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.

Identification and characterisation of a novel Polycomb Repressive Complex 2 associated protein that is an alternatively spliced product of the *LCOR* gene locus

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

2015

Emilia Jerman

Thesis supervisor: Dr. Adrian P. Bracken

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

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

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

Summary

Polycombs are evolutionary conserved epigenetic regulators crucial for specification of cell types during development. They assemble in multiprotein complexes to modify amino terminal tails of histone H3 at lysine 27 to regulate the expression of underlying genes. PRC2 mediates repression by catalysing tri-methylation of histone H3 at lysine 27 (H3K27me) at the promoters of master transcriptional regulators, such as *HOX* and cell cycle and proliferation regulating genes. Recently, it emerged that PRC2 also mediates the mono-and di-methylation of H3K27 and that these histone marks are associated with gene activation and enhancer silencing, respectively. However, the mechanisms of PRC2 recruitment to and repression of their target genes, as well as its function on the active genes remain an area of active research. This work focused on comparing the biological and biochemical properties of the sub-stoichiometric PRC2 associated proteins Polycomblike 1-3 (PCL1-3) in order to gain further insight into the molecular mechanisms of the PRC2 activity. Gene expression analyses in a model of replicative senescence revealed that in contrast to PCL2 and PCL3, PCL1 mRNA expression is sustained in cells with arrested growth, implicating functional specialisation of PCL1-3 proteins. Immunoprecipitation analyses performed to further explore the functions of PCL1-3 proteins, demonstrated that PCL1, but not PCL2 or PCL3, interacts with tumour suppressor p53 independently of PRC2, indicating a unique role of PCL1 in non-proliferating cells. Furthermore, these analyses lead to identification of a novel PCL1 and PCL2 interacting protein called 'BIG', which I characterise to be a product of alternative splicing of the ligand dependent co-repressor *LCOR* gene locus encompassing the *C10ORF12* predicted gene. I demonstrate that 'BIG' is a sub-stoichiometric

component of the PRC2 complex and is required specifically for the global mono- and di-, but not tri-methylation of histone H3 at lysine 27 by PRC2. 'BIG' is the first PRC2 component discovered to modulate the PRC2 action towards the activating H3K27me1 modification. Furthermore, I demonstrate that 'BIG' associates with estrogen receptor alpha in both the presence and absence of the agonist E2 and is required for proliferation of the breast cancer cell line MCF7. The ability of 'BIG' to modulate H3K27me1 and H3K27me2 deposition by PRC2 makes it a prime candidate for further investigation in elucidating the dynamic mechanism of PRC2 activity on both active and repressed genes.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Dr. Adrian Bracken, for his mentorship and support, be it a professional or personal hardship, over the course of this work. I would also like to extend my gratitude to Dr. Gerard Cagney, Giorgio Olivero and Kieran Wynne for their input, expertise and collaboration in the mass spectrometry analysis experiments. I am deeply grateful to my thesis committee, Prof. Kevin Devine and Prof. David McConnell, and the Head of Genetics, Prof. Tony Kavanagh, for their support and for assistance in the financial questions towards the end of this work. Also a thanks is due to Indigo Pratt Kelly for the supporting data related to some aspects of chapters 4 and 5 and her invariable high spirits.

I would like to thank all members of the Bracken laboratory, past and present. A special thanks to Dr. Fiona Lanigan and Dr. Gerard Brien for all the training in the early years. I am deeply grateful to Dr. Gundula Streubel for the stimulating and encouraging discussions we had throughout the years. I would also like to thank Carol O'Brien and Eric Conway for their much needed grammar advice and enjoyable scientific, and non-scientific discussions.

This thesis would not be possible without the funding awarded to me by the Irish Research Council (IRC), for which I would like to express my gratitude.

I would like to thank my friends and family for their love and continuous support. Especially, I would like to thank my fiancé Mark, for his patience, willing ear and unconditional support.

Finally, I will be eternally grateful to my late parents, Anna and Jurij Larkin, for always believing in me and giving me a life filled with opportunities.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Abbreviations

AEBP2	AE binding protein 2
ARF	Alternate open reading frame
ATP	Adenosine triphosphate
BMI1	B-cell-specific Moloney murine leukaemia virus integration site 1
BSA	Bovine serum albumin
C-terminus	Carboxy-terminus
C10ORF12	Chromosome 10 open reading frame 12
C17ORF96	Chromosome 17 open reading frame 96
CBP	CREB-binding protein
CBX	Chromobox
CDK4	Cyclin-dependent kinase 4
CDK6	Cyclin-dependent kinase 6
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
Co-REST	RE1-Silencing Transcription factor
CpG	Cytosine phosphate guanine
cPRC1	Canonical polycom repressive complex 1
CtBP	C-terminal binding protein
DMEM	Dulblecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DTT	DL-Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EED	Embryonic ectoderm development
EHMT1/2	Euchromatic histone-lysine N-methyltransferase 1/2
ERα	Estrogen receptor α
ESC	Embryonic stem cells
EV	Empty vector
EZH1	Enhancer of Zeste homolog 1
EZH2	Enhancer of Zeste homolog 2
FBS	Fetal bovine serum
FDR	False discovery rates
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
Gm340	Mouse predicted gene 340
GNAT	Gcn5-related N-acetyltransferase
H3	Histone H3
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDF	Human diploid fibroblasts
HEK293T	Human Embryonic Kidney 293 cells, transformed with T antigen
HMEC	Human mammary epithelial cells
HMT	Histone Methyl Transferase

HOTAIR	HOX transcript antisense RNA
Hox	Homeotic genes
HRP	Horseradish peroxidase
IP	Immunoprecipitation
iPOND	Isolation of proteins on nascent DNA
JARID2	Jumonji, AT rich interactive domain 2
Kdm2b	Lysine (K)-specific demethylase 2B
LCOR	Ligand dependent nuclear receptor corepressor
me1	Monomethylation
me2	Dimethylation
me3	Trimethylation
MEF	Mouse embryonic fibroblast
MEGM	Mammary Epithelial Cell Growth Medium
MLR2	Mblk1-Related Protein 2
mRNA	Messenger RNA
MTF2	Metal response element binding transcription factor 2
MYST	MOZ, Ybf2/Sas3, Sas2, and Tip60 acetyltransferase family
N-terminus	Amino-terminus
NCBI	National Center for Biotechnology Information
ncPRC1	Non canonical Polycomb repressive complex 1
ncRNA	Noncoding RNA
NEB	New England Biolabs
NR	Nuclear receptor
NuRD	Nucleosome Remodeling Deacetylase
NURF	Nucleosome remodeling factor
ORF	Open reading frame
PBS	Phosphate-buffered saline
PcG	Polycomb
PCGF	Polycomb group ring finger
PCL	Polycomblike
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
pcv	Packed cell volume
PEP	Posterior error probability
PHC1	Polyhomeotic homolog
PHD	Plant homeodomain
PHF19	PHD finger protein 19
PHO	Pleiohomeotic
PMSF	Phenylmethanesulfonylfluoride or phenylmethylsulfonyl fluoride
pRb	retinoblastoma protein
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PRE	Polycomb Response Element
PTM	Post-translational modification

qPCR	Quantitative PCR
RBBP4/8	Retinoblastoma binding protein 4
RING1A	Ring finger protein 1
RING1B	Ring finger protein 2
RIP	RNA-immunoprecipitation
RIP-seq	RNA-immunoprecipitation followed by sequencing
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
RT-qPCR	Reverse transcription quantitative PCR
RYBP	RING1 and YY1 binding protein
SET domain	Suppressor of variegation, Enhance of Zeste, Trithorax
SET protein	SET nuclear proto-oncogene
shRNA	Short hairpin RNA
shSCR	Short hairpin scramble
SIN3A	SIN3 transcription regulator family member A
SUZ12	Suppressor of Zeste 12 Homolog
TK	Thymidine kinase
TRC2	The RNAi consortium
TSA	Trichostatin A
TxG	Trithorax
UAS	Upstream activation sequence
UCSC	University of California Santa Cruz
USP11	Ubiquitin specific peptidase 11
USP22	Ubiquitin specific peptidase 22
UTR	Untranslated region
UTX	Ubiquitously-Transcribed TPR Protein On The X Chromosome
WIZ	Widely interspaced zinc finger motifs
XIST	X inactive specific transcript
YAF2	YY1 associated factor 2
YY1	Yin And Yang 1 Protein
ZNF644	Zinc finger protein 644

Publications

Brien, G. L., Gambero, G., O'Connell, D. J., Jerman, E., Turner, S.A., Egan, C. M., Dunne E.J., Jurgens M.C., Wynne K., Piao L., Lohan A.J., Ferguson N., Shi X., Sinha K.M., Loftus B.J., Cagney G., Bracken, A. P. (2012) Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nature Structural & Molecular Biology, 19:1273–81.

Lanigan, F., Brien G.L., Fan Y., Madden S.F., Jerman E., Aloraifi F., Hokamp K., Dunne E.J., Lohan A.J., Flanagan L., Garbe J.C., Stampfer M.R., Fridberg M., Jirstrom K., Quinn C.M., Loftus B., Gallagher W.M., Geraghty J., Bracken A.P. (2015) Delineating Transcriptional Networks of Prognostic Gene Signatures Refines Treatment Recommendations for Lymph Node-negative Breast Cancer Patients. FEBS Journal, *in press*.

Munawar N., Olivero G., Jerman E., Doyle B., Streubel G., Wynne K., Bracken A.P., Cagney G. (2015) Native gel analysis of macromolecular protein complexes in cultured mammalian cells. Proteomics, *in press*.

Brien G.L., Healy E., Jerman E., Fadda E., O'Donovan D., Krivtsov A.V., Rice A.M., Kearney C.J., Flaus A., Martin S.J., McLysaght A., O'Connell D.J., Armstrong S.A., Bracken A.P. (2015) A Chromatin Independent Role of Polycomb-Like 1 to Stabilize p53 and Promote Cellular Quiescence. Genes and Development, *submitted*.

Chapter 1

Introduction

1.1 Epigenetics

Every multicellular organism develops from a single cell. In humans, about two hundred different cell types, which have distinct functions, with a few exceptions, share the same genome as the fertilised zygote. It is well known that gene expression patterns unique to each cell type determine cell fates. These patterns are established during the process of cellular differentiation, starting as early as embryogenesis, when the signals obtained from the mother and the neighbouring dividing cells are received and interpreted by each cell. Once the fate is determined, the gene expression pattern is maintained, in order to facilitate the correct phenotype, even after the signals that were required for its establishment are gone (Tammen et al, 2013).

The term 'epigenetics' was coined by Conrad Waddington to describe heritable changes in the cellular phenotype that were not dependent on the alterations in the DNA sequence. Epigenetic regulation has a crucial role in establishing and maintaining gene expression patterns during differentiation and is instructed by the master transcription factors as a response to developmental cues. As understanding of the gene regulation mechanisms developed, the term epigenetics evolved to describe chromatin-based events that regulate DNA-templated processes (Dawson and Kouzarides, 2012). Such events involve DNA methylation, the post-translational modification of histones, and RNA-mediated regulatory processes.

1.2 Chromatin and histone post-translational modifications

The chromatin of eukaryotic cells comprises DNA packaged into nucleosomes around octamers of two copies of four different histone proteins: histones H3, H4, H2A and H2B. A stretch of 147 base pairs of DNA double helix is wrapped around one octamer, resulting in what has been observed as a "bead on a string" structure of chromatin (Woodcock et al, 1976, Olins et al, 1977). The histone N-terminal tails protrude from the octamer (Luger et al, 1997) and are subject to multiple covalent post-translational modifications (PTMs), such as acetylation, phosphorylation, ubiquitylation, methylation and sumoylation (Kouzarides, 2007). The amino acid residue and type of histone PTM can be highly informative of the underlying gene expression status (Strahl and Allis, 2000, Ernst et al, 2011). Chromatin immunoprecipitation (ChIP) studies using antibodies specific to a particular histone PTM demonstrated the specific distribution patterns of PTM localisation in the genome. Some modifications and their effects are summarised in Table 1.

Table 1.1 Histone PTMs and gene transcription states

Modification	Expression state	Genomic localisation	Reference
H3K27Ac	active	enhancer	Ernst et al, 2011
HK9Ac	active	promoter	Ueda et al, 2006
H3K4me1	poised/active	enhancer	Heintzman et al, 2007
H3K4me2	active	enhancer, promoter	Pekowska 2011
H3K4me3	active/poised	promoter	Mikkelsen et al, 2007,
H3K9me1	active	gene body	Barski et al, 2007
H3K9me3	repressed	Inter-/intra-genic regions	Bannister et al, 2001, Barski et al, 2007
H3K27me1	active	gene body	Barski et al, 2007, Ferrari et al, 2014
H3K27me2	repressed	inter-/intra-genic regions	Ferrari et al, 2014
H3K27me3	represed/poised	promoter, enhancer	Zentner et al, 2011
H3K36me3	active	gene body	Bannister et al, 2005, Barski et al, 2007, Mikkelsen et al, 2007

Abbreviations used: H3K27 – histone H3 lysine 27; H3K9 – histone H3 lysine 9, H3K4 – histone H3 lysine 4; H3K36 – histone H3 lysine 36; Ac – acetylation; me1 – monomethylation; me2-dimethylation; me3 – trimethylation.

Histone PTMs are deposited by chromatin "writer" enzymes, such as acetylases, kinases, methyltransferases and ubiquitin ligases; recognized and bound to by "reader" proteins, via their bromo-, chromo- or TUDOR domains; and removed by "eraser" enzymes – demethylases, deacetylases and phosphatases (Kouzarides, 2007). These chromatin regulator proteins interplay during differentiation to establish tissue specific gene expression patterns and maintain them during subsequent cell divisions (Laugesen and Helin, 2014). Chromatin regulators often work in multi-protein complexes that contain different 'writers', 'readers' and 'erasers' and determine activation or repression of the underlying genes. Mutations in genes that encode epigenetic regulators lead to aberrant cell development and proliferation and are often found in cancer (You and Jones, 2012, Laugesen and Helin, 2014). Therefore the processes governing the function

of epigenetic regulators, their recruitment and displacement from target genes remain an area of immense interest and investigation towards development of novel cancer therapies and regenerative medicine approaches. The main mechanisms associated with histone acetylation and methylation are discussed in the following section.

1.3 Histone post-translational modifications in regulation of gene expression

Chromatin has two states, euchromatin (at regions of active transcription) and heterochromatin (at repressed regions). In euchromatic regions, the post-translational acetylation of histone tails neutralises the basic charge of lysine residues on which it is present, thus loosening the salt bridge that was existent before between the positively charged lysine residue and the phosphate backbone of DNA. The result of this is a more 'open' chromatin conformation, which is permissive to access of the transcription machinery (Vettese-Dadey et al, 1996, Kouzarides, 2007). Histone tails can be acetylated at various lysine residues. Histone acetyltransferases (HATs), which are subdivided into GNAT (Gcn5 related acetyltransferase), MYST and CBP/p300 families, are responsible for deposition of this PTM (Kouzarides, 2007). On the other hand, histone deacetylases (HDACs), which work as part of multi-protein complexes such as NuRD, SIN3A and Co-REST, mediate repression by removing acetyl groups from histones (Roth et al, 2001, McDonel et al, 2009). Therefore, HAT and HDAC complexes act as co-activators and co-repressors when associated with the specific DNA-binding transcription factors. Interestingly, some nuclear hormone receptors are capable of recruiting HATs or HDACs depending on whether they are activated by an agonist or not (Gurevich et al, 2007, Perissi et al, 2010, Watson et al, 2012).

Histone PTMs such as methylation do not change the charge of the nucleosome. Unlike acetylation, a methylation mark may be present on either active or repressed genes (Table 1.1). Lysines can be mono-, di- or tri-methylated, while arginines can be symmetrically or asymmetrically di-methylated. SET domain (stands for: Suppressor of variegation, Enhance of Zeste, Trithorax, named after proteins it was identified in) is found in proteins responsible for the Histone Methyl Transferase (HMT) activity of around 40 known HMT proteins. The methylation PTMs, which are associated with gene transcription, are recognised by complexes that possess ATP-dependent chromatin remodeler activity, which perform sliding of nucleosomes and facilitate access of the general transcription factors to DNA. For example, CHD1 (via its chromodomain) and NURF (via its PHD finger domain) can both bind to the H3K4me3 mark to promote transcriptional activation (Sims et al, 2005, Gaspar-Maia et al, 2009, Wysocka et al, 2006). In addition to the recruitment of chromatin remodeling factors, histone PTMs serve as contact sites for the complexes that deposited them, for stabilisation and sustainment of their action. In particular, this is the case for the methylation of histones at repressed gene loci (Hansen et al, 2008).

In recent years, the localisation of various PTMs has been extensively studied by genome wide sequencing and this revealed that specific combinations of PTMs define so called "chromatin states", which are associated with gene activation or repression (Table 1.1). For example, chromatin regions are marked with H3K4me1 and acetylated at histone H3 lysine 27 at active enhancers, while H3K4me3 is present on promoters and H3K36me3 is distributed along the gene bodies of

expressed genes. In contrast, H3K9me3 was reported to be present in the condensed heterochromatin regions encompassing repressed genes and intergenic regions. H3K27me3 is also associated with repression, although its distribution is confined to enhancers, gene promoters, and to some extent, gene bodies. These observations suggest that 'crosstalk' between PTMs is occurring on the molecular level (Kouzarides, 2007, Voigt et al, 2012). Firstly, some histone PTMs, such as trimethylation and acetylation of H3K27 and H3K9, are in competition for the histone residue, indicating the inherent mutual exclusivity of the activated and repressed chromatin states. Secondly, the presence of the H3K4me3 or H3K36me3 activation marks, for instance, was reported to inhibit the H3K27me3 deposition (Schmitges et al, 2011, Yuan et al, 2011).

In contrast to the active and repressed states, bivalent chromatin domains, containing both the active H3K4me3 and the repressive H3K27me3 marks were found on promoters of developmental genes in mouse embryonic stem cells that are repressed (Bernstein et al, 2006a, Voigt et al, 2012). These genes were referred to as 'poised', because upon induction of differentiation they can be either activated or permanently repressed depending on the specific lineage path taken (Voigt et al, 2013). H3K4me3 and H3K27me3 are "written" by the Trithorax (TxG) and Polycomb (PcG) group proteins, respectively. The functions of Trithorax and Polycomb proteins are essential for proper development and mechanisms of how these protein complexes regulate developmental genes are a subject of active research (Bracken and Helin, 2009).

1.4 Polycomb Group proteins

1.4.1 Function of Polycomb Group proteins

The Polycomb group proteins are transcriptional repressors that are required for establishing the correct cellular identities during development (Simon and Kingston, 2013). They were originally identified as being essential for maintaining the repression of homeotic genes during development in *Drosophila* (Tolhuis et al, 2006, Ringrose and Paro, 2004). Knock-out studies in mice have shown that loss of most of the PcG genes results in early embryonic lethality (Laugesen and Helin, 2014). In agreement with this, the conditional knock-out of Polycomb proteins leads to posterior to anterior homeotic transformations in *Drosophila*, concomitant with the failure to repress Hox genes (Ringrose and Paro, 2004). The genome wide mapping of Polycomb target genes demonstrated that, in addition to Hox genes, they repress genes encoding developmental regulators in stem cells and in differentiated cells (Bracken et al, 2006, Boyer et al, 2006, Lee et al, 2006, Schwartz et al, 2006). Subsequent studies established a model for dynamic Polycomb association with their target genes during differentiation and development (Figure 1.1) (Pasini et al 2007, Bracken and Helin, 2009). During stem cell differentiation Polycombs are displaced from a small cohort of their target genes that are essential for lineage specification, while also getting recruited to several stem cell genes. The mechanisms of how Polycombs are recruited to or displaced from their target genes during lineage specification and how they function to maintain gene repression during subsequent cell divisions, remain poorly understood.

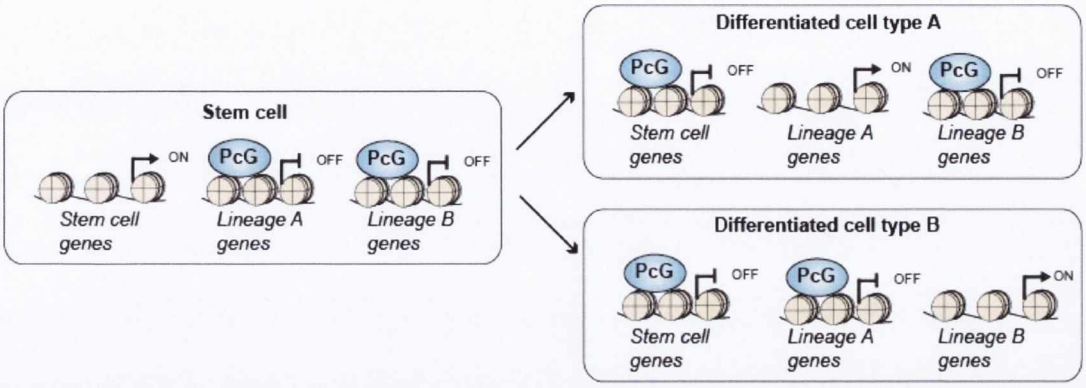


Figure 1.1 Dynamic function of Polycomb proteins during cell fate decisions

A model demonstrating the general mechanism of Polycomb action during development. This models shows that in stem cells Polycomb group (PcG) proteins bind and repress lineage specific genes. As cells commit to a lineage specific differentiation pathway (either to turn into cell type A or cell type B) PcG proteins get recruited to and repress the promoters of genes required for maintaining stem cell identity and are specifically displaced from the genes required for generation of a specific lineage type (A or B). The mechanisms that determine the specific displacement and recruitment of Polycomb are poorly understood and remain a subject of active research.

1.4.2 Polycomb Repressive Complexes

Polycomb proteins are divided into two main multi-protein complexes that have distinct enzymatic activities: the Polycomb repressive complex 1 (PRC1) and the Polycomb Repressive Complex 2 (PRC2). PRC1 ubiquitynates histone H2A at lysine 119 (H2AK119Ub), a PTM, which was proposed to have a role in chromatin compaction (Wang et al, 2004a, Francis et al, 2004, Zhou et al, 2008, Endoh et al, 2012). Core PRC2 complex comprises three proteins: a histone methyltransferase EZH2 or EZH1 as well as EED and SUZ12, which together mediate mono-, di- and tri-methylation of histone H3 at lysine 27 (H3K27me3) (Cao et al, 2002, Kuzmichev et al, 2002, Cao and Zhang, 2004, Margueron and Reinberg, 2011, Ferarri et al, 2014). In addition, PRC2 may contain several sub-stoichiometric components, such as Polycomblike 1-3 (PCL1-3), JARID2 and RBBP4/8 proteins (Figure 1.2, A) (Sauvageau and Sauvageau, 2010).

The initially characterised mammalian canonical PRC1 complex (cPRC1) is composed of the catalytic subunit (RING1A or RING1B), one PCGF protein (PCGF2 or PCGF4), one CBX protein (CBX2, CBX4, CBX6, CBX7, CBX8) and a Polyhomeotic protein (PHC1-3) (Figure 1.2, B) (Buchwald et al, 2006, Li et al, 2006, Gao et al, 2012). It was postulated that cPRC1 complex is recruited to its target genes in a PRC2-dependent manner via a Chromobox domain in the CBX component, which binds to the H3K27me3 mark (Figure 1.3, A) (Wang et al, 2004b, Francis et al, 2004, Bernstein et al, 2006b).

More recently, non-canonical forms of the PRC1 complex (ncPRC1) have been described, which lack the CBX and PHC proteins and instead contain a

RYBP/YAF2 subunit, as well as a RING protein and a PCGF(1-6) (Figure 1.2, C) (Gao et al, 2012, Morey et al, 2013, Wu et al, 2013). Surprisingly, these studies have demonstrated that PRC2 recruitment is dependent on the deposition of the H2AK119Ub by the ncPRC1 (Wu et al, 2013, Blackledge et al, 2014, Kalb et al, 2014). Furthermore, the Pcgf1-containing ncPRC1 is recruited to unmethylated CpG islands by an interaction with lysine demethylase Kdm2B/Fbxl10 via its CXXC motif (Wu et al, 2013). Surprisingly, loss of Pcgf1-ncPRC2 caused reduced PRC2 binding and H3K27me3, while de novo recruitment of Kdm2b or ncPRC1 components was sufficient to recruit PRC2 and cPRC1 (Blackledge et al, 2014). These findings thus propose a model whereby Polycomb recruitment begins with ncPRC1 binding to CpG unmethylated islands *via* Kdm2b, and is followed by PRC2 recruitment (which mediates H3K27me3) thus facilitating a feedback loop for the cPRC1 recruitment (Figure 1.3, B) (Turner and Bracken, 2013, Schwartz and Pirrotta, 2013, Blackledge et al, 2014).

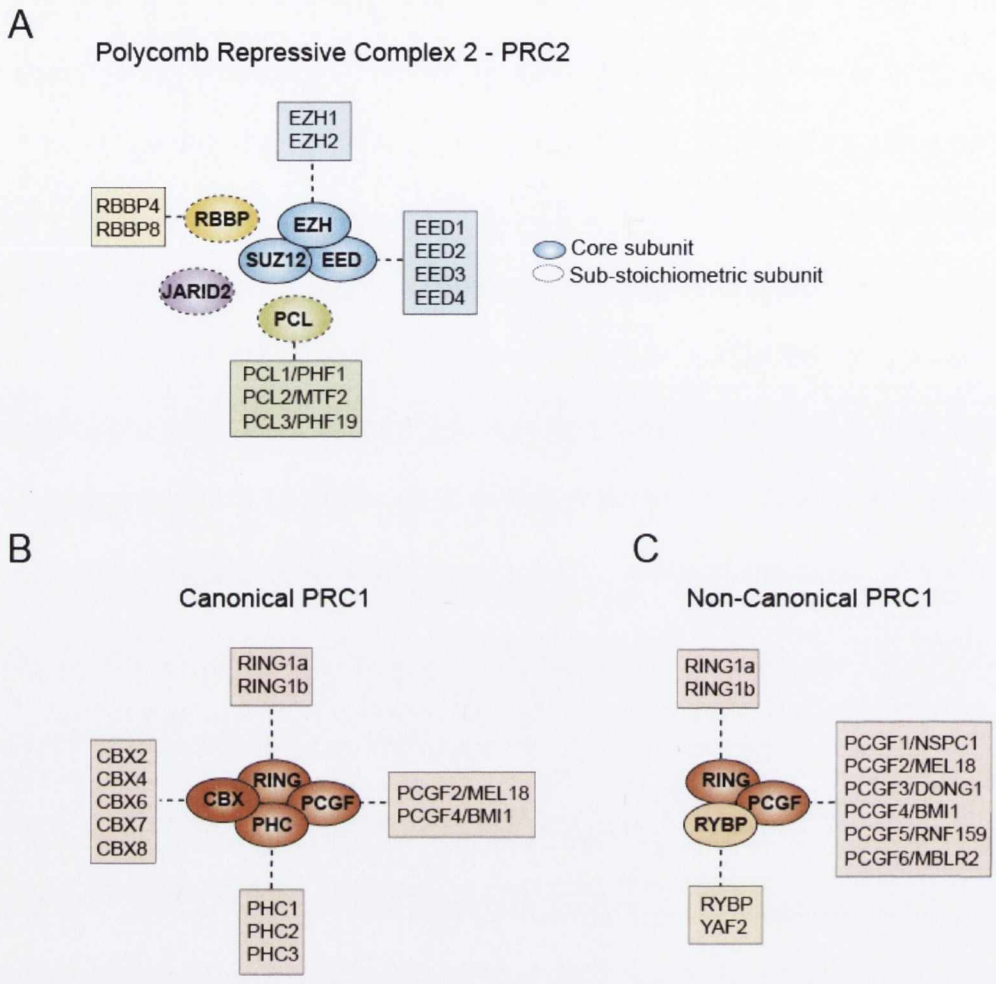


Figure 1.2 Polycomb group proteins form distinct multiprotein enzymatic complexes

Polycomb group proteins form two major types of multi-protein complexes: the Polycomb Repressive Complex 2 (PRC2) and the Polycomb Repressive Complex 1 (PRC1), which possess distinct enzymatic activities.

A. PRC2 is composed of three core components: EZH1/2, EED1-4 and Suz12. The core PRC2 complex may additionally associate with sub-stoichiometric components such as Polycomblike(PCL)1-3, JARID2 and RBBP4/6 proteins.

B. The canonical PRC1 (cPRC1) complex is composed of four core subunits: catalytic subunit RING1A or RING1B, one PCGF protein (PCGF2 or PCGF4), one CBX protein and a Polyhomeotic (PHC) protein.

C. The non-canonical PRC1 (ncPRC1) complex, in contrast to cPRC1, lacks the CBX and PHC components and instead associates with an RYBP protein.

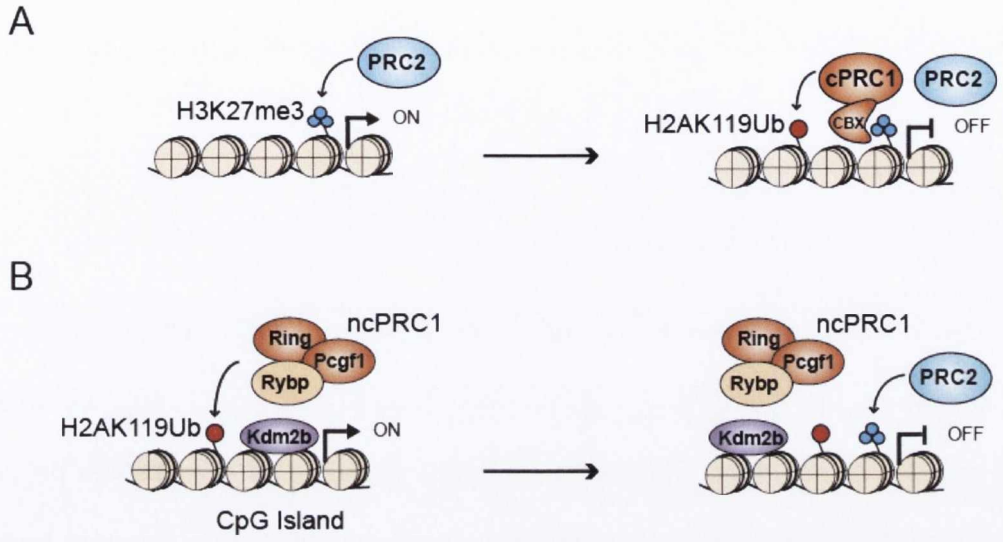


Figure 1.3 PRC1 and PRC2 co-operate in gene repression

A. In the canonical pathway of Polycomb recruitment, PRC2 is recruited to chromatin and mediates the trimethylation of histone H3 at lysine 27 (H3K27me3). H3K27me3 is "read" by the Chromobox domain of the CBX component of the cPRC1 complex leading to cPRC1 mediated ubiquitination of histone H2A at lysine K119 (H2AK119Ub) and chromatin compaction.
B. Recently, a non-canonical pathway for Polycomb recruitment was proposed. The ncPRC1 complex is recruited to unmethylated CpG islands in the presence of the lysine demethylase KDM2B and deposits H2AK119Ub PTM. This in turn leads to recruitment of the PRC2 complex and H3K27 trimethylation.

1.4.3 Potential mechanisms of PRC2 recruitment to target genes

In addition to the sequential model of *de novo* PRC2 recruitment mediated by the ncPRC1 complex discussed above, a number of other factors have been implicated in having a role in targeting PRC2 to its target genes.

Transcription factors

In *Drosophila,* Polycombs are recruited to DNA regions called Polycomb Response Elements (PREs) (Chan et al, 1994, Mohd-Sarip et al, 2002, Orlando, 2003, Sengupta et al, 2004, Tolhuis et al, 2006). Polycombs themselves do not possess specific DNA binding capabilities. In *Drosophila,* transcription factor Pleiohomeotic (PHO) was reported to directly link Polycomb to DNA *in vitro* and mediate repression (Mohd-Sarip et al, 2002, Bloyer et al, 2003). Yy1, an orthologue of Drosophila *Pleiohomeotic* in mammals has been the focus of attention – but despite several reports, its links to Polycomb function in mammalian cells remain unclear (Caretti et al, 2004, Vella et al, 2011). Other data suggests that the interaction of Polycomb with a transcription factor is context-specific and possibly transient (Di Croce and Helin, 2013), therefore making such an interaction difficult to observe using current technologies such as ChIP.

Sub-stoichiometric components

An alternative mechanism by which Polycombs might be recruited to target genes has emerged, when it was shown that the sub-stoichiometric subunits of the PRC2 complex, in particular, JARID2 and AEBP2, have a binding preference for GC-rich regions of DNA (Peng et al, 2009, Kim et al, 2009, Landeira et al, 2010). A model was proposed, in which unmethylated CpG islands on transcriptionally silenced

genes serve as recruitment points for both PRC2 and ncPRC1 (Mendenhall et al, 2010, Di Croce and Helin 2013, Wu et al, 2014). On the other hand, PCL1 and PCL3 were reported as "readers" of the H3K36me3 PTM, which is associated with the gene bodies of active genes, and, in the case of PCL3, to be required for the recruitment of the PRC2 complex to active pluripotency genes during differentiation (Brien et al, 2013, Musselman et al, 2012, Ballaré et al, 2012, Cai et al, 2013, Qin et al, 2013).

Long non-coding RNAs

Long noncoding RNAs (lncRNA) have also been reported to be associated with the process of PRC2 recruitment to chromatin. For example, the lncRNA XIST was reported to be required for the recruitment of the PRC2 complex to inactivate the X-chromosome *in cis* (Zhao et al, 2008a), while the HOTAIR long lncRNA was reported to mediate recruitment of the PRC2 complex to the HoxD locus *in trans* (Rinn et al, 2007). Furthermore, RNA-immunoprecipitation (RIP) studies identified that ~20% of long non-coding RNAs were associated with PRC2 (Khalil et al, 2009, Zhao et al, 2010). However, it is still unclear whether the interaction between the lncRNA and PRC2 is direct, considering that none of the PRC2 components contain any known RNA-binding domains and the issue of unascertained specificity of the RIP-seq assays (Brockdorff, 2013). Therefore further elucidation of the RNA interacting region in PRC2 components will be required for evaluating this model of PRC2 recruitment to target genes.

1.4.4 Roles of PRC2-mediated mono- and di-methylation of H3K27

A recent study demonstrated that in addition to the H3K27me3 mark, which is present on genes that are bound by Polycombs, the PRC2 complex is required for mediation of H3K27me1 and H3K27me2 deposition (Ferrari et al, 2014). Surprisingly, it was found that genome-wide distribution of these PTMs was distinct to that of PRC2, in particular, H3K27me1 was present on the gene bodies of active genes, co-localised with the H3K36me3 activating PTM, and was required for the transcription of its target genes; while H3K27me2 was detected on ~70% of the histone H3 in the intergenic and intragenic regions of the repressed genes (Ferrari et al, 2014, Barski et al, 2007). H3K27me2 was demonstrated to prevent the acetylation of histone H3, and thus to prevent the activation of inactive enhancers. The precise mechanisms of the modulation of the PRC2 towards a particular methylation product (mono-, di-, or tri-methylation of H3K27) remain to be elucidated. The current explanation for the lack of PRC2 occupancy at H3K27me1 and H3K27me2 sites comes from the fact that the conversion rate of H3K27me2 to H3K27me3 is significantly slower than the conversion rate of H3K27me0 to H3K27me1 and H3K27me1 to H3K27me2 (McCabe et al, 2012). Further studies will be required to address this mechanism.

1.4.5 PRC2 function in cell cycle regulation

EZH2 expression is associated with proliferating cells, but is low in differentiated cells and cells with arrested growth (Müller et al, 2001, Bracken et al, 2003, Bracken et al, 2007, Margueron et al, 2008). Furthermore, PRC2 binds and represses the INK4A/ARF locus, which encodes the p16 and ARF proteins – activators of the pRb and p53 checkpoints, respectively (Bracken et al, 2007).

EZH2 is directly regulated by the E2F pathway and is downregulated as cells undergo cellular senescence (Bracken et al, 2003, Bracken et al, 2007). In the model of replicative senescence, the stress caused by passaging of primary mammalian cell lines leads to gradual upregulation of cyclin-dependent kinase inhibitor $p16^{INK4A}$ on the mRNA and protein levels. P16 accumulation in turn inhibits the cyclin dependent kinases CDK4 and CDK6, whose function is to phosphorylate retinoblastoma protein pRb, resulting in the cell cycle arrest at the transition from G1 to S phase (Lanigan et al, 2011). Notably, the upergulation of p16 expression levels is correlated with upregulation of the histone H3K27me2/3 demethylase KBM6B/ JmjD3, suggesting a mechanism by which PRC2 binding to the locus could be diminished (Agherbi et al, 2009). Finally, the overexpression of PRC2 components mediates the enhanced repression of the INK4A/ARF locus and prevents the activation of the senescence checkpoint, thereby contributing to cellular transformation (Bracken et al, 2007, Bracken and Helin, 2009, Lanigan et al, 2011).

1.4.6 PRC2 function in DNA replication

Surprisingly, PRC2 was recently reported to regulate cellular proliferation independently of the p16/p53 checkpoints (Piunti et al, 2014). PRC2 expression was required for proliferation of primary and SV40 antigen immortalised mouse embryonic fibroblasts (MEFs) that were null for p16 or p53. Furthermore, PRC2 expression was required for the transformation of those cells in the mouse xenograft studies. This effect was attributed to the impaired ability of the PRC2 deficient cells to transition from the G1 to S cell cycle phase (Pasini et al, 2004, Piunti et al, 2014). Together with the facts that Polycombs associate with

chromatin during DNA replication (Hansen et al, 2008, Francis et al, 2009, Piunti et al, 2014), and that EZH2 deficiency caused a reduction of the replication fork speed and promoted asymmetric fork progression (Piunti et al, 2014), these data suggested a direct role of PRC2 function in DNA replication.

1.4.7 Deregulation of PRC2 function in cancer

The function of the PRC2 complex is frequently deregulated in cancer (Laugesen and Helin, 2014). Interestingly, both activating (Sneeringer et al, 2010, Zhang et al 2012, Vogelstein et al, 2013) and inactivating (Ntziachristos et al, 2012, Morin et al, 2010, Ernst et al, 2010, Nikoloski et al, 2010) mutations of EZH2 and SUZ12 have been reported, indicating that the PRC2 complex may exert tumour suppressive or oncogenic functions, depending on the cellular context (Koppens and van Lohuizen, 2015). Furthermore, JARID2 was found to be deleted in some leukemia cases, further suggesting that PRC2 has a tumour suppressive function in acute myeloid leukemia (Puda et al, 2012). PCL1 and SUZ12 translocations were detected in endometrial sarcomas (Micci et al, 2006, Li et al, 2007) and PCL3 was reported to be overexpressed in multiple cancer types (Wang et al, 2004). In addition, somatic mutations in genes encoding histone H3.1 and H3.3 (*HIST1H3B* and *H3F3A,* respectively) substituting the lysine with a methionine at residue K27 occur in up to 70% of glioblastoma multiforme cases (Schwartzentruber et al, 2012, Wu et al, 2012). Deactivating mutations of UTX, a H3K27me3 demethylase, were detected in multiple cancer types, including renal carcinomas (Van Haaften et al, 2009, Dalgliesh et al, 2010). Taken together, all these observations suggest that multiple forms of disruption of the PRC2 function

may lead to a deviation from the selected gene expression program of the cell, shifting it to become cancerous.

1.5 Aims of this study

While deposition of the H3K27me3 (Pengelly et al, 2013) and the H2AK119Ub (Endoh et al, 2012) PTMs are essential for mediating repression of PRC2 target genes, the exact molecular mechanisms of Polycomb mediated repression, as well as the mechanisms of its eviction from specific target genes during development remain unclear (Laugesen and Helin, 2014). The study of the function of the PRC2 complex in developmental biology and cancer may lead to a better understanding of the regulation of cellular identity and may contribute to the development of novel strategies for cancer treatment. In this work we decided to explore further the mechanisms of PRC2 recruitment by delineating the molecular and functional properties of the sub-stoichiometric components of the PRC2 complex.

Chapter 2

Materials and methods

2.1 Reagents

2.1.2 Antibodies

The antibodies used in this study their dilutions for Western blotting analysis were as follows: rabbit anti-FLAG, 1:1000 (Sigma, F7425); rabbit anti-PHF1, 1:500 (Proteintech, 15663-1-AP); rabbit anti-MTF2/PCL2, 1:500 (Genway, GWB-FA7207); rabbit anti-PHF19/PCL1, 1:500 (selfmade, Brien et al, 2013); mouse anti-p53, 1:1000 (DO-1); mouse anti-EZH2, 1:8 (BD43 for Western blot and AC22 for IP); mouse anti-BMI1 supernatant, 1:60 (DC9 for Western blot and AF27 for IP); rabbit anti-SUZ12, 1:1000 (Cell Signaling); rabbit anti-beta-tubulin, 1:500 (sc-9104); rabbit anti-histone H3, 1:20,000 (Abcam, ab1791); goat anti-LCOR, 1:500 (sc-163009); rabbit anti-SET/ I2PP2A, 1:500 (sc-25564); rabbit anti-CtBP, 1:1000 (sc-11390); rabbit anti-USP11, 1:1000 (Cambridge Bioscience, A301-613A); rabbit anti-USP22, 1:500 (Abcam, ab4812); mouse anti-CtBP1, 1:2000 (BD, 612042); mouse anti-CtBP2, 1:2000 (BD, 612044); mouse anti-H3K27me1, 1:2000 (Active motif, 61015); rabbit anti-H3K27me2, 1:2000 (Cell Signaling, 9728S); rabbit anti-H3K23me3, 1:2000 (Active Motif, 39155); rabbit anti-G9a, 1:500 (Cell Signaling, #3306S); rabbit anti-GAL4 (sc-577).

2.1.2 'BIG' (C10ORF12) antibody generation

The 'BIG' (C10ORF12) antibody was generated in collaboration with Millipore Corporation. EMD Millipore generated the antibody by immunization of rabbits with epitope peptide LSSRKTARKSTRGYFFNGDC, corresponding to amino acids 491-510 of human 'BIG' protein, which consisted of 5 injections approximately 20 days apart from each other. Antibody containing serum was isolated from three bleeds

that commenced 72 days after the first immunisation, and after a 5-day interval and a subsequent 20-day interval, as well from the final exsanguination. The antibody was then affinity purified from pooled rabbit serum against the antigen. The affinity-purified antibody was then tested for Western blot and ChIP applications as outlined in this work. Four different rabbits were immunised and therefore 4 different batches of antibody received, each performing comparably in the Western blot application.

2.2 Cloning and expression vector generation

2.2.1 Gateway cloning

Full length open reading frame (ORF) of selected human genes, summarised in Table 2.1, were PCR amplified using the forward and reverse primers indicated from the cDNA generated from early passage primary human mammary epithelial cells (HMEC), using Platinum Taq High Fidelity DNA polymerase (Life Technologies) or Phusion High Fidelity DNA polymerase (NEB). The PCL3/PHF19 ORF was previously cloned by Dr. Gerard Brien. The truncated fragments of C10ORF12 ORF were subsequently amplified from the pLENTI-C10ORF12 expression construct and sub-cloned into pLENTI expression vector under my supervision as part of undergraduate final year project by Indigo Pratt Kelly.

Table 2.1. Summary of the genes cloned in this study

Gene name	Accession number	Forward primer	Reverse primer
PCL1/PHF1	NM_024165.2	GATGCAATGGCGCAGCCC	TCAGAAGATGCCCCCTCCTCC
PCL2/MTF2	NM_007358.3	ATGAGAGACTCTACAGGGGCA	TCAGGATGCAGTTGCTCCTTC
LCOR	NM_001170765.1	ATGCAGCGAATGATCCAACAA	CTACTCGTTTTTTGATTCATTTGC
C10ORF12	NM_015652.2	ATGCAGAGTTCAGCTTTAGTAG	TCACTTTGCATCCAGCCG
BIG	-	ATGCAGCGAATGATCCAACAA	TCACTTTGCATCCAGCCG
Fragment 1	1-1239 of C10OR12	ATGCAGAGTTCAGCTTTAGTAG	TCAAGCCTTTCCCAGACTC
Fragment 2	607-1806	CTGCCAACTGTTCGTACACT	TCACAAAGTCTGCCTGGTG
Fragment 3	1240-2475	GAGGACAACCAAAGCATCAG	TCATGTTGGCTGTCCTTCTT
Fragment 4	1813-3075	ATGCTGGACAAAGAAGTCAAG	TCATGGGCATTCCATTTTCTCA
Fragment 5	2476-3744	CCAAGAGCAAGGAACAAATCA	TCACTTTGCATCCAGCCG

The amplified ORFs were gel excised from 1% agarose gel using the Wizard® SV Gel and PCR clean-up system (Promega) according to the manufacturer's instructions and inserted into a pCR8/GW/TOPO Gateway cloning entry vector (Invitrogen). The ORF sequence was verified by Sanger sequencing (GATC) and then sub-cloned into the Gateway cloning-compatible destination vectors: pMINKIO (PCL1, PCL2) or pLENTI (LCOR, C10ORF12, BIG), pCDNA5-FRT-TO-GAL4 (BIG) using the Gateway LR Clonase II enzyme mix (Life Technologies).

2.2.2 shRNA expression vector generation

The complimentary DNA oligonucleotides, containing a targeting sequence of 21 nucleotides designed to form a stem-loop structure and flanked by EcoRI and Age1 overhangs on 5' and 3' ends, respectively, were annealed and ligated into pLKO TRC2 EcoRI and AgeI double-digested lentiviral expression vector. The 21 targeting nucleotide sequences were as follows:

Human short hairpin RNA (shRNA) target sites:

shLCOR+BIG:	GGACGGTGTACTTGATCTGTCC
shLCOR only:	CCATTCATCTCCTGTAGATTT
shBIG:	CAGTTCAGTGGATAGTTTCAC
shC10ORF12 + BIG:	AAGTCAAGGAAGATAGAAACA

Mouse shRNA target sites:

shBIG(Gm340)-4:	AAGATCCATACTGTCTTCTCG
shBIG(Gm340)-5:	GACTCCCAAGCAGACTCTTAC
shBIG(Gm340)-8:	GGGCTAAACTACGAGAGAATC
shBIG(Gm340)-9:	GACACAGAAGCTATGGGCTAA
shCtBP2-1:	CGGATGAATTTGATGGTCTTT
shCtBP2-2:	TACGAAACTGTGTCAACAAAG

The shSCR (MISSION TRC2 pLKO.5- puro Non-Mammalian shRNA Control Vector (Sigma, SHC202)) (CAACAAGATGAAGAGCACC) was used as a negative control.

2.3 Cell Culture

All cells were grown in a humidified incubator at 37°C with 5% CO2.

2.3.1 HMEC cell culture

Primary Human Mammary Epithelial Cells (HMECs) were obtained from the Martha Stamfer laboratory or from reduction mammoplasty tissue by Dr. Fiona Lanigan and Dr. Fatima Aloraifi in the Bracken laboratory and passaged in the M87A medium: 50% MM4 medium [DMEM/F12 (Gibco), 10 µg/ml insulin (Sigma), 10 nM tri-iodothyronine (Sigma), 1 nM β-estradiol (Sigma), 0.1 µg/ml hydrocortisone (Sigma), 0.5% FBS (26140, Gibco), 5 ng/ml epidermal growth factor (Peprotech), 2 mM glutamine (Lonza), 1 ng/ml cholera toxin (Sigma)] and 50% MCDB170 medium [MEGM media (Lonza) supplemented with 5 µg/ml

transferrin (Lozna), 10^{-5} isoproterenol (Sigma), and 2 mM glutamine (Lonza). M87A media was further supplemented with 0.1 nM oxytocin (Bachem) and 0.1% Albu-Max 1 (Invitorgen). For passaging, ~80% confluent cells were trypsinised by STV (5.37 mM KCl, 6.9 mM $NaHCO_3$, 136.9 mM NaCl, 5.55 mM D-Glucose, 0.54 mM EDTA, 500 mg/L Trypsin (1:250)) buffer and split at a 1 to 5 ratio.

2.3.2 MEF cell culture

Mouse embryonic fibroblasts (MEFs) were derived from embryonic day 13.5 C57BL6 mouse embryos by Dr. Gerard Brien, and cultured in Dulblecco's Modified Eagle Medium DMEM media supplemented with 10% FBS (Hyclone), 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco).

2.3.3 HEK293T cell culture

Human embryonic kidney HEK293T cells and producer amphotrophic HEK293T cells, which contain retroviral packaging and envelope genes incorporated into their genome, were cultured in DMEM supplemented with 10% fetal bovine serum (FBS, Hyclone or Sigma), 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco). Cells were passaged at ~80% confluence, by trypsinising using 0.25% Trypsin-EDTA (Gibco) and plated at a ratio of between 1 to 4 and 1 to 8.

2.3.4 MCF7 cell culture

Human breast cancer MCF7 cell line was contained in DMEM/F12 media, supplemented with 5% FBS (Gibco), 100 U/ml penicillin and 100 µg/ml

streptomycin (Gibco). For passaging the cells were split at a 1 to 4 ratio. For estrogen depletion experiments, ~80% confluent MCF7 cells were split at a 1 to 3 ratio until cells reached ~80% confluence, after which cells were washed twice with DPBS and placed into phenol-red free DMEM/F12 (Gibco) supplemented with 5% charcoal-stripped FBS (Sigma, F6765) for 72 hours, for depletion of activated estrogen receptor. The cells were then treated with either ethanol vehicle or 1 mM estradiol (Sigma) for 45 minutes and washed twice with PBS and harvested.

2.3.5 Calcium phosphate transfection of HEK293T cells

Approximately 80% confluent HEK293T cells were split at a ratio of 1:5 approximately 16 hours before the transfection. For transfection of a 10 cm tissue culture dish, 500 µl of 400 mM calcium chloride solution containing 2-10 µg of expression vector plasmid DNA was mixed together with 500 µl 2x HBS solution and incubated for 15 min, after which the DNA-Calcium phosphate mixture was added drop-wise directly onto the media of the HEK293T cells, in the presence of 30 µM chloroquine. Transfected cells were placed into fresh media 7-16 h and harvested 24-74 h post transfection.

2.3.6 Retroviral transduction

For generation of HEK293T cell lines stably expressing FLAG-HA tagged PCL1, PCL2 and PCL3 or the empty vector control, amphotrophic HEK293T retroviral producer cells were transfected by the calcium phosphate method with pMINKIO expression vectors for 7 hours. Media containing viral particles was collected at 48 and 72 hour time points post transfection, filtered through a 0.45 µm filter and placed directly onto target HEK293T cells, which were split at 1:6 ratio ~16 before

infection. The infected cells were incubated with the virus for 8 hours on two consecutive days with addition of 5 µg/ml polybrene (Sigma) and subsequently split into media containing 0.5 µg/ml puromycin (Sigma). The infected cells were selected for 3-5 days until puromycin had killed control non-infected HEK293T cells.

2.3.7 Lentiviral transduction

Lentiviral particles were produced by calcium phosphate transfection of HEK293T cells, which were seeded at 5 million cells per T75 flask the day prior to transfection with 10 µg of pLKO shRNA expression vector, 10 µg of viral packaging vector (pPAX8) and 8 µg of viral envelope vector (pVSVG) DNA. The media containing viral particles was collected at 48 and 72 hours post transfection, filtered through a 0.45 µm filter, and used directly to infect target cells for 8 hours in the presence of 5 µg/ml polybrene. Alternatively, lentiviral particles were purified by a 2 h ultracentrifugation at 20,000 rpm at 4°C, resuspended in 200 µl of DPBS and stored at -80°C. After the infection, target cells were placed into normal media for 24-48 hours, before splitting into media selection with puromycin (0.5 µg/ml for HEK293T cells or 1.0 µg/ml for MCF7 cells).

2.3.8 Generation of inducible GAL4-'BIG' HEK293T luciferase reporter cell line

The inducible Gal4-BIG or empty vector (EV) control cell lines were generated as described previously (Hansen et al, 2008). In brief, pCDNA5-FRT-TO-GAL4-BIG vector (or empty vector) was co-transfected together with the Flp-recombinase vector into the Flp-In T-Rex 293T cell line by the calcium transfection method.

Cells with stable incorporation of Gal4-BIG or empty vector constructs into their genomes were obtained by splitting the transfected cells at a 1 to 20 ratio into selection with 1 µg/ml puromycin for 5 days, and clones deriving from single cells were expanded over the course of 4-5 weeks. The Gal4 cell lines were passaged in DMEM, 10% tetracycline screened FBS (Fisher, HYC-001-333C), 100 U/ml penicillin and 100 µg/ml streptomycin, to prevent the unwanted "leakiness" of the tetracycline inducible promoter, and split at a 1 to 5 ratio. Expression of Gal4-BIG and Gal4-EV was induced by treating the cells with 1 ug/ml of doxycycline for 48 or 72 h before harvesting for luciferase or ChIP assays.

2.4 Real-time quantitative PCR (RT-qPCR)

For mRNA expression assays, RNA was extracted from cells using the RNeasy kit (Qiagen) and cDNA was generated by reverse transcription PCR from 1-2 µg of RNA using the TaqMan Reverse Transcription kit (Applied Biosystems) and the oligo dT primers that recognise the poly-A tail of the mRNA. Relative mRNA expression levels were determined by the SYBR Green I detection chemistry (Applied Biosystems) on the ABI Prism 7500 Fast Real–Time PCR System. In the mRNA expression analyses, the ribosomal constituent *RPLPO* or housekeeping *GAPDH* levels were used as a normaliser. Error bars indicate standard deviation of triplicate or duplicate qPCR data, while in the ChIP-qPCR experiments, 1% of Input material served as the normaliser. The qPCR primers used are summarised in Table 2.2.

Table 2.2. qPCR primers used for detection of mRNA levels of the listed genes.

mRNA	Forward (5' to 3')	Reverse (5' to 3')
hPCL1	GAGCTTTCAGACACCCCCAAAG	GGCTGTTAAGAGCAGAGAGGAGC
hPCL2	AACGCTCATTCTACCCCCAAC	GCACATATGCACGCACAAACC
hPCL3	GAAGGACATACAGCATGCCGG	CCCTAGGCAGATGTTGCACTTG
hEZH2	GGGACAGTAAAAATGTGTCCTGC	TGCCAGCAATAGATGCTTTTTG
hEZH1	TAAATTGCACGCGTTTAGGCTG	TCAGATACCCTCTGCCAGTGTG
hINK4A	GAAGGTCCCTCAGACATCCCC	CCCTGTAGGACCTTCGGTGAC
hRLPO	TTCATTGTGGGAGCAGAC	CAGCAGTTTCTCCAGAGC
hGAPDH	GCCTCAAGATCATCAGCAATGC	CCACGATAGCAAAGTTGTCATGC
hLCOR	AAGCTTACAGGATGGAACCAGG	CAGTGGAACTTTGAGTGATGTGG
hBIG	TCACAGCCCTCTACACTTGACG	AGGGTTTATCTCAGGAGGAGGC
hC10ORF12+BIG	GAGCTTCAGAGAGTGGAGACCC	TGTTCATCTACCACGGTGTCAAC
mEZH2	ACTGCTGGCACCGTCTGATG	TCCTGAGAAATAATCTCCCCACAG
mEZH1	TCCGATGGAAAGCAAGACGAC	GGCGCTTCCGTTTTCTTGTTAC
mINK4A	GTGTGCATGACGTGCGGG	GCAGTTCGAATCTGCACCGTAG
mPCL1	TTCTGCTCTTAACAGCCACAAGG	TTTTCGCTTTTTAATCTCCCTCC
mPCL2	CTCAAGGGAAGTAAGCAATGGG	AGGACGACCTACAGATTTTTTCTTTC
mPCL3	CATCTCCTCAACGCTCTCAACAG	TCTTAATTTCCTTGCCACACAGG

2.5 Protein extraction and expression analyses

2.5.1 Cell fractionation into cytosolic, nucleosolic and chromatin bound fractions

Approximately 80% confluent HEK293T cells were harvested by washing twice with PBS, scraping in PBS and pelleting by 5 min centrifugation 4°C at 1,500 rpm. The nuclei were isolated by incubating the cell pellet in 2-3 times its packed cell volume (pcv) of Buffer A (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 2 µg/mL aprotinin, 1 µg/mL leupeptin, 10 mM PMSF) for 15 min. This was followed by mechanical bursting of the swollen cells by dounce homogenising using 16 strokes with pestle type A ('loose'). The released nuclei were pelleted by 5 min centrifugation at 1,500 rpm at 4°C. The supernatant was retained as the cytosolic

fraction and nuclei washed once with buffer A, and either used for further fractionation or snap frozen and stored in -20°C (short term) or -80°C (long term). To isolate the nucleosolic fraction, nuclei were resuspended in 2-3 packed cell volume of buffer S1 (10mM HEPES pH 7.9, 120 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA), dounc homogenised with 15 times strokes using a pestle type B (tight) and centrifuged at 14,000 rpm at 4°C for 15 min. The supernatant was collected as the nucleosolic fraction and snap frozen, while the pelleted chromatin was resuspended in Buffer S2 (20 mM HEPES pH 7.9, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 2 µg/mL aprotinin, 1 µg/mL leupeptin, 10 mM PMSF), incubated with rotation at 4°C for 30 min, diluted to final NaCl concentration of 220 mM with buffer D (20 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 2 µg/mL aprotinin, 1 µg/mL leupeptin, 10 mM PMSF) and subjected to digestion with 250 U/mL benzonase-nuclease for 1.5 to 16 hours, prior to preclearing by centrifugation for 15 min at 14,000 rpm at 4°C. The chromatin bound protein containing supernatant was snap frozen and stored at -20°C. The three cellular fractions were then quantified by the Bradford method and analysed by Western blotting.

2.5.2 Preparation of whole or nuclear cell lysates

Nuclear or whole cell pellets were re-suspended in 4-5 pcv of IPH buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40 detergent, 0.5 µM DTT, 2 µg/mL aprotinin, 1 µg/mL leupeptin, 10 mM PMSF) and incubated for 30 min at 4°C with rotation. $MgCl_2$ was then added to a final concentration of 7 mM before the digestion of chromatin with 250 U/mL benzonase-nuclease overnight at 4°C with rotation. The digested lysate was precleared by centrifugation at 14,000 rpm for 15 min at 4°C, twice, to remove traces of cell debris. The protein

concentration of the sample was quantified by the Bradford method and adjusted to be equal between the different samples.

2.5.3 Western blotting

Protein samples were boiled for 6 min in SDS-loading buffer (final concentration: 60 mM Tris pH 6.8, 33% glycerol, 5% 2-mercaptoethanol, 2% SDS, 0.05% bromphenol blue), separated by SDS-PAGE and transferred to nitrocellulose membrane. The membrane was incubated in blocking buffer (5% non-fat dry milk (NFDM) in PBS containing 0.2% Tween) for 1 hour at room temperature or ~16 h at 4°C and subsequently probed with the desired primary and secondary antibodies, diluted in blocking buffer. The protein signal was detected by chemiluminescence using Immobilon Western HRP substrate (Millipore) on Kodak film.

2.6 Immunoprecipitations

2.6.1 Immunoprecipitation of exogenously expressed FLAG-tagged proteins

The IPH whole cell or nuclear lysates were pre-cleared to remove non-specifically binding proteins by incubation with mouse IgG agarose beads (Sigma) for 1 h at 4°C with rotation. FLAG-tagged proteins were immunoprecipitated from these lysates using anti-FLAG-M2 affinity gel agarose (Sigma; A2220) for 4-16 hours at 4°C with rotation. The agarose beads were washed with 1 ml of IPH buffer at 4°C for 5 min with rotation followed by centrifugation at 5,000 rpm at 4°C for 1 min. This step was repeated five times. After the final wash, the pelleted beads were re-suspended in 250 μg/ml 3xFLAG peptide (Sigma) in 0.05% NP40 and incubated

with shaking at room temperature for 30 min, for the competitive elution of FLAG-tagged proteins from the beads. The supernatant was separated from agarose beads by centrifugation at 5,000 rpm for 5 min, collected and then either frozen at -20°C for later use or analysed directly by Western blotting.

2.6.2 Immunoprecipitation of endogenous proteins

For a single immunoprecipitation, 1-4 µg of antibody was coupled to 20 µl of Protein A (Sigma P9424) or protein G-beads (Invitrogen, 101242) sepharose bead slurry by incubation in 0.5 ml PBS (0.1% Tween-20) with rotation for 4-16 hours at 4°C. Beads were collected by centrifugation at 5,000 rpm for 3 min and washed twice in 1 ml 0.2 M Sodium Borate, pH 9.0. Antibodies were then crosslinked to beads by incubation in 1 ml 0.2 M sodium borate pH 9.0 containing 20 mM dimethyl pimelimidate dihydrochloride at room temperature for 30 min with rotation. The reaction was terminated by washing beads once and incubating for 2 h at room temperature with rotation in 1 ml of 0.2M ethanolamine pH 8.0. The beads were then washed twice with PBS-Tween (0.1%) and twice with IPH buffer, blocked in 0.5 µg/ml BSA in IPH buffer for 1-4 h and used for IP on total or nuclear IPH lysates for 3-16 h, washed with 1 ml of IPH buffer at 4°C for 5 min with rotation followed by centrifugation at 5,000 rpm at 4C for 1 min. This step was repeated five times. The complexes were eluted in SDS loading dye (final concentration: 60 mM Tris, 33% glycerol, 5% 2-mercaptoethanol, 2% SDS, 0.05% Bromphenol blue) by boiling for 6 min at 99°C, cooling on ice for at least 2 min and analysed by Western blotting.

2.7 Mass-spectrometry preparation and analysis

2.7.1 In-gel tryptic digest

Immunoprecipitated samples of FLAG-PCL1/2/3 or control IPs were denatured in SDS-loading buffer, separated on 4–12% gradient Nu-PAGE gels (Novex) and stained with GelCode Blue solution (Thermo Scientific). Lanes were excised and cut into 10 fragments from the top to the bottom of the gel, each fragment diced into small (~1 mm^3) pieces and washed three times with 25 mM NH_4HCO_3/50% acetone, dehydrated in acetonitrile/NH_4HCO_3 (3:2) and rehydrated with 50 mM NH_4HCO_3 twice, followed by dehydration with undiluted acetonitrile and reduction of proteins in 10 mM DTT in 25 mM NH_4HCO_3 for 1 hour at 56°C and a subsequent incubation in 55 mM iodoacetamide for 1 h. Gel pieces were washed with 25 mM NH_4HCO_3, then 25 mM NH_4HCO_3/50% acetonitrile and dehydrated. The peptides were then digested with one half of Trypsin Singles reaction (Sigma, T7575) in 25 mM NH_4HCO_3 overnight at 37°C. Finally, peptides were extracted from the gel pieces with 50% acetone / 5% formic acid (Sigma, 94318) and dried by vacuum concentration. The final peptide sample was resuspended in 20 µl 0.1% formic acid.

The sample analysis was performed by collaborators in Dr. Gerard Cagney laboratory (UCD) with the following specifics on a Thermo Scientific LTQ Orbitrap mass spectrometer equipped with a nanoACQUITY UPLC (Waters) liquid chromatography mass spectrometry 31 system and a nanoelectrospray source by Kieran Wynne in the Conway Institute, UCD, Dublin. Each sample was injected onto a nanoACQUITY Symmetry C18 trap (5µm particle size, 180µm x 20mm) 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 1 h 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 x 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 32 ProteinProspector MS-Product program was used to calculate the theoretical masses of fragments of identified peptides for manual validation. The spectral count information for the proteins identified by MaxQuant was then manually analysed using Excel. Any proteins that were detected in the negative control immunopecipitations were regarded as background. Two independent IP experiments were considered, and only proteins scoring in both experiments were considered to be positive candidates. Where spectral counts for a given protein were equal to one in both experiments, the candidate was considered a false positive.

2.7.2 In-solution tryptic digest

In-solution tryptic digest was performed by Giorgio Oliviero in Conway Institute, UCD. After FLAG-LCOR, -C10ORF12 or 'BIG' IP, the proteins, while still bound to FLAG-agarose, were treated with trypsin as described (Wiśniewski et al., 2009). Peptide samples were introduced to Q Exactive mass spectrometer via an EASY-nLC 1000 UHPLC system (Thermo Fisher) coupled to an in-house packed C18 column (New Objective). Parent ion spectra (MS1) were measured at resolution 70,000, AGC target 3e6. Tandem mass spectra (MS2; up to 10 scans per duty cycle) were obtained at resolution 17,500, AGC target 5e4, collision energy of 25. Data were processed using MaxQuant version 1.3.0.5 (Cox & Mann, 2008) under default settings (peak identification and processing) with the human UniProt

database (release 2013_12; 67,911 entries). The following search parameters were used: Maximum Missed Cleavage: 2; Fixed Mod: cysteine carbamidomethylation; Variable Mods: methionine oxidation; Trypsin/P digest enzyme; Precursor mass tolerances 6 ppm; Fragment ion mass tolerances 20 ppm; Peptide FDR 1%; Protein FDR 1%. Immunoprecipitation experiments were carried out in duplicate to address biological reproducibility, while control runs (every 10 sample runs) containing peptide standard mixes were used to assess analytical reproducibility. The spectral count information for the proteins identified by MaxQuant was then manually analysed using Excel. Any proteins that were detected in the negative control immunopecipitations were regarded as background. Two independent IP experiments were considered, and only proteins scoring in both experiments were considered to be positive candidates. Where spectral counts for a given protein were equal to one in both experiments, the candidate was considered a false positive.

2.8 Chromatin Immunoprecipitation (ChIP)

The chromatin was chemically cross-linked by incubating a 70-80% confluent 15 cm tissue culture dish with 15 ml of 1% formaldehyde in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1,8 mM KH_2PO_4, pH 7.4), for 7 min at room temperature. Fixing was stopped by addition of glycine to a final concentration of 0.125 M and incubation for another 5 minutes. Fixed cells were then washed twice with PBS and lysed in 6 ml of SDS lysis buffer (50 mM Tris pH 8.1, 0.5% SDS, 100 mM NaCl, 5 mM EDTA, 1 µg/ml aprotinin, 10 µg/ml leupeptin, 1 mM PMSF), harvested, quick frozen and stored at -20°C until required. Frozen nuclei were thawed, pelleted by centrifugation at 1200 rpm for 5 min at room temperature and

resuspended in ice-cold ChIP Buffer containing protease inhibitors (2:1 mix of SDS lysis buffer and Triton dilution buffer (100 mM Tris pH 8.6, 100 mM NaCL, 5 mM EDTA pH 8.0, 5% Triton-X 100, 1 µg/ml aprotinin, 10 µg/ml leupeptin, 1 mM PMSF). Chromatin was sheared into fragments of approximately 200-1000 bp by sonication using Sonicator150 (12-14 pulses of 30 sec at Amplification 5 per sample). For each ChIP sample, the lysate was pre-cleared for unspecific binding by addition of 15 µl (30 mg/ml) of either protein A (Sigma, P9424) or G (Invitrogen) sepharose beads, which were previously blocked with 0.5 mg/mL lipid-free BSA (Sigma) and 0.2 mg/mL herring sperm DNA (Sigma) in TE, after which samples were centrifuged at 14,000 rpm at 4°C for 30 min to remove any precipitated proteins. The supernatants were immunoprecipitated with 2-5 µg of antibody overnight at 4°C with rotation. ChIP samples were centrifuged at 14,000 rpm at 4°C for 30 min to remove any further non-specific precipitated proteins. The antibody/protein complexes were purified by an incubation with 50 µL of blocked protein A or G beads for 4 h at 4°C. The beads were subsequently washed three times with 1ml of Mixed Micelle Wash Buffer (20 mM Tris, pH 8.1; 150 mM NaCl, 5 mM EDTA, 5% w/v sucrose, 1% Triton X-100, and 0.2% SDS), twice with 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), twice with 1 ml 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 once with 1 ml of TE (pH 8.0). The immunoprecipitated complexes were then eluted in ChIP elution buffer (1% SDS, 0.1M NaHCO) at 65°C for 2 h with continuous agitation and the cross-linking was reversed by an overnight incubation at 65°C. The eluted material was then digested with RNase (Thermo Scientific) for 1 h at 37°C and treated with 0.2 µg/µl proteinase K for 2 h at 55°C,

purified by phenol/chloroform-extraction and ethanol-precipitation and the DNA re-suspended in 100-150 µl of nuclease-free water (Sigma) for subsequent analysis. The DNA content of the ChIP material was determined using RT-qPCR method.

Chapter 3

Conserved and divergent roles of three human homologues of *Drosophila Polycomblike*

3.1 Introduction

The main questions in Polycomb Repressive Complex 2 (PRC2) biology are how these protein complexes are recruited to their target genes in a cell type specific manner and what mechanisms mediate the repression of PRC2 target genes. Since the core subunits of the PRC2 complex do not possess any DNA or histone specific binding functions, the PRC2 complex relies on its sub-stoichiometric components for specificity of binding to target genes as well as enzymatic potency (Bracken and Helin, 2009). One such sub-stoichiometric component is *Polycomblike*.

The Drosophila *Polycomblike* was identified by a mutagenesis screen as having an identical homeotic transformation phenotype as compared to the loss of *Polycomb* itself (Duncan, 1982). Similar to Polycomb, Polycomblike localises to the Polytene chromosome and salivary glands (Lonie et al, 1994). Eventually, *Drosophila* Polycomblike was shown to be a sub-stoichiometric component of PRC2 (Tie et al, 2003), which directly binds to E(z) protein via its plant homeodomain finger (O'Connell et al, 2001). It is required for repression and efficient tri-methylation of histone H3 at lysine 27 on the PRC2 target genes (Nekrasov et al, 2007, Savla et al, 2008). In *Drosophila* development, Pcl is associated with PRC2 at early embryonic stages (0-16 h), while binding to PRC2 could not be detected at a later embryonic stage (18-24 h), despite the sustained expression levels of both Pcl and PRC2 (Tie et al, 2003). This implies that Polycomblike conditionally dissociates with PRC2, while still performing important roles in cells.

In mammals, the *Drosophila* Pcl (dPcl) has three homologues: PCL1/PHF1 (also called Tctex3 in mice), PCL2/MTF2 (also known as M96) and PCL3/PHF19 (Inouye et al, 1994, Coulson et al, 1998, Kawakami et al, 1998, Wang et al, 2004). PCL1 and PCL2 proteins share 41% and 44% amino acid identity with PCL3, respectively (Wang et al, 2004). Similarly to the *Drosophila* Polycomblike, mammalian PCL1 and PCL3 are required for efficient catalysis of H3K27me3, but not H3K27me1 or H3K27me2, by PRC2 on target genes *in vitro* and *in vivo* and for repression of Polycomb target genes (Nekrasov et al, 2007, Sarma et al, 2008, Cao et al, 2008, Brien et al, 2012).

In addition, all three PCL proteins, similarly to their *Drosophila* Pcl counterpart, contain a TUDOR domain and two tandem plant homeodomain (PHD) fingers. Recently, the TUDOR domain of the PCL1 and PCL3 proteins was shown to be required for "reading" the H3K36me3 PTM, which is present on the gene bodies of active genes (Brien et al, 2013, Mussellman et al, 2012, Ballare et al, 2012, Cai et al, 2013, Qin et al, 2013). PCL3 is also required for the recruitment of the PRC2 complex to pluripotency genes during differentiation (Brien et al, 2013). All these observations point to PCL1-3 having important roles in modulating PRC2 function. However, since these studies were conducted in different cell models, whether the roles of PCLs in modulation of PRC2 function are redundant remains unanswered. Another question is whether the three homologues of the dPcl gained any novel functions to dPcl as they diverged during the course of evolution. For example, PCL1 was reported to be involved in DNA double strand break repair and to bind the tumour suppressor p53 (Hong et al, 2008, Yang et al, 2012). However, the involvement of PCL2 and PCL3 in this pathway remains unknown.

The aim of this study was to compare the biological and biochemical properties of the Polycomblike 1-3 proteins. Firstly, it was assessed whether the expression patterns of PCL1-3 resemble that of Polycomb in the model of replicative senescence. Next, in order to compare the interactomes of the PCL1-3 proteins, mass-spectrometry analysis of immunoprecipitated FLAG-HA-tagged PCL1-3 proteins was performed.

3.2 Results

3.2.1 PCL2 and PCL3, but not PCL1, are down-regulated in senescing MEFs and HMECs

In order to establish whether the three homologues of PCL1, PCL2 and PCL3, are regulated similarly to the PRC2 components during the establishment of cellular senescence, reverse transcription quantitative PCR (RT-qPCR) analysis using the primers to the PCL1, PCL2 and PCL3 transcripts was performed. To ensure that the effects observed were specific for the process of cellular senescence, rather than lineage specific changes, two different models for cellular senescence: the primary mouse embryonic fibroblasts (MEFs) and primary human mammary epithelial cells (HMECs) were used. Both cell lines show reduced rates of proliferation and stain positively for beta-galactosidase, a marker for cellular senescence, as they are passaged (Bracken et al, 2007, Lanigan et al, 2015). The cDNA of these cells, taken at three different passages from their establishment into cell culture and to senescence, was subjected to qPCR analysis. As expected, in both MEF and HMEC cells, the expression of the $p16^{INK4A}$ mRNA transcript increased by about 5-10 fold as cells entered cellular senescence (Figure 3.1), which is consistent with previously reported studies and confirms the functionality of the models of cellular senescence used. Also, as expected, the *EZH2* mRNA expression was down-regulated as cells were passaged (Bracken et al, 2007), while *EZH1* expression levels remained unchanged. *PCL2* and *PCL3* expression patterns followed that of *EZH2* and were down-regulated more than 50% in both MEF and HMEC cells, while the expression levels of *PCL1* was increased in senescing MEFs (Figure 3.1, A) and were unaffected in HMEC cells (Figure 3.1,

B). These results indicate that during cellular senescence, PCL2 and PCL3 expression appears to follow the expression patterns of PRC2 components EZH2 and EED and are down-regulated, while PCL1 expression is maintained on the transcription level, suggesting that PCL1 might have a unique role in senescent cells.

3.2.2 Mass spectrometric analysis of PCL1, PCL2 and PCL3 containing protein complexes

It has been previously reported that PCL1 was involved in DNA damage processes and in particular that it bound to the tumour suppressor p53 (Hong et al, 2008). Having observed that PCL1 expression remained unchanged during cellular senescence, while the expression of PCL2 and PCL3 were down-regulated, it was hypothesised that the interaction with p53 could be unique to the PCL1 protein and that this functional interaction with p53 could be the reason for the divergence in PCL1 expression in senescence, compared to other Polycomblike proteins. In order to test this hypothesis, and to identify any other differences among the proteins interacting with PCL1, PCL2 or PCL3, the proteins were affinity purified and analysed by tandem mass spectrometry (Figure 3.2, A). For isolation of the PCL1-3 containing complexes, stable HEK293T cell lines expressing the FLAG-HA-tagged PCL1-3 were generated by retroviral transduction of the pMINKIO expression vectors containing the newly cloned *PCL1* and *PCL2* gene coding sequences and the previously cloned *PCL3* gene (by Dr. Gerard Brien). The exogenous PCL1-3 and their interacting proteins were immunoprecipitated from the nuclear extracts of the

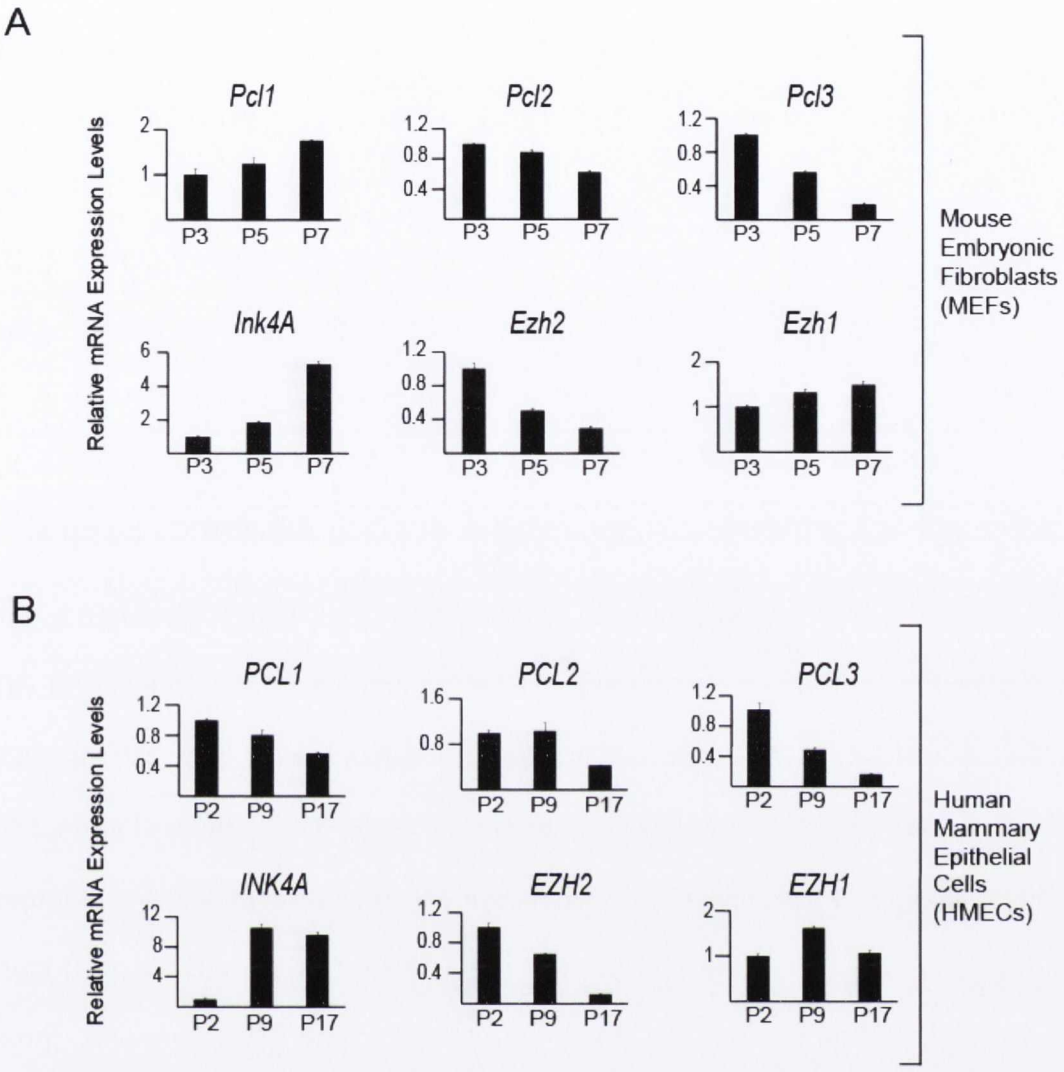


Figure 3.1 PCL2 and PCL3 are down-regulated in senescing MEFs and HMECs, while PCL1 is not.

RT-qPCR analysis of the mRNA expression levels of p16INK4A, EZH2, EZH1 and PCL1-3 in primary mouse embryonic fibroblasts (MEFs) (A) or human mammary epithelial cells (HMECs) (B) at the indicated passage numbers. The mRNA expression levels were normalized to the RPLPO housekeeping gene.

corresponding HEK293T cell lines using anti-FLAG agarose. A negative control IP was performed on nuclear extracts from HEK293T cells transduced with an empty expression vector. PRC2 component EZH2 was immunoprecipitated with all three PCL proteins, and served as a positive control for the approach (Figure 3.2, B). The immunoprecipitated material was then subjected to the in-gel tryptic digest and subsequently to mass-spectrometry analysis (Figure 3.2, C).

3.2.3 Endogenous PCL proteins do not co-exist in the same PRC2 complex

The mass spectrometric analysis of PCL1, PCL2 and PCL3 revealed a few observations. Firstly, core components of PRC2 complex, namely EZH2, EED, SUZ12 as well as EZH1 were all immunoprecipitated with each of the PCL proteins, further validating them as components of the PRC2 complex (Figure 3.3). Secondly, neither of the three PCL proteins appeared to interact with one another, in that PCL1 only pulled down PCL1, but not PCL2 or PCL3 and so on (Figure 3.3). Taken together with the fact that all three bind to the PRC2 complex, these data suggests that PCL1-3 proteins interact with PRC2 in a mutually exclusive manner.

In order to exclude the possibility that this observation is an artifact of exogenous expression of the PCL1-3 proteins, immunoprecipitations were performed using the antibodies to endogenous PCL1-3 proteins. This experiment confirmed that human PCL proteins do not bind to one another (Figure 3.4). These data support the hypothesis that the PCL1, PCL2 and PCL3

proteins do not co-exist in the same PRC2 complex and allows for the interpretation of the PCL1-3 IP-mass spectrometry datasets obtained as three independent interactomes.

3.2.4 PCL1 is unique in that it immunoprecipitates both p53 and EZH2

The tumour suppressor p53 was also identified in IP mass spectrometry analysis of PCL1-3. It immunoprecipitated specifically in PCL1, but did not in the PCL2 IP and was detected at the negative control IP background levels in the PCL3 IP (Figure 3.3). In order to eliminate the possibility of any artifact arising due to retroviral transduction, FLAG-IPs were performed on HEK293T cells that were transiently transfected with the expression vectors for PCL1-3 or empty vector (EV) as a negative control. Western blotting confirmed that only PCL1, but not PCL2 or PCL3, is capable of binding to both p53 and EZH2 (Figure 3.5).

Next, to establish whether the binding of PCL1 to p53 and PRC2 occurs simultaneously in the same protein complex, or whether these binding events are mutually exclusive, immunoprecipitations using antibodies specific to endogenous EZH2 and p53, and the PRC1 component BMI1 as a negative control were performed, followed by Western blotting for PCL1 and the target proteins (Figure 3.6). As expected, EZH2, but not BMI1, pulled down the PCL1 protein, further consolidating that PCL1 is a sub-stoichiometric subunit of PRC2 and not PRC1. On the other hand, neither EZH2 nor BMI1 immunoprecipitated p53 protein, suggesting that p53 does not interact with PRC2 or PRC1. Accordingly, p53 did immunoprecipitate PCL1, but not EZH2 or BMI1, supporting the previous observation that PCL1 binding to p53 and PRC2 is mutually exclusive (Figure 3.6).

Interestingly, the endogenous EZH2 and p53 appeared to pull down PCL1 proteins of different sizes. There may be two explanations for this observation. One possibility is that since PCL1 protein has two isoforms: a shorter isoform 'a' and a longer isoform 'b', the endogenous IP of EZH2 and p53 might suggest that isoform 'a' is preferentially binding to p53, and isoform 'b' – to EZH2. However this scenario is not consistent with the study of the exogenous PCL1 IP, for which isoform 'b' used, suggesting that it is also capable of pulling down p53 as well as PRC2 (Figure 3.5). This implies that on an endogenous level the isoform 'b' is preferred by PRC2 and isoform 'a' by p53, while of the abundance of isoform 'b' may be forced by overexpression of this protein, so that it becomes capable of interacting with p53. Another possibility is that both of the PCL1 bands observed are isoform 'b' (Figure 3.6), but the one that binds to PRC2 is subject to a post-translational modification.

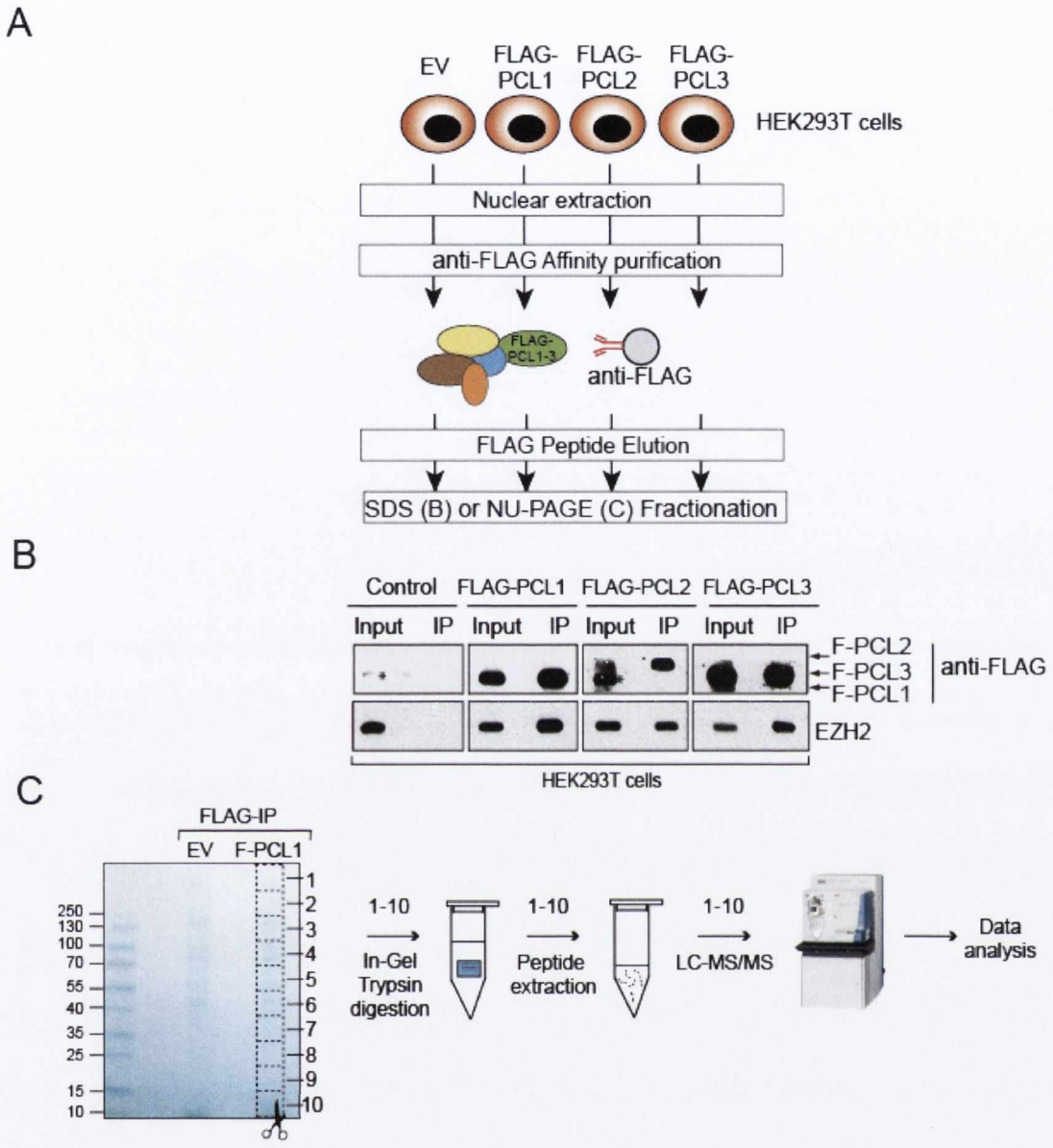


Figure 3.2 Affinity purification of FLAG-PCL1, FLAG-PCL2 and FLAG-PCL3 in HEK293T cells.

A. Experimental outline for FLAG affinity purification of FLAG-HA-tagged PCL1, PCL2 and PCL3. FLAG IPs were performed on nuclear extracts from HEK293T cells stably infected with pMINKIO empty or pMINKIO PCL1-3 expression vectors.
B. Western blot analysis of FLAG affinity purified PCL1-3 and negative control IP, as well as inputs separated on SDS-PAGE gel from A. EZH2 served as positive control for the PCL IPs.
C. A representative example of the in-gel tryptic digestion procedure for FLAG-PCL1 IP. The IP material was separated on a 4-12% NU-PAGE gel. The protein lanes were cut into 10 roughly equally sized pieces, which were then subjected to in-gel trypsin digestion. The extracted peptides were identified by LC-MS/MS and subsequently analysed by MaxQuant software.

		Mass spectrometry analysis of FLAG-PCL1-3 proteins in HEK293T cells				
		PCL1	PCL2	PCL3	Control	Protein name
PCL family	PHF1	33.5	0	0	0	PHD finger protein 1
	MTF2	0	17.5	0.3	0	Metal-response element-binding transcription factor 2
	PHF19	0	0	20.7	0	PHD finger protein 19
PRC2	EZH2	31.5	18.5	29.7	0	Histone-lysine N-methyltransferase EZH2
	SUZ12	33.5	22.5	35.3	0	Polycomb protein SUZ12
	EED	19	14.5	20.7	0	Polycomb protein EED
	EZH1	1.5	0.5	6.7	0	Histone-lysine N-methyltransferase EZH1
	C17ORF96	0	0.5	2.0	0	Uncharacterized protein C17orf96
	LCOR	1.5	0	0	0	Ligand-dependent corepressor
	C10ORF12	22	1.5	0	0	Uncharacterized protein C10orf12
	p53	3	0	0.5	0.5	Cellular tumour antigen p53
	NO66	4.5	0	0	0	Lysine specific demethylase NO66

Figure 3.3 Summary of mass spectrometric analysis of FLAG-PCL1-3 IPs in HEK293T cells.

This table summarises the top scoring proteins and their peptide counts in the FLAG-PCL1-3 IP-mass spectrometric analysis. The peptide counts represent the average of at least two independent experiments.

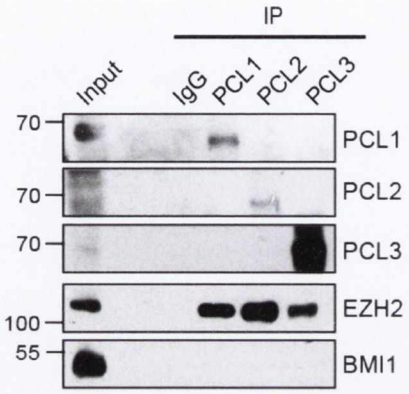


Figure 3.4 Endogenous Polycomblike proteins do not co-exist in the same PRC2 complex

Western blot analysis of immunoprecipitations (IP) using specific antibodies for PCL1, PCL2, PCL3 and rabbit IgG, as a negative control, from nuclear extracts of HEK293T cells. The PRC2 component, EZH2, served as a positive control, while the PRC1 component, BMI1, served as a negative control.

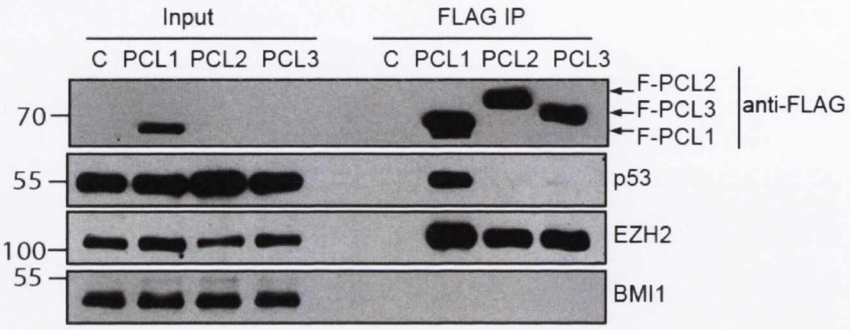


Figure 3.5 FLAG-PCL1 is unique in that it immunoprecipitates both p53 and EZH2

Western blot analysis of FLAG- immunoprecipitations of nuclear lysates from HEK293T cells transiently transfected with FLAG-HA-tagged PCL1, PCL2 and PCL3 expression vectors. The FLAG-IP of nuclear extracts from HEK293T cells transfected with empty vector served as the negative control ("C").

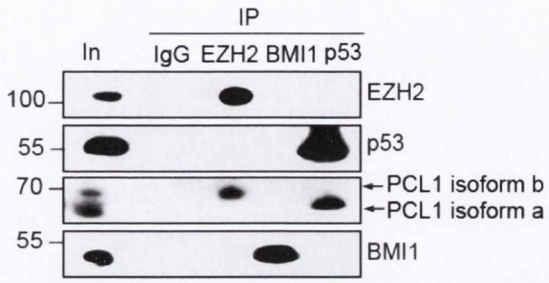


Figure 3.6 Endogenous p53 preferentially interacts with a shorter PCL1 a-isoform, while EZH2 pulls down the longer b-isoform

Western blot analysis of endogenous IP of PRC2 core subunit, EZH2, PRC2 component, BMI1, and endogenous p53 from nuclear extracts of HEK293T cells.

3.3 Discussion

This chapter describes a divergent role of Polycomblike protein PCL1 from its homologues PCL2 and PCL3. It was observed that unlike the case of PCL2, PCL3 and EZH2, the expression of PCL1 is not downregulated and remains relatively unchanged as cells undergo cellular senescence (Figure 3.1). Furthermore, PCL1, PCL2 and PCL3 subuntis are mutually exclusive to each other within the PRC2 complexes. In addition, these results confirmed previously published observations that PCL1 interacts with tumour suppressor p53 (Hong et al, 2008, Yang et al, 2012). The data presented here expands on this observation by showing that the interaction with p53 is exclusive to PCL1 and that it occurs outside the PRC2 complex. These results suggest that PCL1 has a unique PRC2 independent role in proliferating and senescing cells, which might involve the tumour suppressor p53.

However, there is a limitation in studying the exogenously expressed proteins. The study involves over-expression of the protein of interest, which can cause either false positive interactions due to oversaturation of the bait proteins in the cells or false negatives if the protein of interest is expressed on a sub-endogenous level. While the PCL1, PCL2 and PCL3 proteins were successfully over-expressed, the levels of their overexpression were not equal. As reflected in Western blotting analysis and spectral counts of the bait proteins, the expression levels of exogenous FLAG-tagged PCL1 were higher than those of PCL3, while PCL2 was overexpressed the least (Figure 3.3, Figure 3.5). This could explain why the histone demethylase NO66, which was previously reported to bind to PCL3 (Brien et al, 2013), was detected in PCL1 but not PCL2 or PCL3 IPs. Co-immunoprecipitations of endogenous PCL1-3 with p53 would allow to rule out the

possibility that p53 in the PCL1 IP was an artifact of the high overexpression of PCL1 compared to the other two PCLs. In fact, co-immunoprecipitations of endogenous PCL1-3 in human fibroblasts further validated the uniqueness of the PCL1 interaction with p53 (Brien et al, in submission).

The observation that PCL2/3 were down-regulated in cellular senescence, while PCL1 was not, also poses the question whether all three PCLs are involved in Polycomb mediated silencing of the *INK4A/ARF* locus. Interestingly, the study subsequently performed by Dr. Gerard Brien in the laboratory of Dr. Adrian Bracken confirmed that PCL1-3 were capable of binding and silencing the INK4A promoter. Interestingly, while the overexpression of PCL2/3 in primary human fibroblasts conferred growth advantage, PCL1 overexpression only showed such growth advantage in the absence of p53. Furthermore, PCL2/3 and EZH2 were down-regulated in G_0 quiescent cell state, while PCL1 expression remained unchanged throughout the cell cycle, consistent with observations made in senescent cells in this chapter. These data further consolidate the requirement of PCL1 for proper cell cycle progression.

The binding capability of PCL1 to p53 was attributed to evolutionary neofunctionalisation of PCL1 from its homologues PCL2/3 (Brien et al, in submission), while PCL1 still retained its function as a repressor of the *INK4A* locus. Therefore it remains to be answered whether PCL1 is involved in repression of the PRC2 target genes in quiescent cells and what effect, if any, does the interaction with p53 have on this function.

EZH1 was reported to have a methyltransferase activity when in complex with EED and SUZ12 (Margueron et al, 2008, Shen et al, 2008). EZH1 expression levels are comparable between both rapidly proliferating and growth arrested cells and tissues, while EZH2 levels decrease as cellular proliferation decreases (Margueron et al, 2008). Since both PCL1 and the PRC2 component EZH1 are co-expressed in senescent cells (Figure 3.1), it becomes possible that PCL1 and EZH1 could act together in the absence of proliferation in order to maintain the gene repression patterns previously established by EZH2-PRC2. This hypothesis could be tested by co-immunoprecipitations of EZH1 and PCL1 in growth arrested versus proliferating cells, as well as ChIP analysis of these proteins on PRC2 target genes. The precise composition of the PCL1-EZH1 PRC2 complex could then be identified by IP-mass spectrometry analysis of PCL1 and EZH1 in quiescent cells.

Albeit interactomes of the exogenously expressed PCL1-3 proteins may not be directly comparable due to unequal expression levels of the bait proteins, the mass spectrometry analysis performed in this study identified several novel putative PCL interacting proteins. Interestingly, C17ORF96, also known as esPRC2p48, a protein recently found to be associated with PCL2 containing PRC2 (Wang et al, 2011, Alekseyenko et al, 2014, Liefke and Shi, 2015), was detected specifically in the PCL3 and PCL2 IPs, but not in the PCL1 IP, suggesting that these sub-stoichiometiric components of PRC2 work together. In addition, the IP-MS identified the nuclear receptor co-repressor LCOR specifically in the PCL1 IP, while the predicted uncharacterised protein C10ORF12 was detected in PCL1 and PCL2 IPs (Figure 3.3). Simultaneously to this study, these proteins were also

identified in proteomic studies of PRC2 core component EED and EZH2 (Smits et al, 2013, Alekseyenko et al, 2014), suggesting they are all sub-stoichiometric components of the PRC2 complex. Hence our data suggest that C17ORF96 is a sub-stoichiometric component of the PCL2/3-containing PRC2 complexes, LCOR is present in the PCL1-containing PRC2 complex, while C10ORF12 is found in both PCL1- and PCL2-containing PRC2 complexes. Further characterisation of these protein interactions by co-immunoprecipitation followed by Western blotting and functional characterisation of their relevance to PRC2 biology is required.

Chapter 4

Identification of a PCL1/2 interacting protein 'BIG' – a novel alternative splicing variant of *LCOR* gene locus

4.1 Introduction

In Chapter 3 it was found that the peptides for C10ORF12, a product of a predicted gene on human chromosome 10, were significantly enriched in the PCL1 and PCL2 IPs, while none were detected in the PCL3 IP (Figure 3.3). LCOR protein was also detected as a novel PCL1 interacting partner, but did not appear to interact with PCL2 or PCL3 proteins.

No molecular function has so far been attributed to the C10ORF12 protein. The evidence that it exists has so far only come from proteomic studies. Recently, the label-free quantitative mass spectrometry analysis of GFP-tagged PRC2 component EED revealed that the stoichiometry of the C10ORF12 in the PRC2 complex was estimated to be about 6% relative to core PRC2 subunits (Smits et al, 2013). Strikingly, the stoichiometry estimate for PCL1 protein in the PRC2 complex was ~5%. Taken together with the observation made in chapter 3, that the numbers of peptides detected for C10ORF12 in the PCL1 IP-Mass spectrometry analysis was in the same range as the core PRC2 components, these data suggest that PCL1 predominantly interacts with PRC2 complex in the presence of the C10ORF12 protein.

Ligand dependent co-repressor LCOR was initially identified in a yeast two-hybrid screen for the proteins that physically interact with estrogen receptor alpha (ERα) (Fernandes et al, 2003). LCOR protein possesses a nuclear receptor (NR) binding box LXXLL motif on its N-terminus (Fernandes et al, 2003). LXXLL motifs are usually present on co-activators, as the feature that binds to the hydrophobic pocket in the ligand-binding domain of the nuclear receptor when it is activated,

that is bound by an agonist (Gurevich et al, 2007). Surprisingly, LCOR was characterised as a co-repressor of nuclear receptors under activating conditions. Luciferase reporter studies with various nuclear receptor DNA response elements, such as estrogen, thyroid, androgen, progesterone, glucocorticoid and vitamin D receptors, showed that LCOR over-expression mediated repression of the nuclear receptor response elements specifically in the presence of their respective activating ligands (Fernandes et al, 2003, Palijan et al, 2009a, Asim et al, 2011, Song et al, 2012). LCOR biding to ERα in the presence of estradiol was also confirmed *in vitro* and by co-immunoprecipitation studies in MCF7 breast cancer cells (Fernandes et al, 2003). Similarly, LCOR co-localized with ERα on a cohort of ERα target genes upon estradiol treatment, however knockdown of LCOR caused a decrease in expression levels of some ERα target genes, implying that it might possibly also possess an activating function (Palijan et al, 2009b), depending on the chromatin context.

LCOR mediated repression was shown to be histone deacetylase (HDAC) and c-terminal binding protein (CtBP) dependent, since treatment with HDAC inhibitor trichostatin A (TSA) or the mutation of the CtBP-binding protein motifs, which are located in the N-terminus of the protein, were sufficient for abolishing the repression of the reporter genes (Fernandes et al, 2003, Palijan et al, 2009b). In addition, LCOR, also known as MLR2, was identified as a sub-stoichiometric component of PRC2 in the Smits *et al* study, with the stoichiometry of LCOR in PRC2 estimated to be below 1%, which is significantly smaller than that of PCL1 or C10ORF12. Since we detected LCOR as a specific PCL1 interacting partner in chapter 3, these data suggest that LCOR/MLR2 is a sub-stoichiometric component

of PRC2-PCL1 complexes. Furthermore, we hypothesised that the ability of LCOR to directly bind to the nuclear receptors could act as a novel mechanism of PRC2 recruitment to their target genes.

The aim of this chapter was to further characterise the interactions between PCL1 and the novel PRC2 sub-stoichiometric components LCOR and C10ORF12. This lead to the discovery of a new protein that we called 'BIG', which is a product of an alternative splicing of the *LCOR* gene locus, as confirmed by our validations using RT-PCR, RNA-interference and Western blotting analyses.

4.2 Results

4.2.1 The *C10ORF12* gene is downstream of the *LCOR* gene on human chromosome 10

In order to clone the *C10ORF12* gene for overexpression and biochemical studies, the locus encompassing this gene was examined in the UCSC genome browser. Intriguingly, the predicted *C10ORF12* gene is located ~10 kb directly downstream of the *LCOR* gene on human chromosome 10 (Figure 4.1). Four known alternative splicing variants of LCOR have been characterised so far, which mostly differ in their 5' and 3' un-translated regions (UTRs) and in three cases out of four give rise to that same protein. Since C10ORF12 and LCOR both specifically interact with PCL1, we hypothesised that the peptides detected in the PCL1 IP could arise from the previously un-annotated alternative mRNA splicing event from the *LCOR* gene locus, whereby C10ORF12 is incorporated into the LCOR transcript. Indeed, an NCBI database of mRNA sequences (AceView), which was designed to predict alternative mRNA splicing events, contains an entry for the 'LCORandC10ORF12' alternative splice variant, which here is labeled as "Predicted 'BIG'" in Figure 4.1 (Thierry-Mieg and Thierry- Mieg, 2006).

Examination of genome-wide ChIP-Seq tracks of H3K4me3, a histone PTM associated with gene promoters (Ernst et al, 2011), along the *LCOR* gene locus in human embryonic stem (ES) cell and in a panel of human derived cell lines and tissues, revealed that H3K4me3 is present on the *LCOR* gene promoter, as expected, but not anywhere near the 5' of the C10ORF12 coding region (Figure 4.1, bottom). This further supports the idea that 'BIG' is an alternatively spliced

variant of the *LCOR* gene. Furthermore, the *C10ORF12* coding region is transcribed in human ES cells as is evident from the RNA-sequencing tracks (Figure 4.1). Taken together, these data suggest that the 'BIG' protein is encoded by a novel alternative splice variant of the *LCOR* gene locus. The third coding exon of LCOR is skipped in the '*BIG*' mRNA transcript, meaning the C-terminus of the LCOR protein is not part of the 'BIG' protein.

4.2.2 Peptides of LCOR and C10ORF12 are detected at higher molecular masses than expected in FLAG-PCL1 IP-MS

To test whether it is, indeed, the novel 'BIG' protein, and not individual LCOR or C10ORF12 proteins, that are physically associated with PCL1, we referred back to the data obtained in the PCL1 IP mass spectrometry (performed in Chapter 3, Figure 3.2). Using the in-gel tryptic digest peptide preparation method for mass spectrometry, it is possible to tell what approximate location in the gel the identified peptides were detected in. Therefore, it was hypothesised that if the 'BIG' protein was pulled down by PCL1, the peptides from the LCOR protein, which is expected to run at ~50 kDa, and for C10ORF12, which is predicted to be ~130 kDa, would score higher than their predicted sizes. Indeed, the peptides representing LCOR and C10ORF12 were detected at the top of the gel at size above 200 kDa, which is greater than would be expected, while the peptides from PCL1 and the three core PRC2 components, EZH2, SUZ12, EED as well as p53 were identified at their expected sizes on the gel (Figure 4.2). These observations suggest that PCL1 immunoprecipitates the product of the novel LCOR alternative splicing variant 'BIG' protein, but not LCOR in its currently annotated form.

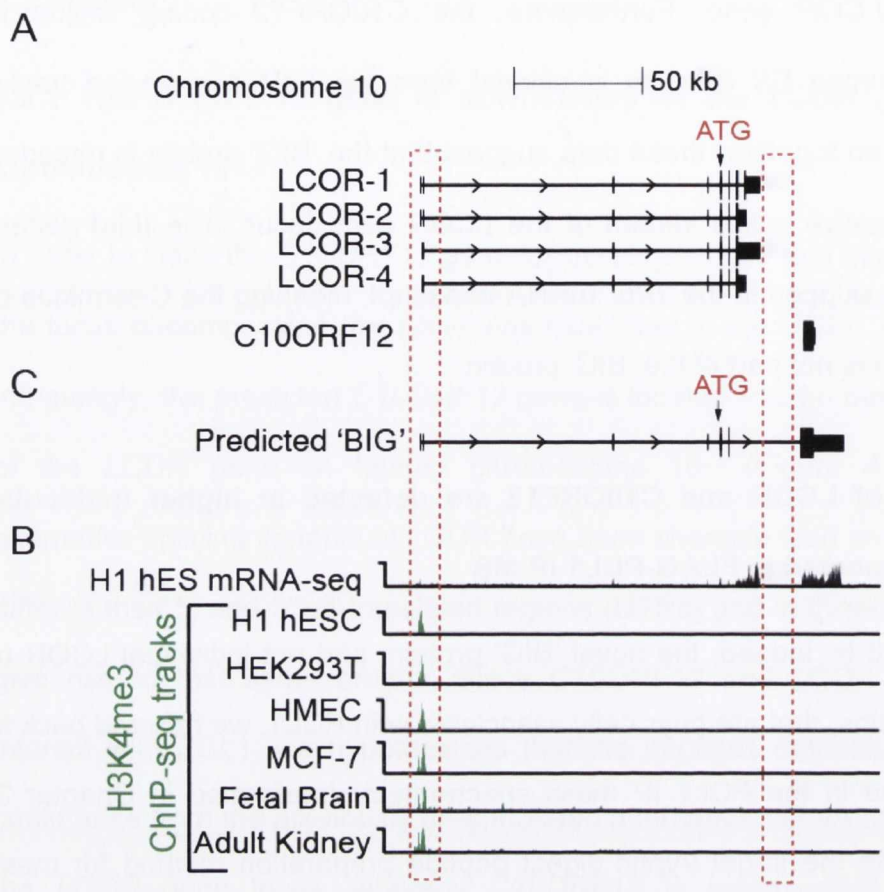
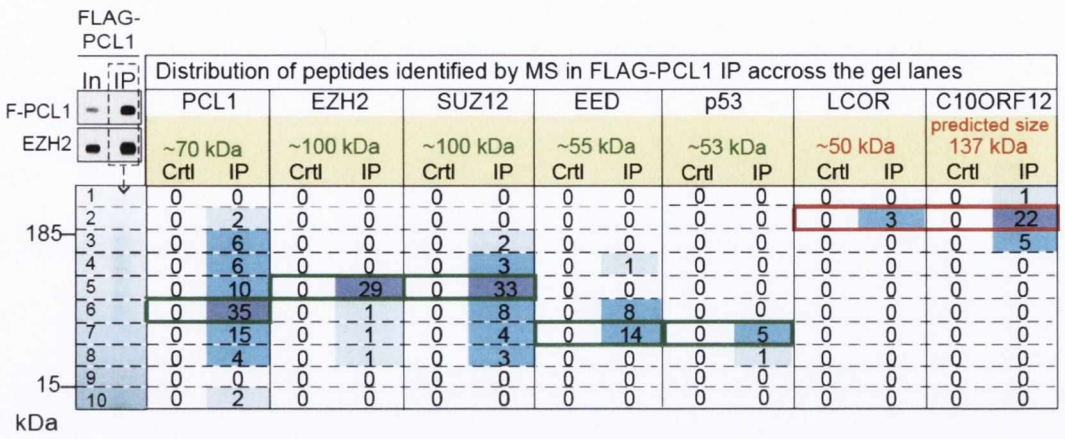


Figure 4.1 The LCOR gene and predicted C10ORF12 are localised on chromosome 10 and likely share a common promoter.

A. A view of the *LCOR* and *C10ORF12* containing locus on human chromosome 10, depicting the non-coding exons (short rectangular boxes) and coding exons (tall rectangular boxes) of the *LCOR* alternative splicing variants (1-4) and the predicted *C10ORF12* gene, with the arrows along the transcript symbolising the direction of transcription and ATG marking the start of the protein coding sequence.

B. Represented are RNA-seq of human embryonic stem cells and H3K4me3 (promoter histone post translational modification) ChIP-seq tracks in various human cell lines and tissues taken from the ENCODE and Epigenetic Roadmap projects along the *LCOR* locus.

C. The sequence of the novel predicted LCOR alternative splicing variant, as proposed by AceView database, which contains the entire C10ORF12 predicted ORF is named 'BIG' here.



Distribution of peptides identified by MS in FLAG-PCL1 IP accross the gel lanes

	PCL1		EZH2		SUZ12		EED		p53		LCOR		C10ORF12	
	~70 kDa		~100 kDa		~100 kDa		~55 kDa		~53 kDa		~50 kDa		predicted size 137 kDa	
	Crtl	IP	Crtl	IP	Crtl	IP	Crtl	IP	Crtl	IP	Crtl	IP	Crtl	IP
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2	0	2	0	0	0	0	0	0	0	0	0	3	0	22
3	0	6	0	0	0	2	0	0	0	0	0	0	0	5
4	0	6	0	0	0	3	0	1	0	0	0	0	0	0
5	0	10	0	29	0	33	0	0	0	0	0	0	0	0
6	0	35	0	1	0	8	0	8	0	0	0	0	0	0
7	0	15	0	1	0	4	0	14	0	5	0	0	0	0
8	0	4	0	1	0	3	0	0	0	1	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	2	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4.2 Peptides of LCOR and C10ORF12 are detected at higher molecular mass than expected in FLAG-PCL1 immunoprecipitations.

Western blot analysis of material immunoprecipitated (IP) with FLAG antibody from nuclear lysates of HEK293T cells stably expressing FLAG-HA-PCL1 is depicted on top left panel. The Coomasie staining of material immunoprecipitated in the FLAG-PCL1 IP is represented below. In the table, the lanes containing the control and FLAG-PCL1 IPs were cut into 10 and each piece was subjected to in-gel tryptic digest and the extracted peptides were subjected to mass spectrometric analysis. The peptide counts of the indicated proteins are listed in the table adjacent to the gel band they were isolated from. Green boxes encompass the peptide count where the expected size correlates with the observed size, while red boxes encompass the peptide counts of LCOR and C10ORF12, which were detected in a band corresponding to higher than expected molecular mass. These results are representative of two independent experiments.

4.2.3 *'BIG'* is a novel splice variant of the *LCOR* gene that extends to incorporate *C10ORF12*

Next it was aimed to characterise the exact gene and amino acid sequence of the novel 'BIG' protein. Provided that the 'BIG' coding sequence started with exons from LCOR and ended with the C10ORF12 stop codon, primers were designed to the beginning and then end of the LCOR and C10ORF12 coding sequences, as they are currently annotated in NCBI, and used them in varying combinations for RT-PCR analysis on HMEC cDNA (Figure 4.3, A). The LCOR and C10ORF12 coding sequences were successfully amplified, indicating that both LCOR and a transcript containing the C10ORF12 sequence are expressed in HMEC cells (Figure 4.3,m B). Importantly, a 4674 bp *'BIG'* open reading frame was also amplified using the forward LCOR primer and the reverse C10ORF12 primer (Figure 4.3, B), suggesting that the novel 'BIG' alternative splice variant of LCOR, generated by incorporation of the C10ORF12 sequence, is indeed expressed in primary human cells.

The 4674 bp *'BIG'* open reading frame (ORF), obtained from the RT-PCR analysis was subsequently cloned, along with the *LCOR* and *C10ORF12* ORFs, and then sequenced by Sanger sequencing. The DNA sequencing confirmed that the 'BIG' coding sequence is obtained by splicing of the first two coding exons of LCOR to C10ORF12 via a region of DNA directly upstream of the C10ORF12 predicted ORF, which we refer to as the 'BIG' unique sequence. Since the third and fourth exons of LCOR are skipped in the alternative splicing of 'BIG', the C-terminus of the LCOR protein is absent in the 'BIG' protein. Therefore the resulting 'BIG' protein sequence starts with the first 111 amino acids of the LCOR protein on the

N-terminus, followed by a 199 amino acid sequence unique to 'BIG' and finally 1247 amino acids encoded by C10ORF12 on the C-terminus.

To seek validation for this protein, we mapped the peptides identified in the PCL1-IP mass spectrometry to the predicted 'BIG' amino acid sequence. Reassuringly, all peptides significantly scoring for LCOR mapped to the region that is shared between the LCOR and 'BIG' proteins (Figure 4.4). Similarly, the coverage on the C-terminus of 'BIG' was also evenly distributed by the mapping of the C10ORF12 peptides. Importantly, the custom search for peptides mapping to the 'BIG' unique sequence, which is not yet present in any databases, yielded two peptides, providing further evidence for the existence of the 'BIG' protein.

4.2.4 The *C10ORF12* predicted ORF is only expressed as part of the *'BIG'* mRNA transcript

While the evidence obtained so far strongly suggests the existence of the novel LCOR alternative splicing variant product 'BIG', it does not exclude the existence of the independent entity of the C10ORF12 protein *per se*. While it appears that LCOR and C10ORF12 genes share a promoter (Figure 4.1), it may also be possible that C10ORF12 is only expressed as a part of the 'BIG' transcript.

To test this hypothesis, four shRNA constructs were designed to target specific exons which would either deplete *LCOR* on its own, *LCOR* and *'BIG'* together, *'BIG'* on its own or *'BIG'* and *C10ORF12* together. Then HEK293T cell lines were generated that stably expressed these shRNAs (Figure 4.5, A). As expected, the shRNA targeting the *LCOR* specific exon reduced the mRNA levels of LCOR,

compared to cells treated with scrambled (scr) control, but did not affect levels of the transcripts recognized by qPCR primer sets 2 ('*BIG*' specific primers) or 3 ('*C10ORF12*' and '*BIG*' primers) (Figure 4.5, B). In contrast, the shRNA designed to target the predicted first translated common exon of the *LCOR* (shLCOR+BIG) and the '*BIG*' mRNA transcripts, reduced mRNA levels of both *LCOR* and the transcripts recognized by qPCR primer sets 2 or 3, again supporting the hypothesis that '*BIG'* is an alternatively spliced variant of the *LCOR* gene. Notably, this shLCOR+BIG construct also reduced the levels of *C10ORF12* (Figure 4.5, B). In addition, the shRNA targeting a 'BIG'-specific region, but not *C10ORF12* (shBIG), reduced both the levels of *'BIG'* and *C10ORF12* mRNA, but not levels of *LCOR* mRNA. Finally, the shRNA designed to target both *C10ORF12* and '*BIG'* (shC10+BIG) reduced both *'BIG'* mRNA (primer set 2) and *C10ORF12* and *'BIG'* mRNA (primer set 3). These data confirm that *C10ORF12* is only expressed as a subset of the *'BIG'* transcript in HEK293T cells.

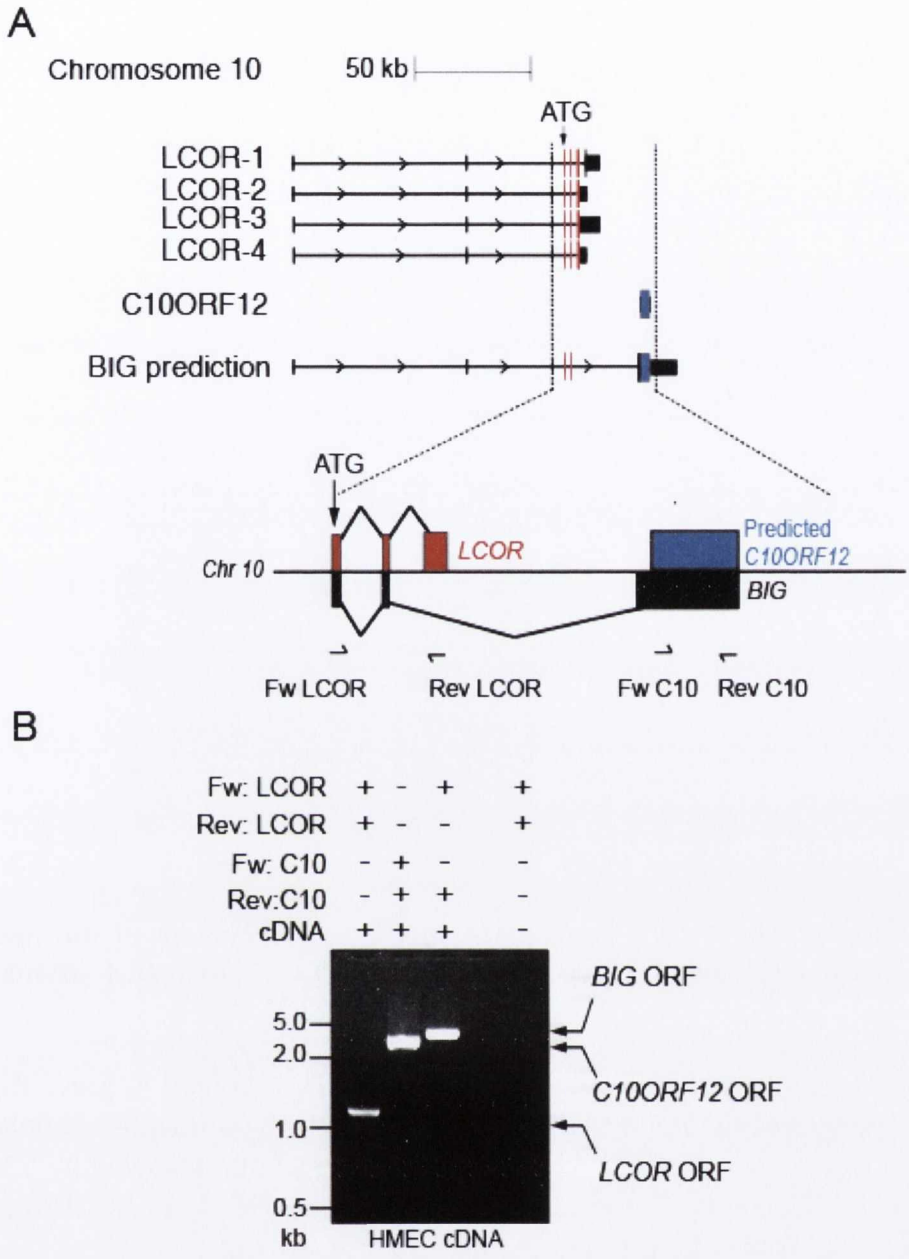


Figure 4.3 RT-PCR confirms that 'BIG' is a novel spice variant of LCOR gene that extends to C10ORF12.

A. UCSC genome browser view of the LCOR gene locus, with a zoom schematic into the coding exons (CDS) of both LCOR (in red) and C10ORF12 (in blue), and the predicted novel alternative splicing variant called 'BIG' (in black). 'BIG' is produced by skipping the third coding exon of LCOR and instead splicing the second coding exon of LCOR to a region just upstream of C10ORF12, as predicted by the AceView database.

B. RT-PCR analysis to confirm that 'BIG' is indeed an alternative splice form of the LCOR gene locus. The locations of the primers used are indicated in panel A. of LCOR, C10ORF12 and 'BIG' transcripts were amplified from primary human mammary epithelial cell (HMEC) cDNA using the primers indicated in panel A. The 'BIG' ORF was subsequently cloned and verified by DNA sequencing.

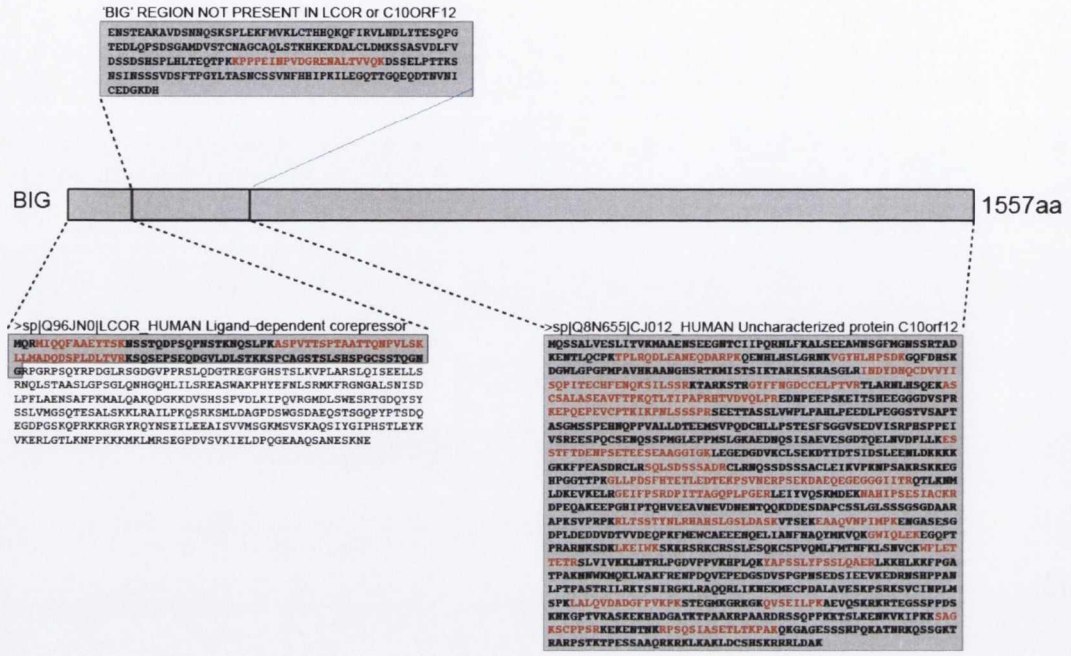
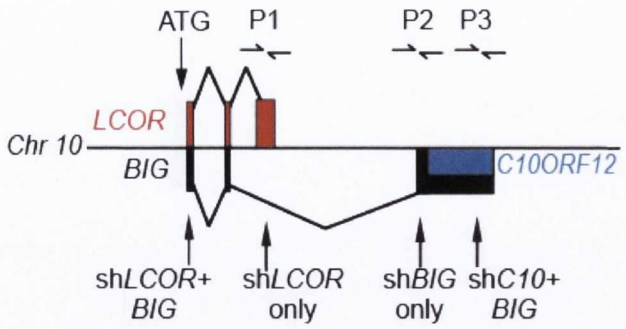


Figure 4.4 Representation of the complete amino acid sequence of the novel 'BIG' protein depicting the peptide coverage in the FLAG-PCL1 IP-mass spectrometry.

A schematic of the complete 'BIG' protein sequence is represented in grey. 'BIG' consists of 111 amino acids of the LCOR protein, followed by a unique sequence of 199 amino acids, not yet annotated in protein databases, and followed by 1247 amino acids of C10ORF12. The peptides identified in PCL1 FLAG-IP-mass spectrometry are highlighted in red.

A



B

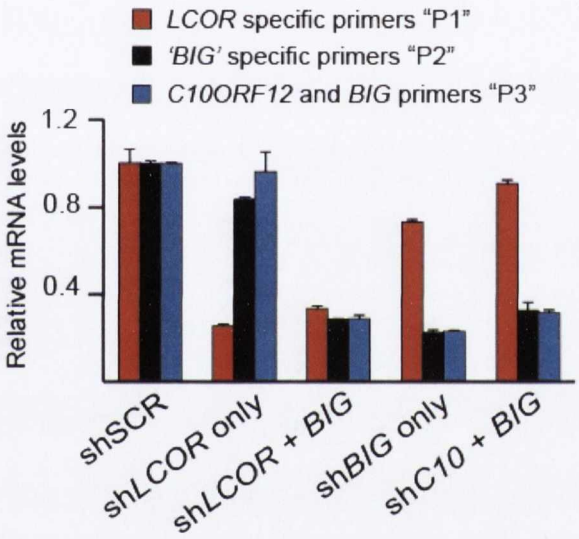


Figure 4.5 The C10ORF12 predicted ORF is only expressed as part of the 'BIG' mRNA transcript.

A. A schematic representation of the alternative splicing of the coding exons of LCOR (red), C10ORF12 (blue) and the novel 'BIG' variant (black). The locations of primer sites used for RT-qPCR and the regions targeted by shRNA are indicated. B. RT-qPCR analysis of LCOR (red), 'BIG' (black) and C10ORF12 (blue) mRNA levels in HEK293T cells stably expressing the shRNA targeting either the exon shared by both LCOR and 'BIG' (shLCOR+BIG), the LCOR specific exon (shLCOR), C10ORF12 predicted gene (sh C10+BIG) or the region of DNA coding sequence unique to 'BIG' ORF (shBIG), as indicated in panel A. The mRNA quantifications were normalised to the mRNA levels of the RPLPO housekeeping gene.

4.2.5 Characterisation of the 3'UTR of the 'BIG' alternative splice form of the LCOR gene

During the analysis of the 'BIG' and C10ORF12 transcripts, some discrepancies between the NCBI and AceView predictions and mRNA-seq from the locus for the C10ORF12 gene locus in the region of the 3' UTR were noted. In particular, the NCBI predicts the shortest 3'UTR. AceView predicts a longer 3'UTR, while the mRNA-seq in human ESCs predicts the 3'UTR to extend by about 10 kb from the stop codon (Figure 4.1). In order to investigate the extent of the 'BIG' mRNA, the data obtained from RNA-seq of early passage HMEC cells (generated by Dr. Gerard Brien) was aligned to the 'BIG' locus (Figure 4.6). As observed in the case of mouse ESCs (Figure 4.1), the 3' UTR of 'BIG' extends past the points predicted by AceView and NCBI. To validate this observation, the cDNA generated from HEK293T cells stably expressing the non-targeting shRNA (SCR) or shRNA targeting *LCOR* only, or *LCOR* and *'BIG'*, or *'BIG'* transcript at C10ORF12 region was analysed by RT-qPCR using the primers to the 3' UTR of 'BIG' transcript along the region downstream of the stop codon (Figure 4.6). This analysis confirmed that the cDNA expressed from the extended 3' UTR regions is down-regulated with knockdown of the 'BIG' transcript expression.

4.2.6 Mouse *Gm340* predicted gene is the orthologue of *C10ORF12*

So far, the mRNA expression analysis suggested that 'BIG' is encoded by an alternative splicing variant of the *LCOR* gene, whereby the third coding exon of *LCOR* is skipped and instead spliced to include the entire *C10ORF12* gene. What is more, the previously predicted *C10ORF12* gene appears to be expressed only as the exon in the 'BIG' transcript, rather than an autonomous transcript. To

enable the validation of these observations on a protein level, antibodies were required that would specifically recognise the protein regions shared with LCOR and the C10ORF12 encoded regions.

ClustalW2 alignment of the 'BIG' protein sequence with human LCOR confirmed that the first 111 amino acids are the same in 'BIG' and in LCOR proteins (Figure 4.7, A). Strikingly, human LCOR protein has 100% evolutionary conservation with its mouse orthologue of LCOR (data not shown), indicating the likely importance of this protein in these species. Therefore, an antibody against the N-terminus of the LCOR gene should in theory recognise two specific bands: a ~50 kDa LCOR protein, and an ~250 kDa 'BIG' in both mouse and human samples.

Similarly, ClustalW2 alignment of the 'BIG' protein sequence to C10ORF12 revealed the expected homology to its entire C-terminus. Strangely, in the mouse genome annotated on NCBI, no orthologue of C10ORF12 was listed. However, on closer inspection of the DNA downstream of the mouse LCOR gene, a predicted *Gm340* gene was identified, which, if translated, is homologous to the C10ORF12 region (Figure 4.7, B). Based on this data, in collaboration with the Millipore Corporation, an antibody to 'BIG' was raised against an epitope that is conserved between the mouse and human C10ORF12 regions (Figure 4.7, B).

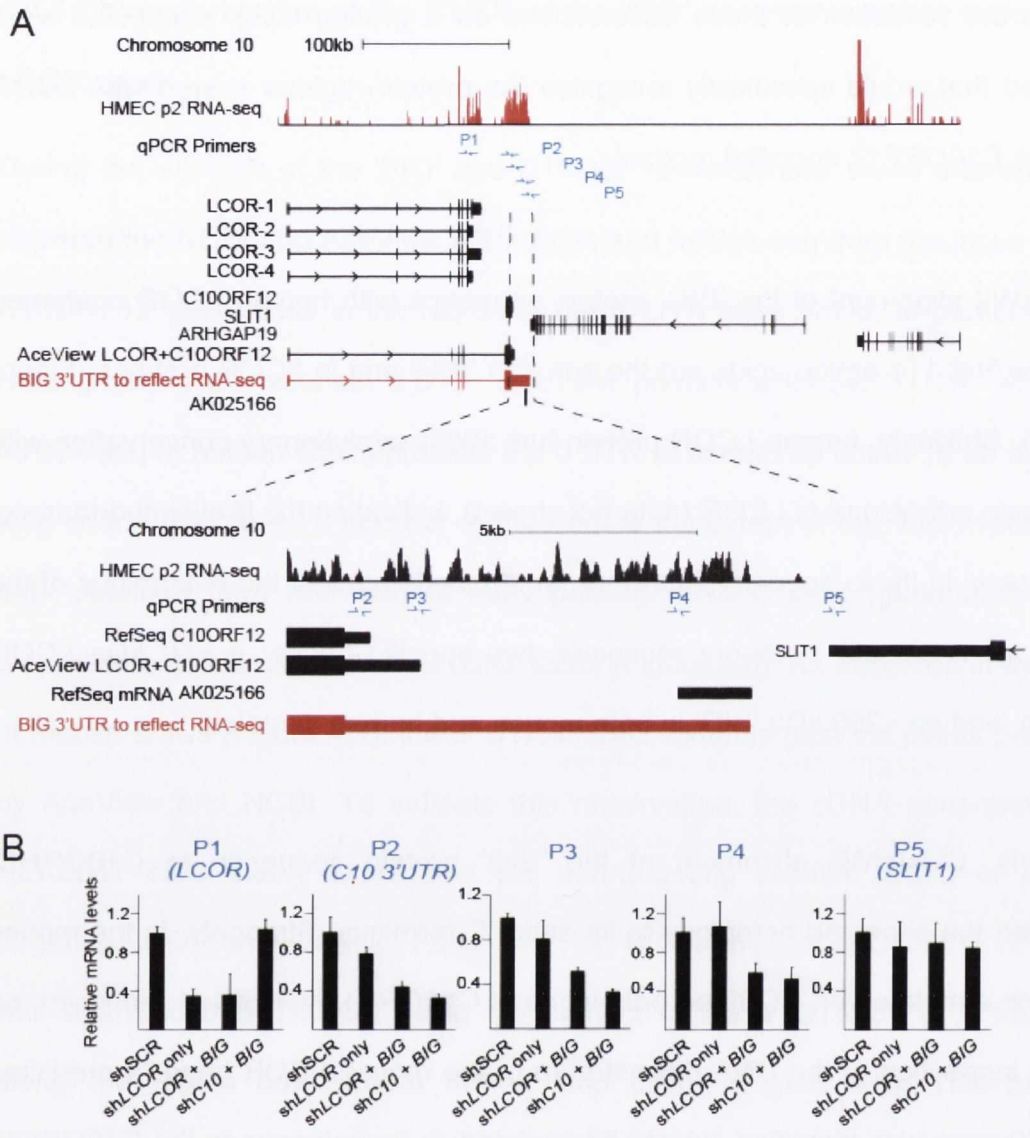


Figure 4.6 Characterisation of the 3'UTR of the 'BIG' alternative splice form of the LCOR gene

A. UCSC genome browser view of the LCOR gene locus, including an alignment to RNA-seq tracks from passage 2 (P2) HMEC cells (RNA-seq provided by Dr. Gerard Brien), with a close-up view of the 3'UTR of the 'BIG' mRNA transcript, including the predictions for the 3'UTR of 'BIG', depicting the locations of the quantitative RT-PCR primers used in panel B.

B. Quantitative RT-PCR analysis using the primers indicated in panel A along the 3' UTR of the 'BIG' mRNA on cDNA from HEK293T cells stably expressing shRNA targeting LCOR mRNA specifically, LCOR and 'BIG' mRNAs or the 'BIG' mRNA at the C10ORF12 region. The mRNA quantifications were normalised to the mRNA levels of *RPLPO* housekeeping gene.

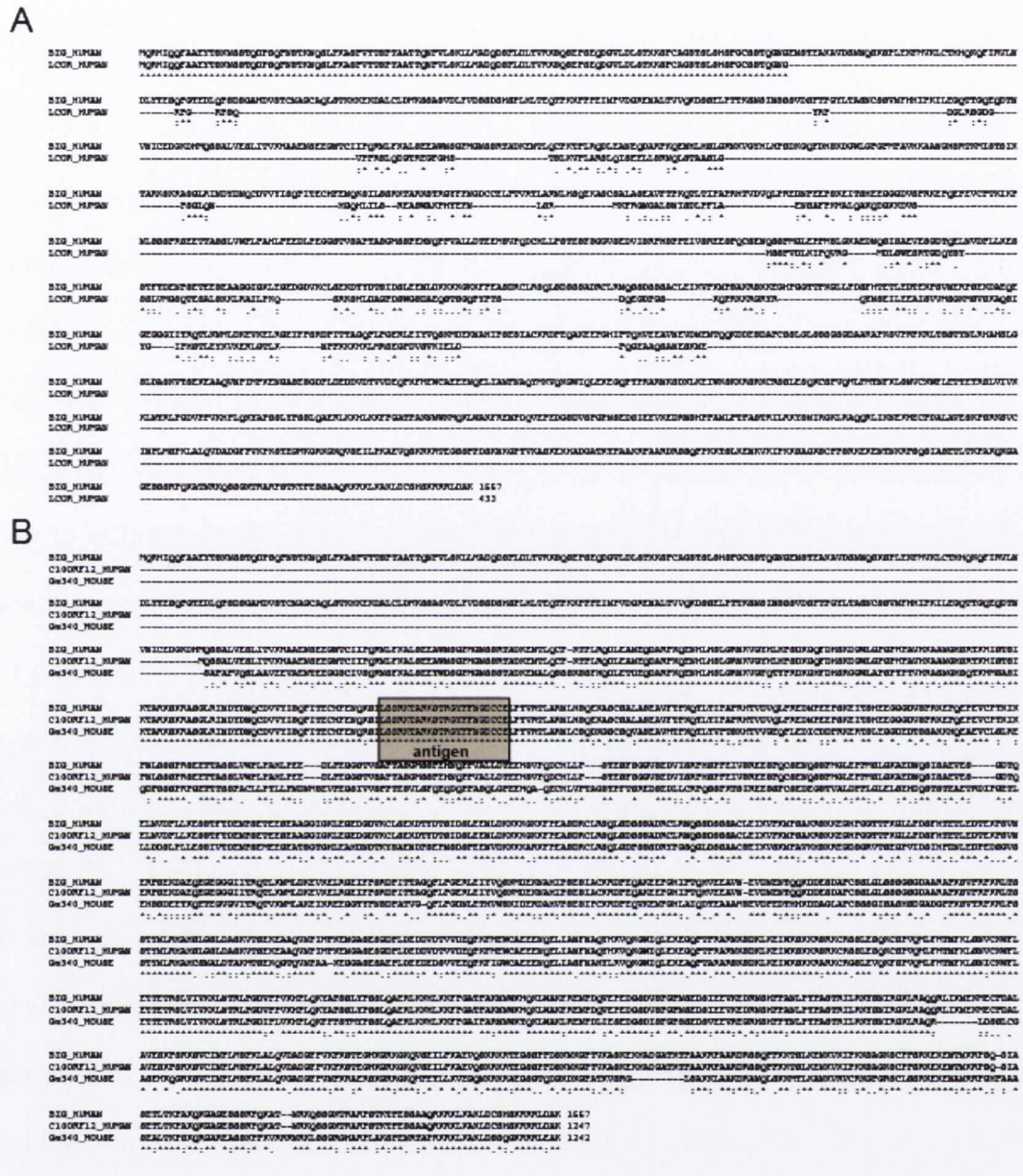


Figure 4.7 The mouse Gm340 ORF corresponds to the human C10ORF12 region of 'BIG'.

ClustalW alignment of human 'BIG' protein sequence (as cloned from HMEC cDNA) with the human LCOR protein sequence (A) and the BIG protein sequence with human C10ORF12 and the mouse orthologue of C10ORF12, annotated as Gm340 (B). The peptide sequence used to raise a rabbit polyclonal antibody against 'BIG', within the C10ORF12 region, is highlighted.

4.2.7 Generation and characterisation of a new antibody specific to the C10ORF12 region of 'BIG'

To test the specificity of the commercial antibody against the N-terminus of the LCOR gene, and the newly generated antibody to the C10ORF12 region of 'BIG' (Figure 4.8, A), Western blotting analysis on HEK293T lysates of the cells stably expressing the shRNA to LCOR only, LCOR and 'BIG' proteins, and two hairpins that target either the 'BIG' unique region or the C10ORF12 region of 'BIG' was performed (Figure 4.8, B). In support of observations made on the mRNA level, anti-LCOR antibody recognised two specific bands. Firstly, the ~50 kDa LCOR band, that was specifically knocked down by shRNA targeting LCOR, but not the ones that target 'BIG' uniquely. Secondly, the ~250 kDa band corresponding to 'BIG' protein, which was depleted in HEK293T cells infected with shRNAs targeting either of the exons included in the '*BIG*' transcript. On the other hand, the anti-C10ORF12 antibody recognised only the 'BIG' protein signal at ~250 kDa, however here it appeared as signal of multiple bands, which nonetheless was weakened by the knockdown of 'BIG' transcript via targeting the exon shared with LCOR and the 'BIG' specific region (Figure 4.8, B). Therefore, the signal recognised by the C10ORF12 antibody is indeed 'BIG' protein, and the diffuse appearance of the band may indicate that this protein is subject to post-translational modifications. Taken together, these data show that 'BIG' protein shares an amino acid sequence with LCOR on the N-terminus, and a C10ORF12 region on the C-terminus. It also suggests that C10ORF12 is not expressed as an autonomous protein.

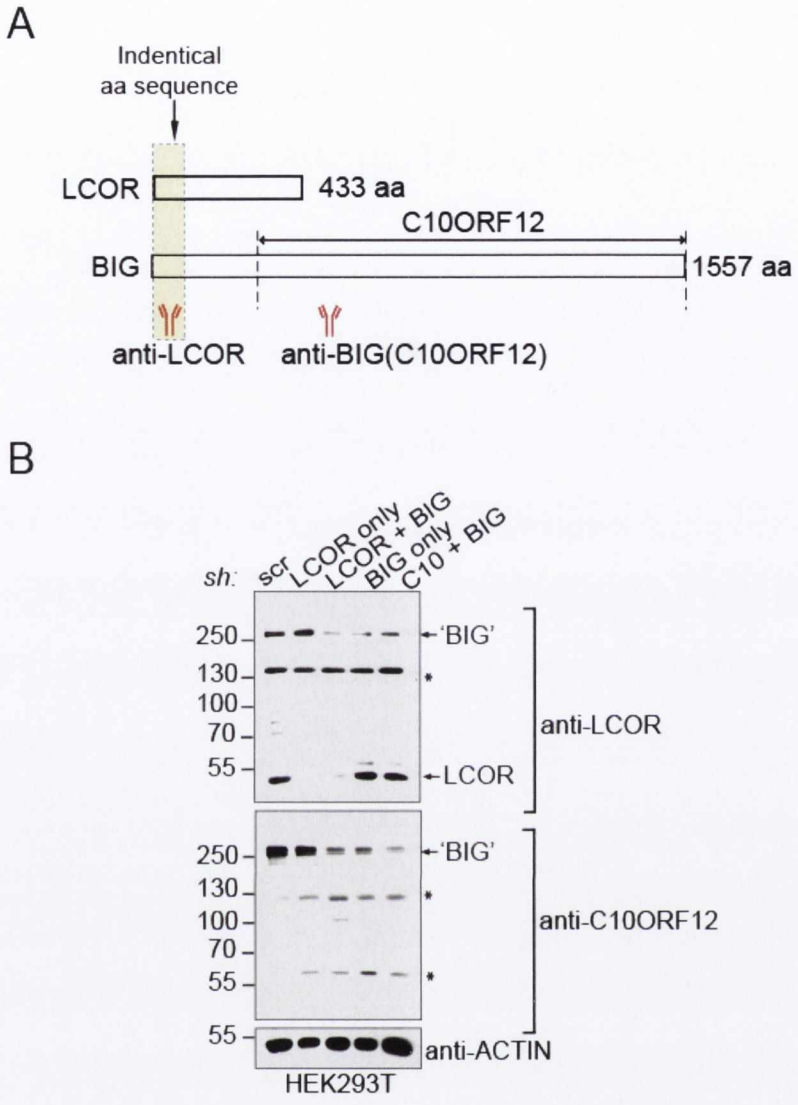


Figure 4.8 Validation of LCOR antibody and a new antibody raised to recognise the C10ORF12 region of 'BIG'.

A. A schematic of the LCOR and 'BIG' proteins with the regions of identical protein sequence enclosed in a shaded box and with the relative epitope regions of antibodies specific to LCOR and 'BIG' indicated.
B. Western blot analysis using the antibodies to LCOR and the C10ORF12 region of 'BIG', as indicated in A, on lysates of HEK293T cells, stably expressing shRNAs targeting the expression of the indicated transcripts. anti-Actin staining serves as loading control. (*) Indicate unspecific antibody binding.

4.2.8 Exogenous and endogenous PCL1 co-immunoprecipitate the 'BIG' protein

Next, to confirm that PCL1 interacts with 'BIG', but not the LCOR protein, as was suggested by the mass spectrometry analysis previously (Figure 4.2 and 4.4), immunoprecipitations of FLAG-HA-tagged PCL1 from nuclear lysates of HEK293T cells were Western blotted using the LCOR antibody. This analysis detected a 'BIG' band ~250 kDa in size, specifically in the FLAG-PCL1 IP, but not the control IP (Figure 4.9, A). As expected, LCOR was not detected in the FLAG-PCL1 IP, confirming that LCOR does not interact with PCL1. The C10ORF12 antibody also detected a band for 'BIG' at ~250 kDa in the FLAG-PCL1 IP. Finally, to exclude the possibility that interaction between PCL1 and 'BIG' is an artifact of PCL1 overexpression, immunoprecipitations using the antibody to endogenous PCL1 in HEK293T cells were performed. Western blotting analysis confirmed that 'BIG', but not LCOR, is immunoprecipitated with the endogenous PCL1 protein.

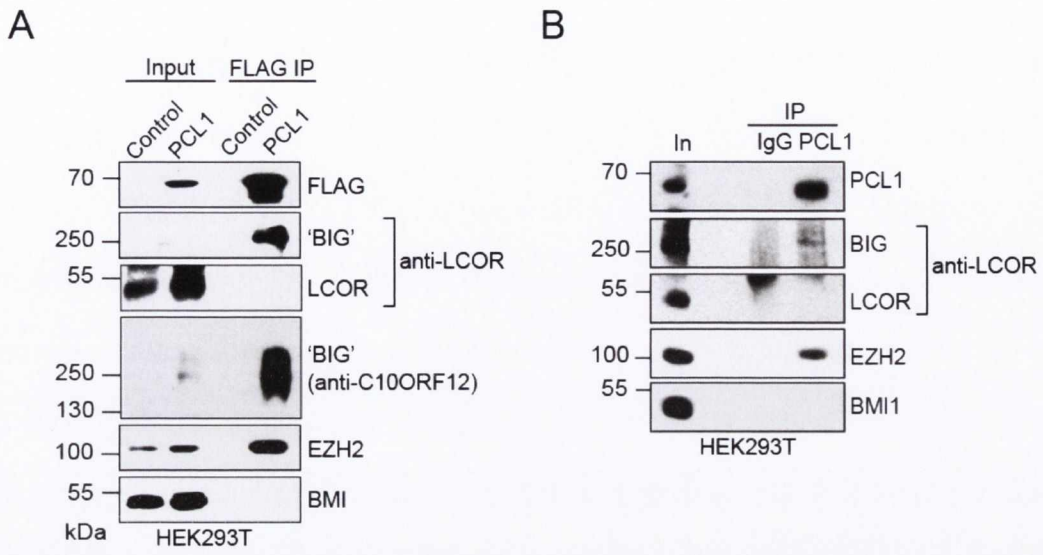


Figure 4.9 Exogenous and endogenous PCL1 co-immunoprecipitate the 'BIG' protein.

A. Western blot analysis of FLAG immunoprecipitations of nuclear lysates of HEK293T, either transfected with empty vector or with FLAG-HA-PCL1. Note that ~55kDa LCOR protein does not co-IP with FLAG-PCL1.
B. Western blot analysis of immunoprecipitations of endogenous PCL1 protein on nuclear from HEK293T cells. The PRC2 component EZH2 is included as a positive control and a PRC1 component BMI1 - as a negative control. Note that LCOR does not co-IP with PCL1.

4.3 Discussion

LCOR and C10ORF12 were previously identified in mass-spectrometry analysis of the core PRC2 component EED (Smits et al, 2013). In this chapter it is demonstrated that the LCOR protein does not bind to the PCL1-PRC2 complex. Instead, the LCOR and C10ORF12 specific peptides detected in PCL1-IP mass-spectrometry and, presumably, in the previous studies of EED mass spectrometry analyses, belong to a novel ~250 kDa protein, which here is called 'BIG'. It is demonstrated that this protein is a product of alternative splicing from the *LCOR* gene locus. The N-terminal region of this 'BIG' protein is comprised of the N-terminus of the LCOR protein, while its C-terminus comprises the entire C10ORF12 protein. The 'BIG' protein has not been characterised previously, but was predicted *in silico* in the AceView database, by an analysis of expressed sequence tags and splicing donor and acceptor sequences throughout the human genome, which aimed to identify novel splice variants of genes (Thierry- Mieg and Thierry-Mieg, 2006).

In this chapter it is also demonstrated that no autonomous C10ORF12-encoding transcript or protein is expressed in HEK293T cells on the mRNA and protein levels. While the AceView database also does not predict the existence of the autonomous C10ORF12 gene, it may not be excluded that an autonomous product may be expressed under certain conditions in certain cell types. Therefore, in any further RNA interference studies conducted to characterize the function of 'BIG', it would be a good precautionary measure to use both shRNA constructs targeting the 'BIG' unique region and the C10ORF12 region of 'BIG' in other cell lines, to confirm the observation made in this study and also to ascertain

the reproducibility of any functional observations.

It is noteworthy that both the LCOR and 'BIG' specific antibodies detected a protein band of ~130 kDa and the expression of this protein was not affected by the RNAi targeting of neither LCOR, nor BIG transcripts (Figure 4.8). It is not uncommon that polyclonal antibodies purified against a specific antigen may detect proteins with similar amino acid composition; LCOR protein has a homologue called LCOR-like (LCORL). In the current annotation this gene locus is expected to produce two isoforms: 35 or 77 kDa in size. However, it is a possibility that a 'BIG'-like protein could be produced from this locus by alternative splicing of the *LCORL* gene. No homologue of the *C10ORF12* predicted gene has been currently annotated in this region therefore a closer inspection of this locus would be required to determine if a homologue of 'BIG' exists.

In addition to the identification of the LCOR alternative splicing variant 'BIG', the RNAi analysis suggests that the 3'UTR of the 'BIG' transcript extends ~10 kb downstream of its predicted termination site (Figure 4.6). However, it may not be excluded that the RNA-seq signal observed past the currently predicted 3' UTR site is an uncharacterised long-non-coding RNA, the expression of which may correlate with 'BIG' expression (hence the observed down-regulation of the transcript from that site). The exact termination site of the 'BIG' transcript can be determined by sequencing of the PCR product obtained on the cDNA of either HEK293T or HMEC, using a forward primer mapping to the end of the 'BIG' coding sequence, and an oligo of 16 tandem repeats of the thymidine nucleotide as a reverse primer, which would recognise the polyadenylated tail of 'BIG' mRNA.

Alternatively, a set of RNA-seq databases could be bioinformatically analysed in search of reads that extend from the 'BIG' coding sequence end down to the polyadenylation tail. Similar strategies could be extended for analysis of the 5' UTR of the 'BIG' transcript, since the exact transcriptional start site for the 'BIG' mRNA is not yet characterised.

In this chapter it is established that the 'BIG' protein in its N-terminus contains two PXDLS CtBP-binding motifs and the nuclear receptor-binding box, suggesting that it might be a potent co-repressor due to CtBP action (Fernandes et al, 2003). Further study would be needed to characterise the possible interaction between 'BIG' and nuclear receptors and the CtBP proteins as well as its action as a co-repressor. Another indication on the action of 'BIG' as repressor is its association with Polycomb. Since it does not appear that LCOR is binding to PCL1 (Figure 4.9), while 'BIG' is, it would suggest that the domain specific for the interaction with the PCL1 proteins, and possibly PRC2, would lie within the 'BIG' specific region, comprising of the 'BIG' unique region and the C10ORF12 region. To test this hypothesis, co-immunoprecipitation and *in vitro* binding analyses of the C10ORF12 region of 'BIG' with PRC2 need to be performed. The characterisation of the nature of 'BIG's interaction with PRC2 function would be useful for understanding the molecular mechanism of the PRC2 action and thus its role in the developmental processes and cancer.

Chapter 5

Characterisation of the 'BIG' protein as a novel PRC2 complex subunit

5.1 Introduction

The novel protein 'BIG' shares functional domains with ligand dependent co-repressor LCOR on its N-terminus and possesses a C10ORF12 region on its C-terminus, which mediates the interaction with the PCL1 protein, and possibly PRC2. The N-terminus of 'BIG' contains the nuclear receptor-binding box LXXLL motif. This motif was shown to be required for the interaction of LCOR with estrogen receptor alpha (ERα) specifically in the presence of the activating agonist estradiol (E2) (Fernandes et al, 2003). Interestingly, EZH2 was reported to associate with ERα specifically in the presence of E2 (Shi et al, 2007), suggesting that 'BIG' might be involved in mediating this interaction.

Similarly, the N-terminus of 'BIG' also shares the PXDLS CtBP binding motifs with the LCOR protein (Fernandes et al, 2003). Previous studies have linked CtBP and Polycomb. In *Drosophila* development, dCtBP was reported to be required for Polycomb recruitment and repression of target genes (Basu et al, 2010, Srinivasan and Atchison, 2004). In addition, the dCtBP protein was reported to bind to several developmentally important transcriptional repressor proteins, such as *hairy*, *knirps, Kruppel, giant* and *snail* via its PXDLS motif and to act as a co-repressor of some of their target pair-ruled genes (Nibu et al, 1998, Poortinga et al, 1998, Strunk et al, 2001). The CtBP proteins have also been implicated as being essential for mammalian development. Mice null for both *CtBP1* and *CtBP2* die by E8.5 (Hildebrand and Soreano, 2002), and display a phenotype similar to that of Suz12 knockout (Pasini et al, 2004). The single *CtBP1* knock-out mice are viable and fertile, however smaller in size than the wild-type heterozygous mice, while *CtBP2* knockout mice exhibit axial patterning defects and die by E10.5 (Hildebrand and

Soreano, 2002). Recently, physical interaction between PRC2 and CtBP2 was described in mouse embryonic stem cells (Kim et al, 2015). Exploring the role of 'BIG' in the modulation of the interaction between CtBPs and PRC2 would lead to a better understanding of the mechanisms involved in PRC2 mediated repression.

The aim of this chapter is to validate that 'BIG' is a sub-stoichiometric component of the PRC2 complex and to explore its interaction with CtBPs and ERα. For this cell fractionation, co-immunoprecipitation and mass-spectrometry analyses were performed. Finally, to investigate the functional link between 'BIG' and PRC2, RNA interference of 'BIG' was employed to detect its potential role in mono-, di- and tri-methylation of histone H3 at lysine 27 in breast cancer cells.

5.2 Results

5.2.1 'BIG' protein is a chromatin associated protein that co-immunoprecipitates with EZH2

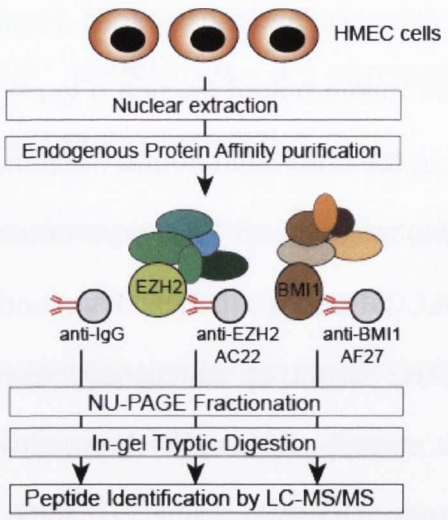
In order to determine if 'BIG' interaction with PCL1 is occurring in a PRC2 dependent context, we searched for peptides constituting 'BIG' in immunoprecipitations of the enzymatic component of the PRC2 complex, EZH2, or a PRC1 component BMI1 as a negative control, from primary human mammary epithelial (HMEC), which were analysed by mass spectrometry analysis, performed by Siobhan Turner as a part of a Masters project (Figure 5.2). In line with the observations made in previous chapters, the peptides for the C10ORF12 protein were among the highest scoring in the EZH2 IP, indicating that the 'BIG' protein is physically associated with PRC2 in HMECs (Figure 5.1 B). Notably, this interaction is specific to PRC2, because BMI1 did not immunoprecipitate either 'BIG' or LCOR.

Next, the interaction between 'BIG' and EZH2 was validated by performing the immunoprecipitations of endogenous EZH2 and BMI1 from nuclear extracts of HEK293T or breast cancer MCF7 cell lines, and Western blotting with the αLCOR antibody. In both cell models EZH2 immunoprecipitated 'BIG' specifically, while it did not pull down LCOR (Figure 5.2, C). The detection of this interaction in three different cell lines suggests its ubiquitous nature.

To investigate which cellular compartment the interaction of 'BIG' with EZH2 was localized to, cytosolic, nucleosolic and chromatin bound fractions of HEK293T

cells were analysed by Western blotting. The fractionation efficiency was validated by enrichment of beta-tubulin in the cytosolic and nucleosolic fractions, while histone H3 served as a marker of the chromatin bound fraction (Figure 5.1, D). As expected, EZH2 was strongly enriched in the chromatin bound fraction, while also present in the nucleosol. Similarly, the majority of 'BIG' was detected in the chromatin bound fraction, using both αLCOR and αC10ORF12 antibodies (Figure 5.1, D), suggesting that 'BIG' and EZH2 might be functionally associated on chromatin. Taken together, these data classify 'BIG' as a chromatin associated protein, which is a novel interacting partner of EZH2.

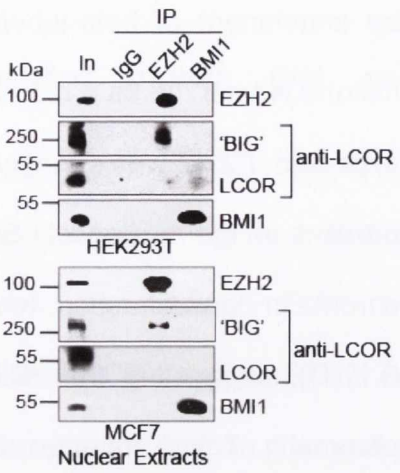
A



B

HMEC	Protein name	IgG	EZH2	BMI1
PRC2 Complex	SUZ12	0	82	0
	EZH2	0	42	0
	JARID2	0	38	0
	C10ORF12	0	29	0
	EED	0	18	0
	PHF1	0	10	0
	AEBP2	0	8	0
	PHF19	0	8	0
Canonical PRC1 Complex	PHC2	0	0	45
	PHC3	0	0	32
	CBX8	0	0	25
	BMI1	0	0	18
	CBX4	0	0	14
	PCGF2	0	0	11
	PHC1	0	0	10
	CBX2	0	0	10

C



D

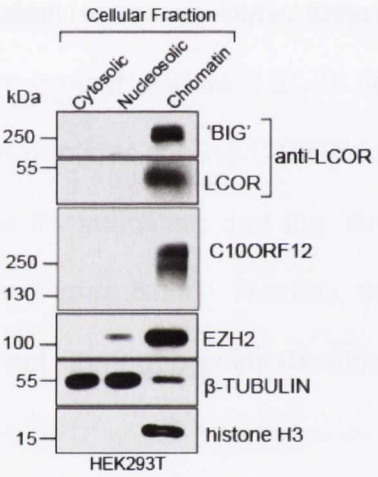


Figure 5.1 The 'BIG' protein is a chromatin associated protein that co-immunoprecipitates with EZH2.

A. Schematic of the experiment designed to identify proteins immunoprecipitated with endogenous antibodies to EZH2 (AC22) and BMI1 (AF27), performed on nuclear lysates from HMEC cells and subsequently processed for mass spectrometry analysis.

B. Table of the Mascot values summarizing the known PRC2 and PRC1 components, detected in the EZH2 and BMI1 IPs, respectively. Note that C10ORF12 is immunoprecipitated specifically with EZH2, but not BMI1 (IP-MS/MS in A and B was performed by Siobhan Turner).

C. Western blot analysis of endogenous immunoprecipitations of nuclear lysates of HEK293T (top panel) and breast cancer MCF7 cells (bottom panel) using the EZH2 and BMI1 antibodies.

D. Western blot analysis of the fractionation of HEK293T cells into cytosolic, nucleosolic and chromatin bound fractions and blotted with the indicated antibodies.

5.2.2 'BIG' is a novel sub-stoichiometric component of the PRC2 complex

Given that the 'BIG' protein, but not LCOR, immunoprecipitated with PCL1 and EZH2 endogenously, it was hypothesised that the N-terminus of 'BIG' would not be responsible for the interaction with PRC2. Therefore this function must be attributed to the C10ORF12 region of 'BIG'. To test this hypothesis, *LCOR* and '*BIG*' genes, as well as the coding sequence for the *C10ORF12* region of 'BIG', were cloned into a FLAG-HA-tagged expression vector, transiently expressed in HEK293T cells, FLAG-immunoprecipitated and analysed by mass spectrometry (Figure 5.2, A). Strikingly, the core PRC2 complex components EZH2, EED and SUZ12, together with PCL2 and PCL1 were all immunoprecipitated with both 'BIG' and C10ORF12, but not with LCOR. Moreover, this approach identified many additional proteins that associate with 'BIG' via its C10ORF12 region, including the G9a co-repressor complex - EHMT1/2, ZNF644 and WIZ (Ueda et al, 2006, Mulligan et al, 2008), the SET repressor and the USP11 deubiquitinase. Interestingly, we also detected the USP22 deubiquitinase in the FLAG-'BIG', but not FLAG-C10ORF12 IP mass spectrometry (Figure 5.2, C). To validate these interactions, Western blot analysis of independent IPs of FLAG-LCOR, FLAG-C10ORF12 and FLAG-'BIG' were performed, which confirmed that 'BIG', like LCOR, but not C10ORF12, associates with CtBP proteins (Figure 5.3, A). The 'BIG' protein shares two CtBP binding motifs with the LCOR protein, which would explain this observation (Fernandes et al. 2003). It was also validated that both 'BIG' and C10ORF12, but not LCOR, associate with the PRC2 complex members EZH2, as well as the SET repressor protein and USP11 deubiquitinase. Finally, it was validated that 'BIG', but not C10ORF12 or LCOR immunoprecipitated the USP22 deubiquitinase, consistent with the idea that the amino acid sequence

unique to 'BIG' is required for this interaction (Figure 5.3, A). These data suggest a model where 'BIG', similarly to LCOR, binds to CtBPs by the N-terminus; to USP22 by its unique region; and the C-terminus of 'BIG', which is here referred to as C10ORF12 region, is accommodating the binding of the PRC2 and G9a complex, as well as USP11, SET and possibly other chromatin regulator proteins (Figure 5.3, B).

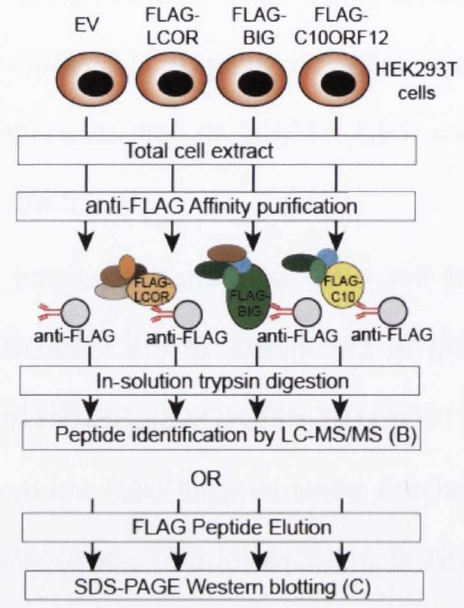
5.2.3 'BIG' interacts with PRC2 and G9a via two distinct regions in its C-terminus

The interaction between the PRC2 and G9a complexes has been previously reported (Mozzetta et al, 2014, Maier et al, 2015). Since the C10ORF12 region of 'BIG' is binding a number of chromatin regulators, including the PRC2 and G9a complexes, in order to delineate the interaction domains for these binding partners, five truncation constructs of the C10ORF12 region were cloned into a FLAG-HA-tagged expression vector for a FLAG-immunoprecipitation analysis in HEK293T cells (Figure 5.4, A). The immunoprecipitations of FLAG-LCOR and FLAG-C10ORF12 served as negative and positive controls, respectively. As expected, EZH2, PCL1, G9a and USP11 were immunoprecipitated by the C10ORF12 region, while BMI1 was not (Figure 5.4 B). Strikingly, the two non-overlapping fragments, namely Fragment 1 and Fragment 4, immunoprecipitated G9a and the PRC2 complexes, respectively (Figure 5.4, B). This suggests that these two complexes are bound to 'BIG' via two distinct regions. Surprisingly, USP11 was also immunoprecipitated by Fragment 1. While amino acids 915-1339 of 'BIG' are sufficient for interaction with PRC2, it remains unclear whether the last 388 amino acids on the C-terminus of 'BIG' are capable of binding PRC2, since

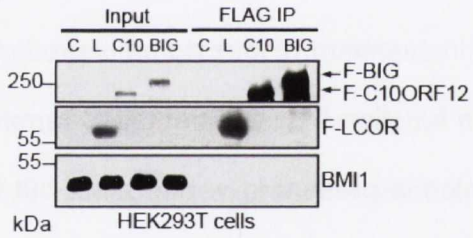
Fragment 5 could not be expressed in this setting. This could possibly be explained by the fact that Fragment 5 is composed of a stretch of protein sequence which is predicted to be unstructured, meaning that this peptide was unlikely to form a stable globular structure, perhaps making it susceptible to degradation (Figure 5.4, C).

Strikingly, the disorder tendency analysis of the 'BIG' sequence predicted two 'ordered', globular-like domains: one mapping to the middle of the G9a/USP11 interaction domain (ASGLRINDYDNQCDVVYISQPITECHFENQKSILSSRKTAR KSTRGYFFNGDCCELPTVRTLARNLHS), and the other to the PRC2 interaction domain (KCRSSLESQKCSPVQMLFMTNFKLSNVCKWFLETTETRSLVIVKK) (Figure 5.4, C), suggesting that these stretches of amino acid sequence could be key for the interaction with the complexes in question. Pfam domain prediction database also detected a domain of unknown function DUF4553 at the C-terminus of 'BIG' (amino acids region 1089-1556), which is present in vertebrates, but has not been functionally characterized yet (Finn et al, 2014). Moreover, the Fragment 4 region showed the highest degree of amino acid conservation between human and zebrafish proteins, suggesting that PRC2 interacting region is evolutionarily conserved (Figure 5.4, C).

A



B



C

FLAG-IP Mass Spectrometry in HEK293T cells				
Gene name	LCOR	BIG	C10ORF12	Control
C10ORF12	0	298	244	3
SUZ12	0	15	17	0
EZH2	0	10	17	0
EED	0	8	11	0
MTF2	0	6	12	0
PHF1	0	1	2	0
RBBP4	0	2	3	0
EHMT1	0	3	5	0
EHMT2	0	4	8	0
ZNF644	0	4	10	0
WIZ	0	2	5	0
PRMT1	0	3	1	0
SET	0	10	6	0
ANKRD52	0	3	3	0
TCEAL1	0	1	2	0
HTATSF1	0	3	2	0
USP11	0	19	19	0
DNAJC7	0	4	5	0
OBSL1	0	7	5	0
MAPK1	0	6	5	0
NAGK	0	2	3	0
CTBP1	39	25	2	0
CTBP2	17	8	2	0
ERLIN2	10	2	0	0
NDUFA9	7	2	0	0
ABCB7	6	1	0	0
INTS2	2	3	0	0
USP22	0	20	0	0
VAPB	0	13	0	0
CTU1	0	9	0	0
MTCH2	0	5	0	0
MID1	0	4	0	0
LCOR	357	78	6	4
PRAME	2	0	0	0
ZNF24	3	0	0	0
HOXD13	5	0	0	0

Figure 5.2 Mass spectrometry analysis of FLAG-LCOR, FLAG-C10ORF12 and FLAG-'BIG' in HEK293T cells.

A. Schematic representation of the experimental outline. HEK293T cells were transfected with pLENTI FLAG-HA empty vector or this vector containing LCOR, 'BIG' or C10ORF12 coding sequences and total cell lysates were subjected to FLAG-IP.

B. Western blot analysis of immunoprecipitations prior to performing mass spectrometry. BMI1 is included as a negative control.

C. Table summarizing the proteins identified by mass spectrometry of material immunoprecipitated with either of FLAG-tagged LCOR, BIG or C10ORF12, but not with all three proteins. The proteins highlighted in red were subsequently validated in independent co-immunoprecipitations (Figure 5.3).

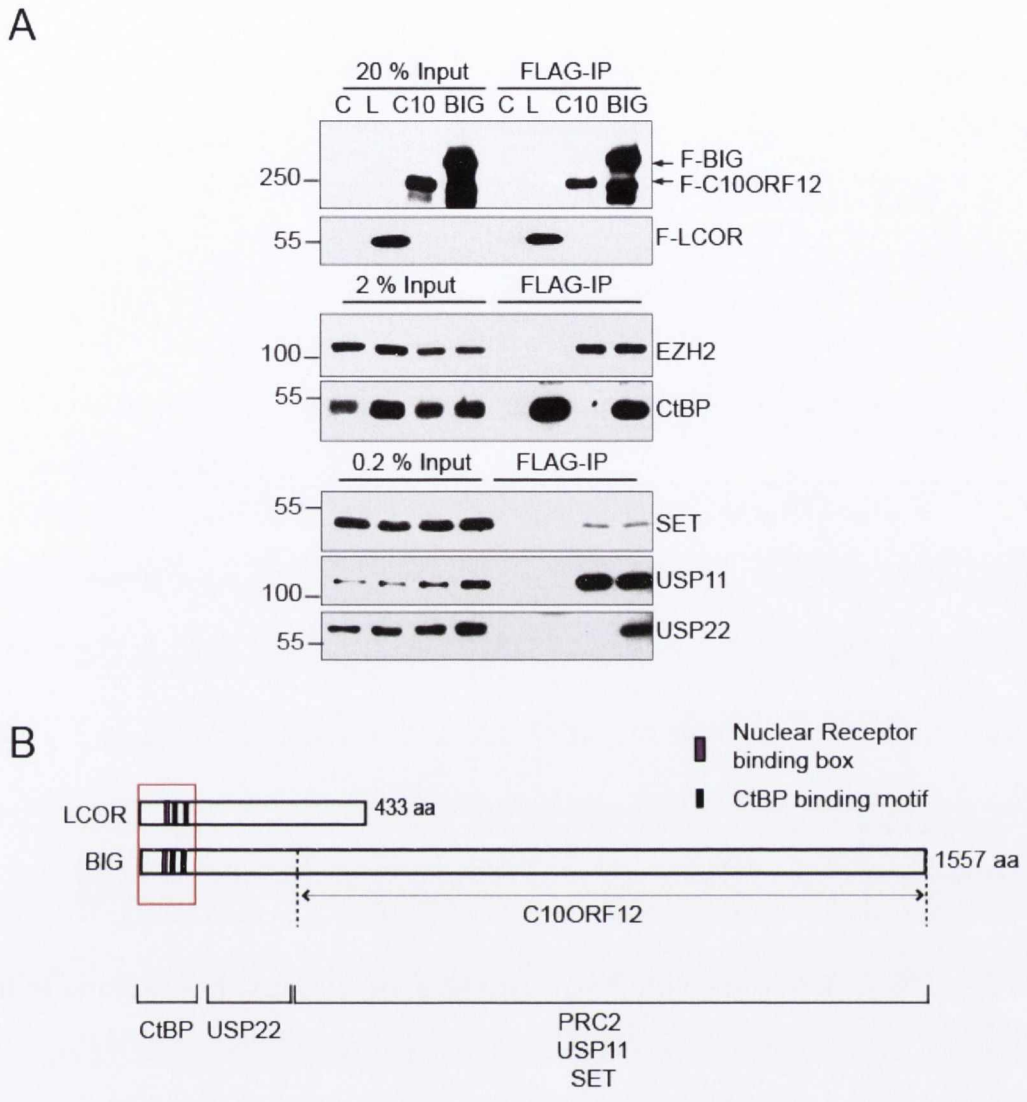


Figure 5.3 'BIG' binds via its CtBP2 on the N-terminus and multiple other chromatin associated proteins via its unique C-terminus.

A. Western blot analysis of FLAG immunoprecipitations performed on total lysates of HEK293T cells transfected with pLENTI empty vector, LCOR, C10ORF12 or 'BIG'.

B. Schematic representation of the 'BIG' protein, along with the LCOR protein, depicting its shared and unique protein interacting regions.

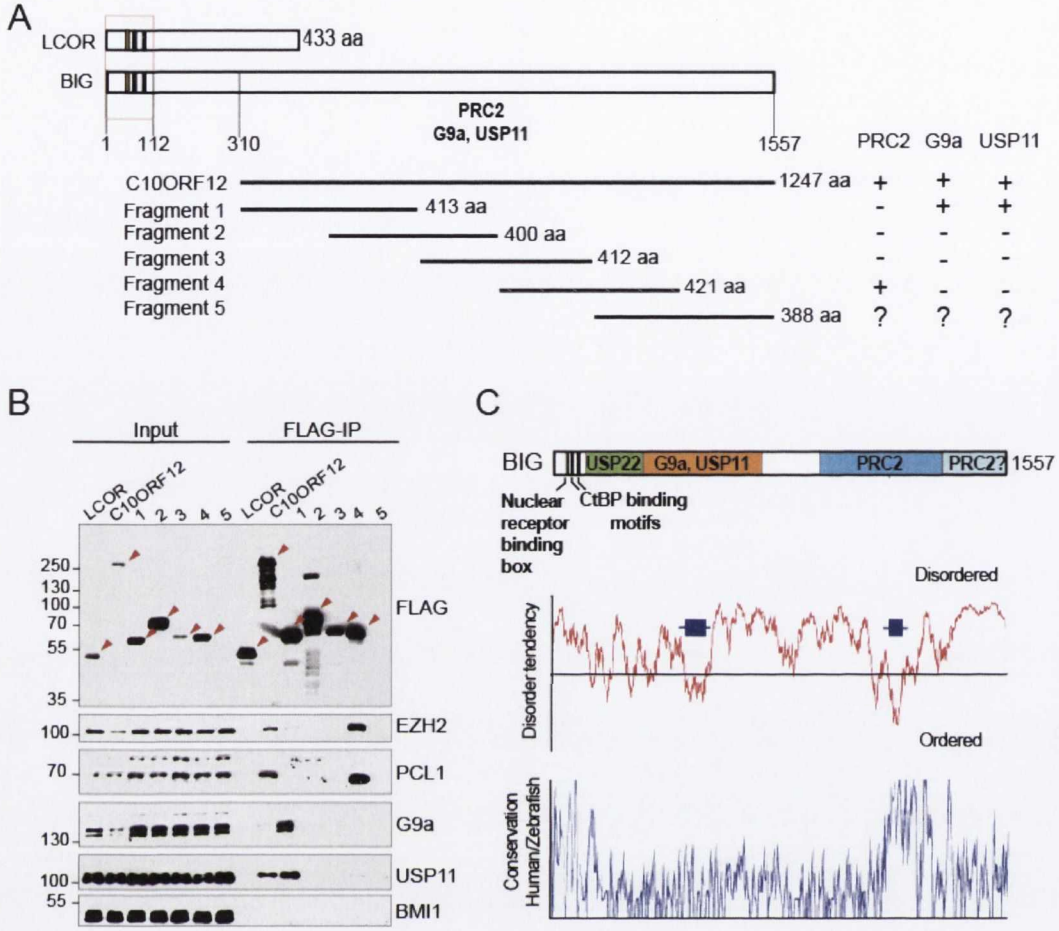


Figure 5.4 'BIG' interacts with PRC2 and G9a via two distinct regions in its C-terminus.

A. A schematic representation of the different truncations of the C10ORF12 region of BIG to identify the regions responsible for the interaction with PRC2, USP11 and G9a proteins.

B. Western blot analysis using the indicated antibodies of the FLAG immunoprecipitations of C10ORF12 and fragments 1 to 5, as depicted in panel A. FLAG-HA-tagged C10ORF12, the truncation fragments 1-4 and LCOR, included as negative control, were transiently expressed in HEK293T cells and immunoprecipitated from total cell extracts using anti-FLAG antibody. The bands corresponding to the expressed proteins are indicated by red arrows. The results are representative of two independent experiments.

C. An allignment of protein disorder plot of the 'BIG' protein, generated using the IUPreD program (Dosztanyi et al, 2005), to the regions of 'BIG' we defined as being sufficient for the interaction with G9a, USP11 and PRC2. The purple boxes in the disorder plot indicate two predicted unknown globular domains. The homology plot was obtained using Vector NTI software by comparing 'BIG' protein sequence from human and zebrafish (Danio rerio).

5.2.4 'BIG' associates with EZH2 together with SUZ12

Identification of two distinct domains for the PRC2 and G9a complexes raised a question: does 'BIG' bind to PRC2 and G9a complexes simultaneously, or whether these complexes interact with different 'BIG' molecules? To answer this question, a tandem immunoprecipitation was performed (Figure 5.5, A). Firstly, FLAG-immunoprecipitations of FLAG-'BIG', or FLAG-LCOR as a negative control, which were FLAG-immunoprecipitations in HEK293T cells were performed. The FLAG-containing complexes were eluted using 3xFLAG peptide by competitive elution, and then immunoprecipitated again, using the antibody to endogenous EZH2. Exogenous FLAG-'BIG' and PRC2 component SUZ12 were both immunoprecipitated by the EZH2 re-IP, confirming that 'BIG' binds to the two core PRC2 components simultaneously (Figure 5.5, B, lane 6). On the other hand, no G9a was detected in the EZH2 re-IP (Figure 5.5, B lane 6), while it was immunoprecipitated with FLAG-'BIG' initially (Figure 5.5, B lane 4). Similarly, no USP11 or USP22 was detected in the EZH2 re-IP, suggesting that EZH2 binds 'BIG' as a PRC2 complex, while G9a, USP11 and USP22 are likely to be binding to 'BIG' independently of PRC2 (Figure 5.5, C).

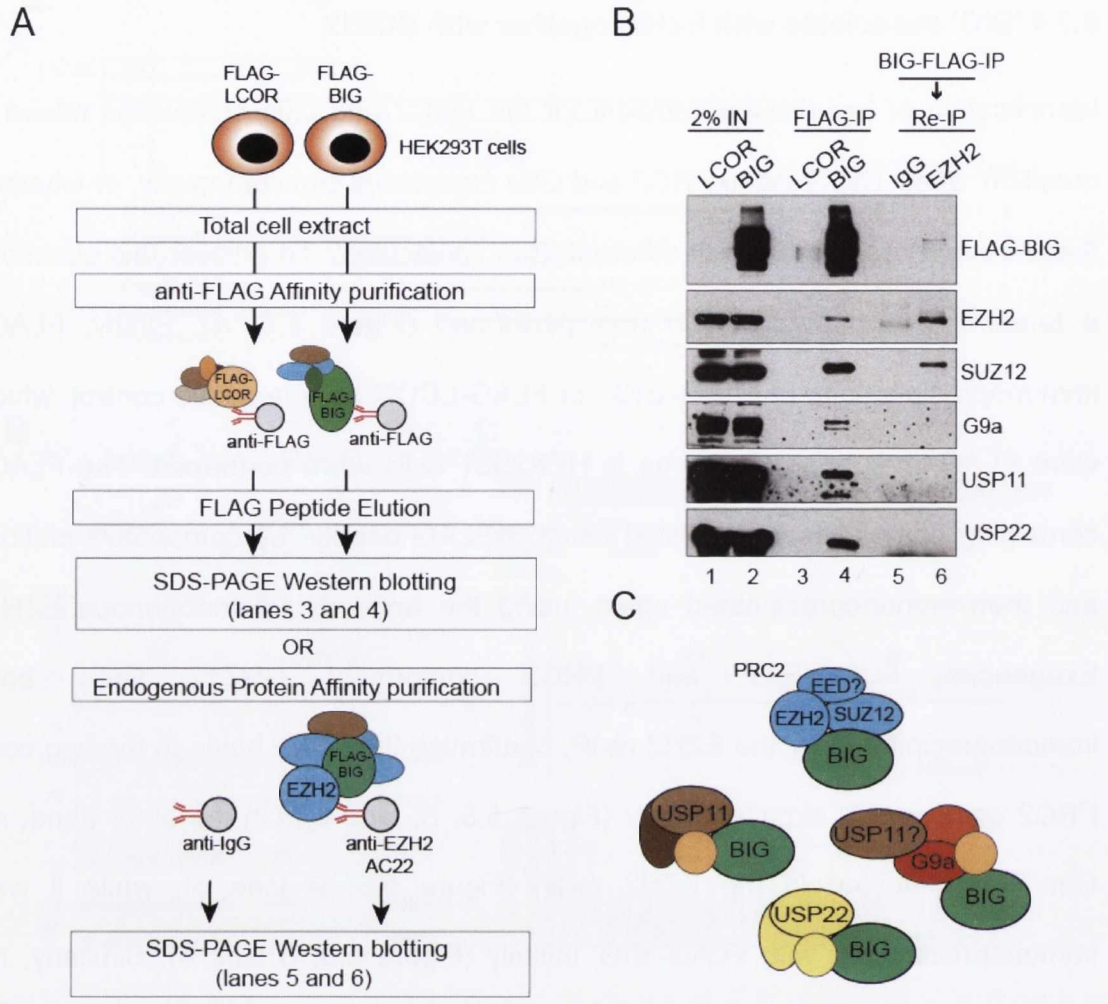


Figure 5.5 'BIG' associates with EZH2 together with SUZ12.

A. Schematic representation of the experimental approach to test if 'BIG' binds PRC2, G9a and USP11 simultaneously. Tandem immunoprecipitation of EZH2-containing PRC2 complexes from the FLAG-purified pool of BIG-containing complexes was performed.

B. Western blot analysis of the experiment in panel A using the indicated antibodies. This experiment was performed once.

C. A preliminary model for 'BIG' containing protein complexes.

5.2.5 Establishment of an inducible 'BIG' reporter system in HEK293T cells

To further explore the functional interplay between 'BIG' and PRC2 and G9a complexes and to evaluate the functionality of the 'BIG' antibody in ChIP, inducible 'BIG' reporter cell line was generated. This system permits the tetracycline or doxycycline inducible targeting of a GAL4 fusion of the 'BIG' protein to the Gal4-binding UAS elements upstream of a thymidine kinase (TK) promoter and luciferase reporter gene, which is integrated into the genome (Figure 5.6, A). Two stable cell lines containing either a doxycycline inducible Gal4-'BIG' fusion gene or Gal4 empty vector incorporated into its genome were generated. Upon induction with doxycycline, the expression of the Gal4-tagged 'BIG' protein was induced (Figure 5.6, B). Concomitantly, the luciferase activity was reduced 2-fold upon Gal4-'BIG' induction, but not in the case of the control Gal4-empty vector control (Figure 5.6, C), suggesting that the 'BIG' protein is associated with gene repression.

To investigate if 'BIG'-mediated repression leads to the recruitment of the PRC2 complex, G9a or CtBP2 to the luciferase promoter, chromatin immunoprecipitations (ChIPs) were performed using antibodies specific to GAL4, EZH2, G9a, CtBP2 and four batches of the rabbit polyclonal 'BIG' (C10ORF12)-specific antibodies and an IgG antibody was used as a negative control, in GAL4-'BIG' cells in presence or absence of doxycycline treatment (Figure 5.6, D). GAL4-'BIG' was enriched on the luciferase promoter upon GAL4-'BIG' induction, but unfortunately neither of the 'BIG' specific antibodies showed enrichment at this site, indicating that they are not suitable for ChIP application.

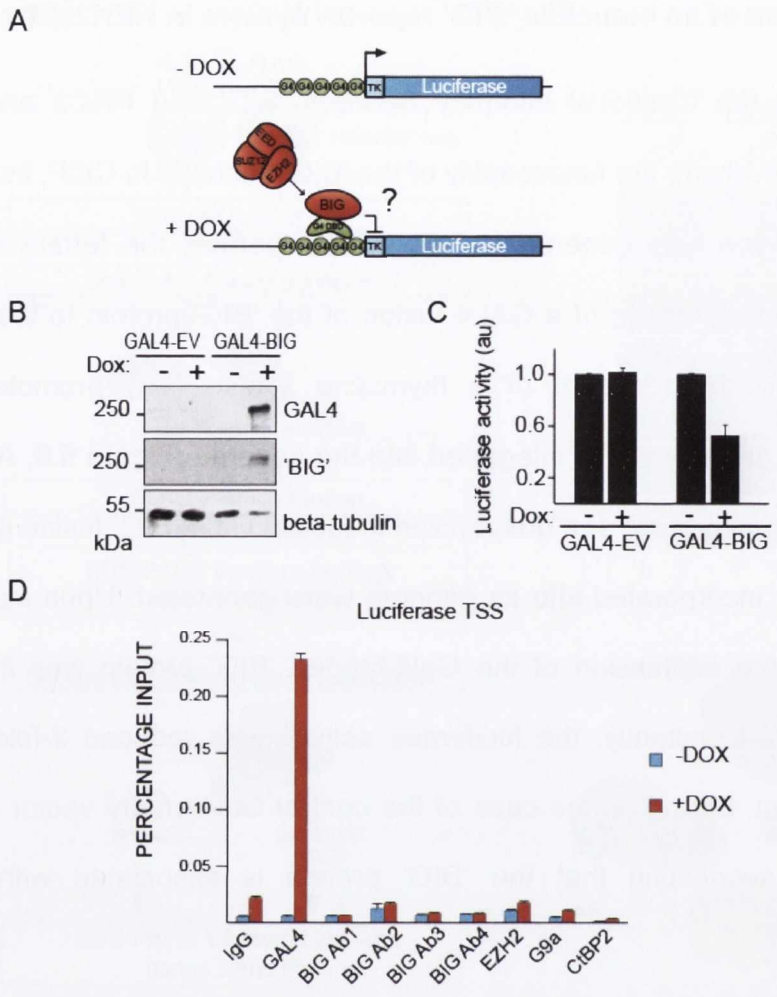


Figure 5.6 Establishment of an inducible 'BIG' reporter system in HEK293T cells.

A. Schematic representation of the GAL4-TK-Luciferase reporter system. Luciferase gene is integrated into the genome and controlled by the Gal4-binding UAS elements and a thymidine kinase promoter. The gene encoding the Gal4-'BIG' fusion protein is also integrated into the genome and induced by the treatment with tetracycline or doxycycline.

B. Western blot analysis showing the induction of GAL4-'BIG' fusion protein upon treatment with doxycycline for 48 h in the GAL4-'BIG' TK-Luciferase cells, but not in the empty vector (EV) control cells.

C. Luciferase reporter activity assay of cells from panel B showing that induction of GAL4-'BIG' leads to repression of the GAL4-TK-Luciferase promoter.

D. Quantitative RT-PCR analysis of chromatin immunoprecipitations using the indicated antibodies on the Luciferase promoter of GAL4-'BIG' cells in presence or absence of doxycycline.

Notably, the binding of GAL4-'BIG' was not accompanied by the recruitment of EZH2, G9a or CtBP2, suggesting that 'BIG' represses the luciferase reporter independently of these proteins, at least in this artificial system.

5.2.6 'BIG' is associated with ERα and EZH2 in breast cancer cells independently of estradiol signaling

To test the hypothesis that 'BIG' might share this function of LCOR and bind to activated ERα, and possibly link the EZH2 and ERα function, co-immunoprecipitations of ERα and EZH2 from an asynchronously growing human breast carcinoma cell line, MCF7, were performed. Reassuringly, 'BIG' was immunoprecipitated using the endogenous antibody to ERa (Figure 5.7, A left panel). Similarly, the EZH2 antibody immunoprecipitated 'BIG' and ERα, but not BMI1 (Figure 5.7, A right panel). However, the 'BIG' band in the ERα IP appeared diffused and the EZH2 band ran higher than the input, which suggested that it might be a background band. Therefore additional negative control immunoprecipitations were performed, in which the protein-specific antibody, such as αEZH2 or αERα, was cross-linked to the beads as usual, however instead of the MCF7 cell lysate, empty lysis buffer was used for the immunoprecipitation. In this way, any background caused by the eluted antibody chains can be detected on the Western blot.

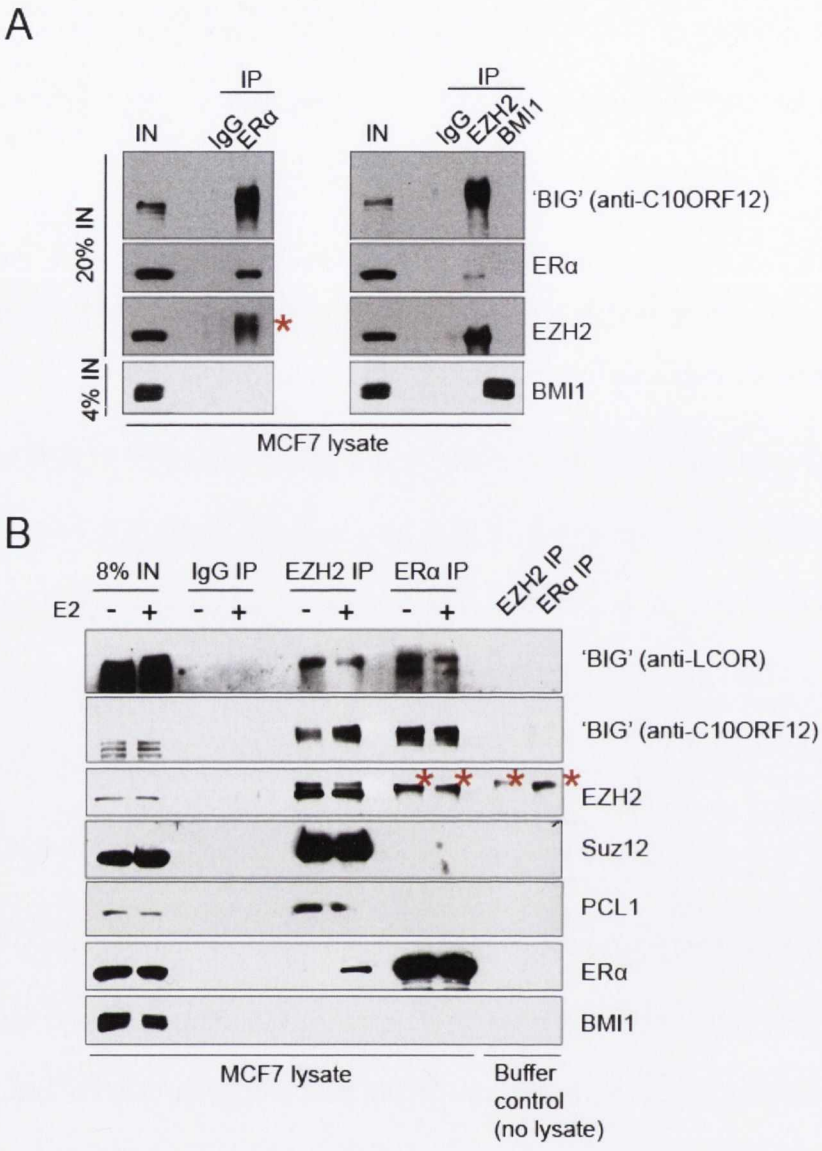


Figure 5.7 'BIG' is associated with ERα and EZH2 in breast cancer cells independent of estradiol (E2).

A. Western blot analysis of endogenous immunoprecipitations of EZH2, BMI1 and ERα from the total extract of MCF7 breast cancer cells.
B. Western blot analysis using the indicated antibodies on immunoprecipitations of EZH2 and ERα performed on the total cellular lysates of MCF7 cells, which were grown in phenol-red free media estradiol free conditions for 3 days and then treated with 100nM estradiol (E2) or equivalent volume of vehicle ethanol for 45 minutes. EZH2 and ERα immunoprecipitations were also performed on just the lysis buffer, without any cell lysate, and this served as negative control for antibody chain specific bands. The stars indicate unspecific antibody background bands, which ran slightly higher than EZH2.

To assess whether the interaction between 'BIG' and ERα was dependent on the activation status of ERα, the MCF7 cells were grown in charcoal stripped medium, in conditions of estrogen depletion, for 3 days. The cells were subsequently treated with estrogen or vehicle for 45 min, after which they were harvested and used for immunoprecipitations of endogenous ERα and EZH2 (Figure 5.7, B). Surprisingly, 'BIG' was detected both in the presence and absence of estradiol in ERα and EZH2 immunoprecipitations, suggesting that 'BIG' binds them both irrespective of ERα activation status. The 'BIG' signal in immunoprecipitations was specific, because no bands of this size were detected in the empty IP controls. However, ERα was immunoprecipitated with EZH2 specifically in the presence, but not in the absence, of estradiol, as was previously reported (Shi et al, 2007), indicating that the estradiol depletion/induction system used here was specific. However, it remained inconclusive, whether EZH2 was pulled down in the immunoprecipitation of ERα, due to the antibody chain background. Notably, 'BIG' appeared as a single band in EZH2 IPs, but as a doublet in ERα IPs, as detected using both αLCOR and αC10ORF12 antibodies (Figure 5.7, B). Therefore, it can be speculated that 'BIG' might be subject to differential post-translational modifications depending on whether it is in a complex with EZH2 or ERα, if, indeed, it is not binding both of them simultaneously.

5.2.7 'BIG' is required for maintenance of H3K27me1 and H3K27me2 levels and proliferation of breast cancer cells

Since the results of the GAL4-reporter experiment indicated that ectopic expression of 'BIG' leads to the repression of the luciferase reporter gene, but does not lead to the recruitment of the PRC2, G9a or CtBP co-repressors, it was

hypothesised that 'BIG' might be affecting the function of PRC2, even if PRC2 binding on the locus cannot be detected, as suggested by Ferrari and colleagues (Ferrari et al, 2014). To investigate the effects that 'BIG' may have on PRC2 function, levels of the PRC2 enzymatic products: mono-, di- and tri-methylation of the lysine residue 27 on histone H3 (H3K27me1, H3K27me2 and H3K27me3, respectively, were assessed in 'BIG' depleted cells. To this end, MCF7 cell lines stably expressing scrambled non-targeting shRNA (shSCR), or two hairpins that target 'BIG' either in 'BIG' unique region (shBIG) or in the C10ORF12 region (shC10+BIG) were generated. The efficiency of the shBIG hairpin was more potent than that of shC10+BIG, however in both cases the 'BIG' protein levels were reduced to at least 20% of the shSCR control levels (Figure 5.8, A). Strikingly, the total levels of H3K27me1 and H3K27me2 were dramatically reduced in 'BIG' depleted cells, while H3K27me3 levels did not appear to show significant changes (Figure 5.8, B). The effect produced on H3K27 mono- and di- methylation could be partially explained by the slight reduction of the PRC2 levels in the 'BIG' depleted cells, however this would also be expected to be reflected on the H3K27me3 levels, which does not seem to be the case. Therefore, these data suggest that 'BIG' is required for the maintenance of the H3K27 mono- and di-methylation, while it does not function on H3K27me3 modulation. Furthermore, the depletion of 'BIG' expression caused a reduction in proliferation rates of MCF7 cells, compared to shSCR control (Figure 5.8, B), implying that 'BIG' is required for cellular proliferation.

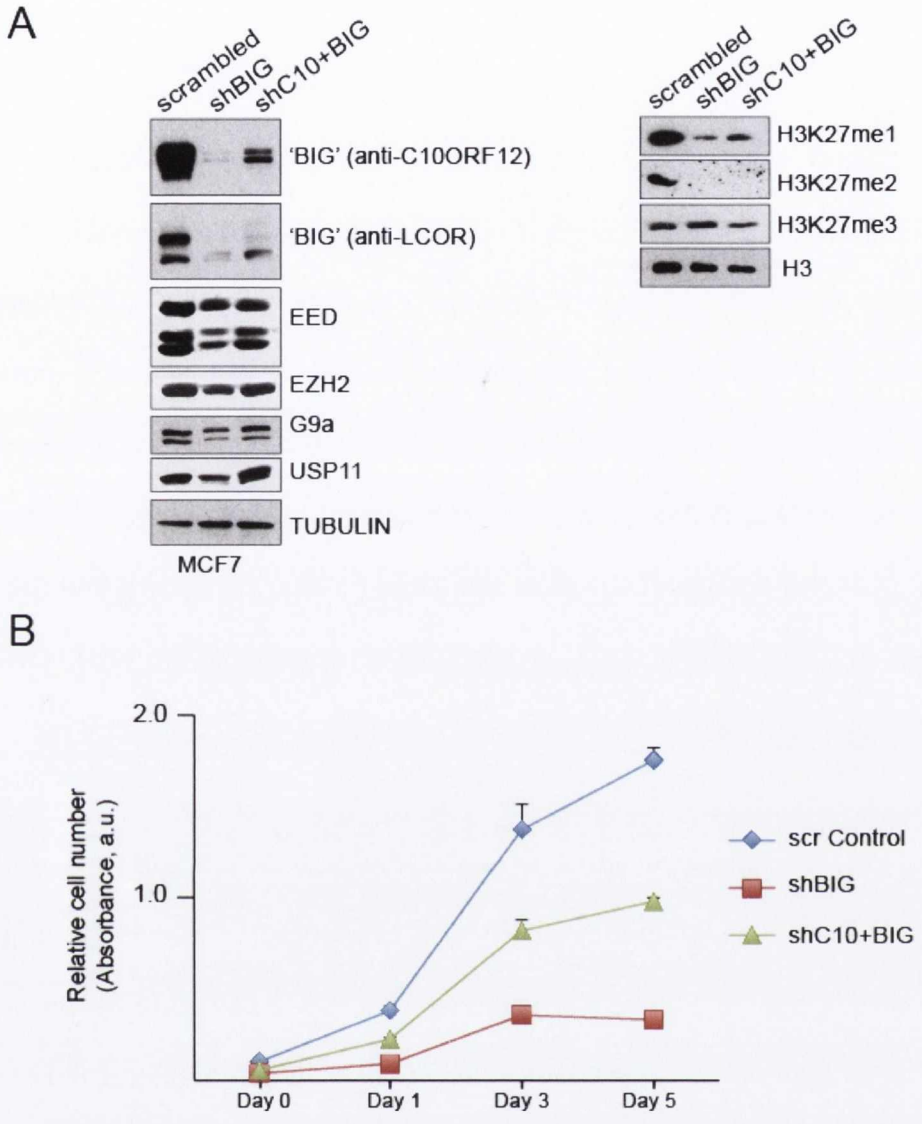


Figure 5.8 'BIG' is required for maintenance of H3K27me1 and H3K27me2 levels and proliferation of breast cancer cells.

A. Western blot analysis of total cell lysates from MCF7 cells, which were stably expressing two independent shRNA to 'BIG', either targeting a BIG unique region or the C10ORF12 encoded region, or non-targeting SCR shRNA. Blotting with beta-tubulin served as a loading control.
B. Growth curve of MCF7 cells from panel A. MCF7 cells were seeded at 200,000 cells per well in 6-well plates in duplicate and stained with crystal violet at the indicated time-points. The relative cell numbers were determined by spectrophotometric assay after the elution of crystal violet stain in 20% acetic acid. The results shown are representative of two independent experiments (n=2).

5.3 Discussion

In this chapter the properties of 'BIG' protein were characterised. It is demonstrated that the C-terminus of 'BIG' contains a previously predicted C10ORF12 ORF, which mediates 'BIG' binding to several members of the PRC2 and G9a repressor complexes, via two different domains. In addition, these domains mediate interaction with de-ubiquitinase USP11 and SET co-repressor. Furthermore, evidence is provided that 'BIG' binds to CtBP co-repressors and USP22 on its N-terminus, and is associated with nuclear receptor ERα in breast cancer cells, regardless of estradiol signaling. Finally, it is shown that the depletion of 'BIG' in breast cancer cells causes down-regulation of the total H3K27 mono- and di-methylation levels and impairs cellular proliferation.

Consistent with mass spectrometry analysis of C10ORF12 and 'BIG' performed in this study, IP- mass spectrometry of bio-tagged C10ORF12 reported by Alekseyenko and colleagues, identified several PRC2 components, as well as G9a complex components, including GLP, ZNF644 and WIZ (Alekseyenko et al, 2014). In addition, they identified the CDYL protein, which acts as a bridging molecule between the transcriptional repressor REST and the G9a complex (Mulligan et al, 2008), which was not detected in our analyses. Interestingly, the PRC2 complex was recently reported to interact with the EHMT1/GLP H3K9me1/2 methyltransferases in mouse ES cells (Mozzetta et al, 2014, Maier et al, 2015). This study expands on this subject to show that G9a/GLP and PRC2 complexes bind to two distinct domains in the C10ORF12 region of 'BIG' (Figure 5.4). However, it is unlikely that a single 'BIG' molecule would act as a bridge between PRC2 and G9a, since our preliminary re-immunoprecipitation experiment showed,

that a single 'BIG' molecule does not bind to these two complexes simultaneously in equal ratios (Figure 5.5). Re-immunoprecipitations of G9a complexes from a pool of 'BIG' immunoprecipitated material need to be performed to further investigate the interplay between 'BIG', G9a and PRC2.

Similarly, the link between CtBPs and 'BIG' can be studied in the PRC2 context. Here we establish that the 'BIG' protein contains two PXDLS CtBP-binding motifs in its N-terminus, which are also present in the LCOR protein (Fernandes et al, 2003). Interestingly, CtBP has previously been reported to interact with EHMTs as well as histone deacetylase (HDAC) proteins (Shi et al, 2003, Ueda et al, 2006). Therefore, to understand the role of 'BIG' with these repressive protein complexes and their possible interplay, size exclusion chromatography and native gel electrophoresis of the complexes immunoprecipitated with FLAG-'BIG' followed by Western blotting analysis for PRC2 components, G9a, CtBPs, USP22, USP11 and SET would need to be performed.

In its N-terminus, in addition to CtBP-motifs, 'BIG' shares a nuclear receptor-binding box with the LCOR protein. LCOR was reported to directly associate with estrogen receptor alpha (ERα) via this nuclear receptor-binding box in the presence of the activating ligand, estrogen, to mediate repression of activated ERα (Fernandes et al, 2003). In addition, LCOR has been reported to act as a co-repressor of androgen, progesterone and thyroid hormone receptors (Song et al, 2012, Asim et al, 2011, Palijan et al, 2009). Consistent with this, it is demonstrated here that 'BIG' binds to ERα (Figure 5.7, A) and that recruitment of 'BIG' to a luciferase reporter gene causes down-regulation of the luciferase activity (Figure

5.6, C). Surprisingly, we find that, unlike LCOR, 'BIG' associates with ERα in both presence and absence of the agonist estradiol (Figure 5.7, B), suggesting that 'BIG' is capable of associating with both activated and inactive ERα. To explore this observation, firstly, it could be tested whether a NR-box is the only point of direct interaction between 'BIG' and ERα, by introducing a mutation into the LXXLL motif in the 'BIG' protein by either CRISPR genome editing technology, followed by ERα immunoprecipitations in presence or absence of estradiol. Secondly, expressing exogenous 'BIG' with the LXXLL motif mutated to LXXAA and performing co-immunoprecipitations with ERα and EZH2 could further address this question. Finally, *in vitro* pull down studies using recombinant ERα and 'BIG' truncation fragments or containing mutations in LXXLL NR box in the presence or absence of estradiol could be performed to delineate the ERα interaction domains.

The data presented in this chapter imply, that 'BIG' is required for PRC2 mediated H3K27me1 and H3K27me2 deposition. So far, 'BIG' appears to be the first PRC2 sub-stoichiometric component to show such specificity. H3K27me1 has been reported to be present in the gene bodies of active genes (Ferrari et al, 2014), suggesting that 'BIG' might be linking PRC2 activity and gene activation. Furthermore, our observation that 'BIG' depletion leads to impaired proliferation of MCF7 cells indicates a possible role of 'BIG' in the recently identified function of PRC2 on the replication forks (Piunti et al, 2014). To explore this possibility further, the effect of 'BIG' depletion on primary cell lines, like human diploid fibroblasts (HDF) and mouse embryonic fibroblasts (MEF) needs to be ascertained, as well as its effects on p16 and p53 levels.

To test whether this effect holds true *in vivo,* we entered into a collaboration with the Koseki laboratory in Japan to generate an inducible 'Big' knock-out mouse. This is important because our current attempts to utilise CRISPR technology for introducing inactivating mutations into the '*Big*' gene in mouse ESCs did not yield any knock-out clones (data not shown). The depletion of 'BIG' in these mice will be tamoxifen inducible, allowing for mouse ESC proliferation and PRC2 function in depositing mono-, di- and tri-methylation of H3K27 analysis, even if the loss of 'BIG' is lethal to the cells.

In agreement with a model of 'BIG' involvement in the regulation of the PRC2 function on sites where PRC2 binding is not detected by ChIP, our ChIP analysis on Gal4-'BIG' luciferase reporter system, showed that PRC2, G9a or CtBPs do not get recruited to the luciferase promoter upon Gal4-directed 'BIG' binding (Figure 5.6, B). However, this observation has alternative explanations. First, is that a technical artifact is observed, whereby the tertiary structure of Gal-4 tagged 'BIG' when tethered to the DNA via the Gal4-tag is forced to assume an unnatural angle, thus impairing the normal association of 'BIG' with the repressors. Second, is that 'BIG' binding to chromatin is downstream of the binding of G9a, PRC2 and CtBPs, therefore Gal4-'BIG' binding does not lead to their recruitment. To address these two points, the 'BIG' target genes need to be determined by ChIP and correlated with PRC2 and G9a binding, in the presence or absence of 'BIG'. Since our 'BIG' specific antibody is not effective in ChIP application, we have devised a strategy to utilise CRISPR technology for introducing an *in vivo* biotinylation tag to the C-terminus of the 'BIG' protein in mouse embryonic stem cells, which would

facilitate ChIP of endogenous 'BIG' (Ran et al, 2013, Kim et al, 2009, Vella et al, 2012) and to ascertain its interplay with PRC2 and G9a complexes.

Chapter 6

General Discussion

6.1 Summary of the results

The aim of the doctoral research presented in this work was to explore the molecular mechanisms involved in the recruitment and function of the Polycomb-like containing PRC2 complex. The initial focus of this study was on delineating the differences between the three mammalian homologues of *Drosophila* Polycomblike proteins: PCL1/PHF1, PCL2/MTF2 and PCL3/PHF19. These three proteins have been characterised as sub-stoichiometric components of the PRC2 complex and implicated in modulation of the PRC2 enzymatic activity and recruitment to target genes (Nekrasov et al, 2007, Sarma et al, 2008, Cao et al, 2008, Li et al, 2011, Brien et al, 2013, Mussellman et al, 2012, Ballare et al, 2012, Cai et al, 2013, Qin et al, 2013). The gene expression analysis performed in this study revealed that PCL2 and PCL3 expression, similarly to PRC2, was down-regulated during the onset of cellular senescence, while the levels of the PCL1 mRNA remained unchanged, suggesting that PCL1 has a unique role in senescent cells. Furthermore, PCL1 was also found to be unique in its association with tumour suppressor p53, and this association was independent of PRC2. This project was pursued further by Dr. Gerard Brien to show that PCL1 plays a PRC2 independent role in stabilizing p53 and affecting G0/1 cell cycle arrest, thereby characterizing the importance of PCL1 in quiescent stem and progenitor cells.

The biochemical analysis of the PCL1-3 proteins performed in this work lead to identification of a novel PRC2 associated factor, the previously uncharacterised 'BIG' protein. I demonstrate that 'BIG' is a product of alternative splicing variant at the *LCOR* gene locus, in which the two coding exons of *LCOR* are spliced to the *C10ORF12* predicted ORF. LCOR and C10ORF12 peptides have been identified

in mass-spectrometric analysis of the core PRC2 component EED (Smits et al, 2013), However, this is the first report to show that it is, indeed, the novel 'BIG' protein, and not the 'canonical' LCOR, that is a sub-stoichiometric of PRC2, and the predicted *C10ORF12* gene is only expressed as a subset of 'BIG'.

It is also demonstrated that, in addition to PRC2, 'BIG' associates with a number of repressive complexes and de-ubiquitinases, including G9a/EHMT histone methyltransferase, the SET protein and de-ubiquitinase USP11 via its C10ORF12 region on the C-terminus and with the CtBP proteins 1 and 2, estrogen receptor alpha (ERα) and USP22 on its N-terminus. Furthermore, we found that 'BIG' is required for PRC2 mediated mono- and di-methylation of histone H3 at lysine 27, but not for the tri-methylation. Finally, depletion of 'BIG' reduced the proliferation rate of breast cancer cells. 'BIG' is the first PRC2 subunit identified so far to differentially modulate the PRC2 activity specifically towards H3K27me1 and H3K27me2, but not H3K27me3, with a potential to bind to nuclear receptors via its LXXLL box, making it an attractive candidate for investigation as mediator of dynamic PRC2 recruitment upon differentiation signaling.

6.2 PRC2 composition in mammalian cells

Understanding the molecular mechanisms of PRC2 recruitment and mediation of repression has been a key question in Polycomb biology. The novel sub-stoichiometric component of the PRC2 complex called 'BIG' is present in about 5-10% of the total PRC2 (Smits et al, 2013). The 'BIG' containing PRC2 complexes are also likely to contain the PCL1 or PCL2 protein, but not PCL3 or C17ORF96 as shown by us and others (Alekseyenko et al, 2014). Accordingly, C17ORF96 is preferentially bound to PCL3 and PCL2, but not PCL1 or 'BIG', indicating that 'BIG' and C17ORF96 are present in mutually exclusive complexes. PCL2 and C17ORF96 containing PRC2 complexes were reported to be highly expressed in mouse ESCs (Zhang et al, 2011, Liefke and Shi, 2015), while PCL1 expression is linked to a more differentiated state. 'BIG', like C17ORF96, does not have any paralogues in *Drosophila*, and appears to have evolutionary evolved in vertebrates, within which the PRC2 interaction domain is strongly conserved. However it is yet to be answered by *in vitro* binding assays, whether the binding of this domain to the PRC2 complex is direct.

It is not yet understood how various PRC2 complexes fulfill a functional specification. For example, PCL1 and PCL3 proteins were required for the tri-methylation of histone H3 at lysine 27 (Sarma et al, 2008, Cao et al, 2008, Hunkapiller et al, 2012, Brien et al, 2012), whereas 'BIG' depletion did not affect the H3K27me3 levels, but resulted in the downregulation of H3K27me1 and H3K27me2. These observations suggest, that while PCL1 associates with 'BIG', their action in modulating PRC2 activity differs, implying a possible antagonistic function between PCL1 and 'BIG'. Perhaps two of these factors may have

opposite allosteric effects toward enzymatic activity of EZH2 or EZH1 in the PRC2 complex. Therefore it would be interesting to determine by in vitro studies the exact binding site and the exact PRC2 component to which 'BIG' is binding, compare it to the PCL interaction domains and determined their effect by histone methyl-transferase assays. This could provide further insight to the mechanisms occurring during the switches between the repressed and activated transcription states.

6.3 Role of 'BIG' in modulation of PRC2 activity

This study demonstrated that 'BIG' is required for production of both H3K27me1 and H3K27me2, but not H3K27me3 in breast cancer cells. Interestingly, a recent study in mouse embryonic stem cells demonstrated that in addition to depositing the H3K27me3 mark on the promoters of repressed genes, PRC2 is required for the deposition of H3K27me1 along the gene bodies of actively transcribed genes (Ferrari et al, 2014). Furthermore, H3K27me1 was required for efficient transcription of these genes. On the other hand, the H3K27me2 was found to be deposited in the intergenic regions and was proposed to b involved in enhancer silencing. Concomitantly, 'BIG' is required for production of both H3K27me1 and H3K27me2, but not H3K27me3. To further investigate the requirement of 'BIG' for the H3K27 methylation, our current work is focused on generation of 'BIG' knockout and knockdown in mouse ESC in order to determine the effects of 'BIG' depletion on PRC2 occupancy and histone H3 mono-, di- and tri-methylation levels at lysine 27. In addition to these *in vivo* studies, *in vitro* histone methylation studies using the reconstituted PRC2-'BIG' complex would be required to ascertain its role in direct modulation of PRC2 enzymatic activity. Similarly, the inter-dependency

between PRC2 and 'BIG' binding can be assessed in mouse ESCs that are depleted of core PRC2 components.

6.4 Biological function of 'BIG'

The phenotype of mouse ESCs deficient of any PRC2 core component results in embryonic lethality at stage E7.5 - 8.5 days post implantation, while PRC2 expression is dispensable for ESC proliferation (Pasini et al, 2007, Brien et al, 2012, Walker et al, 2011, Zhang et al, 2011, Pasini et al, 2010). There is no data available on the phenotype of 'BIG' in this context, therefore we entered into a collaboration with the Koseki laboratory to generate a conditional knockout mouse, in which the entire Gm340 region is targeted. Finally, in order to be able to draw any conclusions regarding its association with PRC2 on target genes, the global DNA binding levels of 'BIG' need to be obtained. To this end we are generating mouse ESCs to contain a biotinylation tag on the C-terminus of the endogenous 'BIG' protein by CRISPR for ChIP studies. The above experiments are expected to reveal further molecular insight into the function of 'BIG'.

6.5 Association of 'BIG' with repressive complexes

We and others found that the C10ORF12 region of 'BIG' in addition to PRC2 is associated with a number of other repressive complexes, including the SET protein and G9a/GLP (a.k.a. EHMT2/1) complexes (Alekseyenko et al, 2014). G9a is responsible for mediating another repressive histone mark, namely di- and tri-methylation of histone H3 on lysine 9 (H3K9me2/3) (Tachibana et al, 2001, Tachibana et al, 2002). While PRC2 and G9a have been reported to interact physically and functionally (Mozzetta et al, 2014), and 'BIG' was detected as a top

hit (as Gm340) in proteomic analyses of both complexes in mouse ESCs (Maier et al, 2015), our preliminary data suggest that 'BIG' does not bind to PRC2 and G9a simultaneously. Furthermore, we showed that PRC2 and G9a interact with 'BIG' via two distinct domains in the C10ORF12 region. The role of 'BIG' in modulating the function of the G9a complex is yet to be elucidated. In particular, this could be investigated by ascertaining the effects of 'BIG' depletion on the total H3K9me2/3 levels, comparing the global 'BIG' and G9a chromatin binding patterns and determining if loss of 'BIG' affects the interaction between PRC2 and G9a complexes with each other and on the common target genes.

Interestingly, de-ubiquitinase USP11, which was previously reported to associate with Polycombs (Maertens et al, 2010), interacted with 'BIG' via the same domain as the G9a complex. USP11 was reported to be involved in modulating sensitivity to the DNA double-strand break repair (Wiltshire et al, 2010), with a mis-regulated recruitment of double strand break repair proteins in the absence of USP11. Considering that PCL1 was also reported to be involved in the response to DNA double-stand breakage (Hong et al, 2008), it would be interesting to ascertain the role of 'BIG' in this context. Since loss of USP11 sensitised cancer cells to PARP inhibitors, finding a link between 'BIG' and this pathway may lead to development of a novel biomarker or a targeted approach.

6.6 Association of 'BIG' with nuclear receptors

'BIG' contains an LXXLL nuclear receptor-binding box on its N-terminus. The same LXXLL motif is present in the LCOR protein and is required for its binding to the agonist activated ERα (Fernandes et al, 2003). Interestingly, EZH2 was

reported to bind ERα specifically in the presence of an agonist (Shi et al, 2007), suggesting that 'BIG' might act as a mediator of PRC2 binding to ERα. Surprisingly, here it is demonstrated that 'BIG' is capable of binding to ERα in either the presence or absence of agonist estradiol (E2). The interaction of 'BIG' with EZH2 was also E2 signaling independent, while ERα and EZH2 needed the presence of E2 to bind to each other. Therefore the role of 'BIG' in modulating EZH2 and ERα interaction remains unclear. Co-immunoprecipitations of ERa and PRC2 in the presence or absence of 'BIG', together with *in vitro* binding assays of truncations of 'BIG' with ERα with or without E2 signaling, as well as such analysis using the 'BIG' protein with mutations in the LXXLL NR binding box motif will be required to dissect the interplay between these proteins. Since the LXXLL motif may have the capacity to bind other nuclear receptors in the cell, they could be identified by the mass spectrometry analysis of immunoprecipitations of 'BIG' breast cancer cell lines and embryonic stem cells.

Interestingly, EZH2 was reported to be required for the activation of some estrogen and androgen receptor target genes in cancer cells (Shi et al, 2007). While recruitment of 'BIG' to a reporter gene caused moderate (~50%) repression, the mechanism of repressive action remains unclear, since no recruitment of PRC2, G9a or CtBPs to the locus was observed. The drawback of this model is that ERα is not present in HEK293T cells. Therefore, an investigation of the levels of the repressive H3K27me2/me3 and H3K9me3 as well as activating H3K27me1 and H3K36me3 histone post-translational modifications, together with their depositing complexes, in a reporter system whereby exogenous 'BIG' gets recruited to an estrogen response element, would be expected to reflect the

effects of 'BIG' action more accurately.

Interaction of 'BIG' with de-ubiquitinase USP22, which is a component of a transcriptional activator SAGA complex (Zhang et al, 2008, Zhao et al, 2008), implies another possible association between activated genes and 'BIG'. USP22 was reported to be upregulated in castration resistant prostate cancer (CRPR) and required for the stabilisation of androgen receptor (Schrecengost et al, 2014). Preliminary data in establishing the link between USP22 and PRC2 via 'BIG' showed that USP22 is not immunoprecipitated in the same complex as PRC2 from FLAG-'BIG' eluted complexes. The characterisation of 'BIG', USP22 and androgen receptor interaction could be extended by performing co-immunoprecipitations of the same in prostate cancer cell lines such as LNCaP. Further investigation of 'BIG' in modulating the action of nuclear receptors such as ERα and androgen receptor potentially would allow development of novel approaches in breast and prostate cancer therapies.

6.7 Role of 'BIG' in cellular proliferation

PRC2 is involved in control of cellular proliferation by repressing the INK4A/ARF locus, which encodes the p16 and ARF proteins, the activators of the pRb and p53 checkpoints, respectively. Interestingly, PRC2 was recently reported to be required for cellular proliferation in a manner independent of the p53 and pRb pathways (Piunti et al, 2014), by a mechanism linked to replication fork progression (Piunti et al, 2014, Hansen et al, 2008). Loss of 'BIG' expression in breast cancer cells caused reduction of cellular proliferation rates. Whether this is due to an effect on the pRb or p53 pathways is yet to be established, along with the validity of this

observation in primary human and mouse embryonic fibroblasts. Furthermore, it was recently suggested that PRC2 might mono- and di-methylate histone H3K27 at the DNA replication fork during S-phase (Ferrari et al, 2014). Given that 'BIG' is required for deposition of these two PTMs, it is tempting to speculate that 'BIG' could be involved in modulating the PRC2 function in this context. Immunofluorescence assays of 'BIG' with DNA replication factor PCNA and PRC2, as well as iPOND techniques could be used to further investigate this interaction (Hansen et al, 2007, Moldovan et al, 2007, Piunti et al, 2014).

6.8 Conclusions

The data presented in this thesis support the following conclusions:

1. p53 binding to PCL1 does not extend to PCL2 and PCL3 and is independent of PRC2
2. *LCOR* gene locus is alternatively spliced to produce a novel 'BIG' protein, which encompasses the *C10ORF12* open reading frame
3. 'BIG' is a sub-stoichiometric component of the PRC2 complex and is required specifically for the PRC2 mediated mono- and di-, but not tri-methylation of histone H3 at lysine 27
4. 'BIG' associates with estrogen receptor alpha in both the presence and absence of the agonist E2
5. 'BIG' is required for proliferation of breast cancer cell line MCF7

References

Agherbi, H., Gaussmann-Wenger, A., Verthuy, C., Chasson, L., Serrano, M., & Djabali, M. (2009). Polycomb mediated epigenetic silencing and replication timing at the INK4a/ARF locus during senescence. *PLoS ONE*, *4*(5), 1–10.

Alekseyenko, A. a, Gorchakov, A. a, Kharchenko, P. V, & Kuroda, M. I. (2014). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. Proceedings of the National Academy of Sciences of the United States of America, 111(7), 2488–93.

Asim, M., Hafeez, B. Bin, Siddiqui, I. A., Gerlach, C., Patz, M., Mukhtar, H., & Baniahmad, A. (2011a). Ligand-dependent corepressor acts as a novel androgen receptor corepressor, inhibits prostate cancer growth, and is functionally inactivated by the Src protein kinase. The Journal of Biological Chemistry, 286(43), 37108–17.

Asim, M., Hafeez, B. Bin, Siddiqui, I. A., Gerlach, C., Patz, M., Mukhtar, H., & Baniahmad, A. (2011b). Ligand-dependent corepressor acts as a novel androgen receptor corepressor, inhibits prostate cancer growth, and is functionally inactivated by the Src protein kinase. The Journal of Biological Chemistry, 286(43), 37108–17.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G. M., Morey, L., Pascual, G., … Di Croce, L. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nature Structural & Molecular Biology, 19(12), 1257–1265.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., … Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell, 129(4), 823–37.

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–95.

Basu, A., & Atchison, M. L. (2010). CtBP levels control intergenic transcripts, PHO/YY1 DNA binding, and PcG recruitment to DNA. Journal of Cellular Biochemistry, 110(1), 62–9.

Berger, S. L., Kouzarides, T., Shiekhattar, R., & Shilatifard, A. (2009). An operational definition of epigenetics An operational definition of epigenetics. Genes & Development, 23, 781–783.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., ... Lander, E. S. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell, 125(2), 315–326.

Bernstein, E., Duncan, E. M., Masui, O., Heard, E., Allis, C. D., & Gil, J. (2006). Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. Molecular and Cellular Biology, 26(7), 2560–2569.

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L. P., ... Klose, R. J. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell*, *157*(6), 1445–1459.

Bloyer, S., Cavalli, G., Brock, H. W., & Dura, J.-M. (2003). Identification and characterization of polyhomeotic PREs and TREs. Developmental Biology, 261(2), 426–442.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., ... Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature, 441(May), 349–353.

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H., & Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes & Development, 20(9), 1123–36.

Bracken, A. P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., ... Helin, K. (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. Genes & Development, 21(5), 525–30.

Bracken, A. P., Pasini, D., Capra, M., Prosperini, E., Colli, E., & Helin, K. (2003). EZH2 is downstream of the pRB-E2F pathway , essential for proliferation and amplified in cancer. The EMBO Journal, 22(20), 5323–5335.

Bracken, A. P., & Helin, K. (2009). Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nature Reviews. Cancer*, *9*(November), 773.

Brien, G. L., Gambero, G., O'Connell, D. J., Jerman, E., Turner, S. a, Egan, C. M., ... Bracken, A. P. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nature Structural & Molecular Biology, 19(12), 1273–81.

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA (New York, N.Y.), 19(4), 429–42.

Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., & Sixma, T. K. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. The EMBO Journal, 25(11), 2465–74.

Cai, L., Rothbart, S. B., Lu, R., Xu, B., Chen, W.-Y., Tripathy, A., ... Wang, G. G. (2012). An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting. Molecular Cell, 1–12.

Cai, L., Rothbart, S. B., Lu, R., Xu, B., Chen, W.-Y., Tripathy, A., ... Wang, G. G. (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Molecular Cell, 49(3), 571–82.

Cao, R., Wang, H., He, J., Tempst, P., Zhang, Y., & Erdjument-bromage, H. (2008). Role of hPHF1 in H3K27 Methylation and Hox Gene Silencing Role of

hPHF1 in H3K27 Methylation and Hox Gene Silencing. Molecular and Cellular Biology, 28(5), 1862–1872.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., ... Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science (New York, N.Y.), 298(5595), 1039–43.

Cao, R., & Zhang, Y. (2004). SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex University of North Carolina at Chapel Hill. Molecular Cell, 15, 57–67.

Caretti, G., Padova, M. Di, Micales, B., Lyons, G. E., & Sartorelli, V. (2004). The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes & Development, 18, 2627–2638.

Chan, C. S., Rastelli, L., & Pirrotta, V. (1994). A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. The EMBO Journal, 13(11), 2553–64.

Connell, S. O., Wang, L., Robert, S., Jones, C. A., Saint, R., & Jones, R. S. (2001). Polycomblike PHD Fingers Mediate Conserved Interaction with Enhancer of Zeste Protein. Biochemistry, 276(46), 43065–43073.

Coulson, M., Robert, S., Eyre, H. J., & Saint, R. (1998). The identification and localization of a human gene with sequence similarity to Polycomblike of Drosophila melanogaster. Genomics, 48(3), 381–3.

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology, 26(12), 1367–1372.

Cui, K., Zang, C., Roh, T. Y., Schones, D. E., Childs, R. W., Peng, W., & Zhao, K. (2009). Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation. Cell Stem Cell, 4(1), 80–93.

Dalgliesh, G. L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., ... Futreal, P. A. (2010). Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature, 463(7279), 360–3.

Dawson, M. a, & Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. Cell, 150(1), 12–27.

Di Croce, L., & Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. Nature Structural & Molecular Biology, 20(10), 1147–55.

Duncan, I. A. N. M. (1982). Polycomblike: a gene that appears t o be required for the normal expression of the bithorax and antennapedia gene complexes of drosophila melanogaster. Genetics, 102(Lewis 1978), 49–70.

Endoh, M., Endo, T. a, Endoh, T., Isono, K., Sharif, J., Ohara, O., ... Koseki, H. (2012). Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS Genetics, 8(7), e1002774.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., ... Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature, 473(7345), 43–49.

Ernst, T., Chase, A. J., Score, J., Hidalgo-Curtis, C. E., Bryant, C., Jones, A. V, ... Cross, N. C. P. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nature Genetics, 42(8),

Fernandes, I., Bastien, Y., Wai, T., Nygard, K., Lin, R., Cormier, O., ... White, J. H. (2003). Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and -independent mechanisms. Molecular Cell, 11(1), 139–50.

Ferrari, K. J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., ... Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Molecular Cell, 53(1), 49–62.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., ... Punta, M. (2014). Pfam: The protein families database. *Nucleic Acids Research*, *42*(D1), 222–230.

Francis, N. J., Follmer, N. E., Simon, M. D., Aghia, G., & Butler, J. D. (2009). Polycomb Proteins Remain Bound to Chromatin and DNA during DNA Replication In Vitro. Cell, 137(1), 110–122.

Francis, N. J., Kingston, R. E., & Woodcock, C. L. (2004). Chromatin compaction by a polycomb group protein complex. Science (New York, N.Y.), 306(5701), 1574–7.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., ... Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. Molecular Cell, 45(3), 344–56.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M. J., Heidersbach, A., ... Ramalho-Santos, M. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature, 460(7257), 863–868.

Goldberg, A. D., Banaszynski, L. a., Noh, K. M., Lewis, P. W., Elsaesser, S. J., Stadler, S., ... Allis, C. D. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. Cell, 140(5), 678–691.

Gurevich, I., Flores, A. M., & Aneskievich, B. J. (2007). Corepressors of agonist-bound nuclear receptors. Toxicology and Applied Pharmacology, 223(3), 288–98.

Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., ... Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nature Cell Biology, 10(11), 1291–1300.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., ... Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, *39*(3), 311–8.

Hildebrand, J. D., & Soriano, P. (2002). Overlapping and Unique Roles for C-Terminal Binding Protein 1 (CtBP1) and CtBP2 during Mouse Development . Overlapping and Unique Roles for C-Terminal Binding Protein 1 (CtBP1) and CtBP2 during Mouse Development . Molecular and Cellular Biology, 22(15), 5296.

Hong, Z., Jiang, J., Lan, L., Nakajima, S., Kanno, S., Koseki, H., & Yasui, A. (2008). A polycomb group protein, PHF1, is involved in the response to DNA double-strand breaks in human cell. Nucleic Acids Research, 36(9), 2939–47.

Hunkapiller, J., Shen, Y., Diaz, A., Cagney, G., Mccleary, D., Ramalho-santos, M., … Reiter, J. F. (2012). Polycomb-Like 3 Promotes Polycomb Repressive Complex 2 Binding to CpG Islands and Embryonic Stem. *PLoS Genetics*, *8*(3).

Inouye, C., Remondelli, P., Karin, M., & Elledge, S. (1994). Isolation of a cDNA encoding a metal response element binding protein using a novel expression cloning procedure: The one hybrid system. DNA & Cell Biology, 13(7), 731–742.

Jung, H. R., Pasini, D., Helin, K., & Jensen, O. N. (2010). Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. Molecular & Cellular Proteomics : MCP, 9(5), 838–850.

Kalb, R., Latwiel, S., Baymaz, H. I., Jansen, P. W. T. C., Müller, C. W., Vermeulen, M., & Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nature Structural & Molecular Biology, 21(6), 569–71.

Kawakami, S., Mitsunaga, K., Kikuti, Y. Y., Ando, A., Inoko, H., Yamamura, K. I., & Abe, K. (1998). Tctex3, related to Drosophila polycomblike, is expressed in male germ cells and mapped to the mouse t-complex. Mammalian Genome, 9(11), 874–880.

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., … Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proceedings of the

National Academy of Sciences of the United States of America, 106(28), 11667–11672.

Kim, H., Kang, K., & Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. Nucleic Acids Research, 37(9), 2940–50.

Kim, J., Cantor, A. B., Orkin, S. H., & Wang, J. (2009). Use of in vivo biotinylation to study protein – protein and protein – DNA interactions in mouse embryonic stem cells. In Vivo, 4(4), 506–517.

Kim TW, Kang BH, Jang H, Kwak S, Shin J, Kim H, Lee SE, Lee SM, Lee JH, Kim JH, Kim SY, Cho EJ5, Kim JH, Park KS, Che JH, Han DW, Kang MJ, Yi EC, Youn HD. (2015). Ctbp2 Modulates NuRD-Mediated Deacetylation of H3K27 and Facilitates PRC2-Mediated H3K27me3 in Active Embryonic Stem Cell Genes During Exit from Pluripotency.

Koppens, M., & van Lohuizen, M. (2015). Context-dependent actions of Polycomb repressors in cancer. Oncogene, (April), 1–12.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell, 128(4), 693–705.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., & Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes & Development, 16(22), 2893–905.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H. F., … Fisher, A. G. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nature Cell Biology, 12(6), 618–24.

Lanigan, F., Geraghty, J. G., & Bracken, A. P. (2011). Transcriptional regulation of cellular senescence. Oncogene, 30(26), 2901–2911.

Laugesen, A., & Helin, K. (2014). Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell, 14(6), 735–751.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., … Young, R. A. (2006). Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. Cell, 125, 301–313.

Li, H., Ma, X., Wang, J., Koontz, J., Nucci, M., & Sklar, J. (2007). Effects of rearrangement and allelic exclusion of JJAZ1/SUZ12 on cell proliferation and survival. Proceedings of the National Academy of Sciences of the United States of America, 104(50), 20001–6.

Li, X., Isono, K., Yamada, D., A, T., Endoh, M., Shinga, J., … Kamijo, T. (2011). Mammalian Polycomb-Like Pcl2 / Mtf2 Is a Novel Regulatory Component of PRC2 That Can Differentially Modulate Polycomb Activity both at the Hox Gene Cluster and at Cdkn2a Genes Mammalian Polycomb-Like Pcl2 / Mtf2 Is a Novel Regulatory Component of PRC2 Tha. Molecular Cell Biology, 31(2), 351.

Li, Z., Cao, R., Wang, M., Myers, M. P., Zhang, Y., & Xu, R.-M. (2006). Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. The Journal of Biological Chemistry, 281(29), 20643–9.

Liao, T.-L., Wu, C.-Y., Su, W.-C., Jeng, K.-S., & Lai, M. M. C. (2010). Ubiquitination and deubiquitination of NP protein regulates influenza A virus RNA replication. The EMBO Journal, 29(22), 3879–90.

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

Lonie, a, D'Andrea, R., Paro, R., & Saint, R. (1994). Molecular characterisation of the Polycomblike gene of Drosophila melanogaster, a trans-acting negative regulator of homeotic gene expression. Development (Cambridge, England), 120(9), 2629–36.

Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature, 389(6648), 251–260.

Maertens, G. N., El Messaoudi-Aubert, S., Elderkin, S., Hiom, K., & Peters, G. (2010). Ubiquitin-specific proteases 7 and 11 modulate Polycomb regulation of the INK4a tumour suppressor. The EMBO Journal, 29(15), 2553–65.

Maier, V. K., Feeney, C. M., Taylor, J. E., Creech, A. L., Qiao, J. W., Szanto, A., … 1. (2015). Functional proteomics defines a PRC2-G9A interaction network and reveals ZNF518B as a G9A regulator. Mol Cell Proteomics, 14(6), 1435–46.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C. L., … Reinberg, D. (2008). Article Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. Molecular Cell, 32(4), 503–518.

Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, *469*(7330), 343–9.

McCabe, M. T., Graves, A. P., Ganji, G., Diaz, E., Halsey, W. S., Jiang, Y., … Creasy, C. L. (2012). Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). Proceedings of the National Academy of Sciences of the United States of America, 109(8), 2989–94.

McDonel, P., Costello, I., & Hendrich, B. (2009). Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. The International Journal of Biochemistry & Cell Biology, 41(1), 108–16.

Melnick, A. (2012). Epigenetic therapy leaps ahead with specific targeting of EZH2. Cancer Cell, 22(5), 569–70.

Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., … Bernstein, B. E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genetics, 6(12), e1001244.

Micci, F., Panagopoulos, I., Bjerkehagen, B., & Heim, S. (2006). Consistent rearrangement of chromosomal band 6p21 with generation of fusion genes JAZF1/PHF1 and EPC1/PHF1 in endometrial stromal sarcoma. Cancer Research, 66(1), 107–12.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., ... Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(August), 553–560.

Mohd-Sarip, A., Venturini, F., Chalkley, G. E., & Verrijzer, C. P. (2002). Pleiohomeotic Can Link Polycomb to DNA and Mediate Transcriptional Repression Pleiohomeotic Can Link Polycomb to DNA and Mediate Transcriptional Repression. Molecular and Cellular Biology, 22(21), 7473–7483.

Moldovan, G. L., Pfander, B., & Jentsch, S. (2007). PCNA, the Maestro of the Replication Fork. Cell, 129(4), 665–679.

Morey, L., Aloia, L., Cozzuto, L., Benitah, S. A., & Di Croce, L. (2013). RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. Cell Reports, 3(1), 60–9.

Morin, R. D., Johnson, N. a, Severson, T. M., Mungall, A. J., An, J., Goya, R., ... Marra, M. a. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nature Genetics, 42(2), 181–5.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., ... Ait-Si-Ali, S. (2014). The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing. Molecular Cell, 53, 277–289.

Müller, H., Bracken, a P., Vernell, R., Moroni, M. C., Christians, F., Grassilli, E., ... Helin, K. (2001). E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. Genes & Development, 15(3), 267–85.

Mulligan, P., Westbrook, T. F., Ottinger, M., Pavlova, N., Chang, B., Macia, E., ... Shi, Y. (2008). CDYL bridges REST and histone methyltransferases for gene repression and suppression of cellular transformation. Molecular Cell, 32(5), 718–26.

Musselman, C. A., Avvakumov, N., Watanabe, R., Abraham, C. G., Lalonde, M., Hong, Z., ... Kutateladze, T. G. (2012). Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nature Structural & Molecular Biology, 19(12), 1266–1272.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Ko, T., ... Wilm, M. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO Journal, 26(18), 4078–4088.

Nibu, Y. (1998). Interaction of Short-Range Repressors with Drosophila CtBP in the Embryo. Science, 280(5360), 101–104.

Nikoloski, G., Langemeijer, S. M. C., Kuiper, R. P., Knops, R., Massop, M., Tönnissen, E. R. L. T. M., ... Jansen, J. H. (2010). Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nature Genetics, 42(8), 665–7.

Ntziachristos, P., Tsirigos, A., Van Vlierberghe, P., Nedjic, J., Trimarchi, T., Flaherty, M. S., ... Aifantis, I. (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nature Medicine, 18(2), 298–301.

O'Connell, S., Wang, L., Robert, S., Jones, C. a., Saint, R., & Jones, R. S. (2001). Polycomblike PHD Fingers Mediate Conserved Interaction with Enhancer of Zeste Protein. *Journal of Biological Chemistry*, *276*(46), 43065–43073.

Olins, D. E., Bryan, P. N., Harrington, R. E., Olins, A. L., Division, B., & Ridge, O. (1977). Nucleic Acids Research, 4(6), 1911–1931.

Orlando, V. (2003). Polycomb, Epigenomes and Control of Cell Identity. *Cell*, *112*, 599–606.

Palijan, A., Fernandes, I., Bastien, Y., Tang, L., Verway, M., Kourelis, M., ... White, J. H. (2009). Function of histone deacetylase 6 as a cofactor of nuclear receptor coregulator LCoR. The Journal of Biological Chemistry, 284(44), 30264–74.

Palijan, A., Fernandes, I., Verway, M., Kourelis, M., Bastien, Y., Tavera-Mendoza, L. E., ... White, J. H. (2009a). Ligand-dependent corepressor LCoR is an attenuator of progesterone-regulated gene expression. The Journal of Biological Chemistry, 284(44), 30275–87.

Palijan, A., Fernandes, I., Verway, M., Kourelis, M., Bastien, Y., Tavera-Mendoza, L. E., ... White, J. H. (2009b). Ligand-dependent corepressor LCoR is an attenuator of progesterone-regulated gene expression. The Journal of Biological Chemistry, 284(44), 30275–87.

Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M., & Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Molecular and Cellular Biology, 27(10), 3769–79.

Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E., & Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. The EMBO Journal, 23(20), 4061–71.

Pasini, D., Malatesta, M., Jung, H. R., Walfridsson, J., Willer, A., Olsson, L., ... Helin, K. (2010). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Research, 38(15), 4958–69.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., ... Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO Journal*, *30*(20), 4198–4210.

Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., & Sidow, A. (2009). Jarid2 / Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. Cell, 139(7), 1290–1302.

Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A., & Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. Science (New York, N.Y.), 339(6120), 698–9.

Perissi, V., Jepsen, K., Glass, C. K., & Rosenfeld, M. G. (2010). Deconstructing repression: evolving models of co-repressor action. Nature Reviews. Genetics, 11(2), 109–23.

Piunti, A., Rossi, A., Cerutti, A., Albert, M., Jammula, S., Scelfo, A., … Pasini, D. (2014). Polycomb proteins control proliferation and transformation independently of cell cycle checkpoints by regulating DNA replication. Nature Communications, 5, 3649.

Poortinga, G., Watanabe, M., & Parkhurst, S. M. (1998). Drosophila CtBP: a Hairy-interacting protein required for embryonic segmentation and hairy-mediated transcriptional repression. The EMBO Journal, 17(7), 2067–78.

Puda, A., Milosevic, J. D., Berg, T., Klampfl, T., Harutyunyan, A. S., Gisslinger, B., … Kralovics, R. (2012). Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. American Journal of Hematology, 87(3), 245–50.

Qin, S., Guo, Y., Xu, C., Bian, C., Fu, M., Gong, S., & Min, J. (2013). Tudor domains of the PRC2 components PHF1 and PHF19 selectively bind to histone H3K36me3. Biochemical and Biophysical Research Communications, 430(2), 547–53.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. a, & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature Protocols, 8(11), 2281–308.

Ringrose, L., & Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annual Review of Genetics, 38, 413–43.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., … Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Non-Coding RNAs. Cell, 129(7), 1311–1323.

Roth, S. Y., Denu, J. M., & Allis, C. D. (2001). Histone Acetyltransferases. Annual Review of Biochemistry, 70, 81–120.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., & Reinberg, D. (2008). Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27 Trimethylation In Vivo Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27 Trimethylation In Vivo. Molecular Cell Biology, 28(8), 2718.

Sauvageau, M., & Sauvageau, G. (2010). Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. Cell Stem Cell, 7(3), 299–313.

Savla, U., Benes, J., Zhang, J., & Jones, R. S. (2008). Recruitment of Drosophila Polycomb-group proteins by Polycomblike , a component of a novel protein complex in larvae. Development, 817, 813–817.

Schmitges, F. W., Prusty, A. B., Faty, M., Stützer, A., Lingaraju, G. M., Aiwazian, J., … Thomä, N. H. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Molecular Cell, 42(3), 330–341.

Schrecengost, R. S., Dean, J. L., Goodwin, J. F., Schiewer, M. J., Urban, M. W., Stanek, T. J., … Knudsen, K. E. (2014). USP22 regulates oncogenic signaling pathways to drive lethal cancer progression. Cancer Research, 74(1), 272–86.

Schwartz, Y. B., Kahn, T. G., Nix, D. a, Li, X.-Y., Bourgon, R., Biggin, M., & Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nature Genetics, 38(6), 700–5.

Schwartz, Y. B., & Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nature Reviews. Genetics*, *14*(12), 853–64.

Schwartzentruber, J., Korshunov, A., Liu, X.-Y., Jones, D. T. W., Pfaff, E., Jacob, K., … Jabado, N. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature, 482(7384), 226–31.

Sengupta, A. K., Kuhrs, A., & Müller, J. (2004). General transcriptional silencing by a Polycomb response element in Drosophila. Development (Cambridge, England), 131(9), 1959–65.

Shen, X., Liu, Y., Hsu, Y., Fujiwara, Y., Kim, J., Mao, X., ... Orkin, S. H. (2008). Article EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. Molecular Cell, 32(4), 491–502.

Shi, B., Liang, J., Yang, X., Wang, Y., Zhao, Y., Wu, H., ... Shang, Y. (2007). Integration of estrogen and Wnt signaling circuits by the polycomb group protein EZH2 in breast cancer cells. Molecular and Cellular Biology, 27(14), 5105–19.

Shi, Y., Sawada, J., Sui, G., & Affar, E. B. (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature*, 735–738.

Simon, J. a, & Kingston, R. E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Molecular Cell, 49(5), 808–24.

Sims, R. J., Chen, C. F., Santos-Rosa, H., Kouzarides, T., Patel, S. S., & Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. Journal of Biological Chemistry, 280(51), 41789–41792.

Smits, A. H., Jansen, P. W. T. C., Poser, I., Hyman, A. a, & Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Research, 41(1), e28.

Sneeringer, C. J., Scott, M. P., Kuntz, K. W., Knutson, S. K., Pollock, R. M., Richon, V. M., & Copeland, R. a. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proceedings of the National Academy of Sciences of the United States of America, 107(49), 20980–5.

Song, Y., Shan, S., Zhang, Y., Liu, W., Ding, W., Ren, W., ... Ying, H. (2012). Ligand-dependent corepressor acts as a novel corepressor of thyroid hormone receptor and represses hepatic lipogenesis in mice. Journal of Hepatology, 56(1), 248–54.

Srinivasan, L., & Atchison, M. L. (2004). YY1 DNA binding and PcG recruitment requires CtBP. Genes & Development, 18(21), 2596–601.

Steiner, L. a., Schulz, V. P., Maksimova, Y., Wong, C., & Gallagher, P. G. (2011). Patterns of histone H3 lysine 27 monomethylation and erythroid cell type-specific gene expression. Journal of Biological Chemistry, 286(45), 39457–39465.

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. Nature, 403(6765), 41–45.

Strunk, B., Struffi, P., Wright, K., Pabst, B., Thomas, J., Qin, L., & Arnosti, D. N. (2001). Role of CtBP in transcriptional repression by the Drosophila giant protein. Developmental Biology, 239(2), 229–40.

Tachibana, M., Sugimoto, K., Fukushima, T., & Shinkai, Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. The Journal of Biological Chemistry, 276(27), 25309–17.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., ... Shinkai, Y. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes & Development, 16(14), 1779–91.

Tammen, S. a., Friso, S., & Choi, S. W. (2013). Epigenetics: The link between nature and nurture. Molecular Aspects of Medicine, 34(4), 753–764.

Thierry-Mieg, D., & Thierry-Mieg, J. (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biology, 7 Suppl 1(Suppl 1), S12.1–14.

Tie, F., Prasad-sinha, J., Birve, A., & Harte, P. J. (2003). A 1-Megadalton ESC / E (Z) Complex from Drosophila That Contains Polycomblike and RPD3. Molecular and Cellular Biology, 23, 3352–3362.

Tolhuis, B., de Wit, E., Muijrers, I., Teunissen, H., Talhout, W., van Steensel, B., & van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nature Genetics, 38(6), 694–9.

Turner, S. a, & Bracken, A. P. (2013). A "complex" issue: deciphering the role of variant PRC1 in ESCs. Cell Stem Cell, 12(2), 145–6.

Ueda, J., Tachibana, M., Ikura, T., & Shinkai, Y. (2006). Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. The Journal of Biological Chemistry, 281(29), 20120–8.

Van Haaften, G., Dalgliesh, G. L., Davies, H., Chen, L., Bignell, G., Greenman, C., ... Futreal, P. A. (2009). Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nature Genetics, 41(5), 521–3.

Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T., & Pasini, D. (2012). Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. Nucleic Acids Research, 40(8), 3403–18.

Vettese-Dadey, M., Grant, P. a, Hebbes, T. R., Crane- Robinson, C., Allis, C. D., & Workman, J. L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. The EMBO Journal, 15(10), 2508–2518.

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. a, & Kinzler, K. W. (2013). Cancer genome landscapes. Science (New York, N.Y.), 339(6127), 1546–58.

Voigt, P., LeRoy, G., Drury, W. J., Zee, B. M., Son, J., Beck, D. B., ... Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell, 151(1), 181–93.

Voigt, P., Tee, W. W., & Reinberg, D. (2013). A double take on bivalent promoters. Genes and Development, 27, 1318–1338.

Walker, E., Manias, J. L., Chang, W. Y., & Stanford, W. L. (2011). PCL2 modulates gene regulatory networks controlling self-renewal and commitment in embryonic stem cells. Cell Cycle, 10(1), 45–51.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature, 431(7010), 873–878.

Wang, L., Brown, J. L., Cao, R., Zhang, Y., Kassis, J. A., Jones, R. S., ... Carolina, N. (2004). Hierarchical Recruitment of Polycomb Group Silencing Complexes. Molecular Cell, 14, 637–646.

Wang, S., Robertson, G. P., & Zhu, J. (2004). A novel human homologue of Drosophila polycomblike gene is up-regulated in multiple cancers. Gene, 343(1), 69–78.

Zhuo Zhang, Amanda Jones, Chiao-Wang Sun, Chao Li, Chia-Wei Chang, Heui-Yun Joo, Qian Dai, Matthew R. Mysliwiec, Li-Chen Wu, Yahong Guo, Wei Yang, Kaimao Liu, Kevin M. Pawlik, Hediye Erdjument-Bromage, Paul Tempst, Youngsook Lee, Jinrong Min, Tom M. Town, H. W. (2011). PRC2 Complexes with JARID2, MTF2, and esPRC2p48 in ES Cells to Modulate ES Cell Pluripotency and Somatic Cell Reprograming. *Stem Cells*, *29*, 229–240.

Watson, P. J., Fairall, L., & Schwabe, J. W. R. (2012). Nuclear hormone receptor co-repressors: structure and function. Molecular and Cellular Endocrinology, 348(2), 440–9.

Wiltshire, T. D., Lovejoy, C. a, Wang, T., Xia, F., O'Connor, M. J., & Cortez, D. (2010). Sensitivity to poly(ADP-ribose) polymerase (PARP) inhibition identifies ubiquitin-specific peptidase 11 (USP11) as a regulator of DNA double-strand break repair. The Journal of Biological Chemistry, 285(19), 14565–71.

Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. Nature Methods, 6(5), 359–362.

Woodcock, C. L., Safer, J. P., & Stanchfield, J. E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. *Experimental Cell Research*, *97*, 101–110.

Wu, G., Broniscer, A., McEachron, T. a, Lu, C., Paugh, B. S., Becksfort, J., ... Baker, S. J. (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nature Genetics, 44(3), 251–3.

Wu, X., Johansen, J. V., & Helin, K. (2013). Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. Molecular Cell, 49(6), 1134–46.

Wysocka, J., Swigut, T., Xiao, H., Milne, T. a, Kwon, S. Y., Landry, J., ... Allis, C. D. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature, 442(7098), 86–90.

Yang, Y., Wang, C., Zhang, P., Gao, K., Wang, D., Yu, H., ... Yu, L. (2012). Polycomb group protein PHF1 regulates p53-dependent cell growth arrest and apoptosis. The Journal of Biological Chemistry, 1–21.

Yang, Y., Wang, C., Zhang, P., Gao, K., Wang, D., Yu, H., ... Yu, L. (2013). Polycomb group protein PHF1 regulates p53-dependent cell growth arrest and apoptosis. The Journal of Biological Chemistry, 288(1), 529–39.

You, J. S., & Jones, P. a. (2012). Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell, 22(1), 9–20.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., & Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. Journal of Biological Chemistry, 286(10), 7983–7989.

Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research*, *21*(8), 1273–1283.

Zhang, J., Ding, L., Holmfeldt, L., Wu, G., Heatley, S. L., Payne-Turner, D., ... Mullighan, C. G. (2012). The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature, 481(7380), 157–63.

Zhang, X., Pfeiffer, H. K., Thorne, A. W., & Mcmahon, S. B. (2008). USP22, and hSAGA subunit and potential cancer stem cell marker, reverses the polycomb-catalyzed ubiquitylation of histone H2A. Cell Cycle, 7(11), 1522–1524.

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., ... Lee, J. T. (2010). Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. Molecular Cell, 40(6), 939–953.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J., & Jeannie, T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X-chromosome. Science, 322(5902), 750–756.

Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., ... Devys, D. (2008). A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. Molecular Cell, 29(1), 92–101.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K. a, Lozach, J., ... Rosenfeld, M. G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Molecular Cell, 29(1), 69–80.

Zhuo Zhang, Amanda Jones, Chiao-Wang Sun, Chao Li, Chia-Wei Chang, Heui-Yun Joo, Qian Dai, Matthew R. Mysliwiec, Li-Chen Wu, Yahong Guo, Wei Yang, Kaimao Liu, Kevin M. Pawlik, Hediye Erdjument-Bromage, Paul Tempst, Youngsook Lee, Jinrong Min, Tom M. Town, H. W. (2011). PRC2 Complexes with JARID2, MTF2, and esPRC2p48 in ES Cells to Modulate ES Cell Pluripotency and Somatic Cell Reprograming. Stem Cells, 29, 229–240.