lethality, which contrasts against the early embryonic lethal phenotype of homozygous *Ezh2* loss-of-function mice (Table 1.1) This observation argues against complete loss-of-function of the Weaver syndrome mutant EZH2. Mice

heterozygous for the *Ezh2* (p.V626M) mutation are viable and born at the expected Mendelian ratio. Previously reported heterozygous *Ezh2* loss-of-function mice were not born at the expected Mendelian ratio, but exhibited fewer than expected live births (O'Carroll et al., 2001). Again, the implication is that the Weaver syndrome *Ezh2* (p.V626M) mutant manifests a milder phenotype than complete *Ezh2* loss-of-function. Nevertheless, both the homozygous and heterozygous *Ezh2* (p.V626M) mutant mice exhibit globally decreased H3K27me2 and H3K27me3 (H3K27me1 levels were not examined). Taken together, the data supports a partial loss-of-function phenotype for the Weaver syndrome mutant EZH2. Importantly, the heterozygous *Ezh2*(p.V626M) mice also exhibit postnatal overgrowth – a key feature of Weaver syndrome.

PRC2 function is frequently deregulated in cancer, and *EZH2* is often mutated in haematological malignancies in particular (Conway et al., 2015; Laugesen et al., 2016). This begs the question of whether people with Weaver syndrome are at higher risk of developing PRC2-associated cancers. It seems probable – the available evidence suggests that they are more prone to developing haematological malignancies – but not enough data is yet available to accurately determine risk; a relatively small number of Weaver syndrome cases have been reported to date and most of these individuals are still at too young an age to manifest malignancy (Tatton-Brown et al., 2013). However, neuroblastoma is a cancer that most commonly occurs in early childhood, with an estimated frequency of 10.2 cases per million children under the age of 15 (Colon and Chung, 2011). At least five cases of neuroblastoma have been reported in children with Weaver syndrome (Cohen et al., 2016; Coulter et al., 2008; Tatton-Brown et al., 2011), with a sixth case described in a child harbouring a Weaver-syndrome-associated mutation (Polonis et al., 2018). In light of this, it is probable that more individuals with Weaver syndrome will develop haematological malignancies as they age. The mutational spectrum of *EZH2* in Weaver syndrome is mostly non-overlapping with its somatic mutational profile in cancer (Laugesen et al., 2016; Tatton-Brown et al., 2013, 2017). For example, the highly recurrent hyperactivating *EZH2* mutations commonly observed in haematological malignancies have not yet been detected in Weaver syndrome (McCabe et al., 2012; Sneeringer et al., 2010; Tatton-Brown et al., 2017). There is, however, some overlap with the loss-of-function *EZH2* mutations observed in myeloid malignancies and leukaemia (Ernst et al., 2010; Tatton-Brown et al., 2017). Therefore, a key motivation for the thorough molecular study of WS-associated *EZH2* mutations is the fact that this information may help clinicians more accurately predict what cancers their Weaver syndrome patients may be predisposed to.

The goal of this research was to define the mutational mechanism(s) of Weaver syndrome. To this end, the primary aim of this research chapter was to use mouse embryonic stem cells to model a wide variety of Weaver-syndrome-associated *EZH2* mutations, and to examine their effects on all PRC2-mediated H3K27 methylation states. In this way, I aimed to provide a more comprehensive biochemical characterization of the enzymatic activity of WS-mutant *EZH2* than has previously been achieved. On top of this, I aimed to provide additional layers of information by examining not just the global levels but also the genome-wide distributions of H3K27 methylation states using quantitative chromatin immunoprecipitation. Furthermore, to my knowledge, only one publication to date has examined a chromatin modification other than H3K27 methylation in the context of Weaver syndrome (Choufani et al. (2020) studied the DNA methylomes of Weaver syndrome patients). Therefore, an additional aim of this research chapter was to explore the consequences of any observed perturbations to PRC2 function on the wider chromatin and transcriptional landscape.

In this chapter, I generate isogenic cellular models of ten different Weaver-syndrome-associated *EZH2* mutations and show that, while the variant *EZH2* proteins are indeed partial loss-of-function, they predominantly exert dominant negative effects on global PRC2 enzymatic activity. I also identify a WS-associated *EZH2* variant that paradoxically exhibits H3K27 hyper-trimethylation at PRC2 target genes, resonant of the change-of-function mutations observed in diffuse large B-cell and non-Hodgkin lymphomas. These findings argue against haploinsufficiency as the major mutational mechanism for *EZH2* mutant Weaver syndrome, and instead suggests functional heterogeneity of WS-mutant *EZH2*. Furthermore, I provide evidence that H3K27 hyper-acetylation is a common feature of *EZH2*-mutant Weaver syndrome cell models. Finally, I identify a core subset of 43 genes whose differential transcript abundance is shared across different Weaver syndrome models, and amongst which one expects to find key drivers of the Weaver syndrome phenotype.

Table 4.1 Weaver-syndrome-associated *EZH2* mutations included in this study

Substitution mutation details are given at both the DNA and protein level: genomic coordinates refer to genome assembly GRCh38/hg38, while amino acid changes are named according to EZH2 isoform 1 (Q15910-1, 746 amino acids) with the equivalent change to EZH2 isoform 2 (Q15910-2, 751 amino acids) given in parentheses. The number of cases (#Cases) for each mutation refers to the total number of individuals with Weaver syndrome identified with this mutation and reported in the literature to date. A mutation is classed as recurrent *de novo* if it has been observed in two or more unrelated individuals.

DNA change (genomic location)	Protein change	Affected Protein Domain	#Cases	Mutation notes	References
C>T (chr7:148809370)	p.R679C (p.R684C)	SET domain	8	Occurs in myeloid malignancies, linked to neuroblastoma, recurrent <i>de novo</i> in Weaver syndrome.	Cyrus et al. (2019); Ernst et al. (2010); Polonis et al. (2018); Tatton-Brown et al. (2011)
A>G (chr7:148811657)	p.K634E (p.K639E)	SET domain	1	Affects putative H3K36 sensing pocket of EZH2.	Jani et al. (2019); Tatton-Brown et al. (2011)
G>A (chr7:148811696)	p.V621M (p.V626M)	SET domain	2	Recurrent <i>de novo</i> .	Lui et al. (2018); Tatton-Brown et al. (2013)
A>G (chr7:148814108)	p.K563E (p.K568E)	CXC domain	1	Affects DNA-binding amino acid of EZH2.	Finogenova et al. (2020); Poepsel et al. (2018); Tatton-Brown et al. (2013)
A>T (chr7:148827194)	p.D233V	SANT domain	2	Inherited familial mutation.	Tatton-Brown et al. (2013)
C>G (chr7:148810407)	p.S647C (p.S652C)	SET domain	2	Recurrent <i>de novo</i> .	Tatton-Brown et al. (2013)
G>A (chr7:148809376)	p.A677T (p.A682T)	SET domain	1	Patient developed neuroblastoma and ALL (13 mo).	Cohen et al. (2016); Tatton-Brown et al. (2011)
G>A (chr7:148807669)	p.E740K (p.E745K)	Post-SET domain	1	Patient developed NHL (12 yr).	Cohen et al. (2016); Tatton-Brown et al. (2011)
C>G (chr7:148807703)	p.Y728* (p.Y733*)	SET domain	1	Mutated in myeloid malignancies, escapes NMD (?).	Ernst et al. (2010), Table S3; Tatton-Brown et al. (2011)
A>G (chr7:148829746)	p.K156E	SRM domain	6	Inherited familial mutation.	Tatton-Brown et al. (2011); Tatton-Brown et al. (2013)

NMD = Nonsense-mediated decay. ALL = Acute Lymphoblastic Leukaemia. NHL = Non-Hodgkin Lymphoma.

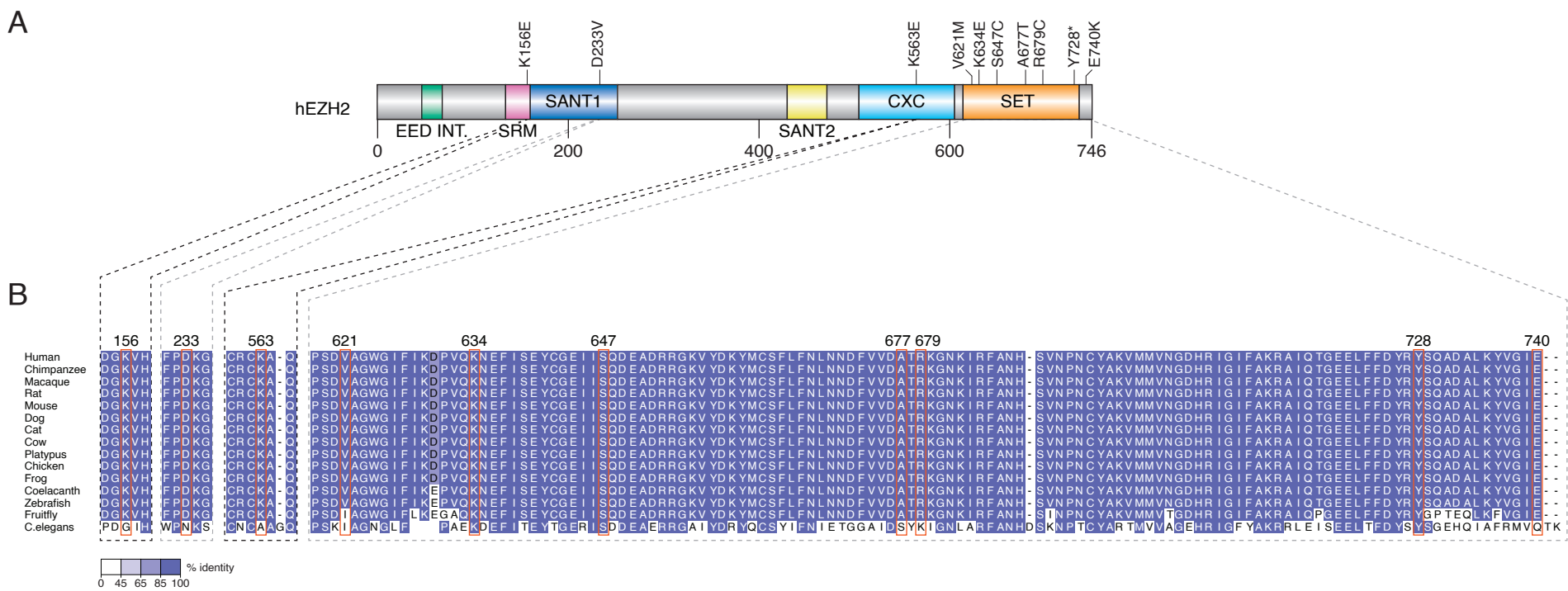


Figure 4.1 Selected Weaver syndrome mutations

(A) Schematic of the human EZH2 protein (isoform 1; Q15910-1) with domains indicated, showing the location and 2D distribution of Weaver-syndrome-associated mutations included in this study. EED int. = EED-interacting domain; SRM = stimulatory response motif.

(B) Alignment of EZH2 protein sequences from the indicated species, highlighting amino acids mutated in Weaver syndrome and included in this study. The percentage amino acid identity is included as a measure of evolutionary conservation.

4.2 Results

4.2.1 Selection of Weaver syndrome-associated *EZH2* mutations for cellular disease modelling

For this study, I decided to model ten different Weaver-syndrome-associated *EZH2* mutations representing approximately one fifth of the causative variants reported to date (Fig. 4.1A; Table 4.1). As mentioned above, most of these variants are missense, though rare truncating mutations occurring in the last exon of *EZH2* have also been reported. Accordingly, I chose to study 9 missense mutations and 1 truncating mutation. All Weaver syndrome mutations included in this study occur at highly conserved residues of the *EZH2* protein (Fig. 4.1B). Weaver syndrome-associated missense mutations predominantly affect the catalytic C-terminal region of *EZH2* comprising the pre-SET (CXC), SET and post-SET domains. However, several mutations also fall within the N-terminal domains which interact with EED (SBD, SRM, SANT1) (Cyrus et al., 2019a). For comprehensiveness, I opted to study missense mutations occurring in the SRM, SANT1, CXC, SET and post-SET domains.

Most cases of Weaver syndrome are sporadic, but rare familial cases have been reported which exhibit an autosomal dominant pattern of inheritance (Fryer et al., 1997; Proud et al., 1998; Tatton-Brown and Rahman, 2013). The ten Weaver syndrome mutations selected for this study include examples of inherited, recurrent *de novo* and unique mutations.

My chosen panel of ten Weaver-syndrome-associated *EZH2* mutations includes two which occurred in children that developed neuroblastoma: A677T and R679C. Two selected mutations have been observed to independently occur in myeloid malignancies: R679C and Y728*. Finally, two selected mutations occurred in individuals who respectively developed acute lymphoblastic leukaemia and non-Hodgkin lymphoma: A677T and E740K.

In sum, the ten *EZH2* mutations selected for study are reflective of the genetic and phenotypic heterogeneity of Weaver syndrome.

4.2.2 No Weaver syndrome mutations are total loss of function

To date, Weaver-syndrome-associated *EZH2* variants have been predicted to be predominantly loss of function. I wished to test this by stably expressing my chosen *EZH2*

mutants in mouse embryonic stem cells that lack both *Ezh1* and *Ezh2* (Fig. 4.2A) – kindly sent to us by Prof Kristian Helin. These so-called *Ezh1/2* double knock-out (dKO) cells are incapable of methylating histone H3 at lysine 27 (H3K27). By re-expressing wild-type or mutant forms of *EZH2* in this isogenic context, I could determine the relative global HMT activity at H3K27 for each variant.

All mutants tested were successfully stably expressed at levels comparable to wild-type EZH2 in *Ezh2^{fl/fl}* cells (Fig. 4.2B). Western blot analyses (Fig. 4.2C-D) confirmed that the *Ezh1/2* dKO cell line alone showed no detectable H3K27 methylation. The control rescue with wild-type (WT) EZH2 behaved as expected with H3K27me1/2/3 restored to levels comparable to the *Ezh2^{fl/fl}* cell line. An additional control rescue with the oncogenic hyperactive Y641F EZH2 mutant also behaved in line with previous reports (McCabe et al., 2012; Sneeringer et al., 2010); Y641F-mutant EZH2 was capable of mono-methylating H3K27, and appeared to convert all subsequently catalysed H3K27me2 to H3K27me3.

These western blot analyses revealed that all Weaver syndrome-associated mutants permit some degree of HMT activity on H3K27. Intriguingly, the D233V missense mutation rescued each H3K27 methylation state to approximately wild-type levels (Fig. 4.2D, dashed lines). By contrast, the Y728* truncating mutation was only capable of modest H3K27 monomethylation, exhibiting H3K27me1 levels between that of the Y641F rescue and the *Ezh1/2* dKO parental cell line (Fig. 4.2D, left panel). All other Weaver mutant rescue cell lines displayed H3K27me1 levels comparable to or greater than the WT control rescue. Eight of ten tested Weaver syndrome mutations were, on average, impaired in their global ability to dimethylate H3K27 compared to the wild-type rescue (Fig. 4.2D, middle panel). Of these eight, two mutants were severely impaired: A677T and Y728*. With the exception of the D233V variant, all Weaver syndrome-associated mutants were globally impaired in their ability to tri-methylate H3K27, showing at least a fifty-percent reduction in global H3K27me3 levels compared to rescue with WT EZH2 (Fig. 4.2D, left panel). The K563E and Y728* mutants displayed the most severe impairment in H3K27 trimethylation.

Collectively, these data support the hypothesis that Weaver-syndrome-associated *EZH2* mutations are loss of function with respect to global PRC2 HMT activity and specify partial rather than total loss of function. The D233V variant was identified as an outlier whose HMT activity was indistinguishable from wild type.

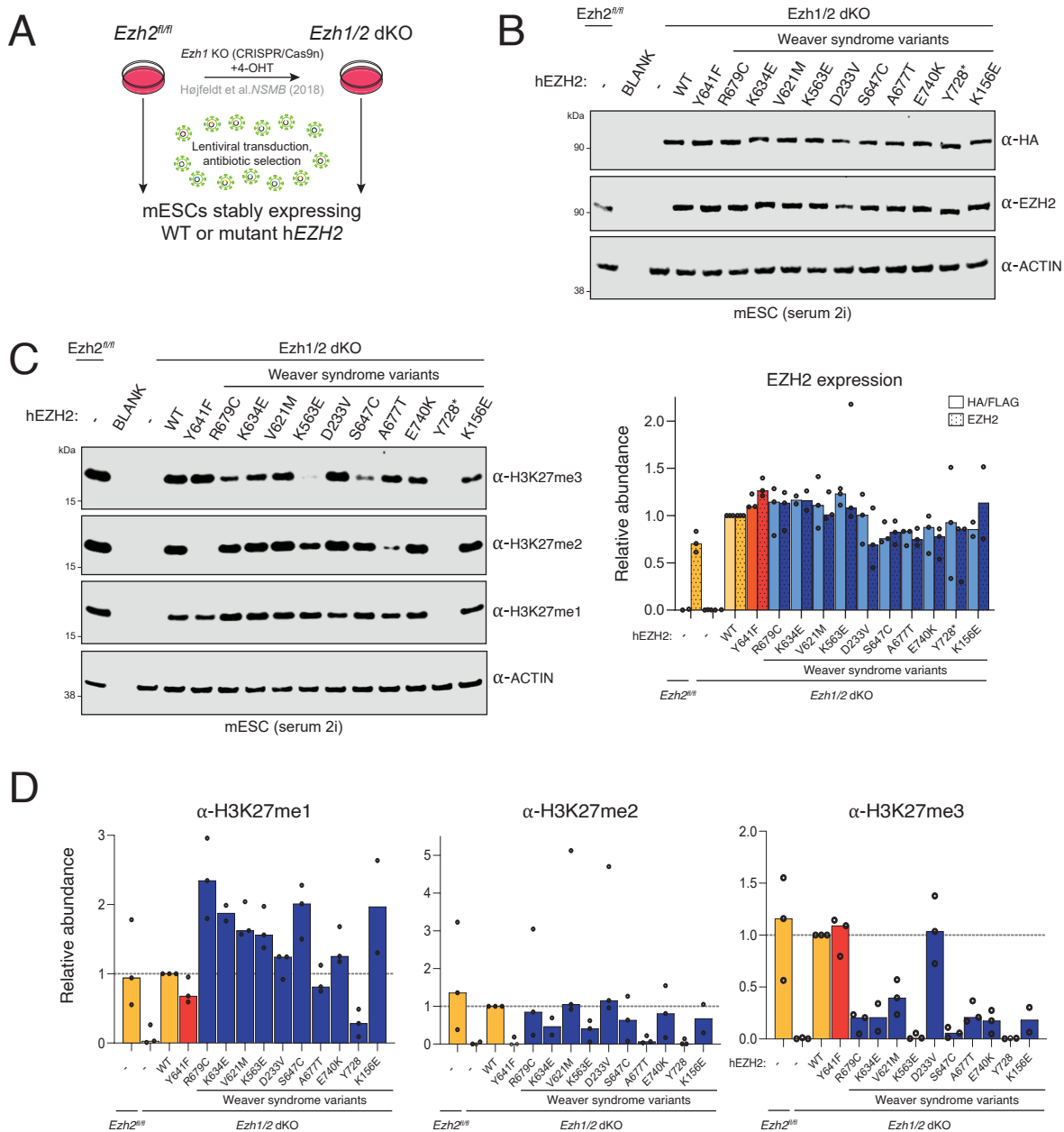


Figure 4.2 Weaver-syndrome-associated EZH2 variants are not total loss of function

(A) Schematic outlining the experimental setup for the data presented in this figure.

(B) Top: Representative western blot for EZH2 and HA, probing whole cell lysates from the indicated cell lines. Bottom: Relative quantification of western blot analyses of EZH2 and the HA or FLAG epitope tag. Signal intensity values were normalized to a loading control and then expressed relative to the *Ezh1/2* dKO + hEZH2 WT cell line.

Data values represent biological replicates and their median values are plotted as bars.

(C) Representative western blots for H3K27 methylation states, probing whole cell lysates from the indicated cell lines.

(D) Relative quantification of western blot analyses of H3K27 methylation abundance. Signal intensity values were normalized to a loading control and then expressed relative to the *Ezh1/2* dKO + hEZH2 WT cell line. The dashed line represents the normalised *Ezh1/2* dKO + hEZH2 WT value. Data values represent biological replicates and their median values are plotted as bars.

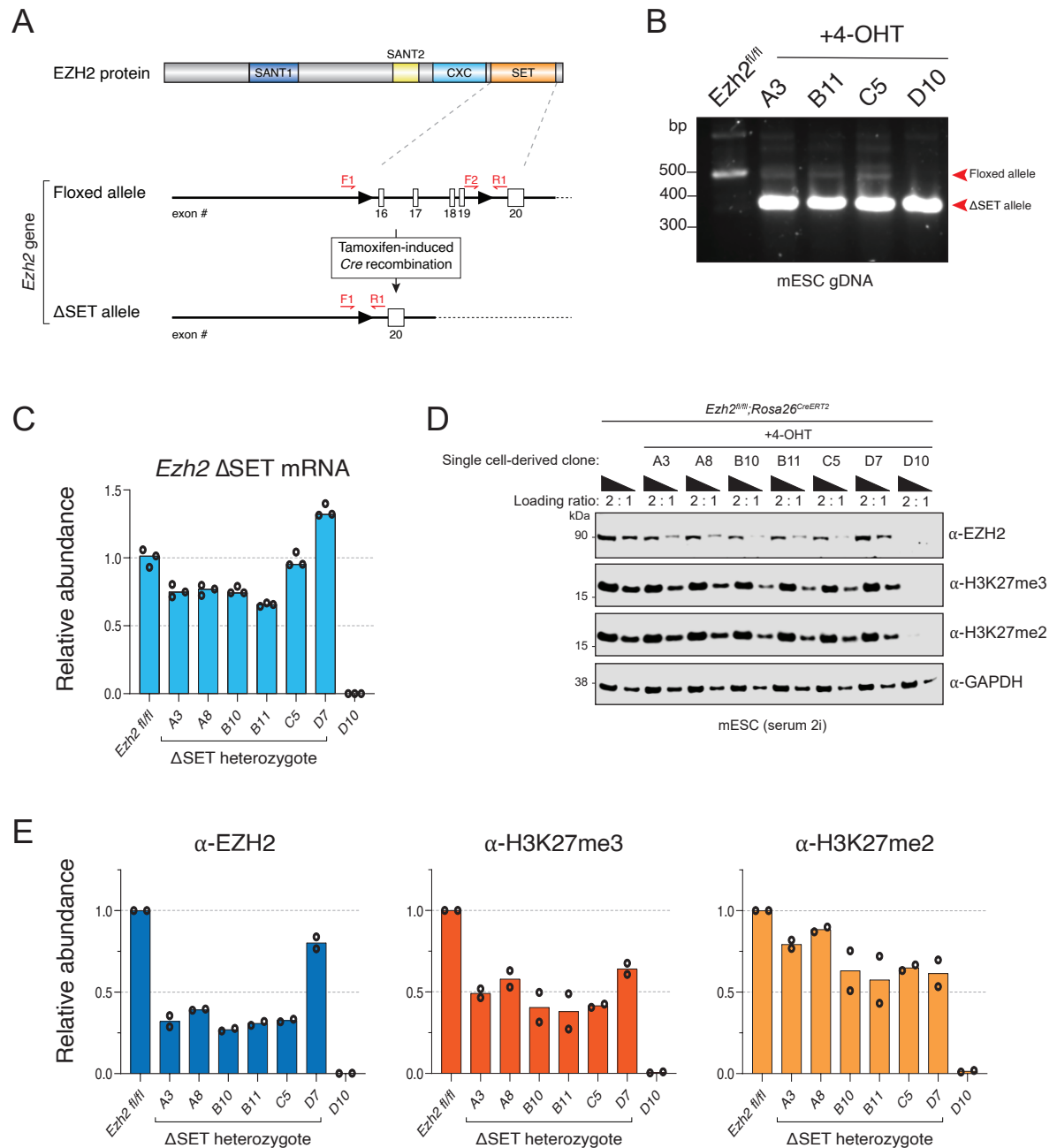


Figure 4.3 Generation and validation of single-cell-derived *Ezh2* heterozygous knockout clones

(A) Top: Schematic of the mouse EZH2 protein showing key domains. Bottom: Schematic representation of the *Ezh2*-targeted locus both before and after Cre-mediated recombination. SET domain-encoding exons are numbered and displayed as open boxes, while *loxP* sequences are depicted by arrows. This gene targeting strategy was originally devised by the Tarakhovsky lab (Su et al. 2003).

(B) A representative genotyping PCR using the primers indicated in panel A, confirming heterozygous (A3, B11, C5) or homozygous (D10) knockout of the *Ezh2* SET domain-encoding exons.

(C) RT-qPCR analysis using primers targeted to the deleted region of *Ezh2* in wild-type (*Ezh2* fl/fl), Δ SET heterozygous or Δ SET homozygous knockout clones. Data values represent technical replicates and their median value is plotted.

(D) Western blot for EZH2, H3K27me3 and H3K27me2, probing whole cell lysates from the indicated cell lines.

(E) Relative quantification of the western blot analyses presented in panel D. Signal intensity values were normalized to a loading control and then expressed relative to the wild-type sample (*Ezh2* fl/fl). Data values represent technical replicates and their median value is plotted.

4.2.3 Generation of an *Ezh2* heterozygous knockout cell line for improved cellular modelling of Weaver syndrome

Weaver syndrome is caused by constitutive heterozygous mutations in *EZH2*, so I wanted to assess the function of my chosen *EZH2* mutations in a more biologically relevant context i.e., in the presence of an endogenous wild-type *EZH2* protein. In lieu of using targeted genome-editing to generate heterozygous “knock-in” cell lines from scratch, I opted to utilize tools already available to me to generate a cell line with a heterozygous knockout of *Ezh2* against which I could re-introduce wild-type or mutant *EZH2* transgenes. The Tarakhovsky lab (Su et al., 2003) generated *Ezh2^{fl/fl}* mice in which exons encoding the catalytic SET domain of *Ezh2* are flanked by *loxP* sites which are recombined in the presence of Cre recombinase, causing a functional knockout of *EZH2* (Fig. 4.3A). While it is possible that a SET-domain-deficient truncated *EZH2* protein (*EZH2*ΔSET) is expressed from the knockout locus, it is likely unstable and/or degraded (Højfeldt et al., 2018; Shen et al., 2008; Su et al., 2003). The Helin lab (Højfeldt et al., 2018) later crossed these *Ezh2^{fl/fl}* mice with *Rosa26^{CreERT2}* mice (B6 strain) to derive tamoxifen-inducible *Ezh2* knockout mouse embryonic stem cells (*Ezh2^{fl/fl};Rosa26^{CreERT2}*). I used these *Ezh2^{fl/fl};Rosa26^{CreERT2}* mESCs – kindly sent to us by Prof Kristian Helin – to generate heterozygous *Ezh2* knockout cell lines.

A three-primer genotyping PCR reaction was used to screen for positive clones; this multiplex approach allowed for the simultaneous detection of both the wild-type (floxed) and deletion (ΔSET) allele from the same template gDNA. I identified several heterozygous *Ezh2* knockout clones (Fig. 4.3B). I took six heterozygous clones forward for further screening, along with one homozygous knockout clone (D10) for comparison. RT-qPCR analysis using primers located within the deleted ΔSET region showed that the heterozygous loss of *Ezh2* at the gDNA level does not necessarily cause a fifty percent reduction in corresponding mRNA abundance relative to the parental *Ezh2^{fl/fl}* cell line (Fig. 4.3C). Nevertheless, western blot analyses confirmed the functional heterozygous loss of *EZH2* across all clones (Fig. 4.3D). At the protein level, *EZH2* abundance was reduced by more than fifty percent compared to the *Ezh2^{fl/fl}* parental cell line – excepting one clone (D7) that showed just a twenty percent reduction (Fig. 4.3E, left panel). Across all six heterozygous knockout clones H3K27me3 levels were reduced by approximately half compared to the *Ezh2^{fl/fl}* parental cell line, while H3K27me2 levels also reduced by between twelve to forty-two percent (Fig. 4.3E, middle and right panels). By comparison, the D10 homozygous knockout clone showed no detectable *EZH2* protein or H3K27me3, and only barely

detectable H3K27me2. On the combined basis of EZH2, H3K27me3 and H3K27me2 levels – I chose *Ezh2*^{fl/Δ} clone C5 as the parental cell line for my experiments going forward. It is worth noting that during this initial clone screening process, I did not detect any truncated EZH2ΔSET protein. However, throughout subsequent analyses I did occasionally detect what appears to be the EZH2ΔSET truncation product (at less than 90 kDa observed molecular weight, data not shown) but consistent with prior reports, it was unstable and present at low levels (Højfeldt et al., 2018).

4.2.4 Isogenic mESC models of Weaver syndrome suggest varying degrees of dominant negative effects on PRC2 catalytic activity

To assess the function of Weaver-syndrome-associated hEZH2 variants in a more biologically relevant context, I repeated the rescue experiments described in section 4.2.2 but this time using the *Ezh2* heterozygous knockout cellular background (Fig. 4.4A) – note that the original parental cell line (*Ezh2*^{fl/fl}) is the same for both experiments. If any WS-associated EZH2 variants operate through a cooperative gain-of-function or dominant negative mechanism, this could be detected by comparing the results obtained in this cellular context versus the last.

Once again, all mutants tested were successfully stably expressed at levels comparable to wild-type EZH2 in *Ezh2*^{fl/fl} cells (Fig. 4.4B). Western blot analyses (Fig. 4.4C-D) confirmed that the *Ezh2*^{fl/Δ} heterozygote cell line exhibits reduced H3K27me3 and H3K27me2 levels compared to both the *Ezh2*^{fl/fl} cell line and the wild-type rescue (+WT) cell line. The +WT control again displayed H3K27me2/3 levels comparable to the *Ezh2*^{fl/fl} cell line, with H3K27me1 levels lower than that of the *Ezh2*^{fl/Δ} heterozygote but higher than the *Ezh2*^{fl/fl} line. The additional control rescue with the oncogenic hyperactive Y641F EZH2 mutant again behaved in line with previous reports (McCabe et al., 2012; Sneeringer et al., 2010); the Y641F H3K27 hyper-trimethylation phenotype was more readily apparent in this cellular context than the last, and very little H3K27me2 was detected.

For each Weaver-syndrome-associated EZH2 mutation tested the relative global patterns of H3K27me1/2/3 were remarkably similar against this *Ezh2*^{fl/Δ} heterozygote cell background compared to the *Ezh1/2* dKO cell background (compare Fig. 4.4D to Fig. 4.2D). This observation argues against any cooperative gain-of-function mutations. Interestingly, the D233V variant still appeared almost indistinguishable from the wild-type EZH2 at the global level. Excluding this variant, average H3K27me3 levels were more

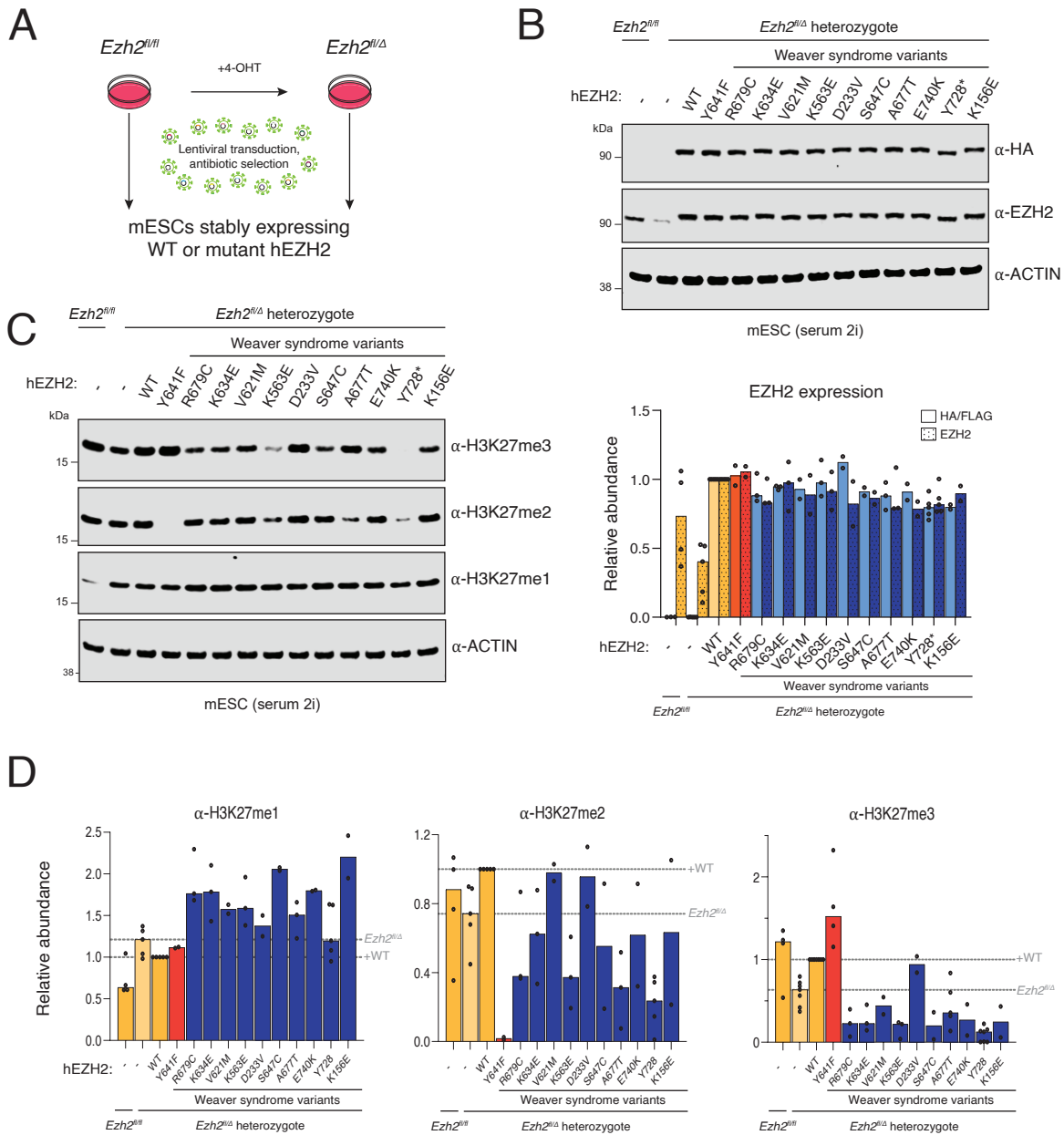


Figure 4.4 Weaver-syndrome-associated EZH2 variants can exert dominant negative effects

(A) Schematic outlining the experimental setup for the data presented in this figure.

(B) Top: Representative western blots for EZH2 and HA, probing whole cell lysates from the indicated cell lines. Bottom: Relative quantification of western blot analyses of EZH2 and the HA or FLAG epitope tag. Signal intensity values were normalized to a loading control and then expressed relative to the *Ezh2*^{fl/Δ} + hEZH2 WT cell line. Data values represent biological replicates and their median values are plotted as bars.

(C) Representative western blots for H3K27 methylation states, probing whole cell lysates from the indicated cell lines.

(D) Relative quantification of western blot analyses of H3K27 methylation abundance. Signal intensity values were normalized to a loading control and then expressed relative to the *Ezh2*^{fl/Δ} + hEZH2 WT cell line. The dashed lines represent the normalised *Ezh2*^{fl/Δ} and *Ezh2*^{fl/Δ} + hEZH2 WT values. Data values represent biological replicates and their median values are plotted as bars.

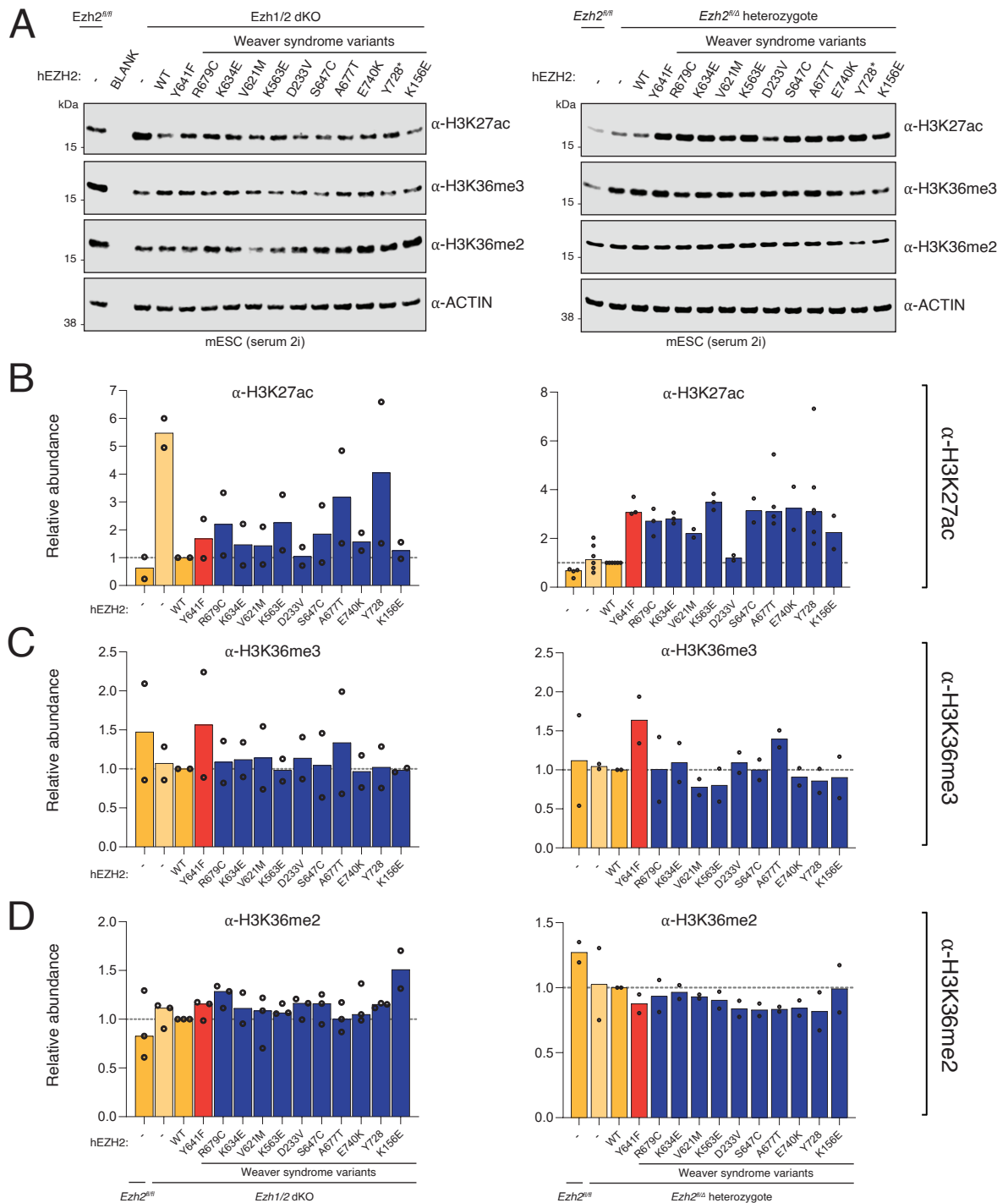


Figure 4.5 Isogenic mESC models expressing Weaver-syndrome-associated EZH2 variants exhibit global H3K27 hyperacetylation

(A) Representative western blots for H3K27ac, H3K36me3 and H3K36me2, probing whole cell lysates from the indicated cell lines.

(B-D) Relative quantification of western blot analyses of (B) H3K27ac, (C) H3K36me3 and (D) H3K36me2 abundance. Signal intensity values were normalized to a loading control and then expressed relative to the +WT rescue cell line. The dashed lines represent the normalised + WT rescue values. Data values represent biological replicates and their median values are plotted as bars.

severely impaired in the Weaver mutant rescue cell lines compared to their parental *Ezh2^{fl/Δ}* heterozygote cell line, (Fig. 4.4D, left panel). Similarly, eight of ten tested Weaver syndrome mutations were, on average, impaired in their global ability to dimethylate H3K27 compared to their *Ezh2^{fl/Δ}* parental cell line (Fig. 4.4D, middle panel). Against this cellular background, all mutants tested showed elevated H3K27me1 levels compared to the +WT rescue.

Taken together with the results of section 4.2.2, these data support to the hypothesis that Weaver-syndrome-associated *EZH2* mutations are partial loss of function through a dominant negative mechanism.

4.2.5 Isogenic mESC models of Weaver syndrome exhibit global H3K27 hyperacetylation

Given the well-established antagonistic relationship between H3K27 methylation and acetylation, I next wanted to assess the effects of my chosen Weaver syndrome mutations on global H3K27ac levels (Lavarone et al., 2019; Pasini et al., 2010; Tie et al., 2009). I was further motivated to do so by a recently published review article which draws attention to the reciprocal skeletal phenotypes of Weaver versus Rubinstein-Taybi syndrome, two congenital growth conditions respectively associated with impaired H3K27 methylation and acetylation – as will be discussed in detail in section 6.1.3 (Gamou and Gibson, 2020). I also sought to explore any potential changes to H3K36 methylation in the presence of Weaver-syndrome-associated mutations, as I had previously hypothesised that functional interplay between H3K27 and H3K36 methylation would be central to the molecular aetiology of Weaver syndrome (section 1.10; Deevy and Bracken, 2019). To address these questions, I re-analysed protein lysates from the experiments described in sections 4.2.2 and 4.2.4 but this time immunoblotting for H3K27ac and H3K36me2/3 (Fig. 4.5).

Looking first at the experiment performed in the *Ezh1/2* dKO context, western blot analyses reveal that H3K27ac levels increased more than five-fold in the *Ezh1/2* dKO cell line compared to the *Ezh2^{fl/Δ}* parental line but returned to normal with the re-introduction of WT EZH2 (Fig. 4.5A-B, left panels). Contrastingly, average H3K27ac levels remained elevated in the Y641F and Weaver mutant rescue cell lines – except for the D233V mutant which closely phenocopied the wild-type rescue. Analysis of H3K36me3 levels did not yield a clear result as duplicate values were inconsistent (Fig. 4.5C, left panel). Similarly, H3K36me2 levels did not appear significantly changed across any condition but trended towards an increase in the Weaver syndrome mutant rescues (Fig. 4.5D, left panel).

Similar results were obtained from the experiment performed in the *Ezh2*^{fl/Δ} heterozygous knockout context. Western blot analyses show that global H3K27ac levels can stay level or increase up to two-fold in the *Ezh2*^{fl/Δ} cell line compared to the *Ezh2*^{fl/fl} parental line (Fig. 4.5A-B, right panels). H3K27ac levels remain comparable between the *Ezh2*^{fl/fl} cell line, the +WT rescue and the +D233V rescue. Notably, H3K27ac significantly increases with all other Weaver-syndrome-associated mutations and, on average, more so than in the *Ezh2*^{fl/Δ} heterozygous knockout alone. This may be a direct consequence of the decreased H3K27me3 in these cell lines though, intriguingly, the Y641F mutant also exhibits elevated H3K27ac despite H3K27 hyper-trimethylation. Turning to H3K36me3, average global levels appear consistent across the *Ezh2*^{fl/fl}, *Ezh2*^{fl/Δ} and +WT rescue cell lines (Fig. 4.5C, right panel). H3K36me3 appears slightly elevated in the +Y641F and +A677T rescue cell lines, but unchanged or marginally decreased in all other mutant rescues. Average global H3K36me2 levels appear to decrease with heterozygous inactivation of *Ezh2* and do not return to normal with the re-introduction of wild-type EZH2 (Fig. 4.5D, right panel). H3K36me2 levels may decrease slightly further in the presence of a Weaver-syndrome-associated mutation.

Taken together, these data support H3K27 hyper-acetylation as a molecular response to Weaver-syndrome-associated EZH2 variants. The effect of these same variants on H3K36 methylation remains unclear.

4.2.6 Weaver syndrome-associated mutant hEZH2 assembles into PRC2 and alters its genome-wide binding profile

I wished to further characterise the molecular nature of Weaver-syndrome-associated *EZH2* mutations, but it became prohibitively labour intensive to continue with the full panel of rescues. For this reason, from this point forward I focussed my attention on just five Weaver-syndrome-associated variants expressed against the *Ezh2*^{fl/Δ} genetic background: R679C, K634E, K563E, A677T and Y728*. These mutants were selected based on their frequency and/or striking phenotype.

The western blot data had suggested that most of my chosen Weaver-syndrome-associated EZH2 variants were negatively interfering with the function of the endogenous wild-type EZH2 protein in the *Ezh2*^{fl/Δ} cellular context (see section 4.2.4). To better understand this apparent dominant negative inhibition, I wanted to confirm that these variant EZH2 proteins were in fact assembling into PRC2. To this end, I performed FLAG-immunoprecipitations and showed that the endogenous EED and SUZ12 proteins

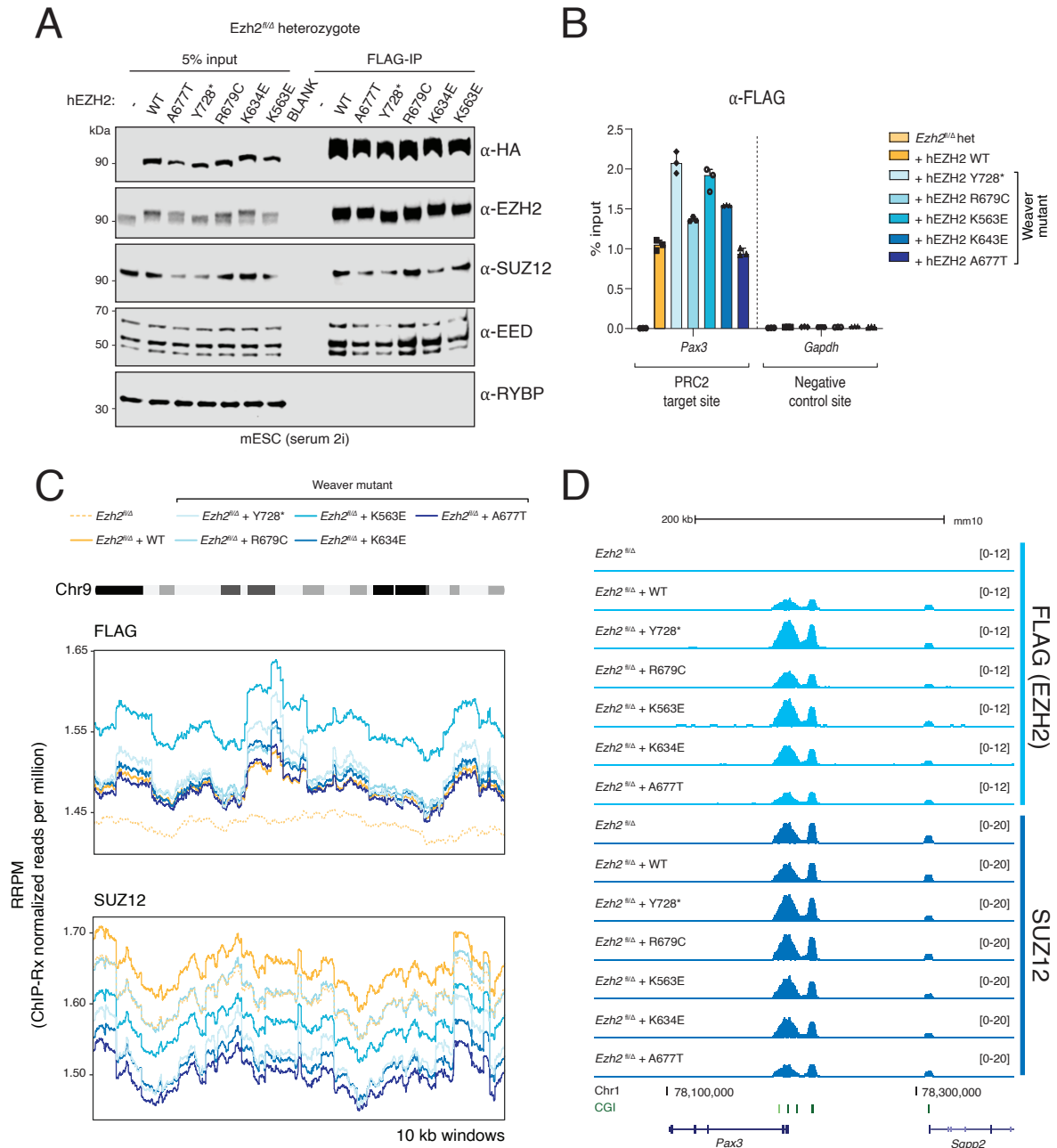


Figure 4.6 Weaver-syndrome-associated EZH2 variants assemble into PRC2 and alter its genome-wide binding profile

(A) Representative western blot for PRC2 subunits of cell extracts (input) and anti-FLAG immunoprecipitated material (FLAG-IP) from the indicated cell lines (*Ezh2*^{fl/fl} mESCs with or without stable transgenic expression of indicated FLAG-tagged hEZH2 constructs). Western blotting for the vPRC1 subunit RYBP is included as a negative control.

(B) ChIP-qPCR signals for α-FLAG in the indicated cell lines (as above) at a PRC2 target site and non-target negative control region. Data values represent technical replicates and their median value is plotted.

(C) Rolling average plots of the normalized reads per million of FLAG and SUZ12 ChIP-Rx enrichments across 12451 10kb windows of chromosome 9 (smoothened over 1000 bins) in the indicated cell lines (as above).

(D) Genome browser representations of ChIP-Rx normalised reads of FLAG and SUZ12 at a representative PRC2 target site (*Pax3*) in the indicated cell lines (as above). CGI = CpG island.

specifically co-immunoprecipitate with the epitope-tagged EZH2 variants (Fig. 4.6A). Fluorescent immunocytochemistry determined that the epitope tagged EZH2 constructs were exclusively localised to the nucleus – as would be expected for a PRC2 component (data not shown). Next, ChIP-qPCR analysis indicated that the variant EZH2 proteins were specifically binding at PRC2 target genes, which again supports the notion that they are assembling into PRC2 (Fig. 4.6B). Interestingly, four out of five mutant constructs demonstrated enriched binding to PRC2 target sites compared to the wild type whilst the relative A677T mutant enrichment was slightly less. Quantitative ChIP-sequencing (ChIP-Rx) for the epitope-tagged constructs confirmed and extended this finding genome-wide (Fig. 4.6C-D). Conversely, ChIP-Rx for endogenous PRC2 (SUZ12) indicated that its enrichment was slightly decreased genome-wide across all cell lines rescued with a Weaver mutant form of EZH2 compared to the cell line rescued with wild-type EZH2 (Fig. 4.6C-D).

These combined findings support the notion that Weaver-syndrome-associated mutant EZH2 can assemble into PRC2 but may alter its genome-wide binding profile.

4.2.7 Weaver syndrome-associated hEZH2 mutations can cause divergent aberrations to H3K27me2/3 deposition at Polycomb target promoters

Once I had established that Weaver-syndrome-associated EZH2 variants could indeed incorporate into PRC2 and localise at its target sites, the next question was whether H3K27 methylation was altered at these target sites. To address this question, I performed genome-wide ChIP-Rx of H3K27me2/3 in *Ezh2^{fl/4}* mESCs rescued with either wild-type or Weaver-mutant forms of EZH2. An average plot profiling H3K27me3 at all PRC2 target sites lead to a surprising discovery: compared to the wild-type rescue, four out of five Weaver mutant rescues display H3K27 hypo-methylation at these sites but one (A677T) shows an opposing H3K27 hyper-trimethylation phenotype (Fig. 4.7 A-B). Although inconsistent with the western blot results, this A677T-associated hyper-trimethylation phenotype replicated across two independent ChIP-Rx experiments. An average plot profiling H3K27me2 at these same PRC2 target sites shows the opposite trend: compared to the wild-type rescue, the four H3K27 hypo-trimethylating EZH2 variants show an accumulation of H3K27me2 at the peak centre whereas the A677T variant exhibits lower H3K27me2 levels (Fig. 4.7 A-B).

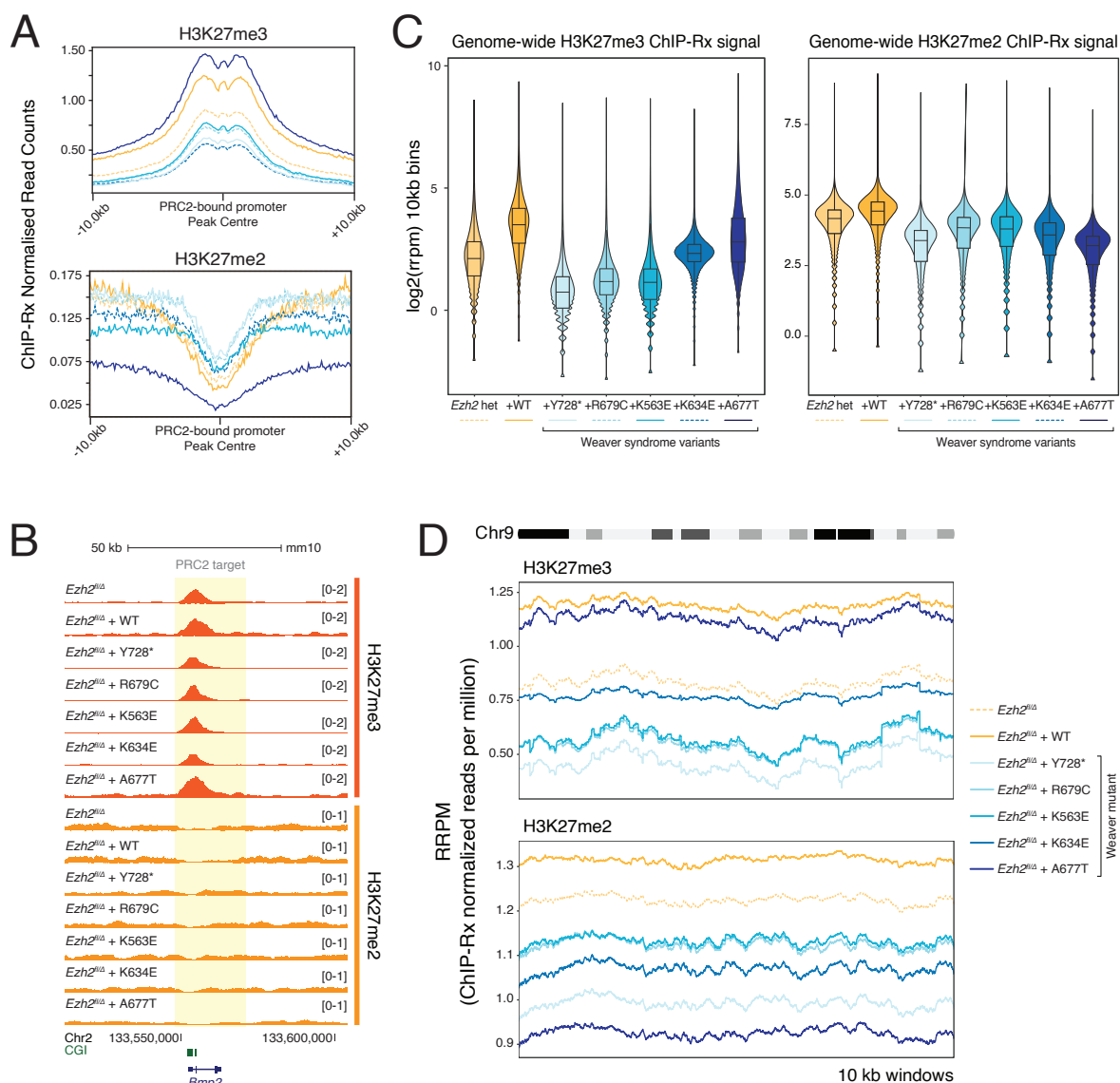


Figure 4.7 Weaver-syndrome-associated EZH2 variants can cause divergent aberrations to H3K27me3 and H3K27me2 deposition at Polycomb target promoters

(A) Average ChIP-Rx signal profiles for H3K27me3 and H3K27me2 at PRC2-bound promoters in *Ezh2*^{flΔ} mESCs with or without stable transgenic expression of indicated hEZH2 constructs. Colour-coded legend as in panel B.

(B) Genome browser representations of ChIP-Rx normalised reads of H3K27me3 and H3K27me2 at a representative PRC2 target site (*Bmp2*) in the indicated cell lines. CGI = CpG island.

(C) Violin plot representations of the genome-wide log2 ChIP-Rx normalized reads per million of H3K27me3 and H3K27me2 enrichments across 272477 10kb bins.

(D) Rolling average plots of the normalized reads per million of H3K27me3 and H3K27me2 ChIP-Rx enrichments across 12451 10kb windows of chromosome 9 (smoothened over 1000 bins) in the indicated cell lines.

Taken together, these data are consistent with a model in which most Weaver-syndrome-associated EZH2 variants are partial loss-of-function but indicate that at least one Weaver-syndrome-associated EZH2 variant exists with a more complex molecular phenotype.

4.2.8 Cellular models of Weaver syndrome share broad reductions in H3K27me2 and H3K27me3 with concomitant global increases in H3K27ac

Having shown that Weaver-syndrome-associated hEZH2 mutations can cause divergent aberrations to H3K27me2/3 deposition at Polycomb target promoters, I next sought to identify any changes common across the five cellular models of Weaver syndrome chosen for study. I also wanted to understand the discrepancy between the observations that the A677T variant exhibits global H3K27 hypo-trimethylation but hyper-trimethylation at PRC2 target sites. To this end, we quantified the global H3K27me2/3 ChIP-Rx signal and found that – consistent with the western blot data – all cellular models of Weaver syndrome exhibit globally reduced levels of H3K27me2/3 compared to the matched wild-type rescue cell line (Fig. 4.7 C-D). Next, we parsed this data out to individually examine gene body and intergenic regions and found that, compared to the wild-type rescue control, all five cellular models of Weaver syndrome showed reduced H3K27me2/3 at both gene body and intergenic regions (Fig. 4.8). All mutants exhibited lower gene body and intergenic H3K27me2 levels compared to the *Ezh2^{f/f}* heterozygote control – highlighting dominant negative mechanisms in effect. Interestingly, the A677T variant was the only rescue that did not also display lower gene body and intergenic H3K27me3 levels compared to the *Ezh2^{f/f}* heterozygote control.

Next, I turned my attention to H3K27ac which was observed by western blot to be globally elevated in cellular models of Weaver syndrome. I performed genome-wide ChIP-Rx of H3K27ac in *Ezh2^{f/f}* mESCs rescued with either wild-type or WS-associated forms of EZH2. Compared to the wild-type rescue control, each cellular model of Weaver syndrome displayed considerable H3K27 hyperacetylation at the global level and at both intergenic and gene body regions. (Fig. 4.9). Four out of five Weaver mutant rescues displayed H3K27 hyperacetylation at promoters and enhancers; the K634E mutant rescue shows reduced H3K27ac compared to wild type at promoter and enhancer peak centres, but increased H3K27ac up- and downstream of this (Fig. 4.9C, top panels). The K563E and Y728* rescue cell lines exhibited the lowest global levels of H3K27me3 and the highest global levels of H3K27ac; genome wide, H3K27ac tended to increase wherever H3K27me3 or H3K27me2 decreased (Fig. 10).

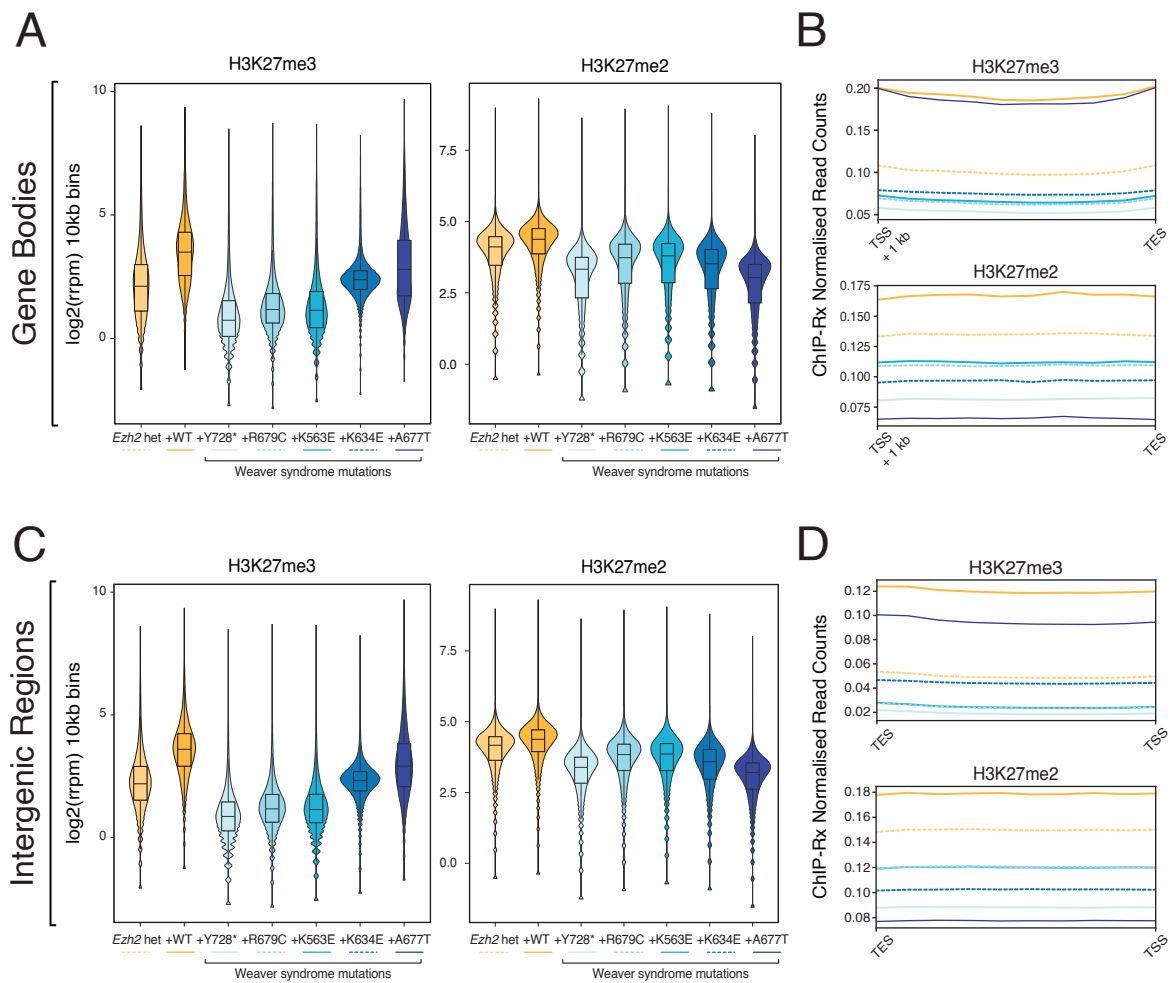


Figure 4.8 Cellular models of Weaver syndrome share broad reductions in H3K27me3 and H3K27me2 outside of Polycomb target promoters

(A) Violin plot representations of the log2 ChIP-Rx normalized reads per million of H3K27me3 and H3K27me2 enrichments in the indicated cell lines across 196505 10kb bins representing all gene bodies.

(B) Average ChIP-Rx signal profiles for H3K27me3 and H3K27me2 at gene bodies in *Ezh2*^{fl/Δ} mESCs with or without stable transgenic expression of indicated hEZH2 constructs. Colour-coded legend as in panel A.

(C) Violin plot representations of the log2 ChIP-Rx normalized reads per million of H3K27me3 and H3K27me2 enrichments in the indicated cell lines across 188538 10kb bins representing all intergenic regions.

(D) Average ChIP-Rx signal profiles for H3K27me3 and H3K27me2 at intergenic regions in *Ezh2*^{fl/Δ} mESCs with or without stable transgenic expression of indicated hEZH2 constructs. Colour-coded legend as in panel C.

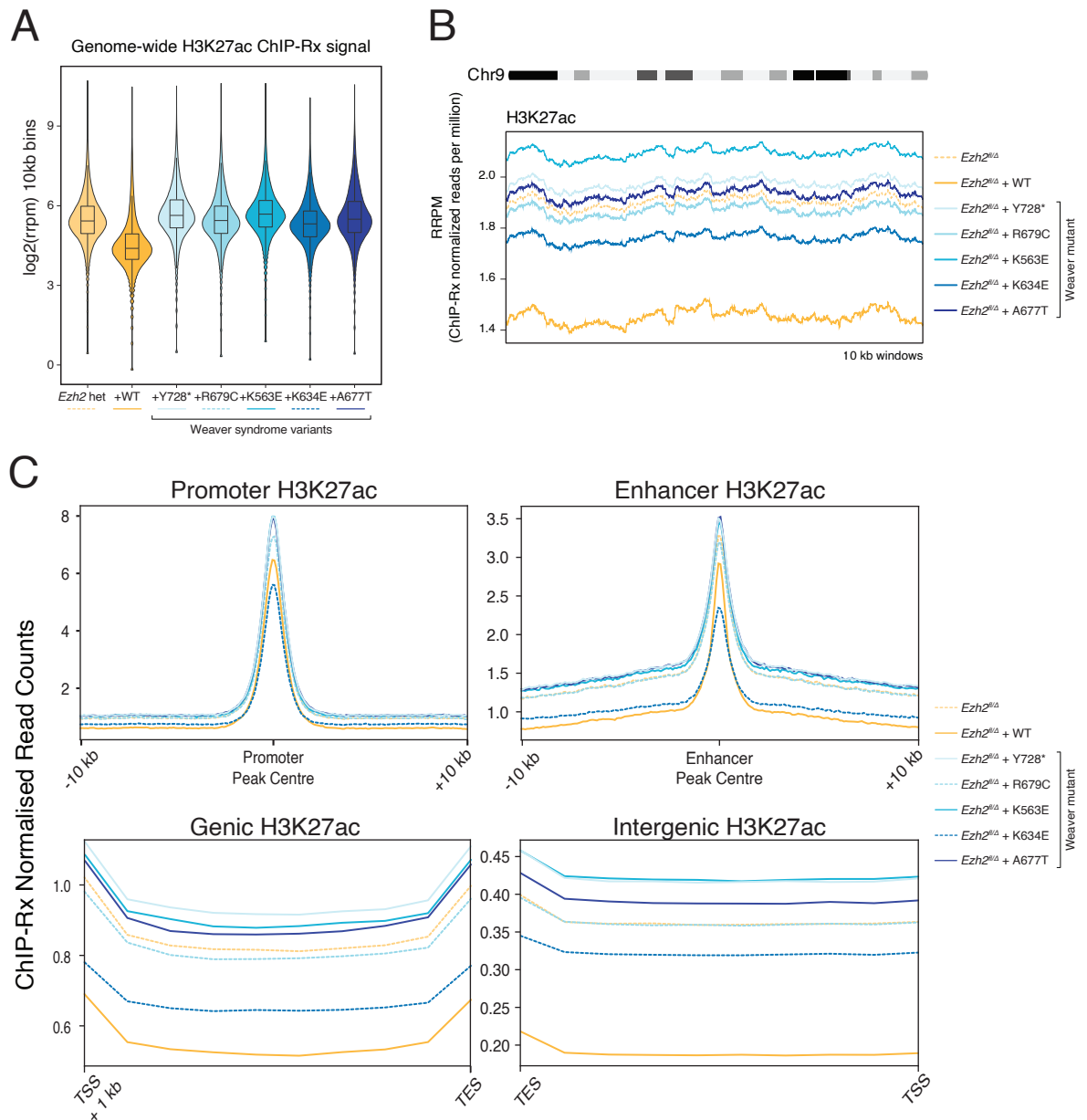
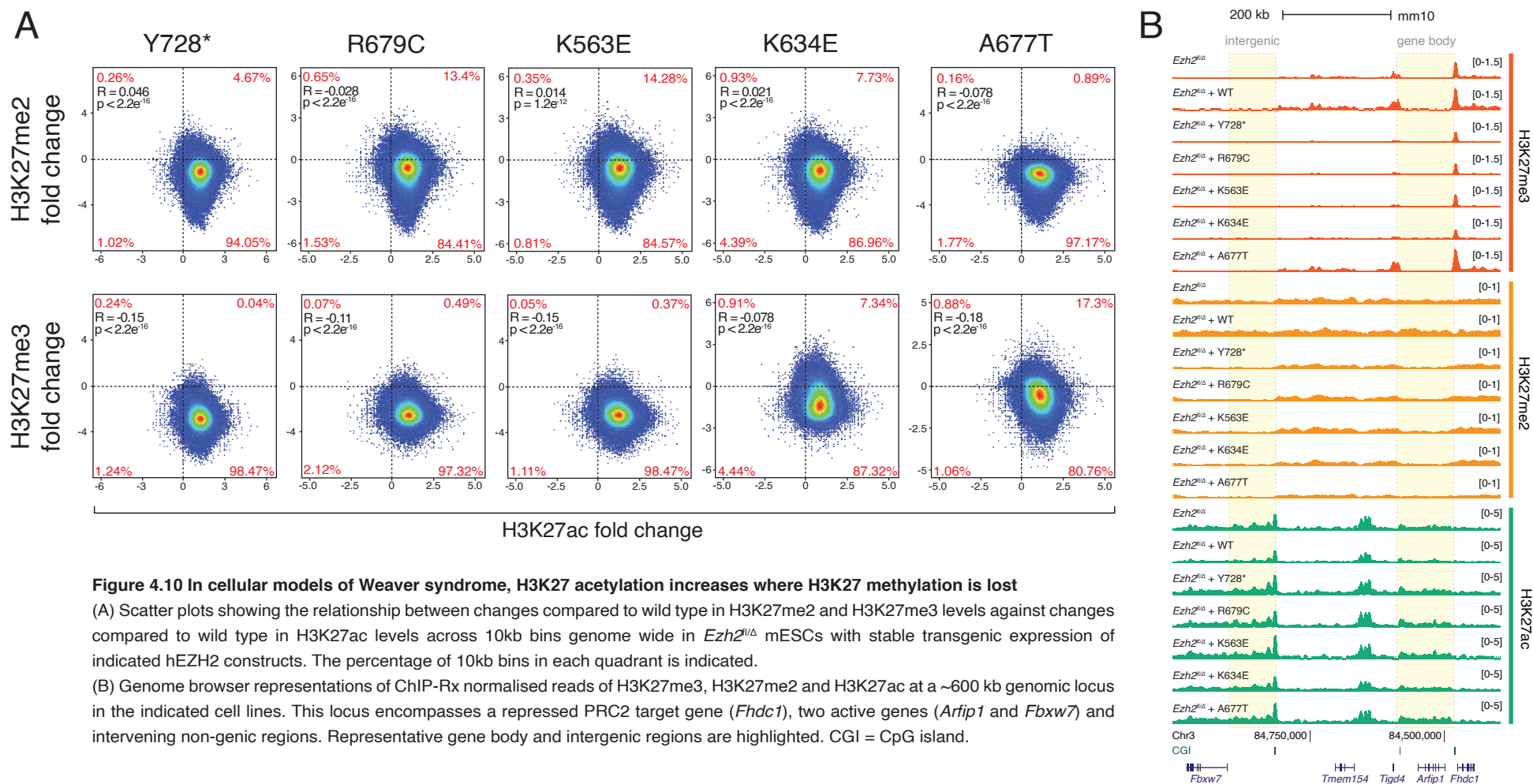


Figure 4.9 Cellular models of Weaver syndrome share broad increases in H3K27ac

(A) Violin plot representations of the log2 ChIP-Rx normalized reads per million of H3K27ac in the indicated cell lines (*Ezh2*^{fl/Δ} mESCs with or without stable transgenic expression of indicated hEZH2 constructs) across 272477 10kb bins genome wide.

(B) Rolling average plots of the normalized reads per million of H3K27ac ChIP-Rx enrichments across 12451 10kb windows of chromosome 9 (smoothened over 1000 bins) in the indicated cell lines (as above).

(C) Average ChIP-Rx signal profiles for H3K27ac at promoter, enhancer, gene body and intergenic regions in *Ezh2*^{fl/Δ} mESCs with or without stable transgenic expression of indicated hEZH2 constructs.



Taken together, the results of these ChIP-Rx analyses support a model in which H3K27ac is globally increased as a response to decreased HMT activity of PRC2 in the presence of a Weaver-syndrome-associated EZH2 variant.

4.2.9 Global H3K27 hyperacetylation does not correlate with an increased abundance of non-genic RNA transcripts in cellular models of Weaver syndrome

I next sought to identify some functional consequences of the observed shift in the balance from H3K27 methylation to acetylation. I hypothesised that the global loss of repressive H3K27me2/3, combined with the accumulation of activating H3K27ac, would affect both chromatin compaction and transcription.

As introduced in chapter one, histone acetylation promotes chromatin accessibility and decompaction. The recruitment of canonical PRC1, which is responsible for PRC1-mediated chromatin compaction, is impaired by the loss of *Ezh2* and H3K27me3 in mouse embryonic stem cells (Kundu et al., 2017). Taken together, I hypothesise that cellular models of Weaver syndrome will display more open chromatin conformations compared to their matched wild-type control. To test this, we have entered collaboration with the lab of Dr Diego Pasini (IEO Milan, Italy) to assess chromatin condensation by stochastic optical reconstruction microscopy (STORM) as described in Conway et al. (2021). The acquisition and analysis of this data is ongoing.

Less condensed chromatin is more accessible to transcriptional machinery and positively correlated with active transcription. As such, I predicted my cellular models of Weaver syndrome would exhibit a global increase in RNA transcript abundance. Specifically, I hypothesised that the broad loss of repressive H3K27 methylation would lead to increased spurious transcription from intergenic sites or repetitive DNA elements. To test these hypotheses, I performed high-depth (>40 million paired reads) total RNA sequencing on these cell lines in biological triplicate (Fig. 11A). Unfortunately, this experiment was not spike-in normalized so did not have the power to detect total signal changes due to global changes (Chen et al., 2016). However, crude analysis of the relative quantification of total RNA extracted from equal cell numbers shows that, on average, a greater yield of total RNA is extracted from cell lines rescued with a Weaver mutant form of EZH2 (except for A677T) compared to those rescued with wild type EZH2 (Fig. 11B). More replicates are needed to determine whether these differences can be considered significant. Furthermore, we found no evidence of increased transcript abundance from intergenic regions (Fig. 11C) or repetitive DNA elements (data not shown).

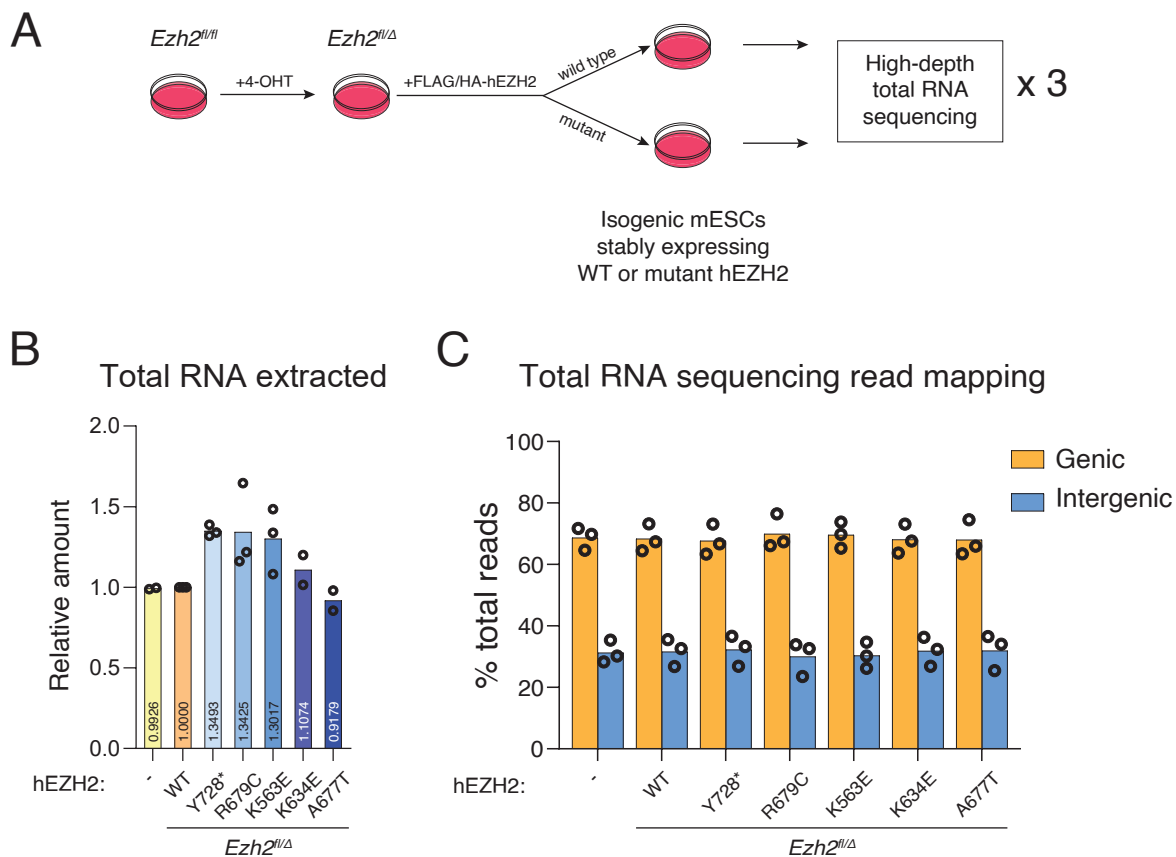


Figure 4.11 Global H3K27 hyperacetylation does not correlate with an increased abundance of non-genic RNA transcripts in cellular models of Weaver syndrome

(A) Schematic outlining the experimental setup for the data presented in panel C of this figure. Greater than 40 million paired reads were sequenced per sample.

(B) A graphical representation of the relative quantification of total RNA extracted from equal cell numbers of the indicated cell lines (*Ezh2^{fl/Δ}* mESCs with or without stable transgenic expression of indicated hEZH2 constructs). RNA concentration was quantified by fluorimetry and the amount of total RNA extracted was expressed relative to the *Ezh2^{fl/Δ}* + hEZH2 WT cell line. Data values represent biological replicates and their mean values are plotted as bars.

(C) A graphical representation of the percentage of total RNA sequencing reads mapping to genic or intergenic regions for the indicated cell lines (*Ezh2^{fl/Δ}* mESCs with or without stable transgenic expression of indicated hEZH2 constructs). Data values represent biological replicates and their mean values are plotted as bars.

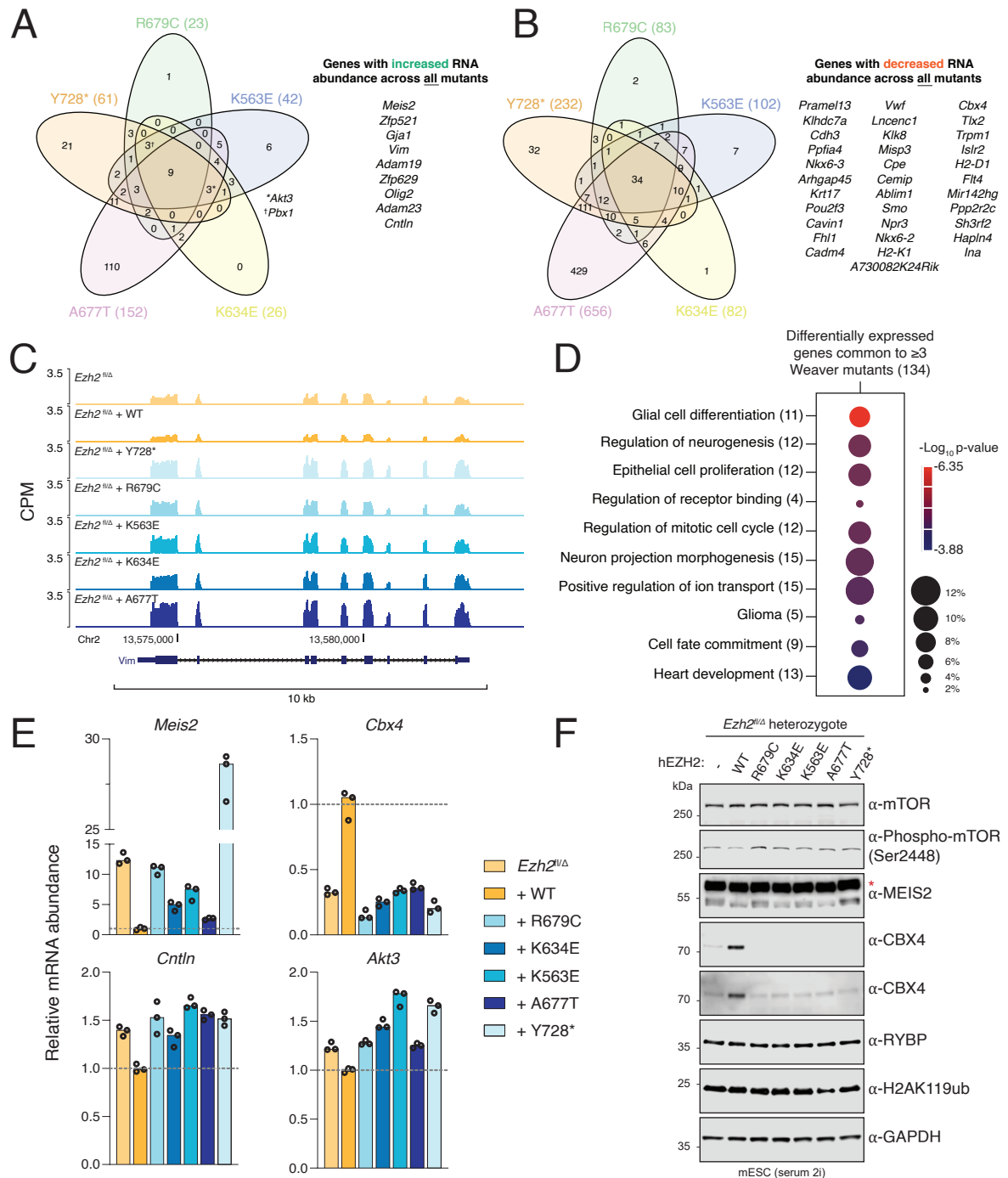


Figure 4.12 Cellular models of Weaver syndrome share a small cohort of differentially expressed genes

(A) Venn diagram presenting the numbers and overlap of genes increased in abundance in the indicated Weaver syndrome models (*Ezh2^{flΔ}* mESCs with stable transgenic expression of indicated hEZH2 constructs) compared to the *Ezh2^{flΔ}* + hEZH2 WT control. Genes specifically increased in all 5 cellular models of Weaver syndrome are listed. The *Akt3* and *Pbx1* genes (increased in abundance in 4 out of 5 mutant cell lines) are also indicated.

(B) Venn diagram presenting the numbers and overlap of genes decreased in abundance in the indicated Weaver syndrome models (*Ezh2^{flΔ}* mESCs with stable transgenic expression of indicated hEZH2 constructs) compared to the *Ezh2^{flΔ}* + hEZH2 WT control. Genes specifically decreased in all 5 cellular models of Weaver syndrome are listed.

(C) Genome browser representation of total RNA sequencing reads at a representative differentially expressed gene (*Vim*) in the indicated cell lines. CPM = counts per million.

(D) Pathway and process enrichment analysis of genes identified as differentially expressed and shared by three or more cellular models of Weaver syndrome.

(E) RT-qPCR validation of the indicated differentially expressed genes in cellular models of Weaver syndrome. Data values represent technical replicates and their median is plotted. The dashed line represents the *Ezh2^{flΔ}* + hEZH2 WT value.

(F) Western blot analyses of the indicated proteins, probing whole cell lysates from *Ezh2^{flΔ}* mESCs with or without stable transgenic expression of indicated hEZH2 constructs. The asterisks denotes a nonspecific band.

4.2.10 Isogenic mESC models of Weaver syndrome share a small cohort of differentially expressed genes

Using the abovementioned RNA-sequencing data, I next wished to identify a transcriptional signature specific to my cellular models of Weaver syndrome. To do this, I examined the transcript abundance of known genes across my different cellular conditions. Differential gene expression analysis determined that there were very few significant changes between the *Ezh2*^{f/Δ} heterozygote knockout compared with the +WT rescue cell line (Table 4.2). On the other hand, each cell line rescued with a Weaver-syndrome-associated mutant EZH2 showed ≥ 106 significant gene expression changes compared to the +WT rescue. Although the transcriptional changes were variable across the different Weaver mutant rescues, a core set of 43 differentially expressed genes was shared across all 5 mutant rescues – 9 consistently upregulated (of which 3 were also upregulated in the *Ezh2*^{f/Δ} condition: *Meis2*, *Adam19* and *Adam23*) and 34 consistently downregulated (of which 1 was also downregulated in the *Ezh2*^{f/Δ} condition: *Ina*) (Fig. 4.12 A-B). I validated the differential transcript abundance of several of these core genes by RT-qPCR analysis, including but not limited to those presented in Figure 4.12E. Pathway and process enrichment analysis revealed that among the differentially expressed genes shared by at least three models of Weaver syndrome were those involved in phenotypically relevant pathways such as regulation of cell differentiation and proliferation (Fig. 4.12D). Attempts to integrate the ChIP-Rx and RNA-seq data failed to yield any clear correlations between a directional change in any histone modification and the expression of an associated gene (data not shown).

The *Meis2* transcription factor was the most highly differentially upregulated gene across each WS cellular condition, exhibiting between 4.3- and 17.3-fold increased abundance compared to WT. *Meis2* is a PcG target gene that is essential for mammalian development (Kondo et al., 2014). *Meis2* upregulation was confirmed at mRNA and protein level (Fig. 4.12 E-F). Conversely, the canonical PRC1 component *Cbx4* was one of the mostly highly consistently downregulated genes across all five cellular models of Weaver syndrome. I confirmed *Cbx4* downregulation at the mRNA and protein level (Fig. 4.12 E-F), while the variant PRC1-mediated H2AK119ub modification and RYBP component appeared comparatively unchanged (Fig. 4.12F). Finally, the *Akt3* gene was called as significantly upregulated (≥1.5-fold increase) in just four out of the five cellular models of Weaver syndrome but was increased 1.4-fold in the remaining (R679C variant) model. *Akt3* encodes a protein kinase which participates in the PI3K/AKT/mTOR signalling pathway which, when activated, promotes cell growth through increased cell metabolism, survival,

turn over and protein synthesis (Engelman et al., 2006; Ersahin et al., 2015). More importantly, *AKT3* was reported to be differentially methylated in Weaver syndrome patients (Choufani et al., 2020). For this reason, I included *Akt3* in the panel of genes for further validation. The slight but consistent upregulation of *Akt3* was confirmed by RT-qPCR (Fig. 4.12E). It has been proposed that AKT phosphorylates mTOR at serine residue 2448 (Edinger and Thompson, 2004; Sekulić et al., 2000). For this reason, I searched for evidence of increased active phospho-mTOR(Ser2448) in my models of Weaver syndrome but the results were inconclusive (Fig. 4.12F).

Table 4.2 Differential transcript abundance analysis

The numbers of genes whose transcripts are significantly (≥ 1.5 -fold) increased (up) or decreased (down) are given for each cell line compared to the *Ezh2^{fl/Δ}* + wild type (WT) EZH2 rescue.

Experiment	Up (# genes)	Down (# genes)	Total (# genes)
WT vs <i>Ezh2^{fl/Δ}</i> heterozygote	9	7	16
WT vs Y728*	61	233	294
WT vs R679C	23	83	106
WT vs K563E	42	102	144
WT vs K634E	26	82	108
WT vs A677T	152	656	808

4.3 Discussion

In this chapter, I confirm that Weaver-syndrome-associated mutations in *EZH2* are predominantly loss-of-function but clarify that this loss-of-function is partial and not complete. Moreover, I demonstrate that both H3K27me3 and H3K27me2 are depleted in cellular models of Weaver syndrome, with H3K27me1 levels concomitantly increased. Finally, I show that the degree of functional impairment on PRC2 HMT activity can vary from mild to severe. These observations are in agreement with the existing literature, and reflective of the variable clinical presentation of Weaver syndrome (Cohen et al., 2016; Imagawa et al., 2017; Lui et al., 2018). However, I additionally identify a single variant (*EZH2* p.D233V) that did not cause any apparent reduction in PRC2 HMT activity compared to wild type, and whose pathogenicity therefore cannot be explained by loss of PRC2 enzymatic activity. It is pertinent to note that the two related individuals with this mutation who were diagnosed with Weaver syndrome present with relatively mild phenotypes (Tatton-Brown et al., 2013). This mutation affects the regulatory SANT1 domain of *EZH2* rather than its catalytic SET domain, so it is possible that it alters rather than impairs the activity of PRC2. Regrettably, I did not include this variant in my FLAG-IP, ChIP-Rx or RNA-seq experiments so I cannot determine its mutational mechanism. Future efforts to understand this variant should focus on assessing whether PRC2-mediated H3K27 methylations are aberrantly redistributed in its presence.

The main goal of this research was to define the mutational mechanism(s) of Weaver syndrome. When inherited, Weaver syndrome exhibits an autosomal dominant mode of inheritance; autosomal dominant loss-of-function mutations can be the result of haploinsufficiency or a dominant negative mutant gene product. I show that most WS-mutant *EZH2* variants impair wild-type PRC2 function, and thus propose dominant negative effects to be a major mutational mechanism for Weaver syndrome. I hypothesise that this dominant negative inhibition may be mediated through several mechanisms. Firstly, I speculate that access to chromatin by wild-type PRC2 is attenuated by prolonged binding of PRC2 containing a WS-associated *EZH2* variant (WS-PRC2). Cohen et al. (2016) showed that Weaver syndrome variants are particularly impaired at H3K27 monomethylation – a prerequisite for H3K27me2/3 catalysis – and thus it is conceivable that WS-PRC2 remains engaged with chromatin for longer while it less efficiently catalyses H3K27 methylation. Supporting this, my data show genome wide that WS-mutant *EZH2* proteins are slightly enriched on chromatin compared to wild type, and vice versa for endogenous *SUZ12*. Secondly, since PRC2 binds to and is allosterically activated by H3K27me3, the depletion of this modification in Weaver syndrome is expected to cause

decreased target engagement and activation of PRC2 – irrespective of whether it contains wild-type or WS-mutant EZH2 (Margueron et al., 2009; Oksuz et al., 2018). This phenomenon could also contribute to the observed decrease in SUZ12 enrichment on chromatin in cellular models of Weaver syndrome. Future detailed biochemical and structural studies would aid in determining the precise mechanisms of dominant negative inhibition of PRC2 by WS-EZH2.

The ChIP-Rx analyses performed in this chapter suggest that most WS-EZH2 variants fail to effectively convert H3K27me2 to H3K27me3 at PRC2-bound sites, and similarly fail to effectively convert H3K27me1 to H3K27me2 at gene bodies and intergenic regions. However, importantly, these analyses also identified an outlier (*EZH2* p.A677T) that appears to operate under different molecular mechanisms. The A677T-EZH2 mutant caused H3K27 hyper-trimethylation at PRC2-bound sites and retained the highest levels of H3K27me3 of all mutants included in these analyses, despite exhibiting the lowest levels of H3K27me2. Collectively, these data suggest that the *EZH2* (p.A677T) mutation confers an enhanced ability to convert H3K27me2 to H3K27me3, similar to the *EZH2* (p.A677G) mutation observed in human B-cell lymphoma (McCabe et al., 2012). In this way, the A677T-EZH2 variant dominant negatively decreases H3K27me2 levels. Notably, the individual with Weaver syndrome and this *EZH2* (p.A677T) mutation developed both neuroblastoma and acute lymphoblastic leukaemia at 13 months old. The key difference between the oncogenic (A677G) and Weaver syndrome (A677T) variant seems to be that A677T-EZH2 is impaired in its overall histone methyltransferase capacity, while A677G-EZH2 is not.

Further analysis of the ChIP-Rx data revealed that despite these apparently different mechanisms at PRC2-bound regions, all WS-models examined share reductions in H3K27me2 and H3K27me3 broadly throughout the genome which correlate with global increases in H3K27ac. Collectively these data suggest that non PRC2-bound gene body and intergenic regions may be more relevant to the molecular aetiology of Weaver syndrome than PRC2-bound CpG islands. This is consistent with my original hypothesis that defects in the landscape of chromatin modifications at intergenic regions will prove relevant to Weaver and related-OGID syndromes (section 1.10; Deevy and Bracken (2019)).

In this chapter I provide the first molecular evidence that dysregulated H3K27 acetylation may be a feature of Weaver syndrome. Additionally, I identify a core set of 43 genes that are commonly dysregulated across the studied cellular models of Weaver syndrome; pathway and process enrichment analysis indicates that these genes are highly relevant to

processes suspected to contribute to the Weaver syndrome phenotype. Specific examples include *Meis2* which encodes a transcription factor expressed in the developing brain and which has been associated with neuroblastoma – a common cancer amongst Weaver syndrome (Groß et al., 2018; Kondo et al., 2014; Zha et al., 2014). It is pertinent to note that the *Pbx1* gene was also observed to be upregulated in four out of five cellular models of Weaver syndrome, since the MEIS family of transcription factors are known to function through dimerization with PBX proteins (Schulte and Geerts, 2019). A second example is *Akt3*, a protein kinase abundantly expressed during human neurogenesis and which functionally links to the mTOR pathway that has been shown to be central to cell growth (Engelman et al., 2006; Ersahin et al., 2015; Poduri et al., 2012). Remarkably, gain-of-function mutations in *AKT3* are reported to cause macrocephaly – a common feature of Weaver syndrome (Alcantara et al., 2017). Moreover, *AKT3* is reported to be differentially methylated in Weaver syndrome patients (Choufani et al., 2020). Similarly, three genes from the core set have been previously linked to aberrant growth phenotypes in humans or mice: *Zfp521*, *Zfp629*, *Gja1* (Bult et al., 2019; Choi et al., 2018; Ohkubo et al., 2014; Pizzuti et al., 2004). Finally, the observed downregulation of *Cbx4* may be relevant as I predict that canonical PRC1 binding decreases in H3K27me3-deficient Weaver syndrome settings. Unfortunately, the datasets generated for this chapter are not sufficient to determine the mechanisms through which these genes become differentially abundant, nor have they provided evidence in support of a specific associated chromatin profile. Future studies should focus on addressing these important open questions.

CHAPTER 5:
DYSREGULATED HISTONE H3K27
ACETYLATION IS A RECURRENT
FEATURE OF CONGENITAL GROWTH
DISORDERS

5.1 Introduction

As introduced in chapter one, Weaver and Sotos syndromes show considerable phenotype overlap. So much so that in the early 2000s they were considered allelic conditions – both thought to be caused by constitutive mutations in *NSD1* (Baujat et al., 2005; Douglas et al., 2003; Rio et al., 2003). However, since then, *EZH2* has been identified as the primary gene implicated in Weaver syndrome and *NSD1* alterations have been shown to be specific to Sotos syndrome (Gibson et al., 2012; Tatton-Brown and Rahman, 2013; Tatton-Brown et al., 2005, 2011). Tatton-Brown et al. (2005) determined that *NSD1* mutations do not cause classic Weaver syndrome, and therefore recommend that a Weaver syndrome diagnosis only be given when an individual has been shown to be negative for *NSD1* aberrations; several individuals with *NSD1* mutations who were initially diagnosed with atypical Weaver syndrome at a young age were later re-classified as having Sotos syndrome.

The main points of phenotypic overlap between Weaver and Sotos syndrome include pre- and post-natal overgrowth, a distinctive facial appearance (specifically including a high, broad forehead and prominent chin), advanced bone age and variable intellectual disability (Tatton-Brown and Rahman, 2013). The facial features of these conditions become easier to differentiate as affected children get older. For example, classic Weaver syndrome cases have a rounder face and hypertelorism, almond shaped eyes with longer palpebral fissures compared to classic Sotos syndrome, in which the face is long and thin with down-slanted palpebral fissures (Tatton-Brown and Rahman, 2006, 2013). Additionally, children with *EZH2*-mutant Weaver syndrome are consistently tall and inconsistently macrocephalic (~50% of cases are not macrocephalic), while *NSD1*-mutant Sotos syndrome cases may or may not be tall and may or may not be macrocephalic (~10% are neither) (Tatton-Brown et al., 2005, 2013). Similarly, bone age is consistently advanced in *EZH2*-mutant Weaver syndrome but ~20% of *NSD1*-mutant Sotos syndrome cases do not share this phenotype (Tatton-Brown et al., 2005, 2013). Finally, although variable intellectual disability is common to both Weaver and Sotos syndromes, mild disability is most frequently reported in *EZH2*-mutant Weaver syndrome whilst moderate disability is most common amongst *NSD1*-mutant Sotos syndrome cases.

Despite the abovementioned differences, the high degree of phenotypic overlap between Sotos/Weaver syndrome can make it challenging for even the most experienced clinicians to accurately distinguish the two conditions. Although molecular testing for alterations in the *NSD1* and *EZH2* genes has greatly facilitated their differential diagnosis, the interpretation of variants of unknown significance remains a challenge. To this end, DNA methylation profiling is emerging as a valuable diagnostic tool. In 2015, Choufani et al.

showed that individuals with Sotos syndrome and heterozygous loss-of-function mutations in *NSD1* display a highly specific and sensitive genome-wide DNA methylation signature that could differentiate them from healthy controls, benign *NSD1* variants and – importantly – Weaver syndrome patients. Then, in a follow up study, the same lead author determined the genome wide DNA methylation signature of Weaver syndrome and showed that it sensitive and specific for PRC2-related overgrowth disorders (Choufani et al., 2020). Remarkably, the signature specificity was such that it could detect somatic mosaicism and distinguish loss-of-function from gain-of-function missense variants in *EZH2*. A key discovery by these researchers was that a novel *EZH2* variant (p.A733T), which was identified in an individual with growth restriction, exhibited increased global H3K27 methylation in an *in vitro* HMT assay and showed “opposite” DNA methylation changes to Weaver syndrome. In other words, they identified an *EZH2* variant which potentially drives opposite deviations in growth, PRC2 activity and DNA methylation compared to Weaver-syndrome-associated variants.

The identification of these distinct DNA methylation signatures for Weaver and Sotos syndromes may help shed further light on the differences between these disorders. However, I am interested in furthering our understanding of their shared molecular biology. As such, the primary aim of this research chapter was to identify common molecular signatures that may contribute to the phenotypic overlap of the genetically distinct Weaver and Sotos OGID syndromes; I hypothesised that such a signature would be centred on intergenic chromatin (section 1.10.5; Deevy and Bracken, 2019). In parallel, I wanted to interrogate the relevance of the Weaver-syndrome-associated molecular changes identified in chapter four through the comparative study of isogenic cellular models of *EZH2*-associated growth restriction, and *Nsd1*-mutant Sotos syndrome. A key goal was to identify cellular and molecular changes shared between the overgrowth models but contrary to the model of *EZH2*-associated growth restriction; I reasoned that those molecular aberrations functionally relevant to the growth phenotypes would conversely deviate from wild type in a cellular model of growth restriction versus overgrowth. A follow-up aim was to test whether any such changes are conserved in patient-derived human cell cultures.

In this chapter, I generate a cellular model of *EZH2*-associated growth restriction that is isogenic to the models of Weaver syndrome presented in chapter four; I show that this growth restriction model displays converse molecular changes – namely, H3K27 hyper-

trimethylation and H3K27 hypo-acetylation. I also provide evidence to suggest that Weaver-syndrome-associated EZH2 variants cause cell size to increase relative to wild type and growth-restriction-associated EZH2. In addition to this, I generate isogenic cellular models of *Nsd1* haploinsufficient Sotos syndrome and show that H3K27 acetylation levels are specifically increased in this context. Furthermore, by performing RT-qPCR analyses on all the above-mentioned cell models, I identify *Cntln* as a key gene whose transcript levels are specifically increased in both the overgrowth syndrome models but decreased in the *EZH2*-associated growth restriction model. I also show that increased *Akt3* transcript abundance is shared by cellular models of Weaver and Sotos syndrome, and identify three additional genes (*Zfp521*, *Cbx4*, *Tlx2*) that show opposing expression profiles in the growth restriction model versus Weaver syndrome models. Finally, I perform western blot and RT-qPCR analyses on non-isogenic human cell cultures derived from Weaver or Sotos syndrome patients or unaffected controls, and the preliminary data recapitulate key findings from mouse embryonic stem cells.

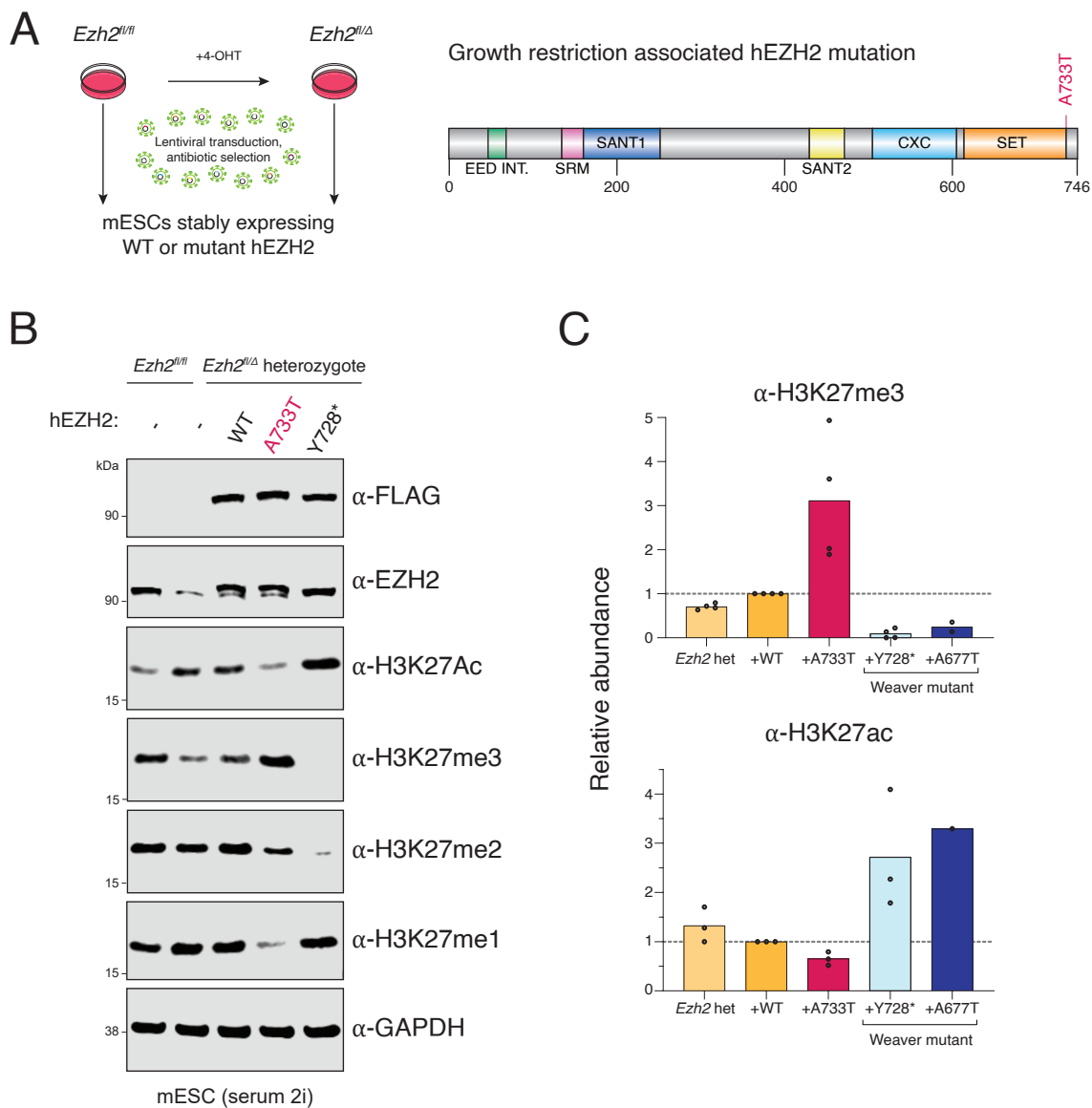


Figure 5.1 An isogenic mESC model of *EZH2*-associated growth restriction displays global H3K27 hyper-trimethylation and hypo-acetylation

(A) Left: Schematic outlining the experimental setup for the data presented in this figure. Right: A schematic of the human EZH2 protein (isoform 1; Q15910-1) with domains indicated, showing the 2D location of the growth-restriction-associated EZH2 mutation included in this study. EED int. = EED-interacting domain; SRM = stimulatory response motif.

(B) Representative western blots for FLAG, EZH2 and H3K27 methylation states, probing whole cell lysates from *Ezh2^{fl/\Delta}* mESCs with or without stable transgenic expression of indicated FLAG-tagged hEZH2 constructs.

(C) Relative quantification of western blot analyses of H3K27me3 and H3K27ac abundance. Signal intensity values were normalized to a loading control and then expressed relative to the *Ezh2^{fl/\Delta}* + hEZH2 WT cell line. The dashed lines represent the normalised *Ezh2^{fl/\Delta}* + hEZH2 WT values. Data values represent biological replicates and their mean values are plotted as bars.

5.2 Results

5.2.1 An isogenic mESC model of EZH2-associated growth restriction displays global H3K27 hyper-trimethylation and hypo-acetylation

Compared to the Weaver-syndrome-associated EZH2 variants studied in chapter four, I hypothesised that the growth-restriction-associated A733T-EZH2 variant might drive “opposite” molecular changes when expressed *in vivo*. To test this hypothesis, I utilized the isogenic rescue system described in section 4.2.4. I expressed either wild-type EZH2, the growth-restriction-associated A733T variant or a Weaver-syndrome-associated variant (Y728* or A677T) against the *Ezh2*^{fl/fl} mouse embryonic stem cell line (Fig. 5.1A). All mutants tested were successfully stably expressed at levels comparable to wild-type EZH2 in *Ezh2*^{fl/fl} cells (Fig. 5.1B).

Western blot analyses confirmed that the A733T variant operates as a gain-of-function mutation in cells, consistent with the prior observation that it exhibits increased HMTase activity *in vitro* (Choufani et al., 2020). Compared to the wild-type rescue, the A733T rescue cell line showed two- to five-fold increased H3K27me3, slightly decreased H3K27me2 and strongly decreased H3K27me1 (Fig. 5.1B-C). The A733T rescue cell line also showed global H3K27 hypo-acetylation, with an estimated twenty to fifty percent reduction compared to wild type (Fig. 5.1B-C). This directly opposes the H3K27 hyper-acetylation phenotype observed in cellular models of Weaver syndrome and, incidentally, the Y641F mutant rescue (see section 4.2.5).

Overall, these data support the notion that global H3K27 hyper-trimethylation and hypo-acetylation is associated with EZH2-mutant growth restriction, and vice versa for Weaver syndrome.

5.2.2 Growth-restriction-associated mutant hEZH2 assembles into PRC2

To provide a more thorough molecular characterization of the A733T-EZH2 mutation, I wanted to confirm that it was assembling into PRC2 in cells. To this end, I performed FLAG-immunoprecipitation and showed that endogenous EED and SUZ12 specifically co-immunoprecipitate with the epitope-tagged A733T variant EZH2 (Fig. 5.2A). Fluorescent immunocytochemistry confirmed that this variant was exclusively localised to the nucleus – as would be expected for a PRC2 component (data not shown). Next, ChIP-qPCR analysis indicated that the A733T-EZH2 variant was specifically bound the *Pax3* PRC2 target gene, which again supports the notion that it assembles into PRC2 (Fig. 5.2B).

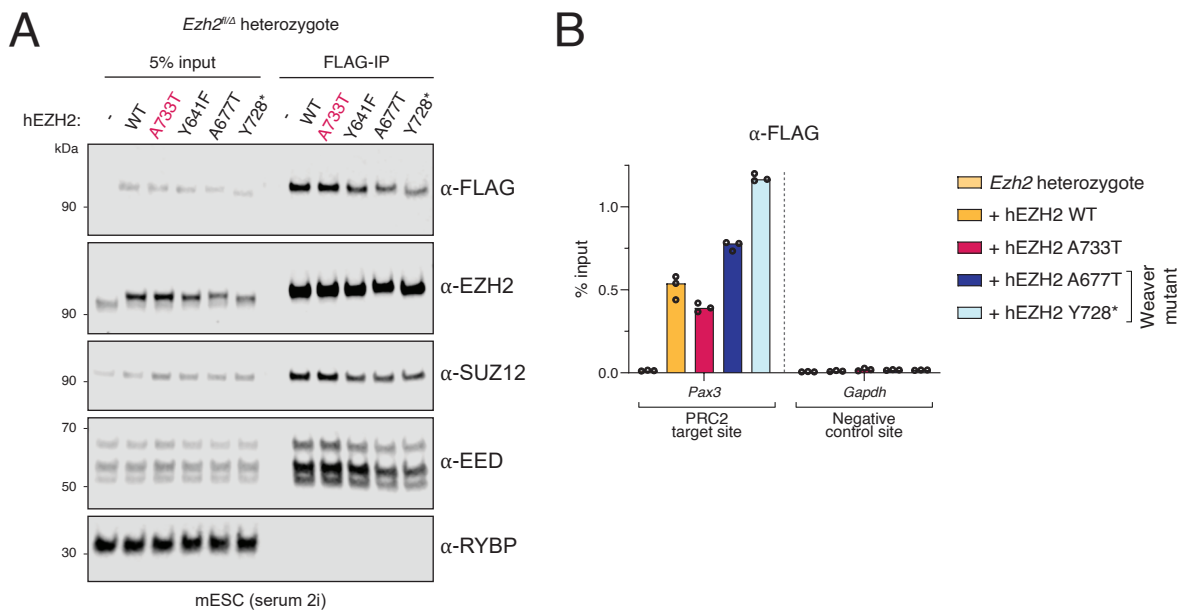


Figure 5.2 The growth-restriction-associated A733T-EZH2 variant assembles into PRC2 and binds its target sites

(A) Representative western blot for PRC2 subunits of cell extracts (input) and anti-FLAG immunoprecipitated material (FLAG-IP) from the indicated cell lines (*Ezh2^{fl/Δ}* mESCs with or without stable transgenic expression of indicated FLAG-tagged hEZH2 constructs). Western blotting for the variant PRC1 subunit RYBP is included as a negative control. The cell line expressing the oncogenic hyperactivating Y641F-EZH2 variant is included as a positive control.

(B) ChIP-qPCR signals for α-FLAG in the indicated cell lines (*Ezh2^{fl/Δ}* mESCs with or without stable transgenic expression of indicated FLAG-tagged hEZH2 constructs) at a PRC2 target site and non-target negative control region. Data values represent technical replicates and their median value is plotted.

Interestingly, it appeared as if the A733T variant EZH2 binds PRC2 target sites with slightly lower affinity than the wild type. The opposite was observed for Weaver-syndrome-associated mutations – they appeared to bind PRC2 target sites with higher affinity than the wild type.

Taken together, these data indicate that the growth-restriction-associated A733T-EZH2 variant assembles into functional PRC2.

5.2.3 Growth-restriction-associated mutant EZH2 drives reciprocal molecular changes compared to Weaver-syndrome-associated mutant EZH2

Once I had established that the growth-restriction-associated A733T-EZH2 variant could assemble into PRC2 and localise to its target sites, I next wanted to determine precisely how the H3K27 hyper-trimethylation phenotype manifests in this context. I also wanted to test whether this was truly a reciprocal molecular phenotype; will sites that lose H3K27me3 in Weaver syndrome instead gain H3K27me3 and/or lose H3K27ac in the growth restriction model? To address these questions, I performed genome-wide ChIP-Rx of H3K27me3, H3K27me2 and H3K27ac in *Ezh2^{d/d}* mESCs rescued with either wild-type, the growth restriction-associated A733T variant or the Weaver-syndrome-associated Y728* or A677T variant EZH2. In agreement with the western blot analyses presented section 5.2.1, the A733T rescue cell line showed a stark global increase in H3K27me3 at the chromatin level, with a concomitant global decrease in H3K27me2; 91.32% of the genome gained H3K27me3 but lost H3K27me2 (Fig. 5.3 A-C). An average plot profiling H3K27me3 at all PRC2 target sites validated previous findings that, compared to the wild-type rescue, the Weaver-syndrome-associated A677T mutant displays local hyper-trimethylation that contrasts against the hypo-methylation observed for the truncated Y728* variant at these sites (Fig. 5.3D). Rescue with the growth-restriction-associated A733T variant caused an approximate two-fold increase in H3K27me3 at PRC2 target sites compared to the wild-type rescue (Fig. 5.3D). It is important to note that this A733T-mediated hyper-trimethylation phenotype was not restricted to PRC2-bound peak centres, but rather extended throughout the genome causing diffuse accumulation of H3K27me3 without typical peaks or troughs (Fig. 5.3E). H3K27me2 levels were accordingly decreased throughout the genome (Fig. 5.3)

Colour coding: ■ *Ezh2*^{flΔ} + hEZH2 WT ■ *Ezh2*^{flΔ} + hEZH2 A733T ■ *Ezh2*^{flΔ} + hEZH2 Y728* ■ *Ezh2*^{flΔ} + hEZH2 A677T

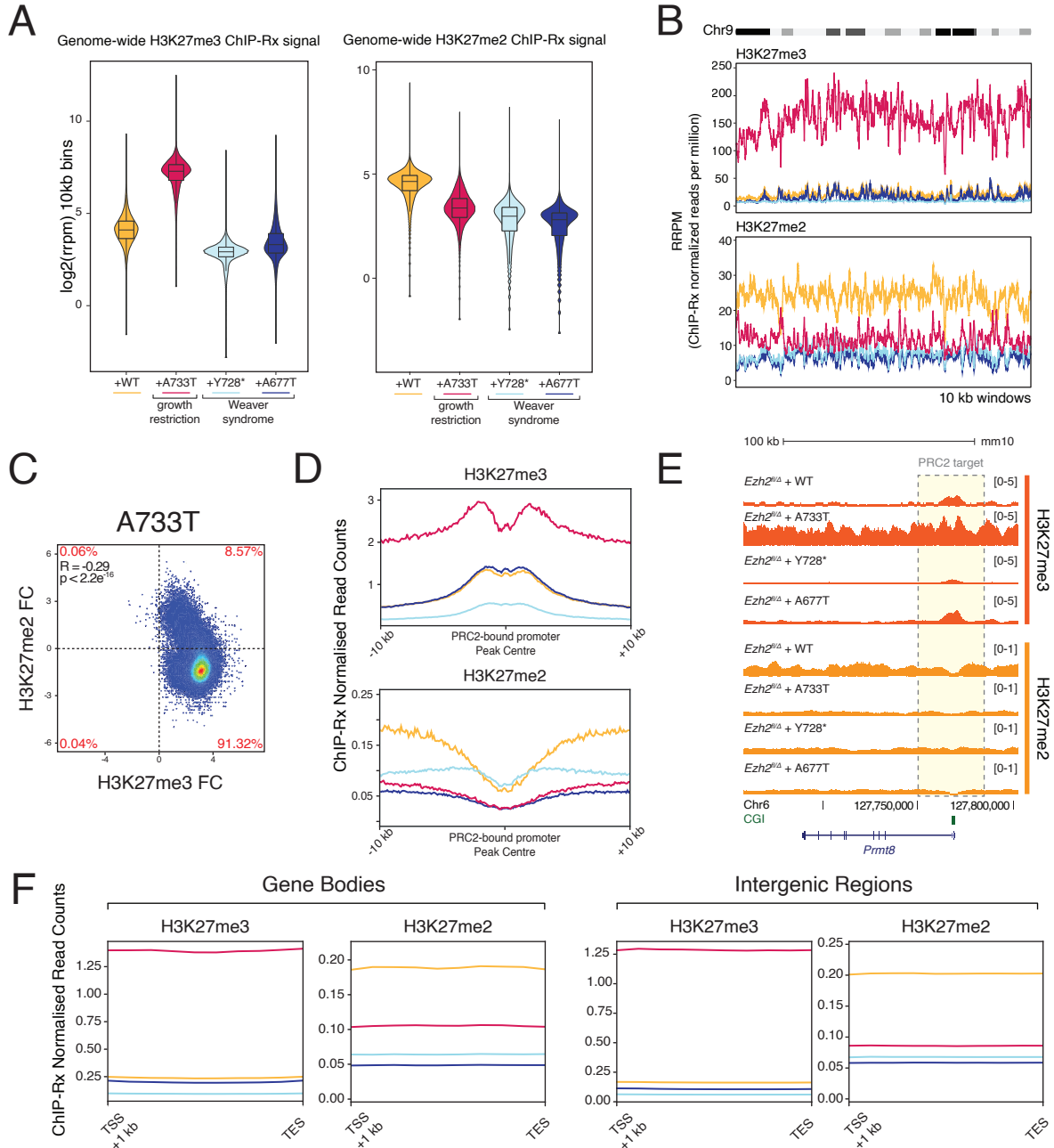


Figure 5.3 The growth-restriction-associated A733T-EZH2 variant causes a diffuse genome-wide increase in H3K27me3 with concomitant global reductions in H3K27me2

(A) Violin plot representations of the genome-wide log2 ChIP-Rx normalized reads per million of H3K27me3 and H3K27me2 enrichments across 1540218 or 1537656 10kb bins, respectively.

(B) Rolling average plots of the normalized reads per million of H3K27me3 and H3K27me2 ChIP-Rx enrichments across 12117 10kb windows of chromosome 9 (smoothened over 50 bins) in the indicated cell lines. Colour-coded legend as indicated at top of page.

(C) Scatter plot showing the relationship between changes compared to wild type in H3K27me2 against changes compared to wild type in H3K27me3 across 254781 10kb bins genome wide in *Ezh2*^{flΔ} mESCs with stable transgenic expression of the growth failure associated A733T-EZH2 variant. The percentage of 10kb bins in each quadrant is indicated. FC = fold-change.

(D) Average ChIP-Rx signal profiles for H3K27me3 and H3K27me2 at PRC2-bound promoters in *Ezh2*^{flΔ} mESCs with or without stable transgenic expression of indicated hEZH2 constructs. Colour-coded legend as indicated at top of page.

(E) Genome browser representations of ChIP-Rx normalised reads of H3K27me3 and H3K27me2 at a representative PRC2 target site (*Prmt8*) in the indicated cell lines. CGI = CpG island.

(F) Average ChIP-Rx signal profiles for H3K27me3 and H3K27me2 at gene bodies and intergenic regions in *Ezh2*^{flΔ} mESCs with or without stable transgenic expression of indicated hEZH2 constructs. Colour-coded legend as indicated at top of page. TSS = Transcription Start Site; TES = Transcription End Site.

Next, I turned my attention to H3K27ac which had appeared decreased at the western blot level in the A733T compared to wild-type rescue; the ChIP-Rx data confirmed the global loss of H3K27ac (Fig. 5.4A-B). However, the peak centres of average plots profiling H3K27ac enrichment at all promoter and enhancer sites revealed respectively increased or unchanged H3K27ac in the growth restriction model compared to the wild-type rescue (Fig. 5.4C top panels, Fig. 5.4D). Similarly, average plots profiling H3K27ac enrichment across all gene bodies showed that genic H3K27ac was just slightly decreased in the growth restriction model compared to the wild-type rescue (Fig. 5.4C, bottom panels). On the other hand, the growth restriction model showed an approximate one-third decrease in H3K27ac across intergenic regions (Fig. 5.4C, bottom panels, Fig 4.5D). A comparative analysis of the observed fold-change in H3K27ac versus H3K27me3 in the growth restriction model compared to wild-type rescue showed that 92.84% of sites examined lost H3K27ac and gained H3K27me3 in the A733T rescue (Fig. 5.4E). A similar comparative analysis of the observed fold-change in H3K27ac versus H3K27me2 showed that that H3K27ac was predominantly lost from sites that also lost H3K27me2 (86.07% of sites examined) (Fig. 5.4E). In other words, H3K27ac was predominantly lost from sites that lost H3K27me2 but gained H3K27me3.

Next, we directly examined sites corresponding to H3K27me3 peaks that were observed (in chapter 4) to lose ≥ 1.5 -fold enrichment averaged across the Weaver syndrome models. This analysis showed that H3K27me3 was indeed reciprocally increased at these sites in the growth restriction model (Fig. 5.5 A+C). These same sites also showed a relative depletion of H3K27ac in the growth restriction mutant compared to Weaver syndrome models (Fig. 5.5 B+D). However, it should be noted that the H3K27ac decrease at these sites was slight compared to wild type, and not concentrated on the peak centre (Fig. 5.5D).

Collectively, these ChIP-Rx data indicate that the growth-restriction-associated A733T-EZH2 variant drives reciprocal molecular changes compared to Weaver-syndrome-associated mutant EZH2, specifically at gene body and intergenic regions.

5.2.4 Assaying the relative cell size and proliferation rates of wild-type, Weaver syndrome and growth restriction mESC models

I next wished to move beyond molecular changes to start assessing relevant cellular phenotypes. I hypothesised that individuals with Weaver syndrome are larger than average because they possess more cells than the average person, and each of these cells is larger than normal. The corollary is that congenital growth restriction is the result of having too few

Colour coding: ■ *Ezh2^{flΔ}* + hEZH2 WT ■ *Ezh2^{flΔ}* + hEZH2 A733T ■ *Ezh2^{flΔ}* + hEZH2 Y728* ■ *Ezh2^{flΔ}* + hEZH2 A677T

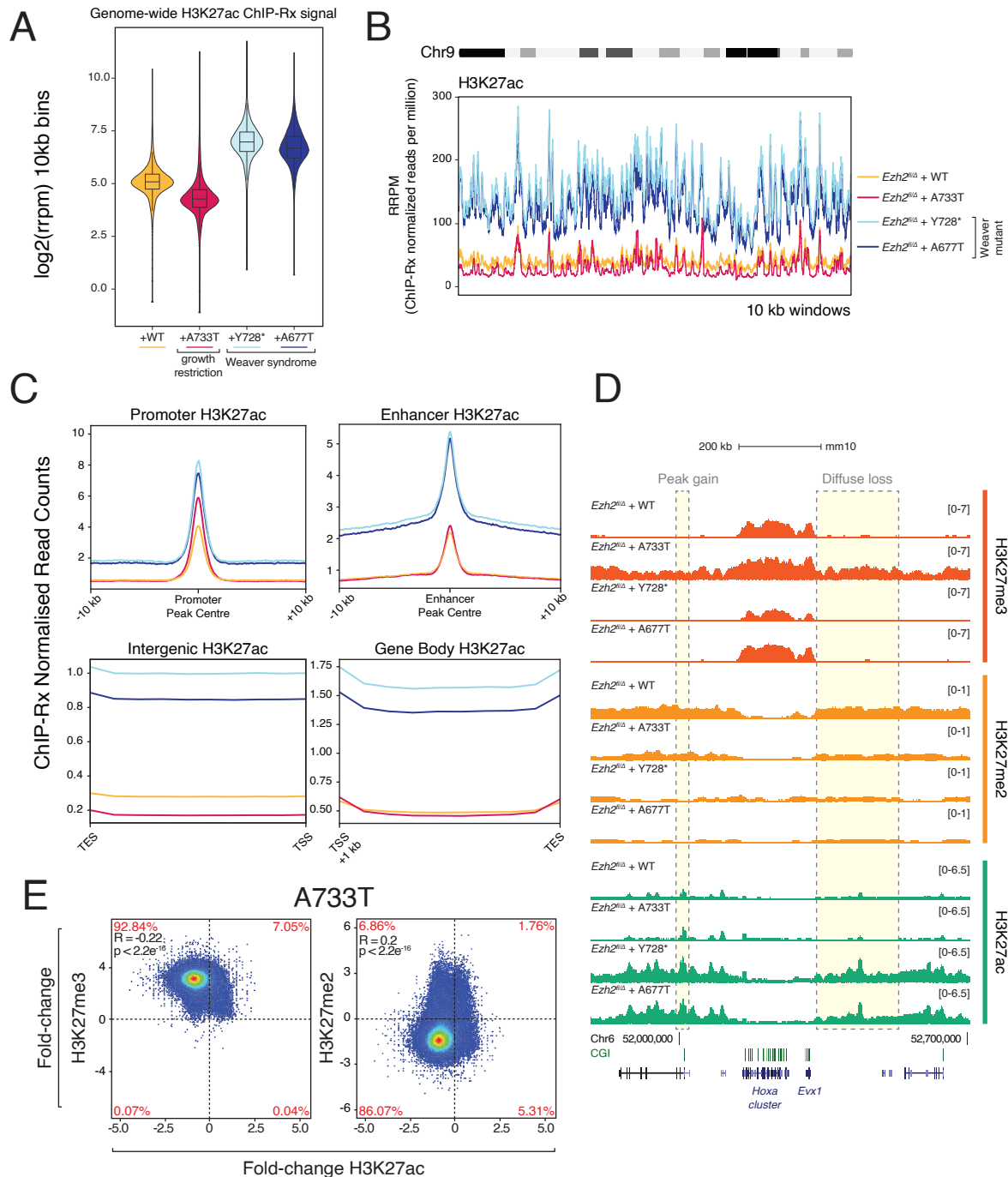


Figure 5.4 Diffuse H3K27ac is lost from the cellular model of *EZH2*-associated growth restriction

(A) Violin plot representations of the genome-wide log2 ChIP-Rx normalized reads per million of H3K27ac enrichments across 1540086 10kb bins.

(B) Rolling average plots of the normalized reads per million of H3K27ac ChIP-Rx enrichments across 12116 10kb windows of chromosome 9 (smoothed over 50 bins) in the indicated cell lines.

(C) Average ChIP-Rx signal profiles for H3K27ac at promoters, enhancers, intergenic regions and gene bodies in *Ezh2^{flΔ}* mESCs with or without stable transgenic expression of indicated hEZH2 constructs. Colour-coded legend as indicated at top of page. TSS = Transcriptional Start Site; TES = Transcriptional End Site.

(D) Genome browser representations of ChIP-Rx normalised reads of H3K27me3, H3K27me2 and H3K27ac at the *Hoxa* gene cluster in the indicated cell lines. Highlighted regions show areas of diffuse H3K27ac loss versus concentrated H3K27ac gain. CGI = CpG island.

(E) Scatter plots showing the relationship between changes compared to wild type in H3K27me3 and H3K27me2 levels against changes compared to wild type in H3K27ac levels across respectively 254825 or 254684 10kb bins genome wide in *Ezh2^{flΔ}* mESCs with stable transgenic expression of the growth failure associated A733T-EZH2 variant. The percentage of 10kb bins in each quadrant is indicated.

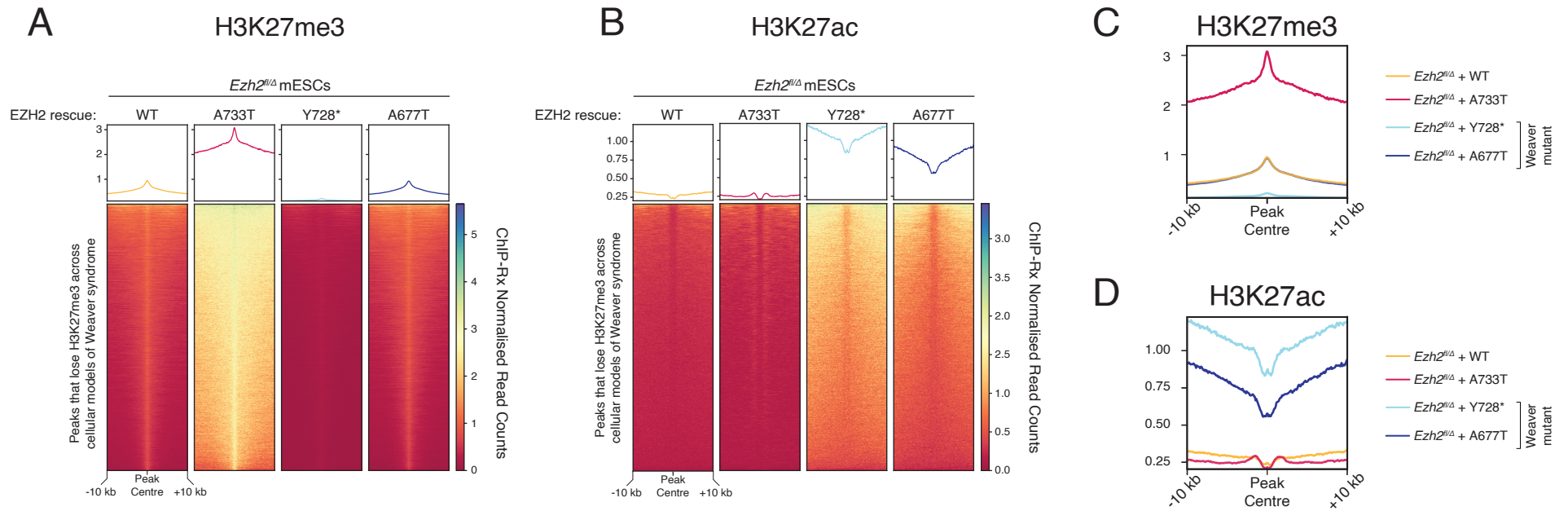


Figure 5.5 Reciprocal H3K27 modification changes in cellular models of *EZH2*-associated growth restriction versus Weaver syndrome

(A) Tornado plots showing average enrichments for H3K27me3 in *Ezh2^{fl/fl}* mESCs with stable transgenic expression of indicated hEZH2 constructs. Indicated are regions corresponding to genomic windows ± 10 kb around H3K27me3 peaks that were present in the *Ezh2^{fl/fl}* + WT EZH2 cell line and which, on average, lost ≥ 1.5 -fold enrichment in the cellular models of Weaver syndrome in the ChIP-Rx experiment presented in chapter 4 of this thesis. Scale bars denote reference-adjusted counts per million ChIP-Rx signal.

(B) Tornado plots showing average enrichments for H3K27ac in *Ezh2^{fl/fl}* mESCs with stable transgenic expression of indicated hEZH2 constructs. Indicated are regions corresponding to genomic windows ± 10 kb around H3K27me3 peaks that were present in the *Ezh2^{fl/fl}* + WT EZH2 cell line and which, on average, lost ≥ 1.5 -fold enrichment in the cellular models of Weaver syndrome in the ChIP-Rx experiment presented in chapter 4 of this thesis. Scale bars denote reference-adjusted counts per million ChIP-Rx signal.

(C) Overlaid average H3K27me3 ChIP-Rx signal profiles from panel A.

(D) Overlaid average H3K27ac ChIP-Rx signal profiles from panel B.

cells and/or smaller cells than normal. To this end, I devised a fluorescent flow cytometry-based assay for the simultaneous relative quantification of cell size and proliferation rates (Fig. 5.6A and described in section 2.1.6). This involved labelling the wild-type rescue cell line with a red fluorophore and three cellular models of Weaver syndrome (the Y728*, K563E and R679C rescues) or the growth restriction model (A733T rescue) with a green fluorophore. I also separately labelled the growth restriction model (A733T rescue) with the red fluorophore. Western blot analyses presented in Figure 5.6B show that the molecular phenotypes of interest for these cell lines (i.e., H3K27 modifications) were not globally altered by the introduction of the fluorescent markers. One exception, however, was the growth-restriction-associated A733T-EZH2 variant which did not express well with either the red or the green fluorophore; this caused some deviation from the expected profile of H3K27 modifications relative to the wild-type rescue cell line. This poor expression was likely because these cell lines were cultured without positive selection for the *EZH2* transgene for several days prior to this protein extraction. It is possible that expression of the A733T variant confers a growth disadvantage such that negative clones will out compete it at the population level in the absence of positive selection. For this reason, I cultured all cell lines under positive selection for expression of the *EZH2* transgene immediately prior to setting up the experiment as depicted in Figure 5.6A. Briefly, I combined red and green cell populations together and then measured their relative cell size (median forward scatter area, FSC-Alin) and abundance (percentage of total population) over time using a benchtop flow cytometer.

The data shows that the cellular models of Weaver syndrome were significantly larger relative to their wild-type counterparts, by ~8-18% on average (Fig. 5.7A). The cellular model of growth restriction (A733T rescue) also appeared significantly larger than the wild-type rescue, though the difference between the means of these two populations ($\Delta = 0.05155$) was smaller than the differences observed between the wild type and Weaver syndrome populations ($\Delta = 0.08399, 0.1253$ and 0.1771). The cellular models of Weaver syndrome were also significantly larger relative to the growth restriction model, by ~12-14% on average (Fig. 5.7B). It is possible that the green fluorophore itself was contributing the observed increases in cell size and so it will be necessary to repeat the experiment with the fluorophores swapped for each cell line. It is also possible that the A733T variant was still poorly expressing throughout this experiment, which would confound the interpretation of the results.

In terms of cellular proliferation, the results are inconclusive (Fig. 5.8). Although there was a general trend for the cellular models of Weaver syndrome to slightly increase in proportion

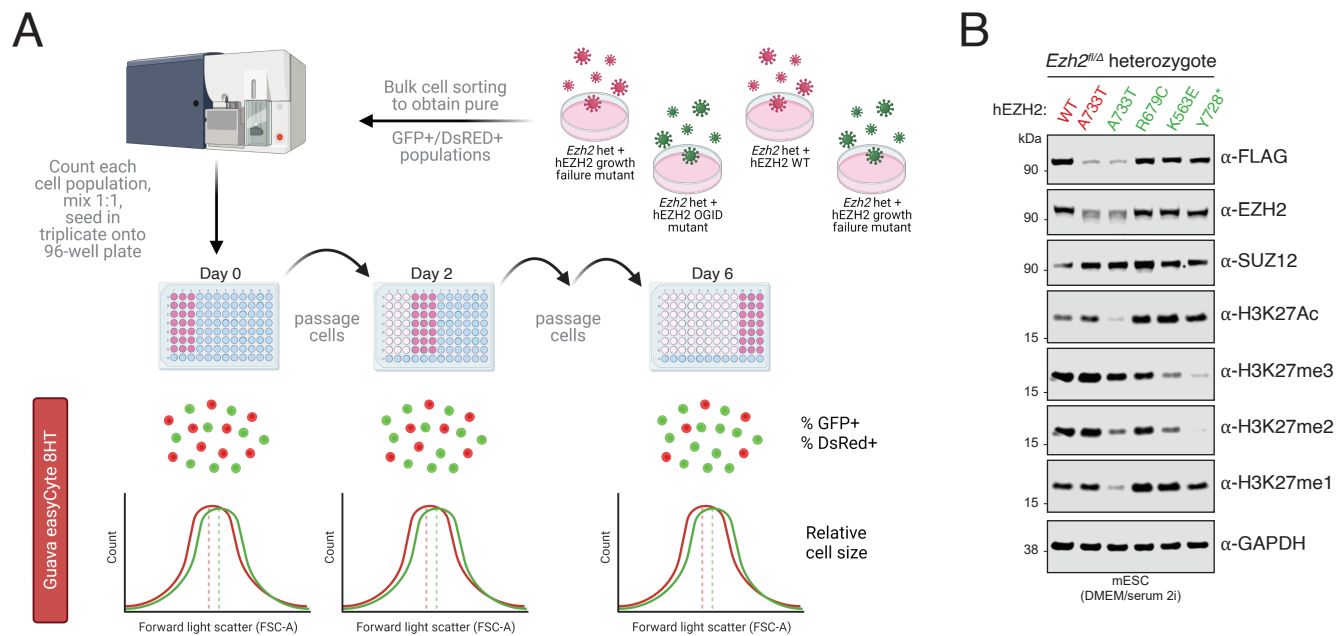


Figure 5.6 Flow-cytometry-based method for assessing relative cell size and proliferation rate

(A) A schematic outlining the experimental procedure for assessing the relative cell size and proliferation rate of my cell lines of interest. Created with BioRender.com.

(B) Western blot analyses of PRC2 subunits and H3K27 modifications, probing whole cell lysates from *Ezh2*^{fl/Δ} mESCs with stable transgenic expression of indicated FLAG-tagged hEZH2 constructs and a red or green fluorescent marker (as per colour coding). Note that in this western blot the A733T mutant was not expressing well with either the red or the green fluorophore, and accordingly the profiles of H3K27 modifications were altered.

Relative Cell Size

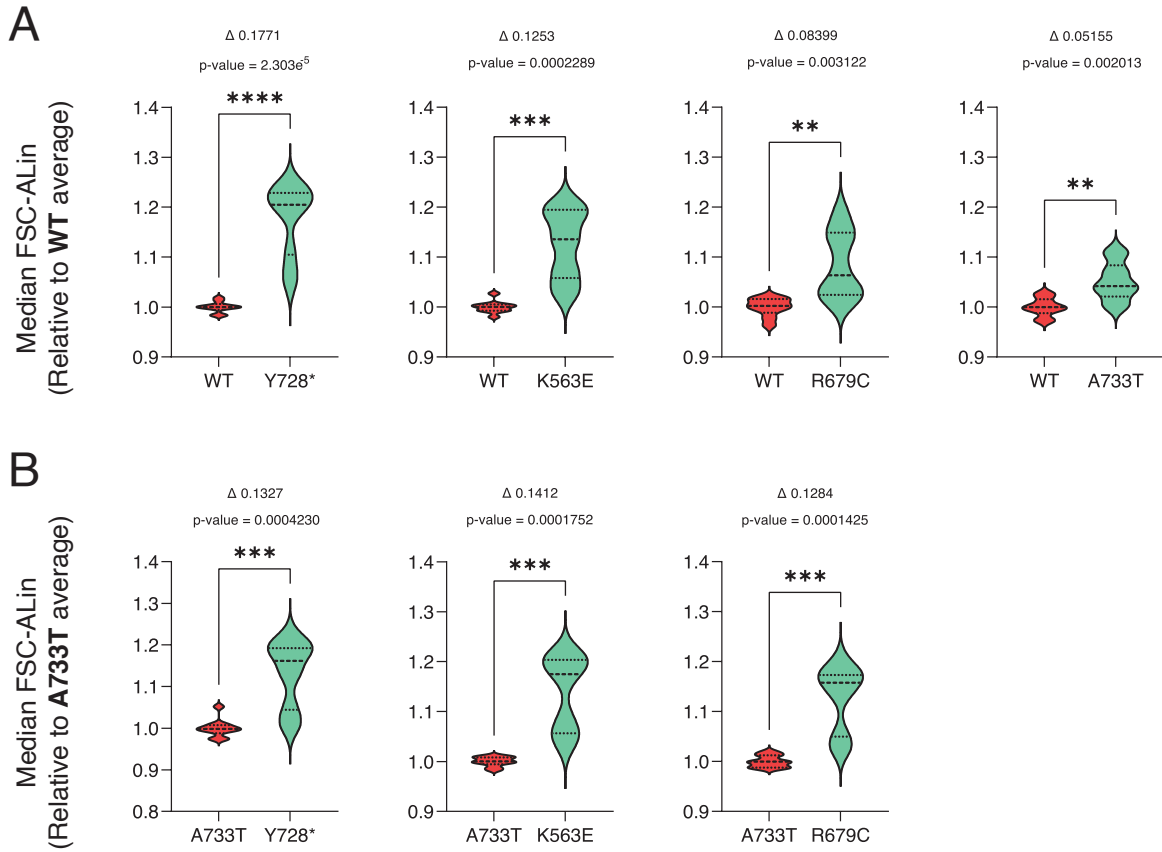


Figure 5.7 Analysis of relative cell size

(A) Violin plots of the relative cell sizes of wild-type versus growth-altered-mutant mESCs. Dashed lines represent the median and quartiles.

(B) Violin plots of the relative cell sizes of growth-restriction-mutant (A733T) versus Weaver-syndrome-mutant mESCs. Dashed lines represent the median and quartiles.

Data values come from a single biological experiment, with ten technical replicate values per cell line for each comparison. The corresponding Δ and p-values are displayed above each graph. The Δ value represents the difference between the means of each pair. The statistical significance of this difference was determined by unpaired t-test with Welch's correction. The asterisks denote the p-value classification as follows: ns = not significant, p-value ≥ 0.05 ; * = significant, p-value 0.01-0.05; ** = very significant, p-value 0.001- 0.01; *** = extremely significant, p-value 0.0001-0.001; **** = extremely significant, p < 0.0001.

Relative Cell Proliferation Rate

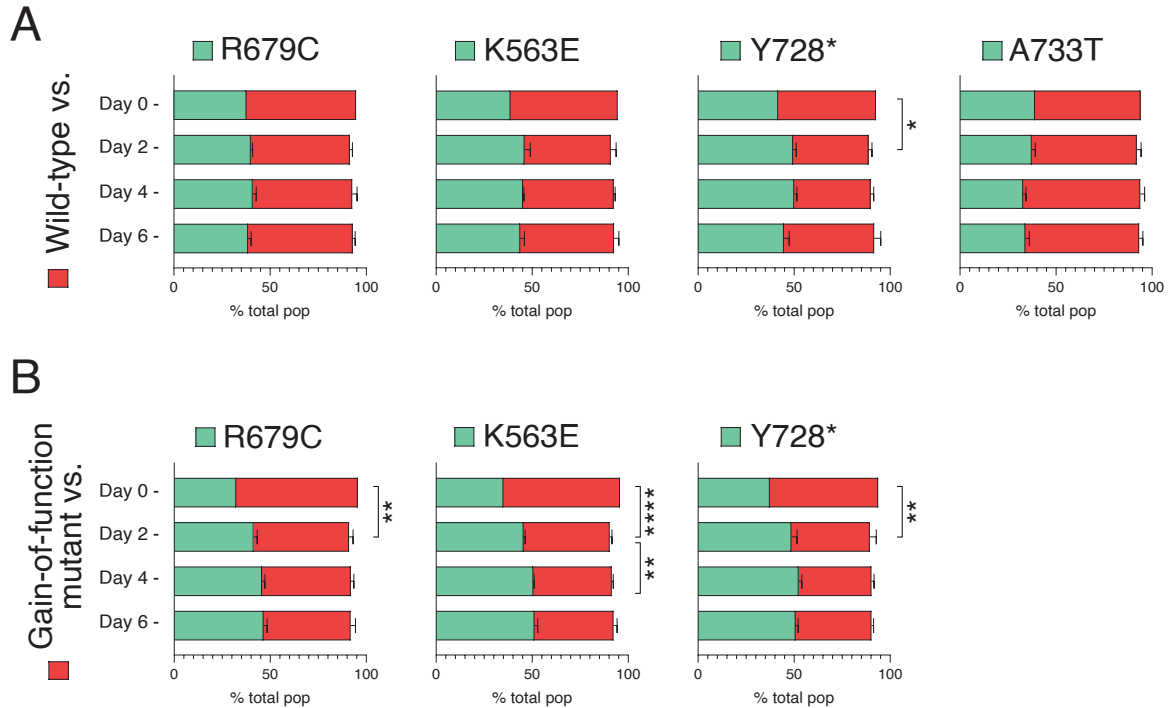


Figure 5.8 Analysis of relative cell proliferation rates

(A) Stacked bar charts of the percentage of wild-type versus growth-altered mutant mESCs at days 0, 2, 4 and 6 of the experiment.

(B) Stacked bar charts of the percentage of gain-of-function mutant (A733T) versus Weaver syndrome mutant mESCs at days 0, 2, 4 and 6 of the experiment.

Data values come from a single biological experiment, with three technical replicates per cell line per time point - excepting day 0 for which only a single measurement was taken. Mean values are plotted with error bars representing the standard deviation. For each cell line within a pair, the mean value of a time point (6, 4 or 2 days) was compared to the mean value of the preceding time point. The statistical significance of any difference between these two mean values was determined by an ordinary two-way ANOVA with Turkey's multiple comparisons test. P-values are displayed by classification only where the difference between the mean values of two consecutive time points was calculated as significant for both cell lines within a pair; the lowest classification of the two p-values is indicated. Asterisks denote p-value classifications as follows: ns = not significant, p-value ≥ 0.05 ; * = significant, p-value 0.01-0.05; ** = very significant, p-value 0.001- 0.01; *** = extremely significant, p-value 0.0001-0.001; **** = extremely significant, p < 0.0001.

over time relative to the wild-type rescue (Fig. 5.8A), the apparent competitive growth advantages were neither statistically significant nor consistent across all time points. When comparing the Weaver syndrome versus growth restriction models, their proportions tended to respectively increase and decrease over time – suggesting that either (a) Weaver-syndrome-associated EZH2 variants confer a competitive growth advantage, (b) the growth-restriction-associated A733T variant confers a proliferative disadvantage or (c) both (Fig. 5.8B). Relative to the wild-type rescue, the growth restriction model decreased in proportion over time, though not in a statistically significant manner.

In summary, these data support the hypothesis that Weaver syndrome affected cells are larger than normal but further validation of the assay is required.

5.2.5 Generation and validation of an isogenic mESC model of *Nsd1* deletion mutant Sotos syndrome

I wanted to test whether my findings regarding congenital growth disorders were specific to *EZH2*-mutant syndromes or more broadly applicable to congenital growth disorders. To this end, I decided to generate a cellular model of Sotos syndrome. Haploinsufficiency of *NSD1* is a major cause of Sotos syndrome (Kurotaki et al., 2002). As such, the heterozygous knockout of *Nsd1* in a cell line is – at least at the genetic level – an appropriate cellular model for Sotos syndrome. I used CRISPR–Cas9-mediated genome editing to generate such a cell line. Importantly, I decided to target *Nsd1* in the *Ezh2*^{fl/fl} mouse embryonic stem cells such that the resulting cellular model of Sotos syndrome would be against the exact same genetic background as my cellular models of Weaver syndrome and EZH2-associated growth restriction.

Figure 5.9A presents the *NSD1* gene locus in humans, highlighting exons which Douglas et al. (2005) showed to be deleted from four unrelated individuals with Sotos syndrome. Figure 5.9B depicts the *Nsd1* gene locus of mice, with the homologous region highlighted. My gene targeting strategy was to delete ~2 kb from this region (Fig. 5.9C) to (a) mimic the partial gene deletions observed in real-life Sotos syndrome cases and (b) destroy the translational start site shared by all annotated *Nsd1* isoforms and, in this way, functionally inactivate the gene.

I first screened clones for the desired deletion at the gDNA level; two separate genotyping PCR reactions were required to independently check for the presence or absence of the wild-type or deletion allele products (Fig. 5.9D). Next, I performed a series of RT-qPCR analyses to assess whether my gene targeting strategy had successfully impaired transcription

of the *Nsd1* gene. Using primer pairs located within and downstream of the deleted region (*Nsd1*Δ), I confirmed that corresponding mRNA transcripts were reduced by about half (Fig. 5.9 E-F). Finally, western blot analyses confirmed that NSD1 abundance was decreased by approximately fifty percent in the *Nsd1*+/ Δ clones, and totally absent from the *Nsd1*Δ/ Δ homozygous knockout clones (Fig. 5.9G).

Overall, I successfully validated four independent *Nsd1* heterozygous knockout cell lines and three independent *Nsd1* homozygous knockout cell lines.

5.2.6 Cellular models of Sotos syndrome phenocopy the H3K27 hyperacetylation signature observed in Weaver syndrome models

Once I had confirmed the successful generation of *Nsd1* knockout cell lines, I next assessed their relative histone modification levels by western blot analyses. Three independent *Nsd1* wild-type (gEV) cell lines (n=3), four independent *Nsd1* +/ Δ cell lines (n=4) and three independent *Nsd1* Δ/ Δ cell lines (n=3) were subject to these analyses. Western blot signals were quantified, and the relative signal intensity values were pooled according to the respective *Nsd1* genotype (WT, Het or KO); the significance of any change between the mean values for each knockout genotype versus wild-type was determined using Brown-Forsythe and Welch ANOVA tests with Dunnett T3 correction for multiple comparisons.

I first investigated H3K36me2 and H3K27me3 levels; we have previously reported that H3K36me2 levels decrease while H3K27me3 levels increase in cells depleted of NSD1 (Streubel et al., 2018). H3K36me2 did not decrease in either heterozygous or homozygous *Nsd1* knockout clones, but H3K27me3 was significantly increased in response to both the heterozygous and homozygous loss of *Nsd1* (Fig. 5.10A). I then examined the H3K27ac and H3K27me2 modifications which had been respectively up- and down-regulated consistently across the different cellular models of Weaver syndrome studied in chapter four (Fig. 5.10B). The heterozygous and homozygous *Nsd1* knockout cells displayed a non-significant increase in H3K27me2. Remarkably, H3K27ac was significantly elevated specifically in the *Nsd1* heterozygous knockout cell lines that model Sotos syndrome; this observation was consistent across four biologically independent *Nsd1* +/ Δ clonal cell lines.

I decided to further analyse all the above histone modifications and SUZ12 by ChIP-qPCR in *Nsd1* wild-type (*Nsd1* +/+), *Nsd1* +/ Δ and *Nsd1* Δ/ Δ cell lines (Fig. 5.10C). In this sensitive assay for local histone modification abundance, H3K36me2 was observed to decrease in a dose-dependent fashion with the loss of *Nsd1*. SUZ12 was enriched at all sites

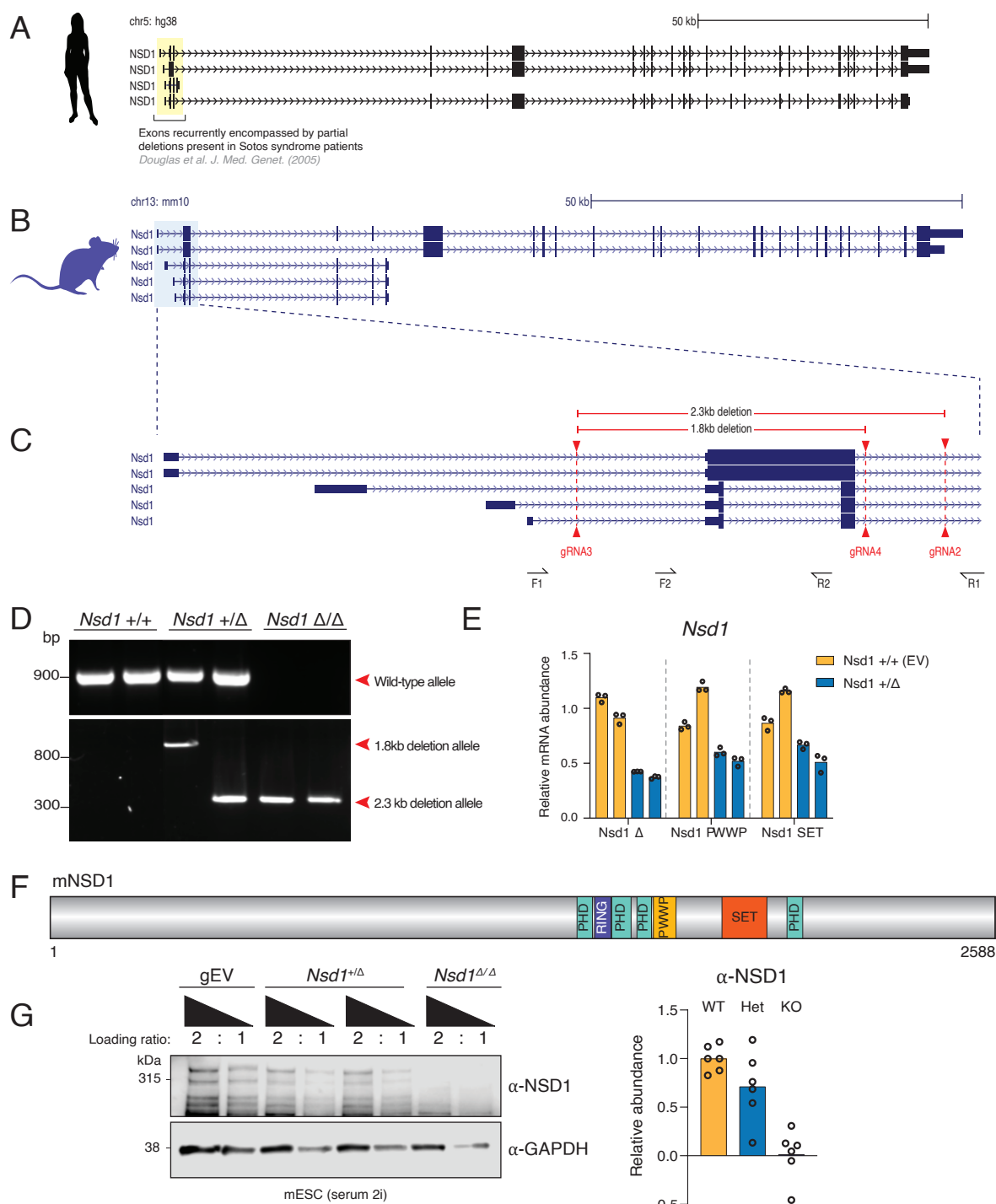


Figure 5.9 Generation and validation of single-cell-derived *Nsd1* heterozygous knockout clones

(A) Graphical depiction of the human *NSD1* gene locus.

(B) Graphical depiction of the mouse *Nsd1* gene locus. The highlighted region is homologous to that in panel A.

(C) A schematic representation of the *Nsd1* targeting strategy; paired gRNAs flank the translational start site to induce a ~2kb CRISPR/Cas9-mediated deletion.

(D) Representative genotyping PCR using the primers indicated in panel C to identify heterozygous and homozygous *Nsd1* knockout clones.

(E) RT-qPCR analysis using primers targeted to the deleted region (*Nsd1* Δ), PWWP or SET domain-encoding exons of *Nsd1* in two biologically independent wild-type (EV, *Nsd1*+/+) and heterozygous knockout (*Nsd1* +/-) clones. Bars represent the mean value and technical triplicate data values are shown.

(F) Schematic of the mouse NSD1 protein (O88491) with domains indicated.

(G) Left: A representative western blot for NSD1, probing whole cell lysates from representative heterozygous and homozygous *Nsd1* knockout clones. Right: Relative quantification of western blot analyses of NSD1 abundance. Signal intensity values were normalised to a loading control and then expressed relative to the wild-type EV control. Data values from biologically independent clones were pooled according to *Nsd1* genotype. Bars represent the mean value for a given genotype.

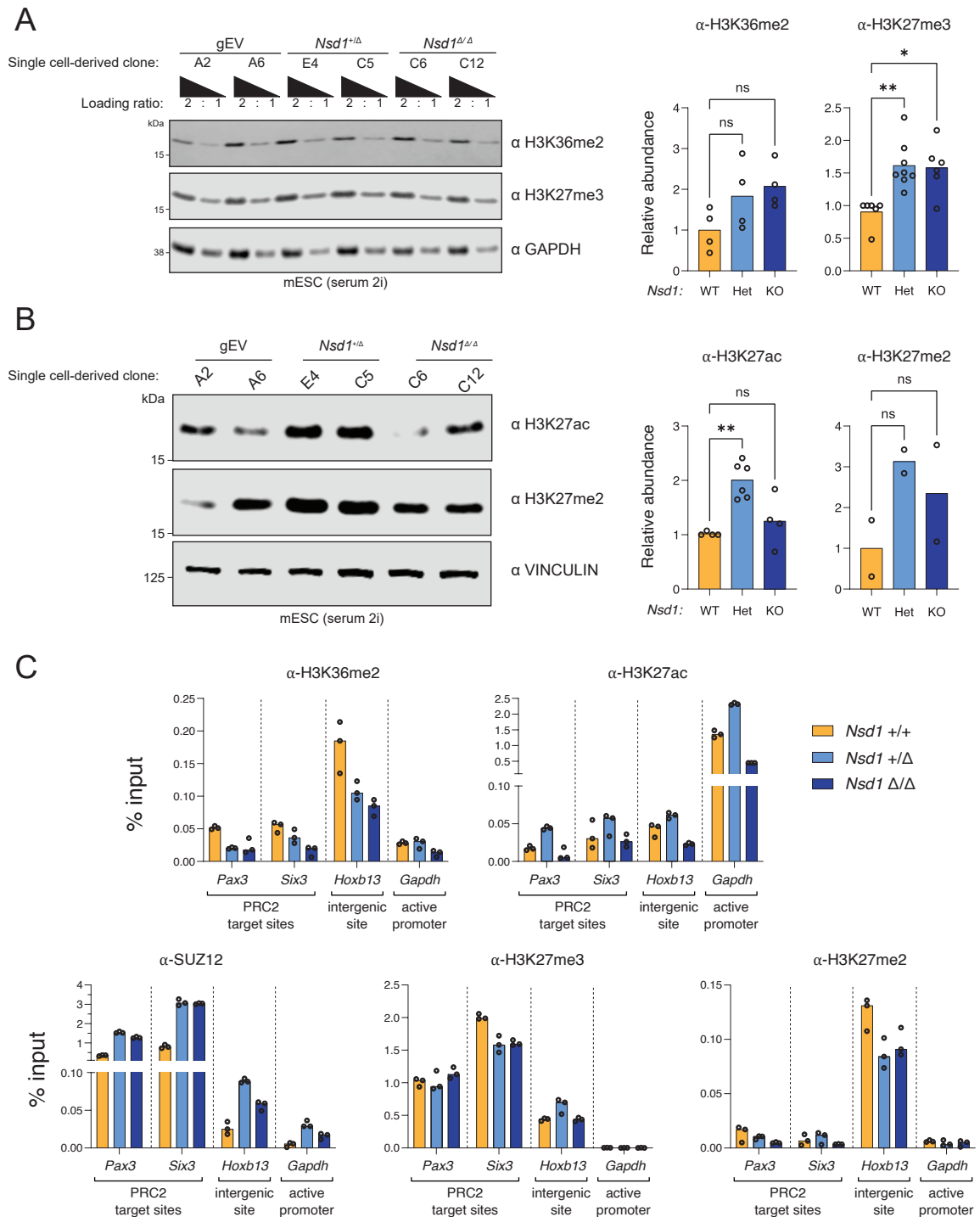


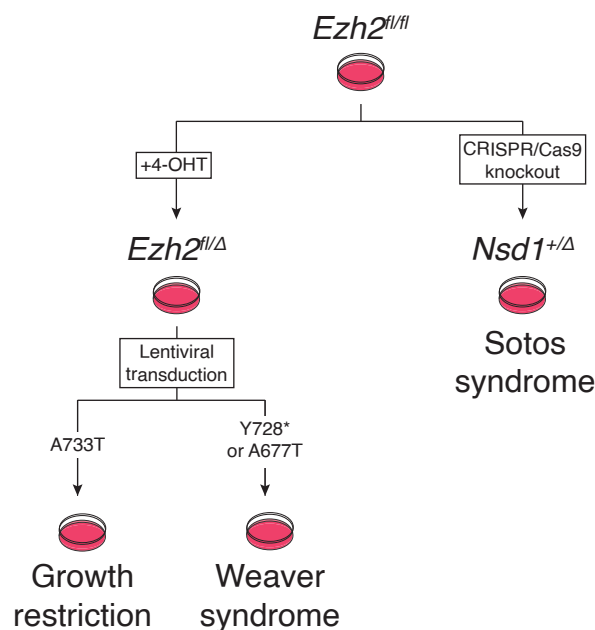
Figure 5.10 Isogenic models of *Nsd1* +/- Sotos syndrome exhibit a specific global increase in H3K27ac

(A) Left: Representative western blots for H3K36me2 and H3K27me3, probing whole cell lysates from *Nsd1* wild type (gEV, WT), heterozygote knockout (*Nsd1* +/-Δ, Het) or homozygote knockout (*Nsd1* Δ/Δ, KO) mouse embryonic stem cell lines. Right: Relative quantification of western blot analyses of H3K36me2 and H3K27me3. Signal intensity values were normalised to a loading control and then expressed relative to the *Nsd1* wild type (gEV) values. Data values have been collated from independent biological replicates and their mean values plotted.

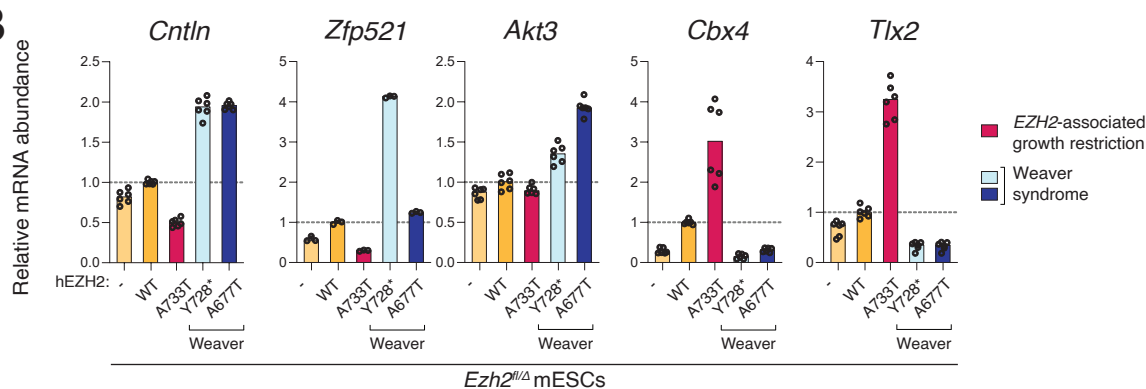
(B) Left: Representative western blots for H3K27ac and H3K27me2, probing whole cell lysates from *Nsd1* wild type (gEV, WT), heterozygote knockout (*Nsd1* +/-Δ, Het) or homozygote knockout (*Nsd1* Δ/Δ, KO) mouse embryonic stem cell lines. Right: Relative quantification of western blot analyses of H3K27ac and H3K27me2. Signal intensity values were normalised to a loading control and then expressed relative to the *Nsd1* wild type (gEV) values. Data values have been collated from independent biological replicates and their mean values plotted.

(C) ChIP-qPCR enrichments for the indicated histone modifications and SUZ12 in *Nsd1* wild type (*Nsd1* +/+), heterozygote knockout (*Nsd1* +/-Δ) or homozygote knockout (*Nsd1* Δ/Δ) cell lines at PRC2 target sites, an intergenic region and an active promoter. Data values come from a representative biological experiment with three technical replicate values shown and the median value plotted.

A



B



C

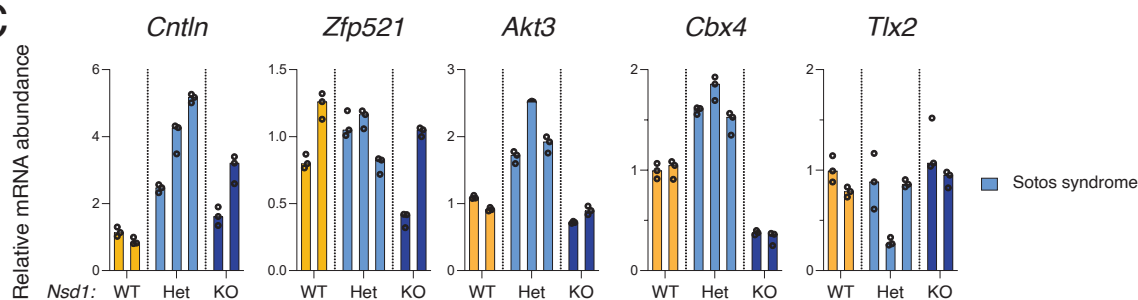


Figure 5.11 Conserved and divergent gene expression changes in isogenic cellular models of Weaver syndrome, Sotos syndrome and *EZH2*-associated growth restriction

(A) Schematic outlining the isogenic nature of the mouse embryonic stem cell models of Weaver syndrome, Sotos syndrome and *EZH2*-associated growth failure.

(B) RT-qPCR analyses of the indicated genes in *Ezh2*^{fl/Δ} mESCs with or without stable transgenic expression of indicated hEZH2 constructs. Data values come from one or two biological replicate experiments with technical triplicate values shown for each. Bars represent the median value. Dashed lines represent the normalised *Ezh2*^{fl/Δ} + hEZH2 WT values.

(C) RT-qPCR analyses of the indicated genes in *Nsd1* wild type (gEV, WT), heterozygote knockout (*Nsd1*^{+/-Δ}, Het) or homozygote knockout (*Nsd1*^{Δ/Δ}, KO) mouse embryonic stem cell lines. Data from two or three biologically independent clones are shown for each genotype with technical triplicate values plotted. Bars represent the median value.

tested in NSD1-depleted cells, but H3K27me3 only increased at the intergenic site in the *Nsd1* +/ Δ cell line. H3K27me2 conversely decreased at this intergenic site, in both the heterozygous and homozygous *Nsd1* knockout cell line. Notably, H3K27ac increased at all sites tested in the *Nsd1* heterozygous but not homozygous knockout cell line

Taken together with my findings in the cellular models of Weaver syndrome and growth restriction, these observations led me to speculate that dysregulated H3K27ac is a recurrent feature of congenital growth disorders.

5.2.7 Conserved and divergent gene expression changes in Weaver syndrome, Sotos syndrome and growth restriction mESC models

Having shown that H3K27ac increases in two genetically distinct cellular models of congenital overgrowth but decreases in a cellular model of growth restriction, I next wanted to investigate the functional relevance of this finding. I had previously identified a core cohort of expressed genes that were differentially abundant in cellular models of Weaver syndrome (section 4.2.10; Fig. 4.12). I selected a few key genes from this cohort and tested their relative abundance by RT-qPCR in the cellular models of growth restriction and Sotos syndrome. Remarkably, even with this low-throughput method I identified 2 genes (*Cntln* and *Zfp521*) that were consistently downregulated in the growth restriction mutant but upregulated in the Weaver syndrome models, and an additional 2 genes (*Cbx4* and *Tlx2*) that were conversely upregulated in the growth restriction model but downregulated in the Weaver syndrome cell lines (Fig. 5.11B). I additionally identified two genes – *Cntln* and *Akt3* – that were specifically upregulated in both Sotos syndrome and Weaver syndrome models (Fig. 5.11C). The *Cbx4* gene was specifically upregulated in the Sotos syndrome cell models despite being downregulated in the Weaver syndrome context.

In summary, I identified a small number of genes whose differential abundance may be directly linked to congenital growth disorders. Although the functional significance of these genes in this context remains to be determined, their further study may generate invaluable insight into the molecular pathology of congenital growth disorders. Furthermore, the data presented herein do not confirm whether these shared and divergent gene expression changes are a direct consequence of altered H3K27ac.

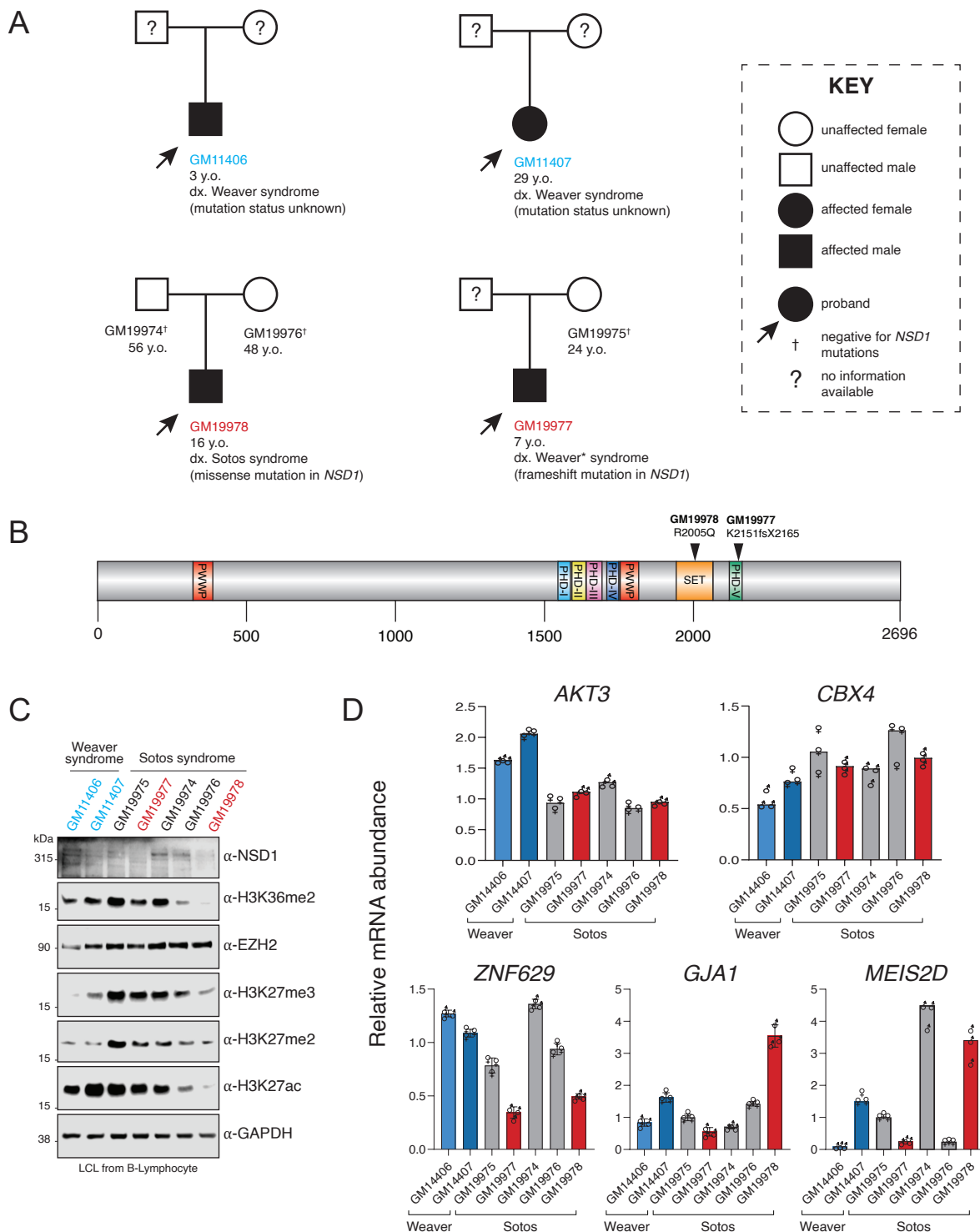


Figure 5.12 Preliminary characterisation of human lymphoblastoid cell lines derived from Weaver and Sotos syndrome-affected individuals

(A) Schematic overview of the individuals from which human lymphoblastoid cell lines (LCLs) were derived and acquired from Coriell Biorepositories. For each individual the age at sampling, original diagnosis and mutation status is given.

*GM19977 was initially diagnosed with Weaver syndrome but would now be considered to have Sotos syndrome.

(B) Schematic of the human NSD1 protein with domains annotated, showing the 2D location of the mutations identified in GM19977 and GM19978 by Douglas et al. (2003).

(C) Representative western blots for the indicated proteins and histone modifications, probing whole cell lysates from the indicated lymphoblastoid cell lines. Affected individuals are coloured in blue (Weaver syndrome) or red (Sotos syndrome).

(D) RT-qPCR analysis of *AKT3*, *CBX4*, *ZNF629*, *GJA1* and *MEIS2D* in the indicated lymphoblastoid cell lines. Data values represent technical replicates and bars represent their median value. Affected individuals are coloured in blue (Weaver syndrome) or red (Sotos syndrome).

5.2.8 Preliminary characterization of human lymphoblastoid cell lines derived from Weaver and Sotos syndrome-affected individuals

I next wanted to explore the relevance my findings from mouse embryonic stem cell models in a more physiologically relevant cell type. To this end, I sourced lymphoblastoid cell lines (LCLs) that had been derived from Weaver or Sotos syndrome patients or their unaffected parents (Fig. 5.12A). Two cultures were derived from individuals diagnosed with Weaver syndrome (GM11406 and GM11407), but unfortunately neither of them had their diagnoses confirmed at the molecular level nor had they matched parental material available. A third culture (GM19977) from an individual diagnosed with Weaver syndrome possessed a heterozygous frameshift mutation in *NSD1*, and cells from their unaffected mother were available (GM19975) (Douglas et al., 2003). As discussed in the introduction to this chapter, this individual should be reclassified as having Sotos syndrome. Finally, there was one culture from an individual diagnosed with classic Sotos syndrome who presented with a heterozygous missense mutation in affecting the SET domain of *NSD1* (GM1998), and for whom matched cell lines from both unaffected parents were available (GM19974 and GM19976).

Western blot analyses (Fig. 5.12B) confirmed that *NSD1* protein levels were decreased in both cell lines derived from *NSD1*-mutant individuals (GM19977 and GM19978). H3K36me2 was also decreased in these cell lines relative to their matched unaffected parent-derived cell lines. At the time of writing – and to the best of my knowledge – this is the first demonstration of impaired H3K36 dimethylation in cells with an overgrowth-associated *NSD1* mutation. Notably, however, H3K27me3, H3K27me2 and H3K27ac also appeared decreased in these *NSD1*-mutant individuals compared to their unaffected parents, which raises questions regarding the specificity of the observed H3K36me2 depletion.

Western blot analyses confirmed that *EZH2* was expressed in both cell lines derived from individuals diagnosed with Weaver syndrome but without a molecular diagnosis (GM11406 and GM11407); considering this, it might be worthwhile to Sanger sequence the *EZH2* open reading frame from both samples to check whether these individuals possess a coding mutation in *EZH2*. It seems likely, since their H3K27me2/3 levels were the lowest of all samples tested. These two cell lines displayed elevated H3K27ac when compared to the unaffected GM19974 and GM19976 cell lines, but the unaffected GM19975 cell line displayed similar levels of H3K27ac.

Finally, I performed RT-qPCRs on RNA isolated from these cell cultures to test the relative transcript abundance of human orthologues of the key genes identified as

differentially expressed in mouse ESC models of Weaver syndrome. Of the genes whose primers were technically successful (Fig. 5.12C), only two showed arguable specific differential abundance – *AKT3* and *CBX4*. Consistent with the observations made in mouse embryonic stem cells, *AKT3* appeared ~1.5 to 2-fold upregulated in the classic Weaver syndrome cases (GM14406 and GM14407) compared to the average of all unaffected controls. Conversely, the same two cell lines showed ~20-50% reduced *CBX4* mRNA levels compared to the average of all unaffected controls. These data are consistent with the findings in mESC models of Weaver syndrome. However, inconsistent with the mESC models of Sotos syndrome, no enrichment of *AKT3* abundance was observed in the Sotos syndrome LCLs.

Overall, it is difficult to draw firm conclusions from this data, due at least in part to the non-isogenic nature of these human cell lines. It is likely that different target genes will be deregulated in differentiated cell types compared to pluripotent stem cells. To this end, RNA-sequencing of these cell lines would be beneficial.

5.3 Discussion

The overarching aim of this research chapter was to identify common molecular signatures that may contribute to the phenotypic overlap of the genetically distinct Weaver and Sotos overgrowth syndromes. I hypothesised that such a signature would be centred on intergenic chromatin, and that the most relevant molecular changes would be conversely altered in an isogenic model of growth restriction.

In this chapter, I first generate a mouse embryonic stem cell model of *EZH2*-associated growth restriction that is isogenic to the cellular models of *EZH2* mutant Weaver syndrome presented in chapter 4. I show that the growth-restriction-associated *EZH2* variant (p.A733T) assembles into the PRC2 multimer and can be detected at a PRC2 target with slightly reduced enrichment compared to wild type. I confirm that the growth-restriction-associated *EZH2* variant (p.A733T) behaves as a gain-of-function mutation in cells, adding to existing evidence from *in vitro* HMT assays (Choufani et al., 2020). Collectively, my data indicate that A733T-*EZH2* exhibits increased HMT activity both at PRC2 target genes and through the ectopic conversion of H3K27me2 to H3K27me3 broadly throughout the genome. As such, the apparent decreased enrichment of A733T-*EZH2* at PRC2 targets could be a function of its increased ectopic engagement with H3K27me2 domains. Accordingly, H3K27me3 is diffusely increased genome wide and global H3K27 dimethyl and acetyl levels are concomitantly reduced. However, H3K27ac is retained (or even slightly increased) at promoter and enhancer peaks, therefore the bulk of its loss is from intergenic regions and – to a lesser extent – gene bodies. Taken together with my findings from chapter four, these data support the hypothesis that imbalanced profiles of H3K27 modifications at intergenic and gene body regions may contribute to the aberrant growth phenotypes of Weaver syndrome and *EZH2*-associated growth restriction. Specifically, these data suggest that increased H3K27me3 and decreased H3K27ac contributes to growth restriction and vice versa for Weaver syndrome. It is worth noting that although decreased H3K27me2 is common to both cellular models of Weaver syndrome and *EZH2*-associated growth restriction, the mechanisms through which this is achieved are not the same. In Weaver syndrome models, H3K27me2 is globally decreased as a result of impaired HMT activity whereas the model of *EZH2*-associated growth restriction shows decreased H3K27me2 because of abnormally efficient HMT activity.

Moving from the molecular to the cellular level, I investigated growth-relevant phenotypes in the isogenic mESC models of Weaver syndrome and *EZH2*-associated growth restriction. I hypothesised that Weaver syndrome affected individuals are larger than average because they possess more cells than the average person and each of these cells is larger than

normal, and that the reverse is true for *EZH2*-associated growth restriction. As such, I designed an assay to test their relative cell size and proliferation rates. I show that cellular models of Weaver syndrome exhibit significantly increased cell size relative to their wild-type and growth-restricted counterparts. Notably, the relative increase in cell size appeared to correlate with the degree of H3K27 hypomethylation observed for that cell line; the severe Y728* truncating mutation caused the largest relative increase in cell size, the moderate K563E mutation caused an intermediate relative increase in cell size while the mild R679C mutation gave the smallest relative increase in cell size. However, I also observed the *EZH2*-associated growth restriction model to exhibit increased cell size relative to wild type. This would argue against a direct link between cell size and the H3K27 methylation changes identified in Weaver syndrome versus growth restriction mESC models. Furthermore, a flaw in my experimental design (discussed in section 5.2.4) means that the observed increases in cell size could be an artefact of the assay. As such, it will be necessary to repeat the experiment under optimized conditions. In designing this assay, I assumed that the potential increased number of cells in Weaver-syndrome-affected individuals would be the result of increased cellular proliferation, and vice versa for *EZH2*-associated growth restriction. My data neither supports nor rejects this hypothesis – more biological replicates are required for a conclusive result. It is possible that an increased number of cells instead results from the delayed differentiation and/or enhanced self-renewal of pluripotent progenitor populations – the converse of what was observed in cerebral organoid models of microcephaly (Lancaster et al., 2013). Therefore, a key parallel future experiment will be to test this hypothesis using relevant *in vitro* differentiation protocols.

A key output of this research project was the generation of mouse embryonic stem cell models of *Nsd1* deletion mutant Sotos syndrome, isogenic to the cellular models of *EZH2* mutant Weaver syndrome presented in chapter four. There are some inconsistencies between mine and previously reported *Nsd1*-depleted mESC models (Streubel et al., 2018). Primarily, the *Nsd1* knockout cells presented in this chapter did not exhibit global loss of H3K36me2 – though a dose-dependent decrease was observed at representative genomic loci by ChIP-qPCR. It is possible that this difference can be accounted for by the alternative NSD1 depletion methods; I generated genetic knockouts of *Nsd1*, whereas the prior study utilised shRNA-mediated knockdown. There is a precedent for such inconsistency: shRNA-mediated knockdowns of *Mtf2* or *Phf19* in mESCs were shown to cause global reductions in H3K27me3 at the western blot level (Ballaré et al., 2012; Brien et al., 2012; Casanova et al., 2011) whereas the genetic knockout of *Mtf2* alone or *Mtf2*, *Phf19* and

Phf1 in combination did not (Healy et al., 2019; Højfeldt et al., 2019). On the other hand, Fang et al. (2021) used CRISPR/Cas9 to induce frameshift indels to homozygously knock *Nsd1* out in mESCs and detected a global decrease in H3K36me2 by western blot. In this case perhaps the discrepancy between mine and their models is because they cultured cells in the absence of two small molecule inhibitors (2i) – a cell culture additive known to induce DNA hypomethylation with downstream effects on PRC2 distribution which could in turn impact NSD1 function (McLaughlin et al., 2019; van Mierlo et al., 2019). Regardless, H3K27ac hyper-acetylation emerged as consistent molecular phenotype specific to the *Nsd1* heterozygous deletion mutant cell lines – a phenotype shared with the cellular models of Weaver syndrome and opposed to that of the growth restriction model. As a priority, the next step in this research project will be to perform ChIP-Rx of H3K36me2 and the H3K27 modifications in the *Nsd1* wild-type and heterozygous mutant cell lines; an important question to address will be whether H3K27ac is lost from target (promoter/enhancer) and/or gene body and intergenic sites in the Sotos syndrome model.

A central goal of this research chapter was to identify molecular changes shared between the overgrowth models but contrary to the model of *EZH2*-associated growth restriction. I identify *Cntln* as a gene whose expression levels meet this criterion. Not much has been published on *Cntln* to date. *Centlein* (*Cntln*) is so-called because it was first identified as a **centrosomal protein** (Fang et al., 2014; Makino et al., 2008). It has also been shown to be microtubule associated protein whose overexpression protects against microtubule depolymerisation and inhibits neurite formation from neuroblastoma-derived cell lines (Jing et al., 2016). Most recently, *Cntln* has been identified as a key component of sperm head-to-tail coupling apparatus (Zhang et al., 2021b). It is not yet clear how differential *Cntln* abundance may contribute to the growth phenotypes, future detailed studies will be required to elucidate its function in this context. A crucial next step will be to perform transcriptomic studies on both the mESC models of *EZH2*-associated growth restriction and Sotos syndrome, as an unbiased approach to identify additional genes that meet the abovementioned criterion. Additionally, other chromatin features (e.g., DNA methylation, alternative histone modifications, chromatin accessibility or chromatin architecture) ought to be explored at these loci to help determine the cause of their differential abundance.

The final aim of this research chapter was to test whether any key molecular changes identified in mouse stem cell models of Weaver and Sotos syndrome are conserved in patient-derived human cell cultures; I acquired an invaluable panel of human lymphoblastoid cell lines to address this question. Unfortunately, it was only possible to complete precursory characterisation of these cell lines within the time constraints of this

PhD research programme and as such, firm conclusions cannot yet be drawn. Nevertheless, I generated important preliminary data that will inform and guide future investigations. Most notably, I provide evidence in support of increased *AKT3* and decreased *CBX4* transcript abundance as a conserved feature of patient-derived and transgenic mESC models of Weaver syndrome. Despite their non-isogenic nature, it would be beneficial to assess the differential transcriptomes of the patient-derived versus unaffected cell lines – indeed, any conserved changes identified in the face of such genetic diversity are likely to be of utmost significance.

CHAPTER 6:

GENERAL DISCUSSION

6.1 Discussion

6.1.1 Summary of findings

In chapter three, I independently validated PALI1 as a positive regulator of PRC2 histone methyltransferase (HMT) activity and determined it likely dispensable for the *de novo* recruitment and targeting of PRC2 to CpG islands in mouse embryonic stem cells. Furthermore, I provided evidence supporting a role for PALI1 in regulating the correct balance of H3K27 methylation states at intergenic chromatin, both *in vitro* and *in vivo*. I established that *Pali1/Lcor* and *Pali2/Lcor1* mRNA transcripts are predominantly co-expressed throughout the developing mouse embryo. Intriguingly, however, the *Pali2* single knockout mice were found to be viable and fertile in contrast to the perinatal lethal phenotype of *Pali1* single knockout and *Pali1/2* double knockout mice. I speculate that PALI1 can compensate for the loss of PALI2 during development, but not vice versa. In addition, I showed that PRC2 HMT activity is similarly impaired in both *Pali1* single knockout and *Pali2* single knockout mouse embryos, and thereby provide the first *in vivo* evidence in support of PALI2 as a positive regulator of PRC2 activity. Notably, I also demonstrated the inefficacy of chromatin immunoprecipitation (ChIP) methods to detect PALI1 enrichment throughout the genome. I hypothesise that this is due to transient PALI1 association with chromatin (explored in detail in section 6.1.2 below), though this remains to be experimentally tested. In sum, the data presented in this chapter provide new insight into the molecular and biological functions of PALI-family proteins.

In chapter four, I modelled 10 different Weaver-syndrome-associated *EZH2* mutations in mESCs and confirm that they are predominantly – though not exclusively – partial loss-of-function with respect to global PRC2 HMT activity; I also determined that these mutations exert varying degrees of dominant negative effects. Intriguingly, I found that the different pathogenic *EZH2* variants can cause increased or decreased H3K27me3 at PRC2-bound CpG islands. Despite these differences, I identified a chromatin signature common across all Weaver syndrome mutations studied: broad reductions in H3K27me2/3 at non-PRC2 target sites, coupled with pervasive H3K27 hyper-acetylation. I additionally identified a small cohort of genes whose differential transcript abundance was shared across the different Weaver syndrome models, and amongst which one would expect to find key drivers of the Weaver syndrome phenotype. Overall, this research has contributed significant new knowledge towards understanding the molecular aetiology of Weaver syndrome

In chapter five, I generated cellular models of *EZH2*-associated growth restriction and *Nsd1* deletion mutant Sotos syndrome to compare against the mESC models of Weaver syndrome

studied in chapter four. Importantly, all cell lines were isogenically matched. I showed that global H3K27 hyper-acetylation was common to both models of congenital overgrowth (Weaver/Sotos syndrome) while an opposing H3K27 hypo-acetylation phenotype was observed in the growth restriction model. I showed that the growth restriction model involved a gain-of-function mutation in *EZH2* which caused indiscriminate increases in H3K27me3 but specific loss of diffuse H3K27ac domains. Taken together with the findings from chapter four, the data supports a model in which imbalances in H3K27 modifications at intergenic chromatin may contribute to aberrant growth phenotypes. I designed and tested a flow-cytometry-based assay to detect aberrant growth phenotypes at the cellular level. I additionally identified several genes whose differential abundance was shared between cellular models of Weaver and Sotos syndrome, or oppositely changed in the cellular models of *EZH2*-associated growth restriction versus Weaver syndrome. Future detailed study of these genes may shed light on the pathways and processes central to congenital growth disorders. Finally, I validate key findings from mESC models in human cell lines derived from Weaver and Sotos syndrome-affected individuals. Collectively, the data presented in chapter five reveal that dysregulated H3K27ac is a recurrent feature of congenital growth disorders and provides molecular insight into the shared biology of Weaver and Sotos syndromes.

6.1.2 *PAL11-PRC2*

I hypothesise that attempts to ChIP PALI1 (as described in chapter three) have yielded no obvious enrichment profile because PALI1 may exhibit less stable binding on chromatin compared to other PRC2 accessory proteins, such as MTF2 or JARID2. I further speculate that PALI1 binds chromatin in such a transient manner because it may be involved in the regulation of PRC2-mediated H3K27 methylation pervasively throughout the genome. The variant PRC1 components PCGF3/5 – which are similarly difficult to capture by ChIP – set a precedent for this hypothesis. Fursova et al. (2019) determined that PCGF3/5 were together responsible for the deposition of genome-wide low-level H2AK119ub, despite being dispensable for the maintenance of H2AK119ub peak enrichment at PRC1-bound sites. Then, in a follow-up study, Huseyin and Klose (2021) showed that PCGF3 was not stably bound to chromatin but rather dynamically and iteratively interacting with it, consistent with its role in depositing H2AK119ub diffusely throughout the genome. Given the inefficacy of ChIP to detect PALI1 on chromatin, as presented in this thesis, I propose that PALI1-PRC2 analogously deposits low-level H3K27 methylation broadly throughout the genome. Supporting this hypothesis, I observed H3K27 methylation to be altered at

intergenic sites in both *Pali1* knockout mESCs and E11.5 *Pali1* single-knockout mouse embryos, despite the high level of functional redundancy among PRC2 accessory proteins (Fig. 3.10 and 3.12C). Furthermore, Højfeldt et al. (2019) showed that only dispersed H3K27me3 enrichment (at non-PRC2 bound sites) remains in a cell line depleted of the MTF2, EPOP, AEBP2 and JARID2 PRC2 accessory components (PE+AJ knockout cells). Of the PRC2 accessory components identified to date, only PALI1/2 and EZHIP are still present in the PE+AJ knockout cell line, but EZHIP is known to impede rather than promote PRC2-mediated H3K27 methylation (Jain et al., 2020). A key experiment to test the proposed hypothesis would be to remove PALI1 from the existing PE+AJ knockout cell line and assess whether H3K27 methylation is further reduced and/or diffused in its absence. An important parallel experiment would be to follow the HaloTag approach used by Huseyin and Klose (2021) to quantitatively determine the chromatin binding kinetics of PALI1-PRC2.

6.1.3 Dysregulated H3K27ac is a recurrent feature of congenital growth disorders

The cellular models of congenital growth disorders studied in chapters four and five of this thesis all share dysregulated H3K27 acetylation; the overgrowth models exhibited H3K27 hyper-acetylation while the growth restriction model displayed H3K27 hypo-acetylation. Remarkably, isogenic cellular models of Imagawa–Matsumoto – generated by another PhD student in the Bracken lab – showed similarly elevated global levels of H3K27ac (unpublished data, not shown). These findings are significant given that none of the introduced pathogenic mutations directly target the p300/CBP histone acetyltransferases (HATs) responsible for H3K27 acetylation (Weinert et al., 2018). Intriguingly, loss-of-function mutations that do directly affect the *CREBBP* or *EP300* genes cause a rare autosomal dominant condition called Rubinstein–Taybi syndrome (RSTS) characterised by short stature, along with broad thumbs and halluces, distinctive facial features, and mild-to-severe intellectual disability (Cantani and Gagliosi, 1998; Kalkhoven et al., 2003; Murata et al., 2001; Roelfsema et al., 2005). Gamu and Gibson (2020) speculate that the reciprocal skeletal phenotypes observed in Weaver syndrome versus RSTS may be related to their opposing H3K27 writer function; the data presented in this thesis support that hypothesis. Collectively, these findings indicate that dysregulated H3K27ac is a recurrent feature of genetically distinct congenital growth disorders.

The RNA-seq data presented in chapter four did not identify an increased abundance of *Crebbp* or *Ep300* transcripts across any cellular model of Weaver syndrome, nor was there

evidence for decreased abundance of transcripts coding for histone deacetylases (HDACs) – the one exception being *Hdac7*, which was approximately two-fold decreased in abundance solely in the A677T-variant model of Weaver syndrome (data not shown). As such, the global increase in H3K27ac observed in these cell lines is likely best explained by the broad loss of H3K27me2/me3 providing additional substrate for p300/CBP to act upon – a similar phenomenon has recently been described in H3.3-K27M-expressing neural stem cells (Brien et al., 2021). Conversely, the significant global increase in H3K27me3 caused by the growth-restriction-associated *EZH2* mutation may dominantly deplete available substrate for p300/CBP. In a similar way, the re-introduction of wild-type EZH2 (and H3K27 methylation) to H3K27 hyper-acetylated *Ezh1/2* double-knockout mESCs results in attenuated global H3K27ac levels (Fig. 4.5; Lavarone et al., 2019). Having said that, transcriptomic analysis remains to be performed on the *EZH2*-associated growth restriction cell model so the possibility of altered HAT/HDAC expression in this context remains open.

It is not clear how NSD1 depletion may lead to increased global H3K27ac levels despite increased global H3K27me3, but it is interesting to note that the Y641F change-of-function EZH2 variant also drives increased H3K27me3 and H3K27ac (data presented herein, and independent unpublished data from the Bracken lab). Both NSD1 depletion and oncogenic change-of-function *EZH2* mutations cause decreased H3K27me2, thus one might speculate that the loss of H3K27me2 drives H3K27ac accumulation (McCabe et al., 2012; Sneeringer et al., 2010; Streubel et al., 2018). Indeed, previous studies have proposed the same (Lavarone et al., 2019). However, arguing against this, the gain-of-function A733T-EZH2 variant studied in this thesis caused global reductions in both H3K27me2 and H3K27ac. As such, alternative mechanisms must be considered. For example, homozygous knockout of *Nsd1* in mESCs was recently reported to promote increased local H3K27ac at active enhancers as a result of impaired HDAC1 (histone deacetylase 1) recruitment (Fang et al., 2021). Thus, it is possible that impaired HDAC1 recruitment similarly contributes to the observed H3K27 hyper-acetylation in my *Nsd1* heterozygous knockout mESCs. Transcriptomic analysis of the Sotos syndrome cell models will provide answers regarding HAT/HDAC expression levels.

Interestingly, EZH2 was recently reported to directly interact with p300 via a transactivation domain (TAD) overlapping its SRM and SANT1 domains (Jiao et al., 2020). This TAD is normally sequestered and therefore unable to function. However, phosphorylation of EZH2 at serine residue 21 (S21) was identified as one possible mechanism through which the TAD domain can be ‘unlocked’ (Jiao et al., 2020). AKT-

family kinases are reported to phosphorylate EZH2 at S21, and *Akt3* transcript abundance was specially increased in my cellular models of both Weaver and Sotos syndromes (Cha et al., 2005; Xu et al., 2012). As such, it would be interesting to explore whether the observed increase in *Akt3* abundance leads to increased phosphorylation of EZH2 at S21 to in turn promote direct interactions between EZH2 and p300 that potentially contribute to the hyperacetylation phenotype observed in cellular models of Weaver and Sotos syndrome.

6.1.4 Potential clinical relevance of findings in cellular models of disease

Weaver syndrome exhibits a highly variable clinical phenotype (Tatton-Brown et al., 2013). Accordingly, my data shows that Weaver-syndrome-associated *EZH2* mutations are functionally diverse; I confirm the pathogenic variants to be partial loss-of-function but demonstrate that the degree of functional impairment on PRC2 HMT activity can range from mild to severe. Moreover, I identify mutations that exert dominant negative effects as well as one with apparent altered substrate specificity (*EZH2* p.A677T). These findings highlight the molecular heterogeneity of Weaver syndrome which, if further delineated through careful biochemical and genotype-phenotype analyses, may reveal distinct sub-classifications of Weaver syndrome affected individuals. For example, the Weaver-syndrome-associated *EZH2* variants that were the most (Y728*) or least (D233V) severely impaired in HMT activity in my analyses were identified in individuals amongst the most or least overgrown, respectively, in the largest study of *EZH2*-mutant Weaver syndrome to date (Tatton-Brown et al., 2013).

The identification of dominant negative effects in Weaver syndrome in this thesis has important implications for the design of potential therapeutic interventions. Consider for example that standard gene replacement therapy would now be ruled out as therapeutic strategy for Weaver syndrome. Having said that, it is difficult to conceive of any gene-therapy approach likely to effectively treat a multi-systemic disorder that manifests *in utero*. Personalised oncogenomics, on the other hand, will likely prove a beneficial therapeutic avenue in the context of Weaver and indeed other overgrowth syndromes (Gamu and Gibson, 2020). To date, the prevailing assumption based off the existing evidence is that the mechanisms driving cancer in Weaver syndrome patients resemble that of malignancies in which PRC2 is functionally inactivated e.g. leukaemia or myeloid disorders (Cohen et al., 2016; Conway et al., 2015). However, the characterisation of the *EZH2* p.A677T variant in this thesis argues that subtly different mechanisms are possible which may predispose towards other cancer types, such as lymphoma or ALL (Conway et al., 2015;

Laugesen et al., 2016). My findings agree with the supposition that EZH2 inhibitors would likely be detrimental if used to treat the cancers of patients with Weaver syndrome since PRC2 enzymatic activity is already constitutively impaired (Cohen et al., 2016; Gamu and Gibson, 2020). On the other hand, Gamu and Gibson (2020) recently speculated that HAT inhibition could be beneficial in the context of Weaver syndrome – the data presented herein firmly support that hypothesis. Specifically, it will be worthwhile to test the highly selective p300/CBP inhibitor A-485 for its ability to “reverse” the molecular phenotype of Weaver syndrome described in chapter four and potentially rebalance the landscape of histone modifications at intergenic chromatin (Lasko et al., 2017; Michaelides et al., 2018). Given the H3K27 hyperacetylation phenotype observed in the *Nsd1* heterozygous knockout cell lines in chapter five, HAT inhibition should also be explored in the context of Sotos syndrome. Conversely, HDAC or EZH2 inhibitors may prove useful should neoplasms arise in the individual with the *EZH2* gain-of-function mutation (p.A733T) identified by Choufani et al. (2020).

Finally, the RNA-seq and RT-qPCR analyses performed in chapters four and five revealed key genes likely to contribute to the OGID phenotype of Weaver and Sotos syndromes; future studies should focus on elucidating their specific contributions which may in turn aid the identification of druggable targets. Notably, I showed *Akt3* to be specifically increased in abundance in isogenic cellular models of Weaver and Sotos syndrome, while *AKT3* was similarly increased in abundance in lymphoblastoid cell lines derived from Weaver syndrome cases (Fig. 4.12, 5.11 and 5.12). AKT family proteins (AKT1–3) are protein kinases which participate in the PI3K/AKT/mTOR signalling pathway which, when activated, promotes cell growth through increased cell metabolism, survival, turn over and protein synthesis (Engelman et al., 2006; Ersahin et al., 2015). Mutations in PI3K/AKT3 pathway genes were reported to cause 3.2% of OGID cases in a recent study (Tatton-Brown et al., 2017). Contrastingly, an undergrowth phenotype was recently shown to result from *NSD1*-overexpression-mediated downregulation of mTOR signalling (Quintero-Rivera et al., 2021). Collectively, these findings suggest that different aberrant growth phenotypes may converge on dysregulated PI3K/AKT/mTOR signalling. Due to the frequent pathologic activation of this pathway in cancers, several PI3K/AKT/mTOR pathway inhibitors are already FDA-approved with many more in development (Janku et al., 2018). Moreover, a PIK3CA inhibitor was successfully used to improve disease symptoms (including hemihypertrophy) in human patients with PIK3CA-related overgrowth syndrome (Venot et al., 2018). I propose that the therapeutic potential of

PI3K/AKT/mTOR pathway inhibitors should be evaluated as a possible treatment to ameliorate the overgrowth and neoplastic phenotypes of Weaver and Sotos syndromes.

6.1.5 Comments on experimental design related to chapters four and five

In chapter four, I used mouse embryonic stem cells (mESCs) as a convenient model system to investigate the effects of heterozygous *EZH2* mutations in Weaver syndrome. mESCs are very useful for fundamental studies due to their quick population doubling times, indefinite self-renewal potential and the availability of genetic tools such as conditional knockouts; I took advantage of pre-existing inducible *Ezh2* knockout mESCs to generate an *Ezh2* heterozygous knockout cell line for my studies (Højfeldt et al., 2018). The lentiviral expression of human *EZH2* variants in this cell line made it possible to isogenically model the heterozygous effects of many different Weaver-syndrome-associated mutations in a relatively short space of time. A similar approach was successfully used to investigate the molecular consequences of cancer-associated heterozygous missense mutations in *SMARCA4* (Hodges et al., 2018). Moreover, the same approach can now be readily applied to study other PRC2-related overgrowth syndromes. Indeed, as mentioned in section 6.1.3, a separate student of the Bracken lab has already followed my method to generate cellular models of Imagawa-Matsumoto syndrome. An alternative but more time-consuming experimental approach would have been to employ CRISPR/Cas9-mediated genome engineering to endogenously introduce a heterozygous mutation to a cell line of choice, or even a mouse as done by Lui et al. (2018). This CRISPR “knock-in” approach is more faithful to the disease genetics and would provide better control over the expression levels of wild-type versus mutant alleles. If I were to employ a CRISPR knock-in strategy in future, I would study pathogenic overgrowth-related mutations in the context of clinically relevant human cell types such as chondrocytes, neural stem cells or lymphoblastoid cell lines. Patient-derived human induced pluripotent stem cells (iPSCs) represent yet another attractive disease modelling approach. Though non-isogenic, iPSCs are arguably the most relevant cellular model of any disease, and they allow researchers to study the pathogenic mutation in the numerous differentiated cell types that can be derived from iPSCs.

In choosing to study cellular models of overgrowth syndromes, how does one determine whether an overgrowth phenotype is present or not? To this end, in chapter five I designed an assay to explore potential cellular overgrowth phenotypes in mESC models of Weaver syndrome – specifically, increased relative cell size or proliferation rate (Fig. 5.6). Although optimization is required to fully validate the assay, preliminary results suggest that cellular

models of Weaver syndrome show increased cell size relative to their wild-type counterparts. As such, this assay may become a useful tool for future experiments which seek to determine the precise molecular changes that drive overgrowth phenotypes in congenital growth disorders.

The mouse embryonic cell lines utilized in this research were cultured in medium containing both serum and 2i, following the same formulation as the Pasini lab (Conway et al., 2021; Lavarone et al., 2019; Scelfo et al., 2019; Tamburri et al., 2020). Mouse embryonic stem cells that are cultured under serum-free 2i conditions exhibit DNA hypomethylation that causes diffuse redistribution of PRC2 compared to mESCs cultured in medium without 2i and containing serum (McLaughlin et al., 2019; van Mierlo et al., 2019). However, to the best of my knowledge, no studies to date have examined the DNA methylome of mESCs grown in the presence of both serum and 2i; this is an important open question that needs to be addressed. Given the interplay between PRC2, NSD1 and DNMT3A (Fig. 1.7), it will be very important to test whether growing the mESC models of Weaver/Sotos syndrome in serum-only medium (lacking 2i) will alter any of the results reported in chapters four and five.

6.1.6 Future directions

In chapter three of this thesis, I was unable to detect PALI1 enrichment on chromatin using chromatin immunoprecipitation (ChIP) methods. As discussed in section 3.3, there are modifications to the ChIP protocol which could still be tested in future – namely, the use of MNase digestion rather than sonication to fragment DNA. It is also possible that the generation of new α -PALI1 antibodies will aid attempts to ChIP it; there are ongoing collaborations between the Bracken lab and industry to develop ChIP-grade α -PALI1 antibodies. In addition, it would be beneficial to explore alternative epigenomic profiling methods (e.g., CUT&RUN or DamID) for their ability to generate genome-wide PALI1 enrichment profiles. Finally, as discussed in section 6.1.2 above, live-cell imaging of HaloTagged-PALI1, for example, may help shed light on the potential reasons why PALI1 is so difficult to capture by ChIP.

In chapter three of this thesis, I also provided a preliminary characterisation of the phenotypes of *Pali1* single knockout, *Pali2* single knockout and *Pali1/2* double knockout mice. Unfortunately, the *Pali2* sKO and *Pali1*^{+/-}*Pali2*^{-/-} mouse colonies have since been closed but their embryos are cryopreserved for future use. The *Pali1*^{+/-} line is currently maintained at Trinity College Dublin and requires much further characterisation to answer

open questions regarding the physiological cause(s) of the *Pali1* homozygous knockout perinatal lethal phenotype. The future detailed study of these mice will provide unique and valuable insight into the functions of PALI1 *in vivo*. For example, as mentioned in section 3.3, efforts are currently ongoing to explore potential links between PRC2, PALI1 and G9a in the context of non-canonical imprinting.

In chapter four of this thesis, I performed comprehensive molecular characterisation of Weaver-syndrome-associated EZH2 variants. These initial studies primarily focussed on examining their consequences on PRC2-mediated H3K27 methylation and the directly opposing H3K27 acetylation modification. It would be beneficial to additionally explore H3K36 and DNA methylation in detail in the cellular models of Weaver syndrome, given the molecular crosstalk between these modifications and H3K27 methylation, and the phenotypic overlap between their associated overgrowth syndromes (Fig. 1.7; Deevy and Bracken, 2019). Furthermore, considering the reductions in global H3K27me3 and CBX4 abundance common to cellular models of Weaver syndrome, a key open question that must be addressed is whether canonical PRC1 (cPRC1) binding to chromatin is impaired in this context. A related question yet to be answered is whether chromatin compaction is altered in cellular models of Weaver syndrome. It may also be worthwhile investigating variant PRC1 localisation and H2AK119ub deposition in cellular models of Weaver syndrome, given the recent identification of vPRC1-dependent recruitment of DNMT3A to CpG islands (Weinberg et al., 2021). I propose that the same abovementioned investigations be applied to the cellular models of Sotos syndrome generated in chapter five, to add to our emerging knowledge of the molecular overlap between Weaver and Sotos syndromes.

6.2 Conclusions

The data presented in this thesis support the following conclusions:

- PALI1 is dispensable for the *de novo* targeting and recruitment of PRC2 to chromatin.
- The combined loss of *Pali1* and *Pali2* in mice phenocopies the perinatal lethal phenotype of *Pali1* single knockout mice, whereas *Pali2* single knockout mice are viable and fertile.
- Both PALI1 and PALI2 promote PRC2 histone methyltransferase activity *in vivo*.
- Weaver syndrome associated mutations in *EZH2* are predominantly partial loss-of-function with dominant negative effects on PRC2 enzymatic activity.
- Heterozygous expression of WS associated *EZH2* variants causes global H3K27 hyperacetylation.
- Heterozygous expression of Weaver-syndrome-associated *EZH2* variants induces a specific transcriptional signature in mouse embryonic stem cells.
- Heterozygous expression of the growth-restriction-associated *EZH2* p.A733T variant causes diffuse global H3K27 hyper-trimethylation with concomitant global reductions in H3K27me2 and loss of diffuse H3K27ac domains.
- Heterozygous deletion of *Nsd1* causes global H3K27 hyperacetylation.

REFERENCES

- Akasaka, T., Kanno, M., Balling, R., Mieza, M.A., Taniguchi, M., and Koseki, H. (1996). A role for mel-18, a Polycomb group-related vertebrate gene, during the anteroposterior specification of the axial skeleton. *Development* *122*, 1513–1522.
- Alcantara, D., Timms, A.E., Gripp, K., Baker, L., Park, K., Collins, S., Cheng, C., Stewart, F., Mehta, S.G., Saggat, A., et al. (2017). Mutations of AKT3 are associated with a wide spectrum of developmental disorders including extreme megalencephaly. *Brain* *140*, 2610–2622.
- Alekseyenko, A. a, Gorchakov, A. a, Kharchenko, P. V, and Kuroda, M.I. (2014). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc. Natl. Acad. Sci.* *111*, 2488–2493.
- Alkema, M.J., van der Lugt, N.M., Bobeldijk, R.C., Berns, A., and van Lohuizen, M. (1995). Transformation of axial skeleton due to overexpression of bmi-1 in transgenic mice. *Nature* *374*, 724–727.
- Allis, C.D., Caparros, M.-L., Jenuwein, T., and Reinberg, D. (2015). *Epigenetics* (Cold Spring Harbor, New York : CSH Press, Cold Spring Harbor Laboratory Press).
- Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* *19*, 229–244.
- Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., et al. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* *356*, 1081–1084.
- Amatori, S., Tavolaro, S., Gambardella, S., and Fanelli, M. (2021). The dark side of histones: genomic organization and role of oncohistones in cancer. *Clin. Epigenetics* *2021* *131* *13*, 1–21.
- Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* *9*.
- Bagert, J.D., Mitchener, M.M., Patriotis, A.L., Dul, B.E., Wojcik, F., Nacev, B.A., Feng, L., Allis, C.D., and Muir, T.W. (2021). Oncohistone mutations enhance chromatin remodeling and alter cell fates. *Nat. Chem. Biol.* *17*, 403–411.
- Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* *19*.
- Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nat.* *2001* *410*, 120–124.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* *129*, 823–837.
- Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* *143*, 470–484.

Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., et al. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* *49*, D480–D489.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* *520*, 243–247.

Baujat, G., Rio, M., Rossignol, S., Sanlaville, D., Lyonnet, S., Merrer, M. Le, Munnich, A., Gicquel, C., Colleaux, L., and Cormier-Daire, V. (2005). Clinical and molecular overlap in overgrowth syndromes. *Am. J. Med. Genet. Part C Semin. Med. Genet.* *137C*, 4–11.

Baylin, S.B., and Jones, P.A. (2016). Epigenetic Determinants of Cancer. *Cold Spring Harb. Perspect. Biol.* *8*.

Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C., et al. (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol. Cell* *66*, 384–397.e8.

Bennett, R.L., Swaroop, A., Troche, C., and Licht, J.D. (2017). The Role of Nuclear Receptor–Binding SET Domain Family Histone Lysine Methyltransferases in Cancer. *Cold Spring Harb. Perspect. Med.* *7*, a026708.

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

Bird, A.P., and Taggart, M.H. (1980). Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acids Res.* *8*, 1485–1497.

Blackledge, N.P., and Klose, R.J. (2021). The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* 1–19.

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

Blackledge, N.P., Fursova, N.A., Kelley, J.R., Huseyin, M.K., Feldmann, A., and Klose, R.J. (2020). PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol. Cell* *77*, 857–874.e9.

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

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science* (80-.). *294*, 2536–2539.

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

Bracken, A.P. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* *20*, 1123–1136.

- Bracken, A.P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., Theilgaard-Monch, K., Minucci, S., Porse, B.T., Marine, J.-C., et al. (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev* *21*, 525–530.
- Bracken, A.P., Brien, G.L., and Verrijzer, C.P. (2019). Dangerous liaisons: interplay between SWI/SNF, NuRD, and Polycomb in chromatin regulation and cancer. *Genes Dev.* *33*, 936–959.
- Brien, G.L., Gambero, G., O’Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., et al. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat Struct Mol Biol* *19*, 1273–1281.
- Brien, G.L., Healy, E., Jerman, E., Conway, E., Fadda, E., O’Donovan, D., Krivtsov, A. V, Rice, A.M., Kearney, C.J., Flaus, A., et al. (2015). A chromatin-independent role of Polycomb-like 1 to stabilize p53 and promote cellular quiescence. *Genes Dev* *29*, 2231–2243.
- Brien, G.L., Bressan, R.B., Monger, C., Gannon, D., Lagan, E., Doherty, A.M., Healy, E., Neikes, H., Fitzpatrick, D.J., Deevy, O., et al. (2021). Simultaneous disruption of PRC2 and enhancer function underlies histone H3.3-K27M oncogenic activity in human hindbrain neural stem cells. *Nat. Genet.* *53*, 1221–1232.
- Brockdorff, N., Bowness, J.S., and Wei, G. (2020). Progress toward understanding chromosome silencing by Xist RNA. *Genes Dev.* *34*, 733–744.
- Brockers, K., and Schneider, R. (2019). Histone H1, the forgotten histone. *Epigenomics* *11*, 363–366.
- Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* *84*, 843–851.
- Bult, C.J., Blake, J.A., Smith, C.L., Kadin, J.A., Richardson, J.E., Anagnostopoulos, A., Asabor, R., Baldarelli, R.M., Beal, J.S., Bello, S.M., et al. (2019). Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* *47*, D801–D806.
- Cantani, A., and Gagliosi, D. (1998). Rubinstein-Taybi syndrome. Review of 732 cases and analysis of the typical traits. *Eur. Rev. Med. Pharmacol. Sci.* *2*, 81–87.
- Casanova, M., Preissner, T., Cerase, A., Poot, R., Yamada, D., Li, X., Appanah, R., Bezstarosti, K., Demmers, J., Koseki, H., et al. (2011). Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. *Development* *138*, 1471–1482.
- Cenik, B.K., and Shilatifard, A. (2021). COMPASS and SWI/SNF complexes in development and disease. *Nat. Rev. Genet.* *22*, 38–58.
- Cha, T.L., Zhou, B.P., Xia, W., Wu, Y., Yang, C.C., Chen, C. Te, Ping, B., Otte, A.P., and Hung, M.C. (2005). Akt-mediated phosphorylation of EZH2 suppresses methylation of lysine 27 in histone H3. *Science* (80-.). *310*, 306–310.
- Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., and Tyler, J.K. (2016). The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Mol. Cell. Biol.* *36*, 662–667.

- Chen, S., Jiao, L., Shubbar, M., Yang, X., and Liu, X. (2018). Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. *Mol. Cell* *69*, 840–852.e5.
- Chen, T., Tsujimoto, N., and Li, E. (2004). The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin. *Mol. Cell Biol.* *24*, 9048–9058.
- Choi, J., Bachmann, A.L., Tauscher, K., Benda, C., Fierz, B., and Müller, J. (2017). DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. *Nat. Struct. Mol. Biol.* *24*, 1039–1047.
- Choi, J., AramYang, Song, A., Lim, M., Kim, J., Jang, J.-H., Park, K.-T., Cho, S., and Jin, D.-K. (2018). Oculodentodigital Dysplasia with a Novel Mutation in GJA1 Diagnosed by Targeted Gene Panel Sequencing: A Case Report and Literature Review. *Ann. Clin. Lab. Sci.* *48*, 776–781.
- Choufani, S., Cytrynbaum, C., Chung, B.H.Y., Turinsky, A.L., Grafodatskaya, D., Chen, Y.A., Cohen, A.S.A., Dupuis, L., Butcher, D.T., Siu, M.T., et al. (2015). NSD1 mutations generate a genome-wide DNA methylation signature. *Nat. Commun.* *6*, 10207.
- Choufani, S., Gibson, W.T., Turinsky, A.L., Chung, B.H., Wang, T., Garg, K., Vitriolo, A., Cohen, A.S., Cyrus, S., Goodman, S., et al. (2020). DNA Methylation Signature for EZH2 Functionally Classifies Sequence Variants in Three PRC2 Complex Genes. *Am. J. Hum. Genet.* *106*, 596–610.
- Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* *2017* *18*, 407–422.
- Cohen, A.S., and Gibson, W.T. (2016). EED-associated overgrowth in a second male patient. *J. Hum. Genet.* *61*, 831–834.
- Cohen, A.S.A., Tuysuz, B., Shen, Y., Bhalla, S.K., Jones, S.J.M., and Gibson, W.T. (2015). A novel mutation in EED associated with overgrowth. *J. Hum. Genet.* *60*, 339–342.
- Cohen, A.S.A.A., Yap, D.B., Lewis, M.E.S.S., Chijiwa, C., Ramos-Arroyo, M.A., Tkachenko, N., Milano, V., Fradin, M., McKinnon, M.L., Townsend, K.N., et al. (2016). Weaver Syndrome-Associated EZH2 Protein Variants Show Impaired Histone Methyltransferase Function In Vitro. *Hum. Mutat.* *37*, 301–307.
- Coleman, R.T., and Struhl, G. (2017). Causal role for inheritance of H3K27me3 in maintaining the off state of a *Drosophila* HOX gene. *Science* (80-.). *356*.
- Colon, N.C., and Chung, D.H. (2011). Neuroblastoma. *Adv. Pediatr.* *58*, 297.
- Conway, E., Healy, E., and Bracken, A.P. (2015). PRC2 mediated H3K27 methylations in cellular identity and cancer. *Curr Opin Cell Biol* *37*, 42–48.
- Conway, E., Jerman, E., Healy, E., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D.J., Mucha, M., et al. (2018). A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. *Mol. Cell* *70*, 408–421.e8.
- Conway, E., Rossi, F., Fernandez-Perez, D., Ponzo, E., Ferrari, K.J., Zanotti, M., Manganaro, D., Rodighiero, S., Tamburri, S., and Pasini, D. (2021). BAP1 enhances Polycomb repression by counteracting widespread H2AK119ub1 deposition and chromatin condensation. *Mol. Cell* *81*, 3526–3541.e8.

- Cooney, E., Bi, W., Schlesinger, A.E., Vinson, S., and Potocki, L. (2017). Novel EED mutation in patient with Weaver syndrome. *Am. J. Med. Genet. Part A* *173*, 541–545.
- Cooper, D.N., and Krawczak, M. (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum. Genet.* 1989 *83*, 181–188.
- Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., et al. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat. Commun.* *7*, 13661.
- Coulter, D., Powell, C.M., and Gold, S. (2008). Weaver syndrome and neuroblastoma. *J. Pediatr. Hematol. Oncol.* *30*, 758–760.
- Di Croce, L., and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* *20*, 1147–1155.
- Cyrus, S., Burkardt, D., Weaver, D.D., and Gibson, W.T. (2019a). PRC2-complex related dysfunction in overgrowth syndromes: A review of EZH2 , EED , and SUZ12 and their syndromic phenotypes. *Am. J. Med. Genet. Part C Semin. Med. Genet.* *ajmg.c.31754*.
- Cyrus, S.S., Cohen, A.S.A., Agbahovbe, R., Avela, K., Yeung, K.S., Chung, B.H.Y., Luk, H., Tkachenko, N., Choufani, S., Weksberg, R., et al. (2019b). Rare SUZ12 variants commonly cause an overgrowth phenotype. *Am. J. Med. Genet. Part C Semin. Med. Genet.* *ajmg.c.31748*.
- Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* *46*, D794–D801.
- Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* *25*, 1010–1022.
- Deevy, O., and Bracken, A.P. (2019). PRC2 functions in development and congenital disorders. *Development* *146*, dev181354.
- Dehal, P., and Boore, J.L. (2005). Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLoS Biol.* *3*, e314.
- Derkacheva, M., and Hennig, L. (2014). Variations on a theme: Polycomb group proteins in plants. *J. Exp. Bot.* *65*, 2769–2784.
- Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* *285*, 26114–26120.
- Dikow, N., Maas, B., Gaspar, H., Kreiss-Nachtsheim, M., Engels, H., Kuechler, A., Garbes, L., Netzer, C., Neuhaus, T.M., Koehler, U., et al. (2013). The phenotypic spectrum of duplication 5q35.2–q35.3 encompassing *NSD1*: Is it really a reversed sotos syndrome? *Am. J. Med. Genet. Part A* *161*, 2158–2166.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.
- Doskočil, J., and Šorm, F. (1962). Distribution of 5-methylcytosine in pyrimidine sequences of

deoxyribonucleic acids. *Biochim. Biophys. Acta* **55**, 953–959.

Douglas, J., Hanks, S., Temple, I.K., Davies, S., Murray, A., Upadhyaya, M., Tomkins, S., Hughes, H.E., Cole, R.P.T., and Rahman, N. (2003). NSD1 Mutations Are the Major Cause of Sotos Syndrome and Occur in Some Cases of Weaver Syndrome but Are Rare in Other Overgrowth Phenotypes. *Am. J. Hum. Genet.* **72**, 132–143.

Douglas, J., Tatton-Brown, K., Coleman, K., Guerrero, S., Berg, J., Cole, T.R., Fitzpatrick, D., Gillerot, Y., Hughes, H.E., Pilz, D., et al. (2005). Partial NSD1 deletions cause 5% of Sotos syndrome and are readily identifiable by multiplex ligation dependent probe amplification. *J. Med. Genet.* **42**, e56–e56.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.

Edinger, A.L., and Thompson, C.B. (2004). An activated mTOR mutant supports growth factor-independent, nutrient-dependent cell survival. *Oncogene* **23**, 5654–5663.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27**, 406–420.

Engelman, J.A., Luo, J., and Cantley, L.C. (2006). The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nat. Rev. Genet.* **7**, 606–619.

Erhardt, S., Su, I.-H., Schneider, R., Barton, S., Bannister, A.J., Perez-Burgos, L., Jenuwein, T., Kouzarides, T., Tarakhovsky, A., and Surani, M.A. (2003). Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. *Development* **130**, 4235–4248.

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

Ersahin, T., Tuncbag, N., and Cetin-Atalay, R. (2015). The PI3K/AKT/mTOR interactive pathway. *Mol. Biosyst.* **11**, 1946–1954.

Escobar, T.M., Loyola, A., and Reinberg, D. (2021). Parental nucleosome segregation and the inheritance of cellular identity. *Nat. Rev. Genet.* **2021** **22**, 379–392.

Ezhkova, E., Lien, W.-H., Stokes, N., Pasolli, H.A., Silva, J.M., and Fuchs, E. (2011). EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair. *Genes Dev.* **25**, 485–498.

Fang, G., Zhang, D., Yin, H., Zheng, L., Bi, X., and Yuan, L. (2014). Centlein mediates an interaction between C-Nap1 and Cep68 to maintain centrosome cohesion. *J. Cell Sci.* **127**, 1631–1639.

Fang, Y., Tang, Y., Zhang, Y., Pan, Y., Jia, J., Sun, Z., Zeng, W., Chen, J., Yuan, Y., and Fang, D. (2021). The H3K36me2 methyltransferase NSD1 modulates H3K27ac at active enhancers to safeguard gene expression. *Nucleic Acids Res.* **49**.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* **1**, e00205.

- Faust, C., Schumacher, A., Holdener, B., and Magnuson, T. (1995). The *eed* mutation disrupts anterior mesoderm production in mice. *Development* *121*, 273–285.
- Faust, C., Lawson, K.A., Schork, N.J., Thiel, B., and Magnuson, T. (1998). The Polycomb-group gene *eed* is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development* *125*, 4495–4506.
- Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 8689–8694.
- Fernandes, I., Bastien, Y., Wai, T., Nygard, K., Lin, R., Cormier, O., Lee, H.S., Eng, F., Bertos, N.R., Pelletier, N., et al. (2003). Ligand-Dependent Nuclear Receptor Corepressor LCoR Functions by Histone Deacetylase-Dependent and -Independent Mechanisms. *Mol. Cell* *11*, 139–150.
- Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* *53*, 49–62.
- Finogenova, K., Bonnet, J., Poepsel, S., Schäfer, I., Finkl, K., Schmid, K., Litz, C., Strauss, M., Benda, C., and Müller, J. (2020). Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. *Elife* *9*.
- Fischer, E.G. (2020). Nuclear Morphology and the Biology of Cancer Cells. *Acta Cytol.* *64*, 511–519.
- Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* *17*, 1870–1881.
- Fryer, A., Smith, C., Rosenbloom, L., and Cole, T. (1997). Autosomal dominant inheritance of Weaver syndrome. *J. Med. Genet.* *34*, 418–419.
- Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* *31*, 2305.
- Fursova, N.A., Blackledge, N.P., Nakayama, M., Ito, S., Koseki, Y., Farcas, A.M., King, H.W., Koseki, H., and Klose, R.J. (2019). Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol. Cell* *74*, 1020–1036.e8.
- Gahan, J.M., Rentzsch, F., and Schnitzler, C.E. (2020). The genetic basis for PRC1 complex diversity emerged early in animal evolution. *Proc. Natl. Acad. Sci. U. S. A.* *117*, 22880–22889.
- Gamu, D., and Gibson, W.T. (2020). Reciprocal skeletal phenotypes of PRC2-related overgrowth and Rubinstein–Taybi syndromes: potential role of H3K27 modifications. *Cold Spring Harb. Mol. Case Stud.* *6*, a005058.
- Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol Cell* *45*, 344–356.
- Gaydos, L.J., Wang, W., and Strome, S. (2014). H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science* (80-.). *345*, 1515–1518.

- Ge, Y.Z., Pu, M.T., Gowher, H., Wu, H.P., Ding, J.P., Jeltsch, A., and Xu, G.L. (2004). Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J. Biol. Chem.* *279*, 25447–25454.
- Gibson, W.T., Hood, R.L., Zhan, S.H., Bulman, D.E., Fejes, A.P., Moore, R., Mungall, A.J., Eyedoux, P., Babul-Hirji, R., An, J., et al. (2012). Mutations in EZH2 cause weaver syndrome. *Am. J. Hum. Genet.* *90*, 110–118.
- Glancy, E., Ciferri, C., and Bracken, A.P. (2021). Structural basis for PRC2 engagement with chromatin. *Curr. Opin. Struct. Biol.* *67*, 135–144.
- Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* *20*, 590–607.
- Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M.R., Szumska, D., McGouran, J.F., Bhattacharya, S., Kessler, B.M., Brockdorff, N., and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. *Development* *143*, 2716–2723.
- GroB, A., Schulz, C., Kolb, J., Koster, J., Wehner, S., Czaplinski, S., Khilan, A., Rohrer, H., Harter, P.N., Klingebiel, T., et al. (2018). Tumorigenic and antiproliferative properties of the TALE-transcription factors MEIS2D and MEIS2A in neuroblastoma. *Cancer Res.* *78*, 1935–1947.
- Gu, T., Lin, X., Cullen, S.M., Luo, M., Jeong, M., Estecio, M., Shen, J., Hardikar, S., Sun, D., Su, J., et al. (2018). DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol.* *19*, 88.
- Gurdon, J.B., Laskey, R.A., and Reeves, O.R. (1975). The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs. *J. Embryol. Exp. Morphol.* *34*, 93–112.
- Hagarman, J.A., Motley, M.P., Kristjansdottir, K., and Soloway, P.D. (2013). Coordinate Regulation of DNA Methylation and H3K27me3 in Mouse Embryonic Stem Cells. *PLoS One* *8*, e53880.
- Hanna, C.W., and Kelsey, G. (2021). Features and mechanisms of canonical and noncanonical genomic imprinting. *Genes Dev.* *35*, 821–834.
- Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep.* *17*, 583–595.
- Healy, E., Mucha, M., Glancy, E., Fitzpatrick, D.J., Conway, E., Neikes, H.K., Monger, C., Van Mierlo, G., Baltissen, M.P., Koseki, Y., et al. (2019). PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol. Cell* *76*, 437–452.e6.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* *38*, 576–589.
- Hentze, M.W., and Kulozik, A.E. (1999). A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* *96*, 307–310.
- Herz, H.-M., Mohan, M., Garrett, A.S., Miller, C., Casto, D., Zhang, Y., Seidel, C., Haug, J.S., Florens, L., Washburn, M.P., et al. (2012). Polycomb repressive complex 2-dependent and -independent functions

of Jarid2 in transcriptional regulation in *Drosophila*. *Mol. Cell. Biol.* **32**, 1683–1693.

Hetzelt, K.L.M.L., Winterholler, M., Kerling, F., Rauch, C., Ekici, A.B., Winterpacht, A., Vasileiou, G., Uebe, S., Thiel, C.T., Kraus, C., et al. (2021). Manifestation of epilepsy in a patient with EED -related overgrowth (Cohen–Gibson syndrome). *Am. J. Med. Genet. Part A*.

Heyn, P., Logan, C. V., Fluteau, A., Challis, R.C., Auchynnikava, T., Martin, C.-A., Marsh, J.A., Taglini, F., Kilanowski, F., Parry, D.A., et al. (2019). Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nat. Genet.* **51**, 96–105.

Hodges, H.C., Stanton, B.Z., Cermakova, K., Chang, C.Y., Miller, E.L., Kirkland, J.G., Ku, W.L., Veverka, V., Zhao, K., and Crabtree, G.R. (2018). Dominant-negative SMARCA4 mutants alter the accessibility landscape of tissue-unrestricted enhancers. *Nat. Struct. Mol. Biol.* **25**, 61–72.

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

Højfeldt, J.W., Hedehus, L., Laugesen, A., Tatar, T., Wiehle, L., and Helin, K. (2019). Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. *Mol. Cell* **76**, 1–14.

Holliday, R., and Grigg, G.W. (1993). DNA methylation and mutation. *Mutat. Res. Mol. Mech. Mutagen.* **285**, 61–67.

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

Hu, M., Sun, X.-J., Zhang, Y.-L., Kuang, Y., Hu, C.-Q., Wu, W.-L., Shen, S.-H., Du, T.-T., Li, H., He, F., et al. (2010). Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proc. Natl. Acad. Sci.* **107**, 2956–2961.

Hübner, J.-M., Müller, T., Papageorgiou, D.N., Mauermann, M., Krijgsveld, J., Russell, R.B., Ellison, D.W., Pfister, S.M., Pajtlér, K.W., and Kool, M. (2019). EZHIP/CXorf67 mimics K27M mutated oncohistones and functions as an intrinsic inhibitor of PRC2 function in aggressive posterior fossa ependymoma. *Neuro. Oncol.* **21**, 878–889.

Hughes, A.L., Kelley, J.R., and Klose, R.J. (2020). Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. *Biochim. Biophys. Acta – Gene Regul. Mech.* **1863**, 194567.

Huseyin, M.K., and Klose, R.J. (2021). Live-cell single particle tracking of PRC1 reveals a highly dynamic system with low target site occupancy. *Nat. Commun.* **12**, 1–20.

Imagawa, E., Higashimoto, K., Sakai, Y., Numakura, C., Okamoto, N., Matsunaga, S., Ryo, A., Sato, Y., Sanefuji, M., Ihara, K., et al. (2017). Mutations in genes encoding polycomb repressive complex 2 subunits cause Weaver syndrome. *Hum. Mutat.* **38**, 637–648.

Imagawa, E., Albuquerque, E.V.A., Isidor, B., Mitsunashi, S., Mizuguchi, T., Miyatake, S., Takata, A., Miyake, N., Boguszewski, M.C.S., Boguszewski, C.L., et al. (2018). Novel SUZ12 mutations in Weaver-like syndrome. *Clin. Genet.* **94**, 461–466.

Ingham, P.W. (1983). Differential expression of bithorax complex genes in the absence of the extra sex

combs and trithorax genes. *Nature* 306, 591–593.

Ingham, P.W. (1985). Genetic Control of the Spatial Pattern of Selector Gene Expression in *Drosophila*. *Cold Spring Harb. Symp. Quant. Biol.* 50, 201–208.

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

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., et al. (2019). PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism. *Nat. Commun.* 10, 2146.

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., et al. (2020). H3 K27M and EZHIP Impede H3K27-Methylation Spreading by Inhibiting Allosterically Stimulated PRC2. *Mol. Cell* 80, 726–735.e7.

Jani, K.S., Jain, S.U., Ge, E.J., Diehl, K.L., Lundgren, S.M., Müller, M.M., Lewis, P.W., and Muir, T.W. (2019). Histone H3 tail binds a unique sensing pocket in EZH2 to activate the PRC2 methyltransferase. *Proc. Natl. Acad. Sci.* 116, 8295–8300.

Janku, F., Yap, T.A., and Meric-Bernstam, F. (2018). Targeting the PI3K pathway in cancer: are we making headway? *Nat. Rev. Clin. Oncol.* 2018 155 15, 273–291.

Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* (80-.). 293, 1074–1080.

Jiao, L., Shubbar, M., Yang, X., Zhang, Q., Chen, S., Wu, Q., Chen, Z., Rizo, J., and Liu, X. (2020). A partially disordered region connects gene repression and activation functions of EZH2. *Proc. Natl. Acad. Sci. U. S. A.* 117, 16992–17002.

Jing, Z., Yin, H., Wang, P., Gao, J., and Yuan, L. (2016). Centlein, a novel microtubule-associated protein stabilizing microtubules and involved in neurite formation. *Biochem. Biophys. Res. Commun.* 472, 360–365.

Kalkhoven, E., Roelfsema, J.H., Teunissen, H., den Boer, A., Ariyurek, Y., Zantema, A., Breuning, M.H., Hennekam, R.C.M., and Peters, D.J.M. (2003). Loss of CBP acetyltransferase activity by PHD finger mutations in Rubinstein–Taybi syndrome. *Hum. Mol. Genet.* 12, 441–450.

Kasinath, V., Beck, C., Sauer, P., Poepsel, S., Kosmatka, J., Faini, M., Toso, D., Aebersold, R., and Nogales, E. (2021). JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. *Science* (80-.). 371, eabc3393.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* 2019 101 10, 1–10.

Kaya-Okur, H.S., Janssens, D.H., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2020). Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.* 2020 1510 15, 3264–3283.

Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia

- mutations in *Drosophila*. *Proc. Natl. Acad. Sci.* **85**, 8136–8140.
- Kim, H., Kang, K., Ekram, M.B., Roh, T.Y., and Kim, J. (2011). Aebp2 as an epigenetic regulator for neural crest cells. *PLoS One* **6**, e25174.
- King, A.D., Huang, K., Rubbi, L., Wang, Y., Pellegrini, M., and Correspondence, G.F. (2016). Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. *CellReports* **17**, 289–302.
- Kleff, S., Andrulis, E.D., Anderson, C.W., and Sternglanz, R. (1995). Identification of a gene encoding a yeast histone H4 acetyltransferase. *J. Biol. Chem.* **270**, 24674–24677.
- Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nat.* **502**, 472–479.
- Kondo, T., Isono, K., Kondo, K., Endo, T.A., Itohara, S., Vidal, M., and Koseki, H. (2014). Polycomb Potentiates Meis2 Activation in Midbrain by Mediating Interaction of the Promoter with a Tissue-Specific Enhancer. *Dev. Cell* **28**, 94–101.
- Kosaki, R., Terashima, H., Kubota, M., and Kosaki, K. (2017). Acute myeloid leukemia-associated *DNMT3A* p.Arg882His mutation in a patient with Tatton-Brown-Rahman overgrowth syndrome as a constitutional mutation. *Am. J. Med. Genet. Part A* **173**, 250–253.
- Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *PLoS Genet.* **4**, e1000242.
- Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol. Cell* **65**, 432–446.e5.
- Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T., Ohashi, H., Naritomi, K., Tsukahara, M., Makita, Y., et al. (2002). Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* **30**, 365–366.
- Lachner, M., O’Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nat.* **410**, 116–120.
- Lancaster, M.A., Renner, M., Martin, C.-A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. *Nat.* **501**, 373–379.
- Landeira, D., and Fisher, A.G. (2011). Inactive yet indispensable: the tale of Jarid2. *Trends Cell Biol.* **21**, 74–80.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.
- Laprell, F., Finkl, K., and Müller, J. (2017). Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. *Science* (80-.). **356**, 85–88.
- Lasko, L.M., Jakob, C.G., Edalji, R.P., Qiu, W., Montgomery, D., Digiammarino, E.L., Hansen, T.M., Risi, R.M., Frey, R., Manaves, V., et al. (2017). Discovery of a selective catalytic p300/CBP inhibitor that

- targets lineage-specific tumours. *Nat.* 2017 5507674 *550*, 128–132.
- Laugesen, A., Højfeldt, J.W., and Helin, K. (2016). Role of the Polycomb Repressive Complex 2 (PRC2) in Transcriptional Regulation and Cancer. *Cold Spring Harb. Perspect. Med.* *6*.
- Laugesen, A., Højfeldt, J.W., and Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell* *74*, 8–18.
- Lavarone, E., Barbieri, C.M., and Pasini, D. (2019). Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. *Nat. Commun.* *10*.
- Lee, H.-G., Kahn, T.G., Simcox, A., Schwartz, Y.B., and Pirrotta, V. (2015). Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome Res* *25*, 1170–1181.
- Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* *125*, 301–313.
- Lee, Y., Song, A.J., Baker, R., Micales, B., Conway, S.J., and Lyons, G.E. (2000). Jumonji, a Nuclear Protein That Is Necessary for Normal Heart Development. *Circ. Res.* *86*, 932–938.
- Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* *276*, 565–570.
- Lewis, P.H. (1947). New Mutants Report. *Drosoph. Inf. Serv.* *21*, 69.
- Li, E., and Zhang, Y. (2014). DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.* *6*, 19133–19134.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
- Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., et al. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* *549*, 287–291.
- Li, X., Isono, K., Yamada, D., Endo, T.A., Endoh, M., Shinga, J., Mizutani-Koseki, Y., Otte, A.P., Casanova, M., Kitamura, H., et al. (2014). Mammalian Polycomb-Like Pcl2/Mtf2 Is a Novel Regulatory Component of PRC2 That Can Differentially Modulate Polycomb Activity both at the Hox Gene Cluster and at Cdkn2a Genes. *Mol. Cell. Biol.* *34*, 2773–2773.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* *30*, 923–930.
- Liefke, R., Karwacki-Neisius, V., and Shi, Y. (2016). EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape. *Mol. Cell* *64*, 659–672.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 1–21.
- Luger, K., Mäder, a W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* *389*, 251–260.
- Lui, J.C., Barnes, K.M., Dong, L., Yue, S., Graber, E., Rapaport, R., Dauber, A., Nilsson, O., and Baron, J. (2018). Ezh2 Mutations Found in the Weaver Overgrowth Syndrome Cause a Partial Loss of H3K27 Histone Methyltransferase Activity. *J. Clin. Endocrinol. Metab.* *103*, 1470–1478.

- Luscan, A., Laurendeau, I., Malan, V., Francannet, C., Odent, S., Giuliano, F., Lacombe, D., Touraine, R., Vidaud, M., Pasmant, E., et al. (2014). Mutations in *SETD2* cause a novel overgrowth condition. *J. Med. Genet.* *51*, 512–517.
- Lynch, M.D., Smith, A.J., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., et al. (2011). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J.* *31*, 317–329.
- Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* *47*, W636–W641.
- Makino, K., Umeda, K., Uezu, A., Hiragami, Y., Sakamoto, T., Ihn, H., and Nakanishi, H. (2008). Identification and characterization of the novel centrosomal protein centlein. *Biochem. Biophys. Res. Commun.* *366*, 958–962.
- Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* *469*, 343–349.
- 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., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* *461*, 762–767.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* *17*, 10–12.
- Martin, B.J.E., Brind'Amour, J., Kuzmin, A., Jensen, K.N., Liu, Z.C., Lorincz, M., and Howe, L.J. (2021). Transcription shapes genome-wide histone acetylation patterns. *Nat. Commun.* *2021* *12*, 1–9.
- McCabe, M.T., Graves, A.P., Ganji, G., Diaz, E., Halsey, W.S., Jiang, Y., Smitheman, K.N., Ott, H.M., Pappalardi, M.B., Allen, K.E., et al. (2012). Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc. Natl. Acad. Sci. U. S. A.* *109*, 2989–2994.
- McClive, P., Pall, G., Newton, K., Lee, M., Mullins, J., and Forrester, L. (1998). Gene trap integrations expressed in the developing heart: Insertion site affects splicing of the PT1-ATG vector. *Dev. Dyn.* *212*, 267–276.
- McLaughlin, K., Flyamer, I.M., Thomson, J.P., Mjoseng, H.K., Shukla, R., Williamson, I., Grimes, G.R., Illingworth, R.S., Adams, I.R., Pennings, S., et al. (2019). DNA Methylation Directs Polycomb-Dependent 3D Genome Re-organization in Naive Pluripotency. *Cell Rep.* *29*, 1974–1985.e6.
- Meers, M.P., Bryson, T.D., Henikoff, J.G., and Henikoff, S. (2019). Improved CUT&RUN chromatin profiling tools. *Elife* *8*.
- Michaelides, M.R., Kluge, A., Patane, M., Drie, J.H. Van, Wang, C., Hansen, T.M., Risi, R.M., Mantei, R., Hertel, C., Karukurichi, K., et al. (2018). Discovery of Spiro Oxazolidinediones as Selective, Orally Bioavailable Inhibitors of p300/CBP Histone Acetyltransferases. *ACS Med. Chem. Lett.* *9*, 28–33.
- van Mierlo, G., Dirks, R.A.M., De Clerck, L., Brinkman, A.B., Huth, M., Kloet, S.L., Saksouk, N., Kroeze, L.I., Willems, S., Farlik, M., et al. (2019). Integrative Proteomic Profiling Reveals PRC2-

Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency. *Cell Stem Cell* **24**, 123–137.e8.

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

Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005). The Murine Polycomb Group Protein Eed Is Required for Global Histone H3 Lysine-27 Methylation. *Curr. Biol.* **15**, 942–947.

Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet.* **7**, 45–49.

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

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

Müller, M.M., Fierz, B., Bittova, L., Liszczak, G., and Muir, T.W. (2016). A two-state activation mechanism controls the histone methyltransferase Suv39h1. *Nat. Chem. Biol.* **12**, 188–193.

Murata, T., Kurokawa, R., Krones, A., Tatsumi, K., Ishii, M., Taki, T., Masuno, M., Ohashi, H., Yanagisawa, M., Rosenfeld, M.G., et al. (2001). Defect of histone acetyltransferase activity of the nuclear transcriptional coactivator CBP in Rubinstein–Taybi syndrome. *Hum. Mol. Genet.* **10**, 1071–1076.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.* **26**, 4078–4088.

O’Carroll, D., Erhardt, S., Pagani, M., and Barton, S.C. (2001). The Polycomb -Group Gene Ezh2 Is Required for Early Mouse Development. *Mol. Cell. Biol.* **21**, 4330–4336.

Ohkubo, N., Matsubara, E., Yamanouchi, J., Akazawa, R., Aoto, M., Suzuki, Y., Sakai, I., Abe, T., Kiyonari, H., Matsuda, S., et al. (2014). Abnormal Behaviors and Developmental Disorder of Hippocampus in Zinc Finger Protein 521 (ZFP521) Mutant Mice. *PLoS One* **9**, e92848.

Okamoto, N., Toribe, Y., Shimojima, K., and Yamamoto, T. (2016). Tatton-Brown-Rahman syndrome due to 2p23 microdeletion. *Am. J. Med. Genet. Part A* **170**, 1339–1342.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99**, 247–257.

Oksuz, O., Narendra, V., Lee, C.H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A., et al. (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol. Cell* **70**, 1149–1162.e5.

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

Olson, E.N., Arnold, H.H., Rigby, P.W., and Wold, B.J. (1996). Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene MRF4. *Cell* **85**, 1–4.

- Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* *448*, 714–717.
- Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* *9*, 1163–1170.
- Pasini, D., Bracken, A.P., Jensen, M.R., Denchi, E.L., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* *23*, 4061–4071.
- Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N., et al. (2010). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.* *38*, 4958–4969.
- Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* *6*, 893–904.
- Pengelly, A.R., Copur, Ö., Jäckle, H., Herzig, A., and Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb. *Science* *339*, 698–699.
- Perino, M., van Mierlo, G., Karemaker, I.D., van Genesen, S., Vermeulen, M., Marks, H., van Heeringen, S.J., and Veenstra, G.J.C. (2018). MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat. Genet.* *50*, 1002–1010.
- Petracovici, A., and Bonasio, R. (2021). Distinct PRC2 subunits regulate maintenance and establishment of Polycomb repression during differentiation. *Mol. Cell* *81*, 2625–2639.e5.
- Pierce, S.B., Stewart, M.D., Gulsuner, S., Walsh, T., Dhall, A., McClellan, J.M., Klevit, R.E., and King, M.-C. (2018). De novo mutation in RING1 with epigenetic effects on neurodevelopment. *Proc. Natl. Acad. Sci.* *115*, 1558–1563.
- Piovesan, A., Pelleri, M.C., Antonaros, F., Strippoli, P., Caracausi, M., and Vitale, L. (2019). On the length, weight and GC content of the human genome. *BMC Res. Notes* *2019* *12*, 1–7.
- Piunti, A., and Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* (80-.). *352*, aad9780.
- Piunti, A., and Shilatifard, A. (2021). The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat. Rev. Mol. Cell Biol.* 1–20.
- Piunti, A., Smith, E.R., Morgan, M.A.J., Ugarenko, M., Khaltayan, N., Helmin, K.A., Ryan, C.A., Murray, D.C., Rickels, R.A., Yilmaz, B.D., et al. (2019). CATAcomb: An endogenous inducible gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. *Sci. Adv.* *5*, eaax2887.
- Pizzuti, A., Flex, E., Mingarelli, R., Salpietro, C., Zelante, L., and Dallapiccola, B. (2004). A homozygous GJA1 gene mutation causes a Hallermann-Streiff/ODDD spectrum phenotype. *Hum. Mutat.* *23*, 286.
- Poduri, A., Evrony, G.D., Cai, X., Elhosary, P.C., Beroukhim, R., Lehtinen, M.K., Hills, L.B., Heinzen, E.L., Hill, A., Hill, R.S., et al. (2012). Somatic Activation of AKT3 Causes Hemispheric Developmental

Brain Malformations. *Neuron* 74, 41–48.

Poepsel, S., Kasinath, V., and Nogales, E. (2018). Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat. Struct. Mol. Biol.* 25, 154–162.

Polonis, K., Blackburn, P.R., Urrutia, R.A., Lomberg, G.A., Kruisselbrink, T., Cousin, M.A., Boczek, N.J., Hoppman, N.L., Babovic-Vuksanovic, D., Klee, E.W., et al. (2018). Co-occurrence of a maternally inherited DNMT3A duplication and a paternally inherited pathogenic variant in EZH2 in a child with growth retardation and severe short stature: atypical Weaver syndrome or evidence of a DNMT3A dosage effect? *Mol. Case Stud.* 4, a002899.

Prendergast, L., and Reinberg, D. (2021). The missing linker: emerging trends for H1 variant-specific functions. *Genes Dev.* 35, 40–58.

Proud, V.K., Braddock, S.R., Cook, L., and Weaver, D.D. (1998). Weaver syndrome: Autosomal dominant inheritance of the disorder. In *American Journal of Medical Genetics*, pp. 305–310.

Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., and Xu, R.-M. (2011). The Structure of NSD1 Reveals an Autoregulatory Mechanism Underlying Histone H3K36 Methylation *. *J. Biol. Chem.* 286, 8361–8368.

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

Quintero-Rivera, F., Eno, C.C., Sutanto, C., Jones, K.L., Nowaczyk, M.J.M., Wong, D., Earl, D., Mirzaa, G., Beck, A., and Martinez-Agosto, J.A. (2021). 5q35 duplication presents with psychiatric and undergrowth phenotypes mediated by NSD1 overexpression and mTOR signaling downregulation. *Hum. Genet.* 1–10.

Raas, M.W.D., Zijlmans, D.W., Vermeulen, M., and Marks, H. (2021). There is another: H3K27me3-mediated genomic imprinting. *Trends Genet.*

Ragazzini, R., Pérez-Palacios, R., Baymaz, H., Diop, S., Ancelin, K., Zielinski, D., Michaud, A., Givélet, M., Borsos, M., Aflaki, S., et al. (2019). EZHIP constrains Polycomb Repressive Complex 2 activity in germ cells. *BioRxiv* doi: 10.1101/619080.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165.

Rayasam, G.V., Wendling, O., Angrand, P., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G.L., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* 22, 3153–3163.

Rea, S., Eisenhaber, F., O’Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nat.* 2000 4066796 406, 593–599.

Reddington, J.P., Perricone, S.M., Nestor, C.E., Reichmann, J., Youngson, N.A., Suzuki, M., Reinhardt, D., Dunican, D.S., Prendergast, J.G., Mjoseng, H., et al. (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* 14, R25.

- Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Mol. Cell* 55, 347–360.
- Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet* 38, 413–443.
- Rio, M., Clech, L., Amiel, J., Faivre, L., Lyonnet, S., Merrer, M. Le, Odent, S., Lacombe, D., Edery, P., Brauner, R., et al. (2003). Spectrum of NSD1 mutations in Sotos and Weaver syndromes. *J. Med. Genet.* 40, 436–440.
- da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol. Cell* 53, 301–316.
- Roelfsema, J.H., White, S.J., Ariyürek, Y., Bartholdi, D., Niedrist, D., Papadia, F., Bacino, C.A., Den Dunnen, J.T., Van Ommen, G.J.B., Breuning, M.H., et al. (2005). Genetic Heterogeneity in Rubinstein-Taybi Syndrome: Mutations in Both the CBP and EP300 Genes Cause Disease. *Am. J. Hum. Genet.* 76, 572–580.
- Rose, N.R., and Klose, R.J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1839, 1362–1372.
- Rosenfeld, J.A., Kim, K.H., Angle, B., Troxell, R., Gorski, J.L., Westemeyer, M., Frydman, M., Senturias, Y., Earl, D., Torchia, B., et al. (2013). Further Evidence of Contrasting Phenotypes Caused by Reciprocal Deletions and Duplications: Duplication of NSD1 Causes Growth Retardation and Microcephaly. *Mol. Syndromol.* 3, 247–254.
- Rothberg, J.L.M., Maganti, H.B., Jrade, H., Porter, C.J., Palidwor, G.A., Cafariello, C., Battaion, H.L., Khan, S.T., Perkins, T.J., Paulson, R.F., et al. (2018). Mtf2-PRC2 control of canonical Wnt signaling is required for definitive erythropoiesis. *Cell Discov.* 4, 21.
- Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S.T., Offer, J., et al. (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol. Cell* 57, 769–783.
- Sayers, E.W., Beck, J., Bolton, E.E., Bourexis, D., Brister, J.R., Canese, K., Comeau, D.C., Funk, K., Kim, S., Klimke, W., et al. (2021). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 49, D10–D17.
- Scelfo, A., Fernández-Pérez, D., Tamburri, S., Zanotti, M., Lavarone, E., Soldi, M., Bonaldi, T., Ferrari, K.J., and Pasini, D. (2019). Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent- and RING1A/B-Independent-Specific Activities. *Mol. Cell* 74, 1037–1052.e7.
- Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol. Cell* 42, 330–341.
- Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* 25, 3110–3122.

- Schuettengruber, B., Bourbon, H.M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* 171, 34–57.
- Schulte, D., and Geerts, D. (2019). MEIS transcription factors in development and disease. *Dev.* 146.
- Schumacher, A., Faust, C., and Magnuson, T. (1996). Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature* 383, 250–253.
- Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* 2006 386 38, 700–705.
- Sekulić, A., Hudson, C.C., Homme, J.L., Yin, P., Otterness, D.M., Karnitz, L.M., and Abraham, R.T. (2000). A Direct Linkage between the Phosphoinositide 3-Kinase-AKT Signaling Pathway and the Mammalian Target of Rapamycin in Mitogen-stimulated and Transformed Cells. *Cancer Res.* 60, 3504 LP – 3513.
- Sendzikaitė, G., Hanna, C.W., Stewart-Morgan, K.R., Ivanova, E., and Kelsey, G. (2019). A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. *Nat. Commun.* 10, 1884.
- Shalom-Barak, T., Lierse, J., Memari, B., Flechner, L., Devor, C.E., Bernardo, T.M., Kim, S., Matsumoto, N., Friedman, S.L., Evans, R.M., et al. (2018). Ligand-Dependent Corepressor (LCoR) Is a Rexinoid-Inhibited Peroxisome Proliferator-Activated Receptor γ -Retinoid X Receptor α Coactivator. *Mol. Cell. Biol.* 38, e00107-17.
- Sharif, J., Muto, M., Takebayashi, S.I., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908–912.
- Shen, W., Heeley, J.M., Carlston, C.M., Acuna-Hidalgo, R., Nillesen, W.M., Dent, K.M., Douglas, G. V., Levine, K.L., Bayrak-Toydemir, P., Marcelis, C.L., et al. (2017). The spectrum of *DNMT3A* variants in Tatton-Brown-Rahman syndrome overlaps with that in hematologic malignancies. *Am. J. Med. Genet. Part A* 173, 3022–3028.
- Shen, X., Liu, Y., Hsu, Y.J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.C., and 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, 491–502.
- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstone, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* 119, 941–953.
- Shirane, K., Miura, F., Ito, T., and Lorincz, M.C. (2020). NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. *Nat. Genet.* 1–11.
- Skene, P.J., Henikoff, J.G., and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* 2018 135 13, 1006–1019.
- Slifer, E.H. (1942). A mutant stock of *Drosophila* with extra sex-combs. *J. Exp. Zool.* 90, 31–40.
- Smallwood, A., Estève, P.-O., Pradhan, S., and Carey, M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* 21, 1169.

- Smigiel, R., Biernacka, A., Biela, M., Murcia-Pienkowski, V., Szmida, E., Gasperowicz, P., Kosinska, J., Kostrzewa, G., Koppolu, A.A., Walczak, A., et al. (2018). Novel de novo mutation affecting two adjacent aminoacids in the EED gene in a patient with Weaver syndrome. *J. Hum. Genet.* *63*, 517–520.
- Smits, A.H., Jansen, P.W.T.C., Poser, I., Hyman, A.A., and Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res.* *41*, 1–8.
- Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci USA* *107*, 20980–20985.
- Soshnev, A.A., Josefowicz, S.Z., and Allis, C.D. (2016). Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome. *Mol. Cell* *62*, 681–694.
- Spellicy, C.J., Peng, Y., Olewiler, L., Cathey, S.S., Rogers, R.C., Bartholomew, D., Johnson, J., Alexov, E., Lee, J.A., Friez, M.J., et al. (2019). Three additional patients with EED-associated overgrowth: potential mutation hotspots identified? *J. Hum. Genet.* *64*, 561–572.
- Stanford, W.L., Cohn, J.B., and Cordes, S.P. (2001). Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.* *2*, 756–768.
- Steensel, B. van, and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* *2000* *18*, 424–428.
- Stewart-Morgan, K.R., Petryk, N., and Groth, A. (2020). Chromatin replication and epigenetic cell memory. *Nat. Cell Biol.* *22*, 361–371.
- Streubel, G., Watson, A., Jammula, S.G., Scelfo, A., Fitzpatrick, D.J., Oliviero, G., McCole, R., Conway, E., Glancy, E., Negri, G.L., et al. (2018). The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol. Cell* *70*, 371–379.e5.
- Struhl, G., and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of *Drosophila*. *EMBO J.* *4*, 3259–3264.
- Su, I.H., Basavaraj, A., Krutchinsky, A.N., Hobert, O., Ullrich, A., Chait, B.T., and Tarakhovsky, A. (2003). Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat. Immunol.* *4*, 124–131.
- Sugishita, H., Kondo, T., Ito, S., Nakayama, M., Yakushiji-Kaminatsui, N., Kawakami, E., Koseki, Y., Ohinata, Y., Sharif, J., Harachi, M., et al. (2021). Variant PCGF1-PRC1 links PRC2 recruitment with differentiation-associated transcriptional inactivation at target genes. *Nat. Commun.* *2021* *12*, 1–12.
- Suri, T., and Dixit, A. (2017). The phenotype of EZH2 haploinsufficiency—1.2-Mb deletion at 7q36.1 in a child with tall stature and intellectual disability. *Am. J. Med. Genet. Part A* *173*, 2731–2735.
- Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., and Higashinakagawa, T. (1995). Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev.* *9*, 1211–1222.
- Takeuchi, T., Kojima, M., Nakajima, K., and Kondo, S. (1999). jumonji gene is essential for the

neurulation and cardiac development of mouse embryos with a C3H/He background. *Mech. Dev.* 86, 29–38.

Tamburri, S., Lavarone, E., Fernández-Pérez, D., Conway, E., Zanotti, M., Manganaro, D., and Pasini, D. (2020). Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Mol. Cell* 77, 840–856.e5.

Tatton-Brown, K., and Rahman, N. (2006). Sotos syndrome. *Eur. J. Hum. Genet.* 2007 153 15, 264–271.

Tatton-Brown, K., and Rahman, N. (2013). The NSD1 and EZH2 overgrowth genes, similarities and differences. *Am. J. Med. Genet. Part C Semin. Med. Genet.* 163, 86–91.

Tatton-Brown, K., Douglas, J., Coleman, K., Baujat, G., Cole, T.R.P., Das, S., Horn, D., Hughes, H.E., Temple, I.K., Faravelli, F., et al. (2005). Genotype-Phenotype Associations in Sotos Syndrome: An Analysis of 266 Individuals with NSD1 Aberrations. *Am. J. Hum. Genet.* 77, 193–204.

Tatton-Brown, K., Hanks, S., Ruark, E., Zachariou, A., Duarte, S.D.V., Ramsay, E., Snape, K., Murray, A., Perdeaux, E.R., Seal, S., et al. (2011). Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height. *Oncotarget* 2, 1127–1133.

Tatton-Brown, K., Murray, A., Hanks, S., Douglas, J., Armstrong, R., Banka, S., Bird, L.M., Clericuzio, C.L., Cormier-Daire, V., Cushing, T., et al. (2013). Weaver syndrome and EZH2 mutations: Clarifying the clinical phenotype. *Am. J. Med. Genet. Part A* 161, 2972–2980.

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., et al. (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat. Genet.* 46, 657–657.

Tatton-Brown, K., Loveday, C., Yost, S., Clarke, M., Ramsay, E., Zachariou, A., Elliott, A., Wylie, H., Ardisson, A., Rittinger, O., et al. (2017). Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability. *Am. J. Hum. Genet.* 100, 725–736.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* (80-.). 272, 408–411.

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* 136, 3131–3141.

Tlemsani, C., Luscan, A., Leulliot, N., Bieth, E., Afenjar, A., Baujat, G., Doco-Fenzy, M., Goldenberg, A., Lacombe, D., Lambert, L., et al. (2016). SETD2 and DNMT3A screen in the Sotos-like syndrome French cohort. *J. Med. Genet.* 53, 743–751.

Tolhuis, B., Muijters, I., de Wit, E., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. *Nat. Genet.* 2006 386 38, 694–699.

Venot, Q., Blanc, T., Rabia, S.H., Berteloot, L., Ladraa, S., Duong, J.P., Blanc, E., Johnson, S.C., Hoguin, C., Boccard, O., et al. (2018). Targeted therapy in patients with PIK3CA-related overgrowth syndrome. *Nature* 558, 540–546.

Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein–DNA interactions

- using DamID in mammalian cells. *Nat. Protoc.* 2007 26 2, 1467–1478.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer Genome Landscapes. *Science* (80-.). 339, 1546–1558.
- Vogler, C., Huber, C., Waldmann, T., Ettig, R., Braun, L., Izzo, A., Daujat, S., Chassignet, I., Lopez-Contreras, A.J., Fernandez-Capetillo, O., et al. (2010). Histone H2A C-Terminus Regulates Chromatin Dynamics, Remodeling, and Histone H1 Binding. *PLOS Genet.* 6, e1001234.
- Waddington, C.H. (1942). Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565.
- Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* 13, 115–126.
- Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004a). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873–878.
- Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004b). Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* 14, 637–646.
- Wang, S., He, F., Xiong, W., Gu, S., Liu, H., Zhang, T., Yu, X., and Chen, Y. (2007). Polycomblike-2-deficient mice exhibit normal left–right asymmetry. *Dev. Dyn.* 236, 853–861.
- Wang, X., Paucek, R.D., Gooding, A.R., Brown, Z.Z., Ge, E.J., Muir, T.W., and Cech, T.R. (2017). Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat. Struct. Mol. Biol.* 24, 1028–1038.
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2–A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.
- Weaver, D.D., Graham, C.B., Thomas, I.T., and Smith, D.W. (1974). A new overgrowth syndrome with accelerated skeletal maturation, unusual facies, and camptodactyly. *J. Pediatr.* 84, 547–552.
- Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* 573, 281–286.
- Weinberg, D.N., Rosenbaum, P., Chen, X., Barrows, D., Horth, C., Marunde, M.R., Popova, I.K., Gillespie, Z.B., Keogh, M.-C., Lu, C., et al. (2021). Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. 53.
- Weinert, B.T., Narita, T., Satpathy, S., Srinivasan, B., Hansen, B.K., Schölz, C., Hamilton, W.B., Zucconi, B.E., Wang, W.W., Liu, W.R., et al. (2018). Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. *Cell* 174, 231–244.e12.
- Wilmot, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H.S. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature* 385, 810–813.
- Woodcock, C.L. (2005). A milestone in the odyssey of higher-order chromatin structure. *Nat. Struct. Mol. Biol.* 2005 128 12, 639–640.
- Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010).

- Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444–447.
- Wu, X., Johansen, J.V., and Helin, K. (2013). Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol. Cell* **49**, 1134–1146.
- Xu, K., Wu, Z.J., Groner, A.C., He, H.H., Cai, C., Lis, R.T., Wu, X., Stack, E.C., Loda, M., Liu, T., et al. (2012). EZH2 Oncogenic Activity in Castration-Resistant Prostate Cancer Cells Is Polycomb-Independent. *Science* (80-.). **338**, 1465–1469.
- Xu, Q., Xiang, Y., Wang, Q., Wang, L., Brind'Amour, J., Bogutz, A.B., Zhang, Y., Zhang, B., Yu, G., Xia, W., et al. (2019). SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat. Genet.* **51**, 844–856.
- Youmans, D.T., Schmidt, J.C., and Cech, T.R. (2018). Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits. *Genes Dev.* **32**, 794–805.
- Yu, J.-R., Lee, C.-H., Oksuz, O., Stafford, J.M., and Reinberg, D. (2019). PRC2 is high maintenance. *Genes Dev.* **33**, 903–935.
- Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286**, 7983–7989.
- Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* (80-.). **328**, 916–919.
- Zha, Y., Xia, Y., Ding, J., Choi, J.H., Yang, L., Dong, Z., Yan, C., Huang, S., and Ding, H.F. (2014). MEIS2 is essential for neuroblastoma cell survival and proliferation by transcriptional control of M-phase progression. *Cell Death Dis.* **5**, e1417–e1417.
- Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I.S. (2008a). Roles of the Ctr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat. Struct. Mol. Biol.* **15**, 381–388.
- Zhang, Q., Agius, S.C., Flanigan, S.F., Uckelmann, M., Levina, V., Owen, B.M., and Davidovich, C. (2021a). PALI1 facilitates DNA and nucleosome binding by PRC2 and triggers an allosteric activation of catalysis. *Nat. Commun.* **12**, 1–18.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008b). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, 1–9.
- Zhang, Y., Liu, C., Wu, B., Li, L., Li, W., and Yuan, L. (2021b). The missing linker between SUN5 and PMFBP1 in sperm head-tail coupling apparatus. *Nat. Commun.* **12**, 1–16.
- Zheng, Y., Sweet, S.M.M., Popovic, R., Martinez-Garcia, E., Tipton, J.D., Thomas, P.M., Licht, J.D., and Kelleher, N.L. (2012). Total kinetic analysis reveals how combinatorial methylation patterns are established on lysines 27 and 36 of histone H3. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13549–13554.
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1–10.

APPENDICES

Appendix I: Primers and Oligonucleotides

I.I Genotyping primers

Primer ID	Forward (5'-3')	Reverse (5'-3')
<i>mPali1</i> geno_f1	GTTGCTGGGATTTTAACTCAGG	–
<i>mPali1</i> geno_r1	–	GTCACATACATCAGCTTCTTAGC
<i>mPali1</i> geno_r2	–	GTTAGGATCAGCTACGGGT
<i>Ezh2</i> 3'	–	CTGCTCTGAATGGCAACTCC
<i>Ezh2</i> 5'	TTATTCATAGAGCCACCTGG	–
<i>Ezh2</i> 5'ext	ACGAAACAGCTCCAGATTCAGGG	–
<i>mNsd1</i> _geno_f1	CCTCCTCCTCCTCCATCACTACC	–
<i>mNsd1</i> _geno_f2	ACGGGCTGTTTCCATGTGGCAG	–
<i>mNsd1</i> _geno_r1	–	CCAGGACTACTGCAGAGCACTG
<i>mNsd1</i> _geno_r2	–	ATGGCCGTTCCACTGCACCATG
<i>Gm340-U20796</i>	TAGATCGTCCGGTATTGGCAG	–
<i>Gm340-U32111</i>	GGGAACGCCAGTTTGAGATGC	–
<i>Gm340-L32466</i>	–	GGGAGTTGTGCTGCGGTCTGC

I.II CRISPR gRNAs

gRNA	Target sequence (5'-3')
<i>Pali1</i> gRNA #1	ATGTAATTAGATTGCAAAGC
<i>Pali1</i> gRNA #2	GATTCCAAGTATAAAACCCA
<i>Pali1</i> gRNA #3	AGTGTAGAAAGTATGAAGTA
<i>Pali1</i> gRNA #4	CCCCGTAACCTACACCGTGAA
<i>Nsd1</i> gRNA #2	TATCAAGCCTAATTTAGTGG
<i>Nsd1</i> gRNA #3	GTAGCGAGCAAAGTTTGCTG
<i>Nsd1</i> gRNA #4	AATTAAAGTTTCCCTTTACG

III Primers for amplifying ISH probe template DNA

Probe target	Forward (5'-3')	Reverse (5'-3')
mPali1 mRNA	GCGTCTAATACGACTCACTATAGGG TCACTGCCATCCTCTGACTCTA	ACCCTGAGGATACCCACCATAA
mLcor mRNA	GCGTCTAATACGACTCACTATAGGG CCCAGCCTCTCCTTTACTTTGT	AAGCTTACAGGATGGAACCAGG
mPali2 mRNA	GAGGCTAATACGACTCACTATAGGG CCAATCAACTGTTCTGTGGCTG	TACAGGCAGTCGGGTTTTAGAC
mLcor1 mRNA	GAGGCTAATACGACTCACTATAGGG GGAGGAGTCTTCAGTGTTCAG	CCACTTAGAAGCCTTACCAGCA
mSix2 mRNA	GCGTAATACGACTCACTATAGGGAA TGGAACAGGCAAGTCTGC	GCCTGCAGGACTCCATACTC
mLhx1 mRNA	GCGTAATACGACTCACTATAGGGTC TCTGTACAACCACGAGCTCC	ATCTCTCCAAGCGATCTGGTTCGC
mPax2 mRNA	GCGTAATACGACTCACTATAGGGCT TCCTCGCGTTTCTCTTCTCAC	GCAAAGCAGACCCCTTCTCCG

Note: bold text signifies non-target sequence included to facilitate subsequent *in vitro* transcription by T7 RNA polymerase. The T7 promoter sequence is underlined.

I.IV RT-qPCR mRNA primers

Target mRNA	Forward (5'-3')	Reverse (5'-3')
m18S rRNA	GTAACCCGTTGAACCCATT	CCATCCAATCGGTAGTAGCG
mRplp0	TTCATTGTGGGAGCAGAC	CAGCAGTTTCTCCAGAGC
mPali1	ATCTGGAAACCCATGTCTGGTCC	CTTTATGGTGGGTATCCTCAGGGTC
mLcor	GATACTGGAGGAAGCCATCTCGG	TTACAGAAACATCTGGCCCCCTCC
mMeis2	GTTCCGTCATTTCTGTTCTCC	CTTTCTCGCTCGCTCTCACT
mCbx4	TGAAATGGAGAGGCTGGTCCCC	TGGGCCCTCTCTTGCGATATCC
mCntln	TAAAGGAGGAGCTGGCCCTGTG	TGAAGACGCCTCCACAGAGACC
mAkt3	AGCGAGTGTCTCTGAGGACCG	TGGTAGCTGCATCTGTGATCCC
mTlx2	TTTGCAAAGGACAGGCTCACGG	TGGTAGGGATGGCCTATGCGAC
mZfp521	GAACCACATCCAGACGGTGCAC	GGTAGGTCTTCTCTCATCGGACTGG
mNsd1 Δ	GCCTCCGAAGACAAGGACAGCC	ATGTTCCACTGGCAGCAGTCGG
mNsd1 PWWP	TGTGCTGTGATTCTTGCCCTGC	AGAGGGAACAGCTCGAGGATGG
mNsd1 SET	TTACAAAGGGCTGGGGTCTCC	TGTCGTGCTCTTGTGCGTAACG
hAKT3	TGTCGAGAGAGCGGGTGTCTC	ATGGTGGCTGCATCTGTGATCC
hCBX4	ACCTTTGCCCCGTCGTTCCAATG	ACGGTTGTCAGTGGAGGAGTCC
hMEIS2D	CGAGCAGTGAGCCAAGGAG	GGCATGCTCTGCAAACCTGCAG
hGJA1	AGATCCCTGCCCACATCAGGTG	CACCAAGGACACCACCAGCATG
hZNF629	TCAGGGAGTCAGTGGGGAAAGC	AAGGAGTCTGGGAGCTGCTGAC

I.V ChIP-qPCR primers

Target locus	Forward (5'-3')	Reverse (5'-3')
<i>mGapdh</i> tss	AGTGTGCACCAAGGACATCCAG	CCCATTTTACTCGGGAAGCAG
<i>mPax3</i> tss	AGAAAGGCGGAAAGAGATTAGGAC	GAGCCAGTGAGGATGGAAAAAG
<i>mPax7</i>	TAAACCTCTTGGCGTCTCTTCCTTTC	AACTTCCTTGTCGCCACCTTCC
<i>mT</i> tss	TCCGCAGAGTGACCCTTTTTTC	TACCCAACAGCCACCTTCACTTC
<i>mGata4</i> tss	TCTTTCCTCCCTACTCTCAGTGGTCC	TCACCTTCTCCTCTACCAGCCCC
<i>mGata6</i> tss	TTTTTCTCCTCCCCTGGATCC	CTCTCTACAAGCTGCTGGGTGC
<i>mEomes</i> tss	AAGAGAAAAATGAAACTCGCAGGCG	TTTGCTATTGGCTGTTAGAGGTGACA
<i>mNkx2.9</i> tss	TAAGGATGGAAGTGCGAGGC	TTCGCTCCAGCACTCATTCA
<i>mChd3</i>	GGTCATCAGCCGCTACAAACATTCG	GCAAGCAGAGGTGGGACAGTG
<i>mGng12</i>	CCCCTCCTCCGATCTCTAGACG	TCTCCACGATCAGCTCCTTACCC
<i>mEvx2</i> (nucleation)	TTTCTCCCACCCTTCTCAAA	GGAGATTCCATTCCCTCCAC
<i>mEvx2</i> (spreading)	CGGGGAGCATTATGGAAATA	GACGCCCTGGGTAGACTTTT
<i>mLhx1</i> (spreading)	TGAGACGTTGGCACTTTTCAG	CCACAGGCAGTGACCCTAGT
<i>mHoxb13</i> intergenic 1	AAGCAACAGCAGCAACACAC	TGCCGTCGTCTTTTCCAAAG
<i>mLrg1</i> intergenic 2	TTTGGCTGGACACTGTGTTG	ACAGGATGGATTGCTGTGC
<i>mHoxa13</i> tss	AGCAAGGGTCCCAGGACATTTTC	TAGAGCCTGAGAACAGGGAAGGG
<i>mMeis2</i> prom	TTGACTGAGGTGTGCATGGC	CAACAGCGAGGTGTCCACAC
<i>mSix3</i>	AGGGGGAAGGATGCCAAATG	GCGAGATGCAGCAAATCCAG

Appendix II: Antibodies

II.1 Primary antibodies

Target	Source	Western blot	ChIP
PAL11	Merck (ABE1367)	1:500	-
EZH2	BD43 (Bracken et al., 2007)	1:10	-
EZH2	Abcam (ab191250)	1:1000	-
SUZ12	Cell Signalling (3737)	1:2000	1 µg
EED	Cell Signalling (51673S)	1:500	-
RYBP	Sigma-Aldrich (ab3637)	1:1000	-
BMI1	DC9 (Bracken et al., 2007)	1:20	-
CBX4	Abcam (ab242149)	1:1000	-
Histone H3	Active Motif (0301)	1:50000	-
Histone H3	Abcam (ab1791)	1:50000	1 µg
H3K27ac	Abcam (ab4729)	1:1000	2 µg
H3K27ac	Sigma-Aldrich (07-360)	1:1000	-
H3K27me1	Active Motif (0321)	1:2000	-
H3K27me2	Cell Signalling (9728S)	1:2000	2.25 µg
H3K27me3	Cell Signalling (custom)	-	1 µg
H3K27me3	Active Motif (0323)	1:2500	-
H3K36me2	Abcam (ab9049)	1:2500	5 µg
H3K36me2	Cell Signalling (2901)	1:5000	-
H3K36me3	Cell Signalling (4909)	1:1000	-
H2AK119ub	Cell Signalling (8240)	1:2000	-
NSD1	Abxexa (abx135901)	1:1000	-
MEIS2	Sigma (clone 1H4)	1 µg/mL	-
mTOR	Cell Signalling (2983)	1:1000	-
p-mTOR (Ser2448)	Cell Signalling (5536)	1:1000	-
HA	Cell Signalling (3724)	1:1000	-
FLAG	Sigma-Aldrich (clone M2)	1:1000	-
ACTIN	Proteintech (60008-1-Ig)	1:5000	-
GAPDH	Proteintech (60004-1-Ig)	1:5000	-
VINCULIN	Cell Signalling (4650)	1:2500	-
Mouse IgG	Millipore (12-371)	-	2 µg

II.II Secondary antibodies

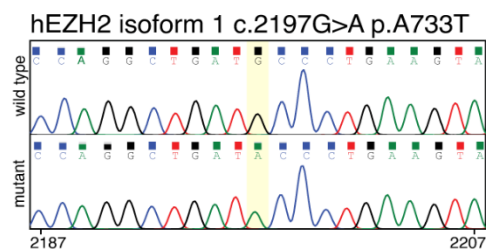
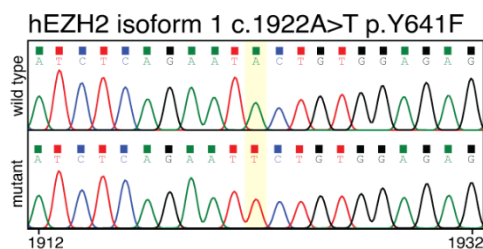
Antibody	Source	Western blot
IRDye 800CW Goat anti-Mouse IgG	Li-COR	1:10,000
IRDye 800CW Goat anti-Rabbit IgG	Li-COR	1:10,000
IRDye 680RD Goat anti-Mouse IgG	Li-COR	1:10,000
IRDye 680RD Goat anti-Rabbit IgG	Li-COR	1:10,000
IRDye 680LT Goat anti-Mouse IgG	Li-COR	1:20,000
IRDye 680LT Goat anti-Rabbit IgG	Li-COR	1:20,000
Goat anti-mouse IgG	Merck (401253)	variable
Goat anti-rabbit IgG	Merck (A0545)	variable

Appendix III: Wild type *hEZH2* ORF

ATGGGCCAGACTGGGAAGAAATCTGAGAAGGGACCAGTTTGTTGGCGGAAGCGTGTA
ATCAGAGTACATGCGACTGAGACAGCTCAAGAGGTTTCAGACGAGCTGATGAAGTAAAGA
GTATGTTTAGTTCCAATCGTCAGAAAATTTTGGAAGAACGGAAATCTTAAACCAAGAA
TGGAACAGCGAAGGATACAGCCTGTGCACATCCTGACTTCTGTGAGCTCATTGCGCGG
GACTAGGGAGTGTTCCGTGACCAGTGACTTGGATTTTCCAACACAAGTCATCCCATTAA
AGACTCTGAATGCAGTTGCTTCAGTACCCATAATGTATTCTTGGTCTCCCCTACAGCAG
AATTTTATGGTGGAAGATGAACTGTTTTACATAACATTCCTTATATGGGAGATGAAGT
TTTAGATCAGGATGGTACTTTCATTGAAGAACTAATAAAAAATTATGATGGGAAAGTAC
ACGGGGATAGAGAATGTGGGTTTATAAATGATGAAATTTTGTGGAGTTGGTGAATGCC
CTTGGTCAATATAATGATGATGACGATGATGATGATGGAGACGATCCTGAAGAAAGAGA
AGAAAAGCAGAAAGATCTGGAGGATCACCGAGATGATAAAGAAAGCCGCCACCTCGGA
AATTCCTTCTGATAAAATTTTGAAGCCATTTCTCAATGTTTCCAGATAAGGGCACA
GCAGAAGAACTAAAGGAAAAATATAAAGAACTCACCGAACAGCAGCTCCCAGGCGCACT
TCCTCCTGAATGTACCCCCAACATAGATGGACCAATGCTAAATCTGTTTCAGAGAGAGC
AAAGCTTACACTCCTTTTCATACGCTTTTCTGTAGGCGATGTTTTAAATATGACTGCTTC
CTACATCCTTTTCATGCAACACCCAACACTTATAAGCGGAAGAACACAGAAACAGCTCT
AGACAACAAACCTTGTGGACCACAGTGTTACCAGCATTTGGAGGGAGCAAAGGAGTTTG
CTGCTGCTCTCACCGCTGAGCGGATAAAGACCCACCAAACGTCCAGGAGGCCGCAGA
AGAGGACGGCTTCCCAATAACAGTAGCAGGCCCAGCACCCCCACCATTAATGTGCTGGA
ATCAAAGGATACAGACAGTGATAGGGAAGCAGGGACTGAAACGGGGGGAGAGAACAATG
ATAAAGAAGAAGAAGAGAAGAAAGATGAACTTCGAGCTCCTCTGAAGCAAATTTCTCGG
TGTCAAACACCAATAAAGATGAAGCCAAATATTGAACCTCCTGAGAATGTGGAGTGGAG
TGGTGCTGAAGCCTCAATGTTTAGAGTCCTCATTTGGCACTTACTATGACAATTTCTGTG
CCATTGCTAGGTTAATTGGGACCAAACATGTAGACAGGTGTATGAGTTTAGAGTCAAA
GAATCTAGCATCATAGCTCCAGCTCCCGCTGAGGATGTGGATACTCCTCCAAGGAAAAA
GAAGAGGAAACACCGGTTGTGGGCTGCACACTGCAGAAAGATACAGCTGAAAAAGGACG
GCTCCTCTAACCATGTTTACAACATCAACCCTGTGATCATCCACGGCAGCCTTGTGAC
AGTTCGTGCCCTTGTGTGATAGCACAAAATTTTGTGAAAAGTTTTGTCAATGTAGTTC
AGAGTGTCAAAACCGCTTTCCGGGATGCCGCTGCAAAGCACAGTGCAACACCAAGCAGT
GCCCCGTGCTACCTGGCTGTCCGAGAGTGTGACCCTGACCTCTGTCTTACTTGTGGAGCC
GCTGACCATTGGGACAGTAAAAATGTGTCCTGCAAGAACTGCAGTATTCAGCGGGGCTC
CAAAAAGCATCTATTGCTGGCACCATCTGACGTGGCAGGCTGGGGGATTTTTATCAAAG
ATCCTGTGCAGAAAAATGAATTCATCTCAGAATACTGTGGAGAGATTATTTCTCAAGAT
GAAGCTGACAGAAGAGGGAAAGTGTATGATAAATACATGTGCAGCTTCTGTTCACCTT
GAACAATGATTTTGTGGTGGATGCAACCCGCAAGGGTAACAAAATTCGTTTTGCAATC
ATTGCGTAAATCCAACTGCTATGCAAAAGTTATGATGGTTAACGGTGATCACAGGATA
GGTATTTTTGCCAAGAGAGCCATCCAGACTGGCGAAGAGCTGTTTTTTGATTACAGATA
CAGCCAGGCTGATGCCCTGAAGTATGTGCGCATCGAAAGAGAAATGGAAATCCCTTGA

Appendix IV: Sanger sequencing of mutant *hEZH2* constructs

CONTROLS



WEAVER SYNDROME ASSOCIATED

