Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.
The role of Polycomb Repressive Complex 2 associated proteins in mediating complex recruitment

Gerard L. Brien

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

Thesis Supervisor: Dr. Adrian P. Bracken

Cancer Epigenetics Laboratory
Department of Genetics
University of Dublin
Trinity College
Dublin 2
DECLARATION

I certify that this thesis, submitted to Trinity College Dublin for the degree of Doctor of Philosophy, has not been submitted as an exercise for a degree at this or any other university. I certify that the work presented here is entirely my own, except where otherwise acknowledged. This thesis may be available for consultation within the university library and may be copied or lent to other libraries upon request.

Gerard L. Brien
September 2013
Acknowledgements

Firstly I would like to thank my supervisor, Dr. Adrian Bracken, for his mentorship and constant support over the course of this work. I would also like to thank Dr. Gerard Cagney for his input and excellent collaboration on all mass spec aspects of this work. Thanks are also due to Dr. David O'Connell, Prof. Brendan Loftus and Dr. Krishna Sinha for their collaborations on Surface Plasmon Resonance (SPR) and ChIP–seq experiments and for generous reagent donations. Eiseart Dunne provided excellent bioinformatics support during ChIP–seq analyses and Emilia Jerman contributed supporting data related to the work presented in Chapter 4, for which I am extremely grateful.

I would like to thank all members of the Bracken laboratory, both past and present, for helpful support and discussions. In particular, I would like to thank Dr. Chris Egan, for his always enjoyable scientific (and non-scientific) discussions.

I would also like to thank all of my friends and family, particularly my girlfriend Rachael, for their constant love and support; and for helping to keep things in perspective.

Finally, and most of all, I would like to thank my parents for giving me a life filled with all of the opportunities they themselves were not fortunate enough to have.
# Table of contents

1 Introduction  
1.1 General Introduction 2  
1.2 Chromatin and Epigenetics 3  
1.3 Covalent histone modifications 4  
1.3.1 Histone Acetylation 6  
1.3.2 Histone Methylation 8  
1.4 Epigenetics and Cancer 10  
1.4.1 Histone Acetylation and Cancer 11  
1.4.2 Histone Methylation and Cancer 13  
1.5 Polycomb group proteins 15  
1.5.1 Polycomb Repressive Complexes 17  
1.5.2 Polycomb target genes 20  
1.5.3 Polycombs and cancer 21  

2 Materials and Methods 24  
2.1 Cell Culture 25  
2.2 Fibroblast cell culture 25  
2.3 NT2/D1 cell culture 25  
2.4 Embryonic stem cell culture 25  
2.5 HMEC cell culture 26  
2.6 Inducible GAL4-PCL3 HEK293 cell line culture 26  
2.7 Retroviral transduction 26  
2.7.1 Retroviral shRNA sequences 27  
2.8 Lentiviral transduction 27  
2.8.1 Lentiviral shRNA sequences 27  
2.9 3T3 assays and BrdU FACs 28  
2.10 Cloning, plasmid generation and mutagenesis 28  
2.10.1 Gateway Cloning 28  
2.10.2 ORF cloning and sub-cloning 29  
2.10.3 ORF mutagenesis 29  
2.11 Real-Time Quantitative PCR 30
3 The Polycomb PCL3 binds H3K36me3 to recruit PRC2 and the demethylase NO66 to embryonic stem cell genes during differentiation

3.1 Introduction 38
3.2 Results 40
3.2.1 PCL3 recruits the lysine demethylase NO66 to chromatin 40
3.2.2 The PCL3 tudor domain binds H3K27me3 and H3K36me3 in vitro 45
3.2.3 PCL3 localises with PRC2 and H3K27me3, not H3K36me3 in vivo 47
3.2.4 PCL3 controls PRC2 and NO66 target gene occupancy 49
3.2.5 PCL3 is necessary for de novo recruitment of NO66 and PRC2 60
3.2.6 A PCL3 W50C Y56A mutant cannot bind H3K27me3 or H3K36me3 65
3.2.7 PCL3 tudor domain is required for PRC2 and NO66 recruitment 67
3.3 Discussion 74

4 Variant PRC2 Complexes between quiescent and proliferating cells defined by PCL members

4.1 Introduction 82
4.2 Results 84
4.2.1 PCL proteins form distinct PRC2 complexes in quiescent and cycling cells 84
4.2.2 PCL proteins play divergent roles in cell growth control 88
4.2.3 PCL1, but not PCL2 or PCL3, stabilizes p53 to promote a G0/1 cell cycle block.

4.2.4 All three PCL proteins are required to prevent the onset of cellular senescence.

4.2.5 PCL protein containing PRC2 complexes cycle on and off chromatin in quiescent and cycling cells.

4.3 Discussion

5 Proteomic analyses of the PRC2 component EZH2 in pluripotent and differentiating cells

5.1 Introduction

5.2 Results

5.2.1 Interaction proteomics of EZH2 in pluripotent mouse ES cells

5.2.2 Interaction proteomics of EZH2 in a dynamic cell fate model

5.2.3 Validation of EZH2 interacting proteins

5.3 Discussion

6 General Discussion

6.1 Introduction

6.2 Do PCL proteins instruct PRC2 recruitment?

6.3 PCL proteins in adult stem and progenitor populations

6.4 PCL proteins as targets for cancer therapy

6.5 Non-histone substrates for EZH2

6.6 Conclusions
Table of Figures

Figure 1.1: The nucleosome ................................................................. 4
Figure 1.2: Sites of histone covalent modifications .......................... 5
Figure 1.3: Epigenetic gene regulatory mechanisms ......................... 6
Figure 1.4: Polycombs regulate body patterning and Hox gene expression 17
Figure 1.5: Polycomb group proteins form distinct multi-protein complexes that collaborate to repress gene expression ........................................... 19
Figure 1.6: Polycomb function dynamically during cell fate decisions 21
Figure 2.1: Gateway cloning system ...................................................... 29
Figure 3.1: Biochemical analysis of PCL3 ............................................. 41
Figure 3.2: Validation of the interaction of PCL3 with PRC2 and NO66 .... 42
Figure 3.3: PCL3 recruits the H3K36me3 demethylase NO66 to chromatin to repress gene transcription ............................................. 44
Figure 3.4: The PCL3 tudor domain binds to H3K27me3 and H3K36me3 modified Histone H3 peptides in vitro ................................................. 46
Figure 3.5: PCL3 co-localises with PRC2 and H3K27me3, but not H3K36me3, in vivo............................................................................................................. 48
Figure 3.6: NO66 binds Polycomb repressed genes in mouse embryonic stem cells ........................................................................................................ 50
Figure 3.7: shRNA mediated knockdown of PLC3 in mouse embryonic stem cells ....................................................................................................... 52
Figure 3.8: PCL3 is required to maintain PRC2 and NO66 binding on target genes in ES cells ...................................................................................... 53
Figure 3.9: Depletion of PCL3 does not significantly affect the ES cell phenotype ........................................................................................................ 54

Figure 3.10: Depletion of PCL3 in ES cell has minimal transcriptional affects on stem cell and differentiation genes ............................................................ 55

Figure 3.11: Depletion of NO66 in mouse ES cells ........................................... 57

Figure 3.12: Depletion of NO66 in mouse ES cells does not affect PCL3 occupancy on target genes ........................................................................... 58

Figure 3.14: PCL3 is required for ES cell differentiation ................................... 61

Figure 3.15: Depletion of PCL3 leads to impaired silencing of ES cell expressed genes following embryoid body differentiation of ES cells .... 62

Figure 3.16: Depletion of PCL3 leads to impaired silencing of ES cell expressed genes following retinoic acid induced differentiation of ES cells .......................................................................................................................... 63

Figure 3.17: PCL3 is required for the de novo recruitment of PRC2 and NO66 to ES cell genes during differentiation ......................................................... 64

Figure 3.18: Mutant PCL3 tudor domain is incapable of binding modified histone peptides .............................................................................................. 66

Figure 3.19: PCL3 knockdown rescue experiment ........................................... 68

Figure 3.20: Wildtype, but not mutant PCL3 is capable of restoring PRC2 and NO66 binding at target genes in PCL3 depleted ES cells ...................... 69

Figure 3.21: PCL3 tudor domain mutant does not rescue the differentiation defects seen following PCL3 depletion ......................................................... 70

Figure 3.22: Wildtype but not mutant PCL3 rescues the de novo recruitment of PRC2 and NO66 to stem cell genes during differentiation ................. 71

Figure 3.23: Model for PCL3 function in ES cells and during differentiation .. 73
Figure 4.1: PCL proteins form distinct PRC2 complexes in quiescent and cycling cells................................................................. 85

Figure 4.2: The expression of the PCL proteins is regulated during cell growth. .................................................................................................. 86

Figure 4.3: The PCL2/3 genes are downstream of the E2F pathway, while PCL1 is downstream of the p53 pathway ........................................ 87

Figure 4.4: All 3 PCL proteins bind the INK4A gene promoter and are dissociated in senescent cells. .................................................. 90

Figure 4.5: Ectopic expression of PCL1–3 leads to reduced p16^{INK4A} levels, but only PCL2 and PCL3 confers a proliferative advantage .......... 91

Figure 4.6: Ectopically expressed PCL proteins are recruited to the INK4A gene promoter.................................................................. 92

Figure 4.7: The PCL3 tudor domain is essential for recruitment to, and repression of, INK4A................................................................. 93

Figure 4.8: The proliferative advantage conferred by PCL2 and PCL3 is dependent on p16^{INK4A} ................................................................. 94

Figure 4.9: PCL1, but not PCL2 or PCL3, interacts with p53 .................................................................................................................. 96

Figure 4.10: Ectopic expression of PCL1 specifically stabilizes p53 protein levels ......................................................................................... 97

Figure 4.11: The G0/1 cell cycle block seen upon PCL1 overexpression is dependent on p53 ......................................................................... 98

Figure 4.12: PCL1 mediated stabilisation of p53 leads to increased p53 binding at target genes................................................................. 99

Figure 4.13: PCL proteins are required to prevent senescence onset in HDFs .......................................................................................... 101
Figure 4.14: PCL proteins are required to maintain Polycomb binding at the INK4A gene in HDFs................................................................. 102

Figure 4.15: PCL proteins are required to prevent cellular senescence in mouse embryonic fibroblasts.............................................................. 103

Figure 4.16: PCL proteins are required to maintain Polycomb binding at the INK4A gene in MEFs ................................................................. 104

Figure 4.17: Depletion of PCL1, but not PCL2 or PCL3, leads to reduced expression of p53 target genes......................................................... 105

Figure 4.18: PCL protein containing complexes cycling on and off target genes in quiescent and cycling cells................................................. 107

Figure 4.19: Polycomb-like proteins and PRC2 in quiescent and cycling cells .......................................................................................... 108

Figure 5.1: The EZH2 protein interactome in mouse ES cells ...................... 117

Figure 5.2: A dynamic model system to study PRC2 recruitment............... 119

Figure 5.3: Purifications of EZH2 in pluripotent and differentiating NT2/D1 cells....................................................................................... 120

Figure 5.4: The EZH2 protein interaction network during neuronal differentiation of NT2/D1 cells............................................................. 123

Figure 5.5: FLAG tagged CEBPZ and PHF14 associate with EZH2......... 125
Summary

All multi-cellular organisms are constructed from a myriad of distinct cell types, each performing specialized functions. With very few exceptions all cell types contain the same genetic information encoded in the genome. Therefore, this specialization is achieved through specific gene expression programs. Cells need these gene expression programs to be maintained during growth and division throughout the entire life of the cell. This implies the presence of a ‘cellular memory’ system that ensures faithful transmission of pre-existing expression programs between mother and daughter cells.

Polycomb group proteins form part of this cellular memory system. Polycombs exist in two large multimeric protein complexes, termed Polycomb Repressive Complex (PRC)1 and PRC2. These complexes act on their target genes by modulating chromatin structure. Once recruited to target genes, Polycombs establish a silent chromatin state that is maintained through cell division. In addition to a role in stably maintaining cell fates, Polycombs also play essential roles during the transitionary phases of stem cell differentiation. Deregulation of Polycomb function, particularly of PRC2 function, can lead to cancer. Thus an improved understanding of PRC2 function/mechanisms will yield fundamental advances in our understanding of developmental and cancer biology.

Chapter III

How Polycomb proteins access active chromatin in order to confer transcriptional silencing during lineage transitions remains unclear. In this chapter I show that the PRC2 component PCL3 binds the active chromatin mark H3K36me3 via its tudor domain. PCL3 associates with the H3K36me3 demethylase NO66, and is required to recruit the PRC2 complex and NO66 to stem cells genes during differentiation, leading to PRC2 mediated H3K27 trimethylation, loss of H3K36me3 and transcriptional silencing. I propose a model whereby PCL3 functions during mouse embryonic stem cell differentiation to transiently bind the H3K36me3 mark via its tudor domain,
forming essential contact points that allow recruitment of PRC2 and H3K36me3 demethylase activity to active gene loci during their transition to a Polycomb-repressed state.

**Chapter IV**

The three PCL proteins (PCL1–3) are ancillary PRC2 components that contribute to complex recruitment through binding the local chromatin environment. However, why three complex components with apparent functional redundancy exist, is unclear. In this chapter I show that PCL1–3 are mutually exclusive, cell growth regulated PRC2 components, which exist in variant forms of the complex in quiescent and cycling cells. In quiescent cells, PCL1 has a dual role as the only PCL present in the PRC2 complex and independently of the complex it promotes a G0/1 block by stabilizing p53. PCL2 and PCL3 expression is induced in cycling cells where all three PCLs exist in mutually exclusive forms of the PRC2 complex, moreover ectopic expression of PCL2 or PCL3 immortalizes cells by repressing INK4A. My results demonstrate that PCL proteins define dynamic cycling forms of PRC2 in cells at different stages of the cell cycle, and that PCL proteins control cell cycle progression both, independently of, and as part of PRC2.

**Chapter VI**

In *D. melanogaster* the PRC2 complex is targeted to chromatin through interactions with a limited number of DNA binding transcription factors. However, evidence in the mammalian system suggests that this mechanism may have diversified to include many additional transcription factors. In this chapter I undertook to perform an in depth proteomic analysis of the PRC2 component EZH2, to identify factors that interact with it in mammalian cells. I performed large-scale purifications of endogenous EZH2 in pluripotent stem cells both before and during differentiation. This is the first work of its type to perform such an analysis at an endogenous level and in a dynamic cell fate model. My analyses identified a large number of DNA binding transcription factors and other chromatin associated proteins that associate with EZH2,
many of which in a cell type specific manner. Preliminary follow-up experiments on two EZH2 associated factors, PHF14 and C/EBPZ, confirmed their interaction with EZH2. Current work is focused on characterizing the biological role of these interactions.
Abbreviations

ADRB2 – Adrenoceptor Beta 2  
AEBP2 – AE Binding Protein 2  
AF4/10/11 – ALL1-Fused Gene From Chromosome 4/10/11  
AML – Acute myeloid leukemia  
AML1-ETO – Acute myeloid leukemia 1–Eight Twenty One Protein  
APL – Acute promyelocytic leukemia  
ATP – Adenosine triphosphate  
ATRA – All-trans retinoic acid  
BCL2 – B-Cell CLL/Lymphoma 2  
BCL6 – B-Cell CLL/Lymphoma 6  
BET – Bromodomain and extra-terminal  
BMI1 – B Lymphoma Mo-MLV Insertion Region 1 Homolog  
BP2TF – Bromodomain PHD finger transcription factor  
BRCA1 – Breast cancer 1  
BRD2–4 – Bromodomain-containing protein 2–4  
BSA – Bovine serum albumin  
C/EBPα/β/γ/δ – CAAT/Enhancer binding proteinα/β/γ/δ  
CBP – CREB binding protein  
CBX1–8 – Chromobox Homolog 1–8  
CCNA2 – Cyclin-A2  
CCNB1 – Cyclin B1  
CDH1 – Cadherin 1  
CDK6 – Cyclin dependent kinase 6  
CDKN1A/B/C – Cyclin dependent kinase inhibitor 1A/B/C  
CDKN2A – Cyclin dependent kinase inhibitor 2A  
cDNA – complimentary DNA  
CHD1 – Chromodomain helicase DNA binding protein 1  
ChIP – Chromatin immunoprecipitation  
ChIP–seq – Chromatin immunoprecipitation–sequencing  
Co–REST – RE1-Silencing transcription factor co–repressor  
DIPG – Diffuse intrinsic pontine glioma  
DLBCL – Diffuse large B-cell lymphoma  
DNA – Deoxyribonucleic acid  
DOT1L – DOT1-Like  
DTT – Dithiothreitol  
E(Z) – Enhancer of zeste  
E1A – Early region 1A  
EBs – Embryoid bodies  
EDTA – Ethylenediaminetetraacetic acid  
EED – Embryonic ectoderm development  
ESC – Extra sex combs  
ES – embryonic stem  
EZH1/2 – Enhancer of zeste homolog 1/2  
FBS – Fetal bovine serum  
FBXL10 – F-Box/LRR–Repeat protein 10  
FGF – Fibroblast growth factor  
FL – Follicular lymphoma  
FOX – Forkhead box  
GATA1/3/4/6 – GATA Binding Protein 1/4/6  
GBM – Glioblastoma  
GNAT – Gcn5-related N-acetyltransferase  
GSH – Glutathione  
GST – Glutathione S-transferases
HA – Haemagglutinin
HAT – Histone acetyltransferase
HCL – Hydrochloric acid
HDAC – Histone deacetylase
HDF – Human diploid fibroblast
HDF – Human diploid fibroblasts
HDM – Histone demethylase
HMECs – Human mammary epithelial cells
HMT – Histone methyltransferase
HOXA1/5/9/11 – Homeobox A1/5/9/11
HSCs – Hematopoietic stem cells
ING1–5 – Inhibitor of growth family, member 1–5
IPTG – Isopropylthio-B-galactoside
JARID1A/B/C – Jumonji, AT rich interactive domain 1A/B/C
JARID2 – Jumonji, AT rich interactive domain 2
JARID2 – Jumonji, AT Rich Interactive Domain 2
JMJC – Jumonji C
JMJD2A – Jumonji domain-containing protein 2A
JMJD3 – Jumonji domain-containing protein 3
KCI – Potassium chloride
KDM5A/B – Lysine (K)-specific demethylase 5A/B
KLH – Keyhole limpet hemocyanin
LC/MS – Liquid chromatography mass spectrometry
LIF – Leukemia inhibitory factor
LSD1 – Lysine specific demethylase 1
MACS – Model-based analysis of ChIP-Seq
MBT – Malignant brain tumour
MDS – Myelodysplastic syndromes
MEF – Mouse embryonic fibroblasts
MEF – Mouse embryonic fibroblasts
Meis1a – Myeloid ecotropic viral integration site 1 homolog A
MgCl – Magnesium chloride
MLL – Mixed-lineage leukemia
MOZ – Monocytic leukemia zinc finger
MPD – Myeloproliferative disorders
MSX1 – Msh Homeo Box 1
MYC – Myelocytomatosis viral oncogene homolog
MYST – MOZ, Ybf2/Sas3, Sas2 and Tip60 protein
NaCl – Sodium chloride
NO66 – Nucleolar protein 66
NP40 – Nonidet P-40
NUP98 – Nucleoporin 98kD
NuRD – Nucleosome remodeling and deacetylase
NURF – Nucleosome remodeling factor
NURF55 – Nucleosome remodeling factor 55
NUT – Nuclear protein in testis
OLIG2 – Oligodendrocyte Lineage Transcription Factor 2
ORF – Open reading frame
PAI1 – Plasminogen activator inhibitor 1
PAX3 – Paired Box 3
PBS – Phosphate buffered saline
Pbx1 – Pre-B-Cell Leukemia Homeobox 1
PC – Polycomb
PCAF – P300/CREB-associated factor
PCGF1–6 – Polycomb group ring finger 1–6
PCL1–3 – Polycomb–like1–3
PCR – Polymerase chain reaction
PH – Polyhomeotic
PHD – Plant homeodomain
PHF23 – PHD finger 23
PLZF – Promyelocytic Leukemia Zinc Finger Protein
PLZF–RARα – Promyelocytic leukemia zinc finger– retinoic acid receptor alpha
PML–RARα – Promyelocytic leukemia–retinoic acid receptor alpha
PMSF – Phenylm ethanesulfonylfluoride
POU5F1 – POU Domain, Class 5, Transcription Factor 1
pRB – Retinoblastoma protein
PRC1/2 – Polycomb repressive complex 1/2
PRE – Polycomb response element
PSC – Posterior sex combs
RB1 – Retinoblastoma 1
RBBP4/7 – Retinoblastoma Binding Protein 4/7
RBP2 – Retinol binding protein 2
RCC – Renal cell carcinoma
RING1B – RING finger protein 1B
RIPA – Radio-immunoprecipitation assay
RNA – Ribonucleic acid
RYBP – RING1 And YY1 Binding Protein
SALL4 – Sal-Like 4
SCE/RING – Sex combs extra/Really interesting new gene
SDS–PAGE – Sodium dodecyl sulfate–Polyacrylamide gel electrophoresis
SET – Suppressor of variegation, enhancer of zeste, trithorax
shRNA – short hairpin RNA
SIN3A – SIN3 transcription regulator homolog A
SPR – Surface plasmon resonance
SUZ12 – Suppressor of zeste 12 homolog
SV40 – Simian virus 40
T–ALL – T-cell acute lymphoplastic leukemia
TE – Tris-EDTA
TFIIID – Transcription initiation factor IID
TGFβ – Transforming growth factor β
TP53 – Tumour protein 53
TSS – Transcription start site
UTF1 – Undifferentiated Embryonic Cell Transcription Factor 1
UTX – Ubiquitously transcribed tetratricopeptide repeat, X chromosome
WNT – Wingless-type MMTV integration site family
ZIC1 – Zic Family Member 1

1 Introduction
1.1 General Introduction

During the development of any multi-cellular organism, a myriad of specialized and molecularly distinct cell types must be generated. From the earliest stages of development this requires progression through a series of intermediate cell types, involving the step-wise reprogramming of cellular fates. Normally development takes place through a unidirectional process, characterized by an ever-decreasing developmental potential (Jaenisch and Young, 2008). This requires the correct spatial and temporal activation of specific gene expression programs. Equally important, a cell must repress genes required for alternative cell fates. Such ‘transcriptional programs’ are specified by DNA binding transcription factors during stem cell differentiation (Brien and Bracken, 2009). These transcription factors may act downstream of external signaling pathways, conveying extrinsic developmental cues from within the stem cell niche (Arnold and Robertson, 2009). Alternatively, asymmetric stem cell division can lead to differential compartmentalization of transcription factors, which initiate transcriptional changes in one daughter cell ultimately leading to differentiation (Knoblich, 2008). Once established, these cell type specific gene expression programs must be maintained in a stable fashion over many cell generations, long after the initial instructive development signals have disappeared. This implies the existence of a ‘cellular memory’ system to ensure faithful transmission of gene expression profiles from mother to daughter cells. Polycomb and trithorax group proteins form the molecular basis of this cellular memory system. These two families of antagonistic proteins maintain genes in a repressed and active state, respectively (Schuettengruber et al., 2007). Polycomb and trithorax group proteins control gene expression by modulating chromatin structure through ‘epigenetic mechanisms’. A major challenge in modern biology is to understand how such epigenetic mechanisms are employed to regulate the developmental decisions taken by stem and progenitor cells. This question has major implications for regenerative medicine as well as our understanding of the molecular events that lead to cancer, which frequently results from improper cellular differentiation (Feinberg et al., 2006).
1.2 Chromatin and Epigenetics

Chromatin is the macromolecular complex of DNA and histone proteins that forms the scaffold for assembling the entire genome of all eukaryotes. The basic functional unit of chromatin is the nucleosome, which consists of two molecules each of the core histone proteins H2A, H2B, H3 and H4. In a single nucleosome 145–147 base pairs of DNA are wrapped around the histone octamer (Fig. 1.1) (Luger et al., 1997). This unit is repeated throughout the genome, and chromatin structure can be further condensed through association of additional proteins, such as the linker histone H1. However, the packaging of eukaryotic genomes into chromatin is not uniform, as different parts of the genome are packaged into more or less accessible domains. It is believed that the specific composition of chromatin is an important determinant of cellular function. In general terms, chromatin can be subdivided into two major subclasses: (1) euchromatin, which exhibits a relatively 'open' structural conformation and contains mostly active genes and (2) heterochromatin, which is highly condensed and primarily contains inactive genes. Heterochromatin can be further subdivided into two categories: (1) Constitutive heterochromatin, which is characteristic of centromeres, telomeres and other repetitive DNA regions and (2) facultative heterochromatin. Unlike domains of constitutive heterochromatin, which exhibit similar patterns of localization from one cell type to the next, the formation of facultative heterochromatic domains is developmentally regulated and varies in a cell type specific manner. Efforts to study the coordinated regulation of chromatin have demonstrated that a major component of regulation is covalent modification of DNA, and of the core histone proteins within the nucleosome (Kouzarides, 2007). Broadly speaking, the concept of epigenetics relates to the processes pertaining to these regulatory mechanisms. The term epigenetics was first coined in 1942 by C. H. Waddington to describe heritable changes in cellular phenotype that occurred independently of alterations at the DNA level. Today epigenetics in the most general sense is used to describe chromatin-based events that regulate DNA templated processes. For the purposes of this thesis, epigenetics will refer
more specifically to heritable modifications to chromatin that affect gene expression patterns, and I will be focusing mainly on modifications to histone proteins.

Figure 1.1: The nucleosome.

The core nucleosome octamer contains a single H2A-H2B tetramer and two H3-H4 dimers. Around this core histone particle, 145–147 base pairs of DNA is wrapped in a 1.67 left handed superhelical turn.

1.3 Covalent histone modifications

Covalent modifications of chromatin, which affect gene expression, are dynamically deposited and removed by chromatin modifying proteins in a highly regulated manner during development and differentiation (Kouzarides, 2007). The core histone proteins within the nucleosome share a similar structural configuration, with a globular hydrophobic internal region and a non-
structured flexible N-terminal region that protrudes from the nucleosome (Fig. 1.1). These N-terminal ‘tails’ can be modified at many residues along their length. Furthermore, a limited number of residues within the globular domains of histones H2A, H2B and H3 are also subject to modification (Fig. 1.2). The most characterized modifications include acetylation, methylation, phosphorylation and ubiquitylation. These chemical modifications are dynamic, and are catalyzed by “epigenetic writer” proteins and removed by antagonistic “epigenetic eraser” proteins. The modifications can affect chromatin structure by altering non–covalent interactions within and between nucleosomes. They may also serve as docking sites for additional “epigenetic reader” proteins, containing specialized domain structures, which bind these modifications (Fig. 1.3). These reader proteins can recruit additional chromatin modifying and remodeling enzymes, which serve as the effectors of the modification. Within the context of transcriptional regulation, the acetylation and methylation of specific lysine and arginine residues have been the most extensively studied histone modifications. In the following sections I will outline the key conceptual paradigms that have been established from these studies.

Figure 1.2: Sites of histone covalent modifications.

The histone amino-terminal tail regions, which account for approximately one quarter of the mass of the nucleosome, are subject to extensive covalent modification. Modifications may also occur at a limited number of residues within the globular domains (boxed regions) as illustrated. In general, active
modifications include lysine acetylation (ac) and certain forms of arginine (me – blue) and lysine methylation (me – green). Repressive modifications generally include the alternative forms of lysine methylation (me – red). Lysine ubiquitylation (ub – brown) can be associated with activation or repression depending on the residue that is modified. Phosphorylation (ph – green/blue) is mainly associated with DNA damage response and may be involved in regulated chromatin compaction during DNA synthesis.

Figure 1.3: Epigenetic gene regulatory mechanisms.

Epigenetic regulatory proteins can be broadly categorized into three groups: 1. Epigenetic writers, which are the enzymes responsible for depositing covalent modifications on histone proteins. 2. Epigenetic readers, which contain specialized binding domains that bind to these covalent modifications. Following recruitment, these proteins themselves and/or additional associated factors can further modify and/or remodel the local chromatin environment. 3. Epigenetic erasers that catalyze the removal of the covalent modifications.

1.3.1 Histone Acetylation

1.3.1.1 Acetylation writers

Histone acetyltransferases (HATs), which ‘write’ the acetyl mark, were the first enzymes shown to modify chromatin (Roth et al., 2001). There are two major sub-classes of HATs: type-A and type-B. The type-B HATs predominantly
modify free histones in the cytoplasm, but not those already deposited in chromatin, and therefore do not affect gene transcription (Lee and Workman, 2007). The type-A HATs primarily function in the nucleus and are classified in three major families: GNAT, MYST and CBP/p300. These enzymes can modify multiple lysine residues, mainly but not exclusively, within histone N-terminal tail regions. Acetylation neutralizes the positive charge of lysine residues, disrupting the electrostatic interaction between histones and the negatively charged DNA. Hence, histone acetylation is most commonly associated with a more 'open' chromatin confirmation (Kouzarides, 2007). Consistent with this, genome-wide studies have confirmed that hyper-acetylated histones are present at promoters, enhancers, and in some cases, throughout the transcribed regions of active genes (Heintzman et al., 2009; Wang et al., 2008).

1.3.1.2 Acetylation readers

Lysine acetylation can also serve as a binding site for epigenetic 'reader' proteins via the intermediacy of a bromodomain. Bromodomains are protein modules found in a diverse group of proteins with roles associated with chromatin remodeling and transcriptional control (Mujtaba et al., 2007). For example, the TAF1 protein, which is a component of the TFIID complex in humans, has been shown to bind acetylated histone H4 via its tandem bromodomain structures (Jacobson et al., 2000). This interaction is important for TFIID assembly and transcriptional activation. Similarly, the bromodomain and extra-terminal (BET) family of bromodomain containing proteins, which includes BRD2, BRD3, BRD4 and BRD7, have been shown to read acetylated histone tails (Dey et al., 2003; Kanno et al., 2004; LeRoy et al., 2008; Morinie et al., 2009). Significantly, these interactions have been shown to be important for RNA Polymerase II mediated transcription and also for the targeting of associated protein complexes to effect transcription.
1.3.1.3 Acetylation erasers

Histone deacetylases (HDACs) are the enzymes that oppose HATs by erasing lysine acetylation. This has the affect of restoring the positive charge of the lysine residue, stabilizing the local chromatin environment. Consistent with this, HDACs have been shown to function predominantly as transcriptional repressors. In general the substrate specificity of the HDAC enzymes is low, as a single enzyme may be capable of deacetylating multiple different lysine residues (Marks et al., 2001). The substrate specificity and targeting of these enzymes is largely mediated by components of the large multi-protein complexes, such as NuRD, SIN3A and Co-REST, in which they are found (McDonel et al., 2009).

1.3.2 Histone Methylation

1.3.2.1 Methylation writers

Methylation of histone proteins mainly occurs on lysine and arginine residues. Unlike acetylation, histone methylation does not effect the electrostatic interactions within or between nucleosomes, and can be associated with transcriptional activation or repression. Furthermore, lysines may be mono-, di- or tri-methylated and arginines symmetrically or asymmetrically mono- or di-methylated (Kouzarides, 2007). Interestingly, different methylation states on a single residue can have different genomic localization and different transcriptional outcomes. For example, mono-methylation of H3K4 (H3K4me1) is generally associated with gene enhancer elements, whereas di- and tri-methylation of H3K4 (H3K4me2/3) are associated with the promoter regions of active genes (Ruthenburg et al., 2007a). Furthermore, asymmetric di-methylation of H3R2 (H3R2me2a) acts as a repressive mark, whereas symmetric di-methylation of H3R2 (H3R2me2s) is associated with activation (Guccione et al., 2007; Migliori et al., 2012). Histone methyl transferases (HMTs), of which there are now approximately 40 described, are the enzymes that ‘write’ these methyl marks. The enzymatic activity of HMTs comes from the characteristic SET (Suppressor of variegation, Enhancer of
Zeste, Trithorax) domain contained in all but one of the HMTs described to date. Unlike HATs, HMTs tend to have a high level of substrate specificity, with most enzymes being capable of modifying only a single amino acid substrate.

1.3.2.2 Methylation readers

As is the case for histone acetylation, the specific status of histone methylation can be 'read' by additional regulatory proteins. It is these proteins, and their associated activities, which act as the effectors of histone methylation. There are several characterized methylation reader domains; these include the PHD finger domain and members of the Tudor domain 'Royal Family' such as chromodomains, Tudor domains and MBT domains (Dawson et al., 2012). For example, the double chromodomain structures of the CHD1 protein, an ATP-dependant chromatin remodeler, have been reported to bind specifically to H3K4me3 (Sims et al., 2005). Significantly, the CHD1 protein has been shown to associate with the H3K4me3 marked promoter regions of active genes in mouse ES cells (Gaspar-Maia et al., 2009). Moreover, CHD1 is required for maintaining an 'open' chromatin environment at these sites. Similarly, the PHD finger domain containing protein BPTF, a component of the NURF chromatin-remodeling complex, binds H3K4me3 (Wysocka et al., 2006). This interaction is required for the association of NURF with chromatin, and subsequent chromatin remodeling.

1.3.2.3 Methylation erasers

It was thought for many years that histone methylation was relatively permanent, and that it could only be removed passively by dilution during DNA replication or by active histone exchange. However, in 2004 Shi and colleagues described for the first time a protein, LSD1, which had histone demethylating activity (Shi et al., 2004). Since this hallmark discovery there have been an additional 20 proteins identified with specific histone demethylating activities (Kooistra and Helin, 2012). These histone
demethylases (HDMs) can act either as transcriptional activators or repressors, depending on the residues that they demethylate, and like HMTs these enzymes are highly specific in terms of their substrate. For example, JARID1B, which specifically demethylates the activating marks H3K4me2/3 is generally thought of as a transcriptional repressor (Schmitz et al., 2011). Whereas, UTX, a demethylase that specifically removes H3K27me2/3 (marks associated with transcriptional repression) is considered to be a transcriptional activator (Agger et al., 2007; Lan et al., 2007). HDMs can be subdivided into two families: the LSD family and the JMJC family. The LSD demethylases are only capable of demethylating mono- and di-methylated lysine residues. The JMJC family employs a distinct catalytic mechanism, which means that JMJC demethylases can remove mono- di- and tri-methyl marks (Kooistra and Helin, 2012).

1.4 Epigenetics and Cancer
Possibly the most stark observation born out of the last 5 years of cancer genome sequencing studies is the sheer number of genes, which encode proteins involved in epigenetic gene regulation that are recurrently mutated in cancer (Ryan and Bernstein, 2012; Vogelstein et al., 2013). In many cases the epigenetic consequences of these genetic changes, and how/why these changes directly contribute to oncogenesis remains unclear. However, emerging evidence does indicate that these genetic changes do lead to epigenetic alterations, which can be directly linked to oncogenesis in some cases (Ryan and Bernstein, 2012). Indeed, aberrant patterning of epigenetic marks, both on a genome-wide scale and at specific loci, is a fundamental hallmark of human cancer. Therefore, epigenetic alterations owing to upstream genetic changes may push the evolution of the neoplastic state (Baylin and Jones, 2011).

The first epigenetic aberrations observed in cancer were genome-wide losses of DNA methylation (Feinberg and Vogelstein, 1983; Feinberg et al., 2006), with such hypomethylation of DNA leading to genomic instability and
increased tumour frequency in vivo (Eden et al., 2003; Gaudet et al., 2003; Holm et al., 2005). Localized hypermethylation of the promoter regions of silenced tumour suppressor genes, such as CDKN2A and RB1, are also characteristic features of many human tumours (Pujadas and Feinberg, 2012). Furthermore, DNA hypermethylation events can also occur over large sub-chromosomal domains encompassing multiple repressed gene loci (Coolen et al., 2010; Frigola et al., 2006). More recently it has emerged that the patterning of certain histone modifications are also perturbed in cancer. For example, global loss of tri-methyl H4K20 (H4K20me3) and acetyl H4K16 (H4K16ac) have been observed in various primary tumours (Fraga et al., 2005). Moreover, global changes in H3K18ac and H3K4me2 have been correlated with increased recurrence risk in prostate cancer (Seligson et al., 2005). Significant progress has been made in understanding how these changes actually contribute to oncogenesis; this has stemmed from specific analysis of the ‘writers’, ‘readers’ and ‘erasers’ of these and other modifications in cancer related cellular processes. Strikingly, analysis of the cancer genome has unveiled an abundance of somatic mutations, gene amplifications, deletions and translocations involving genes encoding these epigenetic regulatory proteins (Vogelstein et al., 2013). Below I will focus on specific examples of changes in histone acetylation and methylation, mediated by disrupted writer, reader and eraser functions and their involvement in transcriptional regulation during oncogenesis.

1.4.1 Histone Acetylation and Cancer

1.4.1.1 Acetylation writers in cancer

Several HAT enzymes are deregulated in human cancers. For example, the viral oncoproteins E1A (adenovirus) and large T-antigen (SV40) primarily target p300, CBP and PCAF, inhibiting their function. Moreover, recurrent somatic mutations in the p300 gene have been observed in up to 17% of cases of primary colorectal, gastric, hepatocellular and breast cancer (Iyer Ng Fau Ozdag et al., 2004). In mouse models, heterozygous deletion of Cbp
leads to hematological cancers, owing to somatic loss of the second \textit{Cbp} allele. \textit{Cbp} and \textit{p300} null mice also develop hematological cancers, suggesting a clear tumour suppressive role for both of these genes. However, chromosomal rearrangements involving these HATs can also have an oncogenic affect. For example, the \textit{CBP} and \textit{p300} genes are both involved in rearrangements involving the \textit{MLL} and \textit{MOZ} genes in leukemia (Krivtsov and Armstrong, 2007). The resultant gene fusion products are thought to promote leukemic transformation due to aberrant histone acetylation, resulting in activation, of oncogenic targets genes. Interestingly, HATs also acetylate certain non-histone proteins, including the cancer relevant transcription factors p53, pRB and E2F. Therefore, deregulated HAT function may also affect gene transcription in cancer indirectly by modulating the activity of these transcription factors on target genes (Wang et al., 2007).

1.4.1.2 Acetylation readers in cancer

Members of the BET family of acetyl lysine readers (\textit{BRD3} and \textit{BRD4}) are commonly genetically altered in certain forms of human cancer (Ryan and Bernstein, 2012; Vogelstein et al., 2013). \textit{BRD3/4} are the most common fusion partners of the \textit{NUT} gene in cases of the invariably fatal NUT midline carcinoma. These translocation events lead to aberrant targeting of NUT to BET target genes and blocked cellular differentiation. Significantly, pharmacological inhibition of BET protein binding to acetyl lysines results in displacement of NUT-BET from chromatin, coupled with down–regulated BET target gene expression, rapid cellular differentiation and proliferative arrest (Filippakopoulos et al., 2010). Suggesting that aberrant binding of the NUT-BET fusion product to acetyl lysines at BET target genes is essential for oncogenesis. Moreover, inhibition of BET binding to acetyl lysines has proven to be generally effective in decreasing cancer cell proliferation. BET inhibition leads to displacement of BRD3/4 and protein complexes required for gene activation from chromatin. Leading to transcriptional repression of BET target genes such as \textit{MYC}, \textit{BCL2} and \textit{CDK6} (Dawson et al., 2011; Delmore et al., 2011; Mertz et al., 2011).
1.4.1.3 Acetylation erasers in cancer
Several oncogenic DNA-binding fusion proteins, such as PML–RARα, PLZF–RARα and AML1–ETO, induce acute promyelocytic leukemia (APL) and acute myeloid leukemia (AML), possibly by blocking myeloid differentiation of immature blood cells (Bolden et al., 2006). Mechanistically, these fusion proteins recruit HDACs to repress the expression of genes, such as C/EBPε, C/EBPβ and HOXA1, which are required for myeloid differentiation as well as anti-proliferative genes such as CDKN1A. BCL6 is a transcription factor that is over-expressed, due to genomic rearrangements, in 40% of diffuse large B-cell lymphomas (DLBCLs) and 5-10% of follicular lymphomas (FLs). BCL6 over-expression leads to blocked plasmacytoid differentiation in vivo and repression of target genes such as CDKN1A through recruitment of HDAC2 (Bolden et al., 2006; Cattoretti et al., 2005). Significantly, in both of the preceding examples, treatment with chemical inhibitors of HDAC activity, either alone or in combination with other regimes, has proven to be therapeutically effective (Esteller, 2008).

1.4.2 Histone Methylation and Cancer
1.4.2.1 Methylation writers in cancer
Many HMT genes are genetically altered in cancer (Ryan and Bernstein, 2012; Vogelstein et al., 2013). The MLL gene, which encodes a HMT specific for H3K4me2/3, was initially identified due to its involvement in recurrent genomic translocations in human myeloid and lymphoid leukemia (Krivtsov and Armstrong, 2007). Genomic rearrangements involving the MLL gene are one of the most common chromosomal abnormalities known in human cancer. Approximately 80% of infant leukemia and 5-10% of adult AML or lymphoid leukemia cases harbour these lesions (Hess, 2004; Krivtsov and Armstrong, 2007). A partial tandem duplication of MLL is the most common form of rearrangement in AML. This causes elevated H3K4me2 levels at, and increased expression of, genes within the HOXA cluster (Dorrance et al., 13
2006, 2008). Genes within this cluster such as \textit{HOXA9}, are normally expressed specifically in immature blood cells and silenced during differentiation. Suggesting that this \textit{MLL} rearrangement may block terminal differentiation of these cells (Krivtsov and Armstrong, 2007). \textit{MLL} fusions are a second \textit{MLL} gene rearrangement seen in leukemia, which result in deletion of a large carboxy terminal portion of the \textit{MLL} protein that includes the SET domain. More than 70 different \textit{MLL} fusion partners have been identified. Mechanistically, the most common \textit{MLL} gene fusions, which include \textit{MLL-AF4} (Guenther et al., 2008; Krivtsov et al., 2008), \textit{MLL-AF9} (Thiel et al., 2010), \textit{MLL-AF10} (Okada et al., 2005) and \textit{MLL-ENL} (Mueller et al., 2007) cause increased H3K79me (a mark associated with transcription activation/elongation) levels at key oncogenic \textit{HOXA} loci through recruitment of the H3K79 HMT DOT1L. Significantly, DOT1L is required for transformation in these \textit{MLL} translocated cancers and specific pharmacological inhibition of DOT1L methyltransferase activity leads to improved survival in mouse xenograph models (Bernt et al., 2011; Daigle et al., 2011).

\subsection*{1.4.2.2 Methylation readers in cancer}

A subset of patients with AML have certain genomic rearrangements that result in the fusion of the nuclear pore complex component gene \textit{NUP98}, with the PHD finger containing genes \textit{PHF23} or \textit{JARID1A}. These fusion gene products are sufficient to induce AML in mouse models, and this requires the ability of the PHD finger to read the H3K4me3 mark (Wang et al., 2009a). Mechanistically, binding of the \textit{NUP98-PHF23} and \textit{NUP98-JARID1A} fusion gene products to H3K4me3 blocks the differentiation of haematopoietic stem and progenitor cells, by preventing the removal of H3K4me3 from developmentally crucial transcription factor genes such as \textit{Hoxa5-11}, \textit{Meis1a}, \textit{Gata3} and \textit{Pbx1}. As a result these genes are locked in an active state and are refractory to Polycomb mediated silencing (Cui et al., 2009; Wang et al., 2009a). The \textit{ING} (\textit{ING1–5}) family of PHD finger containing H3K4me3 readers are potential tumour suppressors, which are often deregulated in cancers. \textit{ING1, 3} and \textit{4} are often found to be downregulated, mutated or deleted in
breast, colorectal, pancreatic, melanoma, glioma, gastric and squamous cell cancers (Chi et al., 2010). *Ing1*−/− mice are predisposed to lymphoma and specific mutations affecting the ability of ING1 to bind H3K4me3 are found in tumours (Coles and Jones, 2009). Suggesting a direct requirement of H3K4me3 reading during ING1 mediated tumour suppression.

### 1.4.2.3 Methylation erasers in cancer

Many HDMs are deregulated at both the transcriptional and genetic level in cancer (Chi et al., 2010; Vogelstein et al., 2013). For example, the H3K4me1/2 demethylase LSD1 represses the expression of genes required for TGFβ signaling, cell growth, migration and invasion in breast cancer cells by removing H3K4me1/2. LSD1 levels are reduced in breast cancer cells, resulting in increased invasiveness and metastatic potential. Owing to de-repression of these target genes (Wang et al., 2009b). LSD1 may also affect gene expression indirectly by acting on p53, inhibiting p53 induced transcriptional activation (Huang et al., 2007). Hence, LSD1 may have both a tumour suppressive and oncogenic role. The *JARID1C* gene which encodes a H3K4me2/3 HDM is recurrently mutated leading to truncation in renal cell carcinoma (RCC). These mutations are correlated with specific transcriptional changes within these tumours (Dalgliesh et al., 2010). The *FBXL10* gene, which encodes a dual specificity HDM for H3K4me3 and H3K36me1/2 is commonly upregulated in T–cell lymphomas (Pfau et al., 2008). Acquired expression of *FBXL10* in cancer is believed to contribute to oncogenesis by directly repressing the *INK4B-ARF-INK4A* tumour suppressor locus by removing H3K36me2/1 and/or H3K4me3, preventing cancer cell senescence (He et al., 2008; Pfau et al., 2008; Tzatsos et al., 2008).

### 1.5 Polycomb group proteins

Polycomb group proteins were originally identified over 30 years ago in *D. melanogaster* as the protein products of genes that are required to prevent inappropriate expression of the homeotic (*Hox*) genes in early development.
(Schwartz and Pirrotta, 2007). In Polycomb mutant flies, Hox genes are expressed outside their normal spatial territories along the anterior–posterior axis (Fig. 1.4). This causes characteristic body patterning mutant phenotypes, which are the hallmark phenotypes that define fly Polycomb proteins. This function is conserved in mammals with several Polycomb gene knockout mouse models exhibiting similar transformation mutant phenotypes (Sparmann and van Lohuizen, 2006). Indeed, a crucial role for Polycombs during development is further underscored by early embryonic lethality in mice after deletion of certain Polycomb genes such as Eed, Ezh2, Suz12 and Ring1B (Margueron and Reinberg, 2011). In addition to a role in early development, recent work has demonstrated that Polycombs are necessary for establishing cell type specific transcriptional programs in, and for subsequently maintaining the correct identities of, adult stem, progenitor and differentiated cell populations (Schwartz and Pirrotta, 2007; Sparmann and van Lohuizen, 2006).
Figure 1.4: Polycombs regulate body patterning and Hox gene expression.

a. Hox gene expression in the Drosophila melanogaster embryo defines the positions of the various appendages along the anterior–posterior axis of the adult body. b. Polycomb proteins maintain the correct spatial and temporal expression pattern of Hox genes through transcriptional repression. Deletion of Polycomb genes causes a de-repression of specific Hox members, which leads to homeotic transformations — transformations of one body segment into the identity of another. The term 'Polycomb' was used to describe the first Polycomb mutants — extra sex combs (Esc) and Polycomb (Pc) — because of additional sex combs on the second and third legs of male flies, a structure that is normally restricted to the first legs. Reproduced from Sparmann and van Lohuizen, Nat Rev Cancer 6, 846-856.

1.5.1 Polycomb Repressive Complexes

At a molecular level, Polycombs have been characterized into two main protein complexes, known as Polycomb Repressive Complex 1 (PRC1) and PRC2 (Simon and Kingston, 2009). In D. melanogaster PRC2 contains the three core components Enhancer of zeste (E(Z)), Suppressor of zeste 12
(SU(Z)12) and Extra sex combs (ESC), in addition to the sub-stoichiometric components, Nucleosome remodeling factor 55 (NURF55) and Polycomb–like (PCL). PRC1 contains at least four core components, including Polycomb (PC), Posterior Sex Combs (PSC), Sex Combs Extra (SCE/RING) and Polyhomeotic (PH) (Fig. 1.5a). However, the concept of two static Polycomb complexes, particularly in the mammalian system, is an over simplification. Apparent gene duplication events during evolution have resulted in the mammalian genome encoding multiple orthologues of many *D. melanogaster* Polycomb proteins. In mammals the core PRC2 consists of one, or possibly both, of the E(Z) orthologues EZH1 and EZH2, the SU(Z)12 orthologue SUZ12 and one of the four isoforms of the ESC orthologue, EED. In addition the complex may contain the sub-stochiometric components RBBP4 and RBBP7 (NURF55 orthologues), PCL1–3 (PCL orthologues), JARID2 and AEBP2 (Sauvageau and Sauvageau, 2010). The canonical mammalian PRC1 complex is comprised of various combinations of the same four core Polycomb components as that in flies, including a PC protein (CBX2, 4, 6–8); a PSC protein (PCGF1–6); a SCE/RING protein (RING1A and RING1B) and a PH protein (HPH1–3). The exact configurations and combinations of Polycomb components within PRC1 and PRC2 are incompletely characterized, however recent evidence indicates that the complexes vary in a cell type and context dependent manner (Gao et al., 2012; Maertens et al., 2009; Margueron and Reinberg, 2011; Morey et al., 2012; O’Loghlen et al., 2012).

EZH1/2 are the catalytically active components of PRC2 (Cao et al., 2002; Kuzmichev et al., 2002; Margueron et al., 2008). Together with EED and SUZ12, they tri–methylate the N-terminal tail of histone H3 at lysine 27 (H3K27me3). This modification acts as docking site for the recruitment of the chromodomain containing CBX proteins (PC orthologues) within canonical PRC1 complexes. However, alternative recruitment pathways must exist for PRC1, since non-canonical complexes, which lack CBX components, have recently been reported (Fig. 1.5a) (Gao et al., 2012; Tavares et al., 2012). PRC1 mediates the mono-ubiquitylation of histone H2A at lysine 119.
(H2AK119ub) through the enzymatic activity of the RING1A and RING1B subunits and this is believed to be important for target gene repression, potentially by mediating chromatin compaction (Fig. 1.5b) (Francis et al., 2004; De Napoles et al., 2004; Shao et al., 1999; Wang et al., 2004).

Figure 1.5: Polycomb group proteins form distinct multi-protein complexes that collaborate to repress gene expression.

a. Polycomb group proteins form 2 major multi-protein complexes, the Polycomb Repressive Complex 2 (PRC2) (Left) and PRC1 (Right). PRC2 is composed of the core subunits EZH1/2, SUZ12 and EED, these core components further associate with additional sub-stiochiometric components such as RBBP4, RBBP7, AEBP2, JARID2, PCL1, PCL2 and PCL3. The canonical PRC1 is composed of different combinations of at least 4 core subunits, including a PC; PCGF; RING and a PHG protein. The existence of a non-canonical form of PRC1, which contains an addition subunit; either RYBP or YAF2, has recently been reported in mammalian systems. These complexes also lack a Pc (CBX) and Phc component. Significant gene duplication events during evolution from D. melanogaster to mammals, particularly amongst genes encoding PRC1 components, has resulted in diversification of these complexes in mammals. b. PRC1 and 2 collaborate to repress gene expression. PRC2 is responsible for catalyzing the tri-methylation of lysine 27 on histone H3 (H3K27me3). Canonical PRC1 complexes are thought to be recruited to Polycomb target genes, at least in part, by the affinity of the chromodomains of the CBX (CBX2/4/6/7/8) proteins within the complex for this methyl mark. Following recruitment of PRC1, the E3-ubiquitin ligase proteins, RING1A and RING1B, monoubiquitylate lysine 119 of histone H2A (H2AK119ub) and this is believed to be important for chromatin compaction and transcriptional repression.
1.5.2 Polycomb target genes

It was postulated more than 20 years ago that Polycombs likely regulate more than just the Hox genes, as Polycombs accumulate on more than one hundred additional sites on polytene chromosomes in D. melanogaster (DeCamillis et al., 1992; Zink and Paro, 1989). The advent of genome-wide chromatin immunoprecipitation (ChIP) techniques, such as ChIP-on-chip and ChIP-seq, allowed researchers to identify the downstream regulatory pathways controlled by Polycombs and has provided important insights into how they regulate developmental processes (Boyer et al., 2006; Bracken, 2006; Bracken and Helin, 2009; Lee et al., 2006; Tolhuis et al., 2006). These seminal studies revealed that in addition to their canonical target genes, the Hox genes, Polycombs regulate about 10% of all genes in any given cell type, and importantly these genes encode proteins with key roles in developmental and differentiation processes (Bracken and Helin, 2009). For example, Polycombs bind the promoters and repress the transcription of genes encoding members of the WNT, FGF, Hedgehog, TGF-β and Notch signalling pathways, in addition to those encoding many developmental essential transcription factors. Significantly, many homologs of these target genes were also identified as Polycomb target genes in D. melanogaster, suggesting that Polycomb function has been generally conserved from flies to mammals (Schwartz and Pirrotta, 2007). The initial studies revealed that Polycombs are displaced from the promoters of lineage commitment genes during differentiation (Boyer et al., 2006; Bracken, 2006; Lee et al., 2006; Nègre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). However, more recent studies have shown that Polycombs are also recruited to a subset of target genes during differentiation (Mohn et al., 2008; Pasini et al., 2007). Based on these observations, we and others, have proposed a model for Polycomb function, whereby they act dynamically to silence genes of alternative cell fates during development and differentiation (Fig. 1.6) (Bracken and Helin, 2009; Brien and Bracken, 2009).
Figure 1.6: Polycomb function dynamically during cell fate decisions.

In stem cells, Polycombs bind and repress genes required for lineage commitment, while transcriptional activators maintain stem cell specific genes in an active transcriptional state. During cell fate transitions specific signals instruct the stem cell to differentiate, and Polycombs become displaced from a subset of target genes required for the specification of the respective cell lineage. This coincides with their transcriptional activation. Furthermore, Polycombs are also recruited to the promoters of genes required for maintaining stem cell identity. It is unclear how Polycombs are displaced and recruited to target genes during development. However, emerging evidence suggests that transcription factors and long non-coding RNA molecules may be important in mediating this dynamic regulation.

1.5.3 Polycombs and cancer

The genes encoding members of the Polycomb family, particularly members of PRC2 complex, are commonly subject to genetic aberration in cancer. The \( \text{EZH2} \) gene is frequently over-expressed due to amplification in human cancers and this is associated with poor patient outcome (Bracken and Helin, 2009; Bracken et al., 2003; Varambally et al., 2002). Importantly, suppression of \( \text{EZH2} \) expression, either by pharmacological means or by RNA interference, decreases cancer cell growth both \textit{in vitro} and \textit{in vivo} (Bracken et al., 2003; Tan et al., 2007; Wilson et al., 2010). Moreover, \( \text{EZH2} \) is also subject to recurrent somatic change-of-function mutations in lymphoma. These mutations, involving the Y641 and A677 residues within the SET domain of \( \text{EZH2} \) have been observed in 22% DLBCL and 7% of FL (Morin et al., 2010). The mutations enable the \( \text{EZH2} \) molecule to more efficiently catalyze the tri-methylation of H3K27 (Sneeringer et al., 2010). \( \text{EZH2} \) mutant
lymphoma cell lines exhibit a global increase in H3K27me3 levels with a concomitant reduction in H3K27me1 levels. Significantly, highly specific pharmacological suppression of EZH2 activity in these cell lines reduces global H3K27me3 levels and inhibits cell proliferation, both in vitro and in vivo (Knutson et al., 2012; McCabe et al., 2012). Therefore, inhibition of EZH2 activity may provide a promising epigenetic therapeutic strategy in lymphoma. These data suggest that unchecked PRC2 activity has a direct oncogenic affect. Possibly owing to silencing of tumour suppressor genes, such as INK4A–BARF–INK4A (Bracken et al., 2007), CDKN1C (Yang et al., 2009), CDKN1B (Ougolkov et al., 2008), CDH1 (Cao et al., 2008; Fujii and Ochiai, 2008), BRCA1 (Gonzalez et al., 2009) and ADRB2 (Yu et al., 2007). Consistent with this, inactivating somatic mutations of the ubiquitously expressed H3K27me2/3 demethylase UTX, have been reported in many types of human cancer, including multiple myeloma (10%), esophageal cancer (8%), renal cancer (1.4%) and transitional cell bladder cancer (20%) (Gui et al., 2011; Van Haaften et al., 2009). Importantly, restoring wildtype UTX in cancer cells with mutated UTX reduced H3K27me3 levels at specific sites and slowed cancer cell proliferation (Van Haaften et al., 2009). These data suggest that UTX may have a direct tumour suppressive role in these cancers. Furthermore, the expression of the related H3K27me2/3 demethylase, JMJD3, is reduced in several cancers, and JMJD3 activates the INK4B–ARF–INK4A tumour suppressor locus during oncogene-induced senescence (Agger et al., 2009; Barradas et al., 2009). Taken together these observations suggest a simple model where increased PRC2 activity contributes directly to oncogenesis.

However, the situation seems to be more complicated than this, as several independent studies have recently reported that the EZH2 gene is subject to inactivating mutations as well as homo/hetero-zygous deletions in malignant myeloid disorders. These events have been seen in myelodysplastic syndromes (MDS) (6%) and myeloproliferative disorders (MPD) (12%) (Ernst et al., 2010; Nikoloski et al., 2010). Furthermore, the genes encoding the other PRC2 core components, SUZ12 and EED, in addition to the sub-
stiochiometric component JARID2, have been reported to be mutated/deleted, albeit to a lesser extent, in MDS (Score et al., 2012). The biological significance of these observations have yet to be established; however, collectively they demonstrate that loss of PRC2 activity is likely to contribute to oncogenesis in the context of myeloid malignancies. Moreover, deletion of Ezh2 in mouse HSCs is sufficient to cause an aggressive T-cell acute lymphoplastic leukemia (T-ALL) in mice. These leukemic cells exhibit markedly reduced H3K27me3 levels, demonstrating that PRC2 activity is compromised in these T-ALLs (Simon et al., 2012). Significantly, mutations have been reported in the EZH2 (18%) and SUZ12 (7%) genes in human T-ALL, and this correlates with diminished PRC2 activity (Ntziachristos et al., 2012). Furthermore, in pediatric T-ALL, the EZH2, SUZ12 or EED genes are mutated/deleted in 42% of cases (Zhang et al., 2012). Taken together, these observations suggest that in certain cancers of the blood lineages PRC2 activity actually has a tumour suppressive role. Collectively, the genetic data implicating PRC2 and UTX activities in cancer strongly points to a pivotal importance for H3K27me3 in cancer.

In line with this assertion, two recent independent genome-wide studies have reported recurrent somatic mutations in histone H3 genes affecting H3K27 (Schwartzentruber et al., 2012; Wu et al., 2012). Schwartzentruber et al., discovered that in 19% of pediatric glioblastomas (GBM) the H3F3A gene, which encodes the histone H3 variant H3.3, is mutated resulting in the lysine at position 27 being changed to a methionine (K27M). Remarkably, Wu et al., found the same K27M mutation to be present in H3F3A in 60% of diffuse intrinsic pontine gliomas (DIPG). This mutation was also found to be present in the HIST1H3B gene, which encodes the canonical histone H3 protein H3.1, in 18% of DIPG cases (Wu et al., 2012). These observations directly link H3K27 with human cancer, and evidence suggests they inhibit EZH2 methyltransferase activity (Lewis et al., 2013). However, further study will be required to fully understand how these mutations promote oncogenesis.
2 Materials and Methods
2.1 Cell Culture
All cells were kept in a humidified incubator at 37°C, 5% CO₂. Cells were routinely grown to 80% confluence prior to passaging in order to maintain exponential cell growth and to avoid undue cellular stress due to excessive levels of confluency.

2.2 Fibroblast cell culture
Mouse embryonic fibroblasts (MEFs) derived from embryonic day 13.5 C57BL6 mouse embryos, primary human diploid fibroblasts (HDFs) and hTERT immortalized human embryonic fibroblasts (TIG3T) were maintained in DMEM media supplemented with 10% (v/v) FBS (Hyclone), 100U/ml penicillin and 100U/ml streptomycin (Gibco).

2.3 NT2/D1 cell culture
NT2/D1 cells were grown in DMEM supplemented with 10% (v/v) FBS (Hyclone), 100U/ml penicillin and 100U/ml streptomycin. To induce neuronal differentiation of NT2/D1 cells, asynchronously growing cells were seeded at 20% confluence and treated 24 hrs later with 10μM all-trans retinoic acid (ATRA) (Sigma). Cultures were treated continuously with ATRA during the differentiation procedure.

2.4 Embryonic stem cell culture
Mouse embryonic stem (ES) cells were maintained on gelatinized culture dishes in GMEM media (Sigma) supplemented with 10% FBS (Gibco) (v/v), 1000U/ml leukemia inhibitory factor (Millipore), 100 U/ml penicillin, 100 U/ml streptomycin (Gibco), 1:100 GlutaMAX (Gibco), 1:100 non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco) and 50 μM β-mercaptoethanol (Sigma). To induce differentiation of ES cells to embryoid bodies (EBs), cells were plated at 10x10⁵ cells/ml in non-adherent bacterial dishes (Greiner) in ESC media without LIF. Media was changed every 2-3 days throughout the differentiation procedure. ESCs were also differentiated in monolayer cultures using media containing 1 μM ATRA without LIF, for 48 hrs. Media was changed after 24 hrs.
2.5 HMEC cell culture

Human Mammary Epithelial cells were maintained essentially as described. Briefly, low passage finite lifespan HMEC lines obtained from reduction mammoplasty tissue were maintained in M87A media (Garbe et al., 2009). M87A medium is a mix of 50% MM4 media, composed of DMEM/F12 media (Gibco), 10ug/ml insulin (Sigma), 10nM tri-iodothyronine (Sigma), 1nM β-estradiol (Sigma), 0.1ug/ml hydrocortisone (Sigma), 0.5% FBS (Gibco), 5ng/ml epidermal growth factor (Peprotech), 2mM Glutamine (Lonza) and 1ng/ml cholera toxin (Sigma) and 50% MCDB170, composed of MEGM media (Lonza), 5ug/ml transferrin (Lonza), 10^{-5} M isoproterenol (Sigma) and 2mM Glutamine (Lonza). M87A media was further supplemented with 0.1nM oxytocin (Bachem) and 0.1% AlbuMAX I (Invitrogen).

2.6 Inducible GAL4-PCL3 HEK293 cell line culture

The inducible GAL4–PCL3 cell line was generated by co-transfecting a tetracycline inducible GAL4 fusion vector (pcDNA5-FRT-TO-GAL4-PHF19) with a Flp-recombinase vector (pOG44) into Flp-In T-Rex 293 cells using the calcium phosphate method (Chen and Okayama, 1987). This co-transfection results in the stable integration of the GAL4-PCL3 fusion construct into a defined locus in the Flp-In T-Rex 293 cell genome. This particular Flp-In T-Rex 293 cell line was described previously and contained a stably integrated luciferase reporter gene downstream of 5XGAL4 DNA binding motifs (Hansen et al., 2008). These cells were maintained in the DMEM medium supplemented with 10% tetracycline-free FBS (Hyclone), 100 U/ml penicillin and 100 U/ml streptomycin (Gibco), blasticidin (300 μg ml^{-1}) to select for the Tet-repressor, neomycin (400 μg ml^{-1}) to select for the 5xGAL4TK–luciferase reporter and hygromycin B (300 μg ml^{-1}) to select for the GAL4–fusion construct. To induce GAL4-fusion protein expression, tetracycline (1 μg ml^{-1}) was added for 48hrs.

2.7 Retroviral transduction

Retroviral particles were produced in ecotropic or amphotrophic HEK293 cells maintained in DMEM media containing 10% FBS (Hyclone), 100U/ml penicillin
and 100U/ml streptomycin (Gibco). These cells were grown to ~60-70% confluence and transfected with a either with a pBABE ORF expression vector, or a pRETROSUPER shRNA expression vector, using the calcium-phosphate method. Media containing viral particles was collected at 48 and 72 hrs post transfection, filtered through a 0.45\mu m filter, and used directly for transduction of target cells after addition of 5\mu g/ml Polybrene (Sigma). Target cells were transduced with viral particles for 4-8 hours on 2 consecutive days and subsequently selected in media containing 2ug/ml Puromycin (Sigma) for 48–72 hours immediately after transduction.

2.7.1 Retroviral shRNA target sequences

shBM1 - 5' -AAGGAATGGTCCACTTCCATT-3' (Bracken et al., 2007).
shEZH2 - 5' - AAGACTCTGAATGCAGTTGCT-3' (Bracken et al., 2007).
shPCL1 - 5' - GATGTGCTGGCCAGATGGA-3'
shPCL2 - 5' - GGACCAAATGCCAAAGCAT-3'
shPCL3 - 5' - CCTCGTGACTTTCGAAGAT-3'

2.8 Lentiviral transduction

Lentiviral particles were produced by transfection of HEK293T cells maintained in DMEM media containing 10% FBS (Hyclone) and 100U/ml penicillin and 100U/ml streptomycin (Gibco). HEK293T cells were grown to ~80% confluence and transfected with a lentiviral vector, and vectors encoding viral packaging (pPAX8) and envelope (pVSVG) proteins using the calcium phosphate method. Media containing viral particles was collected 24 and 48 hours after transfection and filtered through a 0.45\mu M syringe filter. Viral particles were purified from these supernatants by ultracentrifugation at 20,000 rpm for 2 hours at 4°C. Viral particles were resuspended in sterile PBS (Gibco) and added directly to target cell media supplemented with 10\mu g/ml Polybrene (Sigma). ESCs and HMECs were transduced with viral particles for 16-20 hours, washed and selected in media containing 2\mu g/ml Puromycin (Sigma) for 24–36 hours immediately after transduction.

2.8.1 Lentiviral shRNA target sequences

shScr 5' - CCTAAGGTTAAGTCGCCCTCG-3'
shBmi1 5′–CCTAAGGAAGGATGAATGAT–3′
shEzh2 5′–CCGCAGAAGAACTGAAAGAAA–3′
shPcl1 5′–GCTGGGATGTACCTGGAGAGA–3′
shPcl2 5′–CCGTTACAGTGGGTAGATATA–3′
shPcl3 1. 5′–CCTAGCCAGTATATTCTGACTT–3′
shPcl3 2. 5′–CCTCAAGTCCTCTATCACCAA–3′
shNo66 1. 5′–TGTCACCTACCAGGCAATA–3′
shNo66 2. 5′–CCGAGACTTCATGGATTACAT–3′

2.9 3T3 assays and BrdU FACs

3T3 growth assays were conducted as follows, 1x10^6 cells were plated on 100mm plates, 3 days later, the total number of cells was counted and 1x10^6 cells were plated again. The cumulative increase in cell number was calculated according to the formula ‘Log(Nf/Ni)/Log2’ where Ni and Nf are the initial and final numbers of cells plated and counted after 3 days, respectively. For BrdU FACs analyses, cells were pulsed with 33μM BrdU for 45mins. BrdU incorporation was measured by staining with an anti-BrdU antibody followed by FACs analysis. DNA content was measured by propidium iodide staining.

2.10 Cloning, plasmid generation and mutagenesis

2.10.1 Gateway Cloning

Gateway cloning technology is a proprietary cloning system developed by Invitrogen. It allows for the efficient sub-cloning of DNA fragments from a Gateway ‘entry clone’ into a multitude of compatible ‘destination vectors’ to generate an expression clone. Importantly, this sub-cloning method does not introduce sequence changes and maintains the DNA fragment in the correct reading frame and orientation. The system utilizes complimentary attL1/attL2 and attR1/attR2 recombination sequences flanking the DNA fragment. Combining an entry clone and destination vector in the presence of the LR clonase enzyme induces recombination, yielding an expression clone containing the DNA fragment of interest (Fig 2.1).
Addition of the LR clonase enzyme to an attL1/L2 entry clone and attR1/R2 destination vector induces a specific recombination event. This yields two products. 1. an attB1/B2 expression clone and an attP1/P2 donor vector. Importantly, this donor vector cannot be propagated in standard bacteria due to the presence of the ccdB DNA gyrase gene. Therefore, transformation will yield only expression clones containing the DNA fragment of interest.

2.10.2 ORF cloning and sub-cloning

The human full length PCL1–3, PHF14 and C/EBPZ ORF sequences were PCR amplified from cDNA generated from low passage human mammary epithelial cells. The amplified ORFs were inserted into the pCR8/GW/TOPO Gateway cloning™ entry vector (Invitrogen) in accordance with the manufacturers instructions. ORF integrity was subsequently confirmed by DNA sequencing. The sequence verified ORFs were then sub-cloned into Gateway cloning™ compatible expression vectors by recombination using the LR clonase enzyme (Invitrogen).

2.10.3 ORF mutagenesis

Full-length PCL3 pCR8/GW/TOPO and PCL3-tudor (residues 25–95) pGEX-6P1 W50C/Y56A mutant plasmids were generated from wildtype PCL3 pCR8/GW/TOPO and PCL3-tudor (residues 25–95) pGEX-6P1 plasmids, respectively. This was accomplished by PCR mutagenesis using the GeneArt™ Site Directed Mutagenesis system (Invitrogen) in accordance with the manufacturers instructions. Insertion of mutations was confirmed by DNA sequencing. In the case of full-length PCL3 W50C/Y56A the sequence verified ORF was sub-cloned into Gateway cloning™ compatible expression vectors by recombination using the LR clonase enzyme (Invitrogen).
2.11 Real-Time Quantitative PCR

Total RNA was extracted from cells using the RNeasy kit (Qiagen) according to manufacturer's instructions. 1–2μg of total RNA was used to generate cDNA by reverse transcriptase PCR using the PE Applied Biosystems TaqMan Reverse Transcription kit. Relative mRNA expression levels were determined using the SYBR Green I detection chemistry (Applied Biosystems), on the ABI Prism 7500 Fast Real-Time PCR System. The ribosomal constituent RPLPO was used as a control gene for normalization.

2.12 Immunoprecipitations

2.12.1 Immunoprecipitations of exogenous FLAG tagged proteins

Nuclear pellets prepared from retrovirally infected HEK293T and HMEC cells were lysed in low salt buffer containing protease inhibitors (150mM NaCl, 50mM TRIS pH 8.0, 1mM EDTA, 1% NP40, 1μg/ml aprotinin, 10μg/ml leupeptin, 1mM PMSF). Nuclear protein lysates were then subjected to mechanical homogenization by douncing and sonication. The detergent concentration was adjusted to 0.5% by dilution of the lysate with an equal volume of low salt buffer containing protease inhibitors without NP40. Homogenized lysates were pre-cleared by centrifugation at 4°C, 14000rpm for 30 mins. An additional pre-clearing step was performed using 20–60μls of mouse IgG agarose (Sigma) for 1hr at 4°C. Immunoprecipitations of exogenously expressed FLAG tagged proteins were performed using M2 anti-FLAG agarose (Sigma). All IPs were performed overnight at 4°C. Beads were washed extensively using 0.1% NP40 and finally with low salt buffer lacking detergent. Elution of FLAG tagged proteins was performed at 4°C using 250μg/ml 3xFLAG peptide (Sigma) in 0.05% NP40 with horizontal shaking for 30 mins. Eluted protein fractions were separated by SDS–PAGE and analysed by western blotting or liquid chromatography mass spectrometry (LC/MS).

2.12.2 Immunoprecipitation of endogenous proteins

Endogenous immunoprecipitations were performed essentially as previously described (Van Den Berg et al., 2010). Briefly, nuclear pellets prepared from
HDFs, ESCs or NT2/D1 cells were lysed in Buffer C containing protease inhibitors (20 mM HEPES (pH 7.6), 20% (v/v) glycerol, 0.42 M NaCl, 1.5 mM MgCl$_2$, 0.2 mM EDTA, aprotinin 1μg/ml, leupeptin 10μg/ml, PMSF 1mM) and nuclear extracts were dialysed against Buffer C-100 (20 mM HEPES (pH 7.6), 20% (v/v) glycerol, 0.2 mM EDTA, 100 mM KCl, 1.5 mM MgCl$_2$, 0.2 mM EDTA) for 5hrs. For each individual immunoprecipitation, a total of 5–10μgs of antibody was cross-linked to 20–50μls packed volume of protein A Sepharose (Sigma). Antibodies were subsequently covalently coupled to beads using dimethyl pimelimidate (DMP). Coupled beads were equilibrated in Buffer C-100 containing 0.02% NP40 (C-100*), 0.1mg/ml insulin (Sigma), 0.2mg/ml chicken egg albumin (Sigma) and 1% fish skin gelatin (Sigma) and added to 1ml of dialysed nuclear extract containing 250 units of benzonase (Company) for 3 hours at 4°C in “No Stick” microcentrifuge tubes (Eppendorf). The beads were then washed 5 times with C-100* and elutions performed with 2XSDS-loading dye. Eluted protein fractions were separated by SDS-PAGE and analysed by Western blot or liquid chromatography mass spectrometry.

2.13 Western Blotting
Whole cell or nuclear protein samples were prepared by lysing whole/nuclear cell pellets in RIPA buffer containing protease inhibitors (25 mM TrisHCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, aprotinin 1μg/ml, leupeptin 10μg/ml, PMSF 1mM). Protein lysates were separated on SDS-PAGE gels and transferred to nitrocellulose membranes. Membranes were subsequently probed using the relevant primary and secondary antibodies and relative protein levels detected by chemiluminescence.

2.14 Mass Spectroscopy preparation and analysis
Immunoprecipitated samples were separated on 4–12% gradient SDS-PAGE gels (Novex) and stained with GelCode Blue solution (Thermo Scientific). Lanes were excised and diced into small (1mm$^2$) pieces and washed three times with 25mM NH$_4$HCO$_3$/50% acetone. Gel pieces were dehydrated and incubated with 10mM DTT in 25mM NH$_4$HCO$_3$ for 1 hour at 56°C. The supernatant was removed and the gel pieces were incubated with 55mM
iodoacetamide for 60 minutes. Gel pieces were washed with 25 mM NH₄HCO₃, then 25 mM NH₄HCO₃/50% acetone and dehydrated. 10 ng/µl trypsin in 25 mM NH₄HCO₃ was added to the gel pieces and incubated overnight at 37°C. Finally, peptides were extracted from the gel pieces with 50% acetone/5% formic acid and the solvent evaporated. The final peptide sample was resuspended in 20 µl 0.1% formic acid.

Samples were analyzed on a Thermo Scientific LTQ Orbitrap mass spectrometer equipped with a nanoACQUITY UPLC (Waters) chromatography system and a nanoelectrospray source. Each sample was injected onto a nanoACQUITY Symmetry C18 trap (5 µm particle size, 180 µm x 20 mm) in buffer A (0.1% formic acid in water) at a flow rate of 4 µl/min and then separated over a nanoACQUITY BEH C18 analytical column (1.7 µm particle size, 100 µm x 100 mm) over one hour with a gradient from 2% to 25% buffer B (99.9% acetone/0.1% formic acid) at a flow rate of 0.4 µl/min. The mass spectrometer continuously collected data in a data-dependent manner, collecting a survey scan in the Orbitrap mass analyzer at 60,000 resolution with an automatic gain control (AGC) target of 1 × 10⁶ followed by collision-induced dissociation (CID) MS/MS scans of the 10 most abundant ions in the survey scan in the ion trap with an AGC target of 5,000, a signal threshold of 1,000, a 2.0 Da isolation width, and 30 ms activation time at 35% normalized collision energy. Charge state screening was employed to reject unassigned or 1+ charge states. Dynamic exclusion was enabled to ignore masses for 30 s that had been previously selected for fragmentation.

Raw files were processed using version 1.1.36 of MaxQuant. For protein identification the ipi.HUMAN protein database was combined with the reversed sequences and sequences of widespread contaminants, such as human keratins. Carbamidomethylation was set as fixed modification. Variable modifications were oxidation (M) and N-acetyl (protein). Initial peptide mass tolerance was set to 20 ppm and fragment mass tolerance was set to 0.5 Da. Two missed cleavages were allowed and the minimal length required for a peptide was six amino acids. The peptide and protein false
discovery rates (FDR) were set to 0.01. The maximal posterior error probability (PEP), which is the probability of each peptide to be a false hit considering identification score and peptide length, was set to 0.01. Proteins identified in two of three experimental data sets were accepted. Tentative identifications with only one unique peptide, or two (or more) unique peptides in only one experimental data set, were manually validated considering the assignment of major peaks, occurrence of uninterrupted y- or b-ion series of at least 3 consecutive amino acids, preferred cleavages N-terminal to proline bonds, the possible presence of a2/b2 ion pairs and mass accuracy. The ProteinProspector MS-Product program was used to calculate the theoretical masses of fragments of identified peptides for manual validation.

2.15 Purification of recombinant GST tagged protein
The wildtype and W50C/Y56A mutant PCL3--tudor (aa 25-95) pGEX-6P1 plasmids were transformed into the E. coli strain BL21-DE3. BL21-DE3 were grown in 500ml cultures at 37°C for 3−4hrs and protein expression was induced with 0.5 mM IPTG, cultures were grown in the presence of IPTG for a further 16hrs at 20°C. GST-PCL3 fusion proteins were purified over GSH-agarose beads (Pierce).

2.16 In vitro peptide binding assays
Biotinylated histone H3 peptides (1μg) either unmodified or modified at the indicated lysine residues were incubated with 5 μg of GST-PCL3 (aa 25-95) in binding buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 20% glycerol, 0.1% NP-40, 1 mM PMSF) overnight at 4°C. Strepavidin agarose beads (Invitrogen) were added, and samples were incubated for a further 60 mins at 4°C. Beads were washed extensively in binding buffer, and bound protein was eluted using 2×Laemilli dye. Eluted protein was analysed by Western blot.

2.17 Surface Plasmon Resonance
SPR experiments were performed at 25°C using a series S sensor chip SA with a BiaCore T200 surface plasmon resonance instrument (GE Healthcare). All experiments were performed with in HBS-P running buffer [10mM Hepes, 150mM NaCl, 0.05% surfactant P20]. Biotinylated H3, H3K4me3, H3K27me3
and H3K36me3 peptides were diluted in running buffer and then immobilised to a density of 600-770 RU. GST-PCL3 fusion-protein (at concentrations from 50 to 800nM) was injected onto the chip surface for 180 s at a flow rate of 20 \( \mu l/min \). The dissociation phase was monitored for up to 600 s. Individual sensorgrams were double-referenced against injection onto an empty flow cell and GST alone injections at equivalent concentrations. Data were fitted to a 1:1 Langmuir model using Biaevaluate analysis software. The observed results and apparent \( K_d \) values were highly reproducible in replicate experiments.

### 2.18 PCL3 Antibody Generation

PCL3 antibodies were generated in rabbits by intravenous injection of 2 synthetic peptides corresponding to amino acids 362-376 (NSASSELKRKGSKSP +C) and 503-518 (SAEGASVPERPDEGID) of full-length PCL3. These peptides were C terminally conjugated to KLH and subcutaneously injected into rabbits in accordance with standard procedures (Eurogentec). Whole rabbit serum was subsequently affinity purified on both peptide antigens.

### 2.19 Chromatin Immunoprecipitation

Cells for Chromatin Immunoprecipitation (ChIP) analysis were chemically cross-linked using 1% formaldehyde in PBS. This cross-linking reaction was performed for 10 min at room temperature and stopped by addition of glycine to a final concentration of 0.125 M, followed by an additional 5 min incubation. Fixed cells were then washed twice with PBS and lysed in SDS lysis buffer containing protease inhibitors (50 mM Tris pH 8.1, 0.5% SDS, 100 mM NaCl, 5 mM EDTA, 1\( \mu g/ml \) aprotinin, 10\( \mu g/ml \) leupeptin, 1mM PMSF) yielding intact nuclei. Harvested nuclei were quick frozen and stored at -20°C until required. Frozen nuclei were thawed, pelleted by centrifugation and resuspended in ice-cold ChIP Buffer containing protease inhibitors (2:1 mix of SDS lysis buffer (as above) and Triton dilution buffer (100mM Tris pH 8.6, 100mM NaCl, 5mM EDTA pH 8.0, 5% Triton-X 100, 1\( \mu g/ml \) aprotinin, 10\( \mu g/ml \) leupeptin, 1mM PMSF). Chromatin was sheared into fragments of approximately 200-1000bps
by sonication using a Sanyo Soniprep sonicator. For each ChIP replicate, the lysate was pre-cleared by addition of 15μl (30mg/ml) of either protein A or G Dynalbeads (Invitrogen), blocked with 0.5 mg/mL lipid-free BSA (Sigma) and 0.2 mg/mL herring sperm DNA (Sigma) in TE. For the purposes of ChIP-SEQ herring sperm DNA was omitted from this stage. Samples were then centrifuged at 14000rpm at 4°C for 30mins to remove precipitates. The supernatants were immunoprecipitated overnight at 4°C with continuous rotation. The next morning, ChIP replicates were centrifuged at 14000rpm at 4°C for 30 mins to remove any precipitates which formed overnight. 50μL of blocked protein A or G beads were then added to ChIP samples and incubated for 4 hours at 4°C. The beads were extensively washed using 3 x 1ml of Mixed Micelle Wash Buffer (20 mM Tris at pH 8.1, 150 mM NaCl, 5 mM EDTA, 5% w/v sucrose, 1% Triton X-100, and 0.2% SDS), 2 x 1ml of Buffer 500 (50 mM HEPES at pH 7.5, 0.1% w/v deoxycholic acid, 1% Triton X-100, 500 mM NaCl, and 1 mM EDTA), 2 x 1ml of LiCl Detergent Wash Buffer (10 mM Tris at pH 8.0, 0.5% deoxycholic acid, 0.5% NP-40, 250 mM LiCl, and 1 mM EDTA) and 1 x 1ml of TE (pH 8.0). The immune complexes were then eluted in ChIP elution buffer (1% SDS, 0.1M NaHCO3) at 65°C for 2 hrs with continuous agitation. The eluted material was purified by phenol/chloroform-extraction and ethanol-precipitation and the DNA resuspended in nuclease-free water (Sigma) for downstream analysis.

2.20 ChIP–Seq library preparation and analysis

For ChIP-Seq, DNA from 10 independent ChIP experiments was pooled and quantified using a Qubit fluorometer (Invitrogen). Sequencing libraries were generated using 100ng of immunoprecipitated DNA using the ChIP-Seq Sample Prep Kit (Illumina). Amplified library DNA was purified by gel isolation and quality checked to ensure the absence of adaptor dimer contamination using the Bioanalyzer 2100 and DNA High Sensitivity Chip assay (Agilent). DNA libraries were quantified and diluted to 10 pM. Diluted libraries were used directly for cluster generation and sequencing analysis using the Genome Analyser II (Illumina) following the protocol of the manufacturer. Base calling and mapping to the mouse genome (mm8, Feb 2006 release) of the 42-bp
sequences were done using the Bowtie alignment tool allowing for up to 2 mismatches in each read. To avoid any PCR bias only two reads per chromosomal position were allowed, thus eliminating spurious spikes. Peak detection was performed using MACS, and Input DNA was used as a control for normalization.
3 The Polycomb PCL3 binds H3K36me3 to recruit PRC2 and the demethylase NO66 to embryonic stem cell genes during differentiation
3.1 Introduction

The mechanism(s) by which the PRC2 complex is recruited to active genes during lineage specification remain unclear (Bracken and Helin, 2009; Margueron and Reinberg, 2011). The *D. melanogaster* PCL (Polycomb–like) protein is a sub-stoichiometric member of the PRC2 complex and has been reported to be required for the recruitment of the complex to target genes (Nekrasov et al., 2007; Sauvageau and Sauvageau, 2010; Savla et al., 2008; Tie et al., 2003). The three mammalian orthologues of *D. melanogaster* PCL (PCL1, PCL2 and PCL3) also associate with the PRC2 complex and have been implicated in the recruitment and maintenance of the complex on target genes (Boulay et al., 2011; Casanova et al., 2011; Hunkapiller et al., 2012; Li et al., 2011; Sarma et al., 2008; Walker et al., 2010). Like their *D. melanogaster* homologue, all three mammalian PCL proteins lack any known DNA binding domains, suggesting that their association with chromatin is not mediated by a direct association with DNA. All of these proteins contain a conserved domain structure consisting of a single N-terminal tudor domain and tandem PHD finger domains. These domain structures have previously been shown to bind to, or ‘read’, post–translationally modified histone tails and have been demonstrated to be functionally important both for protein targeting to chromatin and for transcription regulation (Chan et al., 2009; Wysocka et al., 2006). This highlights the possibility that the PCL proteins may contribute to PRC2 recruitment through similar histone reading mechanisms.

The chromatin landscape at actively transcribed genes is marked by the histone modifications H3K4me3 and H3K36me3 (Kouzarides, 2007). The H3K4me3 mark is tightly localized around the transcription start site (TSS), while the H3K36me3 modification is localized downstream, throughout the transcribed gene. Both the H3K4me3 and H3K36me3 histone modifications, when present on the N–terminal tails of both H3 components within a nucleosome, are known to inhibit the enzymatic activity and chromatin binding of PRC2 (Schmitges et al., 2011; Voigt et al., 2012). Therefore, it is unclear
how PRC2 initiates repression of hitherto active genes during differentiation. Previous work suggests a potential link between Polycomb repressive mechanisms and histone demethylating activities directed towards the H3K4 and H3K36 residues. For example, Pasini and colleagues reported that PRC2 is associated with the H3K4me3 demethylase RBP2 (also known as KDM5A) in mouse embryonic stem cells, and that this association is important for mediating the removal of H3K4me3 and for silencing of Polycomb target genes (Pasini et al., 2008). Additionally, the related H3K4me3 demethylase (also known as KDM5B) shares a large proportion of PRC2 target genes in mouse embryonic stem cells (Schmitz et al., 2011). Importantly, depletion of JARID1B resulted in higher levels of H3K4me3 at a subset of Polycomb target genes, concomitant with a reduction in PRC2 binding at these sites. Furthermore, in *D. melanogaster*, the H3K36me2 demethylase dKDM2 physically associates with a variant PRC1, enhances Polycomb function and is required for target gene silencing (Lagarou et al., 2008). Taken together, these data indicate that Polycomb repressive mechanisms, in part necessitate the specific concomitant modulation of H3K4 and H3K36 methylation status at target genes.

Here I set out to elucidate the role of the PCL3 molecule within PRC2. I show that a primary function of PCL3 is to recruit PRC2 and the H3K36me3 demethylase NO66 (Kooistra and Helin, 2012; Sinha et al., 2010) to target genes. PCL3 achieves this, at least in part, by binding the H3K36me3 mark via its N–terminal tudor domain. While the function of PCL3 is dispensable for embryonic stem cell self–renewal, it potentiates normal embryonic stem cell differentiation, being required for the silencing of embryonic stem cell genes. Taken together, my results predict a model whereby PCL3 binds to an active histone modification to recruit PRC2 and the H3K36me3 demethylase NO66 to embryonic stem cell genes during differentiation.
3.2 Results

3.2.1 PCL3 recruits the lysine demethylase NO66 to chromatin

In order to elucidate the molecular role of PCL3, I designed a screen to identify associated proteins in mammalian cells. I performed affinity purification mass spectrometry (APMS) on FLAG–immunoprecipitations of nuclear extracts prepared from both HEK293 and HMEC cell lines stably expressing a human FLAG–haemagglutinin (HA)–PCL3 fusion protein (Fig. 3.1). As expected, this analysis identified the core components of the PRC2 complex, SUZ12, EED, EZH1 and EZH2 in addition to many additional proteins (Fig. 3.1c and Appendix 3.4). Interestingly, I also identified a novel interactor, NO66, in both cell types. This protein is a member of the Jumonji C (JmjC) family of demethylases and has been reported to demethylate both di- and tri–methylated histone H3 at lysine 4 (H3K4) and lysine 36 (H3K36) (Kooistra and Helin, 2012; Sinha et al., 2010). Therefore, I speculated that NO66 is a PCL3 co–repressor protein, which may play a role in Polycomb target gene repression. I subsequently validated that NO66 associates with PCL3 by Western blotting of endogenous NO66 on immunoprecipitations of FLAG–HA–PCL3 in both HEK293 and HMEC cells (Fig. 3.2a and 3.2b). Next, I also validated the interaction by Western blotting for endogenous PCL3 in a reciprocal purification of human FLAG–NO66, stably expressed in HEK293 cells (Fig. 3.2c). Finally, I performed immunoprecipitations of NO66, PCL3 and EZH2 in mouse embryonic stem cell nuclear protein lysates and showed that these proteins all interact with each other at the endogenous level, but not with the PRC1 component RING1B, included as a negative control (Fig. 3.2d).
Figure 3.1: Biochemical analysis of PCL3.

(a) Western blot analysis using the indicated antibodies of whole cell protein lysates prepared from control (uninfected) or stably infected HEK293 and HMEC cells expressing an N-terminally FLAG/HA tagged PCL3. (b) Silver-stain analysis of anti-FLAG immunoprecipitates (IPs) performed on nuclear protein lysates from cells in panel (a). (c) Summary of peptides identified by mass spectrometry in duplicate (HEK293) and triplicate (HMEC) anti-FLAG IPs performed on nuclear protein lysates from cells in panel (a). Protein mass, accession number, average observed peptides numbers and average Mascot scores are indicated. All proteins were undetected in matched negative control experiments.
Figure 3.2: Validation of the interaction of PCL3 with PRC2 and NO66

(a and b) Western blot analysis using the indicated antibodies of FLAG IPs on nuclear protein lysates showing that endogenous PRC2 components and NO66, but not the PRC1 component CBX8, associate with FLAG–HA–PCL3 in HEK293 (a) and HMEC cells (b). (c) Western blot using the indicated antibodies of FLAG IPs on nuclear protein lysates showing that endogenous PCL3 and EZH2, but not CBX8, associate with FLAG–NO66 in HEK293 cells. (d) Western blot analysis using the indicated antibodies of IPs using antibodies recognizing the indicated endogenous proteins show NO66, PCL3 and EZH2 co–associate in mouse embryonic stem cell nuclear protein lysates.
I next asked if PCL3 is sufficient to recruit PRC2 and NO66 to chromatin. I generated a tetracycline-inducible human GAL4–PCL3 fusion protein in a HEK293 cell line that contained an integrated heterologous GAL4–luciferase reporter construct (Fig. 3.3a) (Hansen et al., 2008). Addition of tetracycline to these cells resulted in GAL4–PCL3 protein induction and in efficient repression of luciferase activity (Fig. 3.3b and 3.3c). Notably, chromatin immunoprecipitation (ChIP) analysis of the same cells showed that induction of GAL4–PCL3 binding to the luciferase promoter was sufficient to recruit the PRC2 component EZH and to induce a strong enrichment of H3K27me3 (Fig. 3.3d). Moreover, GAL4–PCL3 induction also led to NO66 recruitment, with a concomitant loss of H3K36me3 throughout the luciferase gene body. I included the CCNA2 promoter as a negative control. Taken together, these data strongly suggest that PCL3 is sufficient to recruit both PRC2 and NO66 to a specific target gene locus, concomitant with the transition from active expression, associated with high levels of H3K36me3, to gene repression and an associated increase in the levels of H3K27me3.
Figure 3.3: PCL3 recruits the H3K36me3 demethylase NO66 to chromatin to repress gene transcription.

(a) Schematic representation of the GAL4-TKLuc reporter system. Addition of tetracycline in this system leads to induction of a GAL4 DNA-binding domain-PCL3 fusion protein and recruitment of GAL4-PCL3 fusion to the luciferase reporter promoter. (b) Western blot analysis of GAL4-PCL3 TK-Luc cells, grown in the absence (−) or presence (+) of tetracycline for 48 h showing induction of the fusion protein. (c) GAL4–PCL3 TK-Luc cells, but not control empty vector cells, show repression of luciferase activity upon addition of tetracycline for 48 h. (e) ChIP analyses using the indicated antibodies on GAL4–PCL3 TK-Luc cells grown in the absence (−) or presence (+) of tetracycline for 48 h. Precipitated DNA was analysed by quantitative RT–PCR with primers corresponding to the locations indicated in panel (a), the CCNA2 gene promoter was included as a negative control. ChIP enrichments are presented as percentage bound, normalized to input.
3.2.2 The PCL3 tudor domain binds H3K27me3 and H3K36me3 \textit{in vitro}

The PCL3 protein contains an N-terminal tudor domain and tandem PHD finger domains (Fig. 3.4a). Previously, Hunkapiller and colleagues reported that a mouse PCL3 molecule, mutated in the tudor domain, was unable to rescue the loss of H3K27me3 observed upon depletion of endogenous PCL3 in embryonic stem cells, when compared to wild-type PCL3 (Hunkapiller et al., 2012). The tudor domain is an important ‘reading’ domain found in several epigenetic proteins (Maurer-Stroh et al., 2003). To examine whether the tudor domain of human PCL3 reads modified histone tails, I purified it as a GST-fusion protein and performed \textit{in vitro} peptide pull-down assays on biotinylated histone H3 peptides, modified by tri-methylation of lysine residues at positions 4, 9, 27 and 36 (Fig. 3.4b). This analysis demonstrated that the PCL3 tudor domain binds directly to the active H3K36me3 mark, in addition to binding the Polycomb–associated H3K27me3 mark, albeit to a lesser extent.
Figure 3.4: The PCL3 tudor domain binds to H3K27me3 and H3K36me3 modified Histone H3 peptides in vitro

(a) Schematic representation of the domain structure of the human PCL3 protein. PCL3 contains an N-terminal tudor domain followed by tandem PHD finger domains

(b) Western blot analysis using an anti-GST antibody of an in vitro peptide pull-down assay performed using a GST–PCL3_{25–95} fragment containing the tudor domain region. The western blot demonstrates enrichment of GST–PCL3_{25–95} binding on synthetic Histone H3 peptides trimethylated at lysine 27 (H3K27me3) and 36 (H3K36me3).
3.2.3 PCL3 localises with PRC2 and H3K27me3, not H3K36me3 in vivo

In order to investigate the potential in vivo relevance of these PCL3 histone binding capabilities, I next determined the genome–wide occupancy of PCL3 in mouse embryonic stem cells by ChIP followed by DNA sequencing (ChIP-seq). This analysis showed that PCL3 has a highly comparable binding profile to EZH2 and SUZ12 on Polycomb target gene loci (Fig. 3.5a). Importantly, at a global level, PCL3 shares most of its target genes with SUZ12 and EZH2 (Fig. 3.5b). Several genes enriched for SUZ12 and EZH2 appeared not to have associated PCL3 in my ChIP-seq analysis, suggesting that PCL3 may execute a more restricted function in embryonic stem cells. However, when I subsequently tested several of these genes in quantitative ChIP analysis, I could show that they were in fact bound by PCL3 (Fig. 3.6 (see PCL3 +ve and PCL3 –ve)). Therefore, the smaller number of genes bound by PCL3 is likely a consequence of the lower affinity of the PCL3 antibody for its epitope, rather than a reflection of a restricted PRC2 function for PCL3. Consistent with its overlap with Polycomb target genes and my in vitro peptide binding experiments, I found that PCL3 target genes were enriched for the H3K27me3 modification (Fig. 3.5a, b). Importantly, despite the ability of the tudor domain of PCL3 to bind to the H3K36me3 mark, no PCL3 target genes were observed to be marked with H3K36me3 in mouse embryonic stem cells (Fig. 3.5b). This led me to speculate that in vivo PCL3 might only transiently interact with the H3 tail when tri-methylated at K36. Therefore, I next wished to explore the possibility that the lack of H3K36me3 on PCL3 target genes could be reconciled with the fact that PCL3 physically associates with a H3K36me3 demethylase.
Figure 3.5: PCL3 co-localises with PRC2 and H3K27me3, but not H3K36me3, in vivo

(a) Representative examples of PCL3 ChIP-seq signal tracks for the three Polycomb target genes and germ layer markers, *Gata6* (endoderm), *Brachyury* (mesoderm) and *Fgf5* (primitive ectoderm). The *Ccna2* gene is shown as a negative control. Binding profiles of ChIP-seq results for EZH2, SUZ12, H3K27me3, H3K36me3 and H3K4me3 obtained from previous studies are indicated for comparison (Ku et al., 2008; Mikkelsen et al., 2007; Peng et al., 2009). (b) Heat maps depicting the binding patterns of PCL3, PRC2 components (EZH2 and SUZ12) and histone modifications (H3K27me3 and H3K36me3) within −5 and +5 Kb of the transcription start sites (TSS) of all ‘active’ and ‘bivalent’ genes in mouse embryonic stem cells. ‘Bivalent’ genes were ranked on their H3K27me3 read density, and ‘active’ genes were ranked on H3K4me3 read density around the TSS.
3.2.4 PCL3 and NO66 co-localise on target genes in mouse embryonic stem cells

To explore the potential functional interplay between PCL3 and NO66, I first decided to determine if both proteins co-occupy the same target genes in mouse embryonic stem cells. I performed ChIP analysis followed by real-time PCR on a cohort of Polycomb target genes and found that PCL3 and NO66 co-occupied 8 of the 10 tested Polycomb target gene promoters (Fig. 3.6). I also analyzed a cohort of 10 “non-Polycomb” repressed genes in these ChIP experiments. Consistent with these genes being repressed independently of Polycombs, they lacked both H3K27me3 and H3K4me3, but still retained Histone H3 on their promoters. Importantly, I did not observe PCL3 or NO66 on the promoters of these genes. Taken together, these results suggest that NO66 specifically associates with Polycomb and PCL3 target genes. Interestingly, my observation that NO66 co-occupied bivalent Polycomb target genes, which contained both H3K27me3 and H3K4me3, suggests that NO66 may primarily function as an H3K36me3 demethylase in vivo. Furthermore, these data suggest that the co-occupancy of NO66 and PCL3 on target genes is causally linked and not an indirect consequence of a more general role for NO66 as a transcriptional repressor.
Figure 3.6: NO66 binds Polycomb repressed genes in mouse embryonic stem cells

ChIP analyses of PCL3, NO66, H3K27me3, H3K4me3, Histone H3 and HA (negative control) in mouse embryonic stem cells. 'PCL3-+ve' indicates genes identified as PCL3 targets in our genome-wide ChIP-seq analysis. 'PCL3-ve' indicates previously reported Polycomb target genes that were not identified as PCL3 targets in our ChIP-seq experiment. Precipitated DNA was analyzed by quantitative RT-PCR with primers corresponding to the promoter regions of the indicated actively expressed, Polycomb repressed, and non-Polycomb repressed genes and presented as percentage bound, normalized to input.
3.2.5 PCL3 controls PRC2 and NO66 occupancy on target genes

I next wished to investigate if PCL3 is required for the association of the PRC2 complex and NO66 to target genes. To do this, I established mouse embryonic stem cell lines that stably expressed either scrambled (SCR) or one of two independent Pcl3-specific short hairpin RNAs (shRNAs). Western blot and qPCR analysis confirmed efficient and specific inhibition of PCL3 expression (Fig. 3.7). These cells were only marginally affected in terms of their expression of embryonic stem cell genes, such as *Pou5f1* and *Nanog*, growth rate and alkaline phosphatase activity (Fig. 3.9, Fig. 3.10a, b). However, I did observe a global depletion of H3K27me3 and a global increase of H3K36me3 in PCL3 depleted cells (Fig. 3.7). I next performed ChIP analysis followed by quantitative real-time PCR of these cells and found that inhibition of PCL3 led to its depletion from the promoters of several Polycomb target genes, including markers of ectodermal (*Fgf5* and *Olig2*), mesodermal (*Gata4* and *Gata6*), and endodermal differentiation (*Pax3* and *Brachyury*) (Fig. 3.8). This depletion of PCL3 correlated with a dramatic reduction in the binding of EZH2 and NO66 on the promoters of these genes. However, despite these changes on the chromatin level, I only detected moderate increases in the expression of a subset of these germ cells genes (Fig. 3.10b). This is consistent with previous reports of the activation of only a sub-cohort of Polycomb target genes upon inhibition of components of the PRC2 complex (Bracken, 2006; Pasini et al., 2007). It has been speculated that the DNA binding, lineage specific transcription factors that normally activate this cohort of genes are absent in the undifferentiated cells (Bracken et al., 2006). Taken together, these results strongly suggest that PCL3, while not required for embryonic stem cell proliferation, is required for the maintenance of PRC2 and NO66 on the promoters of differentiation genes in embryonic stem cells.
Figure 3.7: shRNA mediated knockdown of PLC3 in mouse embryonic stem cells

Western blot analysis using the indicated antibodies on whole cell protein lysates prepared from mouse embryonic stem cells expressing scrambled (shSCR), or one of two Pcl3-specific (shPCL3 1. and 2.) shRNAs shows efficient depletion of PCL3, which correlates with a global decrease of H3K27me3 and a global increase of H3K36me3.
Figure 3.8: PCL3 is required to maintain PRC2 and NO66 binding on target genes in ES cells

ChIP analysis of PCL3, NO66, EZH2 and H3K27me3 on promoters of lineage genes in mouse embryonic stem cells expressing a scrambled shRNA (shScr – Red) or one of two independent PCL3 targeting shRNAs (shPcl3.1 – Light green and shPcl3.2 – Dark green). ChIPs are presented as percentage bound, normalized to input. The Ccna2 gene promoter is represented as a negative control.
Figure 3.9: Depletion of PCL3 does not significantly affect the ES cell phenotype

(a) Proliferation of ES cells is not affected by PCL3 depletion. Mouse ES cells expressing either non-specific scrambled (shSCR) or PCL3 specific shRNAs (1 and 2) were seeded for growth assays and stained with crystal violet at the indicated time-points. (b) Alkaline phosphatase staining of cells from (a) shows no significant affect upon inhibition of PCL3.
Figure 3.10: Depletion of PCL3 in ES cell has minimal transcriptional affects on stem cell and differentiation genes

(a) qPCR expression analysis of embryonic stem cell expressed genes in ES cells expressing either control scrambled shRNA (Scr – Red) or one of two independent Pcl3 specific shRNAs (1 - Light green and 2 - Dark green). (b) qPCR expression analysis of germ cell marker genes in cells from (a).
In order to rule out the possibility that NO66 is required for targeting PCL3 and PRC2 to chromatin, I established mouse embryonic stem cell lines that stably expressed either scrambled (SCR) or one of two independent No66–specific short hairpin RNAs (shRNAs). Western blot and quantitative RT-PCR analysis confirmed efficient inhibition of NO66 expression (Fig. 3.11, Fig3.13a). As expected, the depletion of NO66 led to loss of the protein on the promoters of Polycomb target genes, (Fig. 3.12). However, this loss did not lead to the displacement of PCL3 or PRC2. These NO66 depleted cells had no detectable proliferation or pluripotency defects and did not significantly activate the expression of Polycomb target genes (Fig. 3.13a and data not shown). Furthermore, NO66 depleted cells exhibited no differentiation defects and are fully capable of repressing pluripotency markers when differentiated towards embryoid bodies (Fig. 3.13b). Taken together, my results show that NO66 localization is dependent on PCL3, but not the other way round and support a model in which PCL3 is targeted to chromatin independently of NO66.
Figure 3.11: Depletion of NO66 in mouse ES cells

Western blot analysis using the indicated antibodies of mouse embryonic stem cells expressing either non-specific scrambled (shSCR), Pcl3 specific (shPCL3.1) or one of two NO66 specific (shNO66.1 and shNO66.2) shRNAs.
Figure 3.12: Depletion of NO66 in mouse ES cells does not affect PCL3 occupancy on target genes

ChIP analysis of PCL3, NO66, EZH2 and H3K27me3 in mouse ES cells from expressing either a control scrambled shRNA (shScr – Red), a Pcl3 targeting shRNA (shPcl3.1 – Dark green) or one of two independent No66 targeting shRNAs (shNo66.1 – yellow and shNo66.2 – Blue). The precipitated DNA was analysed by quantitative RT-PCR with primers corresponding to the promoter regions of the indicated germ cell marker genes and presented as percentage bound, normalised to input. The Ccna2 gene promoter is presented as a negative control.
Figure 3.13: Depletion of NO66 in mouse ES cells has minimal transcriptional affects in ES cells or during differentiation

(a) qPCR expression analysis of the indicated genes in mouse ES cells expressing either control scrambled shRNA (SCR – Red), a Pcl3 targeting shRNA (shPCL3 – Dark green) or one of two independent No66 targeting shRNAs (NO66.1 – yellow and NO66.2 – Blue). (b) qPCR expression analysis of the indicated genes in either undifferentiated mouse embryonic stem cells from (a) (indicated with a “0”), or the same cells induced to differentiate into embryoid bodies for 8 days.
3.2.6 PCL3 is necessary for de novo recruitment of NO66 and PRC2

To assess whether PCL3 is required for the de novo recruitment of PRC2 and NO66 to target genes during embryonic stem cell differentiation, I generated mouse embryonic stem cells, expressing either scrambled (SCR) or Pcl3-specific shRNAs, and differentiated them to embryoid bodies. Inhibition of PCL3 dramatically affected the ability of these embryonic stem cells to form large embryoid bodies with differentiation features (Fig. 3.14a, b). Consistent with this, PCL3 inhibition led to impaired silencing of embryonic stem cell genes (Fig. 3.15a) and to reduced activation of differentiation genes (Fig. 3.15b). I also observed that PCL3 inhibition led to a defect in silencing of embryonic stem cell genes when the cells were differentiated with retinoic acid addition for 2 days under adherent conditions (Fig. 3.16). ChIP experiments of embryoid bodies, 8 days after they were induced to differentiate demonstrated that inhibition of PCL3 resulted in a decreased recruitment of the PCL3 protein to both the promoters and intragenic regions of the Fgf4 and Pou2f3 genes (Fig. 3.17). Importantly, PCL3 inhibition also led to impaired recruitment of PRC2 and NO66 to these regions, concomitant with reduced accumulation of H3K27me3 and failure to remove the H3K36me3 mark. Conversely, mouse embryonic stem cells, expressing NO66-specific shRNAs, when differentiated to embryoid bodies, were fully capable of repressing embryonic stem cell expressed genes (Fig. 3.13b). My observation that depletion of PCL3 led to a greater increase in H3K36me3 levels than depletion of NO66, suggests that PCL3 is required for the function of additional H3K36me3 demethylases (Fig. 3.11). Taken together, these results show that PCL3 is necessary for the de novo targeting of the PRC2 complex, as well as NO66, during embryonic stem cell differentiation.
Figure 3.14: PCL3 is required for ES cell differentiation

(a) Phase contrast microscope image of embryoid bodies formed after 8 days of non-directed differentiation of embryonic stem cells, expressing either non-specific scrambled (shSCR) or PCL3 specific (shPCL3.1 and shPCL3.2) shRNAs reveals defects in differentiation in the absence of PCL3 (Left panel). Hematoxylin and Eosin (H&E) stains of embryoid bodies formed after 12 days from embryonic stem cell expressing either shSCR or shPCL3 (Right Panel). (b) A total of 100 embryoid bodies were examined and the overall percentage of large embryoid bodies exhibiting structural features of differentiation ‘Large differentiated’ and small embryoid bodies which lacked these features ‘Small undifferentiated’ in shSCR and shPCL3 cells was examined.
Figure 3.15: Depletion of PCL3 leads to impaired silencing of ES cell expressed genes following embryoid body differentiation of ES cells

(a) Defects in embryonic stem cell gene repression during differentiation are observed in PCL3 depleted cells. Presented are qPCR expression analysis of the indicated embryonic stem cell expressed genes in undifferentiated shSCR, shPCL3.1 or shPCL3.2 mouse embryonic stem cells (indicated with a '0'), or the same cells induced to differentiate towards embryoid bodies for 4 and 8 days. (b) qPCR expression analysis of the indicated germ cell differentiation genes in cells from panel (a).
Figure 3.16: Depletion of PCL3 leads to impaired silencing of ES cell expressed genes following retinoic acid induced differentiation of ES cells

qPCR analysis of the indicated genes in mouse embryonic stem cells, expressing either non-specific scrambled (shSCR) or PCL3 specific (1 and 2) shRNAs (indicated with a "0"), or the same cells induced to differentiate with 1μM all-trans retinoic acid (ATRA) for 2 days.
Figure 3.17: PCL3 is required for the de novo recruitment of PRC2 and NO66 to ES cell genes during differentiation

Dramatic chromatin defects are revealed in the absence of PCL3. Presented are ChIP analyses of PCL3, NO66, EZH2, H3K27me3 and H3K36me3 in undifferentiated shSCR or shPCLS mouse embryonic stem cells (indicated with a "0"), or the same cells induced to differentiate towards embryoid bodies for 8 days. Precipitated DNA was analysed by quantitative RT–PCR with primers corresponding to the promoter (labelled 1 in the top panel) or intragenic (2 in the top panel) regions of the embryonic stem cell expressed genes Fgf4 and Pou2f3 and presented as percentage bound, normalized to input.
3.2.7 A PCL3 W50C Y56A mutant cannot bind H3K27me3 or H3K36me3

I next wished to determine if the PCL3 tudor domain is required for targeting PRC2 and NO66 to chromatin. Previously, Hunkapiller and colleagues reported that the mouse PCL3 molecule when mutated at positions W48 and Y54, within the tudor domain, was unable to rescue the loss of H3K27me3 observed upon depletion of endogenous PCL3 (Hunkapiller et al., 2012). I therefore generated the corresponding human PCL3 mutant tudor domain [W50C Y56A (corresponding to mouse W48 and Y54, respectively)], purified it as a GST–fusion protein and performed *in vitro* peptide pull–down assays on biotinylated histone H3 peptides, modified by mono–, di– and tri–methylation of lysine residues at positions 27 and 36 (Fig. 3.18a, b). This revealed that the mutant tudor domain was incapable of binding the H3K27me3 and H3K36me3 modifications. I then performed independent surface plasmon resonance (SPR) to accurately measure the *in vitro* binding properties of the wild–type and mutant GST–PCL3 tudor domain to the H3K36me3 peptide (Fig. 3.18c, d). I found the binding of GST–PCL3 to H3K36me3 was highly reproducible with an apparent $K_d$ of $\sim$188 nM, as determined from independent measurements at 5 concentrations of protein. Consistent with my peptide pull–down experiment, I observed negligible binding of GST–PCL3 mutant tudor domain to the H3K36me3 peptide.
Figure 3.18: Mutant PCL3 tudor domain is incapable of binding modified histone peptides

(a) Protein alignments of the tudor domains of human, mouse and Drosophila PCL proteins. Indicated are the residues targeted in the double point mutated form of PCL3 (b) Western blots of in vitro peptide pull-down assays using wild-type (WT) GST–PCL3<b>25-95</b>, a mutant (W50C/Y56A) GST–PCL3<b>25-95</b> or GST–only show that the point mutated PCL3 tudor does not bind H3K27me3 or H3K36me3 modified peptides (c and d) Representative SPR sensorgrams for WT–PCL3<b>325-95</b> (c), and WT and mutant–GST–PCL3<b>25-95</b> (d) and H3K36me3. Affinity rate constants were determined from a concentration series of WT–PCL3 binding to H3K36me3.
3.2.8 PCL3 tudor domain is required for PRC2 and NO66 recruitment

To address whether recognition of H3K36me3 is important for PCL3 function in vivo, I performed rescue experiments in parallel, using lentiviral expression vectors, either expressing wild-type human PCL3 or human PCL3 containing the two point mutations (W50C Y56A), which result in loss of H3K36me3 recognition. The reintroduction of the wild-type PCL3 restored genome-wide levels of H3K27me3, as previously shown, as well as the association of NO66 and PRC2 to target genes in embryonic stem cells (Fig. 3.19, Fig. 3.20). It also rescued the defects in embryoid body differentiation (Fig. 3.21a, b) and restored the association of PRC2 and NO66 to hitherto active genes during embryonic stem cell differentiation (Fig. 3.22). In contrast, the PCL3 mutant (W50C Y56A) was incapable of associating with Polycomb target genes in embryonic stem cells (Fig. 3.20), did not rescue the embryoid body differentiation defects (Fig. 3.21a, b) and was not recruited to embryonic stem cell genes during differentiation, leading to a failure to recruit both PRC2 and NO66, and correlated with impaired gene silencing (Fig. 3.22).
Figure 3.19: PCL3 knockdown rescue experiment

(a) Western blot analyses using the indicated antibodies in mouse embryonic stem cells expressing either non-specific scrambled (shSCR) or PCL3 specific (shPCL3) shRNAs, where PCL3 knockdown was rescued by over-expression of either wildtype (WT) or mutated (W50C/Y56A) PCL3.
Figure 3.20: Wildtype, but not mutant PCL3 is capable of restoring PRC2 and NO66 binding at target genes in PCL3 depleted ES cells

Rescue of PRC2 and NO66 recruitment to Polycomb target genes with WT, but not mutant PCL3. Presented are ChIP analyses of PCL3, NO66, EZH2 and H3K27me3 in mouse embryonic stem cells expressing shSCR (Red) or shPCL3 alone (Dark green) or in combination with WT (shPCL3+WT−hPCL3 (Light blue)) or mutant (shPCL3+W50C Y56A−hPCL3 (Dark blue)) human PCL3. The precipitated DNA was analysed by quantitative RT-PCR with primers corresponding to the promoter regions of the indicated germ cell marker genes and presented as percentage bound, normalised to input. The Ccna2 gene promoter is presented as a negative control.
Figure 3.21: PCL3 tudor domain mutant does not rescue the differentiation defects seen following PCL3 depletion

(a) Rescue of embryoid body differentiation defects with WT but not mutant PCL3. Presented are phase contrast images of embryoid bodies formed after 8 days differentiation of embryonic stem cells expressing shSCR or shPCL3 alone or in combination with WT (shPCL3+WT-hPCL3) or mutant (shPCL3+W50C Y56A-hPCL3) human PCL3 (b) qPCR expression analysis of the indicated embryonic stem cell expressed genes in ES cells expressing shSCR (Red) or shPCL3 alone (Dark green) or in combination with WT (shPCL3+WT-hPCL3 (Light blue)) or mutant (shPCL3+W50C Y56A-hPCL3 (Dark blue)) human PCL3 (indicated with a “0”), or the same cells induced to differentiate towards embryoid bodies for 8 days (indicated with an “8”).
Figure 3.22: Wildtype but not mutant PCL3 rescues the de novo recruitment of PRC2 and NO66 to stem cell genes during differentiation

Rescue of PRC2 and NO66 recruitment to embryonic stem cell genes during differentiation with WT, but not mutant PCL3. Presented are ChIP analyses of PCL3, NO66, EZH2, H3K27me3 and H3K36me3 in ES cells from (Fig 3.19) (indicated with a "0"), or the same cells induced to differentiate towards embryoid bodies for 8 days. Precipitated DNA was analysed by quantitative RT-PCR and presented as percentage bound, normalized to input.
Taken together, these results demonstrate that the ability of the PCL3 tudor domain to recognise the H3K27me3 and H3K36me3 modifications is essential for its function in vivo. I propose a model in which PCL3 'reads' the H3K36me3 mark at embryonic stem cells genes and recruits the core PRC2 complex, leading to H3K27me3 accumulation and gene repression during differentiation. The presence of H3K27me3 on these genes may contribute to stabilizing the association of PCL3 via its tudor domain. Furthermore, the demethylating activity of NO66 may contribute to stabilizing PRC2 activity on chromatin by preventing accumulation of the antagonistic H3K36me3 modification (Fig. 3.23).
Instructive differentiation cues

Repressive Polycomb target genes e.g. Gata4, Fgf5 etc...

Instructive differentiation cues

Repressive Polycomb target genes e.g. Gata4, Fgf5 etc...

Active stem cell genes e.g. Fgf4, Pou2f3 etc...

Active stem cell genes e.g. Fgf4, Pou2f3 etc...

Figure 3.23: Model for PCL3 function in ES cells and during differentiation

(a) In undifferentiated embryonic stem cells, PCL3 and NO66 are co-bound on Polycomb repressed genes such as the germ cell markers, Olig2 and Gata4. During differentiation PCL3 and other PRC2 components are displaced from differentiation genes as they become activated.

(b) During differentiation of embryonic stem cells, PCL3 binds to H3K36me3-modified histones at active embryonic stem cell genes, such as Fgf4 and Pou2f3. This binding results in the recruitment of the PRC2 complex and the H3K36me3 demethylase NO66 to hitherto active genes, resulting in a switch from K36me3 to K27me3 and consequent transcriptional repression.
3.3 Discussion

A key unanswered question has been how the PRC2 complex is targeted to regions of active chromatin, in order to facilitate subsequent gene repression during lineage transitions. My demonstration that PCL3 confers a H3K36me3 reading capability to PRC2 reveals a new aspect to the transition from an active gene to a Polycomb-repressed gene. Furthermore, I provide the first evidence revealing that a H3K36me3 demethylase activity is associated with the PRC2 complex.

The H3K36me3 and H3K27me3 histone modifications have never been observed to co-localize on chromatin in ChIP-seq experiments (Kouzarides, 2007; Mikkelsen et al., 2007; Zhou et al., 2011). Furthermore, biochemical analyses of histone H3 modifications suggest that both marks rarely, if ever, co-exist on the same histone H3 tail (Yuan et al., 2011). My demonstration that the H3K36me3 demethylase NO66 associates with PCL3 may explain this, at least in part. Tellingly, I found that depletion of PCL3 led to a much greater increase in global H3K36me3 levels, when compared to depletion of NO66. This could suggest that additional H3K36me3 demethylases are dependent on PCL3 and the PRC2 complex for their activity and (or) association on target sites. In addition to NO66, there are at least four other H3K36me3 demethylases known (Kooistra and Helin, 2012), yet we only found NO66 to be associated with the PCL3–PRC2 complex. Therefore, an alternative possibility is that the presence of the PCL3–PRC2 complex on chromatin is refractory to H3K36me3 methyltransferase binding and (or) enzymatic activity. However, previous observations have implicated other H3K36me3 demethylases in Polycomb gene repression. For example, the JMJD2A (also known as KDM2A) H3K36me3 demethylase is recruited to unmethylated CpG islands, which are enriched in Polycomb target genes, via its zinc finger CxxC (ZF–CxxC) domain (Blackledge et al., 2010; Zhou et al., 2011). This suggests that this particular H3K36me3 demethylase may be recruited to Polycomb target genes to collaborate in gene silencing, and it may be indicative of a more general link between Polycomb repressive mechanisms and H3K36me3 demethylase activities. It will be important to
determine if the other two mammalian PCL proteins, PCL1 and PCL2, associate with NO66 or indeed any other H3K36me3 demethylases.

My results also have implications for the roles of the PCL proteins in early mammalian development. The phenotypes observed upon knockdown of PCL3 are similar to those previously observed upon loss of other PRC2 components and associated proteins. For example, the observation that PCL3 inhibition does not affect embryonic stem cell proliferation, but does reduce the association of Polycomb proteins to the promoters of repressed differentiation genes, is in keeping with previous observations on SUZ12 (Pasini et al., 2007), EZH2 (Shen et al., 2008), EED (Montgomery et al., 2005) and JARID2 (Landeira et al., 2010; Pasini et al., 2010; Peng et al., 2009). Similarly, like PCL3, these proteins have been shown (by these and other studies) to be required during embryonic stem cell differentiation for the silencing of genes previously expressed in undifferentiated embryonic stem cells (Sauvageau and Sauvageau, 2010). However, for the first time, my work now offers molecular insights into how the PRC2 complex is targeted to these genes and thereby initiates subsequent repression.

My results also raise some intriguing new questions. For instance, PCL3 is expressed both in embryonic stem cells and during their differentiation towards embryoid bodies. Thus, while I have established that PCL3 functions in the early stages of the transition from active genes to PRC2 repressed genes, the actual trigger that initiates the de novo recruitment of PCL3 remains unknown. For instance, the H3K36me3 mark is present on the loci of these genes in embryonic stem cells, yet PCL3 is not recruited until these cells are induced to differentiate. Thus, additional changes to the local chromatin environment, such as deacetylation, may also be required. Alternatively, nucleosome compaction, lineage specific DNA binding transcription factors and ncRNAs may also play a role (Bracken and Helin, 2009). Furthermore, it is unclear how the demethylase activity of NO66 is sterically compatible with the simultaneous binding of the PCL3 tudor domain. It is possible that NO66 is not required for the removal of the H3K36me3 mark
during initial gene silencing. Instead, other demethylase complexes might act in parallel with PRC2 leading to demethylation. Therefore, the function of NO66 might be to prevent aberrant H3K36me3 accumulation on Polycomb target genes. Answering these questions will undoubtedly shed further light on the molecular mechanisms at the heart of not only embryonic stem cell differentiation, but also lineage changes in general.
3.4 Appendix

<table>
<thead>
<tr>
<th>Rank</th>
<th>HEK293 IP-MS 1.</th>
<th>HEK293 IP-MS 2.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Protein Name</td>
<td>Mascot Score</td>
</tr>
<tr>
<td>1</td>
<td>ANM5</td>
<td>1169</td>
</tr>
<tr>
<td>2</td>
<td>SPTA2</td>
<td>1132</td>
</tr>
<tr>
<td>3</td>
<td>SPTB2</td>
<td>1012</td>
</tr>
<tr>
<td>4</td>
<td>SUZ12</td>
<td>669</td>
</tr>
<tr>
<td>5</td>
<td>PCL3</td>
<td>555</td>
</tr>
<tr>
<td>6</td>
<td>FLNA</td>
<td>459</td>
</tr>
<tr>
<td>7</td>
<td>HNRPU</td>
<td>366</td>
</tr>
<tr>
<td>8</td>
<td>DSG1</td>
<td>351</td>
</tr>
<tr>
<td>9</td>
<td>MEP50</td>
<td>351</td>
</tr>
<tr>
<td>10</td>
<td>EZH2</td>
<td>334</td>
</tr>
<tr>
<td>11</td>
<td>PPM1B</td>
<td>318</td>
</tr>
<tr>
<td>12</td>
<td>ALBU</td>
<td>280</td>
</tr>
<tr>
<td>13</td>
<td>LAP2A</td>
<td>266</td>
</tr>
<tr>
<td>14</td>
<td>LAP2B</td>
<td>266</td>
</tr>
<tr>
<td>15</td>
<td>POTEE</td>
<td>240</td>
</tr>
<tr>
<td>16</td>
<td>RBM10</td>
<td>238</td>
</tr>
<tr>
<td>17</td>
<td>LAS1L</td>
<td>232</td>
</tr>
<tr>
<td>18</td>
<td>MBB1A</td>
<td>227</td>
</tr>
<tr>
<td>19</td>
<td>DDX17</td>
<td>221</td>
</tr>
<tr>
<td>20</td>
<td>G3P</td>
<td>220</td>
</tr>
<tr>
<td>21</td>
<td>EED</td>
<td>216</td>
</tr>
<tr>
<td>22</td>
<td>PELP1</td>
<td>213</td>
</tr>
<tr>
<td>23</td>
<td>ADT3</td>
<td>208</td>
</tr>
<tr>
<td>24</td>
<td>SPIN1</td>
<td>207</td>
</tr>
<tr>
<td>25</td>
<td>POTEF</td>
<td>199</td>
</tr>
<tr>
<td>26</td>
<td>LMNA</td>
<td>189</td>
</tr>
<tr>
<td>27</td>
<td>SFRS1</td>
<td>165</td>
</tr>
<tr>
<td>28</td>
<td>DDX5</td>
<td>161</td>
</tr>
<tr>
<td>29</td>
<td>GFAP</td>
<td>158</td>
</tr>
<tr>
<td>30</td>
<td>HSP71</td>
<td>156</td>
</tr>
<tr>
<td>31</td>
<td>HSP7C</td>
<td>156</td>
</tr>
<tr>
<td>32</td>
<td>DHX9</td>
<td>148</td>
</tr>
<tr>
<td>33</td>
<td>DCD</td>
<td>144</td>
</tr>
<tr>
<td>34</td>
<td>TRY3</td>
<td>140</td>
</tr>
<tr>
<td>35</td>
<td>ZA2G</td>
<td>139</td>
</tr>
<tr>
<td>36</td>
<td>SFRS2</td>
<td>134</td>
</tr>
<tr>
<td>37</td>
<td>HS71L</td>
<td>128</td>
</tr>
<tr>
<td>38</td>
<td>PABP1</td>
<td>125</td>
</tr>
<tr>
<td>39</td>
<td>PABP4</td>
<td>125</td>
</tr>
<tr>
<td>40</td>
<td>HNRPR</td>
<td>123</td>
</tr>
<tr>
<td>41</td>
<td>SENP3</td>
<td>123</td>
</tr>
<tr>
<td>42</td>
<td>PPM1A</td>
<td>116</td>
</tr>
<tr>
<td>43</td>
<td>M20M</td>
<td>115</td>
</tr>
<tr>
<td>44</td>
<td>HSP72</td>
<td>114</td>
</tr>
<tr>
<td>45</td>
<td>LYSC</td>
<td>110</td>
</tr>
<tr>
<td>46</td>
<td>GRP78</td>
<td>108</td>
</tr>
<tr>
<td>Rank</td>
<td>Protein Name</td>
<td>Mascot Score</td>
</tr>
<tr>
<td>------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>1</td>
<td>SFRS7</td>
<td>173</td>
</tr>
<tr>
<td>2</td>
<td>MYH9</td>
<td>166</td>
</tr>
<tr>
<td>3</td>
<td>CCD25</td>
<td>157</td>
</tr>
<tr>
<td>4</td>
<td>PCL3</td>
<td>129</td>
</tr>
<tr>
<td>5</td>
<td>DHX9</td>
<td>105</td>
</tr>
<tr>
<td>6</td>
<td>MATR3</td>
<td>102</td>
</tr>
<tr>
<td>7</td>
<td>SFR14</td>
<td>85</td>
</tr>
<tr>
<td>8</td>
<td>PRP8</td>
<td>74</td>
</tr>
<tr>
<td>9</td>
<td>SF3B1</td>
<td>70</td>
</tr>
<tr>
<td>10</td>
<td>SF3B3</td>
<td>69</td>
</tr>
<tr>
<td>11</td>
<td>POTEEM</td>
<td>67</td>
</tr>
<tr>
<td>12</td>
<td>US20</td>
<td>63</td>
</tr>
<tr>
<td>13</td>
<td>MYO1C</td>
<td>63</td>
</tr>
<tr>
<td>14</td>
<td>MBB1A</td>
<td>58</td>
</tr>
<tr>
<td>15</td>
<td>RBM10</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>16</td>
<td>DDX21</td>
<td>54</td>
</tr>
<tr>
<td>17</td>
<td>ILF3</td>
<td>53</td>
</tr>
<tr>
<td>18</td>
<td>AHNK</td>
<td>46</td>
</tr>
<tr>
<td>19</td>
<td>DHX15</td>
<td>44</td>
</tr>
<tr>
<td>20</td>
<td>HSP70C</td>
<td>43</td>
</tr>
<tr>
<td>21</td>
<td>ANXA2</td>
<td>43</td>
</tr>
<tr>
<td>22</td>
<td>SF3B2</td>
<td>39</td>
</tr>
<tr>
<td>23</td>
<td>SF3A1</td>
<td>37</td>
</tr>
<tr>
<td>24</td>
<td>MYH11</td>
<td>35</td>
</tr>
<tr>
<td>25</td>
<td>US51</td>
<td>32</td>
</tr>
<tr>
<td>26</td>
<td>MYO6</td>
<td>32</td>
</tr>
<tr>
<td>27</td>
<td>MYO1B</td>
<td>31</td>
</tr>
<tr>
<td>28</td>
<td>ADT3</td>
<td>31</td>
</tr>
<tr>
<td>29</td>
<td>SUZ12</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>GRP78</td>
<td>30</td>
</tr>
<tr>
<td>31</td>
<td>SR140</td>
<td>29</td>
</tr>
<tr>
<td>32</td>
<td>PRKDC</td>
<td>29</td>
</tr>
<tr>
<td>33</td>
<td>SPTB2</td>
<td>28</td>
</tr>
<tr>
<td>34</td>
<td>MYH10</td>
<td>27</td>
</tr>
<tr>
<td>35</td>
<td>SPTA2</td>
<td>26</td>
</tr>
<tr>
<td>36</td>
<td>PELP1</td>
<td>26</td>
</tr>
<tr>
<td>37</td>
<td>KV204</td>
<td>26</td>
</tr>
<tr>
<td>38</td>
<td>GNL3</td>
<td>26</td>
</tr>
<tr>
<td>39</td>
<td>DDX5</td>
<td>26</td>
</tr>
<tr>
<td>40</td>
<td>ANM5</td>
<td>26</td>
</tr>
<tr>
<td>41</td>
<td>ADT2</td>
<td>26</td>
</tr>
<tr>
<td>42</td>
<td>SRRM2</td>
<td>25</td>
</tr>
<tr>
<td>43</td>
<td>KV206</td>
<td>25</td>
</tr>
<tr>
<td>44</td>
<td>GRP75</td>
<td>25</td>
</tr>
<tr>
<td>45</td>
<td>MYH14</td>
<td>24</td>
</tr>
<tr>
<td>46</td>
<td>SVIL</td>
<td>22</td>
</tr>
<tr>
<td>47</td>
<td>EZH2</td>
<td>22</td>
</tr>
<tr>
<td>48</td>
<td>DSRAD</td>
<td>22</td>
</tr>
<tr>
<td>49</td>
<td>TR150</td>
<td>21</td>
</tr>
<tr>
<td>50</td>
<td>TEX10</td>
<td>21</td>
</tr>
<tr>
<td>51</td>
<td>LAS1L</td>
<td>21</td>
</tr>
<tr>
<td>52</td>
<td>GELS</td>
<td>21</td>
</tr>
<tr>
<td>53</td>
<td>DESP</td>
<td>20</td>
</tr>
<tr>
<td>54</td>
<td>CPSF1</td>
<td>20</td>
</tr>
<tr>
<td>55</td>
<td>ARP3</td>
<td>20</td>
</tr>
<tr>
<td>56</td>
<td>ILF2</td>
<td>18</td>
</tr>
<tr>
<td>57</td>
<td>HSP72</td>
<td>17</td>
</tr>
<tr>
<td>58</td>
<td>HS90B</td>
<td>17</td>
</tr>
<tr>
<td>59</td>
<td>GFAP</td>
<td>17</td>
</tr>
<tr>
<td>60</td>
<td>STOM</td>
<td>16</td>
</tr>
<tr>
<td>61</td>
<td>RRP12</td>
<td>16</td>
</tr>
<tr>
<td>62</td>
<td>RBM14</td>
<td>16</td>
</tr>
<tr>
<td>63</td>
<td>EPIPL</td>
<td>16</td>
</tr>
<tr>
<td>64</td>
<td>BPA1</td>
<td>16</td>
</tr>
<tr>
<td>65</td>
<td>ANXA1</td>
<td>16</td>
</tr>
<tr>
<td>66</td>
<td>EF1A1</td>
<td>15</td>
</tr>
<tr>
<td>67</td>
<td>TIF1B</td>
<td>14</td>
</tr>
<tr>
<td>68</td>
<td>ROA2</td>
<td>14</td>
</tr>
<tr>
<td>69</td>
<td>RBPP4</td>
<td>14</td>
</tr>
<tr>
<td>70</td>
<td>PRP19</td>
<td>14</td>
</tr>
<tr>
<td>71</td>
<td>LAP2B</td>
<td>14</td>
</tr>
<tr>
<td>72</td>
<td>EED</td>
<td>14</td>
</tr>
<tr>
<td>73</td>
<td>XRN2</td>
<td>13</td>
</tr>
<tr>
<td>74</td>
<td>WDR33</td>
<td>13</td>
</tr>
<tr>
<td>75</td>
<td>VDAC1</td>
<td>13</td>
</tr>
<tr>
<td>76</td>
<td>SF3A3</td>
<td>13</td>
</tr>
<tr>
<td>77</td>
<td>RBP2</td>
<td>13</td>
</tr>
<tr>
<td>78</td>
<td>IF4A3</td>
<td>13</td>
</tr>
<tr>
<td>79</td>
<td>CSTF3</td>
<td>13</td>
</tr>
<tr>
<td>80</td>
<td>PABP1</td>
<td>12</td>
</tr>
</tbody>
</table>
4 Variant PRC2 Complexes between quiescent and proliferating cells defined by PCL members
4.1 Introduction

The expression of several PRC2 complex components is regulated by cell growth. This implies that the composition of the PRC2 complex changes between non-growing and growing cells. However, it is unclear how this regulation affects PRC2 complex composition and function. In quiescent (G0) cells, the EZH2 and SUZ12 proteins are absent, or very low. Both the EZH2 and SUZ12 genes are transcriptionally activated by E2F transcription factors at the G1–S transition in cells re-entering the cell cycle (Bracken et al., 2003; Pasini et al., 2004). The up-regulation of EZH2 and SUZ12 prior to entry into S-phase has been proposed to be an important step in ensuring the maintenance of H3K27me3 on target genes in dividing cells (Hansen et al., 2008). However, it is unclear how target genes are maintained repressed in quiescent cells and if H3K27me3 is required. In contrast to EZH2, EZH1 is predominantly expressed in quiescent and slow cycling tissues suggesting that EZH1 containing complexes may be important in this context (Margueron et al., 2008). Indeed differences in the relative importance of EZH1 and EZH2 in quiescent and cycling populations in vivo have been documented (Ezhkova et al., 2009; Hidalgo et al., 2012; Mochizuki-Kashio et al., 2011).

EZH1 is highly expressed in a number of quiescent adult stem cell populations forming part of the ‘quiescent stem cell expression signature’ (Cheung and Rando, 2013). Consistent with this, evidence suggests that EZH1 is more important than EZH2 in certain quiescent stem cell populations. For example, EZH1 is required to maintain adult HSCs in a slow–cycling, undifferentiated state. Deletion of EZH1 in HSCs leads to HSC loss due to an impairment of self–renewal capacity and an induction of cellular senescence (Hidalgo et al., 2012). Whereas deletion of EZH2 in adult HSCs has no deleterious effect on HSCs, most likely due to compensation by EZH1. However, deletion of EZH2 in fetal HSC populations, which cycle significantly more than their adult counterparts, leads to insufficient expansion of HSC and progenitor populations and defective erythropoiesis (Mochizuki-Kashio et al., 2011). Moreover, EZH2 is primarily expressed in epidermal progenitor cells in the skin where it is essential to maintain the proliferative capacity of these.
cells and for Polycomb target gene silencing (Ezhkova et al., 2009). Taken together these observations suggest that distinct forms of the PRC2 complex are relevant in quiescent and cycling populations and at different developmental stages in vivo.

I and others have shown that the PCL proteins are essential for PRC2 function in various contexts (Ballaré et al., 2012; Brien et al., 2012; Casanova et al., 2011; Hunkapiller et al., 2012; Musselman et al., 2012b; Walker et al., 2010). Here I sought to study the composition and regulation of PCL protein containing PRC2 complexes. I found that PCL1–3 are mutually exclusive, cell growth regulated PRC2 components, which exist in variant forms of the complex in quiescent and cycling cells. In quiescent cells, PCL1 has a dual role as the only PCL present in the PRC2 complex, and independently of the complex, it promotes a G0/G1 block by stabilizing p53. PCL2 and PCL3, like EZH2 and SUZ12, are E2F regulated genes induced in cycling cells. In cycling cells, all three PCL proteins exist in mutually exclusive forms of the PRC2 complex. The ectopic expression of PCL2 or PCL3 confers a proliferative advantage, which is dependent on direct repression of the INK4A gene. My results demonstrate that PCL proteins define alternative forms of PRC2 between quiescent and cycling cells, and that PCL proteins control cell cycle progression both, independently of, and as part of PRC2.
4.2 Results

4.2.1 PCL proteins form distinct PRC2 complexes in quiescent and cycling cells

I first sought to study the composition of PCL1–3 containing PRC2 complexes. Initial IP experiments of endogenous EZH1–2 and PCL1–3 in asynchronously growing human diploid fibroblasts (HDFs) demonstrated the PCL proteins exist in mutually exclusive PRC2 complexes (Fig. 4.1a). Moreover, they suggested that PCL1 predominantly exists within an EZH1 containing PRC2 complex, whereas, PCL2 and PCL3 predominantly exist within EZH2 containing complexes. To further dissect this possibility I performed PCL1–3 IPs in quiescent (serum starved) and cycling (serum stimulated) HDFs (Fig. 4.1b). This clearly demonstrated the existence of distinct forms of the PRC2 complex between quiescent and cycling cells, defined by different PCL members. In quiescent cells, PCL1 is the only PCL protein present and forms part of a PRC2 complex that also contains EZH1 and EED. In contrast, in cycling cells, PCL2 and PCL3 predominantly associate within EZH2–SUZ12–EED containing PRC2 complexes (Fig. 4.1b). Consistent with this, PCL1, like EZH1, is expressed predominantly in quiescent cells, while PCL2 and PCL3, like EZH2 and SUZ12, are expressed only in cycling cells (Fig. 4.2a and b). Moreover, the promoter of the PCL1 gene, like the promoter of the CDKN1A gene, is bound by p53 in quiescent cells. While the promoters of the PCL2 and PCL3 genes, like those of EZH2 and CCNB1 (Bracken et al., 2003), are bound by E2F4 in quiescent cells and E2F1 in cycling cells (Fig. 4.3). Taken together these results demonstrate that the PCL1–3 genes are divergently regulated on the transcriptional level, and their encoded proteins define distinct PRC2 complexes in quiescent and cycling cells.
Figure 4.1: PCL proteins form distinct PRC2 complexes in quiescent and cycling cells.

(a) PCL1 preferentially associates with EZH1, while PCL2 and PCL3 preferentially associate with EZH2. Immunoprecipitations using the indicated endogenous antibodies in asynchronously growing HDFs at population doubling 42. L. ex. – Long exposure (b) Distinct PRC2 complexes, defined by PCL members, exist in quiescent and cycling cells. Immunoprecipitations of endogenous PCL proteins in quiescent of cycling HDFs. HDFs were serum starved for 96 hrs (quiescent) and stimulated to re-enter the cell cycle by the addition of serum for 24 hrs (cycling).
Figure 4.2: The expression of the PCL proteins is regulated during cell growth.

(a) PCL1 is expressed in cells throughout the G0–S transition, whereas PCL2/3 are induced only in S phase. Western blot analyses of the indicated proteins in asynchronously growing HDFs (AS) or the same cells following serum starvation (0 hrs) and at the indicated timepoints following re-stimulation with serum. (b) PCL1–3 gene expression is regulated during the cell cycle. qPCR analyses of the indicated mRNA transcripts in cells from (b).
Figure 4.3: The PCL2/3 genes are downstream of the E2F pathway, while PCL1 is downstream of the p53 pathway

PCL1 is a downstream target of the p53 pathway in G0 cells, while PCL2/3 are regulated by pRB–E2F pathway. ChIP analyses of the indicated proteins in HDFs in G0 or S phase. Precipitated DNA was analyzed by qPCR with primers corresponding to the promoter regions of the indicated genes. ChIP enrichments are presented as the percentage of protein bound, normalized to input.
4.2.2 PCL proteins play divergent roles in cell growth control

To determine if PCL proteins might play divergent roles in growing and non-growing populations I examined their expression in growing and senescent HDF cultures. This demonstrated that PCL1 levels are increased in senescent HDF populations, while in contrast PCL2 and PCL3 decreased in these cells (Fig. 4.4a and b). This suggests that PCL proteins may play divergent roles in controlling cell growth. I next decided to investigate the potential regulation of the \textit{INK4A–ARF} gene locus by PCL1–3. This locus is well established as a key Polycomb target gene central to cell growth control (Bracken et al., 2007, Jacobs et al., 1999). Therefore, I examined the binding of the PCL1–3 proteins on the \textit{INK4A} and \textit{ARF} promoters in both proliferating and senescent populations. This revealed that PCL2 and PCL3, and surprisingly also PCL1, bind the promoter of the \textit{INK4A} gene in early passage cells and become displaced concomitant with transcriptional activation at later passage (Fig. 4.4c). To understand if PCL1–3 can all confer repression of the \textit{INK4A} gene, I ectopically expressed all 3 proteins in HDFs. This confirmed that ectopic expression of all three PCLs caused a reduction in p16\textsuperscript{INK4A} levels (without affecting \textit{ARF} or \textit{INK4B}), via direct recruitment of PRC2 (and PRC1) components to the promoter of \textit{INK4A}, but not to that of the negative control \textit{CCNA2} (Fig. 4.5 and Fig. 4.6). Despite this only ectopic expression of PCL2 and PCL3 resulted in a proliferative advantage. This growth advantage was associated with fewer cells staining positively for the senescence marker SA–β–Galactosidase (Fig. 4.5c). A mutated form of PCL3, with a Tudor domain that is incapable of binding modified histone tails was not targeted to the \textit{INK4A} promoter and could not confer transcriptional repression (Fig. 4.7). Taken together these results demonstrate that all 3 PCL proteins are capable of directly repressing \textit{INK4A} expression.

Next I sought to understand if the proliferative advantage conferred upon ectopic expression of PCL2 and PCL3 was dependent on p16\textsuperscript{INK4A}. Therefore, I ectopically expressed all 3 PCL proteins in HDFs following shRNA mediated depletion of \textit{INK4A} (Fig. 4.8a). This demonstrated that the proliferative advantage conferred by PCL2 and PCL3 was dependent on p16 (Fig. 4.8).
However, following depletion of *INK4A*, ectopic expression of PCL1 caused a proliferative disadvantage and an increase in the proportion of cells in G0/1 (Fig. 4.8b – lower panel). Taken together, these results suggest that PCL1 plays a distinct role from PCL2 and PCL3 in promoting a G0/1 cell cycle block.
Figure 4.4: All 3 PCL proteins bind the INK4A gene promoter and are dissociated in senescent cells.

(a) PCL1 is the predominant PCL in slow/non-growing senescent populations, while PCL2 and PCL3 are predominant in fast growing early passage populations. Western blot analyses of the indicated proteins in asynchronously growing HDFs at increasing population doublings (PD). (b) qPCR analyses of the indicated mRNA transcripts in asynchronously growing HDFs at the indicated population doublings (PD). (c) PCL1–3 bind the INK4A gene promoter in early passage cells and are dissociated in senescent populations. ChIP analyses using the indicated antibodies in cells from (a). Precipitated DNA was analyzed by PCR with primers corresponding to the promoter regions of the indicated genes, the CCNA2 gene promoter was included as a negative control. ChIP enrichments are presented as the percentage of protein bound, normalized to input.
Figure 4.5: Ectopic expression of PCL1–3 leads to reduced p16\textsuperscript{INK4A} levels, but only PCL2 and PCL3 confers a proliferative advantage.

(a) Ectopic expression of PCL1–3 leads to reduced p16\textsuperscript{INK4A} levels. Western blot analyses using the indicated antibodies in HDFs infected with the indicated pBABE retroviral expression vectors. (b) qPCR analyses of the indicated mRNA transcripts in cells from (a). (c) Ectopic expression of PCL2 and PCL3, but not PCL1, delays the onset of replicative senescence. Senescence associated-β-Galactosidase stainings in cells from (a) at PD60. (d) Quantifications of percentage of Senescence associated-β-Galactosidase positive cells in cell from (c). Quantifications were made from 2 independent biological replicates, where 200 cells were counted per experiment.
Ectopically expressed PCL proteins are recruited to the INK4A gene promoter. ChIP analyses using the indicated antibodies in HDFs infected with the indicated pBABE retroviral expression vectors at PD32. ChIPs were performed at PD48. Precipitated DNA was analyzed by PCR with primers corresponding to the promoter regions of the indicated genes, the CCNA2 gene promoter was included as a negative control. ChIP enrichments are presented as the percentage of protein bound, normalized to input.
(a) A PCL3 point mutant is not recruited to the INK4A gene promoter. ChIP analyses using the indicated antibodies in HDFs ectopically expressing BMI1, EZH2 or a wildtype (WT) or double point mutated (Mut) full-length PCL3. HDFs were infected at PD32. Precipitated DNA was analyzed by PCR with primers corresponding to the promoter regions of the indicated genes, the CCNA2 gene promoter was included as a negative control. ChIP enrichments are presented as the percentage of protein bound, normalized to input. (b) qPCR analyses of the INK4A mRNA in cells from (a).
Figure 4.8: The proliferative advantage conferred by PCL2 and PCL3 is dependent on p16\textsuperscript{INK4A}.

(a) Western blot analyses using the indicated antibodies in HDFs expressing either control (shSCR) or INK4A (shINK4A) specific shRNAs concomitantly with the indicated pBABE retroviral expression vectors. HDFs were infected with shRNA vectors at PD28 and subsequently with pBABE expression vectors at PD32. Western blots were performed at PD50 (b) Ectopic expression of PCL1 causes a proliferative disadvantage in HDFs in the absence of 3T3 growth assays (top panel) and BrdU FACs analyses (bottom panel) in cells from (a). BrdU FACs analysis was performed on 2 independent biological replicates where 10,000 cells were analysed in each experiment. *p = <0.05, ** p = <0.01, Student’s T-test.
4.2.3 PCL1, but not PCL2 or PCL3, stabilizes p53 to promote a G0/1 cell cycle block

I next sought to investigate the mechanism by which ectopic expression of PCL1 promotes a G0/1 cell cycle block in HDFs. It has been reported previously that ectopically expressed PCL1 interacts with, and stabilizes p53 to block proliferation of cancer cells (Hong et al., 2008; Yang et al., 2013). Therefore I wished to determine if this is a specific function of PCL1 and whether other PRC2 components are involved. I performed IPs of p53, PCL1–3 and EZH1–2 in asynchronously growing HDFs (Fig. 4.9). This demonstrated that PCL1 is the only PCL protein that interacts with p53. Moreover, PCL1 interacts with p53 independently of the PRC2 complex, as neither EZH1 nor EZH2 were co–purified with, or could co–purify p53. I next demonstrated that the proliferative disadvantage seen upon ectopic expression of PCL1 is completely dependent on the presence of p53. To do this PCL1 was ectopically expressed in HDFs expressing a TP53 targeting shRNA (Fig. 4.10). Strikingly, in this context, PCL1 was equally capable of conferring a proliferative advantage as PCL2 and PCL3 (Fig. 4.11b). Consistent with a specific role for PCL1 in p53 stabilization, p53 protein levels were increased only in cells ectopically expressing PCL1 (Fig. 4.10). This increase in p53 protein levels correlated with increased expression of certain p53 target genes, such as CDKN1A and PAI1, which both have roles in inhibiting cell cycle progression and stabilizing quiescence (Fig. 4.11a). This also correlated with increased p53 binding at the promoters of these genes (Fig. 4.12). Interestingly, PCL1 was not found to be co–bound at any of these sites, suggesting that PCL1 plays its role in stabilizing p53 away from chromatin (Fig. 4.12). Several p53 target genes, with roles in apoptotic cell death, did not display increases in expression. However, these genes typically exhibited negligible p53 binding in control cells suggesting that they are not relevant p53 target genes in this cellular context (data not shown).
Figure 4.9: PCL1, but not PCL2 or PCL3, interacts with p53

Western blots using the indicated antibodies on IPs of endogenous p53, PCL1-3 in nuclear protein lysates prepared from asynchronously growing HDFs (PD42).
Figure 4.10: Ectopic expression of PCL1 specifically stabilizes p53 protein levels

Western blot analyses using the indicated antibodies in HDFs expressing either control (shSCR) or TP53 (shTP53) specific shRNAs concomitantly with the indicated pBABE retroviral expression vectors. HDFs were infected with shRNA vectors at PD28 and subsequently with pBABE expression vectors at PD32. Western blots were performed at PD50.
Figure 4.11: The G0/1 cell cycle block seen upon PCL1 overexpression is dependent on p53

(a) Ectopic PCL1, but not PCL2 or PCL3, expression leads to increased p53 target gene expression. qPCR analyses of the indicated mRNA transcripts in HDFs expressing either control (shSCR) or TP53 (shTP53) specific shRNAs concomitantly with the indicated pBABE retroviral expression vectors. HDFs were infected with shRNA vectors at PD28 and subsequently with pBABE expression vectors at PD32. qPCRs were performed at PD50. (b) The PCL1 mediated G0/1 block (Fig 4.8b) is completely dependent on the presence of p53. 3T3 growth assays (left panel) and BrdU FACs analyses (right panel) in HDFs ectopically expressing PCL proteins with TP53 knockdown (shTP53). BrdU FACs analysis was performed on 2 independent biological replicates where 10,000 cells were analysed in each experiment. *p = <0.05, ** p = <0.01, Student's T-test.
Figure 4.12: PCL1 mediated stabilisation of p53 leads to increased p53 binding at target genes.

Ectopic PCL1, but not PCL2 or PCL3, expression leads to increased p53 binding at the promoters of anti-proliferative genes. ChIP analyses using the indicated antibodies in HDFs infected with the indicated pBABE retroviral expression vectors. Precipitated DNA was analyzed by qPCR with primers corresponding to the promoter regions of the indicated p53 target genes. ChIP enrichments are presented as the percentage of protein bound, normalized to input.
4.2.4 All three PCL proteins are required to prevent the onset of cellular senescence

Since all three PCL proteins are capable of directly repressing INK4A expression, I next sought to understand if they are also required to maintain the repression of INK4A. To do this I depleted PCL1–3 expression in HDFs using specific shRNAs (Fig. 4.13a, b). Similar to depletion of BMI1 or EZH2, which was reported previously (Bracken et al., 2007), depletion of all three PCL proteins lead to increased p16\(^{INK4A}\) levels and senescence onset, as evidenced by increased positivity for SA–β–Galactosidase (Fig. 4.13). Depletion of PCL1–3 also correlated with reduced Polycomb binding at the INK4A gene promoter concomitant with transcriptional activation (Fig. 4.14). Similar results were obtained following shRNA mediated depletion of Pcl1–3 in mouse embryonic fibroblasts (MEFs) (Fig. 4.15 and Fig. 4.16). Taken together these results demonstrate that PCL proteins are required to maintain Polycomb binding at the INK4A gene promoter and the prevention of cellular senescence. Consistent with a specific role for PCL1 in stabilizing p53 protein levels, depletion of PCL1, but not PCL2 or PCL3, resulted in reduced p53 protein levels in both HDFs and MEFs (Fig. 4.13a and 4.15a). Importantly, these reductions did not correlate with reduced TP53 gene expression, implying that this is indeed a direct protein stabilization affect. Reduced p53 protein levels also correlated with reduced expression of several p53 target genes, such as CDKN1A and PAI1, with roles in inhibiting cell cycle progression (Fig. 4.17). Interestingly, despite the down–regulation of these genes, the proliferative capacity of PCL1 depleted cells was still diminished most likely as a consequence of p16\(^{INK4A}\) activation.
Figure 4.13: PCL proteins are required to prevent senescence onset in HDFs

(a) Western blot analysis using the indicated antibodies in HDFs expressing the indicated pRetroSuper shRNA vectors. HDFs were infected at PD32 and western blots performed at PD38 (b) qPCR analyses of the indicated mRNA transcripts in cells from (a). (c) Senescence Associated-β-Galactosidase stainings in cells from (a). (d) Quantifications of the percentage of Senescence associated-β-Galactosidase positive cells in cells from (a). 200 cells were counted for each knockdown.
Figure 4.14: PCL proteins are required to maintain Polycomb binding at the INK4A gene in HDFs

ChIP analyses using the indicated antibodies in HDFs expressing the indicated pRetroSuper shRNA vectors. HDFs were infected at PD32 and ChIPs performed at PD38. Precipitated DNA was analyzed by qPCR with primers corresponding to the promoter regions of the INK4A gene. The CCNA2 gene is presented as a negative control. ChIP enrichments are presented as the percentage of protein bound normalized to input.
Figure 4.15: PCL proteins are required to prevent cellular senescence in mouse embryonic fibroblasts

(a) Depletion of PCL1–3 leads to increased levels of p16\textsuperscript{INK4A}. Western blot analyses in MEFs expressing the indicated shRNAs. MEFs were infected with shRNA expressing lentivirus at passage 2 and western blots performed at passage 4 (b) qPCR analyses of the indicated mRNA transcripts in cells from (a).
Figure 4.16: PCL proteins are required to maintain Polycomb binding at the INK4A gene in MEFs

ChIP analyses using the indicated antibodies in MEFs expressing the indicated shRNAs. MEFs were infected with shRNA expressing lentivirus at passage 2 and ChIPS performed at passage 4 Precipitated DNA was analyzed by qPCR with primers corresponding to the promoter regions of the Ink4a gene. The Ccna2 gene is presented as a negative control. ChIP enrichments are presented as the percentage of protein bound normalized to input.
Depletion of PCL1, but not PCL2 or PCL3, leads to reduced expression of certain p53 target genes but does not affect TP53 gene expression. qPCR analyses of the indicated transcripts in HDFs (top panel) and MEFs (bottom panel) expressing the indicated shRNAs. HDFs were infected at PD32 and qPCRs were performed at PD38. MEFs were infected at passage 2 and qPCRs performed at passage 4.
4.2.5 PCL protein containing PRC2 complexes cycle on and off chromatin in quiescent and cycling cells

I next sought to understand if the distinct forms of PCL protein containing PRC2 complexes switch on and off chromatin between quiescent and cycling cells. To investigate this, I performed ChIP analyses in quiescent and cycling HDF populations and found that only PCL1, EZH1 and EED are present on the promoters of Polycomb target genes in quiescent cells (Fig. 4.18). However, after re-entry into the cell cycle, PCL1 and EZH1 become less prominent and PCL2, PCL3, EZH2 and SUZ12 containing complexes predominate. Taken together, these results demonstrate that PCL1–3 containing complexes cycle on and off chromatin between quiescent and cycling cells.

Based on these ChIP data and the immunoprecipitation and expression analyses I propose a model for PCL1–3 function in quiescent and cycling cells (Fig. 4.19). In this model PCL1 plays a dual role in quiescent cells, maintaining repression of Polycomb target genes as part of the PRC2 complex, and stabilizing p53 independently of the complex. In cycling cells, PCL2 and PCL3 expression is induced and PCL2 and PCL3 containing PRC2 complexes maintain the repression of Polycomb target genes (Fig. 4.19).
Figure 4.18: PCL protein containing complexes cycling on and off target genes in quiescent and cycling cells

Polycomb–like proteins are differentially present on target genes in quiescent and cycling cells. ChIP analyses in quiescent (Q) and cycling (C) HDFs. Precipitated DNA was analyzed by qPCR with primers corresponding to the promoter regions of the indicated genes. ChIP enrichments are presented as the percentage of protein bound normalized to input.
Figure 4.19: Polycomb-like proteins and PRC2 in quiescent and cycling cells

Model for the regulation and function of Polycomb-like proteins in quiescent and cycling cells. In quiescent cells PCL1 containing PRC2 complexes bind Polycomb target genes, while PCL1 independently of the PRC2 complex stabilizes p53. In cycling cells, the PCL2 and PCL3 genes are activated by E2F and the PCL2 and PCL3 containing PRC2 complexes are predominant on Polycomb target genes.
4.3 Discussion

Why 3 PCL proteins with apparent functional redundancies exist has been unclear. In this chapter I demonstrate for the first time that the PCL1–3 proteins define distinct forms of the PRC2 complex between quiescent and cycling cells. This observation suggests that distinct PCL1–3 containing PRC2 complexes are required to maintain Polycomb target gene silencing at different stages of the cell cycle. Furthermore, I've shown that PCL proteins play divergent roles in cell growth control, both independently of, and as part of the PRC2 complex.

The fact that several of the core PRC2 complex components are expressed in a cell growth regulated manner implies that the composition of the complex changes significantly between growing and non-growing cells (Bracken et al., 2003; Margueron et al., 2008; Pasini et al., 2004). This has implications for understanding the mechanisms that maintain H3K27me3 and Polycomb target gene silencing in quiescent and cycling populations. In vitro the EZH1 containing PRC2 complex has been shown to possess lower methyltransferase activity compared to EZH2 containing PRC2 complexes (Margueron et al., 2008; Shen et al., 2008). Consistent with this, I observed reduced levels of H3K27me3 on Polycomb target genes in quiescent cells when compared to cycling cells. However, SUZ12, which is also absent in quiescent cells, is required for the methyltransferase activity of both EZH1–PRC2 and EZH2–PRC2 complexes (Margueron et al., 2008; Pasini et al., 2004; Shen et al., 2008). Therefore it is possible that the PCL1–EZH1–EED containing complex, which lacks SUZ12 in quiescent cells, has no histone methyltransferase activity. Therefore it will be important to characterize the enzymatic activity of the PCL1–EZH1–EED containing complex in order to understand how this complex contributes to the maintenance of Polycomb target gene repression in quiescent cells.

It will be important to examine if my observations have any in vivo relevance by examining the expression and chromatin binding of PCL proteins in quiescent normal, and cancer, stem and cycling progenitor cell populations.
Polycomb proteins are essential for maintaining the proliferative capacity of normal and cancer stem and progenitor cell populations (Ezhkova et al., 2009; Hidalgo et al., 2012; Lessard and Sauvageau, 2003; Mochizuki-Kashio et al., 2011; Molofsky et al., 2003; Park et al., 2003). This is mainly attributed to repression of the INK4A–ARF locus and prevention of cellular senescence (Bracken et al., 2007; Jacobs et al., 1999). My work suggests that distinct PCL protein containing PRC2 complexes may be relevant in maintaining INK4A–ARF repression at different stages of stem cell differentiation. This has implications for cancer therapies as the newly developed specific EZH2 chemical inhibitors, which can induce cellular senescence in cancer cells in certain contexts (Knutson et al., 2013; Qi et al., 2012), may prove ineffective for targeting quiescent cancer stem cells. Therefore, targeting the PCL1–EZH1–EED containing PRC2 complex could potentially prove more effective at inducing cellular senescence in cancer stem cells.

Interestingly, both p53 and p21 are essential for the maintenance of tissue specific stem cells (Cheng et al., 2000; Cheung and Rando, 2013; Liu et al., 2009). Deletion of the genes that encode either of these proteins leads to stem cell exhaustion owing to excessive cell cycle entry (Cheng et al., 2000; Liu et al., 2009). Taken together with my results, this suggests that PCL1 may be required to maintain stem cells in a quiescent state via its ability to stabilize p53. It will be important to study the relationship between PCL1 and p53 in tissue specific stem cells, as modulating PCL1 mediated stabilization of p53 could provide a means to induce cancer stem cell exhaustion.
5 Proteomic analyses of the PRC2 component EZH2 in pluripotent and differentiating cells
5.1 Introduction

My work, and that of others, has demonstrated that PCL1–3 are essential in facilitating PRC2 recruitment to target genes through binding the local chromatin environment (Ballaré et al., 2012; Brien et al., 2012; Casanova et al., 2011; Hunkapiller et al., 2012; Li et al., 2011; Musselman et al., 2012b; Walker et al., 2010). However, it remains unknown how this recruitment is instigated and it is likely that alternative/complimentary mechanisms also contribute to PRC2 recruitment to target genes during cell fate transitions (Bracken and Helin, 2009). In *D. melanogaster* several DNA binding transcription factors, such as Pleiohomeotic (PHO) and the related protein PHO-like (PHOL) (Schwartz and Pirrotta, 2007), associate with the PRC2 complex, recruiting it to target sites. These transcription factors bind DNA sequence elements known as Polycomb Response Elements (PREs). Genetic studies have clearly established that both PREs, and the transcription factors that bind them, are essential for PRC2 recruitment. This demonstration sparked an interest in identifying analogous transcription factors and DNA elements in mammals. However, initial *in silico* attempts at defining mammalian PREs based on sequence homology to those in *D. melanogaster* failed to identify any such elements. Moreover, genome–wide mapping of YY1, the mammalian homologue of PHO, demonstrated in mouse ES cells that YY1 is not localized on Polycomb target genes (Vella et al., 2011). These data suggest that this mechanism of PRC2 recruitment is not conserved in mammals. However, two recent studies have identified candidate mammalian PREs, which are capable of recruiting Polycombs *in vivo* (Sing et al., 2009; Woo et al., 2010). Bioinformatic analyses of these PREs failed to predict any DNA sequence similarity to *D. melanogaster* PREs. However, close inspection of the elements demonstrates some consistent features, such as the presence of multiple different transcription factor binding motifs *vivo* (Sing et al., 2009; Woo et al., 2010). This is significant as it demonstrates that PRE elements do in fact exist in mammals and that the sequence of these PREs has significantly diverged from their ancestral elements. Taken together, this implies that many additional transcription factors are involved in PRE binding.
and PRC2 recruitment. Indeed several mammalian transcription factors, such as E2F6, PLZF, MSX1, GATA1 and IKAROS, have been reported to interact with various Polycomb group proteins and to be involved in Polycomb recruitment to target genes during cell fate decisions (Barna et al., 2002; Boukarabila et al., 2009; Bracken et al., 2007; Trimarchi et al., 2001; Wang et al., 2011; Yu et al., 2009).

Genetic studies have clearly demonstrated that many transcription factors play essential roles during lineage decisions in development and differentiation (Arnold and Robertson, 2009; Orkin and Zon, 2008). Members of the FOX, PAX and GATA transcription factor families' function in multiple different lineages to regulate cell fate decisions both in stem cells and during differentiation (Kouros-Mehr et al., 2008; Myatt and Lam, 2007; Robson et al., 2006). These 'cell fate' transcription factors regulate lineage decisions by 'instructing' the establishment of cell type specific transcriptional programs. We hypothesize that these factors may be involved in PRC2 recruitment mechanisms, directing and displacing the complex from target genes during lineage transitions (Bracken and Helin, 2009).

Interestingly the PRC2 complex has recently been demonstrated to directly methylate non–histone substrates (He et al., 2012). For example, the transcription factor GATA4 is methylated in vivo, and GATA4 interacts with EZH2 in the developing heart. In vitro methyltransferase assays demonstrated that GATA4 can be directly methylated by the PRC2 complex. In vivo this methylation attenuates the ability of GATA4 to activate the expression of genes required for heart development, possibly by reducing its interaction of with activating co–factors (He et al., 2012). EZH2 was also recently shown to interact with the transcription factors STAT3 and RORα (Kim et al., 2013; Lee et al., 2012). Interestingly, both of these proteins are methylated in an EZH2 dependent manner, and these methylations have functional consequences for both factors. However, it remains to be clarified whether EZH2 directly mediates methylation of either protein. These examples illustrate the possibility that EZH2 associated proteins may in fact be substrates for
PRC2/EZH2 mediated methylation. Furthermore, they demonstrate that EZH2 can also control gene expression indirectly by methylating other chromatin associated factors.

In this chapter I sought to identify the transcription factors (and other chromatin regulators) that interact with the PRC2 component, EZH2, in pluripotent stem cells and during their differentiation. The hypothesis is that these factors may play an instructive role in PRC2 recruitment to target genes. Alternatively, they could themselves be methylated by EZH2. To this end, I performed large-scale endogenous immunoprecipitation experiments of EZH2 using a potent and specific monoclonal antibody (Bracken et al., 2006) and subjected the immunoprecipitates to tandem mass spectrometry. These experiments were initially performed in pluripotent mouse ES cells, and later expanded to a dynamic model of directed differentiation. This utilized the retinoic acid induced differentiation of pluripotent NT2/D1 human embryonic carcinoma cells towards a neuronal cell fate. Importantly, I correlated NT2/D1 differentiation with displacement of Polycombs from genes required for neuronal cell fate, such as the \textit{HOXA} genes, and a concomitant recruitment to pluripotency genes such as \textit{NANOG} and \textit{POU5F1} (also known as \textit{OCT4}). These experiments have massively expanded the known EZH2 protein interaction network and identified many transcription factors with roles in cell fate decisions, which may be involved in cell type specific recruitment of PRC2. Significantly, this is the only example to date of such a study being undertaken on an endogenous level and within the context of a dynamic cell fate model. Offering unprecedented opportunities to garner insights into PRC2 biology.
5.2 Results

5.2.1 Interaction proteomics of EZH2 in pluripotent mouse ES cells

I sought to identify EZH2 associated proteins in mouse ES cells. To do this I first established an immunoprecipitation (IP) procedure capable of efficiently purifying endogenous EZH2, and associated PRC2 components, from mouse ES cell nuclear lysates. Significantly, the presence of Histone H3 in my EZH2 IPs demonstrated that this procedure also allows for the purification of EZH2 bound chromatin (Fig. 5.1a). Thereby allowing the identification of potentially functionally relevant proteins/protein complexes co-bound within the same chromatin environment as EZH2, but which do not necessarily interact with EZH2 directly. Using this IP procedure I performed large-scale purifications of EZH2 from mouse ES cell nuclear lysates, and subjected the purified samples to silver stain and mass spectroscopy analysis (Fig. 5.1b and c). Mass spectroscopy results confirmed that all three PRC2 core components were highly represented within purified protein samples. In addition to these core PRC2 components most other previously described PRC2 members including, EZH1, RBBP4, RBBP7, AEBP2, JARID2 and PCL2, were highly represented. Significantly, although being undetectable by Western blot (Fig. 5.1a), members of the PRC1 complex (RING1A, RING1B, CBX7 and RYBP) were also detectable in the mass spec analysis. These proteins most likely do not directly interact with EZH2, but instead are present on the same nucleosomes via the interaction of CBX7 with H3K27me3. This suggests that in addition to identifying proteins that interact directly with EZH2, that this method can also identify functionally relevant yet indirect associations, likely mediated through co-localisation on the same nucleosomes.

This analysis identified a large cohort of DNA binding transcription factors and other chromatin regulators that associate with EZH2 (Fig. 5.1c). Reassuringly, several of the proteins identified, such as UTF1, SALL4 and CHD4, have previously been reported to associate with PRC2. Genome-wide ChIP analyses of the stem cell specific transcription factor UTF1 demonstrated that this factor co-binds the majority of PRC2 target genes in mouse ES cells (Jia
et al., 2012; Kooistra et al., 2010). Moreover, UTF1 regulates PRC2 binding at these shared sites. Similarly, the ES cell transcription factor SALL4 has been demonstrated to co-exist with the PRC2 complex at many sites in the mouse ES cell genome (Yang et al., 2008). The Nucleosome Remodeling and Deacetylase (NuRD) complex, a protein complex containing CHD4, has been linked to PRC2 silencing/recruitment mechanisms. NuRD and NuRD-like complexes have been shown to co-bind a cohort of PRC2 target genes, and the complex(es) potentially contribute to PRC2 recruitment by deacetylating the local chromatin environment (Egan et al., 2013; Reynolds et al., 2011; Sparmann et al., 2013). Moreover, CHD4 has been reported to associate directly with the PRC2 complex (Sparmann et al., 2013). Finally this screen also identified GATA4 (a known non-histone EZH2 substrate) as an EZH2 associated protein in ES cells. Taken together, this demonstrates that this approach has been successful in identifying known PRC2 associated proteins, implying that many of the additional proteins identified may play a important roles in EZH2 biology.
Figure 5.1: The EZH2 protein interactome in mouse ES cells

(a) Western blot analysis of the indicated proteins in triplicate EZH2 and control (IgG) immunoprecipitations performed on mouse ES cell nuclear protein lysates. (b) Silver stain analysis of the purifications in (a). (c) EZH2 protein interaction network identified by mass spectrometry. All proteins presented were detected in at least 2 of the 3 EZH2 purifications, while remaining undetected in control purifications.
5.2.2 Interaction proteomics of EZH2 in a dynamic cell fate model

In a complementary approach, I extended my analyses of the EZH2 protein interaction network to the pluripotent human embryonic teratocarcinoma cell line NT2/D1. These cells are molecularly and functionally similar to human ES cells and importantly can be differentiated along a neural lineage upon treatment with all trans-retinoic acid (ATRA) (Lee and Andrews, 1986; Schwartz et al., 2005). I sought to use this model system to identify the EZH2 protein interaction network both in pluripotent cells and during differentiation. Following ATRA treatment of NT2/D1 cells, pluripotency associated genes such as *NANOG* and *POU5F1* (also known as *OCT4*) are downregulated, with a concomitant upregulation of the genes associated with neuronal cell fate such as *HOXA3* and *ZIC1* (Fig. 5.2a and data not shown). Importantly, I correlated these expression changes with active recruitment and displacement of PRC2 and PRC1 components from these genes (Fig. 5.2b and data not shown).

I next sought to identify all transcription factors and chromatin regulators that associate with EZH2 during ATRA induced neuronal differentiation of NT2/D1 cells. Using my immunoprecipitation procedure I purified endogenous EZH2 from nuclear protein lysates prepared from undifferentiated NT2/D1 cells and cells treated with ATRA for 2, 4 and 8 days, respectively. These purified samples were subjected to Western blot, silver stain and mass spectroscopy analysis (Fig. 5.3).
Figure 5.2: A dynamic model system to study PRC2 recruitment.

(a) qPCR analyses of the indicated transcripts in NT2/D1 cells at the indicated time-points following treatment with 10mM ATRA. (b) Quantitative-ChIP analyses using the indicated antibodies in NT2/D1 cells at the indicated time-points after ATRA treatment. The CCNA2 promoter was included as a negative control. Precipitated DNA was analyzed by qPCR with primers corresponding to the promoter regions of the indicated genes. ChIP enrichments are presented as the percentage of protein bound, normalized to input.
Figure 5.3: Purifications of EZH2 in pluripotent and differentiating NT2/D1 cells.

Western blots for the indicated proteins (left) and silver stain analysis (right) of triplicate EZH2 and control IgG purifications performed in undifferentiated NT2/D1 cells nuclear protein lysates and the same cells at the indicated time points following treatment with 10mM ATRA.
This mass spectroscopy analysis confirmed that all three PRC2 core components were highly represented within purified protein samples. In addition to these, all other previously described PRC2 members including, RBBP4, RBBP7, AEBP2, JARID2, PCL1, PCL2 and PCL3 were highly represented (Fig. 5.4). Moreover, JARID2 which has been shown to be a component of the PRC2 complex specifically in undifferentiated ES cells (Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010; Shen et al., 2009) only co-purifies with EZH2 in undifferentiated NT2/D1 cells and during the early stages of differentiation (Fig. 5.4). This is a significant observation as it confirms that this approach is capable of identifying cell type specific EZH2 interactions.

This approach also identified a large cohort of DNA binding transcription factors, which co-purify with EZH2 in undifferentiated and/or differentiating cells (Fig. 5.4). Significantly, 10 of these factors (CEBPZ, PHF14, ZNF593, ZNF687, UHRF1, HELLS, HMGA2, ZNF532, HMG20A and SALL4) were also found to associate with EZH2 in mouse ES cells (Fig. 5.1 and 5.4). Interestingly, a cohort of factors associated with EZH2 specifically in undifferentiated cells (Fig. 5.4 “Early Interactors”). Amongst these were SALL4, ZNF532 and CEBPZ, which also associated with EZH2 in mouse ES cells. As previously stated SALL4 is an important ES cell transcription factor, that co-binds many target genes with PRC2. Taken together this suggests that ZNF532 and CEBPZ may also be important factors in ES cell biology potentially associating with PRC2.

I also identified a smaller cohort of factors that associate with EZH2 in a stable fashion, both in undifferentiated cells and during neuronal differentiation (Fig. 5.4 “Steady Interactors”). Several of these, such as PHF14, HMG20A and ZNF593, also associated with EZH2 in mouse ES cells. These factors could potentially have more general EZH2 associated functions and hence do not interact with EZH2 in a cell type specific manner, at least in this experimental context. Finally, I identified several factors that interact with EZH2 specifically in differentiated NT2/D1 cells, most notably members of the
SMAD and PBX families (Figure 5.4, “Late Interactors”). The SMAD proteins have well-established roles during development as mediators of external signaling pathways such as the TGF–β pathway (Arnold and Robertson, 2009). The PBX transcription factors function during cell fate decisions and significantly PBX1, which co-purifies with EZH2 only at Day 8 of neuronal differentiation, has been shown to be required for differentiation of mouse P19 teratocarcinoma cells (an analogous cell model to NT2/D1) (Qin et al., 2004). PBX1 depleted cells were shown to be incapable of undergoing ATRA induced neuronal differentiation and this defect was associated with an inability to repress the expression of genes associated with pluripotency. This leads to the exciting possibility that PBX transcription factors may mediate the recruitment of PRC2 to, and silencing of, target genes required for stem cell function during differentiation.
Figure 5.4: The EZH2 protein interaction network during neuronal differentiation of NT2/D1 cells.

Triplicate affinity purifications of EZH2 (see Fig. 5.3) were subjected to analysis by mass spectroscopy. All proteins presented were in at least 2 of 3 experiments and were undetected in controls. Rows depict individual proteins detected in EZH2 purifications and scales are representative of the number of unique non-redundant peptides mapping to each protein. Transcription factors and chromatin regulators co-purifying with EZH2 throughout the differentiation time-course or specifically in the early or later stages of differentiation are labeled as 'Steady', 'Early' or 'Late Interactors' respectively.
5.2.3 Validation of EZH2 interacting proteins

My current efforts are focused on validating the association of EZH2 with a cohort of the newly identified EZH2 protein interaction network. Parallel work in the lab has already validated the association of EZH2 with CHD4, SMC1A, DNMT3A, TAF6 and ZNF593 at an endogenous level. Furthermore, ZNF593 has been found to co-localize on PRC2 target genes in mouse ES cells demonstrating that this approach has identified previously unappreciated EZH2 associated proteins (data not shown). Thus far I have validated the association of exogenously expressed FLAG-tagged CEBPZ and FLAG-tagged PHF14 with EZH2 in HEK293 cells (Fig. 5.5). Follow on experiments will examine whether these proteins interact with EZH2 at the endogenous level, in addition to attempting to validate several more of my EZH2 associated proteins. Ultimately the goal will be to delineate the molecular relationship between EZH2 and these newly identified interaction partners.
**Figure 5.5: FLAG tagged CEBPZ and PHF14 associate with EZH2.**

Western blots of immunoprecipitates showing that endogenous EZH2 associates with FLAG-HA-CEBPZ (Top panel) and FLAG-HA-PHF14 (Bottom panel) in HEK293 cells. Immunoprecipitations were performed on Control (uninfected) HEK293 cells, or cells infected with lentivirus expressing FLAG-HA-tagged CEBPZ or PHF14.
5.3 Discussion

In this chapter I have performed the first ever large-scale, endogenous proteomic interaction analysis of EZH2 in pluripotent cells and during differentiation. By coupling immunoprecipitations of EZH2 with mass spectroscopy, I have hugely expanded the known EZH2 protein interaction network. Many of these interactions were seen to be cell type specific, and EZH2 interactors can be divided into three subgroups. 1. Those that interact in undifferentiated cells. 2. Those that interact in differentiating cells. 3. Those that stably associate with EZH2 in all experimental contexts examined. Excitingly, these data suggest that many of these factors may be involved in EZH2 biology in a cell type specific manner. Importantly, my own work, and that of others in the lab, has now validated many of these novel EZH2 interactors, demonstrating in some cases co–localization with EZH2 on target genes in mouse ES cells (data not shown). This demonstrates that many of these factors are bona fide EZH2 interactors, and current work in the laboratory is focused on delineating the molecular roles of these factors in the context of EZH2 function.

The implications of these results (and of our follow up studies) are significant as the mechanisms underlying the recruitment of PRC2 both in stem cells and during differentiation are unclear. Understanding these mechanisms will contribute to our fundamental understanding of development and differentiation. Emerging evidence points to a role for lineage specific DNA binding transcription factors in regulating PRC2 targeting during cell fate decisions (Bracken and Helin, 2009). Significantly, many of these cell type specific transcription factors are deregulated in cancer. How this deregulation contributes to oncogenesis is poorly understood, however evidence suggests that their normal roles in the regulation of cell fate decisions are perturbed (Garraway and Sellers, 2006; Mullighan et al., 2008; Nerlov, 2007; Ramsay and Gonda, 2008; Zhang et al., 2008). Taken together, this raises the possibility that deregulation of transcription factors involved in regulating cell fate may lead to aberrant PRC2 targeting during differentiation. Therefore, delineating the functional interactions between EZH2, and these newly
identified EZH2 associated proteins, could provide important insights into molecular events leading to cancer.

Recently it has emerged that EZH2 can directly methylate non–histone substrates. This raises the possibility that many of the candidate EZH2 associated proteins may in fact be EZH2 substrates, not necessarily involved in PRC2 recruitment. Several other enzymes originally thought of as epigenetic histone modifiers, particularly acetyltransferases and deacetylases, have also been shown to modify non–histone proteins (Boyes et al., 1998; Trivedi et al., 2010; Yang and Seto, 2008). These modifications, like EZH2 mediated methylation of GATA4, can have profound functional consequences for the target protein (He et al., 2012). This highlights that EZH2 may have significant functional inputs independent of H3K27me3 and direct target gene repression. Therefore, the identification of EZH2 substrates, within the EZH2 interaction network, may yield new insights into the role of EZH2 during cell fate decisions and oncogenesis.
6 General Discussion
6.1 Introduction

The goal of the work presented in this thesis was to begin to delineate the molecular mechanisms that contribute to PRC2 recruitment. The PRC2 complex is of fundamental importance to lineage specification during development and differentiation (Margueron and Reinberg, 2011). The importance of the PRC2 complex in early development is underlined by embryonic lethality in Ezh2\(^{-/-}\), Eed\(^{-/-}\) and Suz12\(^{-/-}\) knockout mice (Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2004). These mice typically die around embryonic day E5.5 as the early stage embryo is incapable of differentiating into more specialized tissues. Deletion of PRC2 components in differentiating adult tissues generally leads to defective tissue regeneration owing to impaired differentiation and/or loss of adult stem cell populations (Ezhkova et al., 2009; Hidalgo et al., 2012; Margueron and Reinberg, 2011; Mochizuki-Kashio et al., 2011). Significant open questions still exist regarding PRC2 function during development and differentiation. The mechanisms that govern the dynamic recruitment of the PRC2 complex during lineage differentiation are still largely unclear (Bracken and Helin, 2009). Delineating these mechanisms is fundamentally important for our understanding of lineage specification and could also aid in the development of regenerative therapies. PRC2 function is also frequently deregulated in cancer (Margueron and Reinberg, 2011), and PRC2 target genes are often aberrantly silenced in cancer cells. This suggests that normal PRC2 recruitment mechanisms may become perturbed during oncogenesis (Bracken and Helin, 2009). Therefore, improving our understanding of the molecular mechanisms regulating PRC2 recruitment has implications for both regenerative medicine and cancer treatment. Targeted modulation of PRC2 recruitment mechanisms could be utilized to aid reprogramming of cells for regenerative therapies, or to re-activate aberrantly silenced gene expression programs in cancer cells.

Initial efforts at delineating PRC2 recruitment mechanisms sought to study the molecular role of the ancillary PRC2 component PCL3. PCL3 had previously been implicated in the maintenance of PRC2 at target sites in mouse ES cells.
(Hunkapiller et al., 2012), however the molecular mechanisms involved were unclear. I found that PCL3 utilizes an N-terminal tudor domain to read modified histone tails, including the H3K27me3 and H3K36me3 modifications. Mechanistically, these interactions are essential for the ability of PCL3 to facilitate both the maintenance of PRC2 binding in ES cells and the de novo recruitment of the complex during lineage transitions. This work revealed the first ever example of a histone reading capability within PRC2 being essential for de novo complex recruitment.

PCL3 is one of three (PCL1–3) related proteins in mammals that have all been proposed to execute similar roles, mainly related to PRC2 complex recruitment (Ballaré et al., 2012; Brien et al., 2012; Casanova et al., 2011; Hunkapiller et al., 2012; Musselman et al., 2012a; Walker et al., 2010). Previously, it was unclear why three PCL proteins exist in mammals given their apparent functional redundancies. Therefore, I next sought to study the composition of PCL protein containing PRC2 complexes. I found that PCL1 preferentially associates with an EZH1 containing PRC2 complex in quiescent cells, while PCL2 and PCL3, which are expressed only in cycling cells, preferentially associate with EZH2 containing complexes in this setting. This observation suggests that PCL proteins may contribute to PRC2 recruitment/maintenance mechanisms in different (cycling and non–cycling) cell types and/or at different stages of the cell cycle in any given cell. Moreover, I also found that PCL1 plays a PRC2 complex independent role in stabilizing p53 and affecting a G0/G1 cell cycle arrest. These observations have significant implications for the role of PCL proteins in quiescent stem and cycling progenitor cells in vivo which are currently being investigated.

Trans–acting DNA binding transcription factors, such as PHO and PHO–L, are essential for PRC2 recruitment in D. melanogaster, and it has been proposed that similar factors may play a role in PRC2 recruitment in mammals (Bracken and Helin, 2009). Consistent with this, genetic evidence in several systems indicates that transcription factors are required to facilitate PRC2
recruitment to target genes during cell fate transitions. However, biochemical evidence for these apparent interactions is still lacking. Interestingly, recent evidence also indicates that EZH2 can directly methylate non-histone substrates, such as the transcription factor GATA4 (He et al., 2012). This observation is significant since it suggests that EZH2 can influence gene expression independently of its canonical role as a H3K27me3 methyltransferase. This is a novel and emerging concept and the identification of additional non-histone EZH2 substrates will undoubtedly yield significant insights into the role of EZH2 beyond H3K27me3. In this context, I undertook to perform the first ever large scale, endogenous proteomic analysis of EZH2 in pluripotent stem and differentiating cells. These analyses identified a large cohort of transcription factors and other chromatin associated proteins that associate with EZH2. Importantly, several previously described interactions were identified suggesting that this approach is capable of finding bona fide EZH2 interactors. Preliminary follow up experiments have validated a panel of these interactions at an exogenous and/or endogenous level and current work is focused at delineating the role of these factors in EZH2 biology. It is expected that these follow up studies will reveal significant insights into the role of these factors in EZH2 biology.

6.2 Do PCL proteins instruct PRC2 recruitment?

One of the key open questions in Polycomb biology in mammals is what are the factors/mechanisms that instruct PRC2 recruitment to target genes during lineage transitions (Bracken and Helin, 2009). I have shown that PCL3 is essential for facilitating PRC2 maintenance at target sites in mouse ES cells, and de novo recruitment during ES cell differentiation. However, based on the current data it is difficult to conclude if PCL3 plays an instructive role in these processes. For instance, a large cohort of genes are marked by H3K36me3 in mouse ES cells but none of these are bound by PCL3. Moreover, PCL3 is recruited only to a subset of these genes during differentiation in order to confer transcriptional repression. This demonstrates that the ability of the PCL3 tudor domain to read H3K36me3 modified nucleosomes, although being
required, is not in itself sufficient to instruct recruitment. Implying that additional layers of regulation that either positively or negatively affect PCL3 recruitment must exist. One potential mechanism could relate to the combinatorial readout of histone modifications in the local chromatin environment surrounding H3K36me3 marked nucleosomes. It has been proposed that cooperative engagement of adjacent modifications, within or between nucleosomes, by proteins with multiple reader domains may provide an additional layer of specificity for factor recruitment (Ruthenburg et al., 2007b). Significantly, the reading capabilities of the tandem PHD domains of PCL3 have not yet been examined. Therefore, it will be interesting to ascertain if these domains can binding modified histones as this could further indicate whether PCL3 has either an instructive, or a more passive facilitating role in complex recruitment. Moreover, the PHD domains of PCL1–3 exhibit significantly less sequence conservation than the tudor domain. This highlights the possibility of functional divergence in their putative reading capabilities. This would raise the intriguing potential for modulation of PRC2 recruitment by substituting different PCL protein components into the complex. Therefore, it will be important to determine the reading capabilities of all three reader domains within PCL1–3, to better understand the role of these proteins in mediating PRC2 recruitment. Alternatively, lineage specific transcription factors that interact with PCL containing PRC2 complexes or long ncRNAs could provide an additional layer of specificity for complex recruitment (Bracken and Helin, 2009).

6.3 PCL proteins in adult stem and progenitor populations

In an adult organism tissue homeostasis is maintained by tissue specific stem cells, most tissues contain long–lived undifferentiated stem cells that can proliferative and differentiate to compensate for tissue loss throughout the life of the organism. These cells have the unique capacity to produce proliferating/differentiating progenitor cells while retaining their own stem cell identify by self–renewal (Weissman, 2000). Tissue specific stem cells generally persist in a quiescent state for prolonged periods of time, and upon stimulation to differentiate asymmetrically divide to produce one daughter cell
that retains stem cell identity and another committed progenitor cell. This progenitor cell rapidly proliferates, amplifying itself before eventually exiting the cell cycle and terminally differentiating. My observations on the expression patterns of PCL1–3 imply that the PCL1 containing PRC2 complex is likely to be the main form of PRC2 present in quiescent stem cells and terminally differentiated cells. Indeed, analysis of the expression profiles of several quiescent stem cell populations indicate that PCL1 (and EZH1) expression is enriched in quiescent stem cells in vivo (Cheung and Rando, 2013). In contrast, my results suggest that PCL2 and PCL3 will be predominantly expressed in cycling progenitor populations. It will be important to confirm this by analyzing both the expression and chromatin binding dynamics of PCL1–3 in purified stem and progenitor populations in order to begin to unravel the role of PCL protein containing PRC2 complexes in vivo. My data also suggest that PCL1 may directly regulate stem cell quiescent independently of PRC2. Both p53 and p21 have been demonstrated to be essential in reinforcing stem cell quiescence (Cheng et al., 2000; Liu et al., 2009). Deletion of either gene in HSCs results in stem cell exhaustion owing to excessive proliferation. Taken together with my results, this implies that PCL1 may also be an essential regulator of stem cell quiescence through stabilizing p53 to promote a G0/1 cell cycle block. Ultimately, tissue specific conditional PCL1/2/3 knockout mouse models will be required to fully delineate the roles of PCL1–3 in stem and progenitor cells in vivo.

6.4 PCL proteins as targets for cancer therapy

PCL1–3 are required for the repression of INK4A and prevention of senescence. Moreover, ectopically expressed PCL proteins are capable of repressing INK4A expression and conferring a proliferative advantage to cells, which in the case of PCL1 requires the absence of p53. This suggests that, in the right context, aberrations of PCL protein function could potentially directly contribute to oncogenesis by repressing INK4A and providing a growth advantage. Indeed PCL1 is trans–located and PCL3 is aberrantly expressed in several different cancer types (Sauvageau and Sauvageau, 2010). However if remains to be determined if these aberrations contribute to cancer
development. In any case, my data demonstrates that PCL proteins are required to prevent senescence onset in normal cells. In addition, preliminary evidence also indicates that PCL proteins are also required to prevent cancer cell senescence (data not shown), implying that targeting PCL proteins in cancer cells could be exploited for therapeutic gain. Interestingly, I have seen that the PCL3 tudor domain is essential for INK4A repression; given the demonstrated functional conservation in the tudor domains of PCL1–3, it remains likely that the PCL1 and PCL2 tudor domains are also essential for INK4A repression (Ballaré et al., 2012; Brien et al., 2012; Cai et al., 2013; Musselman et al., 2012a) and current work is aimed at resolving this possibility. However the preliminary data already highlights the potential utility of inhibiting the PCL tudor domain as a means of inducing cancer cell senescence. Chemical inhibitors of EZH2 methyltransferase activity have recently been developed (McCabe et al., 2012), and have been used in certain contexts to induce senescence in cancer cells (Knutson et al., 2013; Qi et al., 2012). However, the fact that EZH2 is only expressed in cycling cells is significant, as it indicates that EZH2 specific inhibitory drugs will not be effective against quiescent cancer stem cells. Therefore, an alternative/complimentary approach will be needed to inhibit PRC2 function in cancer stem cells. My work highlights that chemical inhibitors of the PCL tudor domain, in particular that of PCL1, may be an excellent alternative cancer therapeutic strategy.

6.5 Non–histone substrates for EZH2

The emerging concept that EZH2 can methylate non–histone proteins to modulate their activity, illustrates a whole new aspect of EZH2 biology that has been previously unappreciated. Identifying and studying additional EZH2 substrates will undoubtedly yield significant insights into the role of EZH2 in development, differentiation and cancer beyond its canonical role as a H3K27me3 methyltransferase. Significantly my proteomic analyses of EZH2 identified GATA4, a known non–histone EZH2 substrate (He et al., 2012), as an EZH2 associated protein. This implies that additional EZH2 substrates exist among these proteins. It will be important to identify EZH2 substrates
within this cohort in order to begin to understand the role of EZH2 outside of H3K27me3.

6.6 Conclusions

The data in this thesis support a number of conclusions, specifically that:

1. PCL3 is an essential PRC2 co-factor that contributes to complex recruitment during lineage differentiation through interactions with modified nucleosomes.

2. PCL1–3 associate with distinct forms of the PRC2 complex in quiescent and cycling cells.

3. PCL proteins have distinct and overlapping roles in controlling the cell cycle, both as part and independently of the PRC2 complex.

4. EZH2 associates (in some cases dynamically) with a large cohort of transcription factors and other chromatin regulators in stem and differentiating cells.
References


Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y.,
Garbe, J.C., Bhattacharya, S., Merchant, B., Bassett, E., Swisshelm, K.,
between Stasis and Telomere Attrition Senescence Barriers Shown by Long­
term Culture of Normal Human Mammary Epithelial Cells. Cancer Res 69,
7557–7568.
Gascar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J.,
Heidersbach, A., Ramalho-Santos, J.P., McManus, M.T., Plath, K., Mesheror,
E., et al. (2009). Chd1 regulates open chromatin and pluripotency of
Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray,
Gonzalez, M.E., Li, X., Toy, K., DuPrie, M., Ventura, A.C., Banerjee, M.,
Ljungman, M., Merajver, S.D., and Kleer, C.G. (2009). Downregulation of
EZH2 decreases growth of estrogen receptor-negative invasive breast
ca fine and requires BRCA1. Oncogene 28, 843–853.
Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlautz,
and H3K4 by an MLL complex are mutually exclusive. Nature 449, 933–937.
Guenther, M.G., Lawton, L.N., Rozovskiaia, T., Frampton, G.M., Levine, S.S.,
Volkert, T.L., Croce, C.M., Nakamura, T., Canaani, E., and Young, R.A.
(2008). Aberrant chromatin at genes encoding stem cell regulators in human
Gui, Y., Guo, G., Huang, Y., Hu, X., Tang, A., Gao, S., Wu, R., Chen, C., Li,
X., Zhou, L., et al. (2011). Frequent mutations of chromatin remodeling genes in
Van Haaften, G., Dalgliesh, G.L., Davies, H., Chen, L., Bignell, G., Greenman,
C., Edkins, S., Hardy, C., O'Meara, S., Teague, J., et al. (2009). Somatic
mutations of the histone H3K27 demethylase gene UTX in human cancer.
Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad,
He, A., Shen, X., Ma, Q., Cao, J., von Gise, A., Zhou, P., Wang, G., Marquez,
V.E., Orkin, S.H., and Pu, W.T. (2012). PRC2 directly methylates GATA4 and
demethylase Jhd, Jhd, M. i, Kud, Kd, 2b regulates cell proliferation and senescence
Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp,
modifications at human enhancers reflect global cell-type-specific gene


lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature 453, 110–114.


