# Polycomb group proteins recruitment during mouse development

**Lígia Tavares**

Corpus Christi College

Dissertation submitted for the degree of Doctor of Philosophy

Department of Biochemistry
University of Oxford

October 2010

# Abstract

Polycomb proteins are a group of repressor proteins that form multiprotein complexes, Polycomb repressor complexes (PRC) that mediate gene silencing. Two major complexes that have been widely studied are PRC1 and PRC2.Both complexes have enzymatic activities that modify specific histone residues, monoubiquitination of H2AK119 (H2AK119u1) and methylation of H3K27 (H3K27me3) respectively, and it is thought that this is the basis for their silencing activity. The basis for recruitment of PcG proteins in mammalian systems remains to be fully elucidated. A key concept is the hierarchical model explaining co-occupancy of PRC1 and PRC2 at target loci. Here it is proposed that a chromodomain in the protein polycomb (Pc), a core PRC1 subunit, recognises the H3K27me3 histone modification catalysed by PRC2. Although some observations appear inconsistent with this idea it is nevertheless widely accepted.

In this study the *Ink4b-Arf-Ink4a* locus was used as a model to study PcG recruitment both by examining PcG localization and by setting up a reporter system that could be used to screen for novel regulators of PcG recruitment. The former analysis revealed that Exon E2, which is shared between *Ink4a* and *Arf* is strongly bound by PcG proteins, suggesting it as a good candidate for a recruitment centre. However, deletion of exon E2 does not inhibit the recruitment of PcG to other regulatory regions within the locus. In the latter study P1-derived artificial chromosome (PAC) recombineering was used to produce a modified PAC in which fluorescent marker proteins replaced endogenous *Ink4a* and *Arf* open reading frames. This will provide a useful tool for screening for factors that contribute to *Ink4a-Arf* activation including PcG recruitment factors.

Preliminary data from the lab indicated that global levels of PRC1 mediated H2AK119u1 are largely unaffected in cells lacking H3K27me3, an unexpected outcome based on the hierarchical recruitment model. Chromatin immunoprecipitation (ChIP) was carried out on *Eed* mutant (H3K27me3 depleted) ESCs and consistent with the hierarchical model, occupancy of the core PRC1 proteins RING1B and MEL18 was strongly reduced. However, H2AK119u1 was clearly detectable at target loci, suggesting PRC1 recruitment independent of H3K27me3. Genome wide analysis of RING1B occupancy substantiated this conclusion, demonstrating significant RING1B levels at polycomb target loci in $Eed^{-/-}$ ESCs.

From this study it can be concluded that PRC1 and PRC2 are recruited independently to sites that they co-occupy. Additionally evidence is provided that the primary function of H3K27me3 is to increase the residency of PRC1 at target loci and thereby to contribute to the stability of PRC1 mediated silencing.

*A ti Pedro*

# Acknowledgments

Almost 5 years ago I started my PhD in Portugal, through the GABBA programme. Since then I joined the Developmental Epigenetics group from Neil Brockdorff at Imperial College, London, and then moved to the Department of Biochemistry in Oxford. In this journey many people were important to me and to this work and without them the completion of this project would not have been possible or it would have at least been more difficult. For them my sincere thank you.

Mentioning some of the most important people in that journey I first want to thanks my supervisor Neil Brockdorff for taking me as his student. I came to him as an immature and insecure student and under his supervision I grew as a scientist. I learned valuable lessons with him and I gained a model, someone whose research I believe. With him I have learned how to be a scientist, how science is exciting and challenging. I want to acknowledge all the present and former members of the Brockdorff's lab for all their support, tutoring, and friendship. I know that I made friend that will be kept for life and I will be waiting for you all in Portugal! A special thank you goes to Tatyana Nesterova. You were with me since day 1 in the lab until the very last minute. You were there for me whenever I needed, no matter how busy you where. You are much more than a colleague; you are a big friend, thanks for it all. I also want to thank Claire Senner. Thanks for being an amazing colleague in the lab but more than that, thanks for being a friend. I will always remember our dinners; we definitely have to repeat them. I also want to give a big big thanks for your help with this thesis, for your amazing work on transforming it into proper English. A special thank you also goes for Heather Coker, Jonathan Godwin and Sarah Cooper. You made me feel in the lab as if it was home (of course it had its drawbacks; I used to forget to go home...). Seriously you are great friends. Tea? Lunch? Pub? I will always remember those times. Thanks also all the members of Rob Klose lab which whom I had a very nice time both socializing and discussing science.

My time in Oxford was great and for that I have to give a special Obrigada to my housemate and friend Sara Duarte. I made there a real friend that will always be kept in my heart. Thanks for listening to me, for the nice meals we cooked, for the nights out. I had a really good time. Thanks to all my other Portuguese friends which whom I had lots of fun. A special thanks goes to Mafalda who handles in this thesis! You took this huge responsibility, thanks a lot! Not only

the friends I gained during my PhD but also the friend that came from a long time where deeply important to me. My dearest friends Catarina Pimentel, Ana Costa e Catarina Santos “o meu adoro-vos”. I know you will be always there for me. Thanks Ana Paula Rebocho for living with me in London. I miss our little flat. And thanks Mariana Cerqueira for introducing me to London. I really enjoyed it!

Aos meus pais, Maria Rosa e António o meu muito, muito obrigada. Vocês fizeram de mim tudo o que eu sou hoje e sem vocês ao meu lado este percurso tinha sido bem mais difícil. Obrigada pais por me terem criado, por terem feito de mim uma apaixonada por aquilo que eu faço e pela vida. E a minha vida e tudo o que eu sou devo-o a vocês. Obrigada pais! E nunca se esqueçam que vos amo muito. E a ti Susana, a minha maninha linda, Obrigada pela paciência. Pelas horas ao telefone, pelas visitas, pelo carinho. Obrigada por estares sempre lá quando eu preciso. Um agradecimento especial vai para o meu afilhado, João. Porque apesar de eu estar longe quando ele nasceu e em muitos momentos da vida dele sempre me recebeu de braços abertos e com um sorriso matreiro.

And last but not the least thanks you Pedro. Obrigada amor, obrigada por TUDO. E só tu sabes o que TUDO significa. Obrigada pelo apoio, pela força, pelo carinho mas também pelas discussões e chatices. Foram 4 anos separados. Conseguimos o impossível, o nosso amor resistiu ao improvável e nós, 4 anos depois estamos mais juntos do que nunca, mais unidos do que alguma vez pensamos possível. E agora unidos o futuro é nosso! E eu quero estar ao teu lado, fisicamente ao teu lado! Amo-te Pedro, e amo-te ainda mais por nao teres desistido e por teres lutado ao meu lado. Obrigada! E não te esqueças que todos os dias “I’m in love with you”.

To you all

**THANK YOU!!!**

# Abbreviations

| | |
|---|---|
| *µ* | *Micro* |
| *5meC* | *5 methyl cytosine* |
| *Ab* | *Antibody* |
| *ac* | *Acethyl* |
| *ActD* | *Actinomycin D* |
| *ADP* | *Adenosine diphosphate* |
| *AEBP* | *AE binding protein* |
| *ATP* | *Adenosine triphosphate* |
| *B2m* | *Beta-2-microglobulin* |
| *BiFC* | *Bimolecular fluorescence complementation* |
| *bp* | *Base pair* |
| *CBB* | *Coomassie brilliant blue* |
| *Cbx* | *Chromobox homolog* |
| *Cdkn* | *Cyclin-dependent kinase inhibitor* |
| *Cdkn2aip* | *Cyclin-dependent kinase inhibitor 2A interacting protein* |
| *Cdx* | *Caudal type homeo box* |
| *Cfp1* | *CXXC finger protein 1* |
| *CGI* | *CpG island* |
| *CHD* | *Chromodomain helicase DNA-binding* |
| *ChIP* | *Chromatin Immunoprecipitation* |
| *ChIP-Seq* | *ChIP-Sequencing* |
| *cKO* | *Conditional knockout* |
| *COMPASS* | *Complex proteins associated with set1* |
| *CpG* | *—C—phosphate—G—* |
| *DNA* | *Deoxyribonucleic acid* |
| *Dnmt* | *DNA methyltransferase* |
| *Dox* | *Doxycycline* |
| *dRaf* | *dRing-associated factors* |
| *DUB* | *deubiquitinase* |
| *E* | *Embryonic day* |
| *E(Z)* | *Enhancer of zeste* |
| *Eed* | *Embryonic ectoderm development* |
| *Ensa* | *Endosulfine alpha* |
| *ESC* | *Embryonic stem cell* |
| *Esc* | *Extra sex combs* |
| *Ezh* | *Enhancer of zeste* |
| *g* | *Gravitational force* |
| *Gapdh* | *Glyceraldehyde-3-phosphate dehydrogenase* |
| *Gata* | *GATA binding protein 1* |

| | |
|---|---|
| *GC* | *Guanine-cytosine* |
| *GO* | *Gene ontology* |
| *H* | *Histone* |
| *h* | *Hour* |
| *HAT* | *Histone acetyltransferase* |
| *HDAC* | *Histone deacetylase* |
| *HDMT* | *Histone demethyltransferase* |
| *Hmbs* | *Hydroxymethylbilane synthase* |
| *HMT* | *Histone methyltransferase* |
| *Hox* | *Homeobox* |
| *HS* | *Heat shock* |
| *Idh* | *Isocitrate dehydrogenase* |
| *IP* | *Immunoprecipitate* |
| *ISWI* | *Imitation switch* |
| *JmjC* | *Jumonji C domain* |
| *K* | *Lysine* |
| *kb* | *kilobase* |
| *kDa* | *Kilo dalton* |
| *KDM* | *Lysine demethylase* |
| *KO* | *knockout* |
| *lncRNA* | *Long ncRNA* |
| *M* | *Molar* |
| *Math* | *Atonal homologue* |
| *MBP* | *Methyl-CpG-binding protein* |
| *me* | *Methyl* |
| *MEF* | *Mouse embryonic fibroblast* |
| *mESC* | *Mouse ESC* |
| *min* | *Minute* |
| *miRNAs* | *Micro RNAs* |
| *ml* | *Millilitre* |
| *MLL* | *Mixed lineage leukemia* |
| *mM* | *Millimolar* |
| *Msx* | *Homeo box, msh-like* |
| *n* | *Nano* |
| *ncRNA* | *Non coding RNA* |
| *Nkx2* | *NK2 homeobox* |
| *nt* | *Nucleotides* |
| *NuRD* | *Nucleosome remodeling and deacetylase* |
| *NuRF* | *Nucleosome remodeling factor* |
| $^{o}C$ | *Celsium degrees* |

| | |
|---|---|
| *ON* | *Overnight* |
| *P* | *Phosphate* |
| *PAC* | *P1-derived artificial chromosome* |
| *Pax* | *Paired box* |
| *PBS* | *Phosphate buffered saline* |
| *Pc* | *Polycomb* |
| *PcG* | *polycomb group* |
| *Pcgf* | *polycomb group ring finger* |
| *PEF* | *Primary embryonic fibroblast* |
| *Ph* | *Polyhomeotic* |
| *PRC* | *Polycomb repressive complex* |
| *PRE* | *PcG response element* |
| *Psc* | *Posterior sex combs* |
| *PTM* | *Posttranslational modification* |
| *qPCR* | *Quantitative real-time PRC* |
| *Rbp* | *Retinoblastoma binding protein* |
| *Recombineering* | *recombination-mediated genetic engineering* |
| *Ring1* | *Sex combs on midleg1* |
| *RNA* | *Ribonucleic acid* |
| *RNAPII* | *Ribonucleic acid polymerase II* |
| *Rnf* | *Ring finger protein* |
| *rpm* | *Rotation per minute* |
| *Rps* | *Ribosomal proteim* |
| *s* | *Second* |
| *siRNAs* | *Small interfering RNAs* |
| *Sox* | *SRY (sex determining region Y)-box* |
| *Suz* | *Suppressor of zeste* |
| *SWI/SNF* | *switch independent/sucrose nonfermenting* |
| *Tm* | *Annealing temperature* |
| *TRE* | *Trithorax response element* |
| *Trx* | *Trithorax* |
| *TrxG* | *Trithorax group* |
| *TSA* | *Trichostatin A* |
| *TSS* | *Transcription start site* |
| *u* | *Ubiquitin* |
| *UV* | *Ultraviolet* |
| *WT* | *Wild-type* |
| *Xa* | *Active X* |
| *Xi* | *Inactive X* |
| *Xist* | *X inactivation specific transcript* |

# Table of Contents


Abstract ..... 2

Acknowledgments ..... 5

Abbreviations ..... 7

Table of Contents ..... 10

1. General Introduction ..... 14

1.1. Epigenetics ..... 14

1.1.1. Historical overview ..... 14

1.1.2. Epigenetics across Development ..... 15

1.2. Chromatin landscape ..... 15

1.2.1. Heterochromatin ..... 17

1.2.2. Euchromatin ..... 19

1.3. Histones ..... 19

1.3.1. Histone variants ..... 19

1.3.2. Histone modifications ..... 21

1.3.3. Histone modifying enzymes ..... 33

1.4. DNA methylation ..... 38

1.4.1. CpG islands ..... 40

1.5. Chromatin remodelling complexes ..... 42

1.6. Non coding RNAs ..... 45

1.6.1. *Xist* (X inactivation specific transcript) ..... 46

1.6.2. Short ncRNAs ..... 47

1.7. Heritability of chromatin landscapes ..... 47

1.7.1. Mitotic heritability ..... 48

1.7.2. Meiotic heritability ..... 50

1.8. Epigenetic gene regulation in mammalian development ..... 50

1.8.1. Zygote development ..... 50

1.8.2. First steps: capturing the pluripotent state ..... 53

1.8.3. Growing into adulthood ..... 58

1.8.4. Cancer ..... 59

1.8.5. Aims of this study ..... 60

2. Polycomb recruitment to the *Ink4b-Arf-Ink4a* locus ..... 61

2.1. Introduction ..... 61

2.1.1. PcG recruitment to target genes ..... 61
2.1.2. Regulation of the *Ink4b-Arf-Ink4a* locus regulation ..... 63
2.2. Results ..... 66
2.2.1. Polycomb proteins are recruited to the *Ink4b-Arf-Ink4a* locus in ESCs ..... 66
2.2.2. Role of Polycomb in MEF senescence ..... 68
2.2.3. Role of *Ink4a*/*Arf* exon E2 in PcG recruitment ..... 72
2.2.4. siRNA screen to unveil new factors that recruit Polycomb ..... 74
2.3. Discussion ..... 76
2.3.1. PcG regulation of the *Ink4*/*Arf* locus ..... 76
2.3.2. PcG recruitment to the *Ink4b-Arf-Ink4a* ..... 78
3. PRC1 activity in *Eed* mutant mouse embryonic stem cells ..... 80
3.1. Introduction ..... 80
3.1.1. Importance of PcG proteins in mouse development ..... 80
3.1.2. Hierarchical recruitment model ..... 81
3.2. Results ..... 84
3.2.1. H2AK119u1 levels are maintained in $Eed^{-/-}$ ESCs ..... 84
3.2.2. *Eed* conditional knockout mimics constitutive depletion ..... 87
3.2.3. *de novo* establishment of H2AK119u1 ..... 91
3.3. Discussion ..... 94
3.3.1. Recruitment of PRC1 and H2AK119u1 in *Eed* depleted cells ..... 94
3.3.2. *de novo* monoubiquitination of H2A ..... 94
3.3.3. Role of PRC1 and PRC2 in gene silencing ..... 95
4. H3K27me3 independent recruitment of RING1B ..... 97
4.1. Introduction ..... 97
4.1.1. Genome-wide distribution of PRC2 and PRC1 ..... 97
4.1.2. PcG proteins in flies and mammals ..... 98
4.2. Results ..... 100
4.2.1. ChIP-Seq data analysis ..... 100
4.2.2. PRC1 bound regions in $Eed^{-/-}$ ESCs ..... 107
4.2.3. PcG recruitment independent of RNAPII transcription ..... 111
4.3. Discussion ..... 118
4.3.1. RING1B recruitment independent of H3K27me3 ..... 118
4.3.2. A role for PRC2 in PRC1 dynamics ..... 119
4.3.3. Role of CpG islands in PRC1 recruitment ..... 120

4.3.4. PRC1 and PRC2 recruitment independent of RNAPII transcription ..... 121
5. General Discussion ..... 123
5.1. Polycomb recruitment to chromatin ..... 123
5.2. The role of PRC1 in gene silencing ..... 128
5.3. PRC1 mechanism of action ..... 131
6. Material and Methods ..... 133
6.1. Tissue culture techniques ..... 133
6.1.1. Cell lines and media ..... 133
6.1.2. Growing, thawing and freezing cells ..... 134
6.1.3. Treatments ..... 136
6.1.4. 3T3 protocol ..... 137
6.1.5. Feeders ..... 137
6.1.6. Lipofection of ESCs with PAC DNA ..... 137
6.2. Mouse work ..... 140
6.2.1. Mice genotyping ..... 140
6.2.2. ESC derivation ..... 141
6.2.3. PEFs derivation ..... 141
6.2.4. PAC microinjection into mice ..... 142
6.3. Senescence-associated β-galactosidase staining ..... 143
6.4. Alkaline Phosphatase staining ..... 143
6.5. Genomic DNA extraction ..... 144
6.5.1. Polymerase chain reaction ..... 144
6.6. RNA related techniques ..... 145
6.6.1. RNA extraction ..... 145
6.6.2. Complementary DNA synthesis ..... 146
6.6.3. Real-time quantitative analysis ..... 146
6.7. Chromatin immunoprecipitation ..... 148
6.7.1. Crosslinked ChIP ..... 148
6.7.2. Short fragments ChIP ..... 153
6.8. Deep-sequencing ..... 154
6.8.1. Library preparation and sequencing ..... 154
6.8.2. Data analysis ..... 154
6.9. Protein work ..... 156
6.9.1. Protein extraction ..... 156

6.9.2. Protein quantification ..... 157
6.9.3. Nuclear Fractionation ..... 157
6.9.4. Protein immunoprecipitation ..... 158
6.9.5. Western blot ..... 160
6.10. Bacteria work and plasmid construction ..... 162
6.10.1. Plasmids ..... 162
6.10.2. Media and culture conditions ..... 162
6.10.3. Plasmid/PAC DNA extraction ..... 163
6.10.4. Restriction enzyme digestion ..... 165
6.10.5. Arms of homology ..... 165
6.10.6. Ligation ..... 166
6.10.7. Transformation ..... 167
6.10.8. Recombination-mediated genetic engineering ..... 167
7. Bibliography ..... 172
Appendix 1 ..... 199
Appendix 2 ..... 208

# 1. General Introduction

## 1.1. Epigenetics

### 1.1.1. Historical overview

The term epigenetics was introduced into modern biology by Conrad Waddington (1905-1975) when in the 1940's he referred to the "epigenetic landscape" (Slack 2002). His work as an embryologist made him look at embryonic development as a dialogue between cytoplasm and

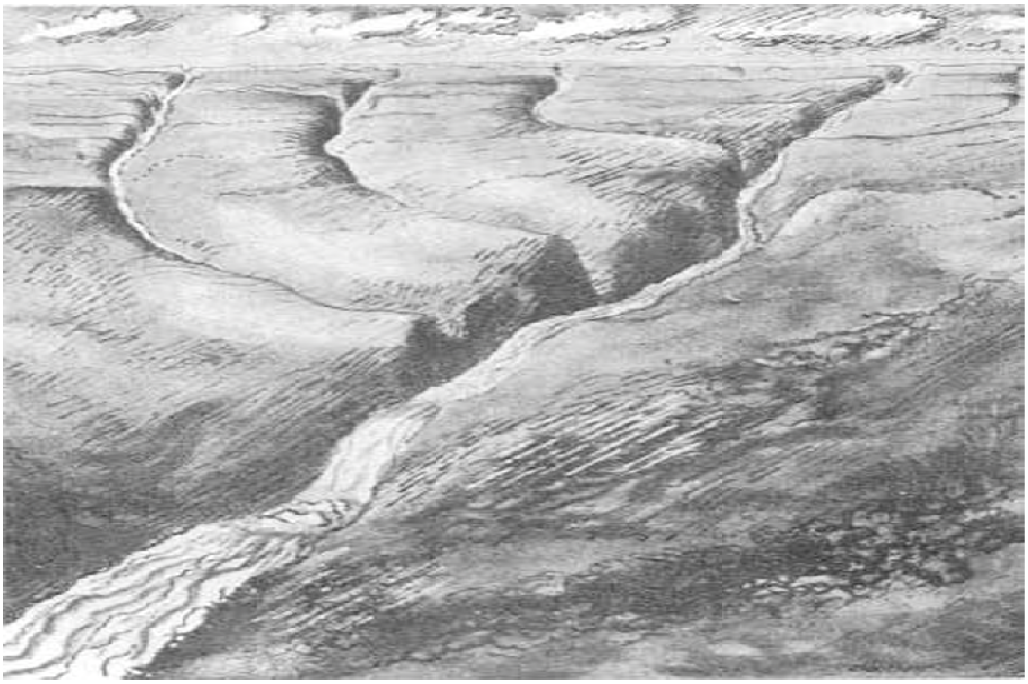


**Figure 1.1 – The Epigenetic Landscape.** Painting by John Piper used as the frontispiece of Waddington's book *Organisers and Genes*. The picture represents an allegory for the Waddington's definition of epigenetic landscape where the water in the river represents each cell in the embryo that at any point can choose different pathways towards the sea at the end. Adapted from Slack (Slack, 2002).

genes. For him, changes in the cytoplasm of a cell would create differential gene expression when the cell divides. Division after division would result in diverse gene expression patterns and a differentiated organism would be formed. It is this action of genes in development that Waddington defined as "epigenetics" (Morange, 2009). In Figure 1.1, the "epigenetic landscape" represents cellular metabolism as a multidimensional space. At each point the cell

has a binary decision to make and its decision is a result of cellular metabolism and the consequent gene action that cooperate to form the landscape as a whole (Slack, 2002).

Over the years all the phenomena that could not be explained by genetic mutation started to be included in the "epigenetics" category: paramutation in maize, imprinted genes and position-effect variegation (Goldberg et al., 2007). Research showed that many of the non-mendelian modes of inheritance had common underlying mechanisms that are important for all cellular processes. Thus, epigenetics gained recognition, and today it is a well established area of biology. The initial definition of epigenetics has changed substantially and today epigenetics is viewed as a bridge between genotype and phenotype - a phenomenon that controls the final outcome of a locus without modifying the underlying deoxyribonucleic acid (DNA) sequence (Goldberg et al., 2007).

#### 1.1.2.Epigenetics across Development

Comparison of the genomes of different species revealed a paradox, that an increase in the complexity of an organism was not reflected in an increase in gene number (Lander et al., 2001; Venter et al., 2001; Waterston et al., 2002). It is now known that it is the many mechanisms of regulation of those same genes that reflects the complexity of an organism. Regulatory regions receive, and interpret signals in the form of regulatory proteins, transcription factors, non coding ribonucleic acids (ncRNAs), histone modifying complexes and histone modifying enzymes, giving a final outcome depending on the balance between those factors (Carroll, 2000; Tautz, 2000).

### 1.2. Chromatin landscape

In 1880 Walther Flemming observed a nuclear substance that he called chromatin (Maeshima et al., 2010). The basic unit of chromatin (Figure 1.2), the nucleosome, consists of 147 basepair

(bp) of DNA wrapped around an octamer of histones, spaced every 157-240 bp. Each histone octamer is formed by two H2A-H2B dimers on each side of an (H3-H4)2 tetramer (Davey et al., 2002; Luger et al., 1997). Thirty years ago, Finch and Klug proposed that nucleosome arrays folded into a highly compacted 30 nm chromatin fibre in a process facilitated by the linker histone H1 (Finch and Klug, 1976; Thoma et al., 1979). The chromatin fibre would then form large-scale configurations that make up chromosomes (Maeshima et al., 2010).

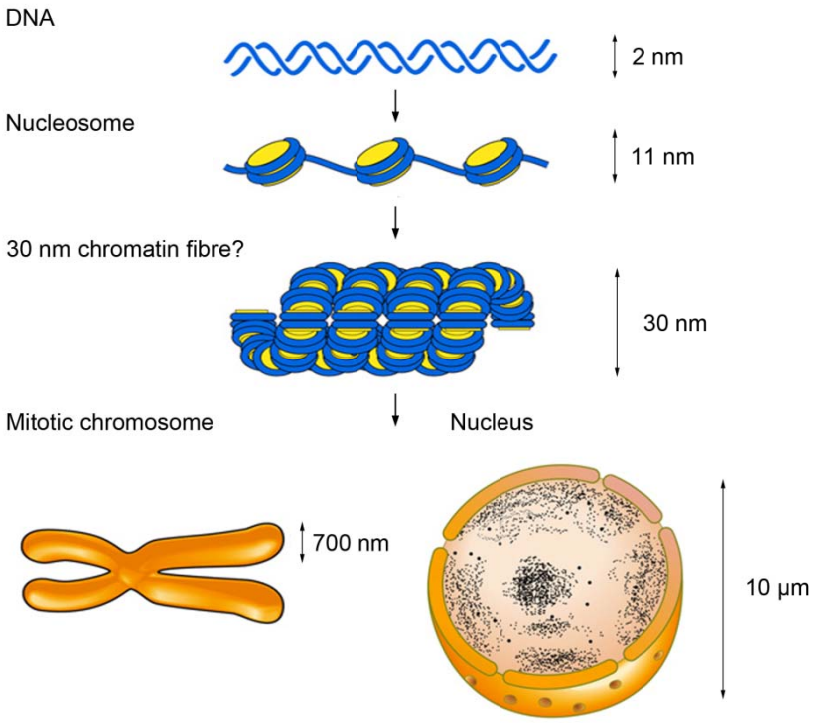



**Figure 1.2 – Chromatin organization in the nucleus.** The 2 nm DNA molecule is wrapped around a histone octamer forming a nucleosome, the basic unit of chromatin, with a diameter of 11 nm. Nucleosomes fold up into a highly compacted 30nm chromatin fibre; such a structure has been recently challenged in vivo. The chromatin fibre would then form large-scale configurations that make up chromosomes allowing genetic information to be kept in the nucleus. Figure adapted from Maeshima at al. (Maeshima et al., 2010).

Recently the 30 nm fibre structure has been challenged. Eltsov and colleagues propose a dynamic disordered model for chromatin compaction, similar to a polymer melt that could

exist together with areas of more compact organization, more similar to the 30 nm fibre (Eltsov et al., 2008). This relaxed and dynamic structure would facilitate biological processes such as transcription and DNA replication. The discussion about the existence of the 30 nm chromatin fibre is still open as recently Li and colleagues generated a 30 nm chromatin fibre *in vitro* that, they believe, mimics the *in vivo* compacted fibre while possessing a dynamic structure that responds to transcription stimulus (Li et al., 2010a).

Thirty years on from the first visualization of chromatin another important discovery was made. Emil Heitz detected that the same chromosomal regions stained brightly in the same cell type of different individuals, while other regions stained fainter, he associated them with condensed and decondensed chromatin states, respectively (Trojer and Reinberg, 2007). Thus, chromatin could be divided into functional areas: heterochromatin and euchromatin (Figure 1.3).

### 1.2.1.Heterochromatin

Heterochromatin can be divided into constitutive and facultative heterochromatin (Figure 1.3). Constitutive heterochromatin corresponds to the highly compact chromatin blocks that remain condensed throughout the cell cycle such as those found at centromeres (pericentromeric heterochromatin) and telomeres. Here heterochromatin plays an important role in sister-chromatid cohesion and chromosome segregation as well as inhibiting recombination between homologous repeats (Grewal and Moazed, 2003). Facultative heterochromatin, on the other hand, corresponds to regions that can respond to cellular stimuli and includes developmental regulated loci (Grewal and Jia, 2007). It is transcriptionally silent by definition but with the potential to convert between heterochromatin and euchromatin. An example is the inactive X (Xi) chromosome (Avner and Heard, 2001).

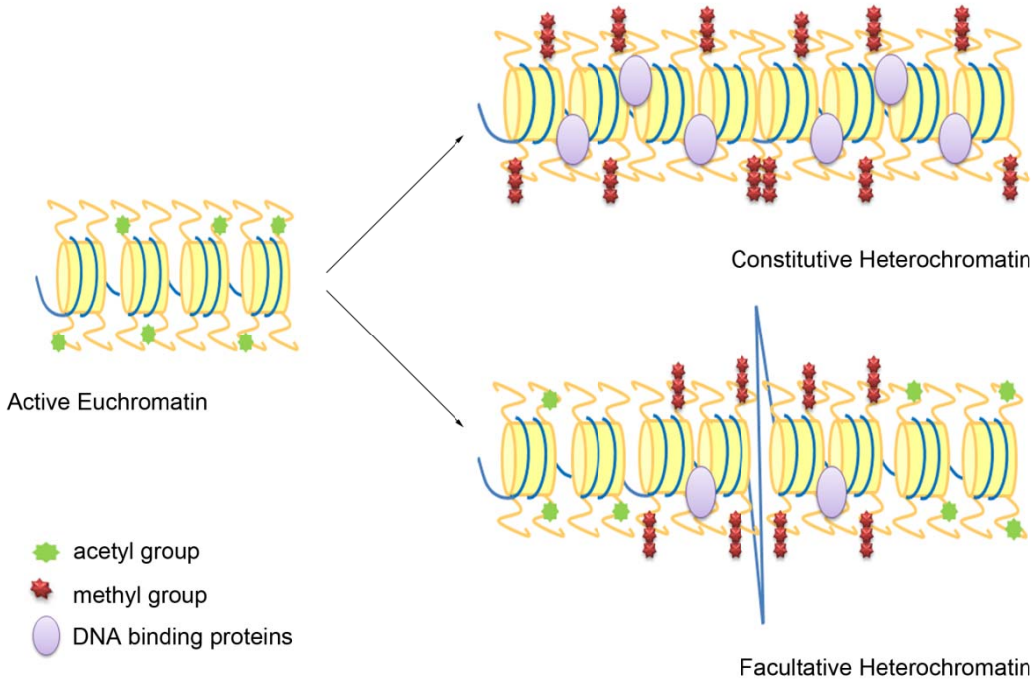



**Figure 1.3 – The different chromatin states: Euchromatin, Constitutive and Facultative Heterochromatin.** Euchromatin corresponds to the active open state of the chromatin (high level of histone acetylation). Opposing it is constitutive heterochromatin which corresponds to the highly compact chromatin blocks that remain condensed throughout the cell cycle, such as those found at centromeres. DNA binding proteins may have a role in chromatin compaction. Facultative chromatin corresponds to an intermediate state with regions that can respond to cellular stimuli. Figure adapted from Trojer and Reinberg (Trojer and Reinberg, 2007).

Although it is not clear how heterochromatin is formed it seems that it involves interplay between regulatory DNA sequences and DNA binding proteins, repetitive DNA elements and ncRNAs (Grewal and Moazed, 2003). These features are thought to be needed to recruit machinery essential for the transcriptional silencing mediated by histone (H) H3 lysine (K) 9 (H3K9), H3K27 and H4K20 methylation, histone deacetylation, DNA methylation and HP1 (heterochromatin protein 1) binding. Nucleation of heterochromatin is important for cell regulation and because it can spread along the chromosome, leading to silencing of euchromatic regions adjacent to heterochromatin, it is important to keep heterochromatin within specific limits. Insulators or DNA barriers have been found to delimit and stop the spreading of heterochromatin (Grewal and Jia, 2007).

### 1.2.2.Euchromatin

In contrast to heterochromatin, euchromatin is gene rich and active. At any given time, genes can be either repressed or activated in response to cellular stimuli making euchromatin a very dynamic structure. It needs to allow access to transcription factors and RNA Polymerase II (RNAPII), DNA repair and DNA replication machinery (Quina et al., 2006). Local modifications reflect these dynamic properties with repressed areas exhibiting low levels of DNA methylation, acetylation and phosphorylation while active regions are hyperacetylated and have high levels of trimethyl (me3) H3K4 (H3K4me3), H3K36me3 and H3K79me3 (Figure 1.3). Further to this, canonical histones can be replaced by histone variants in areas of specialization (Talbert and Henikoff, 2010).

Thus chromatin is a dynamic structure, with specialized areas defining expression, regulation and maintenance of cell identity. Tight regulation is required with several processes controlling assembly and regulation of chromatin: histone modifications, DNA methylation, adenosine triphosphate (ATP)-dependent chromatin remodelling complexes and ncRNAs.

## 1.3. Histones

There are two types of histones: canonical histones and non-canonical histones (histone variants). Changes in the core histones of the nucleosome can induce alterations in the organization and spacing of nucleosomes and affect chromatin dynamics as well as cellular processes).

### 1.3.1.Histone variants

Histone variants differ from their canonical counterparts in their amino acid sequence. They have a wider spectrum of functions including DNA repair, chromosome segregation, sex chromosome condensation, transcription initiation and termination, meiotic recombination

and sperm chromatin packaging (Talbert and Henikoff, 2010). Evidence so far indicates that they are important for specialization of certain areas as they are deposited actively in a replication independent manner (Ahmad and Henikoff, 2002).

H2A.Z is the most studied being highly conserved across species and essential in flies and mice development (Faast et al., 2001; van Daal and Elgin, 1992). Opposing roles have been described for H2A.Z: transcription and gene repression, a barrier for the spreading of heterochromatin and heterochromatin formation, amongst others (Talbert and Henikoff, 2010). Possible explanations for this disparity might be the presence of post translational modification (PTM) on H2A.Z nucleosomes, effects of chromatin remodelling complex responsible for its deposition, binding by H1 or variation in H2A.Z nucleosomes composition (Talbert and Henikoff, 2010).

Macro H2A can be encoded by 2 genes in vertebrates: *mH2A1* (2 isoforms, *mH2A1.1* and *mH2A1.2*) and *mH2A2*. Both mH2A.1 and mH2A.2 are enriched on the Xi chromosome, having a role in transcriptional silencing and dosage compensation in mammals (Chadwick and Willard, 2001; Costanzi and Pehrson, 1998, 2001; Mermoud et al., 1999). In addition, both isoforms were found in human male pluripotent cells at the promoters of developmentally regulated genes, overlapping with polycomb repressive complex (PRC) 2 target genes (Buschbeck et al., 2009).

H3.3 varies only in 4 amino acids from the core histone H3 (H3.1). Genome wide Chromatin immunoprecipitation (ChIP)-Sequencing (ChIP-Seq) studies in mouse embryonic stem cells (mESCs) showed that H3.3 is present at promoters with high guanine-cytosine (GC) content, the body of transcribed genes, transcription factor binding sites and telomeres (Goldberg et al., 2010). Its presence at active genes was confirmed by another study in HeLa cells where it

was shown that H3.3 is present at the transcription start site (TSS) in combination with H2A.Z (Jin et al., 2009). This association with highly dynamic nucleosomes at promoters and gene bodies of active genes links this variant with RNAPII engagement (Jin et al., 2009).

### 1.3.2. Histone modifications

Histones can be covalently modified on their N-terminal tails by methylation, acetylation, ubiquitination, phosphorylation, ADP-ribosylation and sumoylation (Figure 1.4). These modifications can induce a change in global chromatin environments either by compartmentalisation of DNA into euchromatin and heterochromatin or by changing DNA specific functions at the local level, where modifications contribute to the wrapping/unwrapping of DNA to facilitate specific functions such as gene expression or DNA replication (Sims et al., 2003).

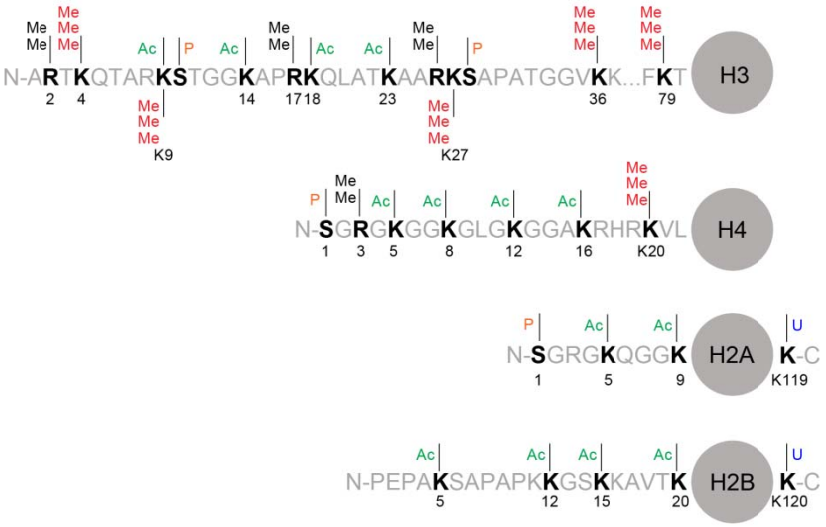



**Figure 1.4 – Most of post-translational modifications on histones H3, H4, H2A and H2B.** Histones can be covalently modified on their N-terminal tails by methylation, acetylation, ubiquitination and phosphorylation. Lysine (L) methylation (Me) is indicated in red while arginine (R) methylation (Me) is represented in black. Phosphorylation (P) is represented in orange. Acetylation (Ac) is represented in green and Ubiquitination (u) is represented in blue. The highest degree of methylation possible for each residue is shown. Residues are annotated below the histone tail. Figure adapted from Sims and colleagues (Sims et al., 2003).

#### 1.3.2.1. Methylation

Early studies of methylation showed that it occurs on the ε-amino group of the lysine on the nucleosome tail in a reaction catalysed by a histone methyltransferase (HMT) using the methyl (me) group donor S-adenosyl-L-methionine (Shi and Whetstine, 2007). HMTs share a SET domain but catalyse methylation reactions in a very specific way.

For many years it was thought that histone methylation was an extremely stable modification as histone demethylases had not been identified (Sims et al., 2003). That vision changed abruptly with the discovery of histone demethylases that can regulate the methylation level on specific lysine residues (Ng et al., 2009). Histone methylation is therefore a highly dynamic epigenetic mark. Both HTMs and histone demethylases are involved in regulating the methylation status of specific lysine residues involved in transcriptional regulation, genome integrity, transcriptional memory and development (Bedford, 2007; Swigut and Wysocka, 2007).

The first lysine demethylases (KDM) described was KDM1 (LSD1 - lysine-specific demethylase) that acts on mono- and dimethylated H3K4 (Shi et al., 2004), but which requires the interaction with the co-repressor Co-REST (Shi and Whetstine, 2007). The existence of this enzyme facilitated the discovery of other demethylases that can act on mono- di- and trimethylated residues. A group of histone demethylases is the dioxygenases represented by the Jumonji C domain (JmjC). Several JmjC enzymes have been found to possess high specificity to demethylate specific methylated residues on nucleosomes (Shi and Whetstine, 2007). For example, KDM3A (JHDM2A - JmjC domain containing histone demethylase) can demethylate mono- and dimethylated H3K9 (Yamane et al., 2006), KDM2A (FBXL11 (F-box and leucine-rich repeat protein)/JHDM1A) and KDM2B, (FBXL10/JHDM1B) mono- and dimethylates H3K36

(Tsukada et al., 2006), while KDM5 (Jarid1) is a demethylase specific for H3K4me3 (Seward et al., 2007).

Some of the other histone PTMs induce charge alterations in chromatin that lead to changes in nucleosome dynamics. So far, methylation seems to be a relatively inert modification that functions as a nucleation site for the binding of "effector" proteins (Taverna et al., 2007) that recognize the methyl groups through chromo-like domains of the Royal family - chromo, tudor, MBT - and nonrelated PHD domains (Kouzarides, 2007). Depending on the methylated residue various associated proteins can have roles as positive or negative regulators of transcription. Another layer of regulation is added as each residue can be mono- , di-, or trimethylated and both HMTs and histone demethylases show high specificity to the level of methylation that they catalyse (Sims et al., 2003).

There are two types of histone methylation: arginine (R) methylation and lysine methylation. Within lysine methylation, the most studied modifications are H3K4, K9, K27, K36, K79 and H4K20. From those, H3K4, K36 and K79 methylation are usually associated with active or poised chromatin whereas H3K9, K27 and H4K20 methylation are found in repressed regions (Ng et al., 2009).

H3K4me localization studies show correlation with places of active transcription, but is also found at "bivalent" genes (see section 1.8.2.2.1) being important for the poised state (Azuara et al., 2006; Bernstein et al., 2006a). Recent work using mass spectrometry corroborated previous *in vivo* studies showing that H3K4 is mainly found on hyperacetylated H3 but never found together with the hallmark of repressed chromatin, H3K9 (Garcia et al., 2007). Its role as a positive regulator of transcription can be exerted through the disruption of repressive chromatin remodelling complexes association, such as the nucleosome remodeling and

deacetylase (NuRD) complex (Xue et al., 1998) which besides its role of chromatin remodeler possesses deacetylase activity (Lee et al., 2005; Martin and Zhang, 2005). In addition H3K4me facilitates the binding of nucleosome remodeling factor (NURF) and chromodomain helicase DNA-binding (CHD) 1 chromatin remodelling complexes that are associated with open chromatin states (Li et al., 2006a; Sims and Reinberg, 2006).

H3K36me3 associates at the gene body with serine (S) 2 phosphate (P), the elongating form of RNAPII, and histone deacetylases (Lee and Shilatifard, 2007). It suppresses initiation from cryptic start sites within the coding region by recruiting EAF3 that in turn recruits Rpd35 deacetylases, removing all the acetyl (ac) groups in the coding region after transcription, resetting the chromatin state (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005).

H3K79me3 is an unusual modification because it lies within the nucleosome core, rather than in the tail. Very little is known about this modification but it has been implicated in activation of *Hox* (Homeobox) *a9* and has a role in restraining heterochromatin to confined regions (Sims et al., 2003).

H3K9 methylation is involved in chromatin silencing, as well as heterochromatin formation and X-chromosome inactivation. The working model formulated for H3K9me3 involves methylation of H3K9 by SUV39H (suppressor of variegation 3-9 homolog 1) that creates a binding site for HP1, through its chromodomain (Bannister et al., 2001; Lachner et al., 2001). Once HP1 is bound it can move to adjacent nucleosomes by association with SUV39H, spreading silencing to the neighbouring regions (Nakayama et al., 2001). H3K9me3, together with H3K20me3 are associated with transposable elements as well as with repetitive sequences. These elements are deleterious for the genome and association of modifications linked with heterochromatin may lead to the stable silencing of those regions. In addition, basal ncRNA transcription from

those regions may be important for H3K9me3 targeting (Ng et al., 2009). For years H3K9 was only associated with chromatin silencing but this association may not be completely true as H3K9me3 and HP1γ have been found in regions of active transcription (Vakoc et al., 2005).

H3K27me3 is important for the silencing of developmentally regulated genes such as *Hox* genes and on the Xi chromosome. Silencing is mediated by enhancer of zeste (E(z)) 1 or 2 (EZH1/EZH2) methyltransferases, part of the Polycomb group (PcG) of repressor proteins (Simon and Kingston, 2009) (See 1.3.3.1 section).

Relatively little is known about H4K20me. It has a role in heterochromatin formation and DNA repair and paradoxically a role in transcription. It has been shown to be one of the most stable histone marks (Nightingale et al., 2007).

#### 1.3.2.2. Acetylation

One of the best characterized PTM is histone acetylation, catalyzed by histone acetyltransferases (HATs). HATs function by transferring an acetyl group from acetyl-coenzyme A to the ε-amino group of a lysine on the N-terminal histone tail (Sterner and Berger, 2000), with the level of acetylation correlating with the level of transcription (Choi and Howe, 2009). Acetyl groups neutralize the basic charge of lysine, disrupting interactions between nucleosomes and leading to chromatin unfolding and increased DNA accessibility, consistent with its role in transcriptional activation (Kouzarides, 2007). HAT complexes are usually composed of one catalytic subunit and auxiliary proteins that potentiate, target and regulate the activity of the complex. HATs are generally targeted to the genome through interactions with transcriptional activators, RNAPII and FACT (facilitates chromatin transcription) (Choi and Howe, 2009). There are three main families of acetyltransferases: GNAT, MYST and CBP/p300 (Sterner and Berger, 2000). It was demonstrated that (as expected due to their role in

transcription), HATs are required for normal development and differentiation into the different embryonic tissues (Roth et al., 2001).

With the use of sequencing H3ac was mapped in human T cells. Acetylated domains were detected that correspond to active TSS and cytosine-P-guanine (CpG) islands (CGIs) in gene-rich areas (Roh et al., 2005). A recent study mapped the genome-wide distribution of 18 histone acetylation marks. Some modifications associate mainly with TSS regions, for example H2AK9ac, H2BK5ac, H3K9ac, H3K18ac, H3K27ac, H3K36ac and H4K91ac while others are found in promoter and transcribed regions of active genes: H2BK12ac, H2BK20ac, H2BK120ac, H3K4ac, H4K5ac, H4K8ac, H4K12ac and H4K16ac, demonstrating clearly that different modifications regulate different regions in the genome (Wang et al., 2008). The different distributions correlate with specificity of HATs for certain chromosomal regions (Cho et al., 1998).

An interesting study showed cooperativity between HATs and HDACs in the regulation of gene expression. At active genes, high levels of HATs and histone deacetylase (HDACs) are recruited by the elongating form of RNAPII. Here HDACs reset the chromatin state by removing acetyl groups added by HATs. At poised genes marked with active H3K4me3, HDACs and HATs are present at low levels. HDACs restore the silent state counteracting the action of HATs and inhibiting RNAPII elongation. On the other hand, when H3K4me3 is not detected and genes are repressed, neither HDACs nor HATs can be detected (Wang et al., 2009).

A question that remains unsolved is whether the effect of acetylation in gene expression is direct or mediated by effector proteins such as proteins with bromodomains that may recognize histone acetylation groups.

### 1.3.2.3. Ubiquitination

Ubiquitin (u) groups are large polypeptides (76 amino acids) that increase the size of histones by approximately two thirds. Histones are usually monoubiquitinated which distinguishes them from the protein polyubiquitination that targets proteins for proteosomal degradation. It can have a role either in activation or silencing of gene expression depending on the modified residue (H2BK120 or H2AK119 respectively).

Ubiquitination generally involves three enzymatic reactions mediated by three different enzymes. First ubiquitin is activated by an ATP-dependent reaction through an ubiquitin activation enzyme (E1). Secondly, the conjugation of the ubiquityl group to a cysteine residue of an ubiquitin-conjugating enzyme (E2) occurs followed by the transfer of the ubiquitin group from the E2 enzyme to the lysine residue on the respective histone by the ubiquitin-protein isopeptidase ligase (E3) (Figure 1.5) (Weake and Workman, 2008).

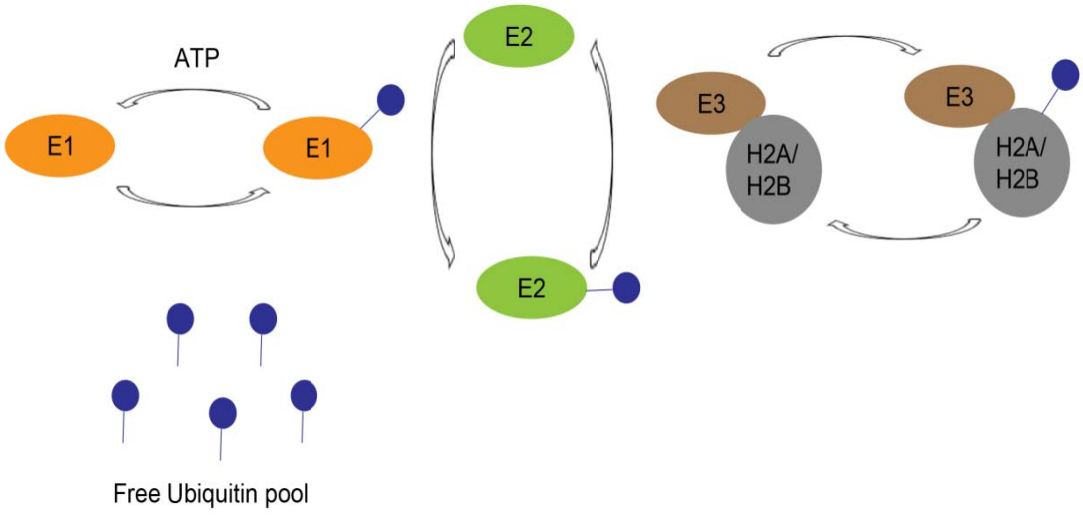



**Figure 1.5 – Histone ubiquitination pathway.** Ubiquitin groups are large polypeptides (76 amino acids) kept in a free pool in the cytoplasm. Ubiquitination pathway is mediated by an activation enzyme (E1) that binds the ubiquitin group in a reaction activated by ATP. Then the ubiquitin group is transferred to an ubiquitin-conjugating enzyme (E2) being then transferred to the histone substract by an ubiquitin-protein isopeptidase ligase (E3).

#### 1.3.2.3.1. H2B ubiquitination

H2B monoubiquitination was first identified in *Saccharomyces cerevisiae* (Robzyk et al., 2000) where RAD6, in conjunction with the E3 ubiquitin ligase Bre1, catalyzes the monoubiquitination of lysine 123 (H2BK123u1) (Weake and Workman, 2008). In mammals H2B monoubiquitination occurs on lysine 120 (H2BK120u1) and is mediated by HR6A (also known as UBE2A - ubiquitin-conjugating enzyme E2A) and HR6B (also known as UBEB2B) (Koken et al., 1991; Roest et al., 1996). Another ubiquitin ligase UBCH6 (also known as UBE2E1) acts in a complex with the E3 ubiquitin ligase ring finger protein (RNF) 20/RNF40 but its role *in vivo* has not been completely proven (Zhu et al., 2005).

Deubiquitination is also tightly regulated with many ubiquitin proteases (Ubps) being found from yeast to humans. Efficient turnover between ubiquitination and deubiquitination of H2B is required for correct activation of gene expression. UBP8 in yeast is associated with SGF11 in the SAGA (Spt-Ada-Gcn5-Acetyltransferase) complex. This association has also been found in *Drosophila Melanogaster* and humans and it is required to regulate H2BK120u1 levels (Weake and Workman, 2008).

For many years the role of H2BK120u1 in transcription was inconclusive. More recently, robust evidence has started to emerge. ChIP-Seq has demonstrated genome-wide location of H2BK120u1 strongly linked with highly transcribed genes in human cells (Minsky et al., 2008). H2BK120u1 is regulated by the phosphorylation of RNAPII carboxy terminal domain (CTD), being established once transcription is initiated (Campos and Reinberg, 2009). It can itself regulate deposition of active marks such as H3K4, H3K36 and H3K79 methylation. Further evidence for the link between H2B and RNAPII came from *in vivo* studies that show that RNAPII elongation is dependent on the histone chaperone FACT, recruitment of PAF (transcription elongation complex) and H2B machinery, independently of H3K4me3 (Pavri et al., 2006).

#### 1.3.2.3.2. H2A ubiquitination

Histone H2A can become modified by monoubiquitination at lysine 119 (H2AK119u1) (Bohm et al., 1980). This modification is mediated in humans by at least two different E3 ubiquitin ligases, RING1B ( also known as RNF2) and 2A-HUB, while in flies and mice it is mediated by RING1A and RING1B (de Napoles et al., 2004; Wang et al., 2004a; Zhou et al., 2008). Knockdown of *RING1B* in human cells (Wang et al., 2004a) or gene deletion in mice (de Napoles et al., 2004) strongly reduces H2AK119u1 levels indicating that this is the main E3 ligase both in humans and in mice. RING1B associates with three major repressive complexes: PRC1, E2F-6.com-1 and FBXL10-BcoR (FBXL10-Bcl6 co-repressor).

PRC1 (see section 1.3.3.1.2) was shown to possess E3 ligase activity specific for histone H2A (Wang et al., 2004a). PRC1 dependent ubiquitination is present on the Xi chromosome where *Ring1B* knockout (KO) or knockdown removes ubiquitin moieties from the chromosome (de Napoles et al., 2004; Fang et al., 2004), as well as in polycomb target genes including the developmental regulators *Hox*, *Sox* (SRY-box containing gene) and *Gata* (GATA binding protein 1), in a *Ring1A/B*-dependent manner (Cao et al., 2005; Stock et al., 2007; Wang et al., 2004a; Wei et al., 2006).

RING1A/B are also components of the E2F-6.com-1 complex that includes also non PcG proteins. This complex has been shown to repress E2F- and c-myc target genes, but no relation with H2AK119u1 has been identified so far (Weake and Workman, 2008).

Recently, proteomic analysis allowed the identification of a third complex containing RING1A and RING1B, the FBXL10-BcoR complex (Gearhart et al., 2006; Sanchez et al., 2007). In addition to RING1B this complex includes the BCL6 corepressor (BcoR), CK2α (casein kinase 2 subunit), SKP1 (component of the SCF (Skp/cullin/F-box) ubiquitin ligase complex), LSD1 (H3K4

demethylase), NSPC1 (nervous system Polycomb 1)/PCGF1 (polycomb group ring finger1) and KDM2B (FBXL10/JHDM1B-H3K36 demethylase) among others. HEK293 cells transfected with *Ring1B* and *Nspc1* or *Ring1B* and *Kdm2B* stimulated H2A ubiquitination providing evidence that the FBXL10-BcoR complex might ubiquitinate the histone H2A *in vivo* (Sanchez et al., 2007). A similar complex, also with H2A ubiquitinase activity, and found in *Drosophila* embryos was dKDM2 (homologue of mammalian KDM2B). dKDM2 associates with dRING and Psc (Posterior sex combs) in a dRING-associated factors complex (dRAF). PcG proteins Pc (Polycomb) and Ph (Polyhomeotic) are not part of this complex (Lagarou et al., 2008). In human cells, in addition to RING1B, 2A-HUB/hRUL138 can also ubiquitinate H2A (Zhou et al., 2008). 2A-HUB is recruited by the nuclear receptor co-repressor (N-CoR)/HDAC1/3 complex to a set of chemokine genes in macrophages suggesting that different E3 ubiquitin ligases can be recruited by different co-repressor complexes.

H2A deubiquitination is tightly regulated. In recent years several deubiquitinases (DUBs) have been identified: ubiquitin-specific peptidase (USP) 21, USP22, UBP-M/USP16, 2A-DUB/MYSM1 and polycomb repressive deubiquitinase (PR-DUB) (Joo et al., 2007; Nakagawa et al., 2008; Scheuermann et al., 2010; Zhao et al., 2008; Zhu et al., 2007). They have a role in activation of gene expression as shown with UBP-M/USP16 which activates *Hox* genes during development. 2A-DUB/MYSM1 also facilitates transcription through interaction with the p300/CBP-associated factor (p/CAF) HAT that leads to release of linker histone H1 and opening of chromatin facilitating transcriptional initiation and elongation (Zhu et al., 2007). Interestingly work with *Drosophila* CALYPSO showed that it interacts with the PcG protein Asx (additional sex combs) and forms the PR-DUB that is bound at PcG targets. BAP1 (BRCA1 associated protein 1) and ASXL1 (additional sex combs like 1) homologues have been found in humans, and as in *Drosophila*, specifically deubiquitinate the histone H2A. PR-DUB is needed for the

correct maintenance of gene repression as a mutation on the catalytic subunit of CALYPSO or absence of Asx activates the *Hox* genes *in vivo* (Scheuermann et al., 2010).

In contrast to H2BK120u1 that has a role in gene activation, H2AK119u1 is important for gene repression, DNA repair, and regulation of the cell cycle. Several examples have been reported where H2AK119u1 is a key player in the initiation or maintenance of gene silencing such as the *Hox* gene clusters, the *Drosophila Ubx* gene and the Xi chromosome (Cao et al., 2005; de Napoles et al., 2004; Fang et al., 2004; Wang et al., 2004a) even though the mechanism is not fully understood. One of the main silencing mechanisms attributed to PRC1 is mediated by H2AK119u1 and RNAPII elongation inhibition. The first indication of RNAPII elongation involvement came from the evidence that RNAPII and transcription factors were being recruited to bind PcG repressed promoters (Breiling et al., 2001; Dellino et al., 2004). More recently it was shown that in mESCs, bivalent genes were expressed at low levels from the 5' end where RNAPII was found stalled at the serine 5 phosphorylated form (Stock et al., 2007). Deletion of *Ring1A/B* depletes cells of H2AK119u1 and allows RNAPII to elongate (becomes phosphorylated on serine 2) and genes become active. This provides evidence of a general role of H2A monoubiquitination in repression of RNAPII elongation and in keeping PcG targets in a poised state (Stock et al., 2007). Another study in macrophages showed that H2AK119u1 (mediated by 2A-HUB) blocks RNAPII elongation by inhibiting the recruitment of FACT (Zhou et al., 2008). This finding is interesting and highlights the differences between H2A and H2B ubiquitination as H2BK120u1 appears to facilitate FACT function and stimulate transcriptional elongation (Pavri et al., 2006).

In addition to the well documented role of H2AK119u1 in gene repression, other roles are starting to emerge. It is important for DNA repair were it appears after an ultraviolet (UV)-induced lesion, mediated by the action of the nucleotide excision repair (NER) pathway

(Bergink et al., 2006; Kapetanaki et al., 2006). Several E3 ubiquitin ligases have been implicated specifically in this process, namely RING1B, DDB1-CUL4A$^{DDB2}$, DDB1-CUL4B$^{DDB2}$ and RNF8 (Bergink et al., 2006; Guerrero-Santoro et al., 2008; Kapetanaki et al., 2006; Mailand et al., 2007). In addition, H2AK119u1 levels have a role in cell cycle regulation as deubiquitination by Ubp-M is required for H3S10 phosphorylation by Aurora (Joo et al., 2007). This finding fits with previous studies that show that H2AK119u1 and H3S10 phosphorylation inversely correlate during the cell cycle, with ubiquitinated H2A being absent from metaphase chromosomes (Joo et al., 2007; Matsui et al., 1979; Muller et al., 2002; Wu et al., 1981).

#### 1.3.2.4. Histone crosstalk

Even though the correlation between histone modifications and the transcriptional state of chromatin has been demonstrated, it is not linear. Evidence has been emerging that a single histone mark is not enough to commit transcription. It is rather the "chromatin environment" that regulates the final outcome. For instance, a study carried out in human T cells shows the existence of a modification module where 17 modifications act cooperatively to prepare chromatin for transcription. Those modifications include among others H2A.Z, H3K4ac, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me1, H3K18ac (Wang et al., 2008). In addition it has become clear that cooperation occurs between different residues to change a specific chromatin environment or regulate a specific DNA process. For example, KDM6A/UTX (Ubiquitously transcribed X chromosome tetratricopeptide repeat) a di- and tri-demethylase of the JmjC family was shown to remove H3K27me3 with concomitant methylation of H3K4. UTX knockdown leads to a more repressed state of *Hox* genes with increased recruitment of PRC1 and H2AK119u1 (Lee et al., 2007b). Further, COMPASS (complex proteins associated with Set1) , the yeast homologue of mammalian MLL (mixed lineage leukemia) complex, is responsible for trimethylation of H3K4 but its activity is dependent on H2B monoubiquitination (Lee et al., 2007a). A study carried out in flies showed that the dRAF complex links histone H3K36

demethylation with H2A ubiquitination increasing the repressive state of the target genes (Lagarou et al., 2008). Another study links H3K4 and H2AK119u1. UPS21 overexpression decreases H2AK119u1 and up-regulates H3K4 di- and trimethylation at regions of transcriptional initiation, and therefore may be important to regulate transcriptional initiation (Nakagawa et al., 2008).

The histone code hypothesis was formulated in 2000 (Strahl and Allis, 2000; Turner, 2000) and states that "multiple histone modifications, acting in combinatorial or sequential fashion on one or multiple histone tails, specify unique downstream functions" (Strahl and Allis, 2000). It has subsequently been shown that histone modifications are highly dynamic and can probably vary for the same function depending on the gene location, cell stage, environmental and cellular stimuli between others.

### 1.3.3.Histone modifying enzymes

As described above, there are many histone modifying complexes. The polycomb group of repressor proteins and the trithorax group (TrxG) of activators are a major focus of this study. Both PcG and TrxG proteins can bind to the same regions regulating transcription by modulation of chromatin structure and changing the spatial accessibility of regulatory regions through histone modifications, interaction with chromatin remodelers and other enzymes (Simon and Kingston, 2009). In mouse PcG and TrxG are needed for axial patterning, cellular proliferation and differentiation (Kerppola, 2009).

#### 1.3.3.1. Polycomb group of repressor proteins

In the 1940's, genetic mutation analysis in *Drosophila* identified the first PcG proteins: extra sex combs (Esc) and Pc (Figure 1.6a). Mutations caused the appearance of sex combs on the second and the third legs of males instead of only on the first one because of mis-expression of

*Hox* genes, transforming one body segment into the other (homeotic phenotype) (Kennison, 1995; Kennison, 2004; Ringrose and Paro, 2004). Many more PcG proteins have subsequently been identified. It was not until the biochemical purification of the first proteins in 1999 that a Polycomb complex of 1-2 KDa, PRC1, was identified in *Drosophila* embryos (Shao et al., 1999). *Drosophila* and mouse PRC1 includes Polycomb (Pc), Polyhomeotic (Ph), Posterior Sex Combs (Psc) and Sex combs on midleg (Ring) (Grimaud et al., 2006b). It was later shown that PRC1 can repress chromatin through the monoubiquitination of H2AK119 (de Napoles et al., 2004; Fang et al., 2004; Wang et al., 2004a).

Later, a second approximately 600 KDa complex was purified, PRC2, containing Extra Sex Comb (Esc) and Enhancer of Zeste (E(z)) (Ng et al., 2000). It was later shown that Suppressor of Zeste (Su(z)12) was also part of core PRC2, together with other proteins which are not essential for its catalytic activity (Ringrose and Paro, 2004). PRC2 possess histone methyltransferase activity through the E(z) protein (EZH1/EZH2 homologues in mammals), being able to trimethylate histone H3 on lysine 27 (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002; Tie et al., 2001). Core members are conserved between mammals and yeast (Ringrose and Paro, 2004).The numbers of PcG homologues has expanded in mammals suggesting an increase in regulation and in functions related to those proteins (Figure1.6b).

In *Drosophila* PcG proteins bind to polycomb repressive elements (PREs). PREs range from a few hundred to about a thousand basepairs and are able to silence linked genes in a PcG dependent manner. The underlying sequence is however, not well defined (Ringrose and Paro, 2004, 2007).

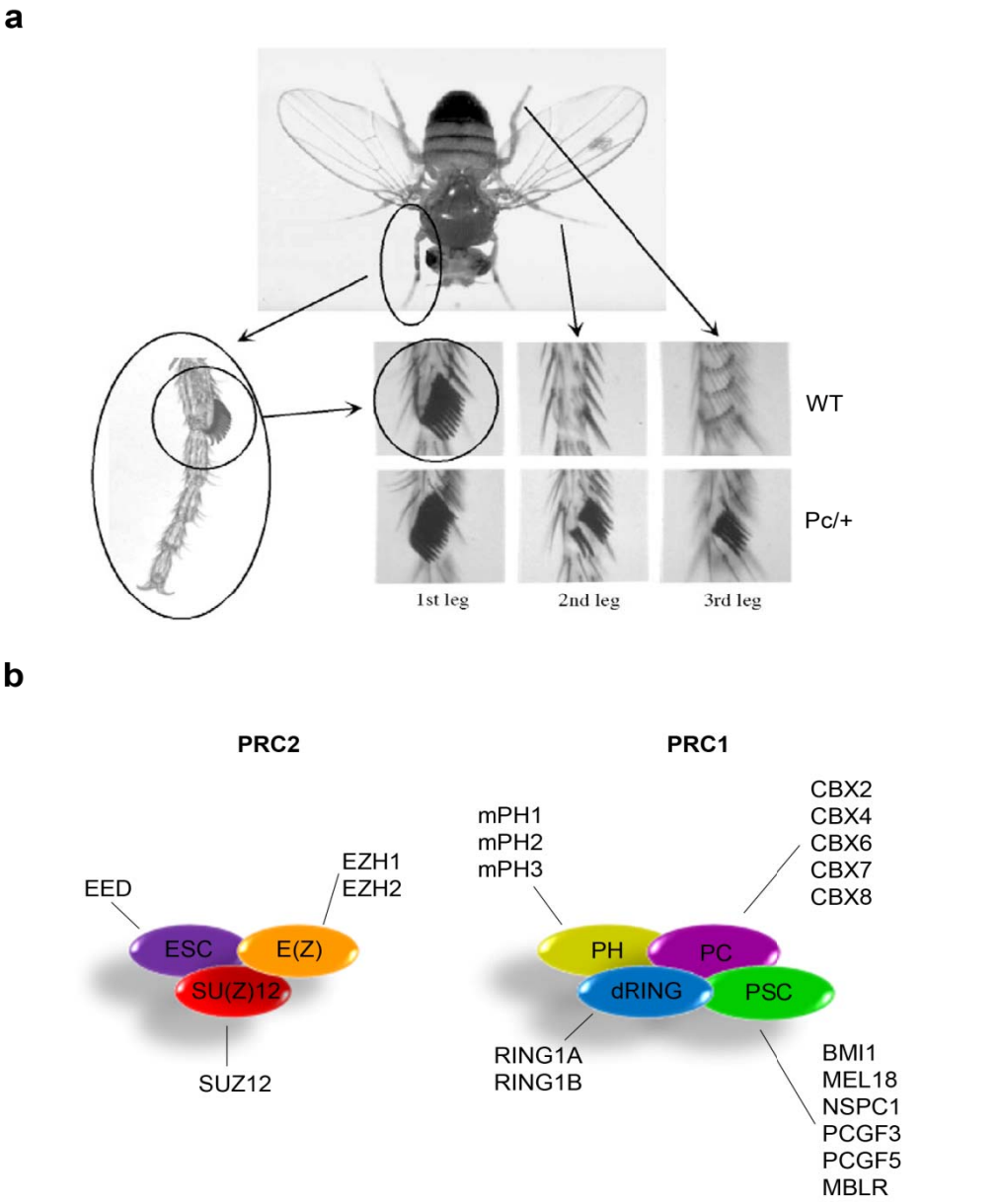



**Figure 1.6 – Polycomb proteins and mutant phenotype.** (**a**) The Scr gene specifies the identity of the adult first leg. The top figure shows an adult male fly. Magnified views of the distal leg (bottom left) and the first tarsal segments of the three pairs of legs for both WT and Pc heterozygous mutant male (Pc/+) (bottom right) are shown. The first leg of the WT fly shows a complete sex comb (10 bristles). Mutations that affect silencing of the Scr gene (as Pc mutations) cause ectopic expression of the gene and differentiation of ectopic sex combs on the second and third legs. This is known as a "sex comb phenotype" and is representative of PcG mutations. Figure adapted from Kennison (Kennison, 1995). (**b**) Schematic representation of core components of the two major PRCs, PRC2 and PRC1. Drosophila proteins are identified in the coloured ovals. Mouse homologues of each protein are listed. Figure adapted from Whitcomb et al. (Whitcomb et al., 2007).

#### 1.3.3.1.1. Polycomb repressive complex 2

As mentioned above PRC2 catalyses trimethylation of histone H3 on lysine 27. The three core mammalian components of the complex (Extra Sex Combs (EED), Suppressor of Zeste 12 (SUZ12) and Enhancer of zeste (EZH2)) are essential for enzymatic activity (Figure 1.6b) (Cao and Zhang, 2004; Ketel et al., 2005; Nekrasov et al., 2005; Pasini et al., 2004). Because of the large number of mammalian PRC2 accessory proteins there are several PRC2 variants. They can have different specificity or be expressed in different developmental stages (Simon and Kingston, 2009).

Each of PRC2 core protein may have specific functions that are beginning to be uncovered. A recent study showed the role of the EED aromatic cage in the spreading of H3K27me3 (Margueron et al., 2009). EZH1 and EZH2 are HMTs where EZH2 has a major activity in ESCs and dividing cells in which it can repress chromatin, even though is not able to compact nucleosomes (Margueron et al., 2008; Shen et al., 2008). On the other hand, EZH1 is fundamental in adult tissues and non dividing cells were it can repress chromatin in a non-catalytic manner in addition to the role as HMT (Margueron et al., 2008; Shen et al., 2008). In addition to the silencing mediated by H3K27me3 and chromatin compaction described above, PRC2 can recruit the H3K4 demethylase RBP2 (retinoblastoma binding protein 2; also known as Jarid1a) (Pasini et al., 2008) as well as DNA methyl transferases (DNMTs) (Vire et al., 2006). Recently it was also reported that it can inhibit the binding of the acetyltransferases p300 and CBP decreasing acetylation levels that are increased in *Suz12*-mutant ESCs (Pasini et al., 2010b).

#### 1.3.3.1.2. Polycomb repressive complex 1

In mammals there are 4 core PRC1 components each with various paralogues (Figure 1.6b): PH1 (polyhomeotic 1, homologue of *Drosophila* Ph), BMI1 (also known as PCGF4; or MEL18

(also known as PCGF2), homologues of *Drosophila* Psc), CBX (chromobox homolog) 4 (or CBX2, CBX7 among others, homologue of *Drosophila* Pc) and RING1B (or RING1A, homologue of *Drosophila* dRing), in addition to many other proteins (Simon and Kingston, 2009). The catalytic function of this complex is to monoubiquitinate H2AK119. RING1A/B have the crucial E3 ligase activity that can be stimulated by the Psc homologues, BMI1 and MEL18 (Cao et al., 2005; Elderkin et al., 2007; Li et al., 2006b; Wei et al., 2006). In addition, it has been shown that MEL18 is regulated by phosphorylation for the correct targeting of ubiquitination to the histone H2A (Elderkin et al., 2007).

PRC1 can silence chromatin in multiple ways although the precise mechanisms have not yet been fully elucidated (Simon and Kingston, 2009). It has been shown to be involved in X-chromosome inactivation, inhibition of ATP-chromatin remodelers and nucleosome compaction, in addition to its role in direct transcriptional repression (mediated by H2AK119u1 and discussed in section 1.3.2.3.2) (de Napoles et al., 2004; Fang et al., 2004; Francis et al., 2004; Francis et al., 2001; King et al., 2005). PRC1 knockout studies emphasize its role in gene repression, cell cycle progression and cell survival (de Napoles et al., 2004; Endoh et al., 2008; Leeb and Wutz, 2007; Stock et al., 2007) with *Bmi1* and *Ring1B* mutants showing severe proliferative defects that can be rescued by silencing the DNA damage response *Chk2* (CHK2 checkpoint homolog) or the tumour suppressor *Ink4*/*Arf* locus (Liu et al., 2009; Voncken et al., 2003).

#### 1.3.3.2. Trithorax group of activator proteins

TrxG proteins were first identified in the 1980's. Mutations in these genes led to the silencing of *Hox* genes working as suppressors of the extra sex combs phenotype (Ringrose and Paro, 2004). Trithorax (Trx) was the first protein identified by a natural mutation that causes translocation of body segments anteriorly (Ingham, 1998). It corresponds to the mammalian

homologue MLL. Later on Ash (absent, small or homeotic-like) 1 and Ash2 were also isolated and evidence emerged that are involved in *Hox* regulation in a similar way to the Trx protein (Ringrose and Paro, 2004). Meanwhile other TrxG proteins were found and were shown to form complexes. In *Drosophila* six complexes were identified (Grimaud et al., 2006b).

In *Drosophila* TrxG proteins bind to the same elements as PcG proteins (PRE/TREs (trithorax response elements)) (Ringrose, 2007).

TrxG complexes trimethylate H3K4me3 and, together with PcG activity, form bivalent domains at developmentally regulated loci (see section 1.8.2.2.1) (Azuara et al., 2006; Bernstein et al., 2006a). Their main function is so to counteract PcG repressive activity promoting open chromatin domains, through the activity of chromatin remodelers, resulting in increased accessibility for transcription factors and the polymerase complex to initiate and/or elongate through DNA (Grimaud et al., 2006b). It is under debate whether they have a wider role as general activators of gene expression (Simon and Kingston, 2009).

### 1.4. DNA methylation

DNA methylation involves the deposition of a methyl group at position 5 of the cytosine ring of CpG dinucleotides. It is an important modification despite not being conserved throughout evolution. It is not found in budding or fission yeast, nematodes or adult *Drosophila* (Zhu, 2009). In mammals it is essential for embryonic development as *Dnmt1* null mice die early in development (Li et al., 1992; Okano et al., 1999). DNA methylation is normally maintained with fidelity in the range of 95-99 %.

Initially, DNA methylation at genes was correlated with X-inactivation (Mohandas et al., 1981; Venolia et al., 1982) and silencing of imprinted genes (Li et al., 1993), which led to an

association of DNA methylation with silencing. It was further suggested that methylation would lock genes in a permanent repressed state because of the prevailing view was that it could not be actively removed (DNA demethylases had not at the time been identified) (Suzuki and Bird, 2008). New technologies allowed refinement of methyl group detection and views changed. Mammalian genomes are predominantly methylated, with all DNA regions being able to exhibit this modification (genes, including gene bodies, transposons and intergenic DNA) (Suzuki and Bird, 2008). The exception to these global methylation patterns are CGIs. These account for 1-2 % of the total genome (Bird et al., 1985; Bird, 1986; Illingworth et al., 2008).

While promoter DNA methylation is associated with gene silencing, it corresponds to a small fraction of CpG methylation with the majority being at non-promoter DNA regions. Here its role is more controversial. One study using a reporter system showed that intragenic gene body methylation reduces RNAPII capacity to transcribe through the body of a gene (Lorincz et al., 2004). However, a recent study in neural stem cells gave a different perspective. The authors describe how, cell-type specific genes are silenced by promoter methylation, mediated by DNMT3a. In addition, DNMT3a has a role on keeping neurogenesis genes active by antagonizing PcG proteins through the methylation of non-promoter DNA. *Dnmt3a* conditional knockout cells show non-promoter DNA demethylation leading to the silencing of critical developmental genes (Wu et al., 2010). A surprising study finds that active X (Xa) chromosome displays global methylation more than two times higher than in Xi. Xa chromosome enrichment is not however global: CGI promoters are found unmethylated while gene bodies become hypermethylated. From two equally methylated chromosomes, hypomethylation of promoters and hypermethylation of gene bodies is needed for the maintenance of the Xa (Hellman and Chess, 2007). In addition, a new theory for the role of DNA methylation has recently emerged: hypomethylation can lead to the increase of transpositions. Evidence from DNA methyltransferase null cells and hypomethylated cancer cells show an increase in

transpositions compared normal cells due to widespread demethylation of the genome (Suzuki and Bird, 2008).

There are two ways that promoter methylated CpGs appear to silence DNA: by directly inhibiting the binding of factors that actively promote transcription or through the recruitment of Methyl-CpG-Binding Proteins (MBPs) that will elicit a silencing cascade through the recruitment of repressors and modifications of surrounding chromatin (Klose and Bird, 2006).

### 1.4.1.CpG islands

CGIs are short CpG rich regions (0.5 to 2 kb) that lack DNA methylation. Approximately 56 % of human genes have unmethylated CpG islands promoters. They were first identified from unexpected patterns of genome digestion with methyl-sensitive and insensitive restriction enzymes. Methyl-sensitive enzyme HpaII gave rise to undigested fragments as opposed to a complete digestion of the genome by methyl-insensitive enzyme MspI. Analysis of small fragments (from the HpaII digestion) showed fragments with high guanine (G) + cytosine (C) content that were CpG rich (Bird et al., 1985; Bird, 1986). It is estimated that 19 % and 23 % of the mouse and human genomes respectively would be made up of CpG dinucleotides if distribution was by chance. In reality the percentages are much lower. This is because deamination of cytosine transforms it into uracil which is recognized as a mismatch and replaced. In contrast, deamination of 5 methyl cytosine (5meC) gives rise to thymine, making it more prone to mutation and depletion from the genome (Fazzari and Greally, 2004). Thus unmethylated CGIs are protected against mutational loss of CpGs. There is a great proportion of non-promoter CGIs but it is interesting to note that investigation of those regions has led to the discovery of new promoters. This has raised the question whether all CGIs correspond to transcription initiation regions that can be used as alternative promoters expressed during

narrow windows in development or transcribing regulatory ncRNAs (Illingworth and Bird, 2009).

How do CGIs maintain an unmethylated state? A few hypotheses can be considered (reviewed in (Illingworth and Bird, 2009)) but the most likely one is that transcription factors or proteins that bind unmethylated CpGs protect CGIs from the action of DNMTs. Evidence for this comes from studies where the binding site for the transcription factor SP1 (specificity protein 1) is mutated leading to *de novo* methylation of the *Aprt* (adenine phosphoribosyltransferase) promoter CGI (Brandeis et al., 1994; Macleod et al., 1994). More recently proteins which can recognise and bind unmethylated CpGs via their zinc finger CXXC (ZF-CXXC) domains have been proposed to maintain CGIs chromatin in a unique epigenetic state (Blackledge et al., 2010; Thomson et al., 2010). Indeed CFP1 (CXXC finger protein 1) modulates H3K4me3 levels though the recruitment of the H3K4 methyltransferase complex SET domain containing (SETD) 1 (Thomson et al., 2010) while KDM2A demethylates H3K36me1/2 (Blackledge et al., 2010) keeping CGIs as regions of high H3K4me3 and low H3K36me2 levels.

It is commonly accepted that CGIs are unmethylated, regardless of their transcriptional state (Suzuki and Bird, 2008). However, there are some cases where they became methylated during development, correlating with gene silencing. A major example is the Xi chromosome, with active methylation of promoter CGIs, but also imprinted genes, where one copy of the gene is silenced, in a sex-specific manner in the germ line (Stoger et al., 1993; Sutcliffe et al., 1994). Unexpectedly, it was shown that 3 % of mouse promoter-CGIs are hypermethylated. Genes regulated in this way include testis and oocyte specific genes (Fouse et al., 2008) but also developmental genes such *Rhox2* (reproductive homeobox) which are expressed in extraembryonic tissues (Oda et al., 2006) or weak CGI promoters as they can show *de novo* DNA methylation in somatic cells (Weber et al., 2007). An increasing number of studies

showed that differential methylation occurs in normal somatic cell and recently it was shown that developmentally regulated genes as *Pax* (Paired box) and *Hox* can have differentially methylated CGIs in different tissues (Illingworth et al., 2008).

### 1.5. Chromatin remodelling complexes

Chromatin remodelling complexes can "read" histone modifications, bind to chromatin and can induce changes in the chromatin environment or facilitate DNA processes. They are grouped into 4 classes, depending on their functional domains. In Figure 1.7 are schematized the 3 most studied ones.

SWI/SNF (switch independent/sucrose nonfermenting) in one of best studied. This family includes the *Drosophila* Brahma (*BRM*) and mammalian *BRM* and Brahma-lik*e* (*BRG1*) complexes (de la Serna et al., 2006). Each family member has a conserved ATP synthases (ATPase) subunit and additional 5 core components. They also contain a C-terminal bromodomain that specifically binds to acetylated histone tails. Some complexes can even have multiple bromodomains with specificity to certain acetylated residues in histone tails, which may target the remodelling complexes to specific nucleosomes (reviewed in (Saha et al., 2006)). Their main function is to promote gene activation through a reorganization of nucleosomes in order to allow binding of transcription factors. In specific contexts they can promote gene repression as well (Martens and Winston, 2002). One of the SWI/SNF remodelling complexes is the Brahma-associated proteins complex (BAP). BAP complexes include several proteins that are specific to different cell stages and cell types. For instance, deletion of human *BAF155* (homologue of *Drosophila* MOIRA and mouse SRG3; belongs to the esBAP complex) causes pre-implantation embryo lethality (Kim et al., 2001; Klochendler-Yeivin et al., 2000) and its knockdown prevents silencing of *Nanog* during ESC differentiation. It is therefore important in the regulation of core pluripotency transcriptional network in ESCs

(Schaniel et al., 2009). In *Drosophila* it was shown that BAP complexes and PcG complexes are mutually exclusive in salivary-gland polythene chromosomes and that PcG complexes could be important to counteract the chromatin remodelling activity and recruitment of BAP complexes to chromatin (Ho and Crabtree, 2010).

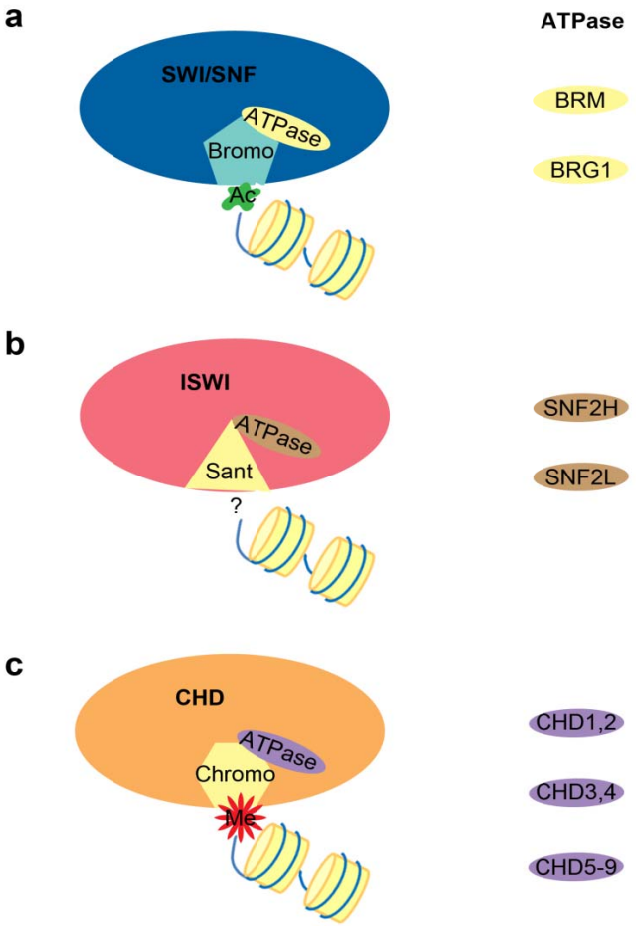



**Figure 1.7 –ATP-dependent chromatin remodelling complexes.** The three best characterized classes of ATP-dependent chromatin remodelling complexes are shown. (**a**) Each member of the SWI/SNF family has a conserved ATPase subunit (BRM and BRG1) and a C-terminal bromodomain that specifically binds to acetyl (Ac) groups on histone tails. (**b**) All the members of the ISWI family share an identical ATPase domain (SNF2H and SNF2L). The SANT domain helps the recognition of histone tails. (**c**) The CHD (NURD/Mi-2/CHD) complex includes 9 chromodomain containing members of CHD family, divided in 3 subfamilies: subfamily I (CHD1 and 2); subfamily II (CHD 3 and 4) and subfamily III (CHD5 to 9). The chromo domain helps in the recognition of the methyl (Me) groups in chromatin. Figure adapted from de la Serna and collaborators (de la Serna et al., 2006).

The ISWI (imitation switch) complex class is also one of best studied being the largest class of mammalian remodelers. All share an identical ATPase domain associated with a unique specialized protein that confers specificity to each ISWI complex: NuRF, CHRAC (chromatin-accessibility factor) and ACF (ATP-utilizing chromatin-assembly and remodelling factor) (Saha et al., 2006). They contain two domains known as SANT and SLIDE that may help ISWI remodelers recognize histone tails and linker DNA respectively. Their main function is in organizing and spacing nucleosomes after DNA replication. In mammals there are 2 core ISWI ATPases: SNF2H (also known as SMARC (SWI/SNF related, matrix associated, actin dependent regulator of chromatin) A5) and SNF2L (also known as SMARCA1). They do not show overlapping expression patterns in mice and are found in independent and functionally different complexes. The effect of knocking out *Snf2l* has not yet been studied but *Snf2h* null mice die shortly after implantation (E 5.5 to E 7.5) (Stopka and Skoultchi, 2003). In mammals SNF2H containing chromatin-remodelling complex NoRC (nucleolar chromatin remodeling complex) recruits histone-modifying enzymes and DNA methyltransferases that establish heterochromatin, leading to the silencing of ribosomal DNA. This silencing is dependent on MOF (also known as MYST1) and SIRT1 (sirtuin (silent mating type information regulation 2 homolog) 1) acetylation and deacetylation. Deacetylation enhances NoRC association with promoter associated RNA and increases heterochromatic histone marks (Zhou et al., 2009).

NuRD/Mi-2/CHD is another complex that includes 9 chromodomain containing members of CHD family, CHD1 to CHD9. Members are grouped in 3 subfamilies depending on their domains: subfamily I (CHD1 and 2); subfamily II (CHD3 and 4) and subfamily III (CHD5 to 9).

In mESCs, CHD1 depletion leads to a loss of pluripotency. It has been hypothesised that CHD1 is important in keeping chromatin in an open state, preserving plasticity characteristic of pluripotent cells (Gaspar-Maia et al., 2009). Subfamily II contains members of the NuRD

complex. These function as transcriptional repressors through the action of HDACs. Complexes are formed of a core ATPase (CHD3 or CHD4) and three main accessory subunits, MTA (metastasis-associated), MBD (methyl binding domain) and RbBP (retinoblastoma-associated-binding protein). Combinations of different accessory units confer specificity and unique characteristics to the different complexes (reviewed in (Ho and Crabtree, 2010)). CHD complexes of subfamily II are also important for the differentiation of ESCs in culture with *Mbd3* mutants failing to silence pluripotent genes (Kaji et al., 2006).

INO80 (INO80-like proteins) complexes are involved in telomere regulation, chromosome segregation, checkpoint control and DNA replication (Ho and Crabtree, 2010). Complexes in mammals including INO80 and SWRI, are characterized by the presence of a split ATPase domain. Recent studies in yeast have shown that SWR1 remodelers are responsible for the incorporation of the H2A.Z histone variant by replacing H2A-H2B dimers with H2A.Z-H2B through the use of ATP hydrolysis (Kobor et al., 2004; Krogan et al., 2003; Mizuguchi et al., 2004). Some of these complexes are dependent on acetylation of histones H4 or H2A for their function, highlighting the specificity of this process (Altaf et al., 2010). SCARP and TIP60-p400 complexes are human homologues and were shown to have a similar role in histone exchange *in vitro* and *in vivo* (Kusch et al., 2004; Ruhl et al., 2006; Wong et al., 2007). TIP60-p400 complexes were shown to be important in the maintenance of pluripotency in ESCs as their downregulation leads to premature differentiation and arrest of ESCs (Fazzio et al., 2008).

### 1.6. Non coding RNAs

Recent advances have shown that only 1.2 % of human genome encodes proteins even though the majority is transcribed. It was also shown that about 98 % of RNAs produced in human cells do not encode proteins (Mattick, 2005). The explanation for this is the existence of ncRNAs. These can be divided into different categories: short and long ncRNA (lncRNAs)

(longer than 200 nt), as well as transfer RNA (tRNAs), ribosomal RNA (rRNA) and small nuclear RNA (snRNA) (Margueron and Reinberg, 2010). Short RNAs include processed RNAs, smaller than 200 nt such as piwi interacting RNAs (piRNAs), small interfering RNAs (siRNAs), micro RNAs (miRNAs) as well as some bacterial regulatory RNAs (Brosnan and Voinnet, 2009). It was indeed shown that as well as protein-coding regions, non coding regions also have cryptic promoters that produce antisense transcripts and ncRNAs (Grewal, 2010). Many of these RNAs may not have a known function but in some cases it is known that they are expressed in response to developmental or environmental stimulus (Dutrow et al., 2008; Wilhelm et al., 2008). ncRNAs are important chromatin regulators. It has been shown that they can recruit chromatin modifying complexes as exemplified by their role in heterochromatin formation and the role of the lncRNA *Xist* in the initiation of X chromosome inactivation (Grewal and Elgin, 2007; Senner and Brockdorff, 2009).

### 1.6.1.*Xist* (X inactivation specific transcript)

*Xist* is expressed in female cells in a developmentally regulated way to silence one of the X-chromosomes, a mechanism that compensates for the gene dose between males and females. *Xist* expression from the future Xi is the first event of X-inactivation. It is quickly followed by the enrichment of repressive histone PTMs such as H3K27me3, H3K9me2, H3K9me3, H4K20me1, and H2AK119u1 as well as incorporation of the histone variant macroH2A and recruitment of repressive complexes such as PcG proteins. On the other hand, marks commonly associated with gene activation (acetylated lysines and H3K4me2/3 and RNAPII) are erased (Chow and Heard, 2009). It is still arguable whether *Xist* is the major player in the active recruitment of repressive complexes/modifications but it seems clear that it has a major role. Recent work shows that integration of *Xist* in an autosome can lead to *Xist* coating of the autosome and gene silencing (Tang et al., 2010).

Besides *Xist*, there are an increasing number of lncRNAs involved in gene silencing. *Kcnqot1* (Pandey et al., 2008) and *Air* (Nagano et al., 2008) are two lncRNAs that interact with KMT1/G9a (also known as EHMT2 - euchromatic histone-lysine N-methyltransferase 2) to maintain monoallelic gene silencing within imprinted gene clusters. Disruption of these lncRNAs leads to bi-allelic expression (Mancini-Dinardo et al., 2006; Sleutels et al., 2002).

### 1.6.2.Short ncRNAs

Short ncRNAs are also important for gene regulation. Small RNAs in eukaryotes are products from the processing of longer precursors and function as sequence-specific guides for AGO (Argonaute) proteins. Dicer (RNAse III endoribonuclease) processes RNA hairpins producing miRNAs, double stranded small RNAs, capable of being incorporated into RISC (RNA-induced silencing complex) which in turn elicits transcriptional silencing or messenger RNA (mRNA) cleavage. siRNAs are also generated by Dicer and target complementary mRNAs for degradation (reviewed in (Matzke and Birchler, 2005)). Several examples are starting to emerge showing the importance of small RNAs in development. In fact, studies carried out with Dicer mutant mice show that it is crucial for development as mice die at E 7.5 (Bernstein et al., 2003; Murchison et al., 2005). Dicer conditional knockout ESCs were generated and show impaired proliferation perhaps due to an up-regulation of normally silenced genes. Cell differentiation is also compromised (Kanellopoulou et al., 2005), even though initiation and spreading of lncRNA *Xist* occurs normally (Kanellopoulou et al., 2009; Nesterova et al., 2008).

## 1.7. Heritability of chromatin landscapes

The heritability of chromatin landscapes became apparent when it was realized that different cells had different gene expression programmes active at different times, despite sharing the same DNA sequence, and that those programmes were transmitted from mother to daughter

cell through division. The epigenetic factors that perpetuate expression patterns fall into two main mechanisms, epialleles and chromatin modifications.

### 1.7.1. Mitotic heritability

From yeast to mammals, cells show evidence of mitotic heritability of gene expression patterns (reviewed in (Margueron and Reinberg, 2010)). Gene expression profiles are maintained from cell to cell through division. It is not known how epigenetic information is maintained through DNA replication but 2 models have been proposed, the random and the conservative model.

In the random model new and old histones (old histones carrying original PTMs) are assembled randomly and epigenetic marks are diluted. This model stresses the importance of chromatin domain inheritance. As PTMs get diluted, it is only when the marks are enriched in various nucleosomes that they can be efficiently transmitted as, despite the dilution, they still mark a chromatin domain. In that case histone modifying machinery can be recruited to spread modifications using neighbouring histones as a template (Margueron and Reinberg, 2010; Probst et al., 2009). On the other hand the conservative model suggests that histones are equally distributed between both strands. This means that the hemi-modified nucleosomes (old H3-H4 dimer coupled with new one) or adjacent nucleosomes (if deposition happens as tetramers) provide a template for the modifications to happen in the newly deposited H3-H4 dimers or tetramers. In that case modifications are accurately transmitted (Margueron and Reinberg, 2010).

While there is not a consensus on which model is correct, some research has been carried out to try to shed some light onto the mechanisms and proteins involved. For example, PCNA (proliferating cell nuclear antigen) has been implicated in the heritability of histone modifications during replication. Evidence suggests that DNMT1 interacts with PCNA (Chuang

et al., 1997) and NP95, (nuclear protein 95, also known as UHRF1 - ubiquitin-like with PHD and ring finger domains 1) (Bostick et al., 2007; Sharif et al., 2007). That interaction could be important for DNA methylation inheritance as deletion of NP95 causes global loss of DNA methylation in ESCs (Sharif et al., 2007). Other histone modifying proteins known to associate with PCNA are G9A and SETDB1, both catalyse H3K9 methylation, and PRSET7 (also known as SETD8) that catalyses H4K20me1 (Margueron and Reinberg, 2010). H3.3 may also have a role in epigenetic memory as shown by experiments where endoderm donor cells were transplanted into a *Xenopus* enucleated egg. The maintenance of the active state of the *MyoD* (myogenic differentiation) gene was due to the presence of H3.3 variant histone on its promoter. The authors suggest that H3.3 can probably carry active marks such as H3K4me as mutations on lysine 4 of H3.3 fail to maintain the active memory of the *MyoD* gene (Ng and Gurdon, 2008). Examples of other PTMs involved in epigenetic inheritance are starting to emerge but it is still arguable whether they are indeed kept through replication or established *de novo* after the 2 cells are formed (Campos and Reinberg, 2009; Kerppola, 2009). Recent work has also implicated PRC1 and PRC2 complexes in heritability. H3K27me3 was shown to be maintained in proliferating cells. In fact, EZH2 colocalizes with H3K27me3 in G1 and at sites of ongoing DNA replication which helps to maintain the epigenetic mark in the newly synthesized nucleosomes (Hansen et al., 2008). In addition to PRC2, *in vitro* work showed that PRC1 bound chromatin can replicate keeping the complexes bound to DNA (Francis et al., 2009). A recent study has helped to understand the maintenance of H3K27me3 domains by providing evidence that the CTD of EED can recognize H3K27me3 leading to the activation of the methyltransferase activity of PRC2 and the spread of H3K27me3 across wide domains preserving the normal pattern. Mutations that abolish recognition of methylated residues led to a global reduction of H3K27me3 and disruption of normal *Drosophila* development (Margueron et al., 2009).

As discussed above, DNA methylation is highly reproduced after replication through activity of the maintenance methyltransferase, DNMT1 (Klose and Bird, 2006).

### 1.7.2.Meiotic heritability

Most epigenetic marks are erased in the germ line although some can be transmitted between individuals. As an example, individuals can inherit methylation of imprinted genes from parental germ cells. Another example comes from intracisternal A particles (IAPs) insertions. IAPs can insert in several places in the genome and control the expression of nearby genes. Moreover, this expression is dependent on the epigenetic status of the IAP (methylated versus unmethylated) that resists reprogramming during primordial germ cells and pre-implantation development, thus its status is inherited across generations (Reik, 2007). Evidence of meiotic heritability also comes from epigenetic alleles (epialleles) that are transmitted through generations. An example is the human *MLH1* epiallele where DNA hypermethylation of coding sequence is associated with silencing leading to a predisposition for tumour formation (Richards, 2006). However, the mechanism by which the epigenetic state of a gene can be transmitted through generations is far from understood.

## 1.8. Epigenetic gene regulation in mammalian development

### 1.8.1.Zygote development

In mammals, the fusion of gametes (oocyte and sperm) forms the zygote. The zygote is a pluripotent cell that will divide and give rise to a new organism. The zygote and germ cells undergo extensive epigenetic reprogramming to erase epigenetic marks (Figure 1.8) (with a few exceptions reviewed by Sasaki and Matsui (Sasaki and Matsui, 2008)).

After fertilization, reprogramming of the paternal and maternal genomes takes place with the zygote undergoing several dramatic changes to become totipotent. The paternal genome goes

from a highly repressed state to a more open configuration due to the replacement of protamines by acetylated histones and active DNA demethylation takes place (Oswald et al., 2000; Reik et al., 2001). After this, both maternal and paternal genomes undergo passive demethylation by exclusion of Dnmt1 from the nucleus (Howell et al., 2001). This process leads to erasure of all DNA methylation with the exception of the IAP retrotransposons and imprinted loci. This mechanism may involve PGC7/STELLA (also known as DPPA3 - developmental pluripotency associated 3) that binds to some imprinted genes and IAPs, protecting them from demethylation after fertilization (Nakamura et al., 2007). Even though it is accepted that the first demethylation event in the zygote is an active process, the underlying mechanism remain unknown. Recently, activation induced cytidine deaminase (AID or AICD) was found to deaminate 5meC *in vitro* and *in vivo* (Bhutani et al., 2010), and together with its family member APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) is co-expressed with STELLA and NANOG in the oocyte, ESCs and embryonic germ cells (Morgan et al., 2004). TET proteins may also have a role in demethylation as TET (tet oncogene family member) 1 protein could catalyze demethylation of 5meC (Tahiliani et al., 2009). A further indication that TET proteins (TET1, TET2 and TET3) are involved in demethylation in the pre-implantation embryo comes from *Tet1* knockdown experiments where differentiation towards the trophoblast lineage was observed (Ito et al., 2010).

After demethylation of the zygotic genome, de novo methylation begins in the inner cell mass (ICM) of the blastocyst (Figure 1.8). This confers a unique pattern in the blastocyst that distinguishes the methylated primitive ectoderm (which will give rise to the embryo), from the unmethylated trophoblast and primitive endoderm (Li, 2002). For this methylation to occur, *de novo* methyltransferases *Dnmt3a* and *Dnmt3b* are crucial as their inactivation blocks *de novo* methylation and causes embryonic lethality (Okano et al., 1999). *Dnmt1* is also important

because once methylation patterns are established they need to be maintained (*Dnmt1* null mice die in late gastrulation, at E 9.5).

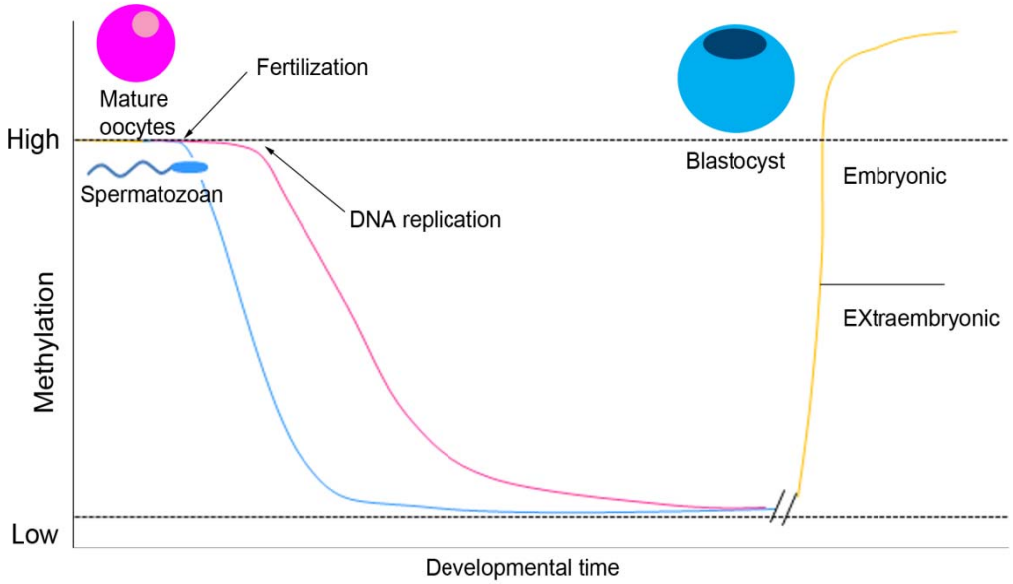



**Figure 1.8 – Methylation across zygote development.** In mammals, the fusion of gametes (oocyte and sperm) forms the zygote. The first change in the level of methylation is the fast demethylation of the paternal genome (blue) followed by both maternal (pink) and paternal passive demethylation by exclusion of DNMT1 from the nucleus (at the time of DNA replication). After demethylation of the zygotic genome, *de novo* methylation begins in the inner cell mass of the blastocyst. This confers a unique pattern in the blastocyst that distinguishes the embryonic from the unmethylated extraembryonic tissues. Figure adapted from Reik et al. (Reik et al., 2001).

Waves of demethylation and *de novo* methylation are probably needed for the erasure of epigenetic marks from the gametes to prevent aberrant epigenetic modifications (Santos et al., 2002). Demethylation will also lead to proper activation of the pluripotency genes *Nanog* and *Oct4* needed for the establishment and maintenance of the ICM of the blastocyst. Their promoters are methylated in sperm so demethylation is necessary for their proper expression (Smith, 2001).

### 1.8.2.First steps: capturing the pluripotent state

The ICM of the embryo gives rise to all the cells of the organism, with the exception of the extraembryonic trophoblast lineage (Rossant, 2008). In these cells chromatin is in an open configuration (Ahmed et al., 2010). ESCs can be derived from the ICM that retain the pluripotent state and can be kept in culture retaining self-renewal and differentiation capacity (Yu and Thomson, 2008). They express high levels of pluripotency genes *Oct4*, *Nanog* and *Sox2* that, together with chromatin remodelers and ncRNAs, establish a regulatory network that retains the pluripotent state (Chambers and Tomlinson, 2009). ESCs have been an important tool for understanding epigenetic processes and differentiation pathways (Murry and Keller, 2008).

#### 1.8.2.1. Pluripotency markers

Expression of the pluripotent markers *Oct4*, *Nanog* and *Sox2* is essential for maintaining the pluripotent state (Silva and Smith, 2008).

*Oct4*, also known as *Pou5f1*, is part of the POU family of transcription factors that interact with DNA through two DNA binding domains (Chambers and Tomlinson, 2009). Conditional deletion of *Oct4* confirmed its role as a master regulator of pluripotency as deletion triggered cells to differentiate into trophectoderm (the only lineage that ESCs do not differentiate into in normal development (Rossant, 2008)) (Niwa et al., 2000).

NANOG protein binds to DNA through a singe homeodomain (Chambers and Tomlinson, 2009). It was identified in 2003 to be one of the key players in ESC self-renewal (Chambers et al., 2003; Mitsui et al., 2003). Interestingly, later studies showed that although reduction of *Nanog* levels increases the predisposition to differentiate, *Nanog* null ESCs can be kept undifferentiated (Chambers et al., 2007).

The last pluripotency marker is *Sox2* that belongs to a superfamily of transcription factors characterized by the presence of a high mobility group (HMG) box DNA-binding domain (Chambers and Tomlinson, 2009).

SOX2 and OCT4 bind at the same genes as their binding regions often occur adjacent to each other. *Sox2* deletion also causes differentiation towards trophectoderm (Masui et al., 2007) and evidence points to a role for *Sox2* in maintaining *Oct4* expression, as overexpression of OCT4 in *Sox2* null cells can rescue ESCs from differentiation (Masui et al., 2007).

Several recent studies have attempted to elucidate the complex network of transcription factors in ESCs and its role in activation/repression of gene expression, but firm conclusions cannot yet be drawn (Chen et al., 2008; Endoh et al., 2008; Kim et al., 2008). It is a complex pathway where binding of activators and repressors can have different outcomes depending on which factors are bound and the chromatin environment. Nevertheless it is widely accepted that pluripotency factors are important for transcriptional regulation and the maintenance of chromatin plasticity to keep the undifferentiated state of an ESC (Chambers and Tomlinson, 2009; Ng et al., 2008).

#### 1.8.2.2. Chromatin plasticity in ES cells

In addition to a core network of pluripotency factors, ESCs have a unique chromatin signature (Figure 1.9). Several factors have a pivotal role in maintaining the chromatin in a highly dynamic state that allows the cell to maintain its pluripotency and to differentiate whenever the appropriate stimulus is received.

The main players in chromatin plasticity are the pluripotency factors (referred in the previous section). They regulate transcription maintaining activity of genes needed for pluripotency and inactivating differentiation associated genes (Chambers and Tomlinson, 2009).

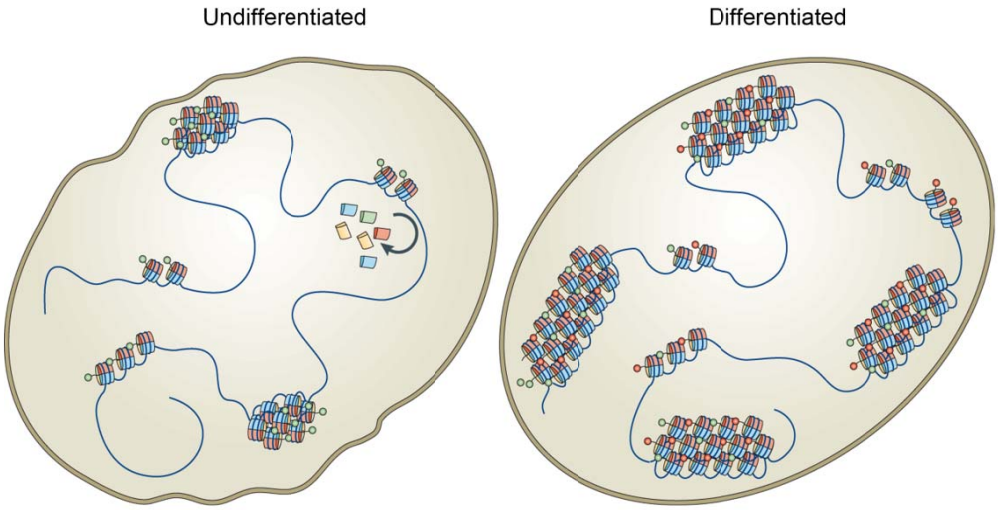



**Figure 1.9 – Chromatin during ESC differentiation.** ESCs chromatin has unique features being characterized by general decondensation with high levels of active marks and proteins highly dynamic in their binding to chromatin. As cells differentiate (right), regions of heterochromatin are formed, silence marks accumulate and proteins became more stably associated to chromatin (Meshorer and Misteli, 2006).

The second players are the ATP-dependent chromatin remodelers. As discussed in section 1.5, several studies showed that CDH1, NuRD, Tip60-p400 and esBAF (embryonic stem BAF) are important in the maintenance of pluripotency. Whether their function is independent of one another is under debate (Ho and Crabtree, 2010). esBAF binds to about a quarter of promoters in mESCs, correlating with gene expression. Some of the promoters bound by esBAF complex include pluripotent genes such as *Oct4*, *Nanog*, *Sox2*, *Sal4*, *REST* and PcG (Kidder et al., 2009). Knockout studies with BAF250a demonstrated a failure to form proper germ-layers at gastrulation (Gao et al., 2008) and RNAi-mediated knockdown of *Brg1* (core ATPase domain) causes pluripotency genes to be downregulated, reinforcing the role of SWI/SNF remodelers in maintaining pluripotency (Ho et al., 2009; Kidder et al., 2009). Tip60-p400 is also important for

the maintenance of mESCs as it was shown to bind promoters enriched in H3K4me3, whether they are silent or active in ESCs. It integrates *Nanog* and H3K4me3 signals to regulate pluripotency, keeping developmental genes silenced (Fazzio et al., 2008). CDH1 was shown recently to be needed for the maintenance of the highly dynamic euchromatin in mESCs. Reduction of CHD1 levels reduces pluripotency by increasing heterochromatin. In this case impairment in pluripotency occurs due to changes in chromatin configuration, rather than changes on histone methylation levels, which are kept the same (Gaspar-Maia et al., 2009).

PTMs are the third factor important for chromatin plasticity. Chromatin in mESCs is hyperacetylated, consistent with an open configuration and high transcription levels (both protein coding and ncRNA) (Efroni et al., 2008). Even developmental genes which are important once cells start to differentiate are expressed at low levels (Efroni et al., 2008; Stock et al., 2007). Also *Setdb1*, an H3K9 methyltransferase was recently shown to be important for ESC plasticity because its loss by knockdown causes destabilization of the ESC state. SETDB1 binds to a subset of developmental regulated genes and has a role in keeping them repressed (Bilodeau et al., 2009).

DNA methylation is another player in mESC chromatin plasticity. The majority of promoter-CGIs are unmethylated correlating with the poised state of the genes. This may be why *Dnmt* knockouts do not affect the pluripotency and plasticity of stem cells (Tsumura et al., 2006). DNA methylation is nevertheless required for differentiation as triple *Dnmt* null ESCs fail to differentiate (Tsumura et al., 2006).

Finally, ncRNAs may be important for the maintenance of pluripotency on ESCs (Dinger et al., 2008). However, cells keep their self-renewal capacity when they lack Dicer (Murchison et al., 2005). That may be because enough small RNAs are still produced to keep genome

architecture, or because their major role is in cell differentiation (Bernstein et al., 2003; Murchison et al., 2005).

#### 1.8.2.2.1. Bivalent Domains in ES cells

In 2006 two independent studies reported the existence of a subset of genes in mESCs with an unusual combination of histone modifications at their promoters (Azuara et al., 2006; Bernstein et al., 2006a). H3K4me3 is catalysed by TrxG proteins and is commonly associated with acetylation of histones and both modifications are usually present at active regions of the genome. On the other hand H3K27me3 is commonly associated with repressed regions with absent or low transcriptional levels. Surprisingly both H3K4me3 and H3K27me3 are found at the same locations at developmentally regulated promoters in mESCs. The presence of bivalent domains (Azuara et al., 2006; Bernstein et al., 2006a), together with the early replication of these genes (Azuara et al., 2006) (commonly associated with active genes) led to a hypothesis that those genes were repressed but poised for transcription and that the bivalent state would be resolved as genes become activated or repressed depending on the lineage commitment in differentiation (Azuara et al., 2006; Bernstein et al., 2006a). Further studies led to a better understanding of those domains and comparative studies with adult stem cells, mouse embryonic fibroblasts (MEFs), T-cells and lung fibroblasts provided evidence that bivalent domains are not a unique feature of ESCs as they can be maintained in differentiated cells (Barski et al., 2007; Cui et al., 2009; Mikkelsen et al., 2007; Pan et al., 2007). This raises the question of whether “bivalency” is a feature of cells in general and whether the extent of “bivalency” correlates with their capacity to reprogram and give rise to other cell types.

Another question remains to be elucidated: how are bivalent domains formed? H3K4me3 in mESCs coincides with CGIs (Bernstein et al., 2006a). This can be explained by the existence, in

TrxG proteins, of domains that bind unmethylated CGIs (Birke et al., 2002; Lee and Skalnik, 2005). H3K27me3 was also shown to correlate to CGIs buts does not have such a linear distribution (Bernstein et al., 2007). It is also not known how bivalent domains resolve in differentiation. Recent work shed some light on this by showing a role for H3K9me3 and DNA methylation in resolving the bivalent chromatin of key developmental regulators as *Sox*, *Hox* and *Gata* upon trophoblast lineage commitment (Alder et al., 2010). Such promoters are targeted with H3K9me3 and DNA methylation in the trophoblast lineage suggesting an alternative repressive pathway. In pluripotent lineages they still retain RING1B.

An important feature of bivalent domains is basal transcription. RNAPII is stalled at the 5' end of developmental genes producing short transcripts (Bernstein et al., 2006a; Stock et al., 2007). Whether this transcription is a cause or consequence of this unique chromatin landscape is not known. Nevertheless, the resulting stochastic flux in gene expression or transcriptional noise allows the cell to respond quickly to a variety of developmental cues and may be important for differentiation.

### 1.8.3.Growing into adulthood

Several extrinsic and intrinsic stimuli can trigger ESC differentiation. The primary events after those signals are received by the cell are not fully known, but one of the first changes is the downregulation of the pluripotency genes *Oct4*, *Nanog* and *Sox2* (Chambers and Tomlinson, 2009). As cells differentiate from pluripotent ESCs to lineage committed cells, major changes occur. The chromatin landscape changes from an open, highly dynamic, decondensed structure, enriched in active PTMs and highly transcribed to a more closed, stable and condensed chromatin structure that accumulates silent marks and reduced transcription (Figure 1.9) (Meshorer and Misteli, 2006).

Differentiation from ESCs is accompanied by an increase in repressive marks such as H3K9me3 and a decrease in active marks such as acetylated H3 and H4 (Meshorer and Misteli, 2006). The relevance of acetylation in differentiation is highlighted by the fact that Trichostatin A (TSA) treated cells cannot differentiate as they stay highly acetylated (Lee et al., 2004). This occurs possibly due to an inability to repress transcription. Numerous other factors are important for differentiation. *Dnmts*, Dicer, chromatin remodelers and histone modifying proteins all have an extremely important role in differentiation as cells that lack those proteins usually fail to differentiate *in vitro* and knockout embryos die during development (Bernstein et al., 2003; Faust et al., 1998; Kaji et al., 2006; O'Carroll et al., 2001; Okano et al., 1999; Pasini et al., 2007; Pasini et al., 2004; Tsumura et al., 2006)

### 1.8.4.Cancer

It is essential that cells maintain their epigenetic profile to avoid aberrant gene expression programmes detrimental to cell identity. Misregulation of the epigenetic profile can give rise to mutant cells that in turn can lead to diseases such as cancer.

Changes in DNA methylation are commonly associated with cancer. The first clue arose in 1983 when global genome hypomethylation was detected in cancer cells relative to their normal cells (Robertson, 2005). Loss of methylation occurs initially at repetitive sequences causing genome instability with an increase in transpositions and genome rearrangements. It is thought that active demethylation happens early in cancer ontogeny and thus represents a key factor in the severity and metastatic potential of a cancer. In contrast to the global demethylation of the genome, some regions become hypermethylated. This is the case for promoter CGIs (generally unmethylated in normal cells) leading to gene silencing of many tumour suppressor genes. Genes involved in cell-cycle regulation, DNA repair, chromatin remodelling, transcription regulation, cell adhesion and apoptosis are also commonly affected

(Costello et al., 2000). These alterations transform normal cells into cells with increased dividing ability, a failure to respond to checkpoints or apoptotic signals, and consequently, greater metastatic potential.

Two genes consistently misregulated in cancer are *Ink4a* (also known as *Cdkn2a*) and *Arf* (also known as *Cdkn2aip*). When activated they trigger senescence by arresting the cell cycle in G1-S. They work as sensors for the correct regulation of the cell, being kept under tight control. Their link with cancer emerged many years ago when they were described as tumour suppressor genes required for successful progression through the mitotic checkpoint. These genes are suppressed in many tumours, by promoter hypermethylation, through suppression of its activators (MYC, RAS), gene coding mutations or up-regulation of its inhibitors, namely *Ring1B* and *Mel18* (Gil and Peters, 2006).

### 1.8.5. Aims of this study

Determining how PcG proteins are targeted to defined loci is of major importance to understanding mouse development, regulation of pluripotency, and misregulation of gene expression in tumurogenesis. The aims of this study are as follows:

- To analsyse the *INK4b-ARF-INK4a* locus with the aim of identifying which regions are essential for PcG recruitment.
- To develop an *Ink4b-Arf-Ink4a* reporter cell system for use in a genome wide small interference (siRNA) screen to unveil novel regulators of PcG recruitment.
- To analyse how PcG proteins are recruited to target genes in ESCs and to understand the basis for co-occupancy of PRC1 and PRC2 targets.

# 2. Polycomb recruitment to the *Ink4b-Arf-Ink4a* locus

## 2.1. Introduction

### 2.1.1. PcG recruitment to target genes

Since PcG proteins were identified in mammals (Levine et al., 2002), determining how they are recruited to chromatin has been a major challenge. Several hypotheses have been put forward. First, sequence specific recruitment was considered, but following identification of target loci in a number of studies, no consensus sequences has been found (Ringrose and Paro, 2007). This has led to the consideration that recruitment may be mediated by other proteins that recognize DNA sequences and silence gene expression. Recently some candidates have emerged from different studies. YY1 (Yin Yang-1) is a possible candidate as it shares some similarities with *Drosophila,* Pho but a direct role of YY1 in PcG targeting in vertebrates has not been established (Kerppola, 2009). Another candidate for PcG recruitment and binding to YY1 is RYBP (Ring1 YY1 binding protein). RYBP was found to interact with RING1A/B and Pc proteins and mediate gene silencing (Garcia et al., 1999; Woo et al., 2010). Recently, crystal structure studies revealed that Pc and RYBP can bind the same C-terminal domain of RING1B. The authors suggest that the targeting of RING1B to chromatin may depend on which protein associates to its C-terminal domain (Wang et al., 2010).

RING1A/B also co-purified with the BcoR complex containing PCGF1/NSPC1 and FBXL10/JHDM1B (Sanchez et al., 2007). This complex may work in the recruitment of PcG as its overexpression increases PcG and H3K27me3 at the *Ink4*/*Arf* locus (Tzatsos et al., 2009).

JARID2 was recently identified as a PRC2 component (Landeira et al., 2010; Li et al., 2010b; Pasini et al., 2010a; Peng et al., 2009; Shen et al., 2009). It was shown to be important for PRC2 recruitment as enrichment of core PRC2 components to targets is diminished in *Jarid2* knockdown and knockout cells. *Jarid2* knockout mESCs show impaired differentiation, a phenotype shared with other PRC2 knockouts (Landeira et al., 2010; Li et al., 2010b; Pasini et al., 2010a; Peng et al., 2009; Shen et al., 2009).

AEBP2 (AE binding protein 2) is a zinc finger protein that initial studies identified as a repressor protein (He et al., 1999). It was found both in *Drosophila* and mammals and it was shown to interact with PcG proteins (Culi et al., 2006; McClure and Schubiger, 2008). It was later co-purified with PRC2 having been demonstrated that interaction between AEBP2 and different PRC2 members enhances the catalytic activity of the complex (Cao and Zhang, 2004).

In addition to DNA binding proteins, evidence for ncRNAs playing a role in PcG recruitment has started to emerge. The first evidence of lncRNA recruiting PcG proteins in *cis* came from studies with *Xist* RNA. Initial studies showed a co-localization of PcG proteins with *Xist* (Mak et al., 2002; Plath et al., 2003; Silva et al., 2003). Further evidence has suggested a link between *Xist* and PRC2 recruitment (Maenner et al., 2010). Recently another lncRNA, *ANRIL*, has been found to mediate *cis*-silencing of the *Ink4*/*Arf* locus by mediating the recruitment of both PRC2 and PRC1 proteins to the locus (Yap et al., 2010). In addition to lncRNAs a recent study suggested that short ncRNAs transcribed from PcG target loci may regulate *cis*-recruitment or maintenance of PRC2 at those targets loci (Kanhere et al., 2010). RNAs may also mediate PcG recruitment in *trans*. Evidence came from a study in which more than 200 ncRNAs were identified at the *Hox* loci and it was shown that one of these, *HOTAIR*, interacts with PRC2 and is required for PRC2 occupancy and H3K27me3 of the *Hoxd* cluster (Rinn et al., 2007).

### 2.1.2. Regulation of the *Ink4b-Arf-Ink4a* locus regulation

The *Ink4b-Arf-Ink4a* locus was found to have a major role in senescence due to its role in cell cycle progression (Serrano et al., 1993). Senescence was first described as the "Hayflick phenomenon", referring to the cell cycle arrest observed in cultured cells after a limited number of divisions (Hayflick, 1961), but which can also occur *in vivo* (Krishnamurthy et al., 2004). Senescence enables cells to remain intact and metabolically viable although incapable of further divisions.

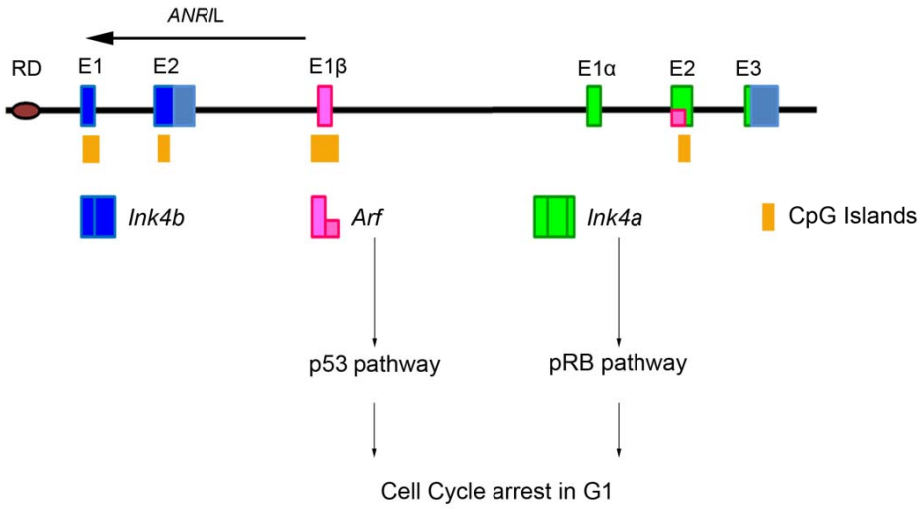



**Figure 2.1 – Schematic representation of the human/mouse *Ink4b-Arf-Ink4a* locus shown (3' to 5').** The genomic locus is represented by a bold line with exons in coloured boxes (not to scale). Exon numbers are represented above coloured boxes. *Ink4b* coding regions are represented in dark blue, those of *Arf* in purple and *Ink4a* coding regions represented in green. Untranslated regions are shown in light blue. CpG islands are represented as oranges boxes. The *cis*-acting regulatory domain (RD) and *ANRIL* transcript are shown. The ncRNA *ANRIL* was found only in humans and was shown to be transcribed in opposite direction to the *Ink4b*/*Arf*/*Ink4a* genes, overlaping with the two *Ink4b* exons. RD can regulate expression from the entire locus. Schematic of *Ink4a* and *Arf* downstream pathways are shown.

The *Ink4/Arf* locus is located on mouse chromosome 4 and comprises three different, although related, genes: *Ink4b* (also known as *cdkn - cyclin-dependent kinase inhibitor 2b*), *Arf* (also known as *Cdkn2aip – cdkn2a interacting protein*) and *Ink4a* (also known as *cdkn2a*) (Figure

2.1). *Ink4b* encodes the $p15^{Ink4b}$ protein and when this gene is disrupted in mice (*Ink4b*-null mice) development is normal, although mice show a predisposition to some specific tumors (Herman et al., 1996; Latres et al., 2000). *Arf* encodes the tumor suppressor protein $p19^{Arf}$. Its coding sequence partially overlaps with *Ink4a*, although it has an Alternative Reading Frame, from which *Arf* takes its name (Quelle et al., 1995). *Arf*-null mice develop a wide variety of cancers in the first year of life. MEFs derived from these embryos are able to grow indefinitely in culture without undergoing senescence (reviewed by (Sherr, 2001)). *Ink4a* encodes a 16 KDa protein, $p16^{Ink4a}$, which has an important role as a tumor suppressor. *Ink4a*-null mice do not have such an extreme phenotype as *Arf*-null mice although the incidence of tumors is also increased (Krimpenfort et al., 2001; Sharpless et al., 2001).

The pathways that these proteins are involved in have now been elucidated. $p15^{Ink4b}$ and $p16^{Ink4a}$ expression triggers a cascade of events that leads to dephosphorylation of retinoblastoma (pRB) protein. When pRB is unphosphorylated it binds to the transcription factor E2F, inhibiting cell cycle progression and inducing G1 cell cycle arrest (reviewed by (Gil and Peters, 2006)). On the other hand, when $p19^{Arf}$ is expressed it binds to MDM2, blocking its ability to poli-ubiquitinate p53 increasing p53 stability. Genes activated by p53 include the cyclin dependent kinase inhibitor, $p21^{CIP1}$, that mediates cell cycle arrest in G1 and G2 phases (Gil and Peters, 2006; Sherr, 2001).

#### 2.1.2.1. *Ink4b-Arf-Ink4a* expression regulated by PcG

It has been shown that the *Ink4/Arf* locus is regulated through a number of mechanisms, with a tight and complex network of repressors and activators influencing its expression. Stress factors and oncogenic signals can induce or repress the locus differently (Gil and Peters, 2006). PcG proteins have also been shown to have a major role in its regulation (Gil and Peters, 2006).

*Bmi1* was the first PcG protein implicated in the regulation of the *Ink4/Arf* locus. It was found that *Bmi1*$^{-/-}$ mice develop neurologic and haematological disorders due to defects in cell proliferation (van der Lugt et al., 1994). *Bmi1* was subsequently found to repress all 3 genes in the *Ink4b/Arf*/*Ink4a* locus. It was also shown that *Bmi1* deficient primary MEFs are impaired in progression into S-phase of the cell cycle and undergo premature senescence which correlates with an increased expression of p16$^{Ink4a}$ and p19$^{Arf}$. When *Bmi1* overexpression is induced, the cells become immortalized, *Ink4a* and *Arf* expression is downregulated and, in combination with *H-RAS* expression, neoplastic transformation is triggered (Jacobs et al., 1999). It has also been shown that *Bmi1* promotes self-renewal and persistence, through the repression of *Ink4a*, of diverse types of postnatal stem cells (Molofsky et al., 2003; Park et al., 2003).

Studies carried out in murine knockouts for polycomb proteins, *Cbx2*, *Cbx7*, *Cbx8*, *Mel18*, *Phc2* and *Ring1B* have all shown defects in haematopoiesis or in MEF proliferation. These defects were reported as being, at least in part, due to *Ink4/Arf* derepression (Dietrich et al., 2007; Gil and Peters, 2006; Miki et al., 2007).

*Ink4b-Arf-Ink4a* is a well studied locus. In addition to its regulation by PcG proteins, many other activators/repressors are known which can easily be manipulated *in vitro*. Thus the *Ink4b-Arf-Ink4a* locus was chosen as a model to study PcG recruitment. In this chapter PcG interactions with the locus will be characterised, both in ESCs and MEFs, with the aim of shedding light on PcG recruitment mechanisms and developing a strategy for a siRNA screen to find novel PcG recruiting proteins.

## 2.2. Results

### 2.2.1. Polycomb proteins are recruited to the *Ink4b-Arf-Ink4a* locus in ESCs

The vast majority of studies focusing on the role of PcG proteins in the regulation of the *Ink4*/*Arf* locus have been carried out in fibroblasts as senescence is easily reproduced in this cell type *in vitro*. Stem cells, in contrast to fibroblasts, can be kept in culture indefinitely without going into senescence and arresting division. For this reason, it is thought that the *Ink4a*/*Arf* locus is repressed in ESCs. As PcG proteins play such an important role in keeping genes repressed in ESCs, several PcG associated epigenetic marks were analysed across the *Ink4*/*Arf* locus in ESCs and compared with MEFs to understand if PcG proteins are important to keep the locus repressed as they have been shown to do in adult stem cells (Ezhkova et al., 2009; Molofsky et al., 2003; Park et al., 2003).

First RNA and protein expression levels for *Ink4a* and *Arf* were assessed. As shown in Figure 2.2c and d, *Ink4a* is not detected in ESCs while *Arf* is expressed at very low, but still detectable levels. Despite this, ESCs do not show any senescence associated features, including staining blue with the senescence associated β-galactosidase (SA-β-galactosidade) marker (Figure 2.2b). Several epigenetic marks associated with PcG (H2AK119u1 and H3K27me3) and TrxG activity (H3K4me3), together with the PRC1 protein MEL18, were analysed by ChIP (Figure 2.3). *Oct4*, active in mESCs, together with *Nkx2-2* (*NK2 homeobox* 2) and *Sox1*, both bivalent in mESCs (Ku et al., 2008), were used as controls.

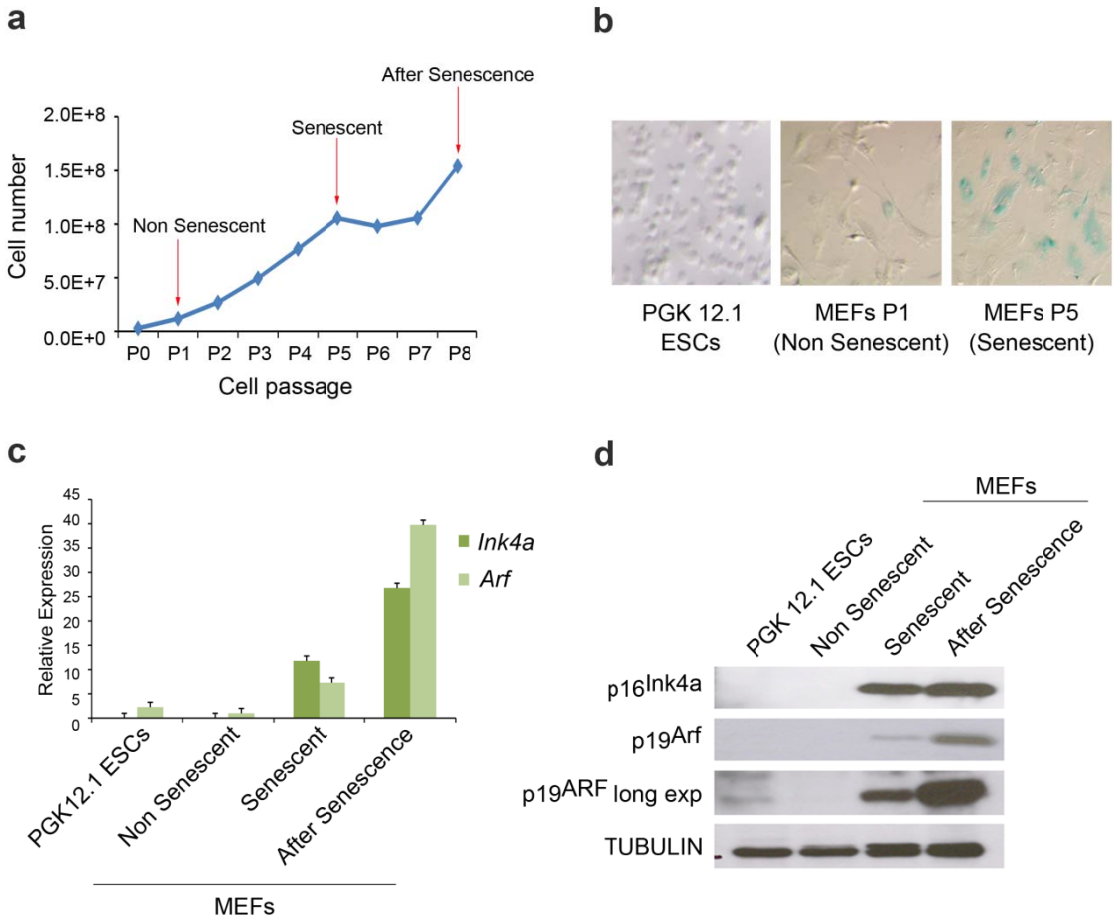



**Figure 2.2 – Characterisation of mESCs, non senescent MEFs, senescent MEFs and MEFs after senescence MEFs used in this study.** (**a**) Cell number with passage. Non senescent MEFs were harvested at passage 1 (P1) while senescent MEFs had undergone cell cycle arrest (when cell number remained constant). MEFs that escape senescence (after senescence) divide normally. (**b**) SA-β-galactosidase staining of PGK12.1 ESCs, non senescent and senescent MEFs. SA-β-galactosidase stains senescent cells blue. (**c**) Expression analysis of *Ink4a* and *Arf* in PGK12.1 ESCs, non senescent MEFs, senescent MEFs and MEFs after senescence. Data was normalized to average expression of *Hmbs (hydroxymethylbilane synthase)*, *Rps (Ribosomal proteim) 14*, *Idh (Isocitrate dehydrogenase) 2* and *B2m (Beta-2-microglobulin)*. Relative expression was determined by normalizing to non senescent *Arf* levels. Bars show average + s.d., n=3. (**d**) Western blot analysis of $p16^{Ink4a}$, $p19^{Arf}$ in whole cell extracts from PGK12.1 ESCs, non-senescent MEFs, senescent MEFs and MEFs after senescence. TUBULIN was used as a loading control.

Two distinct regions can be distinguished within the locus. The 3' region including $RD^{Ink4/Arf}$ region, *Ink4b* and *Arf* promoter show decreased enrichment in repressive marks, while H3K4me3 is detected. On the other hand, the 5' region presents much higher enrichment in

H2AK119u1, as well as MEL18 and H3K27me3, but no H3K4me3 is detected (Figure 2.3b). This region includes the *Ink4a* promoter as well as *Ink4a*/*Arf* exon *E2* (region 10 in Figure 2.3a), an unusual exon shared by the two proteins (Figure 2.3a). Unexpectedly the highest enrichments are indeed at this exon which may be a consequence of the CGI or it might be a PcG recruitment region. Analysing the same epigenetic marks in passage 1 (P1) MEFs a very similar picture is detected (Figure 2.3c). The existence of two different regions marked by the presence/absence of activators/repressors is maintained from mESCs to low passage MEFs. One of the few differences found is the absence of H3K4me3 at the *Arf* promoter (Figure 2.3c). This difference may explain the low expression levels of $p19^{Arf}$ detected in mESCs but absent in MEFs P1 (Figure 2.2c and d). The profile of the *Ink4a-Arf-Ink4b* locus was compared to that of the *Oct4* gene. *Oct4* is a pluripotency gene highly expressed in mESCs. Its expression is downregulated during differentiation and it becomes repressed in differentiated tissues (See section 1.8.2.1). *Oct4* shows enrichment of H3K4me3 in mESCs, and no repressive marks. In P1 MEFs H3K4me3 is lost and expression is silenced (Figure 2.3b and c).

It can be concluded that PcG regulation of the *Ink4*/*Arf* locus is similar in mESCs to non senescent MEFs. PcG bind both *Arf* and *Ink4a* genes but enrichment is higher at the 5' end of the locus which the 3' end prominently enriched for H3K4me3.

### 2.2.2.Role of Polycomb in MEF senescence

Having identified two differently regulated regions at the *Ink4*/*Arf* locus, an analysis of the same epigenetic marks during senescence was carried out. The aim was to analyse how both regions respond to culture stress and to correlate PcG catalytic activity with expression levels.

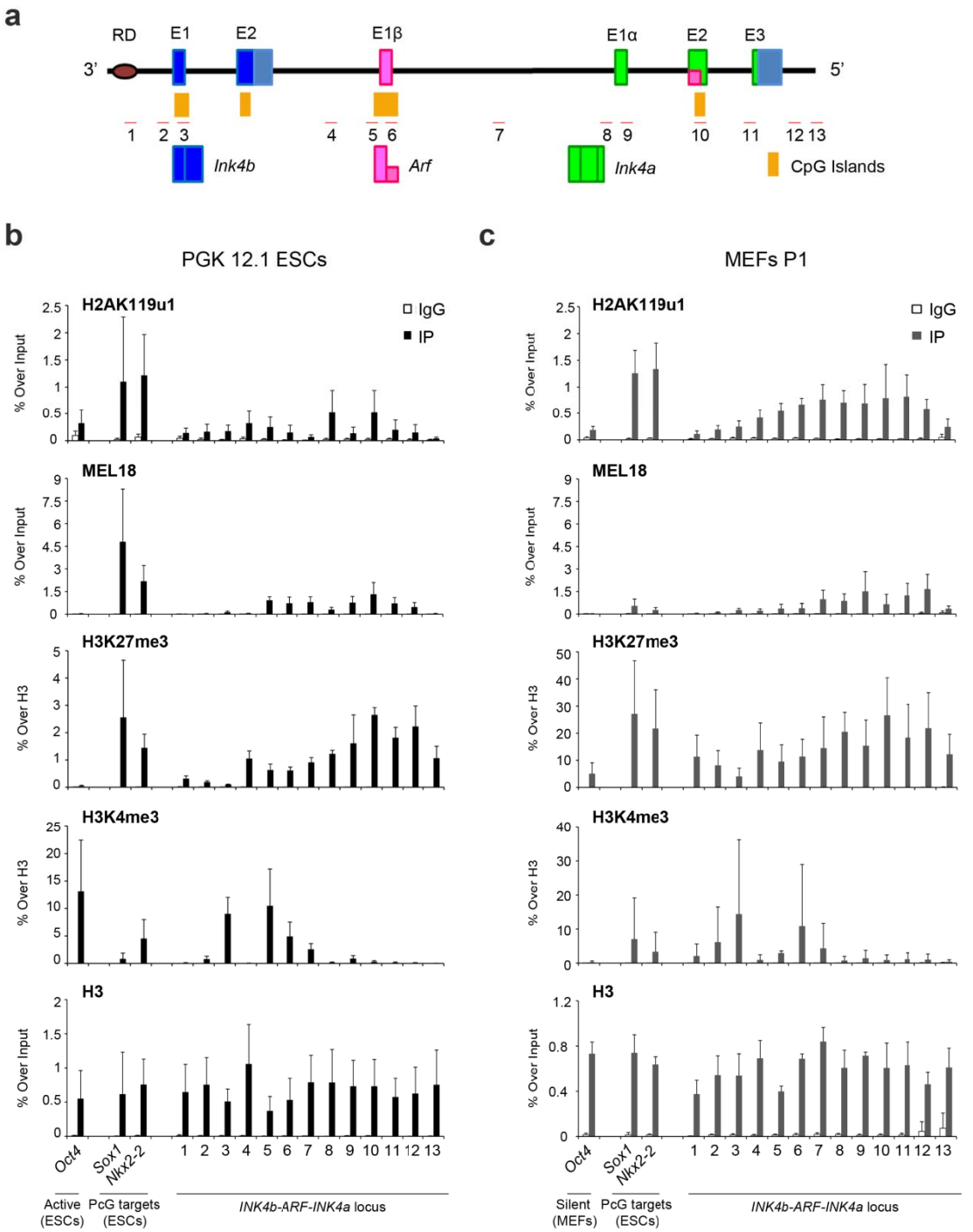



**Figure 2.3 – mESCs and P1 MEFs exhibit the same epigenetic profile at the *Ink4b-Arf-Ink4a* locus.** (**a**) Schematic representation of the *Ink4*/*Arf* locus with primer locations (1 to 13) shown in red (not drawn to scale). (**b**) and (**c**) ChIP analysis of H2AK119u1, MEL18, H3K27me3, H3K4me3 and H3 in PGK12.1 ESCs (**b**) and P1 MEFs (**c**). *Oct4* (active in mESCs and silent in MEFs), *Nkx2-2* and *Sox1* (known PcG targets in mESCs) were used as controls. Bars show average + s.d., n=3.

The experiment (Figure 2.2a) was to harvest MEFs from E 13.5 embryos and expand them in culture. A fraction of the cells (non senescent) were harvested after 1 passage for ChIP and expression analysis while the other fraction was kept in culture. Cells were passaged every third day until they stopped dividing. At this point a fraction of these cells (senescent) were harvested while the other fraction was kept in culture until cells escaping senescence would take over the culture and an established cell line could be obtained and harvested (after senescence). Senescence was accessed by cessation of division but also by senescence associated characteristics. Cells were enlarged, with big nuclei and positively stained for β-galactosidase (Figure 2.2b). In addition, *Ink4a* and *Arf* mRNA and protein levels were upregulated as shown in Figure 2.2c and d respectively. This gene expression upregulation was most likely caused by the loss of PcG proteins from the locus and consequent loss of H2AK119u1 and H3K27me3 enrichment (Figure 2.4b). H3K4me3 increased levels could be detected at the *Ink4a* promoter as well as *Ink4b* and *Arf* (Figure 2.4a and b). The difference between 3' and 5' regions is less evident, although detectable levels of H2AK119u1, MEL18 and H3K27me3 could still be found at the 5' end. An interesting feature noticed in senescent cells is increased levels of histone H3 (Figure 2.4b bottom). This may reflect that senescent cells arrest the cell cycle and chromatin is compacted in senescent-associated foci (Bandyopadhyay et al., 2002; Narita et al., 2003).

MEFs can overcome senescence through various means, but generally they become an immortalized cell line after bypassing cell cycle arrest (Hayflick, 1961). In this study, the population of cells that overcome senescence highly expresses $p16^{Ink4a}$ and $p19^{Arf}$ (Figure 2.2 c and d), despite exhibiting active cell division (Figure 2.4a). Epigenetic marks reflect the high level of activation of the locus as repressive marks cannot be detected and H3K4me3 is highly enriched especially at gene promoter regions (Figure 2.4) but not at the furthest 5' regions, previously enriched for H3K27me3.

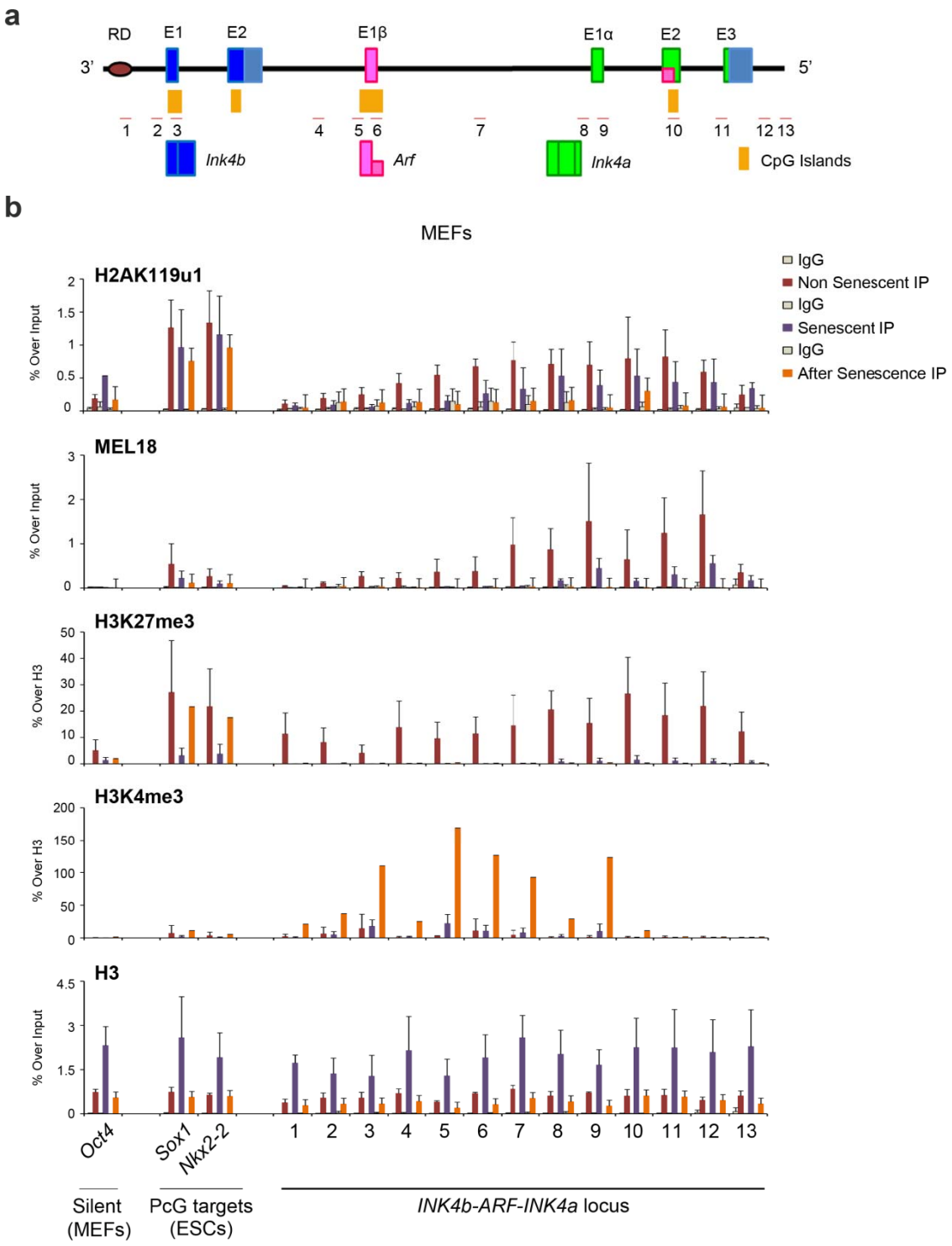



**Figure 2.4 – Epigenetic profile of *Ink4b-Arf-Ink4a* in MEFs during senescence.** (**a**) Schematic representation of the *Ink4*/*Arf* locus with primer locations (1 to 13) shown in red (not drawn to scale). (**b**) ChIP analysis of H2AK119u1, MEL18, H3K27me3, H3K4me3 and H3 in Non-senescent MEFs, senescent MEFs and MEFs after senescence. *Oct4* (silent in MEFs), *Nkx2-2* and *Sox1* (known PcG targets in mESCs) were used as controls. Bars show average + s.d., n=3.

In summary, these results show that PcG protein displacement occurs concomitantly with *Arf* and *Ink4a* expression and H3K4me3 enrichment during MEF senescence.

### 2.2.3. Role of *Ink4a*/*Arf* exon E2 in PcG recruitment

Due to the high enrichment of repressive marks in exon E2 (shared by both *Arf* and *Ink4a*), it was hypothesised that this region could be important for PcG recruitment. This was supported by the existence of a CGI (Figure 2.1), hypothesised to be important for recruitment of PcG proteins (Ku et al., 2008). Previously, a mouse carrying a deletion on *Ink4a*/*Arf* exons E2 and E3 was described. Mice are viable but prone to tumor formation early in life (Serrano et al., 1996). These mice were used to further elucidate the mechanism of PcG recruitment to the locus. MEFs were derived from homozygous mutant E 13.5 embryos (Figure 2.5d) and ChIP was carried out to analyse whether PcG could be recruited to other regions in the absence of exon E2. Figure 2.5c illustrate that MEL18 and H2AK119u1 are still enriched in regions downstream of exon E2 (Figure 2.5b), clearly demonstrating that exon E2 is not necessary for PcG recruitment. Enrichment of H3K27me3 and H3 can still be detected at the depleted region. This may be due to some primer non specificity as this is a small and GC-rich region. To overcome this issue, alternative primer pairs were used but the problem persisted (data not shown). Nevertheless, the profile in the remaining part of the locus is the same as observed for wild-type (WT) MEFs. Therefore it can be concluded that exon E2 is not essential for PcG recruitment to the *Ink4*/*Arf* locus.

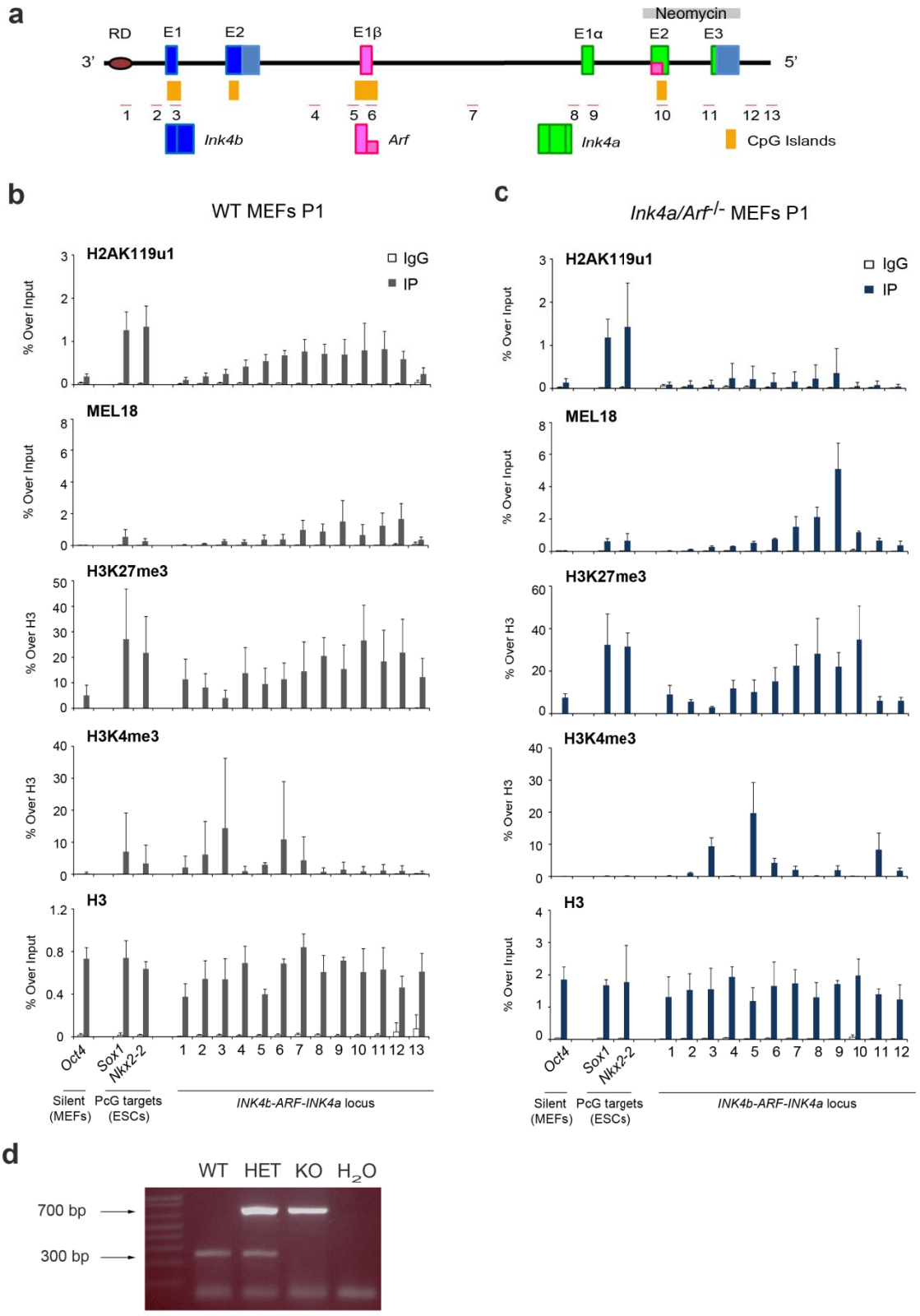



**Figure 2.5 – Exon E2 and E3 deletion does not change the epigenetic profile at the *Arf* and *Ink4a* promoters.** (**a**) Schematic representation of the *Ink4*/*Arf* locus with primer locations (1 to 12) shown in red (not drawn to scale). The grey box represents the deleted region that is replaced by *Neomycin* gene. (**b**) ChIP analysis of the *Ink4*/*Arf* locus for mESCs from figure 2.3c is presented here for comparison with

*Ink4a*/*Arf*$^{-/-}$ MEFs data. (**c**) ChIP analysis of H2AK119u1, MEL18, H3K27me3, H3K4me3 and H3 in *Ink4a*/*Arf*$^{-/-}$ MEFs P1. *Oct4* (silent in MEFs), *Nkx2-2* and *Sox1* (known PcG targets in mESCs) were used as controls. Bars show average + s.d., n=3. (**d**) Mice genotyping PCR showing *Ink4a*/*Arf* heterozygous (HET) and knockout (KO) alleles. Water ($H_2O$) was used as a negative control.

### 2.2.4.siRNA screen to unveil new factors that recruit Polycomb

To find new factors involved in PcG recruitment the *Ink4b-Arf-Ink4a* locus was used as a model to set up an siRNA screen. The aim was to engineer a P1-derived artificial chromosome (PAC), containing the entire locus (Matheu et al., 2004), and including fluorochromes under the control of endogenous *Ink4a* and *Arf* promoters (Figure 2.6a). The PAC could then be integrated in *Ink4*/*Arf*$^{-/-}$ mice or cells, and a genome-wide siRNA screen could be performed. The read-out would be the fluorochromes that would be expressed when the locus loses PcG repression. Many advantages can be seen in the use of this locus. First, the KO mice are available from which MEFs or ESCs can be derived (Serrano et al., 1996). Second, the locus has been thoroughly described and it is known that PcG proteins play an important role in its regulation (Bracken et al., 2007; Dietrich et al., 2007; Jacobs et al., 1999; Miki et al., 2007). Third, it is easy to induce the locus by UV, γ-irradiation or simply by keeping primary fibroblasts in culture (Lloyd, 2002; Zindy et al., 1997), *Arf* can also be highly upregulated in MEFs after treatment with TSA, an HDAC inhibitor (Matheu et al., 2005).

The recombineering (recombination-mediated genetic engineering) strategy was designed with the aim of introducing dsRed at the ATG start site of *Arf* and a construct containing ECFP, IRES and Neomycin at the start site of *Ink4a* (Figure 2.6a). Both constructs were successfully integrated and positive clones analysed for the correct integration of the external sequences (Figure 2.6b). PAC microinjections into *Ink4a*/*Arf*$^{-/-}$ blastocysts were performed as well as lipofection of PACs into *Ink4a*/*Arf*$^{-/-}$ ESCs and 129.1 WT ESCs. One embryo with a PAC

integration was obtained together with one 129.1 ESCs clone. PCR analysis confirmed the presence of different parts of the PAC (Figure 2.6c).

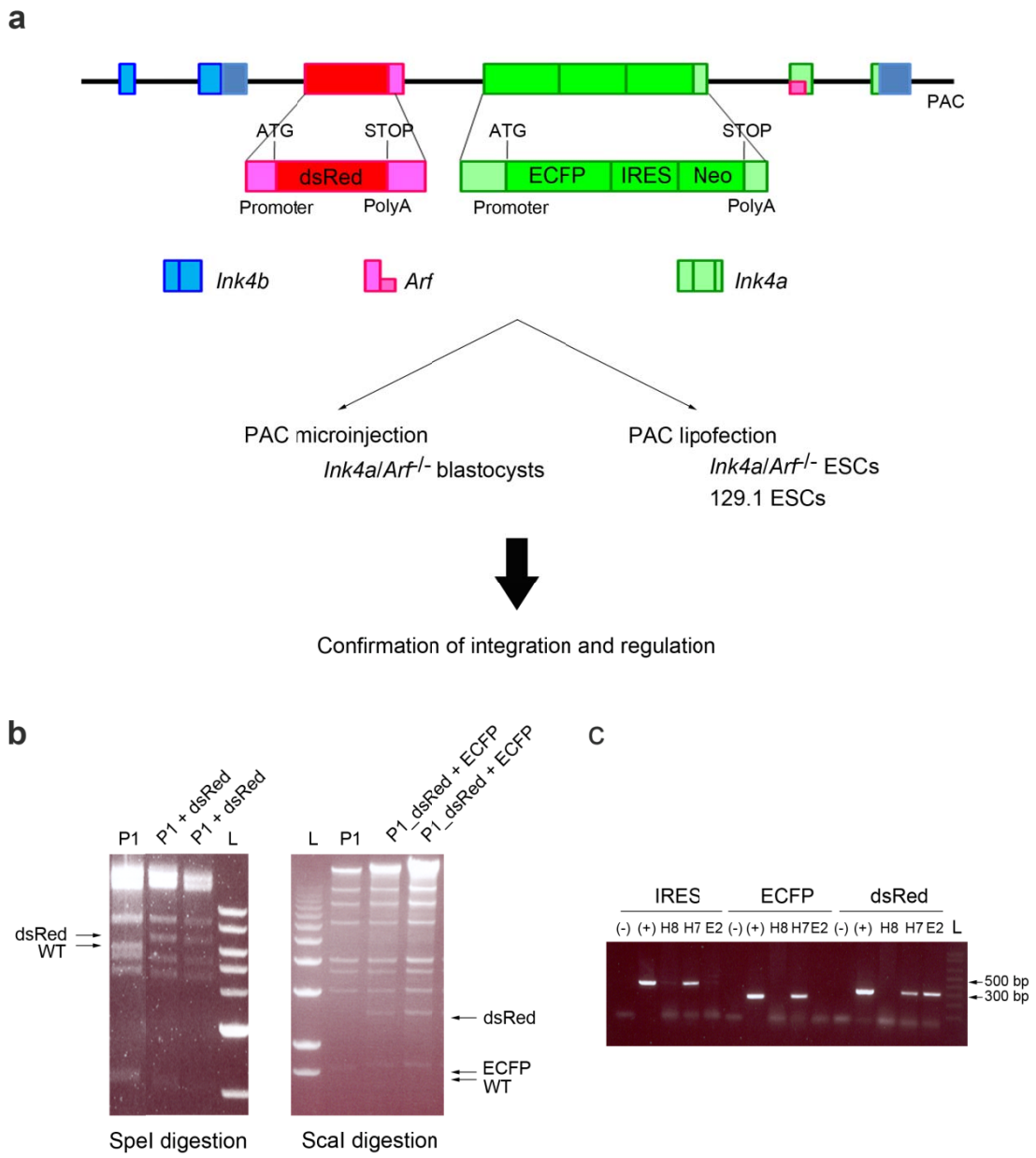



**Figure 2.6 – siRNA screen experimental plan.** (**a**) PAC containing the *Ink4b-Arf-Ink4a* locus is represented, highlighting genetic modifications undertaken. dsRed with a polyA sequence was added at the *Arf* ATG and a construct containing ECFP-IRES-Neomycin was added after the *Ink4a* ATG. (**b**) Enzymatic digestion of PAC containing dsRed and dsRed with ECFP contruct. P1 is parental PAC; L is the molecular weight marker ladder; WT and dsRed diagnostic bands are highlighted. (**c**) PCR analysis of 129.1 ESC containing PAC. Amplification of dsRed, ECFP and IRES regions are shown. L is ladder; E2 is a clone containing only dsRed; H7 is a positive clone for all analysed regions; H8 is a negative clone; (+) is a positive PCR control with PAC DNA; (-) is a negative water control.

To test whether the PAC was properly regulated and could be activated when the *Ink4a*/*Arf* locus was upregulated, UV, γ-radiation and TSA treatment were performed. None of the treatments gave positive results (results not shown) and it has been speculated that PAC could have undergone rearrangement. Alternatively, it is possible that the PAC randomly inserted in a region that disrupted its normal regulation. Production of additional cell lines and/or transgenic animal was not possible within the time frame of this study.

## 2.3. Discussion

### 2.3.1. PcG regulation of the *Ink4*/*Arf* locus

Keeping cells actively dividing while avoiding unrestrained proliferation and cancer is the major role of the *Ink4b-Arf-Ink4a* locus. Activation of the locus triggers cellular senescence and mediates DNA damage pathways leading to oncogenic-induced senescence (Bartkova et al., 2006; Di Micco et al., 2006). Therefore *Ink4* and *Arf* levels need to be tightly regulated. The *Ink4*/*Arf* locus mediates senescence both in mice and humans even though it is speculated that the relative importance of *Ink4a* and *Arf* is different between species (Bracken et al., 2007; Gil and Peters, 2006). Several lines of evidence have demonstrated that genetic manipulation of PcG proteins can interfere with *Ink4a* and *Arf* expression. It is thought that PcG proteins may bind the mouse *Arf* and *Ink4a* locus and regulate its expression (Bruggeman et al., 2005; Jacobs et al., 1999; Miki et al., 2007; Molofsky et al., 2003). In humans it was shown that PcG can regulate the expression of both genes, although its direct recruitment to the *ARF* promoter has only been demonstrated in one study so far (Bracken et al., 2007) while others do not detect it (Barradas et al., 2009; Maertens et al., 2009). In contrast, PcG recruitment to the *INK4a* locus is well documented (Agherbi et al., 2009; Bracken et al., 2007; Maertens et al., 2009; Miki et al., 2007).

In this work mESCs and MEFS were used to understand the role of PcG in mouse *Ink4b-Arf-Ink4a* regulation. Two differently regulated regions were found. The 3' region encompasses $RD^{Ink4/Arf}$, *Ink4b* and *Arf* promoters where low but still detectable levels of PcG direct binding could be observed both in mESCs or MEFs (Figure 2.3). On the other hand, the 5' region encompasses *Ink4a* promoter and exon E2 where MEL18 is highly bound and a strong enrichment of H2AK119u1 and H3K27me3 is detected, suggesting the recruitment of PRC2 as well (Ku et al., 2008). Data presented here is in accordance with other studies where MEL18 and RING1B (Miki et al., 2007), as well as both EZH2 and M33 (Pc homologue) are detected despite the low levels of H3K27me3 (Agherbi et al., 2009). In addition, several PRC1 and PRC2 proteins have been detected at *Arf* and *Ink4a* promoters (Bracken et al., 2007). PcG regulates both *Arf* and *Ink4a* although the increased binding of PcG to *Ink4a* points towards a tighter regulation of this gene, an argument confirmed by the minor upregulation of *Arf* in comparison to *Ink4a* in $Bmi1^{-/-}$ MEFs (Agherbi et al., 2009; Bruggeman et al., 2005; Jacobs et al., 1999). Comparing results presented here with the human locus it can be hypothesised that the influence of PcG in *Arf* regulation was lost between mice and humans.

Interestingly there is no difference in PcG recruitment to the locus in mESCs or MEFs P1 (Figure 2.3). Therefore it seems that, at least for the regions analysed, differential PcG recruitment does not provide an explanation for the inability for ESCs to senesce. Further work needs to be carried out with the aim to understand different regulations of the *Ink4*/*Arf* locus in different cell types.

A region upstream of *Ink4b* has been described as a regulatory domain of the *Ink4/Arf* locus ($RD^{Ink4/Arf}$). This region is involved in the global regulation of the locus (Gonzalez et al., 2006) through the binding of the replication protein CDC6, followed by recruitment of histone deacetylases which are responsible for the repression of the locus (Gonzalez et al., 2006). In

addition EZH2, CBX7 and BMI1 are found at this region even though H3K27me3 is not enriched (Agherbi et al., 2009). In this work enrichment of H2AK119u1 or MEL18 was not detected even though some H3K27me3 was present but subsequently lost in senescence (Figure 2.4b). It may be that a different PRC1 complex (Simon and Kingston, 2009) lacking MEL18 and its catalytic activity is present and has a role in repressing the locus by alternative means (Eskeland et al., 2010; Francis et al., 2004; Shao et al., 1999).

### 2.3.2.PcG recruitment to the *Ink4b-Arf-Ink4a*

Despite a large number of studies addressing PcG regulation of the *Ink4b-Arf-Ink4a* locus and PcG recruitment, advances are slow. In this study a striking enrichment of PcG was found in exon E2, an exon shared by both *Arf* and *Ink4a*. Concentration of PcG at this site had been previously observed but perhaps because it was in an exon it was not considered as a regulatory element (Agherbi et al., 2009; Miki et al., 2007). This strong enrichment together with the presence of a CGI and the fact that exon E2 is shared between two genes in an unusual way suggested a role in PcG recruitment. Nevertheless it was shown that mice lacking this exon can still recruit PcG proteins to *Arf* and *Ink4a* promoter and so it can be concluded that it is not essential for PcG recruitment (Figure 2.5c). Alternative means of PcG recruitment to the locus, such as DNA binding proteins and ncRNAs, are possibly and this is supported by recent findings.

PcG recruitment may be mediated by DNA binding proteins, resembling PREs in *Drosophila*. One hypothesis is that the pRB protein, connects cell cycle arrest with *Ink4*/*Arf* repression by PcG recruitment (Kotake et al., 2007). Another candidate came from a yeast two-hybrid screen that recently identified the protein ZFP277 as a BMI1 interacting partner (Negishi et al., 2010). It was shown that ZFP277 binds to BMI1 and recruits PRC1 to the *Ink4*/*Arf* locus in MEFs. It was

also shown that levels of ZFP277 increase as cells senesce and knockout MEFs undergo premature senescence, losing both PRC1 and PRC2 binding (Negishi et al., 2010).

In human cells the presence of a lncRNA, *ANRIL*, antisense to the *Ink4b-Arf-Ink4a* locus was demonstrated (Pasmant et al., 2007). A recent study showed that in human cells the presence of this transcript was essential to keep the locus repressed and that locus upregulation in senescence coincides with *ANRIL* downregulation (Yap et al., 2010). In addition it was shown that CBX7 can bind H3K27me3 and *ANRIL*. Therefore it was hypothesised that *ANRIL* is needed for PcG recruitment. While the authors argue that RNAPII nascent transcripts are essential for PRC1 and PRC2 recruitment they fail to show its specificity to *ANRIL* and PcG as has been previously shown for miRNA *miR-24*, also involved in *Ink4a* regulation (Lal et al., 2008). Furthermore an inverse correlation between PcG (as BMI1 and H3K27me3) and RNAPII at the *Ink4*/*Arf* locus was demonstrated. This was shown to be mediated by KDM2B binding to the locus and resulting H3K36me2 demethylation (Tzatsos et al., 2009).

So far many PcG recruitment mechanisms for the *Ink4b-Arf-Ink4a* locus have been suggested but the precise mechanism is still unknown. Unbiased studies, such as RNAi screens are a powerful tool to uncover PcG recruitment and to open unexplored pathways and the reagents developed in this work will facilitate such studies in the future.

# 3. PRC1 activity in *Eed* mutant mouse embryonic stem cells

## 3.1. Introduction

Polycomb group (PcG) proteins were first found in *Drosophila* where they contribute to heritable repression of *Hox* genes (Francis, 2009). The *Hox* cluster encodes several genes that are expressed in a segmented manner along the anterior-posterior axis of the embryo. Once initiated in the embryo, *Hox* gene expression is maintained until adult life. It is therefore important to maintain expression or silencing in a heritable manner, a role for which PcG proteins are essential. Subsequently, homologs of the *Drosophila* PcG proteins were also identified in mammals (see section 1.3.3.1).

### 3.1.1. Importance of PcG proteins in mouse development

Gene knockout studies have shown the major importance of PRC2 and PRC1 in development. Because many KO show early embryo lethality it has been difficult to understand the role of PcG complexes later in development with only a few clues emerging from conditional knockout (cKO) embryos (Kerppola, 2009). Nevertheless, the importance of both PcG complexes in the initial stages of embryo development and differentiation has become clear.

#### 3.1.1.1. PRC2 and PRC1 mutant studies

Null mutations in core PRC2 proteins result in loss of H3K27me3. Embryos suffer premature death soon after gastrulation (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004). Even though embryos die around E 6.6-7.5, it is possible to isolate ESCs that maintain the capacity to self-renew *in vitro* (Chamberlain et al., 2008; Pasini et al., 2007; Shen et al., 2008). Although PRC2 deficient ESCs all show self-renewal characteristics they present different patterns of

differentiation. *Suz12*$^{-/-}$ ESCs completely fail to differentiate due to their inability to silence pluripotency genes and to activate developmental genes (Pasini et al., 2007). On the other hand both *Eed*$^{-/-}$ and *Ezh2*$^{-/-}$ ESCs differentiate albeit with an altered expression profile compared to WT differentiated ESCs (Boyer et al., 2006; Chamberlain et al., 2008; Shen et al., 2008).

Mutations in most PRC1 proteins result in transformations of segments of the axial skeleton in addition to neurologic and haematopoietic defects (Akasaka et al., 1997; Lessard and Sauvageau, 2003; Molofsky et al., 2003; Ohta et al., 2002), without affecting embryo viability. That is because of the high degree of redundancy with each core protein having two or more homologues (Figure 1.6b). In combined mutant animals phenotypes are generally more severe. An example is *Mel18* and *Bmi1*, homologues of *Drosophila Psc*, for which single mutants are viable but in the double KO embryos die at E 9.5 (Akasaka et al., 2001). Although *Bmi1* single mutant mice are viable, they show progressive postnatal growth retardation as *Bmi1* is important for self-renewal of adult stem cells in the peripheral and central nervous systems (Lessard and Sauvageau, 2003; Molofsky et al., 2003). There are only two homologues of the catalytic PRC1 subunit *dRing* in mammalian cells, *Ring1A* and *Ring1B*. *Ring1B* single mutants have a severe phenotype with embryo lethality occurring during gastrulation (Voncken et al., 2003). Conditional deletion of *Ring1B* in *Ring1A* mutant background is the only PcG mutant incompatible with ESCs self-renewal capacity (Endoh et al., 2008).

### 3.1.2.Hierarchical recruitment model

The hierarchical recruitment model was formulated by Wang and colleagues in 2004 (Wang et al., 2004b) and postulates that PRC2 is recruited to target genes by *Pho* and *Phol* (in *Drosophila*) where it methylates histone H3 on lysine 27 through the action of *Drosophila* E(Z). Given the reports that the chromodomain of Pc (*Drosophila*) and CBX (mammals) family

proteins can bind trimethylated residues on H3K27 (Cao et al., 2002; Czermin et al., 2002; Fischle et al., 2003; Kuzmichev et al., 2002; Min et al., 2003) it was hypothesised that PRC1 is recruited after PRC2, in a process mediated by H3K27me3. Although this model is widely accepted there is some evidence that appears inconsistent.

#### 3.1.2.1. Arguments in favour

The main argument in favour of the hierarchical recruitment of PcG complexes comes from genome-wide localization studies. PRC1 and PRC2 co-localize on many target genes and almost all PRC1 binding sites have PRC2 whereas the inverse is not true (Ku et al., 2008). In addition, knockdown of *Drosophila* core PRC2 protein Esc (as well as Pho and Phol) reduces H3K27me3 levels and Pc recruitment at *Drosophila Ubx* PRE (Cao et al., 2002; Wang et al., 2004b). This is also observed in mammals with *Eed* deletion affecting PRC1 binding (Boyer et al., 2006) while *Bmi1* deletion does not affect levels of H3K27me3 (Cao et al., 2005). Further, in human cells a correlation was found between high levels of H3K27me3 (mediated by knockdown of demethylase Utx) and low levels of the trimethyl mark (mediated by overexpression of demethylase JMJD3) with increased or decreased enrichment of PRC1 proteins respectively (Agger et al., 2007; Lee et al., 2007b). A recent study demonstrated that trimethylation of histone H3K27 via a heterologous viral methyltransferase, restores PRC1 targeting to the human *HOX* promoter (Mujtaba et al., 2008).

#### 3.1.2.2. Arguments against

It was shown that the dissociation constant between CBX proteins and H3K27me3, that it has been proposed to recognise, is between 20 and 500 μM. That is relatively high compared with interaction between mouse HP1α and H3K9me3 (2 ± 1 μM). In addition, the binding affinity is similar for both H3K27me3 and H3K9me3 residues (Bernstein et al., 2006b).

Further evidence comes from studies on the Xi chromosome. Initial work showed that both PRC1 and PRC2 are recruited to the Xi in an *Xist* dependent manner (de Napoles et al., 2004; Mak et al., 2002; Plath et al., 2003; Plath et al., 2004; Silva et al., 2003). Nevertheless in *Eed* knockout embryos PRC1 can be recruited to the Xi chromosome independently of *Eed* (Schoeftner et al., 2006).

*Ezh2*$^{-/-}$ embryos have also provided insight into PRC1 independent recruitment. PRC1 proteins were found in heterochromatic regions of the male pronucleus after fertilization of two *Ezh2*$^{-/-}$ gametes, without detectable H3K27me3 (Puschendorf et al., 2008), and RING1B was found to be repressing the imprinted gene *Knq1ot1* (Terranova et al., 2008). Furthermore, when *Suz12*$^{-/-}$ ESCs differentiate PRC1 components such as BMI1 and CBX8 can be recruited to some target loci despite the low levels of H3K27me3 (Pasini et al., 2007) and in *Eed*$^{-/-}$ ESCs, RING1B is still found at some loci and repeats (Leeb et al., 2010).

Two interesting studies using bimolecular fluorescence complementation (BiFC) and imaging analysis showed that mutations in the chromodomain or chromobox of CBX proteins do not affect their chromatin association or RING1B binding ability (Ren et al., 2008; Vincenz and Kerppola, 2008). Moreover binding to chromatin was not affected in *Eed*$^{-/-}$ cells indicating that H3K27me3 is dispensable for the CBX-chromatin complex formation (Vincenz and Kerppola, 2008). Also PRC1 proteins are still bound to polytene chromosomes after *Esc* and *Escl* (Esc like) depletion, and consequent H3K27me3 loss, even though the majority of sites lose PRC1 with time (Ohno et al., 2008). Genome-wide analysis of H3K27me3, RING1B and MEL18 demonstrated the existence of some PRC1 bound genes that do not have detectable H3K27me3 (Boyer et al., 2006; Bracken et al., 2006). Additionally a study in *Drosophila* showed that both PRC1 (Pc and Psc) and PRC2 (E(Z)) core proteins are found at *Ubx* gene PREs depleted in H3K27me3 and not in areas where the methyl mark was abundant (Kahn et al., 2006). The

latter result is arguable as other studies demonstrate H3K27me3 enrichment at PREs (Cao et al., 2002; Ringrose et al., 2004; Wang et al., 2004b).

It is important to note that those and others arguments against the hierarchical model are limited and could represent exceptions or technical limitations and most findings relate to specific targets rather than genome wide. To further test the hierarchical model an $Eed^{-/-}$ ESC line and an *Eed* conditional knockout ESC line (Ura et al., 2008) were studied to determine PRC1 activity and occupancy.

## 3.2. Results

### 3.2.1.H2AK119u1 levels are maintained in $Eed^{-/-}$ ESCs

As described in the previous section, PcG recruitment to target genes has been proposed to be hierarchical (Cao et al., 2002; Czermin et al., 2002; Fischle et al., 2003; Kuzmichev et al., 2002; Min et al., 2003; Wang et al., 2004b). Previous studies carried out in the Brockdorff laboratory by Mariana de Napoles and Ruth Appanah, showed that in *Eed* mutant ESCs global levels of H2AK119u1 remained high (Figure 3.1a) and additionally that accumulation of H2AK119u1 on the Xi occurs in differentiating $Eed^{-/-}$ ESCs (Figure 3.1b). These observations were demonstrated in independent studies (Leeb and Wutz, 2007; Schoeftner et al., 2006).

To investigate the importance of H3K27me3 in recruitment of PRC1, ChIP analysis was carried out for selected PcG target loci in $Eed^{-/-}$ mESCs that are deficient for the core PRC2 protein EED. $Eed^{-/-}$ ESCs were derived in the laboratory from mouse blastocysts carrying a point mutation in the *Eed* gene (*eed3354SB*). The mutation causes a change in the WD40 domain structure (Schumacher et al., 1996), disrupting interaction with EZH2 (Denisenko et al., 1998; Han et al., 2007) and providing the molecular basis for the null phenotype. The *eed3354SB* point mutation has the same phenotype (embryonic lethality around E 8.5), as complete deletion of

the *Eed* locus (Faust et al., 1995; Schumacher et al., 1996) and both disrupt PRC2, resulting in complete loss of H3K27me3 (Silva et al., 2003). In addition to loci repressed by PcG proteins, loci that are specifically expressed in mESCs (*Oct4*, *Nanog* and *Sox2*), loci that are ubiquitously expressed (*B2m*, *Rps14*, *Ensa(endosulfine alpha)*) and a locus that is repressed in mESCs independent of PcG activity (*Gata1*) were included in the analysis. As illustrated in Figure 3.2, levels of H3K27me3 were fully depleted in *Eed* mutant mESCs and there was a greatly reduced occupancy of the PRC1 core proteins RING1B and MEL18 compared to *Eed* WT mESCs. These observations are consistent with hierarchical recruitment of PRC1 by H3K27me3. However, H2AK119u1 levels, which are dependent on PRC1 occupancy, were broadly similar in mutant compared to WT mESCs (Figure 3.2, lower panel).

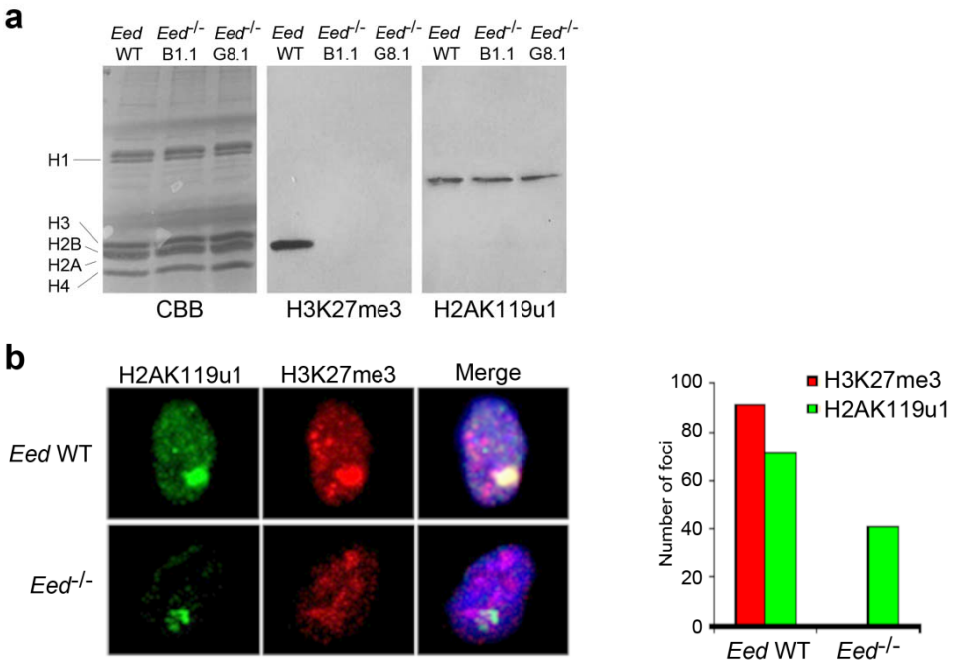



**Figure 3.1 – H2AK119u1 maintenance in *Eed*$^{-/-}$ cells.** (**a**) Protein blot analysis of histone extracts showing H2AK119ub1 present in two *Eed* KO cells lines, *Eed*$^{-/-}$ B1.1 and *Eed*$^{-/-}$ G8.1. H3K27me3 is absent in those two knockouts. Coomassie brilliant blue (CBB) is used as loading control. This work was performed by Mariana de Napoles. (**b**) Immunofluorescence of differentiated *Eed* WT and *Eed*$^{-/-}$ ESCs for H3K27me3 and H2AK119ub1 (left) and quantification of X-chromatin foci (right). This work was carried out by Ruth Appanah.

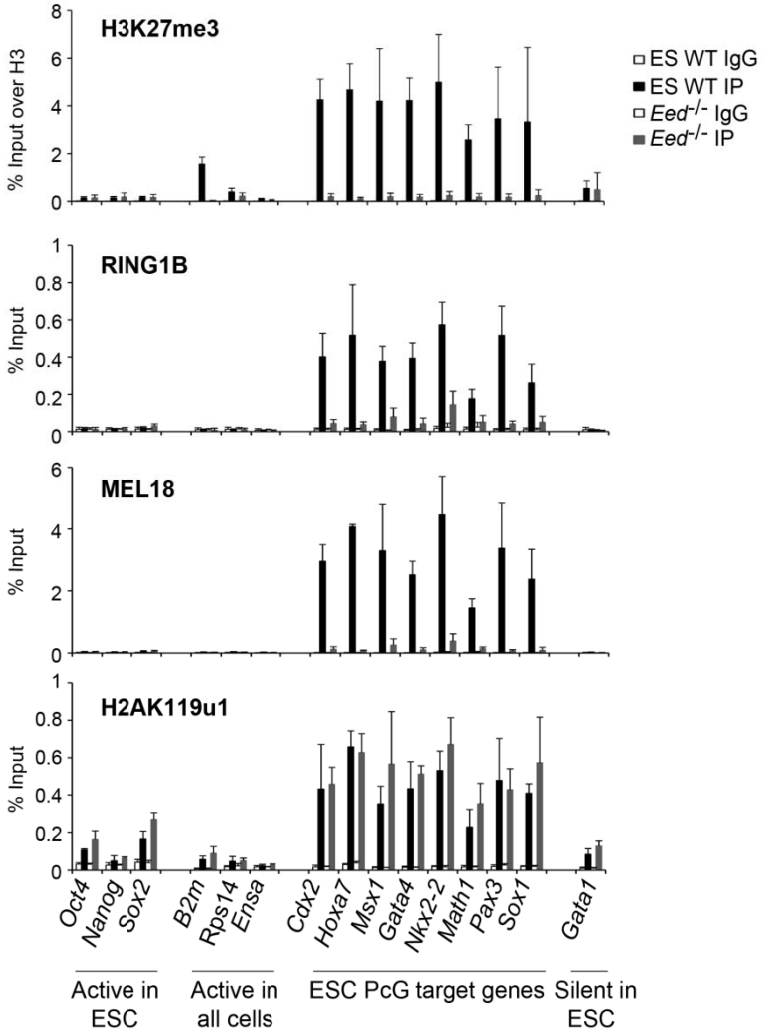



**Figure 3.2 – Targeted H2AK119u1 is retained in *Eed*$^{-/-}$ mESCs.** ChIP analysis of H3K27me3, RING1B, MEL18 and H2AK119u1 in *Eed* WT and *Eed*$^{-/-}$ mESCs. *Oct4*, *Nanog*, *Sox2* (active in ESC), *B2m*, *Rps14*, *Ensa* (active in all cells) and *Gata1* (silent in ESC) were used as controls. *Cdx2*, *Hoxa7*, *Msx (Homeo box, msh-like) 1*, *Gata4*, *Nkx2*-2, *Math (Atonal homologue) 1*, *Pax3* and *Sox1* are known PcG target loci. Bars show average + s.d., n=3.

On the one hand these results corroborate the model due to the absence of PRC1 core components. Conversely, H2AK119u1 global levels are unchanged as is its targeting to PcG repressed genes demonstrating that PRC1 catalytic activity is present in the absence of PRC2.

### 3.2.2. *Eed* conditional knockout mimics constitutive depletion

Maintenance of H2AK119u1 could be an adaptational characteristic of *Eed*$^{-/-}$ ESCs, acquired as a consequence of deriving cell lines from mutant embryos. To test this a cKO *Eed* ESC line was analysed, in which *Eed* is repressed when doxycycline (DOX) is added to the culture medium (Ura et al., 2008).

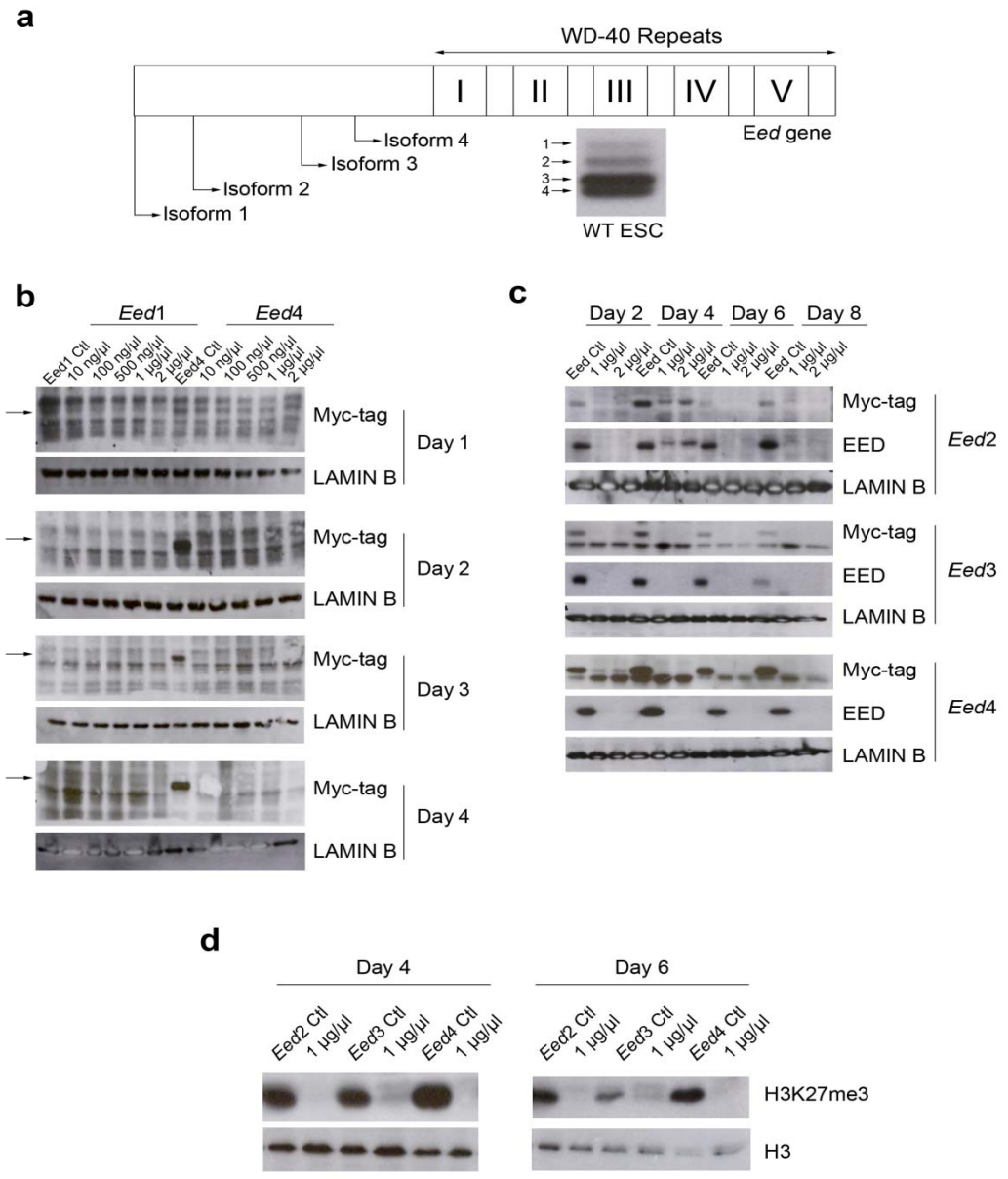



**Figure 3.3 – Doxycycline conditionally depletes *Eed* isoforms from *Eed* conditional ESCs.** (**a**) Schematic representation of the *Eed* gene with the initiation site of each isoform indicated and its representation on a western blot (nuclear extract from WT ESC). (**b**) Analysis of myc-tag expression by western blot on total cell extracts. Analysis was carried out in *Eed*4 and *Eed*1 cell lines treated with 10, 100, 500 ng/µl, 1and 2 µg/µl of DOX during 1, 2 and 3 days. LAMIN B was used as a loading control. (**c**) Analysis of myc-

tag and EED expression by western blot on nuclear extracts. *Eed*2, *Eed*3 and *Eed*4 ESC lines were treated with 1 and 2 µg/µl of DOX during 2, 4, 6 and 8 days. LAMIN B was used as a loading control. (**d**) H3K27me3 analysis was carried by western blot on nuclear extracts from *Eed*2, *Eed*3 and *Eed*4 WT ESCs and cells treated with 1 µg/µl of DOX (Eed4 cKO) for 4 and 6 days. H3 was used as a loading control.

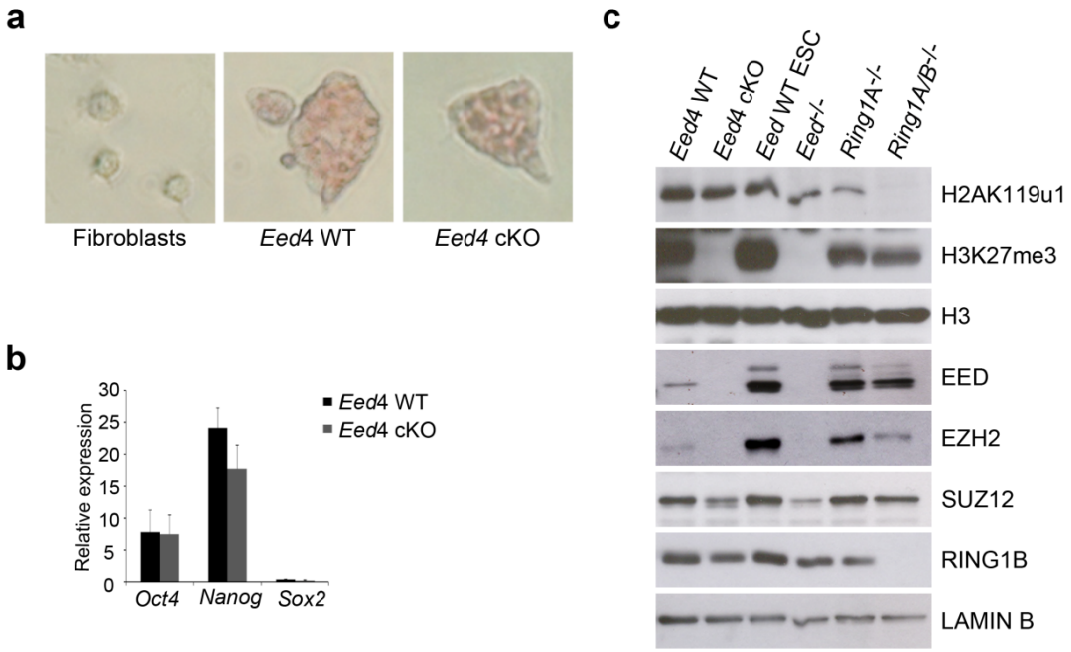



**Figure 3.4 – H2AK119u1 is retained following conditional deletion of *Eed* in mESCs.** (**a**) Alkaline phosphatase staining showing that *Eed*4 WT and cKO cells, 15 days after treatment with 1 µg/ml DOX maintain mESC characteristics. Fibroblasts were used as negative control. (**b**) Expression analysis of pluripotency genes *Oct4*, *Nanog* and *Sox2* in *Eed*4 WT and *Eed*4 cKO 15 days after treatment with 1 µg/ml DOX. Expression was normalized against the average of 3 housekeeping genes, *Hmbs*, *Gapdh (Glyceraldehyde-3-phosphate dehydrogenase)* and *Idh2*. Bars show average + s.d., n=3. (**c**) Western blot analysis on histone extracts (H2AK119u1, H3K27me3, H3) and nuclear extracts (EED, EZH2, SUZ12, RING1B and LAMIN B) in WT and mutant mESCs cell lines as indicated.

In WT ESCs there are four *Eed* isoforms that differ in their N terminal, because of the use of alternative start sites (Kuzmichev et al., 2004), as exemplified in Figure 3.3a. Conditional *Eed* ESCs (kind gift of Hiroki Ura) have a deletion of the endogenous *Eed* gene with myc tagged versions of each *Eed* isoform then transfected independently (Ura et al., 2008). Initially doxycycline (Dox) concentrations were titrated between day 1 to 4 on *Eed*1 and *Eed*4 cells lines (reconstituted with isoform 1 and isoform 4 respectively). As shown in Figure 3.3b, myc-tagged

*Eed* is poorly expressed from the *Eed*1 cell line. Next a comparative analysis of *Eed*2, *Eed*3 and *Eed*4 was carried out. All of the cell lines showed depleted levels *Eed* protein 6 days after initiation of DOX treatment (Figure 3.3c). In Figure 3.3c it can be seen that some bands with similar size as EED were detected in the *Eed*2 cell line treated with DOX. These are most likely contaminants as their size is slightly higher that the WT band. As expected H3K27me3 is efficiently depleted following treatment with DOX (Figure 3.3d). The best cell line in terms of EED expression, H3K27me3 levels and depletion after DOX treatment is *Eed*4. Further experiments were performed using this cell line.

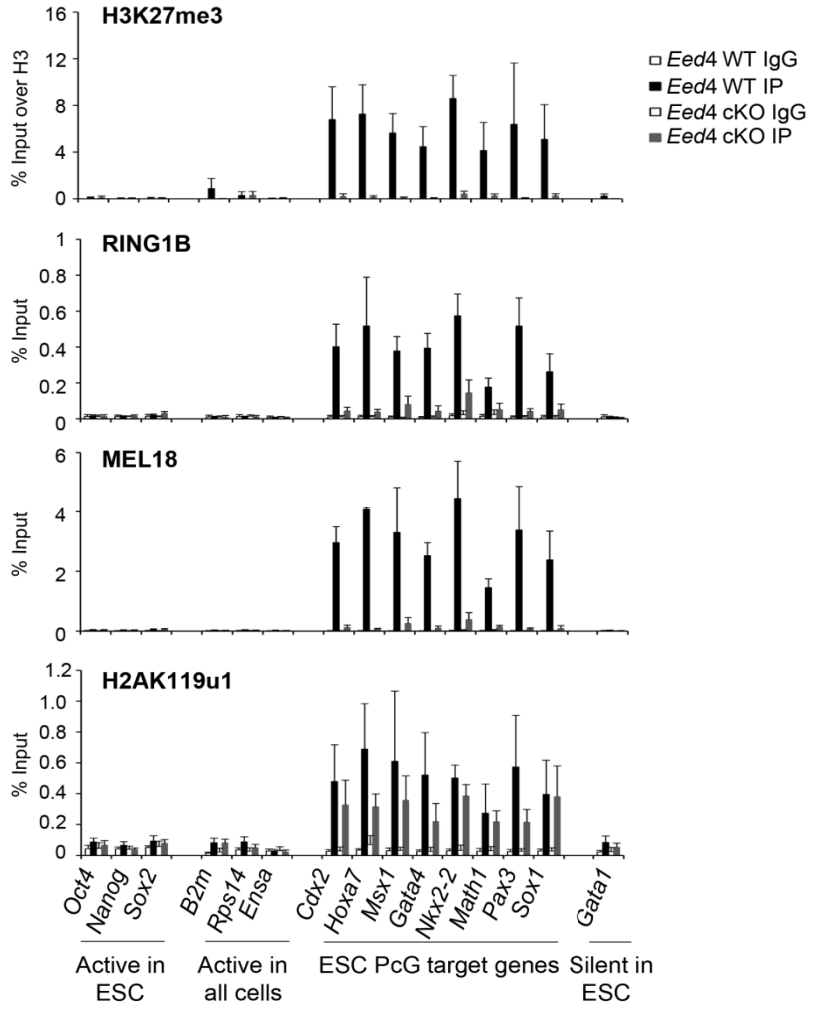



**Figure 3.5 – H2AK119u1 is correctly targeted following conditional deletion of *Eed* (15 days doxycycline treatment) in mESCs.** ChIP analysis in *Eed*4 WT and *Eed*4 cKO (15 days Dox treatment) for H3K27me3, RING1B, MEL18 and H2AK119ub1 as indicated. Bars show average + s.d., n=3. *Oct4*, *Nanog*, *Sox2* (active in ESC), *B2m*, *Rps14*, *Ensa* (active in all cells) and *Gata1* (silent in ESC) were used as controls. *Cdx2*, *Hoxa7*, *Msx1*, *Gata4*, *Nkx2*-2, *Math1*, *Pax3* and *Sox1* are known PcG target loci.

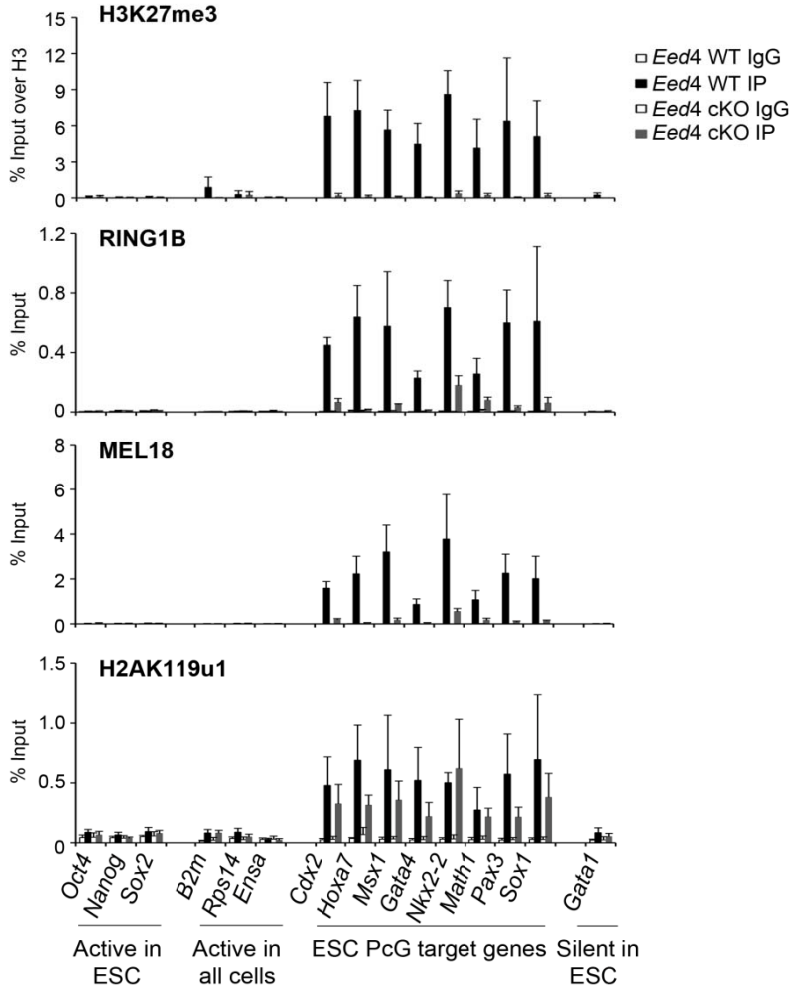



**Figure 3.6 – H2AK119u1 is correctly targeted following 7 days doxycycline treatment of *Eed*4 ESCs.** ChIP analysis in *Eed*4 WT and *Eed*4 cKO (7 days DOX treatment) for H3K27me3, RING1B, MEL18 and H2AK119ub1 as indicated. Bars show average + s.d., n=3. *Oct4*, *Nanog*, *Sox2* (active in ESC), *B2m*, *Rps14*, *Ensa* (active in all cells) and *Gata1* (silent in ESC) were used as controls. *Cdx2*, *Hoxa7*, *Msx1*, *Gata4*, *Nkx2*-2, *Math1*, *Pax3* and *Sox1* are known PcG target loci.

In order to determine if *Eed* depletion affected ESC characteristics, *Eed*4 cells were treated with DOX for 15 days. The treatment did not affect ESC pluripotency as assessed by alkaline phosphatase staining (Figure 3.4a) and quantitative RT-PCR (qPCR) analysis of pluripotency transcription factors (Figure 3.4b). Western blot (WB) analysis demonstrated that EED protein and H3K27me3 were completely depleted, and additionally, levels of PRC2 core proteins EZH2 and SUZ12 were reduced (Figure 3.4c). This latter observation is consistent with previous studies demonstrating instability of core PRC2 components when formation of the

holocomplex is disrupted (Pasini et al., 2004; Schoeftner et al., 2006). In contrast global levels of H2AK119u1 were broadly unchanged, as were levels of RING1B (Figure 3.4c). Similarly, analysis of defined PcG target genes by ChIP demonstrated retention of H2AK119u1, albeit at slightly reduced levels (Figure 3.5). Occupancy of core PRC1 proteins (MEL18 and RING1B) was again strongly reduced following depletion of H3K27me3. ChIP analysis was performed in *Eed*4 cells treated both for 7 and 15 days with similar results (Figure 3.6). These results substantiate the conclusion that maintenance of H2AK119u1 in ESCs occurs independently of PRC2 and associated H3K27me3.

### 3.2.3.*de novo* establishment of H2AK119u1

It is possible that H3K27me3 is required to recruit PRC1 but that H2AK119u1, once established, functions as a signal for ongoing maintenance of PRC1 occupancy. That could be mediated by RYBP as it was shown to bind RING1B and interact with ubiquitinated and non-ubiquitinated H2A through its ubiquitin binding domain (Arrigoni et al., 2006). To test this, the reversible proteasome inhibitor MG132 (Dantuma et al., 2006) was used to deplete H2AK119u1 in *Eed*4 cells both in the presence or absence of PRC2. It was then determined whether restoration of H2AK119u1 occurs following withdrawal of the inhibitor. As shown in Figure 3.7a treatment of cells with MG132 for 6h efficiently depleted global H2AK119u1 (lanes 2 and 6). Allowing cells to recover for 3 days after withdrawal of the inhibitor resulted in restoration of H2AK119u1 both in the presence and absence of H3K27me3 (lanes 4 and 8 respectively). Because MG132 treatment inhibits the proteasome pathway it will have broad effects on cellular processes. For this reason cells were normally recovered for 3 days before analysis was carried out. Nevertheless, results were confirmed after only 24 h recovery as shown in Figure 3.8. ChIP analysis demonstrated that H2AK119u1 accumulates appropriately at PcG target loci after recovery, with levels being only slightly reduced in the absence of H3K27me3 (Figure 3.7b).

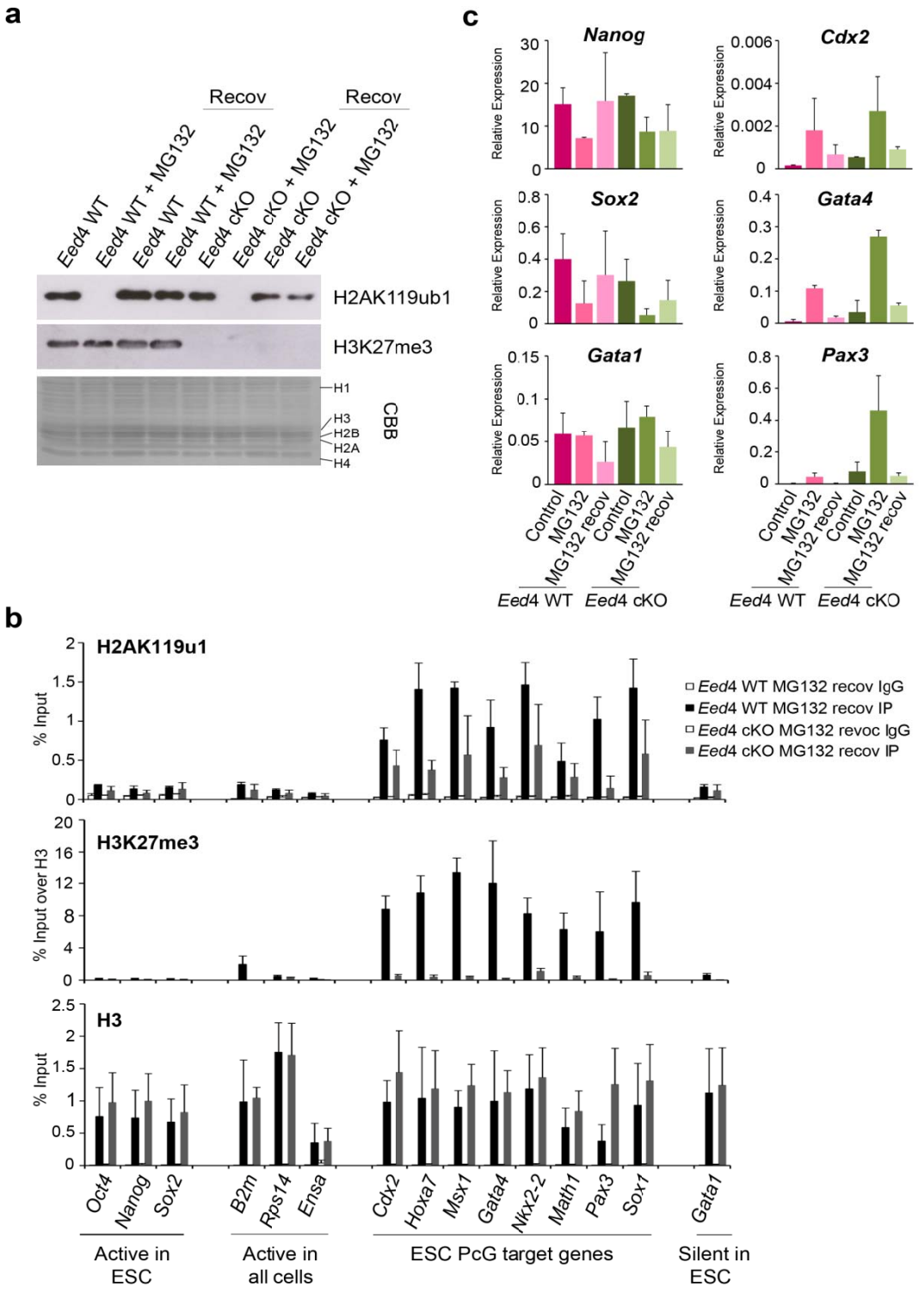



**Figure 3.7 – H2AK119u1 is established *de novo* following depletion in WT and *Eed*4 cKO mESCs.** (**a**) Western blot for H2AK119u1 and H3K27me3 on histone extracts. Coomassie brilliant blue (CBB) staining is used as a loading control. Following treatment for 6 h with 10 µM MG132 H2AK119u1 is completely depleted and is then restored when cells are left to recover for 3 days after removal of the inhibitor, irrespective of presence of H3K27me3. (**b**) ChIP for H2AK119u1 and H3K27me3 in *Eed*4 WT and *Eed*4 cKO cells. In recovered cells (3 days after removal of MG132), H2AK119u1 is correctly targeted in both

*Eed*4 WT and *Eed*4 cKO. H3 is shown as a control. Bars show average + s.d., n=3. (**c**) Expression analysis of selected loci in *Eed*4 WT and *Eed*4 cKO before, during and following MG132 treatment and recovery. Expression is normalized against the average of 3 housekeeping genes, *Hmbs*, *Gapdh* and *Idh2*. Bars show average + s.d., n=3.

Expression analysis illustrated in Figure 3.7c demonstrated that treatment with MG132 derepresses PcG target loci both in the presence and absence of H3K27me3 and that silencing is restored following withdrawal of the inhibitor. No effect was seen at the *Gata1* locus at which repression is PcG independent. Derepression of PcG targets was higher when H2AK119u1 was absent, compared to *Eed* depletion alone (Figure 3.7c, columns 1 and 4) but was enhanced further in the absence of both H3K27me3 and H2AK119u1, consistent with lower H2AK119u1 when MG132 treatment was initiated (see Figure 3.5). Levels of expression of pluripotency factors were reduced following MG132 treatment, consistent with initiation of differentiation due to H2AK119u1 depletion (Endoh et al., 2008; Stock et al., 2007). Taken together these results demonstrate that *de novo* deposition of H2AK119u1 in ESCs in the presence or absence of H3K27me3 occurs appropriately at selected PcG target loci.

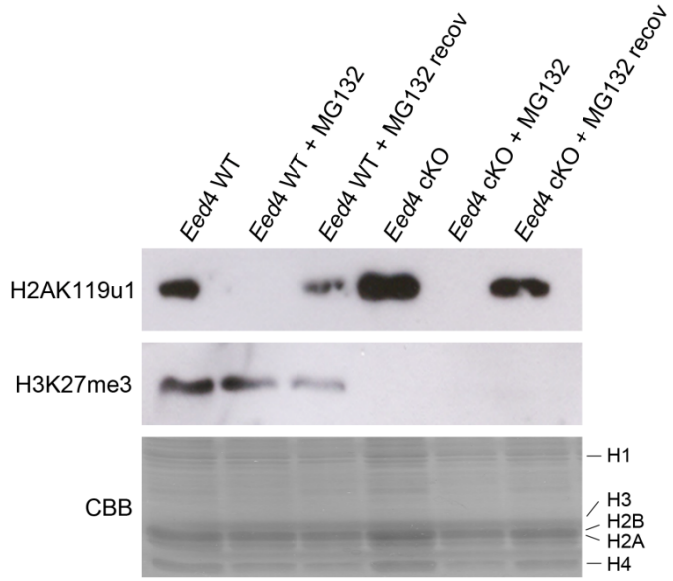



**Figure 3.8 – H2AK119u1 levels are re-established 24 h after removal of MG132 inhibitor from the culture medium.** Western blot on histone extracts shows H2AK119u1 levels before treatment, during 6 h treatment with 10 µM MG132, and 24 h after removal of MG132, both in *Eed*4 WT and *Eed*4 cKO ES cells.

## 3.3. Discussion

### 3.3.1. Recruitment of PRC1 and H2AK119u1 in *Eed* depleted cells

Based on evidence that PRC1 and PRC2 co-localise, together with Pc chromodomain affinity for trimethylated K27 residues a model was proposed in which PRC2 and PRC1 are suggested to be recruited hierarchically to target loci (Cao et al., 2002; Czermin et al., 2002; Fischle et al., 2003; Kuzmichev et al., 2002; Min et al., 2003; Wang et al., 2004b). The ChIP results presented here partially corroborate the hierarchical recruitment model. Core PRC1 proteins, RING1B and MEL18, are not detected in *Eed* depleted ESCs (Figure 3.2 and 3.5). However, H2AK119u1 levels remain unchanged (Figure 3.2 and 3.5). It could be argued that constitutive knockout cells adapt in culture and gain compensatory mechanisms that would maintain H2AK119u1. Results obtained with conditional knockout cells argue against this and strengthen the idea that H2AK119u1 activity is targeted independent of PRC2. In $Eed^{-/-}$ ESCs, ubiquitination is catalyzed at normal PcG target genes with enrichment levels similar to the WT (Figure 3.2 and 3.5). This unexpected result shows that H2AK119u1 is correctly targeted not only on the Xi (Schoeftner et al., 2006) but also on autosomal gene loci.

The demonstration that H2AK119u1 is detected at target loci despite the absence of H3K27me3 may explain the different phenotypes of PRC2 knockout ESCs and *Ring1A/B* mutant ESCs (section 3.1.1.1). Specifically these results suggest that continued recruitment of H2AK119u1 activity following removal of H3K27me3 maintains silencing of PcG target genes in PRC2 knockout cells, at least in part.

### 3.3.2.*de novo* monoubiquitination of H2A

It could be hypothesised that in $Eed^{-/-}$ embryos residual maternal PRC2 is sufficient to target PRC1 and once H2AK119u1 was established it could be self-maintained in ESCs that are derived from these embryos. Utilising a conditional knockout in combination with chemical inhibition

of H2AK119u1 the mechanisms of PcG recruitment could be more thoroughly addressed. The results presented here provide the first evidence that H2AK119u1 can be mediated *de novo* in the absence of PRC2. MG132 blocks the proteolytic activity of the 20S core of the proteasome and has been previously used to deplete H2AK119u1 levels due to an accumulation of high molecular weight ubiquitinated proteins and reduction of free ubiquitin groups in the cytoplasm (Dantuma et al., 2006; Mimnaugh et al., 1997; Xu et al., 2004). Previous studies in mESCs demonstrate that its use for a limited time (8 h) does not significantly affect cell viability, cell morphology and propensity for differentiation (Szutorisz et al., 2006). The data obtained in this study show that H2AK119u1 levels are completely depleted from ESCs when it is added to the medium (Figure 3.7 and 3.8). Most importantly, once the proteasome inhibitor is removed from the medium, H2AK119u1 levels are restored (Figure 3.7 and 3.8) and placed correctly at PcG target genes (Figure 3.7). It follows that H2AK119u1 is not self-maintaining in the *Eed* depleted cells. In fact, because H2AK119u1 is restored after complete depletion it can be concluded that H2AK119u1 can occur *de novo* and independently of H3K27me3. In addition, ubiquitination seems to be completely independent of PRC2 as besides *Eed* depletion, SUZ12 and EZH2 levels are also greatly decreased (Figure 3.4). This result is in agreement with previous observations (Pasini et al., 2004; Schoeftner et al., 2006) and reflects the importance of PRC2 core components in stabilising the complex.

### 3.3.3. Role of PRC1 and PRC2 in gene silencing

Since early work in *Drosophila* and mammals, there are many examples that emphasise the role of PRC2 and PRC1 in gene silencing (Simon and Kingston, 2009). Both complexes have been described to act through hierarchical recruitment with H2AK119u1 being the crucial silencing mark. This is consistent with reports pointing towards an important role for this modification in RNAPII elongation inhibition (Breiling et al., 2001; Dellino et al., 2004; Stock et al., 2007; Zhou et al., 2008). Results presented here indeed highlight the importance of

H2AK119u1 for PcG gene silencing as depletion of H2AK119u1, in addition to *Eed*, upregulate polycomb targets more than 3 times in comparison with *Eed* depletion alone (Figure 3.7c). The results also stress the difference between PRC1 and PRC2 in regulation of gene expression, emphasizing an independent recruitment of both complexes.

Taken together with the evidence that proteasome inhibition leads to activation of cryptic promoters (Szutorisz et al., 2006), results presented here suggest a role for H2AK119u1 in the inhibition of permissive transcription from PcG targets in mESCs. Despite the fact that the proteasome binds to target genes (Szutorisz et al., 2006), evidence that H2AK119u1 is lost when the proteasome is inhibited (Dantuma et al., 2006; Mimnaugh et al., 1997; Xu et al., 2004) and transcription up-regulated (Figure 3.7c) (Szutorisz et al., 2006), together with the link between H2AK119u1 and RNAPII elongation inhibition (Breiling et al., 2001; Dellino et al., 2004; Stock et al., 2007; Zhou et al., 2008) suggests the possibility that is rather H2AK119u1 than the proteasome having a role in silencing intergenic transcription, as suggested by Szutorisz at al., 2006 (Szutorisz et al., 2006).

Recently it was suggested that in ESCs it is RING1B protein rather than the H2AK119u1 modification that is the key player in keeping chromatin compact and maintaining gene silencing (Eskeland et al., 2010). Data presented here would argue rather that H2AK119u1 causes upregulation of PcG target gene expression (Figure 3.7c). This could only happen if H2AK119u1 plays an important role in gene silencing. *Ring1B* single KO cells studied by Eskeland and colleagues (Eskeland et al., 2010) cannot be used to make definitive conclusions about the role of H2AK119u1 in gene silencing as low but detectable levels of this modification can be observed in *Ring1B* depleted cells due to the action of *Ring1A* as an E3 ubiquitin ligase (de Napoles et al., 2004; Endoh et al., 2008).

# 4. H3K27me3 independent recruitment of RING1B

## 4.1. Introduction

It is commonly accepted that in *Drosophila Melanogaster* PcG proteins are recruited to PREs. This process is mediated by several DNA binding proteins such as Pho, Phol, GAGA factor and DSP1 (Morey and Helin, 2010; Ringrose and Paro, 2007). In mammals, despite the efforts of several laboratories, very little is known about polycomb recruitment. Genome wide studies on PcG protein distribution may help to elucidate recruitment sites, sequences, and/or other proteins involved.

### 4.1.1. Genome-wide distribution of PRC2 and PRC1

Genome-wide ChIP-on-chip studies carried out in mouse and human ESCs and MEFs identified over 500 sites that are targets of several PcG proteins (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006). The main findings from these studies were: there is a clear overlap between PRC2 and H3K27me3, and bound DNA sequences fall into highly conserved regions. In addition a strong bias towards developmental regulated genes has been detected. Despite the expectation that such studies could help to find regulatory regions or binding sites for PcG or PcG recruitment proteins, no such sequences were identified. ChIP-on-chip was subsequently superseded by ChIP-Seq as the method of choice for genome wide studies. The resulting profiles have more depth, results are more robust, and hybridization problems and array selection bias are avoided. However, ChIP-Seq studies into PcG localisation again come up with broadly the same conclusions as earlier ChIP-Chip studies. Ku and collaborators (Ku et al., 2008) used ChIP-Seq to analyse genome-wide distribution of PRC2 components SUZ12 and EZH2 as well as PRC1 component RING1B, together with histone modifications (H3K4me3, H2K27me3 and H3K36me3) in mouse and human ESCs. In this study more than 2500 bivalent

promoters containing both H3K4 and H3K27me3 were identified. These promoters could be divided into three classes: bound by both PRC2 and PRC1, bound by PRC2 only (exhibiting H3K27me3 but no PRC1), or not bound by PRC1 or PRC2 (22 %). PRC2 bound promoters overlap extensively with H3K4me3 and H3K27me3 while PRC1 occupies only a subset of bivalent promoters. Interestingly, regions bound by both complexes show more extensive H3K27me3 regions and are evolutionary more conserved, corresponding to promoters of developmental genes at which bivalency is more efficiently preserved during differentiation. On the other hand PRC2-only regions encompass smaller areas that are poorly retained through differentiation. It should be noted that, absence of PRC1 at these sites may in part reflect the depth of coverage in the analysis and criteria used to define peaks. Despite the failure to identify sequence elements able to recruit PcG proteins, a correlation between bivalent domains/PcG bound regions and CG-rich DNA was identified in this study.

### 4.1.2. PcG proteins in flies and mammals

In flies, initial ChIP data showed that, with the exception of Pc protein, PRC1 and PRC2 proteins bind narrow regions that represent PREs. H3K27me3 and the PRC1 protein Pc were present in wider domains (Negre et al., 2006; Schwartz et al., 2006). Interestingly, the employment of a different technique not involving crosslinking, DNA, adenine methyltransferase identification (DamID), allowed the detection of PRC1, PRC2 and H3K27me3 co-localizing at the same domains (Tolhuis et al., 2006). This difference may reflect high mobility of PcG proteins precluding their capture by crosslinking fixation (Ficz et al., 2005). All of the aforementioned studies agree that PcG proteins bind to genes encoding transcription factors, but also genes coding all major developmental and signalling pathways (Negre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). Contrary to initial hypothesis that PcG function was only to regulate embryonic patterns of gene expression during development, analysis of PcG distribution in

*Drosophila* embryo and differentiated cells showed that PcG proteins are dynamic being able to be displaced from embryonic targets and be recruited to new genes (Negre et al., 2006).

In mammals initial studies also highlighted the importance of PcG proteins in maintenance of embryonic pluripotency (Boyer et al., 2006). Nevertheless PcG proteins were soon detected in human fibroblasts where PRC2, PRC1 and H3K27me3 were found to co-occupy more than 1000 sites (Bracken et al., 2006). It was then hypothesised that PcG targets in ESCs could undertake two different pathways with differentiation. On the one hand genes actively transcribed in differentiated cells would lose PcG binding, on the other hand, genes repressed during differentiation would remain bound by PcG from ESCs to differentiated cells (Bracken et al., 2006). Those observations were corroborated by comparison of human ESCs with primary differentiated muscle cells (Lee et al., 2006) and mouse ESCs with neural progenitors and primary fibroblasts (Ku et al., 2008). A recent genome-wide study of BMI1 and H2AK119u1 in mouse MEFs found BMI1 dependent H2AK119u1 both overlapping and not overlapping with H3K27me3, suggesting an alternative mechanism for BMI1 targeting in MEFs (Kallin et al., 2009).

In the previous chapter evidence was presented demonstrating H2AK119u1 targeting to PcG bound genes in the presence of very low RING1B levels. It was suggested that H2AK119u1 was indeed catalysed by RING1B but its detection was not possible by conventional ChIP. Several hypotheses could be put forward to explain the absence of MEL18 and RING1B detected by ChIP in *Eed*$^{-/-}$ ESCs, despite the maintenance of global RING1B levels detected by western blot. MEL18 absence could be explained either by the presence of another homologue (Figure 1.6b), fulfilling its role as the E2 ubiquitin-conjugating enzyme, or it could be another complex altogether, rather than PRC1, catalysing H2A ubiquitination (Weake and Workman, 2008). In mice there are two ubiquitin ligases, *Ring1A* and *Ring1B*, so absence of RING1B could be

explained by RING1A fulfilling the role of ubiquitinating H2A in the absence of *Eed* (de Napoles et al., 2004). The latter hypothesis is unlikely as *Ring1B*$^{-/-}$ ESCs have substantially reduced H2AK119u1 levels, much more than in *Eed*$^{-/-}$ ESCs (de Napoles et al., 2004). It has not been possible to directly test this hypothesis due to a lack of antibodies able to efficiently Immunoprecipitate (IP) RING1A. An alternative explanation may be that H3K27me3 has a role in RING1B spreading similar to EED aromatic cage mutants that have reduced H3K27me3 spreading (Margueron et al., 2009). If that was the case RING1B may not have been detected because the analysis was carried out on very narrow and specific regions by qPCR analysis of ChIP. High resolution techniques were therefore used to increase the depth of the analysis and give a genome wide profile that would answer the question of whether PRC1 could be recruited independently of H3K27me3.

## 4.2. Results

### 4.2.1. ChIP-Seq data analysis

Deep-sequencing of DNA immunoprecipitated with a RING1B antibody from *Eed* WT and *Eed*$^{-/-}$ ESCs was performed. An Input sample was also sequenced as a control for sequencing bias. A very similar profile was obtained for RING1B in *Eed*$^{-/-}$ and *Eed* WT ESCs but with a much lower enrichment in the mutant cells. To obtain a similar enrichment level it was necessary to increase the depth of the analysis and two more lanes of deep-sequencing were performed with a total number of 70 million reads for the *Eed*$^{-/-}$ sample, compared to 17 million reads for the *Eed* WT sample. To have a similar control, the input sample was also further sequenced giving a total of 52.5 million reads. ChIP-Seq data analysis was performed in collaboration with Steve Taylor (Weatherall Institute of Molecular Medicine, Oxford, UK). The complete dataset can be viewed at ww.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE23716. A general analysis of the different datasets was performed, to understand the overall distribution of reads and to detect if there was any bias. The program SeqMonk was used and probes over contiguous sets

of data were automatically generated. Each probe was associated with a value that depends on the pattern of reads. As can be observed in Figure 4.1a top panel, the probe value histograms show a normal distribution as expected for an unbiased dataset. Box-whisker plots show the higher enrichment detected in the *Eed* WT sample as it is the sample with the highest median value.

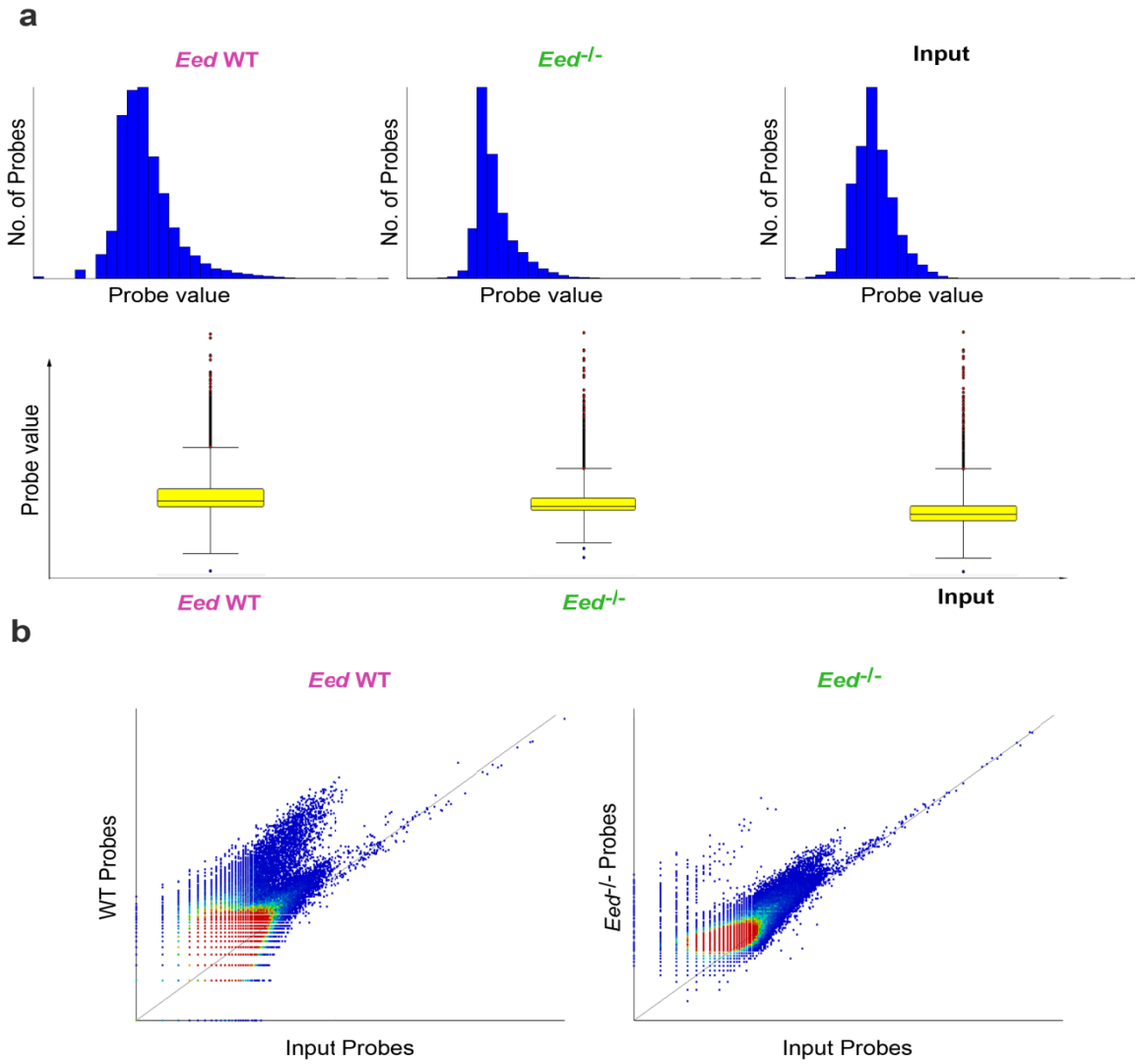



**Figure 4.1 – ChIP-Sequencing data shows a normal probe distribution, without any bias towards Input.** (**a**) Probe value histograms for *Eed* WT, *Eed*$^{-/-}$ and Input (top panel) represent the total number of probes per probe value. They represent a normal distribution of the data. Probes were automatically generated over contiguous sets of data. To each probe was associated a value dependent on the patterns of reads. Lower panel represents box-whisker plots comparing the three datasets. They represent the distribution of probe values where the line across the middle of the yellow box corresponds to the median and the upper and lower extremities of the box represent the 25$^{th}$ and 75$^{th}$ percentile. The upper and lower black whiskers shows the median plus/minus the interquartile (25-75%) range multiplied by 2 and the

individual points falling outside this range are represented by filled circles. (**b**) Scatter plot representation of *Eed* WT vs Input and *Eed*$^{-/-}$ vs Input contigs (clusters of probes) comparison. It represents a 2-dimensional plot where each point represents a probe with an associated value.

The Input dataset shows the lower probe values, which is expected as no IP was performed in that sample (Figure 4.1a, lower panel). In the scatter plot analysis (Figure 4.1b) each point represents a probe and probes shared by the two samples fall diagonally between the two samples. It can be seen that the majority of probes are unique for *Eed* WT and *Eed*$^{-/-}$, giving confidence to further analysis.

#### 4.2.1.1. RING1B localization

Examples of RING1B occupancy at specific loci are shown in Figure 4.2. Both *Irx2* and *Sox1* are defined PcG targets in ESCs (Mikkelsen et al., 2007) and RING1B occupancy is clearly demonstrable in *Eed*$^{-/-}$ cells (Figure 4.2a). Conversely *Oct4*, a gene that is expressed in ESCs and *Gata1*, a gene that is silenced independently of PcG mediated repression do not show RING1B occupancy either in *Eed* WT or *Eed*$^{-/-}$ ESCs (Figure 4.2a). Comparing the *Eed* WT with *Eed*$^{-/-}$ dataset, the existence of 3 different populations of contigs (clusters of probes) can be detected: the largest population comprises contigs present in both datasets (diagonal between *Eed* WT and *Eed*$^{-/-}$), followed by a population with *Eed* WT only contigs (above diagonal) and lastly contigs present only in *Eed*$^{-/-}$ (below diagonal) (Figure 4.2b).

Using Model-based Analysis of ChIP-Seq (MACS) to identify peaks, 2347 sites (1945 genes) where RING1B is enriched in WT cells were found, in comparison to 1810 sites (1601 genes) in *Eed*$^{-/-}$ cells. Both in *Eed* WT and *Eed*$^{-/-}$, analysis includes promoter and non-promoter peaks. The WT RING1B peak list was also validated by comparison with previous data for WT ESCs (Ku et al., 2008) with all previously identified locations being found in the *Eed* WT ESCs peak list. 862 new regions were identified, probably due to the increase in the number of reads.

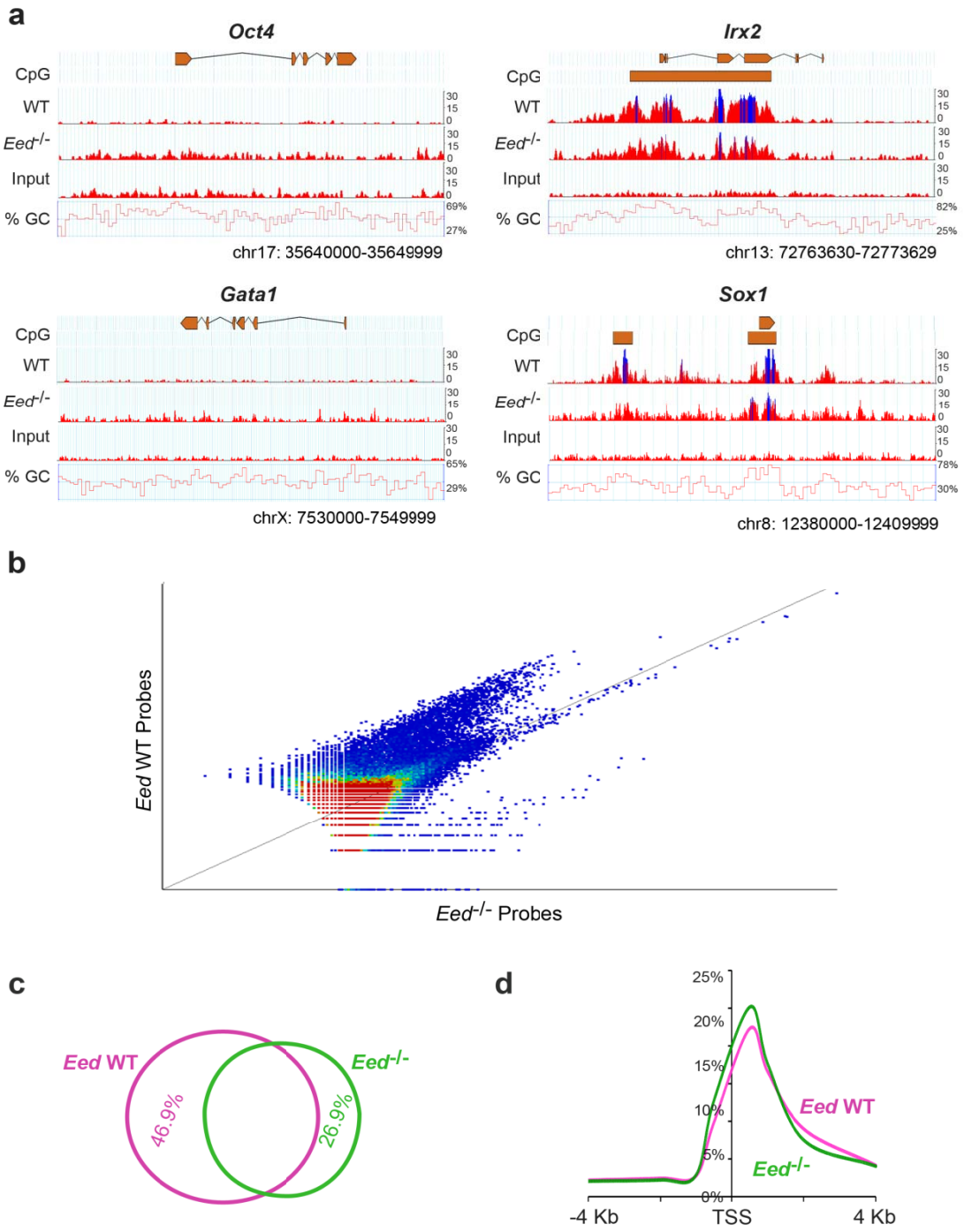



**Figure 4.2 – Analysis of RING1B peaks in *Eed* WT and *Eed*$^{-/-}$ mESCs demonstrates similar RING1B patterns for the majority of RING1B peaks.** (**a**) Example screen shots illustrate gene structure, CpG islands and % GC tracks in addition to distribution of ChIP-Seq reads in *Eed* WT IP and *Eed*$^{-/-}$ Input and IP. Control genes are non-PcG targets *Oct4* (active) and *Gata1* (repressed). PcG targets are *Irx2* and *Sox2*. Enrichment peaks were extracted using a Model-based Analysis of ChIP-Seq (MACS) with a false discovery rate (FDR) < 2 % and number of tags in the peak > 100. The blue bars represent statistically significant peaks. (**b**) Scatter plot comparing contigs identified in *Eed* WT and *Eed*$^{-/-}$ mESCs. The majority of contigs fall into the diagonal line showing the overlap between the 2 datasets. Additionally, two other

groups can be identified, one above the diagonal representing RING1B peaks found only in *Eed* WT sample and the other group below the diagonal representing the few RING1B contigs found only in *Eed*$^{-/-}$ sample. (**c**) Comparison of overlap of peaks identified in WT and *Eed*$^{-/-}$ mESCs where 53 % of WT RING1B peaks are also found in the *Eed*$^{-/-}$ ESCs. (**d**) Peak location in relation to transcriptional start site (TSS). The majority of peaks in both datasets fall between - 0.5 and + 2 kb around the TSS corresponding to promoter regions.

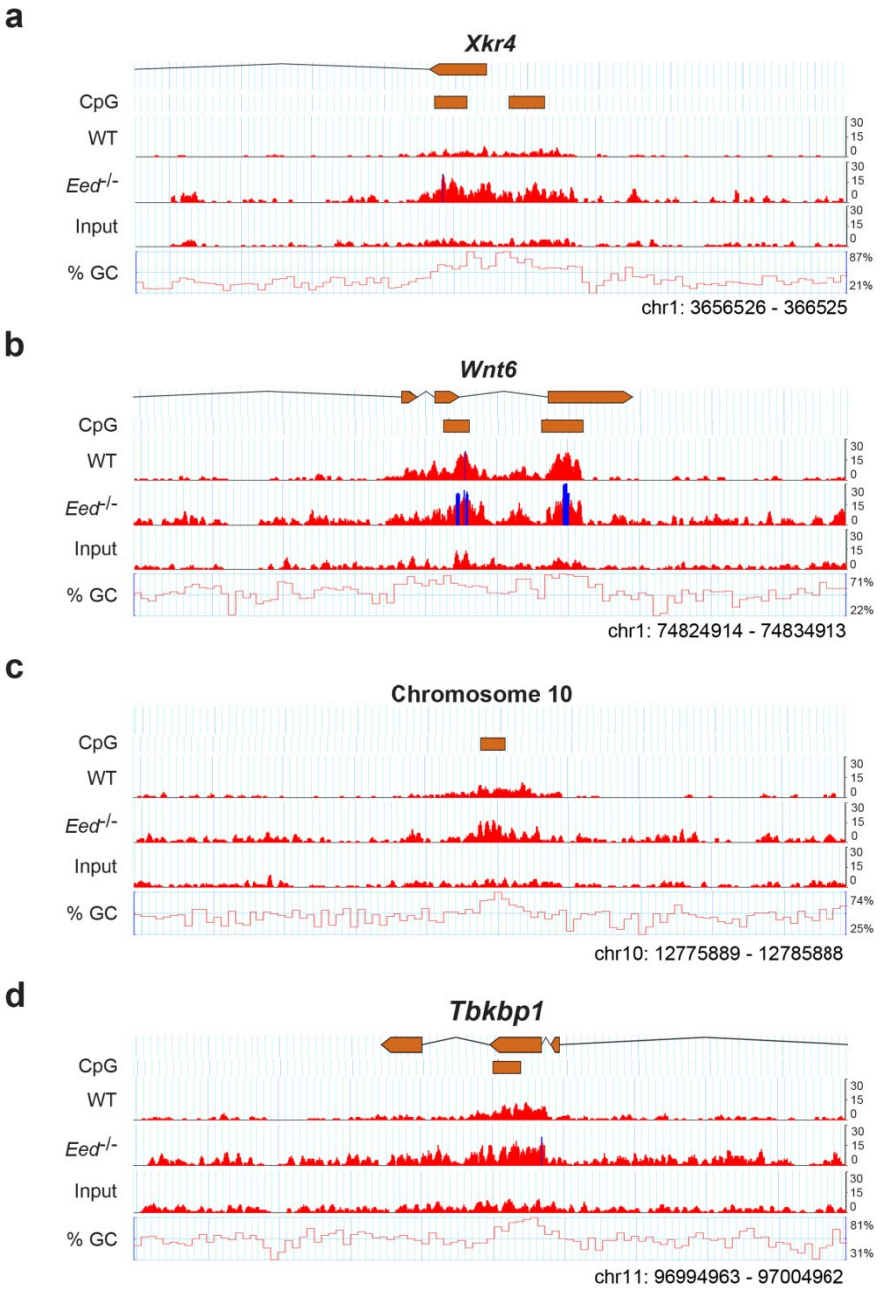



**Figure 4.3 –MACs fails to detect all peaks visible by eye.** Examples are shown where MACs detect peaks in the *Eed* WT dataset but fail to detect them in the *Eed*$^{-/-}$ sets. (**a**) 5' end of *Xkr4* gene; (**b**) 3' end of *Wnt6* gene; (**c**) Intergenic region on chromosome 10; (**d**) 3' end of *Tbkbp1* gene on chromosome 11. Screen shots illustrate gene structure, CpG islands and % GC tracks in addition to distribution of ChIP-Seq reads in *Eed* WT IP and *Eed*$^{-/-}$ Input and IP. The blue bars represent statistically significant peaks.

53 % of regions bound by RING1B in the WT ESCs were also targets in the mutant cells (Figure 4.2c). This is probably an underestimate as the thresholds applied likely fail to detect some RING1B peaks over background in *Eed*$^{-/-}$ cells. Examples are shown on Figure 4.3 where despite efforts to achieve appropriate cut-offs some peaks identifiable by eye are not selected. Random examples of enriched regions at the 5' end of genes (Figure 4.3a), the 3' end (Figure 4.3b and d) and intergenic regions (Figure 4.3c) are shown where MACs identifies *Eed* WT peaks but fails to detect RING1B peaks in the *Eed*$^{-/-}$ dataset. In this study a false discovery rate (FDR) lower than 2 % and a number of tags in the peak higher than 100 was used. This was a compromise as reducing the stringency would increase the number of false positives. When the distribution of peaks across the TSS was compared the percentage of peaks in the *Eed*$^{-/-}$ sample was higher than for *Eed* WT cells within - 0.5 to 2 Kb around the TSS (Figure 4.2d) indicating that the majority of peaks lost in *Eed*$^{-/-}$ cells lie outside promoter regions. Furthermore, gene ontology (GO) analysis demonstrates no significant differences between WT and KO samples (Figure 4.4a), with the majority of genes in both datasets being regulators of transcription.

From evidence presented here it seems that the majority of RING1B bound genes in the *Eed* WT remain bound in the *Eed* KO ESCs. If that is the case, the majority of genes should maintain the expression profile of the WT ESCs. What would then explain the phenotype presented by the PRC2 knockouts? Why are PRC2 KO mice not viable? Analysing GO of genes bound by H3K4me3/H3K27me3 (PRC2 bound genes) and comparing it with GO of RING1B bound genes, it can be seen that a proportion of them are different (Figure 4.4b). PRC2 only genes are enriched for embryonic processes such as embryonic development and morphogenesis which could explain the phenotype observed in mutant embryos.

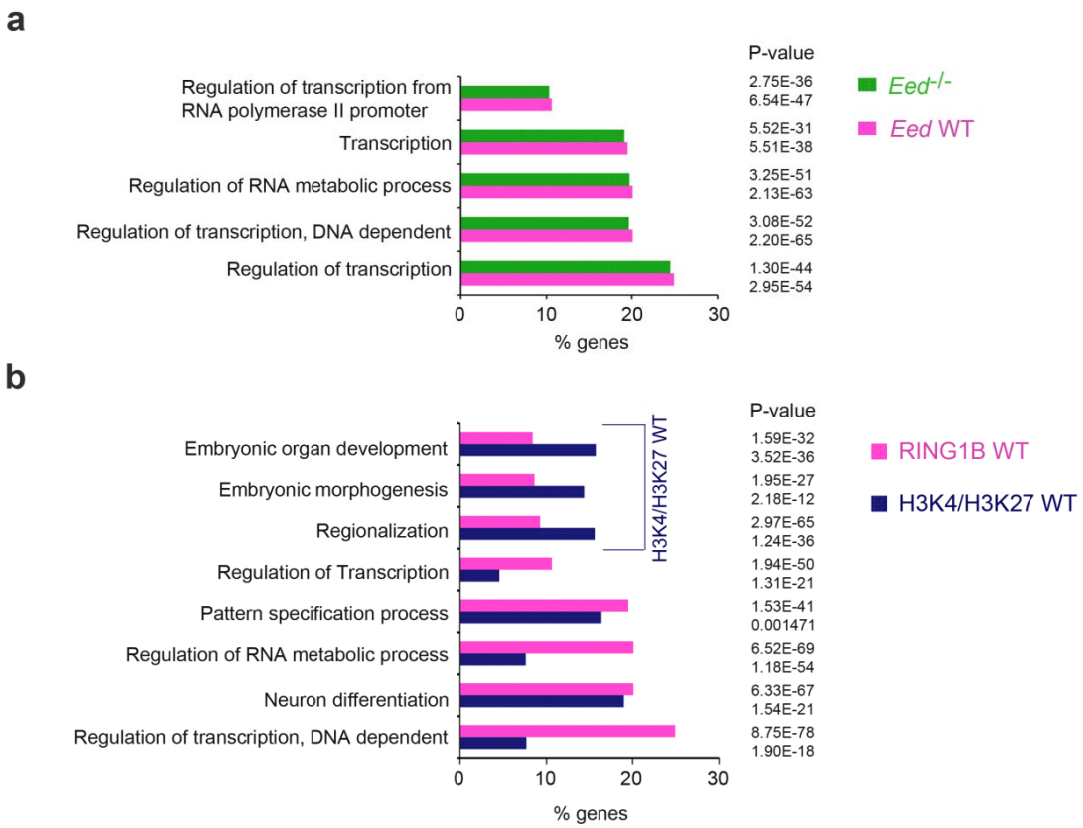



**Figure 4.4 – Gene ontology analysis shows differences between H3K4me3/H3K27me3 bound genes and RING1B genes but not between RING1B genes in *Eed* WT and *Eed*$^{-/-}$ mESCs.** (**a**) Comparison showing similarity in gene ontogeny distribution for RING1B peaks in *Eed* WT and *Eed*$^{-/-}$ mESCs. *P value* < 0.01. (**b**) Graphical demonstration of GO analysis between H3K4me3/H3K27me3 promoter enriched genes and RING1B enriched genes in *Eed* WT mESCs. *P value* < 0.05. Highlighted are categories enriched in H3K4me3/H3K27me3 bound genes.

A recent study suggested that RING1B has a role in silencing genomic repeats, notably endogenous murine leukemia virus (MLV) (Leeb et al., 2010). No evidence for RING1B enrichment at MLV or any other repeat class either in *Eed* WT or *Eed*$^{-/-}$ mESCs was found (appendix 1).

Together with H2AK119u1 detection in *Eed*$^{-/-}$ cells (chapter 3), these results present robust evidence demonstrating PRC1 recruitment at the majority of target sites in ESCs occurs independently of H3K27me3.

### 4.2.2.PRC1 bound regions in *Eed*$^{-/-}$ ESCs

#### 4.2.2.1. RING1B bound genes shared with PRC2

Previously the existence of a small subset of genes in ESCs with RING1B but no H3K4me3/H3K27me3 domains or PRC2 proteins was described. In this study more than 900 RING1B peaks that do not overlap with bivalent domains were found. This difference can be explained by the fact that the bivalent dataset used was from a significantly lower number of reads (Mikkelsen et al., 2007). In addition, the difference may also be attributed to peaks falling outside promoter regions as previous datasets included only bivalent regions corresponding to promoters. PRC2 independent RING1B targets that are genuine should not be affected by *Eed* depletion. As the *Eed*$^{-/-}$ dataset has an average of three times more reads that the *Eed* WT, that should reflect in a three times higher enrichment of PRC2-independent targets. No region was found to show that enrichment level (data not shown). Because H3K27me3 depletion apparently reduces RING1B levels at all target loci, PRC2 seems to be universally important for PRC1 occupancy in ESCs.

#### 4.2.2.2. PRC1 binding overlaps with CGIs

As mentioned above, the broad distribution of RING1B across target loci is retained in *Eed*$^{-/-}$ ESCs. Moreover RING1B distribution closely maps to CGIs associated with target loci. This is clearly illustrated at the *Hoxd* cluster (Figure 4.5a) at which RING1B peaks for a large number of co-regulated genes correlate with the positions of CGIs across the entire locus. Many of the PcG target regions are CGIs, with 99 % of peaks shared between WT and *Eed*$^{-/-}$ cells overlapping with CGIs (Figure 4.5b). The few exceptions that we found (13 peaks) are regions with high GC-content, not annotated as CGIs. This link has already been suggested in the literature although the correlation was less striking (Ku et al., 2008; Mikkelsen et al., 2007).

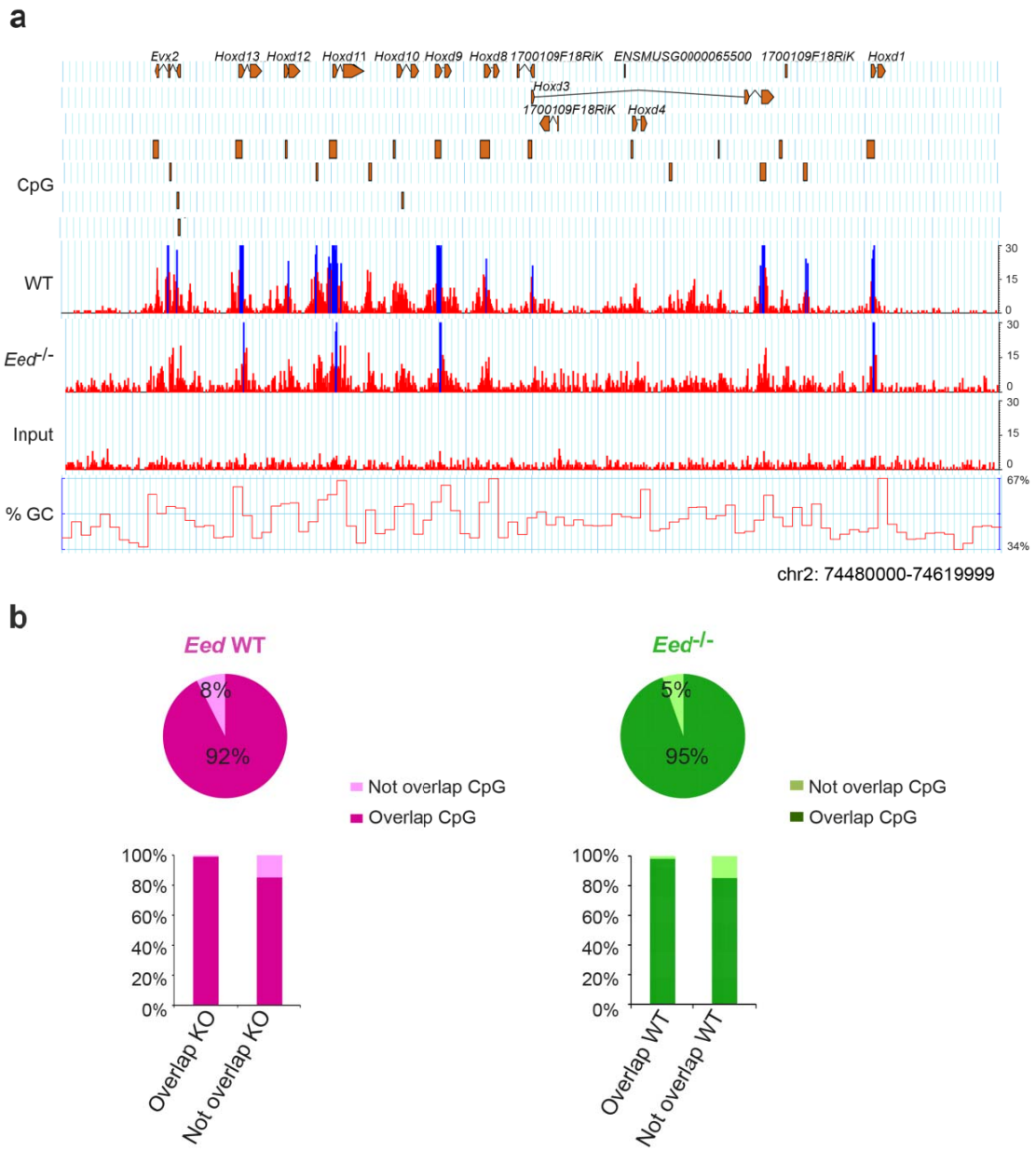



**Figure 4.5 – RING1B enrichment corresponds to CGIs.** (**a**) ChIP-Seq data for the *Hoxd* locus (chr2: 74480000-74619999). CpG islands and % GC tracks in addition to distribution of ChIP-Seq reads in *Eed* WT IP and *Eed*$^{-/-}$ Input and IP are illustrated. (**b**) Overlap of peaks with CGIs for *Eed* WT and *Eed*$^{-/-}$ mESCs. Peaks found both in WT and KO overlap almost 100 % with CGIs.

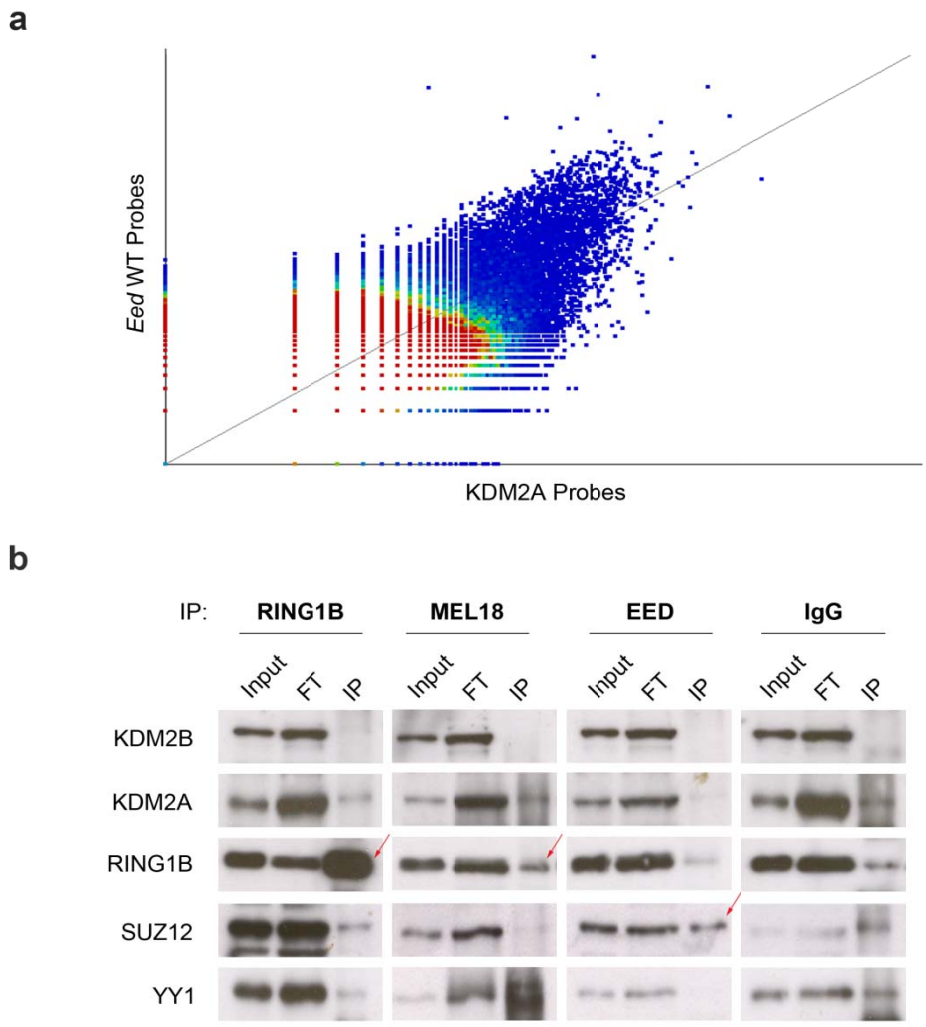



**Figure 4.6 – KDM2A shows some overlap with RING1B but does not Co-IP with PRC1.** (**a**) Scatter plot analysis of RING1B contig probes in *Eed* WT mESCs and KDM2A contig probes in murine V6.5 ESCs (Blackledge et al., 2010). A big number of contig probes fall into the diagonal line showing the overlap between the 2 datasets. Nevertheless RING1B and KDM2A unique probes are identified above and below diagonal line respectively. (**b**) Immunoprecipitation on nuclear extracts with RING1B, MEL18, EED and control antibody IgG. Western blot analysis performed with KDM2A, KDM2B, RING1B, SUZ12 and YY1. Input, Flow through (FT) and Immunoprecipitation (IP) are shown. A low level of KDM2A is found in PRC1 IPs but also in IgG control and therefore looks to be background. Experiment was controlled by Co-IP of PRC1 components (MEL18 - RING1B; RING1B - RING1B) and PRC2 (EED - SUZ12), indicated by arrows, which levels stand out when compared with IgG control.

Recent studies have demonstrated that proteins with CXXC domains localize to CGIs by virtue of interaction between CXXC domains and unmethylated CpG dinucleotides (Blackledge et al., 2010; Thomson et al., 2010). One such protein is KDM2A, a demethylase specific for

H3K36me1/2 (Blackledge et al., 2010). Interestingly, studies in *Drosophila* found that the KDM2 homologue is in a complex with dRing (Lagarou et al., 2008) and similarly that mammalian KDM2B is in a complex with RING1B (Sanchez et al., 2007). ChIP-Seq data for KDM2A (Blackledge et al., 2010) and RING1B was analysed by contig probe generation and comparison of probe datasets. Figure 4.6a illustrates a scatter plot analysis where each dot represents a probe. KDM2A probes clearly overlap with RING1B probes as the majority of probes are represented in the intersection between KDM2A and RING1B (diagonal line). Nevertheless there are many KDM2A regions that do not contain RING1B, as there are many RING1B peaks not bound by KDM2A (Figure 4.6a).

Because KDM2B was found in a complex with RING1B (Sanchez et al., 2007), Co-IP analysis was carried out using nuclear extract from ESCs. As illustrated in Figure 4.6b, KDM2B does not co-immunoprecipitate with PRC1 or PRC2 proteins. A low level of KDM2A is found in PRC1 IPs but also in IgG control and therefore is likely to be due to background. This experiment was controlled by Co-IP of PRC1 components (MEL18 and RING1B) and PRC2 (SUZ12 and EED) (Figure 4.6b) levels of which are significant in the IP compared with IgG control. This result argues against KDM2A and/or KDM2B being involved in the CGIs specific distribution of RING1B in mESCs although the experiment needs repeating and further studies are needed to substantiate this conclusion.

#### 4.2.2.3. PRC1 shows different binding strength to chromatin

The fact that RING1B enrichment occurs at the majority of targets in the absence of H3K27me3 demonstrates that PRC2 is not required to target PRC1. Moreover the distribution of RING1B at target loci is not dependent on H3K27me3 ruling out a role for PRC2 in long-range spreading of PRC1 complexes. However there is a strong reduction in RING1B occupancy in *Eed*$^{-/-}$ ESCs suggesting H3K27me3 may function to stabilize PRC1 occupancy on chromatin and thereby

increase efficiency of silencing. To test this idea, nuclear fractionation was carried out in *Eed*4 WT and *Eed*4 cKO cells to determine if RING1B distribution was affected by loss of H3K27me3. Preliminary results show that in *Eed*4 WT cells RING1B was found predominantly in the chromatin bound fraction (pellet), with low levels present in 100 to 300 mM salt extracted fractions (Figure 4.7 left). In contrast in *Eed*4 cKO cells RING1B distribution shifts away from the pellet fraction into the soluble fractions S2 and 100-300 mM salt (Figure 4.7 right). This result indicates a role for H3K27me3 in increasing RING1B (PRC1) levels on chromatin of target loci.

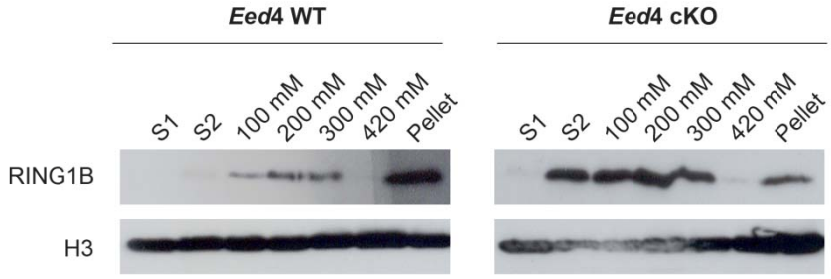



**Figure 4.7 – RING1B binds differently to chromatin in *Eed*4 WT and Eed4 cKO.** Cells were fractionated into different fractions: S1, S2, 100 mM, 200 mM, 300 mM, S3 and Pellet. Western blot analysis was performed with RING1B antibody. H3 was used as control. In *Eed*4 WT cells RING1B was found predominantly in the chromatin bound fraction (pellet) while in *Eed*4 cKO cells RING1B distribution shifts away from the pellet fraction into the soluble fractions S2 and 100-300 mM salt.

### 4.2.3.PcG recruitment independent of RNAPII transcription

Recently several studies suggest a connection between RNAPII transcription and PcG recruitment. In *Jarid2*$^{-/-}$ ESCs, RNAPII is lost from PcG targets promoters and PRC1 targeting and H2AK119u1 levels are also reduced (Landeira et al., 2010). Short RNAs for which transcription is associated with RNAPII transcribed from 5' end of PcG target genes form secondary structures that interact with PRC2 leading to the idea that these transcripts may be important for PcG recruitment in *cis* (Kanhere et al., 2010). With this in mind it was hypothesised that transcription at PcG targets in ESCs targets both PRC2 and PRC1

independent of one another. To test this ESCs were treated with transcriptional inhibitors, whether both in the presence and absence of PRC2/H3K27me3. Inhibitors that interfere with initiation of transcription (α-amanitin, actinomycin D and heat shock) and transcription elongation (DRB) were all tested (Brueckner and Cramer, 2008; Chodosh et al., 1989; Hieda et al., 2005; Kaplan et al., 2008; Mitchell and Fraser, 2008; Sobell, 1985).

#### 4.2.3.1. Effects of transcription inhibitors on PcG silencing in ESCs

α-amanitin had been successfully used to reduce 5' transcription from PcG targets (Stock et al., 2007) and lncRNAs such as *ANRIL* (Yap et al., 2010). In this work, *Eed*4 WT ESCs were treated for 2, 6 and 20 h with α-amanitin and transcription inhibition tested by qPCR. After 2 h treatment highly transcribed RNAs such as *β-actin* and *Mel18* are found at normal levels, presumably reflecting presence of previously synthesised RNA. Transcripts levels then decrease after 6 h and 20 h treatment (Figure 4.8a). Transcription of PcG targets (*Hoxa7* and *Nkx2-2*) and a repressed gene (*Gata1*) also decrease but initial levels of those transcripts were already very low. Global levels of the PcG mediated histone modifications H3K27me3 and H2AK119u1 were unaltered both in *Eed*4 WT and *Eed*4 cKO cells (Figure 4.8b). A slight decrease was observed for H2BK120u1, consistent with the role H2BK120u1 in transcription initiation and elongation (Osley, 2006). ChIP analysis demonstrated that the recruitment of PRC1 and PRC2 PcG proteins and associated histone modifications at target genes (*Hoxa7*, *Nkx2-2*, *Pax3* and *Sox1*) was largely unaffected (Figure 4.8c). *Sox2* (pluripotency gene highly expressed in ESCs) and *Gata1* (repressed gene in ESCs) were used as negative controls for PcG recruitment (Figure 4.8c).

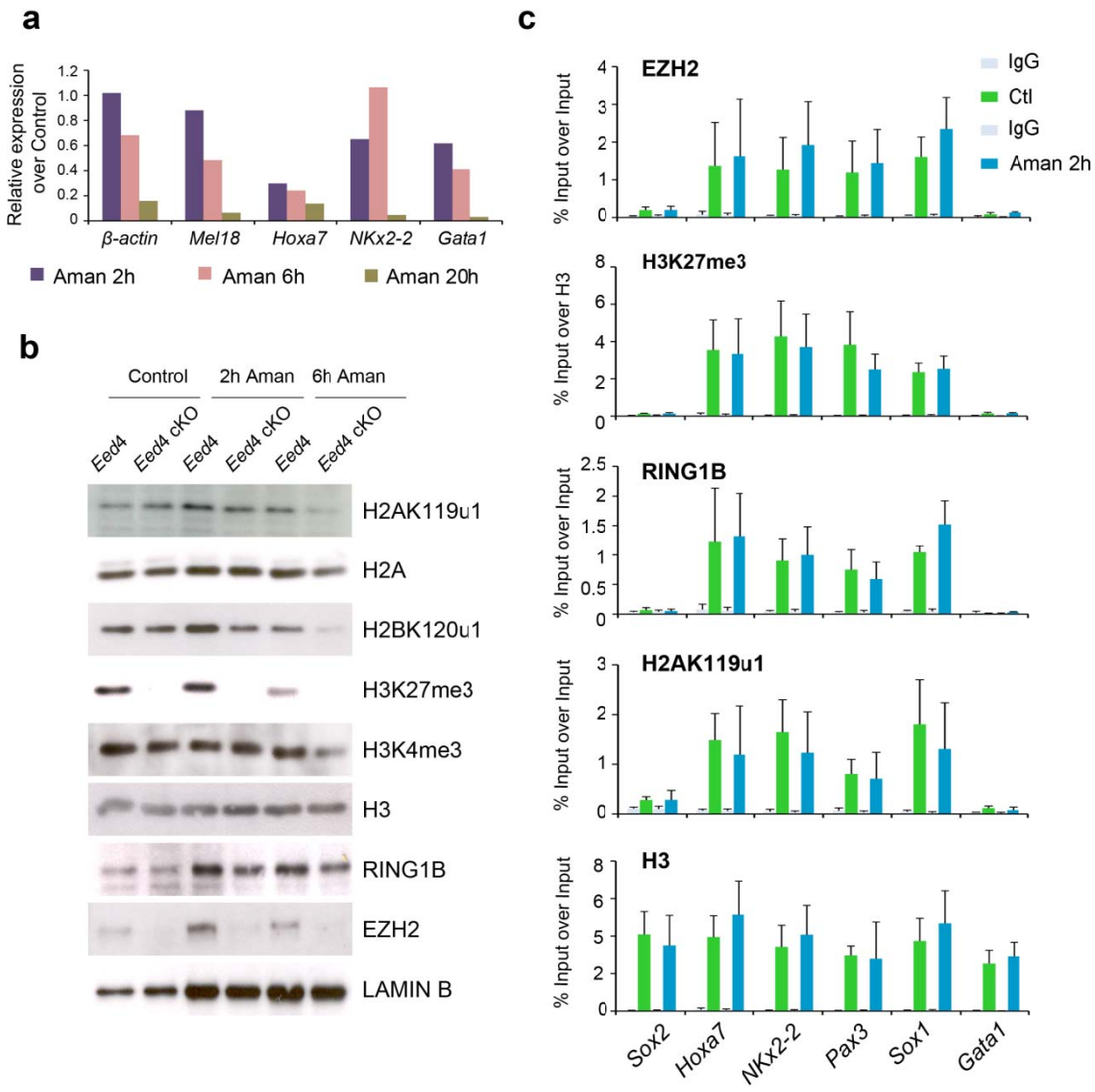



**Figure 4.8 – Recruitment and activity of PRC1 and PRC2 are not affected by α-amanitin (Aman) transcription inhibition of *Eed*4 WT and *Eed*4 cKO ESCs.** (**a**) qRT-PCR analysis of *β-actin*, *Mel18*, polycomb targets genes (*NKx2-2* and *Hoxa7*) and a repressed gene in ESCs (*Gata1*) after 2, 6 and 20 h treatment with α-amanitin. Transcription was normalized to *Eed*4 WT untreated ESCs. Bars represent one experiment with 3 real-time replicas. (**b**) Western blot analysis of *Eed*4 WT and *Eed*4 cKO subject to α-amanitin treatment for 2 and 6 h. Acidic histone extracts were analysed for H2AK119u1, H2BK120u1, H3K27me3, and H3K4me3. H2A and H3 were used as loading controls. Nuclear extract analysis of RING1B and EZH2 was also performed. LAMIN B was used as a loading control for nuclear extracts. (**c**) ChIP analysis in *Eed*4 WT control (Ctl) and treated for 2 h with α-amanitin for EZH2, H3K27me3, RING1B, H2AK119ub1 and H3 as indicated. No changes between treated and untreated samples are observed. Bars show average of at least 3 replicas + s.d. *Sox2* (active in ESC) and *Gata1* (silent in ESC) were used as controls. *Hoxa7*, *Nkx2*-2, *Pax3* and *Sox1* are known PcG target loci.

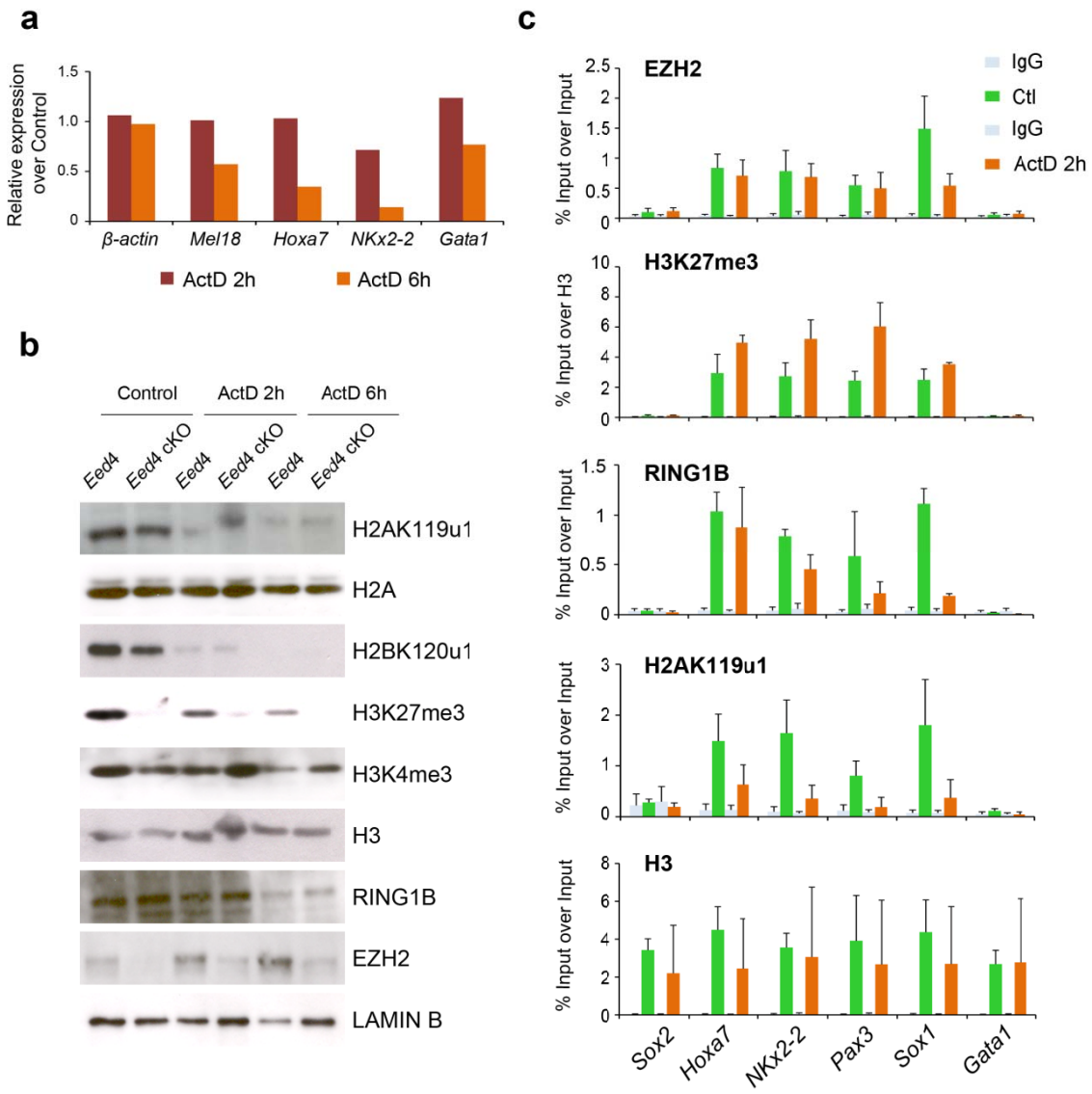



**Figure 4.9 – H2AK119u1 levels are decreased after Actinomycin D (ActD) treatment of *Eed*4 WT and *Eed*4 cKO mESCs but RING1B recruitment to PcG targets is not affected.** (**a**) Transcription analysis of *β-actin*, *Mel18*, polycomb targets genes (*NKx2-2* and *Hoxa7*) and a repressed gene in ESCs (*Gata1*) after 2 and 6 h treatment with ActD. Bars represent one experiment with 3 real-time replicas. Results show expression over *Eed*4 WT untreated ESCs. (**b**) Protein blot analysis of histone extracts showing H2AK119u1 and H2BK120u1 levels decreasing in 2 and 6 h treatment of *Eed*4 WT and *Eed*4 cKO ESCs with ActD. H3K27me3, H3K4me3 remain unchanged. H2A and H3 were used as loading controls. Nuclear extract analysis showed RING1B and EZH2 also unchanged. LAMIN B was used as loading control. (**c**) ChIP analysis in *Eed*4 WT control (Ctl) and treated for 2 h with ActD for EZH2, H3K27me3, RING1B, H2AK119u1 and H3 as indicated. Despite changes in H2AK119u1, RING1B is still detected at chromatin. Bars show average of at least 3 replicas + s.d.. *Sox2* (active in ESC) and *Gata1* (silent in ESC) were used as controls. *Hoxa7*, *Nkx2*-2, *Pax3* and *Sox1* are known PcG target loci.

The results obtained using α-amanitin inhibition were confirmed using two other methods of inhibiting transcription initiation, Actinomycin D (ActD) (Sobell, 1985) and heat shock (HS) (Hieda et al., 2005; Mitchell and Fraser, 2008). For ActD experiments cells were treated for 2, 6 and 20 h although only 2 and 6 h treatment were considered as after 20 h the majority of cells had died. Treatment efficiency was confirmed by analysing levels of specific transcripts (Figure 4.9a). Interestingly after 2 h of ActD treatment a strong reduction of H2AK119u1 levels was observed both globally and at PcG targets (Figure 4.9b and c). Conversely, 2h after ActD treatment RING1B levels were normal (Figure 4.9b) and RING1B was still recruited to target loci (Figure 4.9c). PRC2 levels and H3K27me3 as H3K4me3 levels were not affected by this 2h ActD treatment (Figure 4.9b and c).

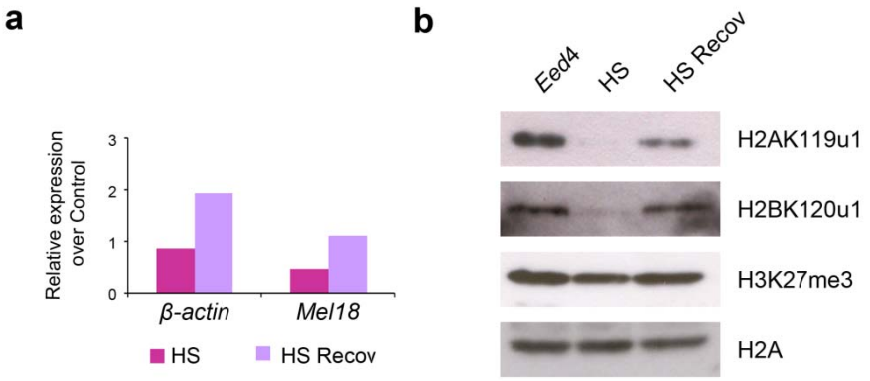



**Figure 4.10 – Heat-shock (HS) treatment affects monoubiquitination of both H2A and H2B in *Eed4* ESCs.** (**a**) qPCR analysis of *β-actin* and *Mel18* after 1 h heat-shock (HS) at 42 °C and recovery after HS for 2 h at 37 °C. Transcription was normalized to *Eed4* WT untreated ESCs. Bars represent one experiment with 3 real-time replicas. (**b**) Acidic histone extracts analysis by western blot show depletion of H2AK119u1 and H2BK120u1 in the *Eed4* WT ESCS treated by HS in comparison with *Eed4* and HS Recovered ESCs. H3K27me3 levels remain unchanged. H3 and H2A were used as loading controls.

Treatment of cells by HS at 42 °C for one hour also inhibits transcription initiation and RNAPII physically dissociates from DNA. Using this approach transcription levels were seen to be reduced (Figure 4.10a) and as seen for ActD treatment, global H2AK119u1 was depleted from cells (Figure 4.10b). This is a reversible phenomenon as recovery of cells for 2 h at 37 °C restored normal levels (Figure 4.10b). In both ActD and HS treatments a similar depletion was

observed for H2BK120u1 suggesting that the mechanism is not H2AK119u1 specific (Figure 4.9b and 4.10b).

DRB is a transcription inhibitor that blocks RNAPII transcription making the transition from initiation to elongation (Chodosh et al., 1989; Hieda et al., 2005). *Eed*4 ESCs were treated for 2, 6 and 20 h with this inhibitor. Levels of specific RNAs were assessed (Figure 4.11a) as were global levels of histone modifications and PcG proteins (Figure 4.11b). As shown in Figure 4.11b, the only modification affected by DRB treatment was H2BK120u1 consistent with the fact that transcription elongation in needed for H2B monoubiquitination (Osley, 2006).

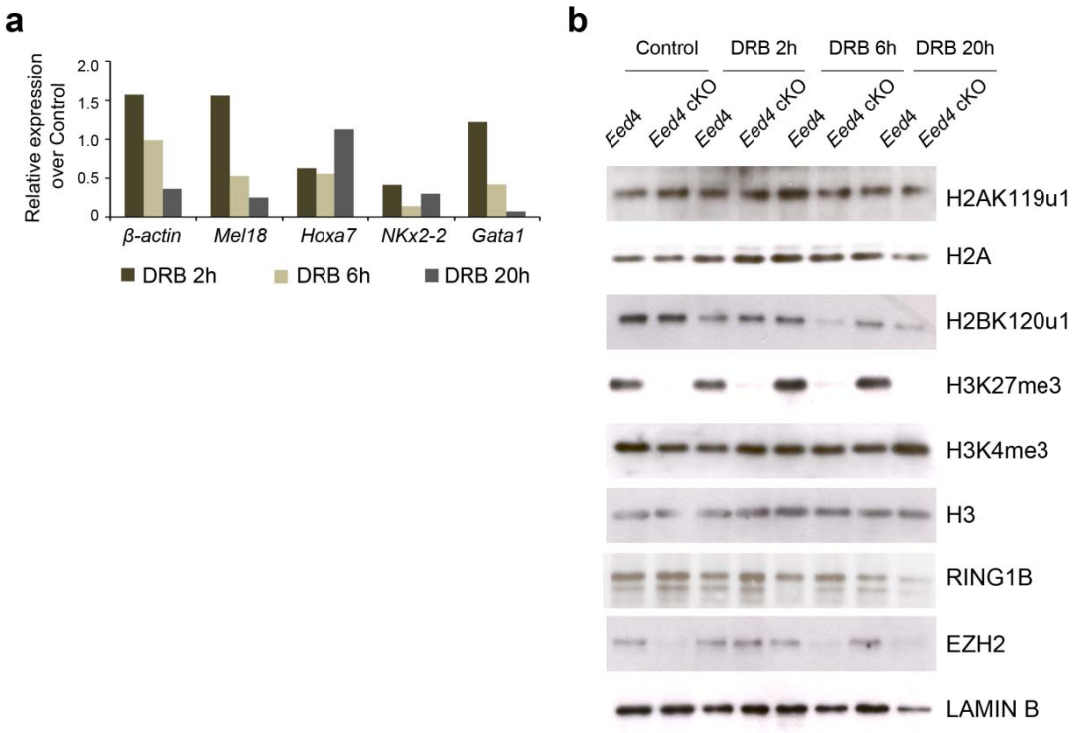



**Figure 4.11 – Transcription elongation inhibition with DRB in *Eed*4 WT and *Eed*4 cKO ESCs affects only H2BK120u1 levels.** (**a**) Transcription analysis of *β-actin*, *Mel18*, polycomb targets genes (*NKx2-2* and *Hoxa7*) and a repressed gene in ESCs (*Gata1*) after 2, 6 and 20 h treatment with DRB. Bars represent one experiment with 3 real-time replicas. Results show expression over *Eed*4 WT untreated ESCs. (**b**) Western blot analysis on histone extracts (H2AK119u1, H2A, H2BK120u1, H3K27me3, H3K4me3 and H3) and nuclear extracts (RING1B, EZH2 and LAMIN B) in untreated, 2, 6 and 20 h treatment with DRB in *Eed*4 WT and *Eed*4 cKO ESCs cell lines as indicated. The only modification affected by DRB treatment is H2BK120u1. EZH2 levels on Eed4 cKO 2h DRB were unexpectedly normal.

#### 4.2.3.2. Transcription inhibition in primary MEFs

Treatments with transcription inhibitors were repeated in a differentiated cell type to evaluate whether results were specific to ESCs. P1 MEFs were used and α-amanitin (2 and 6 h), ActD (2 and 6 h), HS (HS and Recovery) and DRB (2 and 6 h) treatments performed. The results are shown in Figure 4.12. Transcription inhibition was monitored by analysing transcripts levels of *Mel18* (Figure 4.12a). Global levels of H2AK119u1 were reduced following ActD and HS, but not other treatments, as observed in ESCs. On the other hand, H2BK120u1 levels were affected by all the different inhibitor treatments (Figure 4.12b). H3K27me3 levels were unaffected with all treatments. These results demonstrate that transcription inhibition in MEFs has similar outcomes to those observed in ESCs.

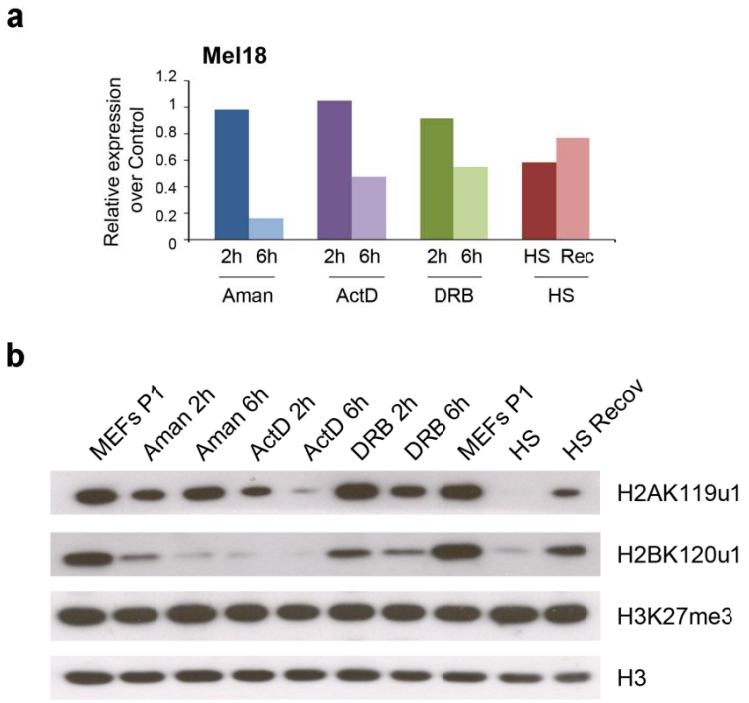



**Figure 4.12 – Transcription inhibitors treatment in MEFs P1 mimics results obtained for ESCs.** (**a**) qPCR analysis of *Mel18* in MEFs P1 after α-amanitin (2 and 6 h), ActD (2 and 6 h), DRB (2 and 6 h) and HS (HS and Recovery) treatments. Transcription was normalized to *Eed*4 WT untreated ESCs. Bars represent one experiment with 3 real-time replicas. (**b**) Western blot analysis of MEFs P1 treated with α-amanitin (2 and 6 h), ActD (2 and 6 h), DRB (2 and 6 h) and HS (HS and Recovery). Acidic histone extracts were analysed for H2AK119u1, H2BK120u1 and H3K27me3 as indicated. H3 was used as loading control.

Overall the transcription inhibitor experiments performed indicate that PRC1 and PRC2 can be recruited independently of both initiation and elongation forms of transcription by RNAPII.

## 4.3. Discussion

In summary, genome-wide analysis of RING1B binding demonstrates that recruitment can occur independent of H3K27me3. In addition a striking correlation has been shown between PRC1 and CGIs that needs to be further analysed. The data suggest a new model for the role of H3K27me3, increasing PRC1 occupancy on chromatin at target loci. Furthermore it was demonstrated using transcription inhibitors that both PRC1 and PRC2 recruitment are independent RNAPII transcription.

### 4.3.1. RING1B recruitment independent of H3K27me3

Previous studies detected reduced levels of RING1B at selected PcG target genes, specifically *Wnt (wingless-type MMTV integration site family, member) 5a*, *Dlk (delta-like 1 homolog) 1* and *Olig (oligodendrocyte lineage transcription factor) 2* in *Eed*$^{-/-}$ mESCs (Boyer et al., 2006; Leeb et al., 2010). Here the authors assumed RING1B was absent and concluded that its recruitment is PRC2 dependent (Boyer et al., 2006). In contrast, Leeb and colleagues find a low level of RING1B present at selected target loci in *Eed*$^{-/-}$ ESCs and suggest PRC1 recruitment may be in part independent of H3K27me3 (Leeb et al., 2010). As has been shown in *Drosophila* studies, methods that do not take into consideration the fast dynamics of PcG proteins (Ficz et al., 2005) fail to detect PcG domains efficiently (Tolhuis et al., 2006). Genome-wide distribution of RING1B in *Eed*$^{-/-}$ cells presented here is in line with recruitment of PRC1 at most, possibly all sites occurring independent of H3K27me3. In addition, normal distribution of RING1B over large domains seen to occur (Figure 4.2a and 4.5a), ruling out the idea that H3K27me3 is important to define the large domains of PRC1 spreading at target genes.

One key factor arguing against the hierarchical model is the contrasting phenotypes between PRC2 and *Ring1A/B* mutants (Kerppola, 2009). In light of results presented here that difference can be understood. RING1B is retained at more than 50 % of PcG targets in *Eed*$^{-/-}$ cells (Figure 4.2), and although levels are strongly reduced, H2AK119u1 levels are still sustained similar to *Eed* WT. In addition GO analysis of PRC2 and PRC1 bound genes may explains why PRC2 deficient embryos die during gastrulation as it highlights the role of bivalent genes (PRC2 bound) in embryonic development with PRC1 predominantly binding to transcription related genes (Figure 4.4). Based on these findings it is concluded that interaction of the chromodomain of the core PRC1 protein Pc with H3K27me3 is not the primary basis for PRC1 recruitment as previously proposed (Cao et al., 2002; Fischle et al., 2003; Min et al., 2003). It cannot be ruled out that H3K27me3 plays a role in recruitment at some loci although this seems unlikely. These surprising results provide a robust explanation for detection of global H2AK119u1 (Leeb and Wutz, 2007) and X-chromosome foci (Schoeftner et al., 2006) in *Eed*$^{-/-}$ ESCs.

### 4.3.2.A role for PRC2 in PRC1 dynamics

Results presented here suggest that the function of H3K27me3 is to increase the time of residence of PRC1 complexes at loci subsequent to primary targeting. The primary targeting of PRC1 is sufficient to establish significant H2AK119u1 levels and maintain silencing at many PcG targets. This provides an explanation for the relatively mild phenotype and limited derepression of target loci in ESCs in which PRC2 core proteins are deleted (Chamberlain et al., 2008; Leeb et al., 2010; Montgomery et al., 2005; Pasini et al., 2007). A decrease in RING1B detected at target loci in *Eed*$^{-/-}$can be argued to be the result of the chromodomain of PRC1 Pc protein having some, albeit low, binding affinity (>10 μM) for H3K27me3 (Bernstein et al., 2006b). This binding may be too low to recruit PRC1, when compared to HP1 binding to H3K9me3 for instance (2 ± 1 μM), but is probably enough to stabilize residency of the complex.

It is worth noting that there is evidence that interactions with a duration of less than 5 s (sufficient for an enzymatic reaction to take place), may not be captured by formaldehyde crosslinking used in methods such as ChIP therefore cannot detect such interactions efficiently (Schmiedeberg et al., 2009) while other methods can (Tolhuis et al., 2006).

### 4.3.3. Role of CpG islands in PRC1 recruitment

Many questions remain; notably what recruits polycomb proteins to chromatin in mammals? Is it a DNA sequence? It is a DNA binding protein? Could it be ncRNAs? Are several mechanisms involved? Many suggestions have been made (Kanhere et al., 2010; Ku et al., 2008), some locus specific PREs have been described (Sing et al., 2009; Woo et al., 2010), but a more general mechanism has not been described. From the data presented here, striking evidence has emerged: the relation between CpG content and RING1B enrichment. In a previous study, BMI1 was found at the same sites as RING1B (Ku et al., 2008), a dataset that entirely matches the WT ESC RING1B data presented here. It is likely therefore that the relationship between RING1B and CGIs extends to all PRC1 components. 92 % of RING1B found in *Eed* WT ESCs overlap with CGIs (Figure 4.5) and data indicates that the remaining 7 % fall into high GC regions not defined as CGIs. Criteria used to define a CpG island define a minimum length of 500 bp, G + C content of 55 % and CpG frequency of 0.65 (Takai and Jones, 2002). This definition eliminated many false positives that were previously annotated as CGIs. Nevertheless the annotation is not ideal as mathematical models do not take into considerations many variations that occur in nature. The remaining 7 % may relate to these false negative regions. New methodologies such as purification of unmethylated regions of the genome followed by deep-sequencing are being used so that in the future the definition of CGI will be refined (Illingworth and Bird, 2009) and a fine mapping of PcG to CGIs will be possible.

Despite the fact that the majority of PRC1 binding sites co-localise with CGIs, the reverse is not true which makes a good argument in favour of PcG recruitment being mediated by a DNA binding protein rather than a DNA sequence. Neither KDM2A nor KDM2B (both binding unmethylated CGIs) interact with RING1B in ESCs (Figure 4.6b). The analysis presented here also demonstrates that RING1B and KDM2A only co-occupy a fraction of their binding regions (Figure 4.6a). This result is in agreement with published work showing that KDM2A binds all unmethylated CpGs regardless of its expression status at a given moment (Blackledge et al., 2010). Further conclusions on KDM2B cannot be drawn as a better antibody needs to be developed and ChIP-Seq analyses carried out. If however KDM2A/B are excluded it remains possible that another factor with a CXXC domain plays a role in PcG recruitment/spreading.

### 4.3.4.PRC1 and PRC2 recruitment independent of RNAPII transcription

Given that PRC1 recruitment is independent of PRC2 and in light of recent studies suggesting a link between PcG recruitment and RNAPII transcription (Kanhere et al., 2010; Landeira et al., 2010), analysis of PcG recruitment in cells treated with transcription inhibitors was performed. Interestingly treatment of ESCs with ActD (Figure 4.9) and HS (Figure 4.10), but not treatment with α-amanitin (Figure 4.8) or DRB (Figure 4.11) decreases global levels of H2AK119u1. This result could be explained by the action of the different drugs; α-amanitin does not inhibit RNAPII recruitment, only slows down its movement across DNA (Brueckner and Cramer, 2008; Kaplan et al., 2008), whereas DRB inhibits elongation to termination but not RNAPII recruitment at promoters (Chodosh et al., 1989; Hieda et al., 2005). Nevertheless recruitment to chromatin of PRC1 core component RING1B was still detected in ActD treated ESCs (Figure 4.9c), the same being true for the PRC2 core component EZH2 (Figure 4.9c), suggesting that both PRC1 and PRC2 can be recruited in the absence of transcription. Three different hypotheses may explain H2AK119u1 levels seen in ActD and HS treatments. First, a post-translational modification of any of PRC1 components, for example MEL18 phosphorylation

(Elderkin et al., 2007) or Ph glycosylation observed to occur in *Drosophila* (Gambetta et al., 2009), may be dependent on RNAPII and despite PRC1 being recruited to target loci it may be catalytically inactive. Secondly a deubiquitinase complex, similar to *Drosophila* CALYPSO (Scheuermann et al., 2010), might have increased activity or accessibility to chromatin in the absence of RNAPII. Strikingly H2BK120u1 levels were also affected both in ESCs and in MEFs in all treatments. That H2AK119u1 levels are depleted with normal recruitment of RING1B to PcG target loci following ActD treatment suggests that the ubiquityl system could be coupled with transcription as previously reported for H2BK120u1 (Minsky et al., 2008; Osley, 2006) or H2AK119u1 and H2BK120u1 (Zhao et al., 2008). Thirdly transcription inhibitors may work as factors of proteotoxic stress, increasing the accumulation of ubiquitinated substrates that coincides with the depletion of free ubiquityl groups and consequently depletion of ubiquitinated histones (Dantuma et al., 2006). These different mechanisms are not exclusive and a combination of both may result in the observations described in this work. Further work needs to be performed to gain a better understanding of the role of transcription in monoubiquitination of histone substrates. Despite the open questions, results presented here show that both PRC1 and PRC2 recruitment are independent of RNAPII transcription both in ESCs and in MEFs.

In summary, four major conclusions can be drawn. First, normal recruitment of RING1B occurs independently of H3K27me3 in more than 50 % of its target genes. Second, a strong correlation between RING1B occupancy, both in *Eed* WT and *Eed*$^{-/-}$ESCs, with CGIs is highlighted. Third, both PRC1 and PRC2 recruitment occur independently of RNAPII transcription both in ESCs and in MEFs. Lastly, a novel function of H3K27me3 is starting to emerge as evidence points towards a role in decreasing PRC1 mobility and increasing its stabilisation on chromatin.

# 5. General Discussion

## 5.1. Polycomb recruitment to chromatin

Polycomb proteins can bind and regulate many developmental genes and for this reason understanding how they are targeted is an important question.

Both PRC1 and PRC2 complexes bind to chromatin even though no individual core component has a DNA binding domain. Genome-wide studies have provided large amounts of data which show an overlap of PRC1 with PRC2 binding sites. This observation argues in favour of a hierarchical recruitment of PRC1 to H3K27me3 chromatin-rich areas (Ku et al., 2008). In this context understanding targeting of PRC2 is sufficient to understand PRC1 recruitment. However a number of indications suggested that hierarchical recruitment does not apply in all cases (see section 3.1.2.2). Evidence presented here clearly demonstrates that in most cases RING1B and H2AK119u1 exhibit normal targeting in *Eed* depleted cells, i.e., PRC1 targeting occurs independent of H3K27me3 (chapter 3 and 4). In previous PRC2 knockout studies both in *Drosophila* (Cao et al., 2002; Wang et al., 2004b) and mammals (Boyer et al., 2006) RING1B and/or other PRC1 proteins were not found at target loci. As shown here in the case of Boyer et al. study (Boyer et al., 2006) it is likely that limited sensitivity of conventional ChIP led to the conclusion that RING1B is absent. Indeed more recent studies on $Eed^{-/-}$ ESCs did report low levels of RING1B at selected target loci although here H2AK119u1 was not analysed (Leeb et al., 2010).

In light of observations presented here it is clear that future studies need to define primary targeting mechanisms for both PRC1 and PRC2 complexes. It is possible that the same factors independently recruit both PRC2 and PRC1 and this would seem to be supported by the high

frequency of co-occupancy and the striking similarity in distribution patterns of PRC1 and PRC2 core components/associated histone modifications. These similar distributions are unusual in that they encompass relatively broad domains that correlate closely with CpG density at the target loci (Ku et al., 2008), seemingly arguing against a simple interaction between PRC1/2 complexes and transcription factors with a defined DNA recognition sequence. Factors such as ncRNAs or DNA binding proteins may have a role in this PRC1 and PRC2 independent recruitment process.

Evidence for trans-recruitment of PcG proteins by long ncRNAs was first provided by studies on *Xist* (Figure 5.1a) (de Napoles et al., 2004; Mak et al., 2002; Plath et al., 2003; Plath et al., 2004; Silva et al., 2003). Subsequently the 2-D structure of the A-repeat region of *Xist* was suggested to mediate the binding with PRC2 and to increase recruitment of SUZ12 establishing a structural link between *Xist* and PcG recruitment (Maenner et al., 2010). A link between PRC2 and *HOTAIR*, a 2.2Kb ncRNA expressed by the *Hoxc* locus has also been characterized. This association is needed for PRC2 occupancy and H3K27me3 with concomitant gene silencing of the *Hoxd* locus (Rinn et al., 2007). Interestingly *ANRIL*, another lncRNA expressed by the *INK4b-ARF-INK4a* locus (Pasmant et al., 2007), was the first ncRNA establishing a link between lncRNAs and PRC1 recruitment. Interaction between *ANRIL*, CBX7 and H3K27me3 is needed to keep the locus silenced (Yap et al., 2010). In addition to long ncRNAs, a recent study has revealed a family of short ncRNAs transcribed at PcG target loci in lymphocytes and in mESCs that may have a role *in cis* in recruitment/maintenance of PRC2 (Figure 5.1b) (Kanhere et al., 2010). Whether PcG proteins are directly recruited by short ncRNAs or as a result of RNA interference pathways (Grimaud et al., 2006a; Kim et al., 2006) is unknown. The existence of nascent transcripts recruiting PcG could also explain the existence of bivalent domains as *Drosophila* ncRNAs were reported to recruit the TrxG protein Ash1 at the *Ubx* gene (Sanchez-Elsner et al., 2006). Interestingly results presented here argue against a link between PcG

recruitment in ESCs and RNAPII or nascent transcripts. Although H2AK119u1 levels were affected by ActD and HS treatments (Figure 4.9b and 4.10b respectively), PRC1 (RING1B) and PRC2 (EZH2) recruitment to chromatin were not affected (Figure 4.9c and 4.10c). Additionally H2BK120u1 levels were also decreased following ActD or HS treatments making a strong argument in favour of an involvement of the ubiquityl machinery (Osley, 2006) or stress induced by transcription inhibitors affecting the proteasome and decreasing histone monoubiquitination levels (Dantuma et al., 2006). Further work needs to be performed in order to clarify which mechanisms or if more than one mechanism is affected by those experiments. Nevertheless PRC1 and PRC2 are normally recruited to chromatin which suggests that nascent transcripts or/and RNAPII are not involved in PcG recruitment in ESCs.

An alternative to ncRNAs targeting PRC1/PRC2 is recognition of a specific chromatin landscape, an idea that is supported by the recent discovery of the role of the non-catalytic jumonji domain protein JARID2 (Figure 5.1c) in PRC2 targeting in ESCs (Landeira et al., 2010; Li et al., 2010b; Pasini et al., 2010a; Peng et al., 2009; Shen et al., 2009). JARID2 is also important for co-recruitment of PRC1 and RNAPII (Landeira et al., 2010). When *Jarid2* is depleted RING1B and MEL18 are still recruited but enrichment levels are decreased. Thus JARID2 recognition of chromatin landscape is unlikely to be the primary recruitment mechanism. Similarly, *Pcl2* a mammalian homologue of *Drosophila* Pcl that forms a sub- stoichiometric PRC2 complex (Walker et al., 2010), may be involved in targeting PRC2 in ESCs via recognition of chromatin landscape because it has two PHD finger domains and a tudor domain, all of which are implicated in chromatin interactions. ChIP-Seq studies (Walker et al., 2010) show co-localization with a subset of PRC2 targets although Knockdown studies indicate only a minor role in targeting in ESCs.

As discussed previously, proteins with a CXXC domain, binding unmethylated CpG dinucleotides may also be important for PcG targeting (Figure 5.1d). In this work evidence is provided that neither KDM2A nor KDM2B fulfil this role in mESCs. A previous study did however detect the binding of KDM2B with RING1B in murine erythroleukemia (MEL) cells that express high levels of PcG proteins (Sanchez et al., 2007). Further work is required to address these contrasting results but a possible explanation may be that interaction is specific for MEL cells.

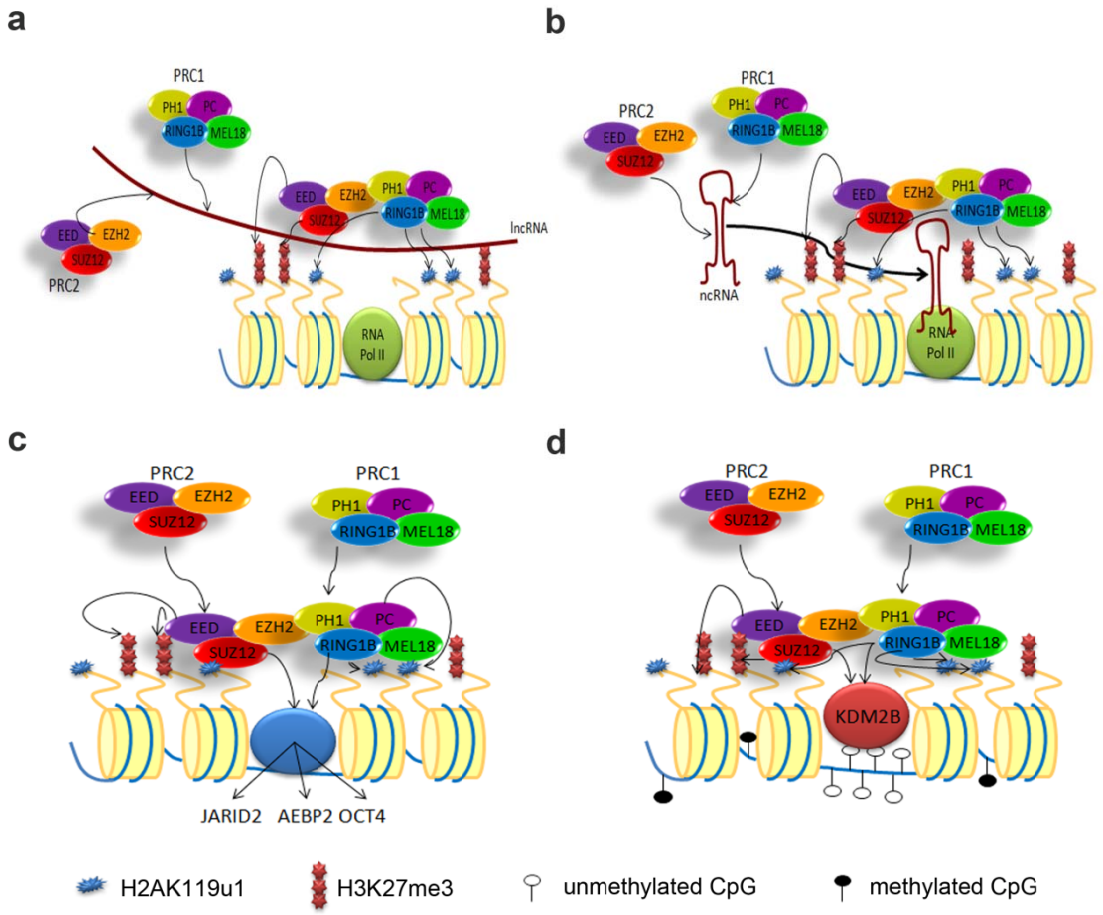



**Figure 5.1 – Possible mechanisms of PRC1 and PRC2 recruitment to chromatin.** (**a**) PRC1 and PRC2 recruitment could be mediated by a lncRNA that target both complexes to the locus to be silenced (e.g. *Xist*, *ANRIL*, *HOTAIR*). (**b**) PcG complexes could be recruited by a small ncRNA transcribed either in another location or at the locus to be silenced. (**c**) Recruitment of both complexes could happen through DNA binding proteins such as by JARID2, AEBP2 or the pluripotency factor OCT4. (**d**) CGIs may have a role in recruitment of PRC1 and PRC2 through the binding of DNA binding proteins such as KDM2B that have a CXXC domain capable of binding unmethylated CpGs.

Since it was found that DNA binding proteins mediate PcG silencing at *Drosophila* PREs (Ringrose and Paro, 2007), huge efforts have made in order to find mammalian equivalents. In *Drosophila* many proteins associate with PREs and interestingly, are able to recruit both PcG and TrxG proteins (Ringrose and Paro, 2004). For example Pho can recruit Pc and E(z) but also the BRM ATP-dependent chromatin remodelling complex through the interaction of 2 different domains of the Pho protein (Mohd-Sarip et al., 2002; Wang et al., 2004b). Despite that, the Pho mammalian homologue *YY1* (yin and yang 1), has only very low levels of co-localisation with PRC2 (Squazzo et al., 2006) suggesting that it is not essential for polycomb recruitment despite the displacement of SUZ12 from target genes and the loss of H3K27me3 when it is knocked down in mouse myoblasts (Thomas and Seto, 1999). Another protein that has been linked to PcG recruitment is AEBP2 (Figure 5.1c). It has been shown to interact with and increase the catalytic activity of PRC2 (Cao and Zhang, 2004). The co-localization of AEBP2 and SUZ12 together with the existence of a DNA binding domain in AEBP2, suggest it as a possible targeting protein for the mammalian PcG (Kim et al., 2009).

Finally, there may be a role for pluripotency factors in polycomb recruitment as their localisation shows a very similar pattern. It has also been shown that RING1B recruitment is *Oct4* dependent in mESCs (Figure 5.1c) as it is lost when cells lose *Oct4* (Endoh et al., 2008). Although the pluripotency network could help to recruit polycomb in undifferentiated cells this would not explain occurrence of Polycomb domains in other cells types (Kallin et al., 2009; Ku et al., 2008).

The different recruitment modes described here are not mutually exclusive and it is plausible that efficient recruitment of PRC1/PRC2 depends on a combinatorial effect of these mechanisms. It cannot be ruled out that PcG targeting results from a sum of weak interactions that on their own are not enough to target. A similar mechanism can be found in the targeting

of the Recombination activating gene 2 (Rag2) protein (Ramon-Maiques et al., 2007). Rag2 PHD domain can only bind H3K4me3 when the adjacent H3R2 residue is dimethylated.

Whatever the mechanism(s) involved in PcG recruitment it needs to be able to respond to developmental signals and different cellular stimuli. This is well illustrated by the *Ink4*/*Arf* locus where senescence is triggered by release of PcG from the locus and upregulation of gene expression while other targets maintain PcG binding (Figure 2.4). What is the mechanism that releases PcG from the locus? How are PcG counteracted? The use of unbiased screens, such as the siRNA screen described in section 2.2.4 could help to answer some of these outstanding questions.

## 5.2. The role of PRC1 in gene silencing

Early studies in *Drosophila* showed that polycomb proteins play a role in regulating gene expression, inferred from the observation that polycomb mutant phenotypes include homeotic transformations associated with *Hox* gene misexpression. Whether both PRC1 and PRC2 complexes are needed for efficient silencing is still an open question. In this study co-localization of both PRC1 and PRC2 with the *Ink4b/Arf/Ink4a* locus in both mESCs and MEFs when the locus is silent (Figure 2.3) was shown. It was also demonstrated that when PRC2 gets deleted silencing is still possible as PRC1 is still recruited and H2A monoubiquitinated at PcG targets in mESCs (chapter 4 and 5).

Until now the PcG hierarchical recruitment model has been widely accepted, and the role of PRC1 in the *Ink4a*/*Arf* regulation has been associated with PRC2 binding and recruitment. Only recently genome-wide localization of BMI1-dependent H2AK119u1 demonstrated that in MEFs H2AK119u1 can be found in genomic loci where H3K27me3 is absent (Kallin et al., 2009). As H2AK119u1 associates with gene silencing in MEFs it suggests that, at least in some regions,

PRC1 only can have a role in gene silencing. Here it was demonstrated that silencing of the *Ink4/Arf* locus reflects PcG recruitment and associated histone modification patterns (chapter 2). Both in mESCs and in primary MEFs the locus is bound with PcG proteins that maintain repression. Activation of the locus occurs concomitantly with both PRC1 and PRC2 displacement from the locus when cells enter senescence (Figure 2.2 and 2.4), underlining the role of PcG in *Ink4b/Arf/Ink4a* regulation in mice (Agherbi et al., 2009; Bracken et al., 2007; Maertens et al., 2009; Miki et al., 2007). Co-localization of PRC1 and PRC2 does not answer the question of whether both complexes are indeed needed for its silencing. This question is difficult to analyse further in MEFs as PRC2 knockout fibroblasts cannot be isolated from mice as the embryos die soon after gastrulation (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004).

If both complexes were required for gene silencing all the genes bound by PRC2 should become derepressed in PRC2 knockout mESCs. As that is not the case, evaluation of gene expression in mESCs depleted in PRC2 and H2AK119u1 was performed by taking advantage of *Eed*4 cKO cells and the proteasome inhibitor MG132 that depletes cells from H2AK119u1 (Dantuma et al., 2006). In Figure 3.7c it can be seen that the highest PcG target derepression happens in cells lacking both PRC2 and H2AK119u1 when compared with derepression when PRC2 only or H2AK119u1 only are lost. This result suggests an important role for H2AK119u1 in gene silencing as previously suggested (Stock et al., 2007; Wang et al., 2004a), but goes against a recently published study that compares chromatin structure in $Eed^{-/-}$ and $Ring1B^{-/-}$ suggesting that it is RING1B rather than H2AK119u1 that is the key to maintain chromatin in a compacted and silent form in mESCs (Eskeland et al., 2010). The authors of this study attribute the role for chromatin compaction to *Ring1B* as they detect a similar decompaction of chromatin both in $Eed^{-/-}$ and $Ring1B^{-/-}$ ESCs. Assuming previously published results where by conventional techniques PRC1 had not been detected in $Eed^{-/-}$ mESCs (Boyer et al., 2006) while H2AK119u1

had already been detected in the same cells (Schoeftner et al., 2006), Eskeland et al., 2010 suggest Ring1B as the key factor for gene silencing. Additionally the authors used *Ring1B*$^{-/-}$ cells and therefore disregard residual H2AK119u1 levels that can still be detected and result from *Ring1A* catalytic activity (de Napoles et al., 2004; Endoh et al., 2008).

Contrary to a major role of PRC1 in gene silencing a recent study demonstrates that both PRC1 and PRC2 play a role in silencing genomic repeats (Leeb et al., 2010) with the authors showing that both complexes can redundantly silence endogenous repeats. H3K27me3 levels were seen to be elevated at endogenous retrovirus. In the work presented here RING1B was not detected at any repeats (appendix 1) which is in line with previous work showing H3K27me3 in transposon free regions (Bernstein et al., 2007). Further studies are required to resolve this apparent paradox.

Given the apparent importance of PRC1 in silencing at PcG targets how can we account for the fact that plants and *Caenorhabditis elegans* have only PRC2 homologues? One possibility is that plants and worms have unknown PRC1 homologues. LHP (Like heterochomatin protein) 1 a chromodomain protein in *Arabidopsis* was found to bind H3K27me3 residues both *in vitro* and *in vivo* (Whitcomb et al., 2007). In this scenario H3K27me3 would be contributing to silencing through a completely independent mechanism compared to that found in *Drosophila* and vertebrates. Bioinformatic analysis in *C. elegans* also uncovered a putative Pc homologue, the *C. elegans* CEC-1(chromobox 1) even though little is known about this factor (Whitcomb et al., 2007). Further work needs to be performed to clearly demonstrate that plants and worms have evolved entirely independent readers of the H3K27me3 histone modification.

## 5.3. PRC1 mechanism of action

It has been shown in *Drosophila* that polycomb proteins are dynamic, having different association kinetics even within the same cell, depending on the chromatin context (Ficz et al., 2005). In the work presented here several lines of evidence are shown that shed some light on the dynamics of PcG binding to chromatin. First H2AK119u1 is detected at almost normal levels at correct polycomb targets despite the absence of RING1B detection by ChIP (Figure 3.3 and 3.5) (Boyer et al., 2006). Second, the fact the RING1B is only detected by increasing the depth of ChIP-Seq and not using conventional qPCR detection provides evidence suggesting that protein dynamics are faster than conventional crosslinking capturing times (Schmiedeberg et al., 2009). Third nuclear fractionation of RING1B shows that it is easier to remove it from chromatin in *Eed* depleted cells than in WT counterparts (Figure 4.7). With this evidence in mind a model of PRC1 regulation is proposed. As shown in Figure 5.2, PRC1 and PRC2 can both be independently targeted to chromatin through the same or different mechanisms, leading to establishment of both H3K27me3 and H2AK119u1. The ability of Pc to recognize the H3K27me3 modification and to bind it, albeit with low binding affinity (Bernstein et al., 2006b), is proposed to increase the association of PRC1 with chromatin, increasing the binding time and the formaldehyde crosslinking efficiency. Either in the presence or absence of H3K27me3, PRC1 is present in chromatin and is able to catalyse H2AK119u1 and to silence gene expression.

Interestingly BiFC analysis in which fluorescent proteins were fused with CBX family members did not show any difference in fluorescence in *Eed*$^{-/-}$ compared to *Eed* WT mESCs. There are however some doubts about the correct regulation of these complexes as they do not co-localize with H3K27me3 staining in the WT cells (Vincenz and Kerppola, 2008). Another study did not detect any differences in CBX protein dynamics when the chromodomain was deleted (Ren et al., 2008). This study was carried out in HEK293T cells with endogenous CBX proteins

still present, that could participate in multimeric complexes with mutant proteins and compensate function. Given the caveats associated with analysis of PcG dynamics future studies should directly examine mobility of fluorescent tagged PRC1 proteins in cells with and without H3K27me3, using for example Fluorescence recovery after photobleaching (FRAP) methodologies. The improvement of technologies that detect fast enzymatic interactions is needed and may change the picture for other chromatin associated proteins for example those that recognize the methyl groups by chromo-like domains - chromo, tudor, MBT and - PHD like domains (Kouzarides, 2007).

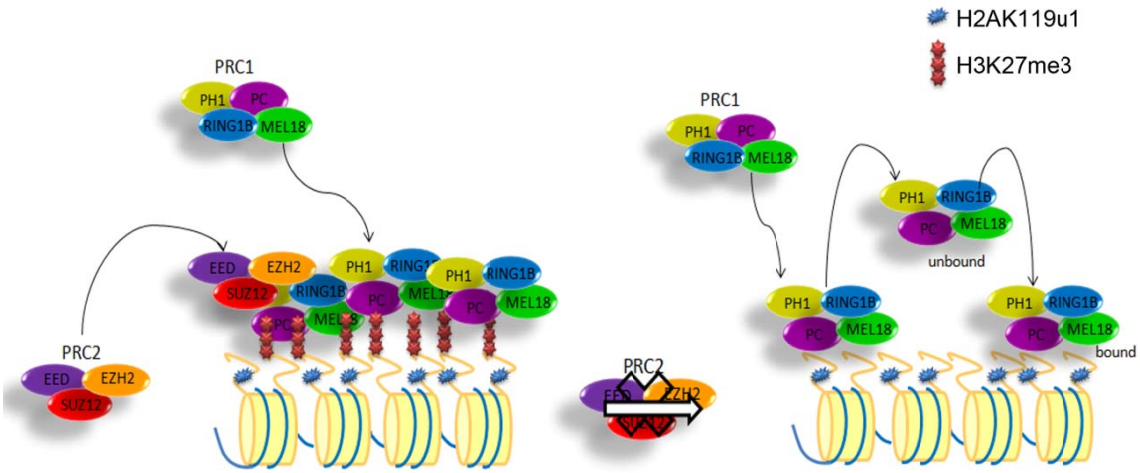



**Figure 5.2 – Schematic model of PRC1 dynamics.** PRC1 and PRC2 are recruited to chromatin and mediate H2AK119u1 and H3K27me3 respectively. Through the binding of Pc protein of PRC1 to H3K27me3, PRC1 is stabilized on chromatin. In the absence of PRC2 and H3K27me3, PRC1 is recruited to chromatin and monoubiquitinates histone H2A on lysine 119. Because H3K27me3 is not present, PRC1 dynamics increase with the complex being found both in bound and unbound forms.

Taking these different studies in consideration, the hypothesis that best fits all the evidence raised so far is that H3K27me3 stabilizes PRC1 in chromatin increasing PRC1 dynamics in the absence of H3K27me3. Further studies will be needed though, with improved methodologies and tools, to better understand those dynamics and provide a clear picture of the different polycomb complexes recruitment and function.

# 6. Material and Methods

## 6.1. Tissue culture techniques

### 6.1.1. Cell lines and media

All cell lines used in this work are listed in Table 6.1.

| | Cell line | Medium | Feeders | Splitting factor | Reference |
|---|---|---|---|---|---|
| Embryonic Stem Cells | PGK12.1 | ES | No | 1:10 to 1:20 | (Mak et al., 2002; Penny et al., 1996) |
| | 129.1 | ES | SNLs | 1:6 | (Kay et al., 1993) |
| | *Eed*1 | ES | No | 1:10 | (Ura et al., 2008) |
| | *Eed*2 | ES | No | 1:10 | (Ura et al., 2008) |
| | *Eed*3 | ES | No | 1:10 | (Ura et al., 2008) |
| | *Eed*4 | ES | No | 1:10 to 1:30 | (Ura et al., 2008) |
| | ES-ERT2 *Ring1A*$^{-/-}$ | ES | PEFs | 1:3 | (Ura et al., 2008) |
| | *Eed* WT | KO | PEFs | 1:10 to 1:15 | (Landeira et al., 2010) |
| | *Eed*$^{-/-}$ | KO | PEFs | 1:3 to 1:6 | (Landeira et al., 2010) |
| | Ink4/Arf$^{-/-}$ | KO | PEFs | 1:8 | |
| MEFs | MEFs WT | EC10 | No | 1:4 | |
| | MEFs *Ink4a*/*Arf*$^{-/-}$ | EC10 | No | 1:4 | (Serrano et al., 1996) |

**Table 6.1 – Description of cell lines used in this work.** SNL is a mouse fibroblast STO cell line transformed with neomycin resistance and murine LIF genes.

Primary embryonic fibroblasts (PEFs)/MEFs were derived from E 13.5 F2 embryos. *Ink4*/*Arf*$^{-/-}$ ESCs were derived in the laboratory as part of this study.

Reagents were from Invitrogen except where indicated.

ES medium was composed of Dulbecco's modified Eagle's medium (DMEM) (without sodium pyruvate) supplemented with 15 % fetal calf serum (FCS) (Globe Pharm), 100 µg/ml penicillin/streptomycin, 2 mM L-glutamine, 1 x non-essential amino acids (NEAA), 50 µM 2-mercaptoethanol and 1 ml leukaemia inhibitory factor (LIF; prepared in house) per 500 ml medium.

KO medium was composed of knockout DMEM supplemented with 10 % FCS, 5 % Knockout serum replacement (KSR), 100 µg/ml penicillin/streptomycin, 2 mM L-glutamine, 1 x NEAA and 50 µM 2-mercaptoethanol and 1 ml LIF per 500 ml medium.

EC10 medium was DMEM, 10 % FCS, 100 µg/ml penicillin/streptomycin, 2 mM L-glutamine, 1 x NEAA and 50 µM 2-mercaptoethanol.

Derivation medium was MultiCell medium (DAINIPPON SUMITOMO PHARMA, Japan), supplemented with 0.1 M 2-mercaptoethanol and 200 µl/50 ml of LIF.

Medium was stored at 4 $^{o}$C.

### 6.1.2. Growing, thawing and freezing cells

*Eed* WT, *Eed* $^{-/-}$ and ES-ERT2 *Ring1A*$^{-/-}$ ESCs were grown on inactivated PEFs. 129.1 ESCs were grown in a monolayer of inactivated SNLs (mouse fibroblast STO cell line transformed with neomycin resistance and murine LIF genes).

All the cell lines were cultures in flasks or dishes from Nunc or BD Biosciences. ESCs were grown in flasks/dishes coated with 0.1 % gelatin to increase adherence. All cells were incubated at 37 $^{o}$C in a 5 % $CO_2$ atmosphere. Medium was warmed to 33 $^{o}$C before use. ESCs were fed daily. MEFs/PEFs were fed every third day.

Cells were passaged when 80-100 % confluent. To split cells, medium was removed, cells were briefly washed with phosphate buffered saline (PBS). MEFs/PEFs were washed twice as they are more resistant to trypsin. 0.05 % trypsin/0.53 mM ethylenediaminetetraacetic acid (EDTA) supplemented with 2 % chicken serum (CS) was added to cover the cells that were incubated for 3 to 10 min at 37 $^{o}$C. Once the cells detached, they were resuspended in fresh medium and plated into a new vessel according to the splitting ratio applied (Table 6.1).

To freeze cells they were trypsinized, as described above, and resuspended in fresh medium. They were then transferred to falcon tubes and spun down at 1000 rpm for 3 min at room temperature (RT; 25 $^{o}$C) in MultiFuge 3L centrifuge (Heraeus). Cells were resuspended in FCS with 10 % Dimethyl Sulfoxide (DMSO) (Sigma). Cells were aliquoted (1 ml) into 1.8 ml Cryotubes (Nunc) and frozen down at -80 $^{o}$C in “Mr Frosty” isopropanol containers (Nalgene). This allowed cells to cool down by approximately 1 $^{o}$C per minute ensuring cell viability after recovery. Frozen vials were stored in liquid nitrogen canisters.

To thaw cells, cryotubes were thawed in a 33-37 $^{o}$C water-bath and added to a falcon tube containing 9 ml of fresh medium per 1 ml of frozen cells to dilute the DMSO. Cells were spun down at 1000 rpm for 3 min at RT to remove medium with DMSO. Cells were resuspended in fresh medium and transferred to a tissue culture vessel. Medium was changed the following day as some cells die during the freeze-thaw process.

### 6.1.3.Treatments

*Eed*1, *Eed*2, *Eed*3 and *Eed*4 ESCs were treated for various lengths of time with varying concentrations of doxycycline (Sigma) (10 ng/ml, 100 ng/ml, 500 ng/ml, 1 µg/ml or 2 µg/ml). Final conditions were set at 1 µg/ml of doxycycline for 15 days. Deletion of *Ring1B* from ES-ERT2 *Ring1A*$^{-/-}$ cells was carried out by adding 800 nM of 4-hydroxytamoxifen (TMX) (Sigma) for 3 days as determined previously (Stock et al., 2007).

MG132 treatment of Eed4 cKO was carried out for 6 h with 10 µM MG132 (Merck Biosciences Ltd).

Trichostatin A (TSA) treatment of MEFs and ESCs was performed by adding TSA (Errant Gene Therapeutics) to a final concentration of 100 ng/ml overnight (ON). MEFs were gamma irradiated with 5 or 10 Gy (137 Cs source, Graviton RX30/55M, Graviton Ltd).

UV treatment of MEFs was performed at 50 J/m$^2$ in a UV stratalinker 2400 (Stratagene).

Transcription inhibitor treatments were performed in *Eed*4 cells, with and without treatment with Dox for 15 days (as described above). DRB (Sigma) was resuspended in ethanol and added to fresh medium to a final concentration of 25 µg/ml (stock solution of 20 mg/ml). Cells were harvested after 2, 6 and 20 h treatment. Actinomycin D (Enzo Life Sciences) was resuspended in DMSO (5 mg/ml) and added to fresh medium to a final concentration of 5 µg/ml). Cells were harvested after 2 and 6 h treatment. α-amanitin (Sigma) was resuspended in water (1mg/ml) and added to cells to a final concentration of 75 µg/ml. Cells were treated for 2, 6 and 20 h. Heat shock was performed by incubating cells at 42 $^{o}$C for 1 h. Cells were harvested by scrapping the plates at 42 $^{o}$C and immediately placing on ice. The subsequent steps were performed on ice, as described in each specific protocol. HS recovery was performed by incubating HS treated cells for 2 h at 37 $^{o}$C.

6.1.4.3T3 protocol

MEFs were counted and $1x10^6$ MEFs were plated per 90 mm tissue culture dish. On the third day of culture, cells were trypsinised and counted. The same initial number of cells was plated again. This procedure was repeated every 3 days.

6.1.5.Feeders

PEFs for feeders were derived from E 13.5 embryos, passaged once in culture and inactivated. SNLs were grown and expanded as much as required before inactivation. Inactivation was performed by inhibition of the cell cycle with mitomycin-C. Both PEFs and SNLs were incubated with 10 µg/ml mitomycin C (Sigma) in EC10 for 2 h at 37 °C in a 5 % $CO_2$ atmosphere. Cells were then washed 3 times with PBS and trypsinized (section 6.1.2). Cells were counted and $3 \times 10^6$ or $9 \times 10^6$ cell aliquots were frozen down. When thawed, SNLs were plated at $1 \times 10^6$ and PEFs at $0.6 \times 10^6$ per 25 $cm^2$.

Cells growing on feeders were trypsinized and pre-plated for 30 min in order to remove contaminating feeder cells prior to harvesting. PEFs and SNLs are fibroblasts and attach faster than ESCs and are therefore depleted from the floating population of cells in 30 min.

6.1.6.Lipofection of ESCs with PAC DNA

For lipofection into cells PAC DNA was extracted with NucleoBond BAC 100 Kit (Fisher) to avoid cytotoxic effects of phenol contamination on ESCs.

*Ink4*/*Arf*$^{-/-}$ and 129.1 ESCs were passaged the day before into a 6-well dish to be in the exponential growth phase and about 90 % confluent. Medium was replaced with 1.5 ml of KO medium without penicillin/streptomycin several hours before transfection.

0.25, 0.5, 1 or 3 µg of PAC DNA was added to individual eppendorf tubes, together with 0.5 ng of hygromycin containing plasmid, in a total of 250 µl of OPTIMEM (Invitrogen) and incubated for 5 min at RT. 0.75, 1.5, 3 or 9 µg of lipofectamine (Invitrogen) (3 x DNA concentration) was added to another eppendorf tube with OPTIMEM to make a final volume of 250 µl. The PAC DNA solution was added to the lipofectamine, mixed and incubated for 30 min at RT. The mixture was then added drop by drop to the cells and incubated ON at 37 $^{o}$C in a 5 % $CO_2$ atmosphere. The following day cells were passaged into a 90 mm dish. Selection began 48 h after lipofection with 150 µg/ml of hygromycin. 10 to 12 days after lipofection single colonies were picked.

#### 6.1.6.1. ESCs colony picking

Each 90 mm dish with colonies to be picked was washed twice with PBS. 2-3 ml of PBS was added to the dish to prevent cells from drying. Individual colonies were picked and transferred to 25 µl trypsin/EDTA/2 % chicken serum in a well of a 96 well plate. Once the entire plate was picked it was incubated for 30-45 min in a $CO_2$ incubator at 37 $^{o}$C. Cells were resuspended and transferred to a pre-prepared 96 well plate with 3 x $10^6$ feeders. The plate was incubated at 37 $^{o}$C and the media changed every day.

#### 6.1.6.2. 96-well plate splitting

When a 96-well dish was confluent it was passaged into 4 x 96-well plates. Two plates had feeders on the bottom while 2 plates were only gelatinized so cells could be harvested directly for genomic DNA extraction. To split cells, the medium was removed and cells were washed twice with PBS. 50 µl trypsin/EDTA/2 % chicken serum was added to each well and cells were incubated for 25-30 min at 37 $^{o}$C. 150 µl of medium was added to stop trypsinization, cells were well resuspended and divided into 4 plates. The plates were incubated at 37 $^{o}$C in a $CO_2$ incubator.

#### 6.1.6.3. 96-well plate freezing

Of the plates described above, one was frozen when more than 50 % of clones were confluent and the other plate frozen 1-2 days later. To freeze cells a round-bottom plate was prepared with 50 µl of 2 x freezing medium (80 % FCS with 20 % DMSO) and kept at 4 $^{o}$C. Cells were trypsinized as described above but with 25 µl of trypsin/EDTA/2 % chicken serum. Once cells were trypsinized 25 µl of FCS was added to the cells and they were resuspended. 50 µl of 2 x freezing medium was then added, resuspended and transferred to the round-bottom plate. The plate was sealed with sticky film (Nunc), wrapped in bubble wrap and placed at -80 $^{o}$C in a steroplast box.

#### 6.1.6.4. 96-well plate DNA extraction

When the feederless 96-well plates were confluent DNA was extracted. Cells are washed 3 x with PBS and 50 µl of Lysis Buffer (10 mM NaCl, 10 mM Tris-HCl pH 7.5 (prepared in house), 10 mM EDTA, pH 8.0 (prepared in house), 0.5 % Sarcosyl (BHD)) with 1 mg/ml Proteinase K (PK; Roche) (added just before use) was added per well. The plate was sealed with sticky film and incubated at 55 $^{o}$C on a rotating platform ON. The following day 2 volumes of ice-cold 100 % ethanol were added and the plate was left for 1 h on a rotating platform at RT. Liquid was then removed by inverting the plate very gently. The plate was washed 3 x with 100 µl 70 % ethanol that was again removed by inverting the plate. After the last wash, the ethanol was completely removed by tapping the plate on absorbent paper until the DNA was dry. The DNA was resuspended in 100 µl TE and 2 µl were used per polymerase chain reaction (PCR) reaction. PCR was performed to check the integration of PAC DNA. Primers are listed in table 6.2.

| Name | Forward (5'-3') | Reverse (5'-3') | Tm (°C) | Size (bp) |
|---|---|---|---|---|
| **dsRed1** | CGAGGACGTCATCAAGGAGT | CCGTCCTCGAAGTTCATCAC | 60 | 294 |
| **dsRed2** | GGACGGCTCCTTCATCTACA | GCTCCACGATGGTGTAGTCC | 60 | 296 |
| **ECFP1** | ACGTAAACGGCCACAAGTTC | AAGTCGTGCTGCTTCATGTG | 60 | 187 |
| **ECFP2** | CACATGAAGCAGCACGACTT | TCGGCGGTGATATAGACGTT | 60 | 236 |
| **IRES** | CTAACGTTACTGGCCGAAGC | TACGCTTGAGGAGAGCCATT | 60 | 392 |
| ***Neomycin*** | AGACAATCGGCTGCTCTGAT | CAATAGCAGCCAGTCCCTTC | 60 | 203 |
| **PolyA** | TAGTTGCCAGCCATCTGTTG | GCGATGCAATTTCCTCATTT | 60 | 104 |

**Table 6.2 – Primers used for PAC PCR/Sequencing**. Primers were used to check PAC recombineering by sequencing and to analyse PAC integration into ESCs/MEFs. Tm refers to annealing temperature.

## 6.2. Mouse work

All mice used in this study were bred and maintained in the animal facilities at the University of Oxford. All procedures were approved by a local Ethical Review Committee and licensed by the Home Office under the Animals (scientific procedures) Act 1986. All regulated procedures were carried out by Jonathan Godwin of the Department of Biochemistry. Pregnant females were sacrificed on the required *day post coitum* (*dpc)*, the uterus was removed and embryos were dissected from the uterus.

F2 mouse embryos were derived from crosses between B6CBF1 (from crosses between C57BL/6 female and male CBA), bought from Charles River, UK. *Ink4*/*Arf* heterozygous mice were obtained from Jesus Gil (Imperial College, London, UK), after previous agreement with Ronald DePinho (Dana-Farber Cancer Institute, Boston, USA).

### 6.2.1. Mice genotyping

DNA was extracted from ear pieces using the DNeasy Blood & Tissue kit (Qiagen), according to manufacturer instructions. Genotyping PCR primers and annealing temperatures are listed in Table 6.3. WT embryos gave a 237 bp band and mutants gave a band of approximately 700 bp. Heterozygous embryos had both bands.

| Name | Forward (5'-3') | Reverse (5'-3') | Reference |
|---|---|---|---|
| **P16INK4a** | ATGATGATGGGCAACGTTC | CAAATATCGCACGATGTC | (Serrano et al., 1996) |
| **P16Neo** | CTATCAGGACATAGCGTTGG | AGTGAGAGTTTGGGGACAGAG | (Serrano et al., 1996) |

**Table 6.3 – Mice genotyping primers.** P16Ink4a amplifies the WT band while p16Neo amplifies the mutant one. Primers Tm goes from 65 $^{o}$C to 55 $^{o}$C, decreasing 0.5 $^{o}$C each cycle.

### 6.2.2. ESC derivation

The ESC derivation protocol was kindly provided by a collaborator (Haruhiko Koseki, personal communication).

4-well dishes were gelatin-coated and prepared with a feeder layer of 1.25 x $10^5$ PEFs in derivation medium per well. Blastocysts were harvested at 3.5 *dpc*. One blastocyst was placed per well and incubated in a $CO_2$ incubator at 37 $^{o}$C until they hatched from the *zona pellucida* and attached to the bottom of the well. Approximately 5 days later the ICM outgrowth was dissected out from the surrounding trophoblast cells. The ICM was dissociated manually into small clumps and single cells. These were transferred to a well of a 4-well dish prepared as described above. A few days later, individual ESC colonies were visible and were picked and expanded to establish a cell line.

### 6.2.3. PEFs derivation

PEFs/MEFs were collected from mice at E 13.5. After removal of internal tissues and organs, embryos were rinsed with sterile PBS and digested with 1 ml trypsin/EDTA/2 % chicken serum per embryo. The tissue was dissociated by passing the whole embryo through an 18 G needle. Homogenised tissue was incubated for 10 min at 37 $^{o}$C. Trypsin was inactivated by the addition of an equal volume of EC10 medium containing 20 % FCS. After centrifugation cells were resuspended in EC10 medium containing 20 % FCS and dissociated again by passing them through a 21 G needle.

These were plated at the density of approximately one embryo per 25 $cm^2$. This was considered passage zero.

### 6.2.4. PAC microinjection into mice

PAC DNA was injected into *Ink4*/*Arf*$^{-/-}$ mice egg pronucleus.

Females were given a 0.1 ml intraperitoneal injection of Follicle Stimulating Hormone (Folligon) or Pregnant Mare Serum (PMS) at 2 pm. 48 h later the same females were given the same dose of Human Chorionic gonadotrophin (HCG) at 2 pm. Females were then set up with the studs and checked for a vaginal plug the next morning (0.5 *dpc*). The females were then killed by cervical dislocation and the oviduct collected in M2 media (Millipore) at 37 $^{o}$C. Eggs were released from the ampulla and treated with Hyaluronidase (Sigma) to release them from the cumulus mass. Eggs were collected and placed in M16 media (Millipore) under mineral oil until use.

To do PAC microinjection PAC DNA was extracted with NucleoBond BAC 100 Kit to avoid the toxic effects of phenol:chloroform (Phe:$CHCl_3$) contamination. DNA was then linearised by digestion with a restriction enzyme that cuts once within the backbone. The DNA was purified by the addition of 0.7 volumes of isopropanol (VWR). DNA was spun down at 12000 g for 20 min at 4 $^{o}$C, washed with 70 % ethanol and spun again at 12000 g for 5 min at 4 $^{o}$C. The supernatant was removed, the DNA air dried and resuspended in Microinjection Buffer (10 mM Tris-HCl pH 7.5, 0.1 mM NaCl (VWR), 0.1 mM EDTA, 30 µM spermine (Sigma) and 70 µM spermidine (Calbiochem), filter sterilized) to a final concentration of 1 µg/µl. 2 pl of DNA in microinjection buffer was used to inject into the 0.5 *dpc* egg pronucleus.

Recipient females were made pseudo-pregnant by mating with T145h sterile males or by vasectomised males. Under general anaesthesia a single/bilateral flank or midline laparotomy was performed and genetically modified embryos implanted into the oviduct or uterus. Recipients were sacrificed at E 13.5 (*dpc*).

### 6.3. Senescence-associated β-galactosidase staining

MEFs and PGK 12.1 ESCs were grown in 90 mm diameter dishes. After washing with PBS, cells were fixed in 0.5 % Glutaraldehyde (Sigma) in PBS (stored in dark), for 15 min at RT. After fixation, cells were washed twice with PBS/ $MgCl_2$ solution (1 mM $MgCl_2$ (VWR) in 1 x PBS, pH 6.0) and incubated for 1-24 h at 37 $^{o}$C with 1 x X-gal staining solution (40 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal; Sigma)) and 1 x KC solution (0.1 mM potassium ferrocyanide (Sigma), 0.1 mM potassium ferricyanide (Sigma), in PBS, stored in dark) in PBS/$MgCl_2$ solution. Cells were then washed twice with water and stored in the dark at 4 $^{o}$C. This protocol was kindly provided by a collaborator (Jesus Gil, personal communication).

### 6.4. Alkaline Phosphatase staining

Alkaline phosphatase (AP) staining was carried out as described (Harris and Hopkinson, 1976) with modifications. *Eed*4/*Eed*4 cKO cells were grown on gelatin coated dishes. Cells were washed twice in PBS and air-dried for 20 min. AP solution (AP buffer (100 mM Tris-HCl pH 9.0, 100 mM NaCl, 5 mM $MgCl_2$, filter sterilized, keep at 4 $^{o}$C) with 0.4 µg/ml Naphthol phosphate (Sigma) and 1 µg/ml Fast Violet B Salt (Sigma)) was prepared and immediately added to the cells. The cells were then incubated for 10 to 30 min at RT in the dark. Cells were rinsed 3 times in water and analysed under the microscope.

## 6.5. Genomic DNA extraction

Cells were trypsinised, collected in a 15 ml falcon tube, washed in PBS and spun down at 1000 rpm for 3 min at RT in a MultiFuge 3L centrifuge (Heraeus). Cells were resuspended in 2 ml per 25 $cm^2$ of Lysis Buffer (50 mM Tris-HCl pH 8.0, 10 mM NaCl, 10 mM EDTA pH 8.0). Sodium Dodecyl Sulfate (SDS; made in house) was added to a final concentration of 0.5 %, PK was added to a final concentration of 20 µg/ml and DNAse-free RNAse A (ribonuclease A; Roche) was added to a final concentration of 200 µg/ml and incubated ON at 55 $^{o}$C to lyse cells and digest proteins and RNA. After incubation an equal volume of phenol:chloroform:isoamyl alcohol mixture (25:24:1 saturated with 10 mM Tris-HCl pH 8.0 and 1 mM EDTA, Sigma) was added and vigorously shaken. Phases were separated by centrifugation at 2000 rpm for 5 min at RT in a Heraeus MultiFuge 3SR+ centrifuge (Thermo Scientific). The aqueous phase with DNA was transferred to a fresh tube. Extraction was repeated once with phenol:chloroform:isoamyl alcohol and again with chloroform (VWR) only. DNA was precipitated with 1:25 volumes of 5 M NaCl and 2.5 volumes of ice-cold 100 % ethanol (VWR). Precipitated DNA was spooled out, washed in 70 % ethanol, spun down at 13000 rpm in a tabletop microcentrifuge, air-dryed and resuspended in appropriated volume of TE buffer (10 mM Tris-HCl, 1 mM EDTA) (according with pellet size).

### 6.5.1. Polymerase chain reaction

PCR reactions contained 1 unit of Taq Polymerase (Invitrogen), 0.4 µM of both forward and reverse primers, 0.2 mM of each deoxyribonucletide triphosphate (dNTP; Promega), 1.5 mM $MgCl_2$ and 1 x PCR buffer supplied with the Taq Polymerase. PCR conditions were 94 $^{o}$C for 4 min, followed by 30 to 35 cycles of 94 $^{o}$C for 30 s, a specific annealing temperature for 30 s, 72 $^{o}$C for 30 s, followed by a final incubation of 72 $^{o}$C for 7 min. PCR products were separated by electrophoresis on 1-2.5 % agarose gels, prepared with electrophoresis grade agarose (Invitrogen) in Tris-acetate EDTA (TAE) buffer (40 mM Tris-acetate, 1 mM EDTA) prepared in house. Ethidium bromide (Invitrogen) was

added to a final concentration of 0.2 µg/ml and products were visualised by ethidium bromide illumination under UV light.

## 6.6. RNA related techniques

### 6.6.1. RNA extraction

RNA was extracted using Trizol reagent (Invitrogen) according to the manufacturer's instructions.

RNA was extracted from cell pellets or from cells growing on plates (if feederless). In both cases cells were washed with PBS and Trizol reagent was added at the equivalent to 1 ml per 25 $cm^2$ of cells and incubated for 2-3 min at RT. Trizol was mixed with cells, transferred to an eppendorf tube and 0.2 ml of chlorophorm added per ml of lysate. The mixture was vigorously shaken for 30 s and incubated for 2-3 min at RT. Phases were separated by centrifugation at 12000 g for 15 min at 4 $^{o}$C in a microcentrifuge. The upper aqueous phase containing RNA was transferred to a clean tube kept on ice. An equal volume of isopropanol (about 0.5 ml) was added and mixed. The sample was incubated on ice for 5-10 min. RNA was then precipitated by centrifugation at 12000 g for 30 min at 4 $^{o}$C in a microcentrifuge. The RNA pellet was washed twice in fresh 80 % ethanol and immediately spun down at 12000 g for 5 min at 4 $^{o}$C. All ethanol was removed after the last wash and the pellet was allowed to air dry for 5-10 min. The RNA was resuspended in 90 µl of RNase-free water.

RNA was treated with Turbo DNA-free reagent (Ambion) to exclude the possibility of DNA contamination. Briefly, 10 x buffer was added to each sample to a final concentration of 1 x. 2 units of TURBO Dnase I were added and sample was incubated at 37 $^{o}$C for 30 min. 10 µl of DNase inactivation reagent was added, the sample was gently mixed and incubated for 2 min. The treated RNA was collected by spinning down samples at 10000 g for 1.5 min. Supernatant containing RNA

was transferred to a fresh tube. RNA concentration was measured on a Nanodrop ND-1000 spectrophotometer and RNA integrity was checked by running 2 µl on an agarose gel.

RNA was stored at -80 $^{o}$C.

### 6.6.2.Complementary DNA synthesis

Complementary DNA (cDNA) synthesis was primed using random hexamers (GE Healthcare) and carried out using SuperScript III (Invitrogen) following manufacturer's guidelines. Briefly 2 µg RNA was mixed with 200 ng random hexamers and 10 mM of dNTP mix (Promega) in a final volume of 13 µl. The mix was heated at 65 $^{o}$C for 5 min to denaturate RNA, followed by an incubation on ice of at least 1 min. 5 x first strand buffer (to 1 x final concentration), 5 mM dichlorodiphenyltrichloroethane (DTT), 40 units of recombinant RNase inhibitor (Invitrogen) and 200 units of SuperScript III RT were added to the RNA in a final volume of 20 µl. Another reaction without Superscript III RT enzyme was performed as a control. Samples were incubated at 25 $^{o}$C for 5 min followed by 50 $^{o}$C for 60 min. The reaction was inactivated at 70 $^{o}$C for 15 min. cDNA was diluted 1:5 and stored at -20 $^{o}$C.

### 6.6.3.Real-time quantitative analysis

#### 6.6.3.1. Quantitative real-time PCR

Quantitative real-time PCR (qPCR) analysis was performed on cDNA templates. 5 µl of cDNA was added together with 1 x iQ SYBR Green Supermix (Bio-Rad) and 0.3 mM forward and reverse primers. A sample with water instead of DNA was used to rule out primer dimers formation. Primers are listed in table 6.4. qPCR was performed on Chromo4 DNA engine (Bio-Rad) using Opticon Monitor 3 software (MJ Research Inc. Waltham, MA). PCR conditions were as follows: 3 min at 95 $^{o}$C followed by 40 cycles at 95 $^{o}$C for 20 s and 60 $^{o}$C for 20 s, followed by plate read after each cycle. A melting curve test was performed at the end of each experiment (from 55 $^{o}$C to 90 $^{o}$C, read every 0.5

$^{o}$C for 1 s) to ensure the specificity of amplification. The comparative $C_T$ method was used for relative quantitation of gene expression (User bulletin 2, Applied Biosystems, 2001). Each amplicon was analysed in triplicates on independent cDNA preparations.

Data were normalized with housekeeping genes.

| Name | Forward (5'-3') | Reverse (5'-3') | Size (bp) | Reference |
|---|---|---|---|---|
| ***Hmbs*** | CTGTACCTGACTGGTGGAGTCTGGAGTC | TGATGCCCAGGTTCTCAGCAGCT | 184 | (Stock et al., 2007) |
| ***Idh2*** | AGAAAATGTGGAAGAGCCCTAACG | TGCCAGCTCGATCTACCACAAAAT | 186 | (Stock et al., 2007) |
| ***B2m*** | ACCCGCCTCACATTGAAATCC | CGATCCCAGTAGACGGTCTTG | 199 | |
| ***Gapdh*** | CCCACTTGCCTCTGTATTGG | CTGTGGGGAGTCCTTTTCAG | 255 | |
| ***ß-actin*** | GATATCGCTGCGCTGGTCGT | AGATCTTCTCCATGTCGTCC | 245 | |
| ***Mel18*** | AGCTGAACCCTCACCTCATGTGTGC | TTTGTTGGTCTCCAAGTAGC | 122 | |
| ***Rps14*** | GACCAAGACCCCTGGACCT | CCCCTTTTCTTCGAGTGCTA | 122 | |
| ***Ink4a*** | GTGTGCATGACGTGCGGG | GCAGTTCGAATCTGCACCGTAG | 146 | |
| ***Arf*** | GCTCTGGCTTTCGTGAACATG | TCGAATCTGCACCGTAGTTGAG | 137 | |
| ***Oct4*** | CGTGGAGACTTTGCAGCCTG | GCTTGGCAAACTGTTCTAGCTCCT | 182 | (Stock et al., 2007) |
| ***Nanog*** | GAACTATTCTTGCTTACAAGGGTCTGC | GCATCTTCTGCTTCCTGGCAA | 200 | (Stock et al., 2007) |
| ***Sox2*** | CATGTGAGGGCTGGACTGCG | GCTGTCGTTTCGCTGCGG | 144 | |
| ***Gata4*** | GAGGCTCAGCCGCAGTTGCAG | CGGCTAAAGAAGCCTAGTCCTTGCTT | 199 | (Stock et al., 2007) |
| ***Hoxa7*** | AGTTCAGGACCCGACAGGAA | TTGATCTGGCGCTCAGTGAG | 152 | |
| ***Nkx2-2*** | TGTGCAGAGCCTGCCCCTTAA | GCCCTGGGTCTCCTTGTCAT | 196 | |
| ***Cdx2*** | TCAACCTCGCCACAACCTTCCC | TGGCTCAGCCTGGGATTGCT | 210 | (Stock et al., 2007) |
| ***Pax3*** | CCCAGTGAGAGGGGAGAGAGCATA | CTTCGAACGCAGACAGCAGC | 218 | |
| ***Gata1*** | GTCCTCACCATCAGATTCCACAG | AGTGGATACACCTGAAAGACTG | 168 | |

**Table 6.4 – Primers used for RT-PCR expression analysis.** All the primers best perform at Tm of 60 $^{o}$C.

#### 6.6.3.2. Primer design and testing

Primers were designed using the Primer 3 program (http://fokker.wi.mit.edu/primer3/input.htm) using mouse genome sequences from Ensembl (http://www.ensembl.org/index.html). Primers were designed to be 180-210 bp, melting temperatures of 58-68 $^{o}$C, 20-28 bp long, GC content of 40-60 %, no more than 3 out of 5 bases at 3' end with G or C, no dimers and with secondary structure as weak as possible. Primers were ordered either from Sigma Genosys or Invitrogen. PCR primers were tested by conventional PCR and resolved by gel electrophoresis (section 6.5.1). The reactions yielding a single band were tested by qPCR. Each primer pair was tested for efficiency and quantification with serial 2-fold dilutions of cDNA or genomic DNA. Only primers with a good quantification rate (C(T) versus $Log_{10}$ of DNA concentration with $R^2 \geq 0.99$) and good efficiency (between 1.8 and 2.2) were used.

## 6.7. Chromatin immunoprecipitation

### 6.7.1. Crosslinked ChIP

Crosslinked ChIP was performed as described (Stock et al., 2007) with some modifications.

#### 6.7.1.1. Chromatin preparation

About 1-2 x $10^8$ ESCs/MEFs were used per experiment. Cells were grown on flasks/plates and prior to starting the experiment medium was replaced with 10 ml of fresh medium. 1 % formaldehyde (Sigma) was added to the medium and cells were incubated for 10 min at 37 $^{o}$C. The crosslinking reaction was stopped with the addition of 1.25 M glycine (Sigma) to a final concentration of 0.125 M and incubated for 5 min at RT. Cells were washed 3 x with ice-cold PBS. Swelling Buffer (25 mM Hepes (Invitrogen), 1.5 mM $MgCl_2$, 10 mM KCl, 0.1 % NP-40 (Sigma) with complete EDTA-free protease Inhibitors (Roche) added immediately before use) was added to the cells and incubated for 10 min on ice. Cells were scraped off the plates and nuclei isolated by douncing 50 x with a tight

dounce homogeneizer. Nuclei were spun down at 4500 rpm for 15 min at 4 $^{o}$C in a Multifuge 3SR+ centrifuge (Heraeus). The supernatant was removed and the pellet resuspended in 4-8 ml (depending on pellet size) of Sonication Buffer (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 1 % Triton X-100 (VWR), 0.1 % deoxycholate (Sigma), 0.1 % SDS with complete EDTA-free protease Inhibitors added immediately before use). Sonication was performed using a Bioruptor sonicator (Diagenode) cooled with a refrigerated water bath (3 $^{o}$C) for 1-1.5 h (30 s on, 30 s off at full power) until the samples were clear.

Fragment sizes were analysed by taking a 5 µl aliquot, boiling at 100 $^{o}$C for 10 min and performing Phe/$CHCl_3$ extraction. The upper phase was run on a 1.5 % agarose gel (section 6.5.1). If fragment sizes were lower than 1.5 kb the procedure was continued. Otherwise, the sample was sonicated again. The sample was transferred to eppendorfs and the cell debris was spun down at 14000 rpm for 15 min at 4 $^{o}$C. The supernatant was transferred to a fresh eppendorf and spun again. The supernatant was again transferred to a fresh tube and chromatin (DNA) concentration was measured by alkaline lysis using a Nanodrop ND-1000 spectrophotometer (NanoDrop). For the measurement an aliquot of chromatin was diluted 1:5 in 0.1 M NaOH. A blank measurement was taken with sonication buffer diluted 1:5 in NaOH.

#### 6.7.1.2. Immunoprecipitation

Antibodies (Ab) used were ubiquityl-H2A (05-678, Millipore) 50 µl, Ring1B (Atsuta et al., 2001) 15 µl, H3 (ab1791, Abcam) 3 µl, Mel18 (sc-10744, Santa Cruz) 30 µl, H3K27me3 (CS-069-100, Diagenode) 5 µl, H3K4me3 (07-473, Millipore), 5 µl, IgG antibody (M7023, Sigma) 2 µl, was used as a control except for H2AK119u1 where a no antibody sample was used.

Protein A pre-blocked with salmon sperm DNA beads (Millipore) were used for all antibodies except anti-ubiquityl H2A antibody for which anti-mouse IgM (μ-chain specific)-agarose beads (Sigma) were used. Beads were washed 2 x with a double volume of Sonication Buffer. Beads were collected by centrifugation at 1500 rpm for 3 min at 4 $^{o}$C.

For Protein A beads, the total volume of chromatin for the total number of IPs was pre-cleared with 50 μl per IP of blocked beads and incubated for 2 h at 4 $^{o}$C on a rotating wheel. Pre-cleared chromatin was spun at 1500 rpm for 5 min at 4 $^{o}$C. Immunoprecipitations (IPs) were set up in eppendorfs by adding supernatant corresponding to 200 μg of chromatin together with the antibodies described above. 1/10 of an IP volume of chromatin (Input) was retained for later comparison with pulled down DNA and kept at -20 $^{o}$C.

For anti-ubiquityl H2A antibody (IgM), chromatin was not pre-cleared. An IP was set up by adding 200 μg of chromatin with the antibody and 50 μl of washed beads. A sample with no antibody was used as a negative control and an Input sample from non pre-cleared chromatin (1/10 of IP volume) was also kept at -20 $^{o}$C.

IPs were incubated ON at 4 $^{o}$C on a rotating wheel. The following day 50 μl of Protein A beads were added to the samples (except for anti ubiquityl-H2A and no Ab samples) and incubated for 2 h at 4 $^{o}$C on a rotating wheel. Antibody-bead complexes were collected by centrifugation at 1400 rpm for 3 min at 4 $^{o}$C and resuspended in 1 ml of wash buffer with EDTA-free protease Inhibitors. Samples were incubated for 5 min at 4 $^{o}$C on a rotating wheel and complexes spun down after each wash at 1500 rpm for 3 min at 4 $^{o}$C. Protein A beads were washed 1 x with Sonication Buffer, 2 x with Wash Buffer A (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 500 mM NaCl, 250 mM LiCl (Sigma), 1 % NP-40, 0.5 % Na-deoxycholate), 2 x with Wash Buffer B (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 250 mM LiCl, 1 % NP-

40, 0.5 % Na-deoxycholate) and 2 x with TE Buffer pH 8.0. IgM beads were washed 2 x with Sonication buffer, 1 x with H5 Wash Buffer (2 mM Tris-HCl pH 8.0, 0.02 mM EDTA, 50 mM LiCl, 0.1 % NP-40, 0.1 % Na-deoxycholate) and 1 x with TE Buffer pH 8.0. After the last wash all the supernatant was removed and the beads were resuspended in 250 µl of Elution Buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 % SDS).

#### 6.7.1.3. DNA elution and purification

DNA was eluted from the antibody-bead complexes by heating the sample for 5 min at 65 $^{o}$C. The sample was then incubated for 15 min on a rotating wheel at RT and spun down at 1500 rpm for 5 min in a tabletop centrifuge. The supernatant was collected and elution repeated. At this point Input samples were thawed and their volume was topped up to 500 µl with TE. DNA crosslinks were reversed by adding NaCl, and DNAse-free RNAse A to a final concentration of 160 mM and 20 µg/ml respectively and incubated at 65 $^{o}$C ON, mixing at 800 rpm for 1 min every 2 min. The following day EDTA was added to a final concentration of 4 mM and PK to a final concentration of 200 µg/ml. The samples were incubated for 2 h at 45 $^{o}$C. The samples were transferred to phase lock gel heavy eppendorfs (VWR) where Phe:$CHCl_3$ extraction was performed by adding 500 µl Phe:$CHCl_3$, mixing vigorously and spinning down at 14000 rpm for 5 min. $CHCl_3$ extraction was performed by adding 450 µl $CHCl_3$ to each sample mixing vigorously and spinning down at 14000 rpm for 5 min. Purified DNA was precipitated by adding glycogen blue (Ambion) to a final concentration of 5 µg/ml and 1 ml of 100 % Ethanol. Samples were incubated at -20 $^{o}$C for a few hours to ON after which DNA was pelleted by centrifugation at 14000 rpm for 30 min at 4 $^{o}$C. DNA was allowed to air dry and was resuspended in 150 µl 1x TE. 5 µl was used per qPCR reaction. qPCR was carried out as described in section 6.6.3.1. Enrichment was normalized to input DNA. Primers are shown in Table 6.5 and Table 6.6.

| Name | Forward (5'-3') | Reverse (5'-3') | Size (bp) | Reference |
|---|---|---|---|---|
| ***Oct4*** | GTGAGCCGTCTTTCCACCAGG | GGGTGAGAAGGCGAAGTCTGAA | 88 | |
| ***Nkx2-2*** | CAGGTTCGTGAGTGGAGCCC | GCGCGGCCTCAGTTTGTAAC | 158 | (Stock et al., 2007) |
| ***Sox1*** | ACAAGAGGAGGCAGCGAACC | TCGCAGGTGGAAAGTTTCTCC | 192 | (Stock et al., 2007) |
| **L1** | GCACTTGATCTTTACGTTCAGAGA | TCTTAGTTTGCCCTCTTACTGGTT | 185 | |
| **L2** | GCTTGGTCTAATGCTAACTGTGAA | GCAGAGGTCTTTATTTAGCTCAGG | 196 | |
| **L3** | GGACGCTCACCGAAGCTACT | GGTCTGTGGCAGAAATGGTC | 102 | |
| **L4** | TAAGTTCTAGGAGGAGTGCATGTG | CATACAGGTGGAGAAAGGAGATTC | 194 | |
| **L5** | GCCTCGCCGATCTTCCTATTTTCT | CCCATCGCGGTGACAGC | 227 | |
| **L6** | GTTCTTGGTCACTGTGAGGATTC | CTCAAGATCCTCTCTAGCCTCAAC | 150 | |
| **L7** | CCCGGGAGATGGAGTACTTT | GTTGGGGCCACTCATTTCTA | 166 | |
| **L8** | ATTGCCTTTCTGGTGAGGACTGT | CAACTATAAAAAAGTGACGGGGTGAAG | 249 | |
| **L9** | CGAACTCGAGGAGAGCCATC | ACACTCCTTGCCTACCTGAA | 170 | |
| **L10** | GTGATGATGATGGGCAACGTTC | AAATATCGCACGATGTCTTGATGTC | 249 | |
| **L11** | AACCTCCTACCGTAGATAAACACG | TAGTCAGGTAATGACCAAAGTCCA | 154 | |
| **L12** | CAGCCAAGTTCAAGTACAAGACAA | CCAAGCCAGTAACTGTTTGCTATT | 152 | |
| **L13** | GGCTTTAGAAATCAAGGAATCAGTC | CCTTCCTACAAGTTCCCTACATGA | 153 | |

**Table 6.5 – Primers used for ChIP across *Ink4/Arf*.** All the primers best perform at Tm of 60 °C.

| Name | Forward (5'-3') | Reverse (5'-3') | Size (bp) | Reference |
|---|---|---|---|---|
| ***Sox2*** | CCATCCACCCTTATGTATCCAAG | CGAAGGAAGTGGGTAAACAGCAC | 92 | (Stock et al., 2007) |
| ***Nanog*** | CCCAGGTTTCCCAATGTGAAG | AAAGAGTCAGACCTTGCTGCCA | 97 | (Stock et al., 2007) |
| ***Rps14*** | GACCAAGACCCCTGGACCT | CCCCTTTTCTTCGAGTGCTA | 122 | |
| ***Ensa*** | ATGCAACATTCAGCGACTTG | CTAGAACACCAACCGGGAAA | 162 | (Blackledge et al., 2010) |
| ***B2m*** | GCATCAACAGCTAGGGACTGGTGAC | CCGAGTAGCAGCCACTGAAATGAG | 119 | (Stock et al., 2007) |
| ***Gata4*** | AAGAGCGCTTGCGTCTCTA | TTGCTAGCCTCAGATCTACGG | 110 | (Stock et al., 2007) |
| ***Math1*** | CCTTCTTTGACTGGGCAGAC | ACTCGGAGATCGCACACC | 157 | (Stock et al., 2007) |
| ***Hoxa7*** | GAGAGGTGGGCAAAGAGTGG | CCGACAACCTCATACCTATTCCTG | 151 | (Stock et al., 2007) |
| ***Msx1*** | ACAGAAAGAAATAGCACAGACCATAAGA | TTCTACCAAGTTCCAGAGGGACTTT | 148 | (Stock et al., 2007) |
| ***Cdx2*** | GGACTCCGCGAGCCAA | CTCAGCCCACGGTGCTC | 162 | (Stock et al., 2007) |
| ***Pax3*** | CGGACGAGTTTGGTGCGAGT | TGTTCCTGCCTCGTCACCAC | 113 | |
| ***Gata1*** | AGAGGAGGGAGAAGGTGAGTG | AGCCACCTTAGTGGTATGACG | 135 | (Stock et al., 2007) |

**Table 6.6 – Primers used for *Eed*$^{-/-}$ ESCs and *Eed* cKO ChIP.** All the primers best perform at Tm of 60 °C.

### 6.7.2.Short fragments ChIP

ChIP for RING1B, H3, H3K4me3, EZH2 and H3K27me3 was carried out using a different protocol to confirm results. RING1B ChIP for deep-sequencing was also performed using this protocol as fragment sizes are smaller which could improve resolution. Antibodies used were Ring1B (Atsuta et al., 2001) 15 µl, H3 (ab1791, Abcam) 3 µl, H3K27me3 (CS-069-100, Diagenode) 5 µl, H3K4me3 (07-473, Millipore), EZH2 (pAb-039-050, Diagenode) 5µl and IgG antibody (M7023, Sigma) 2 µl.

Cells (1-2 x $10^8$ cells) were fixed in PBS with 2 µM Ethylene glycol bis (succinimidylsuccinate) (EGS) (Thermofisher) for 1 h at RT. After washing, cells were fixed again in ChIP Fixation Buffer (1 % formaldehyde, 5 µM EGTA (Sigma), 10 µM EDTA, 1 mM NaCl and 0.5 mM HEPES in PBS) for 10 min at RT. The fixation reaction was stopped by adding glycine to a final concentration of 125 mM. Cell extracts were lysed in Lysis Buffer (1% SDS, 10 mM EDTA pH 8.0, 5 mM Tris-HCl pH 8.1 with complete EDTA-free protease inhibitors) and sonication was performed using a Bioruptor sonicator to obtain an average DNA fragment size of 300 bp. Chromatin sizes were determined as described in section 6.7.1.1. Chromatin was diluted in dilution buffer (1 % Triton X-100, 2 mM EDTA pH 8.0, 150 mM NaCl, 20 mM Tris-HCl pH 8.1 containing complete protease inhibitors). Protein G Sepharose (for Ring1B antibody) and rProtein A Sepharose (both GE Healthcare) were blocked for 1 h at 4 $^{o}$C with 1 mg/ml bovine serum albumin (BSA; Sigma) and 1 mg/ml yeast tRNA (Sigma). Chromatin was pre-cleared with blocked beads for 1 h at 4 $^{o}$C. 150 µg of chromatin were incubated with antibodies ON at 4 $^{o}$C with rotation. Protein-antibody complexes were pulled down by adding beads to the solution for 2 h. Complexes were washed 4 times with Low Salt Buffer (1 % SDS, 1 % triton X-100, 2 mM EDTA pH 8.0, 150 mM NaCl and 200 mM Tris-HCl pH 8.0 with complete EDTA-free protease inhibitors), followed by 1 wash with High Salt Buffer (0.1 % SDS, 1 % triton X-100, 500 mM NaCl and 20 mM Tris-HCl pH 8.0 with complete EDTA-free protease inhibitors) and a final wash in 1 x TE. Samples were treated with 150 µg PK and 50 µg RNase A and incubated for 2 h at 37 $^{o}$C. Crosslinks were reversed

ON at 65 $^{o}$C with mixing at 800 rpm for 1 min every 2 min. DNA was then purified using a PureLink PCR micro column (Invitrogen). When DNA was used for qRT-PCR it was diluted to a final volume of 150 µl and 5µl were used per PCR reaction (see section 6.6.3.1). Primers are listed in Table 6.6.

## 6.8. Deep-sequencing

### 6.8.1. Library preparation and sequencing

ChIP DNA was end-repaired, A-tailed and adapter-ligated using a ChIP-Seq DNA Sample prep Kit (Illumina) following manufacturer's instructions. A Qiagen PCR purification step was performed following each enzymatic reaction. The adapter-ligated material was then PCR amplified, using 18 cycles of PCR before size selection of 200-500 bp fragments on an agarose gel. The library was then extracted using a Qiaquick gel extraction kit and checked for concentration and integrity. DNA was then sequenced using a Solexa 2G instrument. To give more depth to the analysis three lanes were used for the *Eed*$^{-/-}$ sample and input and one lane for WT ESCs.

### 6.8.2. Data analysis

ChIP-Seq data analysis was performed in collaboration with Steve Taylor (Weatherall Institute of Molecular Medicine, Oxford, UK).

51 bp single-end tags were mapped to the mouse genome (mm9) using bowtie (Langmead et al., 2009). Tags with more than one overlap were excluded from further analysis using the m = 1 parameter. Final datasets included 16,814,546 tags for the WT sample, 52,569,530 for the Input and 70,343,706 for the *Eed* KO. Tags aligned to the mouse genome were visualized in GBrowse (Stein et al., 2002). The quality of the tags was analysed using SeqMonk version 0.11 (www.bioinformatics.bbsrc.ac.uk/projects/seqmonk/). Raw data was quantitated by designing probes using the contig probe generator. To reduce noise only probes 5 fold enriched and wider

than 100 bp were considered. That allowed comparison of probe patterns of distribution between the different samples. KDM2A data (Blackledge et al., 2010) was also exported and analysed as described. Enrichment peaks were extracted using a Model-based Analysis of ChIP-Seq (MACS) (Zhang et al., 2008) with a false discovery rate (FDR) < 2 % and number of tags in the peak > 100. Nearest gene and overlaps to location was done using Cisgenome (Ji et al., 2008) and custom scripts. Peak intersection analysis was done using intersectBed (Quinlan and Hall, 2010) between wild-type and KO peaks. IntersectBed was also used for peak overlap with CpG islands (CpG data downloaded from UCSC mm9 table browser) and H3K4me3 and H3K27me3 data from ESCs (Mikkelsen et al., 2007).

#### 6.8.2.1. Repeat sequences analysis

The Repbase Mouse version 15.04 database was searched using bowtie (option –m 1) in KO, WT and Control samples. The instances of the number of hits of each repeat were counted and then normalized by dividing by the number of sequences in each group. No significant differences in enrichment for repeats were found.

#### 6.8.2.2. Gene ontology functional analysis

GO enrichment was carried out to compare *Eed* WT with *Eed* KO RING1B peaks and WT RING1B with WT H3K4/H3K27 peaks (Mikkelsen et al., 2007) using DAVID gene ontology functional annotation tool (Dennis et al., 2003; Huang et al., 2009). Bonferroni correction for multiple hypotheses was used for correcting P-values.

## 6.9. Protein work

### 6.9.1. Protein extraction

#### 6.9.1.1. Total protein extracts

Cells growing on dishes (when feederless) were washed twice with ice-cold PBS. 6 x SDS Loading Buffer (360 mM Tris-HCl pH 6.8, 34 % glycerol (VWR), 12 % SDS, 30 % β-mercaptoethanol (Sigma), 120 µg/ml bromophenol blue (Sigma)) was added to the plate to a final concentration of 2 x, incubated 1-2 min and mixed well. When cells were grown on feeder cells, cells were pre-plated for 30 min, the supernatant containing the floating cells removed to a falcon tube, spun down, washed twice with ice-cold PBS and 6 x SDS Loading Buffer was added to a final concentration of 2 x. Cell lysate was boiled at 95 $^{o}$C for 10 min and vigorously shaken.

#### 6.9.1.2. Nuclear extract

Cells were harvested by trypsinization (section 6.1.2) and cell pellets were either processed immediately or snap-frozen on dry ice and stored at -80 $^{o}$C. The cell pellets, washed with PBS, resuspended in Buffer H (20 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl (VWR), 1 mM DTT (Sigma)) and the plasma membrane was then disrupted by passing cells through a 27 G needle. Nuclei were spun down (5 min x 13000 rpm) and resuspended in Buffer D (20 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.42 M KCl, 20 % Glycerol and 1 mM DTT). Extraction of nuclear proteins was performed at 4 $^{o}$C for two hours with vigorous shaking. The nuclear fraction was cleared by centrifugation at 13000 rpm for 5 min at 4 $^{o}$C.

#### 6.9.1.3. Acidic histone extraction

Cells were harvested by trypsinization (section 6.1.2) and cell pellets were either processed immediately or snap-frozen on dry ice and stored at -80 $^{o}$C. Cell pellets were washed in RSB buffer (10 mM HCl (VWR) pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$), then lysed in RSB plus 0.5 % NP-40 for 10 min

on ice. Histone extraction was performed by resuspending the nuclear pellet in 0.5 ml 5 mM $MgCl_2$ and 0.5 ml 0.8 M HCl and incubating for 20 min on ice. The sample was spun for 20 min at 13000 rpm and the supernatant containing the histones collected. Histones were precipitated by incubation on ice for 30 min after addition of trichloroacetic acid (Fisher) to a final concentration of 25 %. Histones were pelleted by centrifugation at 13000 rpm for 15 min, washed twice with acetone and resuspended in 1 x SDS loading buffer.

### 6.9.2. Protein quantification

Protein quantification was determined by the colorimetric Bradford method with Bradford reagent (Bio-Rad) following manufacturer's instructions. Briefly Bradford reagent was diluted 5 x and 998 µl of that reagent was added to 2 µl of sample. A standard curve was generated with known concentrations of BSA diluted in diluted Bradford reagent (from 0 to 14 µg/ml). Absorbance of all samples was read in a Ultrospec 3300pro spectrophotometer (Amersham Biosciences) in triplicates. Sample concentration was extrapolated from the standard curve.

### 6.9.3. Nuclear Fractionation

Nuclear fractionation was performed as described previously (Elderkin et al., 2007) with some modifications.

Cells were trypsinized (section 6.1.2) and washed twice in PBS with complete EDTA-free protease inhibitors. The cell pellet was resuspended in 5 pellet volumes of Buffer A (50 mM Tris-HCl pH 7.9, 30 % glycerol, 1 mM EDTA, 0.5 mM DTT and complete EDTA-free protease inhibitors added fresh) and lysed by freeze-thawing once in liquid nitrogen. The sample was spun down at 1400 g for 10 min at 4 $^{o}$C and the supernatant was removed (S1 – cytoplasmic fraction). The pellet was washed in ice-cold PBS with complete EDTA-free protease inhibitors and resuspended in 5 pellet volumes of Hypertonic

Buffer (10 mM Tris-HCl pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 2 mM DTT with complete EDTA-free protease inhibitors added immediately before use). The sample was incubated for 5 min at 4 $^{o}$C and nuclei were lysed by passing the sample through a 0.5 x 16 mm syringe 4 times. The S2 fraction was collected after centrifugation at 3000 g for 8 min at 4 $^{o}$C. The pellet was resuspended in 5 pellet volumes of 100 mM buffer (50 mM Tris-HCl pH 7.5, 10 % sucrose (VWR), 100 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 20 % glycerol with complete EDTA-free protease inhibitors added fresh). The sample was incubated on ice for 20-30 min, spun down at 3000 g for 5 min at 4 $^{o}$C and the supernatant was removed (100 mM fraction). The same procedure as for 100 mM buffer was repeated using 200 mM buffer (the same as 100 mM buffer except for KCl that is present at 200 mM) and 300 mM buffer (the same as 100 mM buffer except for KCl that is present at 300 mM) and the 200 mM and 300 mM fractions were collected. The remaining pellet was resuspended in High Salt Buffer (50 mM Tris-HCl pH 7.5, 10 % sucrose, 420 mM KCl, 5 mM MgCl2, 0.1 mM EDTA, 20 % glycerol, 2 mM DTT with complete EDTA-free protease inhibitors added fresh) and spun down at 10000 g for 20 min at 4 $^{o}$C. The supernatant was removed (S3 fraction) and the pellet resuspended in 5 pellet volumes of 2 x SDS Loading Buffer. The pellet was sonicated for 30 min using a Bioruptor sonicator (30 s on 30 s off, set to high) to break up chromatin-protein complexes. Fractions were equally loaded on a SDS-PAGE gel. Cell fractionation samples were analysed by western blot (WB) with a RING1B antibody (section 6.9.5).

#### 6.9.4. Protein immunoprecipitation

The protein IP protocol was kindly provided by a collaborator (Rob Klose, personal communication).

Nuclear extracts were performed as described in section 6.9.1.2 and protein quantified as described in section 6.9.2. Each IP was performed in an eppendorf tube with 500 µg of nuclear extract. The volume was topped up to approximately 550 µl with BC150 Buffer (150 mM KCl, 10 % Glycerol, 50

mM HEPES, 0.5 mM EDTA, 0.5 mM DTT (added immediately before use) and complete EDTA-free protease inhibitors added immediately before use). To remove any precipitates, the sample was spun down at 13000 rpm for 5 min at 4 $^{o}$C. The supernatant was transferred to a fresh tube and 50 µl of the sample was removed as input. To the remaining sample (≈ 500 µl) the antibody of interest was added and incubated ON at 4 $^{o}$C on a rotating wheel. Antibodies and volumes used are listed on table 6.7. On the following day 20 µl of appropriate packed beads (40 µl of 50 % beads slurry spun down to remove supernatant) were added to the sample and incubated for 1 h at 4 $^{o}$C on a rotating wheel. Antibody-bead complexes were spun down at 1000 g for 1 min. The supernatant was removed and transferred to a fresh tube (flowthrough sample). The complexes were washed 3 x with BC300 Buffer (300 mM KCl, 10 % Glycerol, 50 mM HEPES, 0.5 mM EDTA, 0.5 mM DTT (added immediately before use) and complete EDTA-free protease inhibitors added immediately before use). Washes were performed by adding 500 µl of cold BC300 Buffer to the complexes and spinning down at 1000 g for 1 min at 4 $^{o}$C. After the final wash, the pellet was resuspended in 30 µl of 2 x SDS Loading Buffer (IP sample). The IP sample was boiled at 95 $^{o}$C for 5 min. The beads were spun down and sample was analysed on a SDS-PAGE gel. For input and flowthrough samples 6 x SDS Loading Buffer was added to a final concentration of 1 x and samples were incubated at 95 $^{o}$C for 5 min before loading into the SDS-PAGE gel. The IP was analysed by WB (section 6.9.5).

| Antibody | Volume used | Protein beads | Ab Catalogue no. | Ab Company |
|---|---|---|---|---|
| **Ring1B** | 5 µl | Protein G | | (Atsuta et al., 2001) |
| **Eed** | 5 µl | Protein G | | (Sewalt et al., 1998) |
| **Mel18** | 5 µl | Protein A | sc-10744 | Santa Cruz |
| **IgG** | 5 µl | Protein G | M7023 | Sigma |

**Table 6.7 – Antibodies and conditions used for protein IP.**

### 6.9.5. Western blot

Proteins were run on SDS-PAGE gels that were either 15 % or 10 % depending on the protein size to be detected. SDS-PAGE gels were made up of a separating and a stacking phase. The separating gel was 15 or 10 % acrylamide, (National Diagnostics), 125 mM Tris-HCl pH 8.8, 0.1 % SDS, 0.1 % ammonium persulphate (APS; Sigma) and 0.1 % tetra methyl ethylene diamine (TEMED; VWR). The stacking gel was 4 % acrylamide, 25 mM Tris-HCl pH 6.8, 0.1 % SDS, 0.1 % APS and 0.1 % TEMED. Proteins were mixed with SDS Loading Buffer to a final concentration of 1 x, boiled at 95 $^{o}$C for 5 min, spun down and loaded onto the gel with Precision Plus Protein Standards (Bio-Rad) as molecular a weight markers. The gel was run at 150 V (constant voltage) in running buffer (5 mM Tris-HCl, 38.7 mM glycine, 0.1 % SDS) after which proteins were either transferred to a membrane (WB) or stained with Coomassie. Coomassie staining was performed with Instant Blue (Triple Red) following manufacturer's instructions. Briefly, the gel was washed in water, a small amount of Instant Blue was added to cover the gel, and it was agitated for 1 h after which it was washed in water.

For WB, the gel was soaked in Transfer Buffer (20 % methanol (VWR), 9.6 mM Tris, 7.8 mM glycine, 0.0074 % SDS) for 5 min. A Hybond P polyvinylidene difluoride (PVDF) membrane (GE Healthcare) was soaked in 100 % methanol for 1 min, transferred to distilled $H_2O$ for 1 min and then incubated in transfer buffer. Two pieces of extra thick blotting paper (Bio-Rad) were also soaked in transfer buffer. The transfer stack was assembled on the Transblot SD Semidry Transfer Cell (Bio-Rad). One piece of blotting paper was placed on the bottom, then the membrane, followed by the gel and the remaining piece of blotting paper on the top. Excess buffer was removed and proteins were electrophoretically transferred at 12 V for 45 min.

Once transfer was complete, the membrane was incubated in 1 x TBS pH 8.0 (100 mM Tris-HCl pH 8.0, 0.9 % NaCl), with 5 % non-fat milk (Marvel) for 1 h with agitation. The membrane was then incubated ON with rotation at 4 $^{o}$C, in a 50 ml falcon with 3 ml of 1 x TBS pH 8.0 with a specific antibody. Primary antibodies used are listed on Table 6.8. The following day, the membrane was washed 3 x 15 min with 1 x TBS with agitation and incubated with secondary antibody in 1 x TBS pH 8.0, 5 % non-fat milk with rotation for 1 h at RT. The membrane was then washed 3 x 15 min with agitation. Protein detection was performed by enhanced chemiluminescence (ECL) with the ECL Western Blotting Analysis System (GE Healthcare) following the manufacturer's instructions followed by detection on Kodak BioMax XAR Film (Kodak).

| Antibody Name | Dilution | Cat. No. | Company |
|---|---|---|---|
| **Ubiquityl-H2A** | 1:500 | 05-678 | Millipore |
| **Ubiquitinated H2B** | 1:2000 | MM-029 | Médimabs |
| **Ring1B** | 1:500 | | (Atsuta et al., 2001) |
| **H3K27me3** | 1:1000 | CS-069-100 | Diagenode |
| **H3K4me3** | 1:1000 | 07-473 | Millipore |
| **H3** | 1:10000 | ab1791 | Abcam |
| **H2A** | 1:1000 | 07-146 | Millipore |
| **Myc-tag** | 1:1000 | 9B11 | Cell Signalling |
| **Suz12** | 1:1000 | 07-689 | Millipore |
| **Ezh2** | 1:750 | pAb-039-050 | Diagenode |
| **Eed** | 1:650 | | (Sewalt et al., 1998) |
| **KDM2A** | 1:1000 | | (Blackledge et al., 2010) |
| **KDM2B** | 1:500 | sc-101091 | Santa Cruz |
| **Lamin B M20** | 1:1000 | sc - 6217 | Santa Cruz |
| **Lamin B C20** | 1:1000 | sc - 6216 | Santa Cruz |
| **Tubulin** | 1:1000 | 2144 | Cell Signaling |
| **P16** | 1:200 | sc-468 | Santa Cruz |
| **P19** | 1:1000 | ab80 | Abcam |

**Table 6.8 – Antibodies and dilutions used for WB.**

Secondary antibodies were either sheep anti-mouse IgG HPR linked (GE Healthcare) used at a dilution of 1:2000, donkey anti-rabbit IgG HRP linked (GE Healthcare) used at a dilution of 1:20000 or donkey anti-goat IgG HRP linked (Santa Cruz) used at a dilution of 1:2000.

## 6.10. Bacteria work and plasmid construction

### 6.10.1. Plasmids

Several plasmids were used in this work. P1 derived Artificial Chromosome (PAC), with the mouse *Ink4a*/*Arf* sequence was kindly provided by Dr. Gordon Peters and Dr. Manuel Serrano (Matheu et al., 2004). The pGalK plasmid was kindly provided by Dr. Tessarollo and Dr. Court (NCI-Frederick). pDsRed Monomer (Clontech) was a kind gift from Prof. Amanda Fisher (MRC Clinical Science Centre, London, UK). pIRESneo and pECFP-C1 were from Clontech.

### 6.10.2. Media and culture conditions

Bacterial cultures used in this work were grown in liquid suspension cultures on Luria broth (LB) prepared in house. When grown on plates, 1.5 % agar (BD Biosciences) was added to LB medium. Bacteria containing PAC DNA was grown in low-salt LB medium that contained 0.5 % NaCl instead of 1 %. The low salt medium in addition to a lower incubation temperature (30 $^{o}$C) allowed the bacteria to divide slower and avoid recombination of PAC DNA.

Antibiotic selection of bacterial cultures was achieved by adding ampicillin (Sigma) to a final concentration of 100 µg/ml or kanamycin (Fluka) to a final concentration of 25 µg/ml into LB medium.

Long term bacterial cultures were kept at -80 $^{o}$C in LB/glycerol (1:1 dilution of culture and LB/glycerol). To recover a culture, a small fraction of the frozen stock was streaked on an LB agar

plate and allowed to grow. An individual colony was picked and used to inoculate 5 ml of LB. The culture was allowed to grow ON. When large cultures were required, a colony was grown on 3 ml of LB for 6-8 h after which it was transferred to 100-250 ml of LB and allowed to grow ON.

With exception of cultures with PAC DNA (grown at 30 $^{o}$C) all cultures were grown at 37 $^{o}$C with vigorous agitation.

### 6.10.3. Plasmid/PAC DNA extraction

Plasmid DNA extraction was performed using miniprep or maxiprep kits (Qiagen), following manufacturer's instructions.

To extract DNA from PAC small cultures, miniprep kit columns could not be used so a similar protocol was adapted using Qiagen buffers (Bilyana Popova, personal communication). 5 ml of bacterial culture was pelleted by centrifugation at 4500 rpm for 5 min. The pellet was then well resuspended in 250 µl of buffer P1. Lysis was performed by adding 250 µl buffer P2, mixing gently and incubating for 5 min at RT. Lysis was stopped by adding 250 µl of buffer P3, mixing by inversion and incubating on ice for 5 min. The sample was spun down at 13200 rpm for 5 min on a tabletop centrifuge to remove genomic DNA and cell debris. The supernatant was transferred to a fresh eppendorf and centrifugation was repeated. The supernatant was transferred to a clean tube and PAC DNA precipitated by the addition of 750 µl isopropanol. The sample was mixed and incubated on ice for 10 min followed by centrifugation at 13200 rpm for 10 min. The DNA pellet was washed with 70 % ethanol, air dried and resuspended in 30 µl of TE. DNA was used for restriction enzyme digestion or PCR.

Large volume cultures of PAC containing bacteria were grown by inoculating 3 ml of low-salt LB with a single colony and allowing the culture to grow at 37 $^{o}$C ON. The following day 2.5 ml of culture were transferred into 75 ml of low-salt LB and incubated for 1.5 h at 37 $^{o}$C with agitation. Isopropyl-1-thio-β-D-galactopyranoside (IPTG) was then added to a final concentration of 0.5 M and grown for 5 more hours at 30 $^{o}$C. IPTG binds to the lyptic promoter and increases PAC copy number from 1 to approximately 20 per cell. The culture was spun down at 4500 rpm for 15 min at 4 $^{o}$C in a 50 ml falcon tube. The bacterial pellet was resuspended in 12.5 ml buffer P1 followed by lysis with 17.5 ml of buffer P2, gently mixed by inverting and incubated for 5 min at RT. Lysis was stopped with the addition of 17.5 ml of buffer P3, mixed by inverting and incubated on ice for 20 min. The sample was spun down at 13000 rpm for 15 min at 4 $^{o}$C in a SS34 Sorvall centrifuge. The supernatant was removed and spun down again to remove all traces of debris. The supernatant was carefully transferred to a fresh tube to avoid genomic DNA contamination. PAC DNA was precipitated by adding 0.7 volumes of isopropanol. The precipitated DNA was collected by centrifugation at 13000 rpm for 15 min at 4 $^{o}$C in a SS34 Sorvall centrifuge. DNA was washed with 10 ml 70 % ethanol and spun down at 12 000 rpm for 10 min in a SS34 Sorvall centrifuge. A second wash was performed with 1 ml 70 % ethanol and the DNA was transferred to an epperdorf tube. The DNA was then spun down at 13000 rpm for 10 min at 4 $^{o}$C in a microcentrifuge. The pellet was then resuspended in 0.6 ml water or TE. RNAse A was added to the final concentration of 0.2 mg/ml and incubated for 30 min at 37 $^{o}$C. Phe:$CHCl_3$ was performed as described for ChIP DNA (section 6.7.1.3). The pellet was resuspended in 100 µl water.

To confirm that the PAC had not rearranged a large proportion of it was sequenced. Primers used are described in appendix 2.

### 6.10.4. Restriction enzyme digestion

Restriction enzyme digestion was performed on plasmid/PAC DNA by incubating DNA with 1 x buffer (supplied with the enzyme), 0.5 units of enzyme (Roche or NEB), BSA to a final concentration of 100 µg/ml and water for a final volume of 30 µl. For diagnostic digests approximately 500 ng of plasmid DNA was used. For cloning, 2 µg of plasmid DNA was used.

### 6.10.5. Arms of homology

Primer sequences for the entire recombineering process are shown in Table 6.9.

To amplify the galK ORF from the pGalK plasmid, the primers engineered to have an arm of homology to the desired sequence on its 5' end. The galK ORF was amplified using these primers flanked by arms of homology to the PAC. The Expand High Fidelity$^{PLUS}$ PCR System, dNTPack (Roche) was used. The PCR product was run on an agarose gel, purified with a Gel Purification Kit (Qiagen), cloned into a StrataClone vector with the StrataClone PCR cloning kit (Stratagene) and sequenced to check for point mutations.

For the dsRed insert, primers were also engineered to have homology arms to the ARF sequence at its 5' end. The subsequent procedure was the same as described for galK.

ECFP was amplified from pECFP-C1 with BstXI and AgeI on its 5' end and BstXI on its 3' end and cloned into a PCR cloning vector. The plasmid was linearised with BstXI and cloned into pIRESneo also linearised with BstXI. A clone with correct orientation was identified and named pECFP-IRESneo. The 3' homology arm was amplified with XhoI site added to the 5' end and HpaI and XhoI on the 3' end and cloned into a PCR cloning vector. It was linearised with XhoI and cloned into pECPP-IRESneo also linearised with XhoI. A clone with correct orientation was identified and named pECFP-

IRESneo3'. The 5' homology arm was amplified with BamHI inserted to the 5' end and AgeI inserted on its 3' end and cloned into a PCR cloning vector. It was cut with both BamHI and AgeI and was inserted into pECFP-IRESneo3' previously digested with the same enzymes. A positive clone was identified and fully sequenced to confirm that no mutation had arisen during the cloning process. Sequencing primers used were the same as the ones described on table 6.2. The entire insert was cut by digestion with BamHI and HpaI, run on a gel and purified before recombineering.

| Name | Sequence (5'-3') |
|---|---|
| **MouArfgalKForw** | ACCTCGCTTGTCACAGTGAGGCCGCCGCTGAGGGAGTACAGCAGCGGGAGCCTGTTGACAATTAATCATCGGCA |
| **MousArfgalKRev** | GGGCGGCCCGCGCGCTGAATCCTCACAGTGACCAAGAACCTGCGACCCATTCAGCACTGTCCTGCTCCTT |
| **MouArfDsRedForw** | ACCTCGCTTGTCACAGTGAGGCCGCCGCTGAGGGAGTACAGCAGCGGGAGACCATGGCCTCCTCCGAGGA |
| **MouArfDsRedRev** | GGGCGGCCCGCGCGCTGAATCCTCACAGTGACCAAGAACCTGCGACCCATGGGAGGTTTTTTAAAGCAAG |
| **MouINK4agalKForw** | ACTGAATCTCCGCGAGGAAAGCGAACTCGAGGAGAGCCATCTGGAGCAGCCCTGTTGACAATTAATCATCGGCA |
| **MousINK4agalKRev** | ACACGGCCCTGGGCCGCCGCCCTGGCCAGTCTGTCTGCAGCGGACTCCATTCAGCACTGTCCTGCTCCTT |
| **ECFP BstXI AgeI For** | CCAGTGTGCTGGACCGGTCATGGTGAGCAAGGGCGA |
| **ECFP BstXI Rev** | CCAGTGTGCTGGTGAGTCCGGACTTGTACAGCTCGT |
| **MouseINK4a_5_F** | GGATCCTAAGCAGATTGCCCTCCGATGACT |
| **MouseINK4a_5_R** | ACCGGTGCTGCTCCAGATGGCTCTCCTCGA |
| **MouseINK4a_3_F** | CTCGAGATGGAGTCCGCTGCAGACAGACTG |
| **MouseINK4a_3_R** | CTCGAGGTTAACTAGGACTCCAAAGGCAAACTTCTC |

**Table 6.9 – Primer used for recombineering.** Sequences in green correspond to arms of homology or external digestion sites added to the primers.

### 6.10.6. Ligation

To clone 2 fragments of plasmid DNA, ligation was performed. The backbone was first dephosphorylated to avoid self-ligation. This was achieved by adding 1 unit of calf intestine alkaline phosphatise (CIAP) (Invitrogen) to the linearised backbone DNA either in digestion buffer or in 1 x CIAP Buffer. The sample was incubated for 5 min at 37 $^{o}$C for 3' overhangs or 5 min at 50 $^{o}$C for blunt end ligation. DNA was run on a 1.5 % agarose gel to separate different bands. The correct size band was cut and DNA was purified.

Ligations were performed by adding backbone and insert DNA in a 1:1 and 1:3 molar ratios with 3 units of T4 DNA ligase (Promega) and 1 x T4 ligase buffer and incubated for 3 h at RT or ON at 4 $^{o}$C.

### 6.10.7. Transformation

Ligation products were transformed into competent XL2-Blue MRF' ultracompetent cells (Stratagene). Cells were thawed on ice, plasmid DNA was added and cells were incubated on ice for 30 min. 10 to 50 µl of ultracompetent cells were used per transformation. The volume of DNA ligation mixture added was 1/10 of the volume of cells. 1-10 ng of plasmid DNA in 1 µl was generally used. Cells were heat-shocked at 42 $^{o}$C for 30 s to open pores on the bacterial cell wall and allow DNA in. Cells were then incubated for 2 min on ice followed by addition of 250 µl pre-warmed LB medium and recovered by incubation at 37 $^{o}$C for 1h with agitation. Cells were then plated at various densities on LB agar plates with ampicillin and incubated ON at 37 $^{o}$C with agitation. When pBluescript or a PCR cloning plasmid were used a blue/white screen was possible. For colour selection LB plates contained 80 µg/ml of X-gal and 20 mM of IPTG. Individual colonies were analysed for the correct plasmid insert (if blue/white colour screen was performed, only white cells were picked as blue ones do not have the insert). Further analysis was carried out by restriction digestion and sequencing of positive clones.

### 6.10.8. Recombination-mediated genetic engineering

The recombineering (recombination-mediated genetic engineering) technique used in this study is based on *galK* positive/negative selection (Warming et al., 2005). Recombineering was performed according to the protocol available at http://recombineering.ncifcrf.gov/protocol/Protocol3_SW102_galK_v2.pdf.

#### 6.10.8.1. Media and plates

M9 salts medium was composed of 42 mM sodium hydrogen phosphate (VWR), 22 mM potassium hydrogen phosphate (VWR), 17mM ammonium chloride (VWR) and 9 mM NaCl. It was autoclaved before use.

M63 minimal media was needed for M63 minimal plates. 5 x M63 media was 75 mM ammonium sulphate (VWR), 500 mM potassium hydrogen phosphate, and 0.9 mM ferrous sulphate (Sigma) pH 7. It was autoclaved before use. M63 minimal plates contained 1.5 % agar, 1 x M63 minimal medium, 1 µl/ml Biotin (Sigma) and 45 µl/ml leucine (Sigma). When plates were used for positive selection they contained 0.2 % galactose (Sigma) and when used for negative selection contained 0.2 % glycerol and 0.2 % deoxygalactose (DOG) (Sigma). Antibiotic selection was achieved by adding kanamycin to the final concentration of 25 µg/ml.

MacConkey Indicator Plates were made by autoclaving 2.5 % MacConkey agar (BD Biosciences) in water and then adding 2% galactose and autoclaving. Antibiotic selection was achieved by addition of kanamycin to the final concentration of 25 µg/ml.

#### 6.10.8.2. Electroporation of PACs into SW102 cells

The SW102 bacterial strain was used to carry out recombineering of dsRed and ECFP-IRES-Neomycin sequences into the PAC described above. The SW102 bacterial strain was kindly provided by Dr. Tessarollo and Dr. Court (NCI-Frederick).

PAC DNA was electroporated into SW102 cells. SW102 cells were inoculated in 5 ml of low-salt LB and grown ON at 30 $^{o}$C with agitation. The following morning 500 µl of ON culture was inoculated into 25 ml of low-salt LB in a 300 ml flask and incubated at 30 $^{o}$C with shacking until the $OD_{600}$ was

with the range 0.5-0.6. At this point 10 ml of culture was transferred to a 14 ml round bottom tube, chilled on ice and then spun down at 4500 rpm for 5 min at 2 $^{o}$C. Supernatant was removed and 1 ml of ice-cold distilled water ($dH_2O$) was added. Cells were resuspended by swirling the tube in an ice-cold water-bath and 9 ml of ice-cold $dH_2O$ was then added. Cells were spun down at 4500 rpm for 5 min at 2 $^{o}$C. Cells were washed 2 more times. After the last wash as much supernatant as possible was removed and the cells were resuspended in the few microliters of supernatant left (≈ 50 µl). PAC DNA was added to the cells (1 to 5 µg in 1-5 µl water) and the suspension was transferred into a pre-chilled 0.1 cm electroporation cuvette (Bio-Rad). Cells were electroporated at 1.75 kV, 200 Ω and 15 µF. Cells were recovered by adding 1 ml of LB into the cuvette and cells were transferred to a falcon tube where were incubated at 30 $^{o}$C for 1 h with agitation. Cells were then transferred to an eppendorf tube and spun down at 5000 rpm for 3 min. Supernatant was removed except for approximately 200 µl that was left to resuspend the pellet. Cells were spread (20 µl and 180 µl) on low-salt LB agar plates with kanamycin and incubated at 30 $^{o}$C for 24h.

#### 6.10.8.3. Recombineering

The evening before starting 5 ml of low-salt LB with kanamycin were inoculated with a SW102 colony containing the PAC and incubated at 30 $^{o}$C ON with agitation. The following day, 25 ml of low-salt LB with kanamycin were inoculated with 500 µl and incubated at 30 $^{o}$C with agitation until the culture reached $OD_{600}$ of 0.5-0.6. When that happened, 10 ml of culture was transferred to another flask and was heat-shocked at 42 $^{o}$C in a shaking water-bath for 15 min to induce recombineering proteins (induced). The remaining culture was kept at 30 $^{o}$C (uninduced). Heat shock induces the expression of the lambda prophage genes *exo*, *bet* and *gam* that encode proteins required to carry out homologous recombination. After the incubation, both induced and uninduced cultures were briefly cooled down in an ice-water slurry and transferred into 14 ml round bottom falcon tubes. Cells were washed as described in section 6.10.8.2. Once cells were resuspended in ≈ 50 µl of

supernatant, gel purified PCR product (10-30 ng) or linearised plasmid (200 ng) containing *galK* ORF flanked by arms of homology to the targeted region on the PAC was added. Cells were electroporated at 1.75 kV, 200 Ω and 25 µF and recovered by adding 1 ml low-salt LB and incubating for 1 h at 30 $^{o}$C with agitation. Cells were then transferred to an eppendorf and spun down at 13000 rpm for 15 s. Supernatant was removed and cells were resuspended in 1 ml (induced) and 500 µl (uninduced) of M9 salts medium to remove any rich medium prior to selection on minimal medium plates. Cells were spun at 13000 rpm for 15 s. Cells were washed again, resuspended in 1 ml of M9 medium and plated in serial dilutions on M63 minimal medium plates with galactose and kanamycin to select for colonies containing *galk*. Plates were incubated at 30 $^{o}$C for 3 days.

On the third day visible colonies were streaked on MacConkey indicator plate with galactose and kanamycin and incubated ON at 30 $^{o}$C. Bacterial colonies containing the *galk* cassette will be bright pink/red due to a pH change triggered by fermented galactose while bacteria without *galk* will be white/colourless. Colonies containing *galk* were grown, DNA extracted and further analysed by restriction digestion and PCR to confirm the presence of galk.

A positive colony was grown and the recombineering protocol repeated as described above with minor differences. Instead of a plasmid/PCR containing *galk* ORF, a plasmid containing the insert of interest flanked by the homology arms to the PAC DNA was used. In this step the *galk* cassette is replaced by the cassette containing the fluorochrome. After electroporation 1 ml of low-salt LB was added to the cells that were transferred to 9 ml of low-salt LB in a conic flask and incubated for 4 h at 30 $^{o}$C with agitation to recover. 1 ml of cells were transferred to an eppendorf, spun down and washed twice in M9 salts to remove medium as described above. Cells were plated in M63 minimal medium plates containing DOG as a carbon source and kanamycin as antibiotic selection. Plates were incubated for 3 days at 30 $^{o}$C. This step selects cells that loose the *galk* ORF as *galk* containing

bacteria can metabolize DOG into deoxygalactose phosphate which cannot be metabolised further becoming toxic for the cells. Colonies were then expanded, DNA was extracted and analysed by restriction digestion, PCR and sequencing to confirm the insert and its DNA sequence. Primers sequences used to sequence and confirm PAC recombineering are described on Table 6.2.

# 7. Bibliography

Agger, K., Cloos, P.A.C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature *449*, 731-734.

Agherbi, H., Gaussmann-Wenger, A., Verthuy, C., Chasson, L., Serrano, M., and Djabali, M. (2009). Polycomb mediated epigenetic silencing and replication timing at the INK4a/ARF locus during senescence. PLoS One *4*, e5622.

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191-1200.

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., and Bazett-Jones, D.P. (2010). Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PLoS One *5*, e10531.

Akasaka, T., Tsuji, K., Kawahira, H., Kanno, M., Harigaya, K., Hu, L., Ebihara, Y., Nakahata, T., Tetsu, O., Taniguchi, M.*, et al.* (1997). The role of mel-18, a mammalian Polycomb group gene, during IL-7-dependent proliferation of lymphocyte precursors. Immunity *7*, 135-146.

Akasaka, T., van Lohuizen, M., van der Lugt, N., Mizutani-Koseki, Y., Kanno, M., Taniguchi, M., Vidal, M., Alkema, M., Berns, A., and Koseki, H. (2001). Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. Development *128*, 1587-1597.

Alder, O., Lavial, F., Helness, A., Brookes, E., Pinho, S., Chandrashekran, A., Arnaud, P., Pombo, A., O'Neill, L., and Azuara, V. (2010). Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment. Development *137*, 2483-2492.

Altaf, M., Auger, A., Monnet-Saksouk, J., Brodeur, J., Piquet, S., Cramet, M., Bouchard, N., Lacoste, N., Utley, R.T., Gaudreau, L.*, et al.* (2010). NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex. J Biol Chem *285*, 15966-15977.

Arrigoni, R., Alam, S.L., Wamstad, J.A., Bardwell, V.J., Sundquist, W.I., and Schreiber-Agus, N. (2006). The Polycomb-associated protein Rybp is a ubiquitin binding protein. FEBS Lett *580*, 6233-6241.

Atsuta, T., Fujimura, S., Moriya, H., Vidal, M., Akasaka, T., and Koseki, H. (2001). Production of monoclonal antibodies against mammalian Ring1B proteins. Hybridoma *20*, 43-46.

Avner, P., and Heard, E. (2001). X-chromosome inactivation: counting, choice and initiation. Nat Rev Genet *2*, 59-67.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M.*, et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Bandyopadhyay, D., Okan, N.A., Bales, E., Nascimento, L., Cole, P.A., and Medrano, E.E. (2002). Down-regulation of p300/CBP histone acetyltransferase activates a senescence checkpoint in human melanocytes. Cancer Res *62*, 6231-6239.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Barradas, M., Anderton, E., Acosta, J.C., Li, S., Banito, A., Rodriguez-Niedenfuhr, M., Maertens, G., Banck, M., Zhou, M.-M., Walsh, M.J.*, et al.* (2009). Histone demethylase JMJD3 contributes to epigenetic control of INK4a/ARF by oncogenic RAS. Genes Dev *23*, 1177-1182.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bartkova, J., Rezaei, N., Liontos, M., Karakaidos, P., Kletsas, D., Issaeva, N., Vassiliou, L.-V.F., Kolettas, E., Niforou, K., Zoumpourlis, V.C.*, et al.* (2006). Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature *444*, 633-637.

Bedford, M.T. (2007). Arginine methylation at a glance. J Cell Sci *120*, 4243-4246.

Bergink, S., Salomons, F.A., Hoogstraten, D., Groothuis, T.A.M., de Waard, H., Wu, J., Yuan, L., Citterio, E., Houtsmuller, A.B., Neefjes, J.*, et al.* (2006). DNA damage triggers nucleotide excision repair-dependent monoubiquitylation of histone H2A. Genes Dev *20*, 1343-1352.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669-681.

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

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006b). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol *26*, 2560-2569.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. (2003). Dicer is essential for mouse development. Nat Genet *35*, 215-217.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes Dev *23*, 2484-2489.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.

Birke, M., Schreiner, S., Garcia-Cuellar, M.-P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res *30*, 958-965.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. Mol Cell *38*, 179-190.

Bohm, L., Crane-Robinson, C., and Sautiere, P. (1980). Proteolytic digestion studies of chromatin core-histone structure. Identification of a limit peptide of histone H2A. Eur J Biochem *106*, 525-530.

Bostick, M., Kim, J.K., Esteve, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

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

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

Bracken, A.P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., Theilgaard-Monch, K., Minucci, S., Porse, B.T., Marine, J.-C*., et al.* (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. Genes Dev *21*, 525-530.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. Nature *371*, 435-438.

Breiling, A., Turner, B.M., Bianchi, M.E., and Orlando, V. (2001). General transcription factors bind promoters repressed by Polycomb group proteins. Nature *412*, 651-655.

Brosnan, C.A., and Voinnet, O. (2009). The long and the short of noncoding RNAs. Curr Opin Cell Biol *21*, 416-425.

Brueckner, F., and Cramer, P. (2008). Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. Nat Struct Mol Biol *15*, 811-818.

Bruggeman, S.W.M., Valk-Lingbeek, M.E., van der Stoop, P.P.M., Jacobs, J.J.L., Kieboom, K., Tanger, E., Hulsman, D., Leung, C., Arsenijevic, Y., Marino, S*., et al.* (2005). Ink4a and Arf differentially affect cell proliferation and neural stem cell self-renewal in Bmi1-deficient mice. Genes Dev *19*, 1438-1443.

Buschbeck, M., Uribesalgo, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., Morey, L., Guigo, R., Lopez-Schier, H., and Di Croce, L. (2009). The histone variant macroH2A is an epigenetic regulator of key developmental genes. Nat Struct Mol Biol *16*, 1074-1079.

Campos, E.I., and Reinberg, D. (2009). Histones: annotating chromatin. Annu Rev Genet *43*, 559-599.

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

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

Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell *15*, 57-67.

Carroll, S.B. (2000). Endless forms: the evolution of gene regulation and morphological diversity. Cell *101*, 577-580.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M.P*., et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Chadwick, B.P., and Willard, H.F. (2001). Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. Hum Mol Genet *10*, 1101-1113.

Chamberlain, S.J., Yee, D., and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem Cells *26*, 1496-1505.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230-1234.

Chambers, I., and Tomlinson, S.R. (2009). The transcriptional foundation of pluripotency. Development *136*, 2311-2322.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J*., et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Cho, H., Orphanides, G., Sun, X., Yang, X.J., Ogryzko, V., Lees, E., Nakatani, Y., and Reinberg, D. (1998). A human RNA polymerase II complex containing factors that modify chromatin structure. Mol Cell Biol *18*, 5355-5363.

Chodosh, L.A., Fire, A., Samuels, M., and Sharp, P.A. (1989). 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. J Biol Chem *264*, 2250-2257.

Choi, J.K., and Howe, L.J. (2009). Histone acetylation: truth of consequences? Biochem Cell Biol *87*, 139-150.

Chow, J., and Heard, E. (2009). X inactivation and the complexities of silencing a sex chromosome. Curr Opin Cell Biol *21*, 359-366.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.

Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Costanzi, C., and Pehrson, J.R. (2001). MACROH2A2, a new member of the MARCOH2A core histone family. J Biol Chem *276*, 21776-21784.

Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C*., et al.* (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat Genet *24*, 132-138.

Cui, K., Zang, C., Roh, T.-Y., Schones, D.E., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell *4*, 80-93.

Culi, J., Aroca, P., Modolell, J., and Mann, R.S. (2006). jing is required for wing development and to establish the proximo-distal axis of the leg in Drosophila melanogaster. Genetics *173*, 255-266.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185-196.

Dantuma, N.P., Groothuis, T.A.M., Salomons, F.A., and Neefjes, J. (2006). A dynamic ubiquitin equilibrium couples proteasomal activity to chromatin remodeling. J Cell Biol *173*, 19-26.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol *319*, 1097-1113.

de la Serna, I.L., Ohkawa, Y., and Imbalzano, A.N. (2006). Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. Nat Rev Genet *7*, 461-473.

de Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M*., et al.* (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev Cell *7*, 663-676.

Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C.R., and Pirrotta, V. (2004). Polycomb silencing blocks transcription initiation. Mol Cell *13*, 887-893.

Denisenko, O., Shnyreva, M., Suzuki, H., and Bomsztyk, K. (1998). Point mutations in the WD40 domain of Eed block its interaction with Ezh2. Mol Cell Biol *18*, 5634-5642.

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome biology *4*, P3.
Di Micco, R., Fumagalli, M., Cicalese, A., Piccinin, S., Gasparini, P., Luise, C., Schurra, C., Garre, M., Nuciforo, P.G., Bensimon, A*., et al.* (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. Nature *444*, 638-642.

Dietrich, N., Bracken, A.P., Trinh, E., Schjerling, C.K., Koseki, H., Rappsilber, J., Helin, K., and Hansen, K.H. (2007). Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus. Embo J *26*, 1637-1648.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Solda, G., Simons, C.*, et al.* (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res *18*, 1433-1445.

Dutrow, N., Nix, D.A., Holt, D., Milash, B., Dalley, B., Westbroek, E., Parnell, T.J., and Cairns, B.R. (2008). Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. Nat Genet *40*, 977-986.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D.G., Buetow, K.H.*, et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437-447.

Elderkin, S., Maertens, G.N., Endoh, M., Mallery, D.L., Morrice, N., Koseki, H., Peters, G., Brockdorff, N., and Hiom, K. (2007). A phosphorylated form of mel-18 targets the Ring1B histone H2A ubliquitin ligase to chromatin. Mol Cell *28*, 107-120.

Eltsov, M., Maclellan, K.M., Maeshima, K., Frangakis, A.S., and Dubochet, J. (2008). Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. Proc Natl Acad Sci U S A *105*, 19732-19737.

Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y.-i., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M.*, et al.* (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development *135*, 1513-1524.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A.*, et al.* (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol Cell *38*, 452-464.

Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.h., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell *136*, 1122-1135.

Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. Curr Biol *11*, 1183-1187.

Fang, J., Chen, T., Chadwick, B., Li, E., and Zhang, Y. (2004). Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. J Biol Chem *279*, 52812-52815.

Faust, C., Lawson, K.A., Schork, N.J., Thiel, B., and Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. Development *125*, 4495-4506.

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

Fazzari, M.J., and Greally, J.M. (2004). Epigenomics: beyond CpG islands. Nat Rev Genet *5*, 446-455.

Fazzio, T.G., Huff, J.T., and Panning, B. (2008). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell *134*, 162-174.

Ficz, G., Heintzmann, R., and Arndt-Jovin, D.J. (2005). Polycomb group protein complexes exchange rapidly in living Drosophila. Development *132*, 3963-3976.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A *73*, 1897-1901.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev *17*, 1870-1881.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell *2*, 160-169.

Francis, N.J. (2009). Mechanisms of epigenetic inheritance: copying of polycomb repressed chromatin. Cell Cycle *8*, 3513-3518.

Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. Cell *137*, 110-122.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574-1577.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a functional core polycomb repressive complex. Mol Cell *8*, 545-556.

Gambetta, M.C., Oktaba, K., and Muller, J. (2009). Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. Science *325*, 93-96.

Gao, X., Tate, P., Hu, P., Tjian, R., Skarnes, W.C., and Wang, Z. (2008). ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. Proc Natl Acad Sci U S A *105*, 6656-6661.

Garcia, B.A., Pesavento, J.J., Mizzen, C.A., and Kelleher, N.L. (2007). Pervasive combinatorial modification of histone H3 in human cells. Nat Methods *4*, 487-489.

Garcia, E., Marcos-Gutierrez, C., del Mar Lorente, M., Moreno, J.C., and Vidal, M. (1999). RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. Embo J *18*, 3404-3418.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E*., et al.* (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature *460*, 863-868.

Gearhart, M.D., Corcoran, C.M., Wamstad, J.A., and Bardwell, V.J. (2006). Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. Mol Cell Biol *26*, 6880-6889.

Gil, J., and Peters, G. (2006). Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. Nat Rev Mol Cell Biol *7*, 667-677.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

Goldberg, A.D., Banaszynski, L.A., Noh, K.-M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X.*, et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678-691.

Grewal, S.I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. Curr Opin Genet Dev *20*, 134-141.

Grewal, S.I.S., and Elgin, S.C.R. (2007). Transcription and RNA interference in the formation of heterochromatin. Nature *447*, 399-406.

Grewal, S.I.S., and Jia, S. (2007). Heterochromatin revisited. Nat Rev Genet *8*, 35-46.
Grewal, S.I.S., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science *301*, 798-802.

Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006a). RNAi components are required for nuclear clustering of Polycomb group response elements. Cell *124*, 957-971.

Grimaud, C., Negre, N., and Cavalli, G. (2006b). From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. Chromosome Res *14*, 363-375.

Guerrero-Santoro, J., Kapetanaki, M.G., Hsieh, C.L., Gorbachinsky, I., Levine, A.S., and Rapic-Otrin, V. (2008). The cullin 4B-based UV-damaged DNA-binding protein ligase binds to UV-damaged chromatin and ubiquitinates histone H2A. Cancer Res *68*, 5014-5022.

Han, Z., Xing, X., Hu, M., Zhang, Y., Liu, P., and Chai, J. (2007). Structural basis of EZH2 recognition by EED. Structure *15*, 1306-1315.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol *10*, 1291-1300.

Harris, H., and Hopkinson, D.A. (1976). Handbook of Enzyme Electrophoresis in Human Genetics (Amsterdam, North-Holland).

Hayflick, L. (1961). The establishment of a line (WISH) of human amnion cells in continuous cultivation. Exp Cell Res *23*, 14-20.

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

Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. Science *315*, 1141-1143.

Herman, J.G., Jen, J., Merlo, A., and Baylin, S.B. (1996). Hypermethylation-associated inactivation indicates a tumor suppressor role for p15INK4B. Cancer Res *56*, 722-727.

Hieda, M., Winstanley, H., Maini, P., Iborra, F.J., and Cook, P.R. (2005). Different populations of RNA polymerase II in living mammalian cells. Chromosome Res *13*, 135-144.

Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. Nature *463*, 474-484.

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc Natl Acad Sci U S A *106*, 5181-5186.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M., and Chaillet, J.R. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell *104*, 829-838.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J*., et al.* (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol *6*, e22.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS Lett *583*, 1713-1720.

Ingham, P.W. (1998). trithorax and the regulation of homeotic gene expression in Drosophila: a historical perspective. Int J Dev Biol *42*, 423-429.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Jacobs, J.J., Kieboom, K., Marino, S., DePinho, R.A., and van Lohuizen, M. (1999). The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. Nature *397*, 164-168.
Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M., and Wong, W.H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nature biotechnology *26*, 1293-1300.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Joo, H.-Y., Zhai, L., Yang, C., Nie, S., Erdjument-Bromage, H., Tempst, P., Chang, C., and Wang, H. (2007). Regulation of cell cycle progression and gene expression by H2A deubiquitination. Nature *449*, 1068-1072.

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Mol Cell *20*, 971-978.

Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb complexes and the propagation of the methylation mark at the Drosophila ubx gene. J Biol Chem *281*, 29064-29075.

Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V.A., and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol *8*, 285-292.

Kallin, E.M., Cao, R., Jothi, R., Xia, K., Cui, K., Zhao, K., and Zhang, Y. (2009). Genome-wide uH2A localization analysis highlights Bmi1-dependent deposition of the mark at repressed genes. PLoS Genet *5*, e1000506.

Kanellopoulou, C., Muljo, S.A., Dimitrov, S.D., Chen, X., Colin, C., Plath, K., and Livingston, D.M. (2009). X chromosome inactivation in the absence of Dicer. Proc Natl Acad Sci U S A *106*, 1122-1127.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev *19*, 489-501.

Kanhere, A., Viiri, K., Araujo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W*., et al.* (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell *38*, 675-688.

Kapetanaki, M.G., Guerrero-Santoro, J., Bisi, D.C., Hsieh, C.L., Rapic-Otrin, V., and Levine, A.S. (2006). The DDB1-CUL4ADDB2 ubiquitin ligase is deficient in xeroderma pigmentosum group E and targets histone H2A at UV-damaged DNA sites. Proc Natl Acad Sci U S A *103*, 2588-2593.

Kaplan, C.D., Larsson, K.-M., and Kornberg, R.D. (2008). The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. Mol Cell *30*, 547-556.

Kay, G.F., Penny, G.D., Patel, D., Ashworth, A., Brockdorff, N., and Rastan, S. (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. Cell *72*, 171-182.

Kennison, J.A. (1995). The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. Annu Rev Genet *29*, 289-303.

Kennison, J.A. (2004). Introduction to Trx-G and Pc-G genes. Methods Enzymol *377*, 61-70.

Keogh, M.-C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J*., et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

Kerppola, T.K. (2009). Polycomb group complexes--many combinations, many functions. Trends Cell Biol *19*, 692-704.

Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S., and Simon, J.A. (2005). Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. Mol Cell Biol *25*, 6857-6868.

Kidder, B.L., Palmer, S., and Knott, J.G. (2009). SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. Stem Cells *27*, 317-328.

Kim, D.H., Villeneuve, L.M., Morris, K.V., and Rossi, J.J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat Struct Mol Biol *13*, 793-797.

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

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.

Kim, J.K., Huh, S.O., Choi, H., Lee, K.S., Shin, D., Lee, C., Nam, J.S., Kim, H., Chung, H., Lee, H.W*., et al.* (2001). Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. Mol Cell Biol *21*, 7787-7795.

King, I.F.G., Emmons, R.B., Francis, N.J., Wild, B., Muller, J., Kingston, R.E., and Wu, C.-T. (2005). Analysis of a polycomb group protein defines regions that link repressive activity on nucleosomal templates to in vivo function. Mol Cell Biol *25*, 6578-6591.

Klochendler-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C., and Yaniv, M. (2000). The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. EMBO Rep *1*, 500-506.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci *31*, 89-97.

Kobor, M.S., Venkatasubrahmanyam, S., Meneghini, M.D., Gin, J.W., Jennings, J.L., Link, A.J., Madhani, H.D., and Rine, J. (2004). A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. PLoS Biol *2*, E131.

Koken, M.H., Reynolds, P., Jaspers-Dekker, I., Prakash, L., Prakash, S., Bootsma, D., and Hoeijmakers, J.H. (1991). Structural and functional conservation of two human homologs of the yeast DNA repair gene RAD6. Proc Natl Acad Sci U S A *88*, 8865-8869.

Kotake, Y., Cao, R., Viatour, P., Sage, J., Zhang, Y., and Xiong, Y. (2007). pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16INK4alpha tumor suppressor gene. Genes Dev *21*, 49-54.
Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krimpenfort, P., Quon, K.C., Mooi, W.J., Loonstra, A., and Berns, A. (2001). Loss of p16Ink4a confers susceptibility to metastatic melanoma in mice. Nature *413*, 83-86.

Krishnamurthy, J., Torrice, C., Ramsey, M.R., Kovalev, G.I., Al-Regaiey, K., Su, L., and Sharpless, N.E. (2004). Ink4a/Arf expression is a biomarker of aging. J Clin Invest *114*, 1299-1307.

Krogan, N.J., Keogh, M.-C., Datta, N., Sawa, C., Ryan, O.W., Ding, H., Haw, R.A., Pootoolal, J., Tong, A., Canadien, V*., et al.* (2003). A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. Mol Cell *12*, 1565-1576.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S*., et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet *4*, e1000242.

Kusch, T., Florens, L., Macdonald, W.H., Swanson, S.K., Glaser, R.L., Yates, J.R., Abmayr, S.M., Washburn, M.P., and Workman, J.L. (2004). Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. Science *306*, 2084-2087.

Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004). Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Mol Cell *14*, 183-193.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev *16*, 2893-2905.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev *22*, 2799-2810.

Lal, A., Kim, H.H., Abdelmohsen, K., Kuwano, Y., Pullmann, R., Srikantan, S., Subrahmanyam, R., Martindale, J.L., Yang, X., Ahmed, F*., et al.* (2008). p16(INK4a) translation suppressed by miR-24. PLoS One *3*, e1864.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M*., et al.* (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. Nat Cell Biol *12*, 618-624.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W*., et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Latres, E., Malumbres, M., Sotillo, R., Martin, J., Ortega, S., Martin-Caballero, J., Flores, J.M., Cordon-Cardo, C., and Barbacid, M. (2000). Limited overlapping roles of P15(INK4b) and P18(INK4c) cell cycle inhibitors in proliferation and tumorigenesis. Embo J *19*, 3496-3506.

Lee, D.Y., Teyssier, C., Strahl, B.D., and Stallcup, M.R. (2005). Role of protein methylation in regulation of transcription. Endocr Rev *26*, 147-170.

Lee, J.-H., Hart, S.R.L., and Skalnik, D.G. (2004). Histone deacetylase activity is required for embryonic stem cell differentiation. Genesis *38*, 32-38.

Lee, J.-H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J Biol Chem *280*, 41725-41731.

Lee, J.-S., and Shilatifard, A. (2007). A site to remember: H3K36 methylation a mark for histone deacetylation. Mutat Res *618*, 130-134.

Lee, J.-S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R., and Shilatifard, A. (2007a). Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell *131*, 1084-1096.

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007b). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science *318*, 447-450.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.-i.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev *24*, 265-276.

Leeb, M., and Wutz, A. (2007). Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. J Cell Biol *178*, 219-229.

Lessard, J., and Sauvageau, G. (2003). Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature *423*, 255-260.

Levine, S.S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P., and Kingston, R.E. (2002). The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. Mol Cell Biol *22*, 6070-6078.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet *3*, 662-673.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362-365.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.
Li, G., Margueron, R., Hu, G., Stokes, D., Wang, Y.-H., and Reinberg, D. (2010a). Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo. Mol Cell *38*, 41-53.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010b). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev *24*, 368-380.

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006a). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature *442*, 91-95.

Li, Z., Cao, R., Wang, M., Myers, M.P., Zhang, Y., and Xu, R.-M. (2006b). Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. J Biol Chem *281*, 20643-20649.

Liu, J., Cao, L., Chen, J., Song, S., Lee, I.H., Quijano, C., Liu, H., Keyvanfar, K., Chen, H., Cao, L.-Y.*, et al.* (2009). Bmi1 regulates mitochondrial function and the DNA damage response pathway. Nature *459*, 387-392.

Lloyd, A.C. (2002). Limits to lifespan. Nat Cell Biol *4*, E25-27.

Lorincz, M.C., Dickerson, D.R., Schmitt, M., and Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol *11*, 1068-1075.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev *8*, 2282-2292.

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianferani, S., Van Dorsselaer, A., Clerc, P., Avner, P.*, et al.* (2010). 2-D structure of the A region of Xist RNA and its implication for PRC2 association. PLoS Biol *8*, e1000276.

Maertens, G.N., El Messaoudi-Aubert, S., Racek, T., Stock, J.K., Nicholls, J., Rodriguez-Niedenfuhr, M., Gil, J., and Peters, G. (2009). Several distinct polycomb complexes regulate and co-localize on the INK4a tumor suppressor locus. PLoS One *4*, e6380.

Maeshima, K., Hihara, S., and Eltsov, M. (2010). Chromatin structure: does the 30-nm fibre exist in vivo? Curr Opin Cell Biol *22*, 291-297.

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., and Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell *131*, 887-900.

Mak, W., Baxter, J., Silva, J., Newall, A.E., Otte, A.P., and Brockdorff, N. (2002). Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells. Curr Biol *12*, 1016-1020.

Mancini-Dinardo, D., Steele, S.J.S., Levorse, J.M., Ingram, R.S., and Tilghman, S.M. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev *20*, 1268-1282.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

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

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

Martens, J.A., and Winston, F. (2002). Evidence that Swi/Snf directly represses transcription in S. cerevisiae. Genes Dev *16*, 2231-2236.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838-849.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A*., et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol *9*, 625-635.

Matheu, A., Klatt, P., and Serrano, M. (2005). Regulation of the INK4a/ARF locus by histone deacetylase inhibitors. J Biol Chem *280*, 42433-42441.

Matheu, A., Pantoja, C., Efeyan, A., Criado, L.M., Martin-Caballero, J., Flores, J.M., Klatt, P., and Serrano, M. (2004). Increased gene dosage of Ink4a/Arf results in cancer resistance and normal aging. Genes Dev *18*, 2736-2746.

Matsui, S.I., Weinfeld, H., and Sandberg, A.A. (1979). Quantitative conservation of chromatin-bound RNA polymerases I and II in mitosis. Implications for chromosome structure. J Cell Biol *80*, 451-464.

Mattick, J.S. (2005). The functional genomics of noncoding RNA. Science *309*, 1527-1528.

Matzke, M.A., and Birchler, J.A. (2005). RNAi-mediated pathways in the nucleus. Nat Rev Genet *6*, 24-35.

McClure, K.D., and Schubiger, G. (2008). A screen for genes that function in leg disc regeneration in Drosophila melanogaster. Mech Dev *125*, 67-80.

Mermoud, J.E., Costanzi, C., Pehrson, J.R., and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J Cell Biol *147*, 1399-1408.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol *7*, 540-546.

Miki, J., Fujimura, Y.-i., Koseki, H., and Kamijo, T. (2007). Polycomb complexes regulate cellular senescence by repression of ARF in cooperation with E2F3. Genes Cells *12*, 1371-1382.

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

Mimnaugh, E.G., Chen, H.Y., Davie, J.R., Celis, J.E., and Neckers, L. (1997). Rapid deubiquitination of nucleosomal histones in human tumor cells caused by proteasome inhibitors and stress response inducers: effects on replication, transcription, translation, and the cellular stress response. Biochemistry *36*, 14418-14429.

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

Minsky, N., Shema, E., Field, Y., Schuster, M., Segal, E., and Oren, M. (2008). Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nat Cell Biol *10*, 483-488.

Mitchell, J.A., and Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. Genes Dev *22*, 20-25.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science *303*, 343-348.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*, 393-396.

Mohd-Sarip, A., Venturini, F., Chalkley, G.E., and Verrijzer, C.P. (2002). Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. Mol Cell Biol *22*, 7473-7483.

Molofsky, A.V., Pardal, R., Iwashita, T., Park, I.-K., Clarke, M.F., and Morrison, S.J. (2003). Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature *425*, 962-967.

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

Morange, M. (2009). What history tells us XVII. Conrad Waddington and the nature of life. J Biosci *34*, 195-198.

Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. Trends Biochem Sci *35*, 323-332.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem *279*, 52353-52360.

Mujtaba, S., Manzur, K.L., Gurnon, J.R., Kang, M., Van Etten, J.L., and Zhou, M.-M. (2008). Epigenetic transcriptional repression of cellular genes by a viral SET protein. Nat Cell Biol *10*, 1114-1122.

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197-208.

Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S., and Hannon, G.J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc Natl Acad Sci U S A *102*, 12135-12140.

Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell *132*, 661-680.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717-1720.

Nakagawa, T., Kajitani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., Koji, T., Matsuyama, T., Ikura, T., Muramatsu, M*., et al.* (2008). Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes Dev *22*, 37-49.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M*., et al.* (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol *9*, 64-71.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Narita, M., Nunez, S., Heard, E., Narita, M., Lin, A.W., Hearn, S.A., Spector, D.L., Hannon, G.J., and Lowe, S.W. (2003). Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. Cell *113*, 703-716.

Negishi, M., Saraya, A., Mochizuki, S., Helin, K., Koseki, H., and Iwama, A. (2010). A Novel Zinc Finger Protein Zfp277 Mediates Transcriptional Repression of the Ink4a/Arf Locus through Polycomb Repressive Complex 1. PLoS One *5*.

Negre, N., Hennetin, J., Sun, L.V., Lavrov, S., Bellis, M., White, K.P., and Cavalli, G. (2006). Chromosomal distribution of PcG proteins during Drosophila development. PLoS Biol *4*, e170.

Nekrasov, M., Wild, B., and Muller, J. (2005). Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Rep *6*, 348-353.

Nesterova, T.B., Popova, B.C., Cobb, B.S., Norton, S., Senner, C.E., Tang, Y.A., Spruce, T., Rodriguez, T.A., Sado, T., Merkenschlager, M*., et al.* (2008). Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. Epigenetics Chromatin *1*, 2.

Ng, J.-H., Heng, J.-C.D., Loh, Y.-H., and Ng, H.-H. (2008). Transcriptional and epigenetic regulations of embryonic stem cells. Mutat Res *647*, 52-58.

Ng, J., Hart, C.M., Morgan, K., and Simon, J.A. (2000). A Drosophila ESC-E(Z) protein complex is distinct from other polycomb group complexes and contains covalently modified ESC. Mol Cell Biol *20*, 3069-3078.

Ng, R.K., and Gurdon, J.B. (2008). Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. Nat Cell Biol *10*, 102-109.

Ng, S.S., Yue, W.W., Oppermann, U., and Klose, R.J. (2009). Dynamic protein methylation in chromatin biology. Cell Mol Life Sci *66*, 407-422.

Nightingale, K.P., Gendreizig, S., White, D.A., Bradbury, C., Hollfelder, F., and Turner, B.M. (2007). Cross-talk between histone modifications in response to histone deacetylase inhibitors: MLL4 links histone H3 acetylation and histone H3K4 methylation. J Biol Chem *282*, 4408-4416.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.

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

Oda, M., Yamagiwa, A., Yamamoto, S., Nakayama, T., Tsumura, A., Sasaki, H., Nakao, K., Li, E., and Okano, M. (2006). DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. Genes Dev *20*, 3382-3394.

Ohno, K., McCabe, D., Czermin, B., Imhof, A., and Pirrotta, V. (2008). ESC, ESCL and their roles in Polycomb Group mechanisms. Mech Dev *125*, 527-541.

Ohta, H., Sawada, A., Kim, J.Y., Tokimasa, S., Nishiguchi, S., Humphries, R.K., Hara, J., and Takihara, Y. (2002). Polycomb group gene rae28 is required for sustaining activity of hematopoietic stem cells. J Exp Med *195*, 759-770.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Osley, M.A. (2006). Regulation of histone H2A and H2B ubiquitylation. Brief Funct Genomic Proteomic *5*, 179-189.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol *10*, 475-478.

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

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell *32*, 232-246.

Park, I.-k., Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., and Clarke, M.F. (2003). Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. Nature *423*, 302-305.

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

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. The EMBO journal *23*, 4061-4071.

Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010a). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306-310.

Pasini, D., Hansen, K.H., Christensen, J., Agger, K., Cloos, P.A.C., and Helin, K. (2008). Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes Dev *22*, 1345-1355.

Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N.*, et al.* (2010b). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Res *38*, 4958-4969.

Pasmant, E., Laurendeau, I., Heron, D., Vidaud, M., Vidaud, D., and Bieche, I. (2007). Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. Cancer Res *67*, 3963-3969.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell *125*, 703-717.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell *139*, 1290-1302.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131-137.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Plath, K., Talbot, D., Hamer, K.M., Otte, A.P., Yang, T.P., Jaenisch, R., and Panning, B. (2004). Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. J Cell Biol *167*, 1025-1035.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol *10*, 192-206.

Puschendorf, M., Terranova, R., Boutsma, E., Mao, X., Isono, K.-i., Brykczynska, U., Kolb, C., Otte, A.P., Koseki, H., Orkin, S.H.*, et al.* (2008). PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. Nat Genet *40*, 411-420.

Quelle, D.E., Zindy, F., Ashmun, R.A., and Sherr, C.J. (1995). Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. Cell *83*, 993-1000.

Quina, A.S., Buschbeck, M., and Di Croce, L. (2006). Chromatin structure and epigenetics. Biochem Pharmacol *72*, 1563-1569.

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

Ramon-Maiques, S., Kuo, A.J., Carney, D., Matthews, A.G.W., Oettinger, M.A., Gozani, O., and Yang, W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. Proc Natl Acad Sci U S A *104*, 18993-18998.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science *293*, 1089-1093.

Ren, X., Vincenz, C., and Kerppola, T.K. (2008). Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. Mol Cell Biol *28*, 2884-2895.

Richards, E.J. (2006). Inherited epigenetic variation--revisiting soft inheritance. Nat Rev Genet *7*, 395-401.

Ringrose, L. (2007). Polycomb comes of age: genome-wide profiling of target sites. Curr Opin Cell Biol *19*, 290-297.

Ringrose, L., Ehret, H., and Paro, R. (2004). Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. Mol Cell *16*, 641-653.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet *38*, 413-443.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223-232.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E.*, et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

Robertson, K.D. (2005). DNA methylation and human disease. Nat Rev Genet *6*, 597-610.
Robzyk, K., Recht, J., and Osley, M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science *287*, 501-504.

Roest, H.P., van Klaveren, J., de Wit, J., van Gurp, C.G., Koken, M.H., Vermey, M., van Roijen, J.H., Hoogerbrugge, J.W., Vreeburg, J.T., Baarends, W.M.*, et al.* (1996). Inactivation of the HR6B ubiquitin-conjugating DNA repair enzyme in mice causes male sterility associated with chromatin modification. Cell *86*, 799-810.

Roh, T.-Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev *19*, 542-552.

Rossant, J. (2008). Stem cells and early lineage development. Cell *132*, 527-531.

Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone acetyltransferases. Annu Rev Biochem *70*, 81-120.

Ruhl, D.D., Jin, J., Cai, Y., Swanson, S., Florens, L., Washburn, M.P., Conaway, R.C., Conaway, J.W., and Chrivia, J.C. (2006). Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes. Biochemistry *45*, 5671-5677.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nat Rev Mol Cell Biol *7*, 437-447.

Sanchez-Elsner, T., Gou, D., Kremmer, E., and Sauer, F. (2006). Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. Science *311*, 1118-1123.

Sanchez, C., Sanchez, I., Demmers, J.A.A., Rodriguez, P., Strouboulis, J., and Vidal, M. (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. Mol Cell Proteomics *6*, 820-834.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol *241*, 172-182.

Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet *9*, 129-140.

Schaniel, C., Ang, Y.-S., Ratnakumar, K., Cormier, C., James, T., Bernstein, E., Lemischka, I.R., and Paddison, P.J. (2009). Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. Stem Cells *27*, 2979-2991.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Muller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature *465*, 243-247.

Schmiedeberg, L., Skene, P., Deaton, A., and Bird, A. (2009). A temporal threshold for formaldehyde crosslinking and fixation. PLoS One *4*, e4636.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H.H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. Embo Journal *25*, 3110-3122.
Schumacher, A., Faust, C., and Magnuson, T. (1996). Positional cloning of a global regulator of anterior-posterior patterning in mice. Nature *384*, 648.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat Genet *38*, 700-705.

Senner, C.E., and Brockdorff, N. (2009). Xist gene regulation at the onset of X inactivation. Curr Opin Genet Dev *19*, 122-126.

Serrano, M., Hannon, G.J., and Beach, D. (1993). A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. Nature *366*, 704-707.

Serrano, M., Lee, H., Chin, L., Cordon-Cardo, C., Beach, D., and DePinho, R.A. (1996). Role of the INK4a locus in tumor suppression and cell mortality. Cell *85*, 27-37.

Sewalt, R.G., van der Vlag, J., Gunster, M.J., Hamer, K.M., den Blaauwen, J.L., Satijn, D.P., Hendrix, T., van Driel, R., and Otte, A.P. (1998). Characterization of interactions between the mammalian polycomb-group proteins Enx1/EZH2 and EED suggests the existence of different mammalian polycomb-group protein complexes. Mol Cell Biol *18*, 3586-3595.

Seward, D.J., Cubberley, G., Kim, S., Schonewald, M., Zhang, L., Tripet, B., and Bentley, D.L. (2007). Demethylation of trimethylated histone H3 Lys4 in vivo by JARID1 JmjC proteins. Nat Struct Mol Biol *14*, 240-242.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell *98*, 37-46.

Sharif, J., Muto, M., Takebayashi, S.-i., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Sharpless, N.E., Bardeesy, N., Lee, K.H., Carrasco, D., Castrillon, D.H., Aguirre, A.J., Wu, E.A., Horner, J.W., and DePinho, R.A. (2001). Loss of p16Ink4a with retention of p19Arf predisposes mice to tumorigenesis. Nature *413*, 86-91.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.-C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303-1314.

Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell *32*, 491-502.

Sherr, C.J. (2001). The INK4a/ARF network in tumour suppression. Nat Rev Mol Cell Biol *2*, 731-737.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shi, Y., and Whetstine, J.R. (2007). Dynamic regulation of histone lysine methylation by demethylases. Mol Cell *25*, 1-14.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H.F.M., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell *4*, 481-495.

Silva, J., and Smith, A. (2008). Capturing pluripotency. Cell *132*, 532-536.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol *10*, 697-708.

Sims, R.J., Nishioka, K., and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. Trends Genet *19*, 629-639.

Sims, R.J., and Reinberg, D. (2006). Histone H3 Lys 4 methylation: caught in a bind? Genes Dev *20*, 2779-2786.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D., and Cordes, S.P. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell *138*, 885-897.

Slack, J.M.W. (2002). Conrad Hal Waddington: the last Renaissance biologist? Nat Rev Genet *3*, 889-895.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810-813.

Smith, A.G. (2001). Embryo-derived stem cells: of mice and men. Annu Rev Cell Dev Biol *17*, 435-462.

Sobell, H.M. (1985). Actinomycin and DNA transcription. Proc Natl Acad Sci U S A *82*, 5328-5331.

Squazzo, S.L., O'Geen, H., Komashko, V.M., Krig, S.R., Jin, V.X., Jang, S.-w., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2006). Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res *16*, 890-900.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A.*, et al.* (2002). The generic genome browser: a building block for a model organism system database. Genome Res *12*, 1599-1610.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev *64*, 435-459.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol *9*, 1428-U1174.

Stoger, R., Kubicka, P., Liu, C.G., Kafri, T., Razin, A., Cedar, H., and Barlow, D.P. (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. Cell *73*, 61-71.

Stopka, T., and Skoultchi, A.I. (2003). The ISWI ATPase Snf2h is required for early mouse development. Proc Natl Acad Sci U S A *100*, 14097-14102.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Sutcliffe, J.S., Nakao, M., Christian, S., Orstavik, K.H., Tommerup, N., Ledbetter, D.H., and Beaudet, A.L. (1994). Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. Nat Genet *8*, 52-58.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465-476.

Swigut, T., and Wysocka, J. (2007). H3K27 demethylases, at long last. Cell *131*, 29-32.

Szutorisz, H., Georgiou, A., Tora, L., and Dillon, N. (2006). The proteasome restricts permissive transcription at tissue-specific gene loci in embryonic stem cells. Cell *127*, 1375-1388.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A *99*, 3740-3745.

Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol *11*, 264-275.

Tang, Y.A., Huntley, D., Montana, G., Cerase, A., Nesterova, T.B., and Brockdorff, N. (2010). Efficiency of Xist-mediated silencing on autosomes is linked to chromosomal domain organisation. Epigenetics Chromatin *3*, 10.

Tautz, D. (2000). Evolution of transcriptional regulation. Curr Opin Genet Dev *10*, 575-579.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol *14*, 1025-1040.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H.F.M. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev Cell *15*, 668-679.

Thoma, F., Koller, T., and Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol *83*, 403-427.

Thomas, M.J., and Seto, E. (1999). Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? Gene *236*, 197-208.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D*., et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082-1086.

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., and Harte, P.J. (2001). The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. Development *128*, 275-286.

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

Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell *28*, 1-13.

Tsukada, Y.-i., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S.-i., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R*., et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells *11*, 805-814.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.

Tzatsos, A., Pfau, R., Kampranis, S.C., and Tsichlis, P.N. (2009). Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus. Proc Natl Acad Sci U S A *106*, 2641-2646.

Ura, H., Usuda, M., Kinoshita, K., Sun, C., Mori, K., Akagi, T., Matsuda, T., Koide, H., and Yokota, T. (2008). STAT3 and Oct-3/4 control histone modification through induction of Eed in embryonic stem cells. Journal of Biological Chemistry *283*, 9713-9723.

Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell *19*, 381-391.

van Daal, A., and Elgin, S.C. (1992). A histone variant, H2AvD, is essential in Drosophila melanogaster. Mol Biol Cell *3*, 593-602.

van der Lugt, N.M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., and van Lohuizen, M. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. Genes Dev *8*, 757-769.

Venolia, L., Gartler, S.M., Wassman, E.R., Yen, P., Mohandas, T., and Shapiro, L.J. (1982). Transformation with DNA from 5-azacytidine-reactivated X chromosomes. Proc Natl Acad Sci U S A *79*, 2352-2354.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Vincenz, C., and Kerppola, T.K. (2008). Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. Proc Natl Acad Sci U S A *105*, 16572-16577.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.-M.*, et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

Voncken, J.W., Roelen, B.A.J., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J., and van Lohuizen, M. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. Proc Natl Acad Sci U S A *100*, 2468-2473.

Walker, E., Chang, W.Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., Krogan, N.J., Reiter, J.F., and Stanford, W.L. (2010). Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. Cell Stem Cell *6*, 153-166.

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

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004b). Hierarchical recruitment of polycomb group silencing complexes. Mol Cell *14*, 637-646.

Wang, R., Taylor, A.B., Leal, B.Z., Chadwell, L.V., Ilangovan, U., Robinson, A.K., Schirf, V., Hart, P.J., Lafer, E.M., Demeler, B.*, et al.* (2010). Polycomb group targeting through different binding partners of RING1B C-terminal domain. Structure *18*, 966-975.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell *138*, 1019-1031.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897-903.

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A., and Copeland, N.G. (2005). Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res *33*, e36.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Weake, V.M., and Workman, J.L. (2008). Histone ubiquitination: triggering gene activity. Mol Cell *29*, 653-663.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Wei, J., Zhai, L., Xu, J., and Wang, H. (2006). Role of Bmi1 in H2A ubiquitylation and Hox gene silencing. J Biol Chem *281*, 22537-22544.

Whitcomb, S.J., Basu, A., Allis, C.D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. Trends Genet *23*, 494-502.

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J., and Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature *453*, 1239-1243.

Wong, M.M., Cox, L.K., and Chrivia, J.C. (2007). The chromatin remodeling protein, SRCAP, is critical for deposition of the histone variant H2A.Z at promoters. J Biol Chem *282*, 26132-26139.

Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J., and Kingston, R.E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. Cell *140*, 99-110.

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science *329*, 444-448.

Wu, R.S., Kohn, K.W., and Bonner, W.M. (1981). Metabolism of ubiquitinated histones. J Biol Chem *256*, 5916-5920.

Xu, Q., Farah, M., Webster, J.M., and Wojcikiewicz, R.J.H. (2004). Bortezomib rapidly suppresses ubiquitin thiolesterification to ubiquitin-conjugating enzymes and inhibits ubiquitination of histones and type I inositol 1,4,5-trisphosphate receptor. Mol Cancer Ther *3*, 1263-1269.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Cote, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol Cell *2*, 851-861.

Yamane, K., Toumazou, C., Tsukada, Y.-i., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell *125*, 483-495.

Yap, K.L., Li, S., Munoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.-M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell *38*, 662-674.

Yu, J., and Thomson, J.A. (2008). Pluripotent stem cell lines. Genes Dev *22*, 1987-1997.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W*., et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.

Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., Suzuki, E., Le Guezennec, X., Stunnenberg, H.G., Krasnov, A*., et al.* (2008). A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. Mol Cell *29*, 92-101.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol Cell *29*, 69-80.

Zhou, Y., Schmitz, K.-M., Mayer, C., Yuan, X., Akhtar, A., and Grummt, I. (2009). Reversible acetylation of the chromatin remodelling complex NoRC is required for non-coding RNA-dependent silencing. Nat Cell Biol *11*, 1010-1016.

Zhu, B., Zheng, Y., Pham, A.-D., Mandal, S.S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2005). Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Mol Cell *20*, 601-611.

Zhu, J.-K. (2009). Active DNA demethylation mediated by DNA glycosylases. Annu Rev Genet *43*, 143-166.

Zhu, P., Zhou, W., Wang, J., Puc, J., Ohgi, K.A., Erdjument-Bromage, H., Tempst, P., Glass, C.K., and Rosenfeld, M.G. (2007). A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. Mol Cell *27*, 609-621.

Zindy, F., Quelle, D.E., Roussel, M.F., and Sherr, C.J. (1997). Expression of the p16INK4a tumor suppressor versus other INK4 family members during mouse development and aging. Oncogene *15*, 203-211.

## Appendix 1

| | WT | | | KO | | | Input | | | %WT/%Input | %KO/%Input |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Name | Matches | Total no. sequences | % Num matches/num sequences | Matches | Total number sequences | % Num matches/num sequences | Matches | Total number sequences | % Num matches/num sequences | | |
| B1 | 6277 | 23088911 | 0.000271862 | 20559 | 73937472 | 0.000278 | 30091 | 68826807 | 0.000437199 | 0.621827151 | 0.636001867 |
| B1F | 0 | 23088911 | 0 | 3 | 73937472 | 4.06E-08 | 442 | 68826807 | 6.42E-06 | 0 | 0.00631818 |
| B2 | 5797 | 23088911 | 0.000251073 | 18113 | 73937472 | 0.000245 | 31707 | 68826807 | 0.000460678 | 0.545007301 | 0.53177542 |
| BGLII | 105 | 23088911 | 4.55E-06 | 350 | 73937472 | 4.73E-06 | 435 | 68826807 | 6.32E-06 | 0.719538795 | 0.748982744 |
| BGLII_C | 5 | 23088911 | 2.17E-07 | 8 | 73937472 | 1.08E-07 | 14 | 68826807 | 2.03E-07 | 1.064623728 | 0.531930602 |
| BGLII_Mus | 254 | 23088911 | 1.10E-05 | 822 | 73937472 | 1.11E-05 | 825 | 68826807 | 1.20E-05 | 0.917770176 | 0.927493541 |
| C573 | 8 | 23088911 | 3.46E-07 | 639 | 73937472 | 8.64E-06 | 209 | 68826807 | 3.04E-06 | 0.114103213 | 2.846083234 |
| CYRA11_MM | 5 | 23088911 | 2.17E-07 | 775 | 73937472 | 1.05E-05 | 200 | 68826807 | 2.91E-06 | 0.074523661 | 3.607154396 |
| ERVB2_1-I_MM | 225 | 23088911 | 9.74E-06 | 842 | 73937472 | 1.14E-05 | 813 | 68826807 | 1.18E-05 | 0.824985177 | 0.964083324 |
| ERVB2_1A-I_MM | 202 | 23088911 | 8.75E-06 | 725 | 73937472 | 9.81E-06 | 705 | 68826807 | 1.02E-05 | 0.85411515 | 0.957286456 |
| ERVB3_1-I_MM | 51 | 23088911 | 2.21E-06 | 178 | 73937472 | 2.41E-06 | 235 | 68826807 | 3.41E-06 | 0.646928801 | 0.70509099 |
| ERVB3_1-LTR_MM | 15 | 23088911 | 6.50E-07 | 71 | 73937472 | 9.60E-07 | 107 | 68826807 | 1.55E-06 | 0.417889688 | 0.617685769 |
| ERVB4_1-I_MM | 182 | 23088911 | 7.88E-06 | 661 | 73937472 | 8.94E-06 | 672 | 68826807 | 9.76E-06 | 0.80733966 | 0.915640958 |
| ERVB4_1-LTR_MM | 32 | 23088911 | 1.39E-06 | 127 | 73937472 | 1.72E-06 | 154 | 68826807 | 2.24E-06 | 0.619417442 | 0.767672574 |
| ERVB4_1B-I_MM | 377 | 23088911 | 1.63E-05 | 1314 | 73937472 | 1.78E-05 | 1539 | 68826807 | 2.24E-05 | 0.730225346 | 0.794785198 |
| ERVB4_1B-LTR_MM | 553 | 23088911 | 2.40E-05 | 1884 | 73937472 | 2.55E-05 | 1693 | 68826807 | 2.46E-05 | 0.973693668 | 1.03589793 |
| ERVB4_2-I_MM | 1388 | 23088911 | 6.01E-05 | 4711 | 73937472 | 6.37E-05 | 4819 | 68826807 | 7.00E-05 | 0.858591753 | 0.910016366 |
| ERVB4_2-LTR_MM | 504 | 23088911 | 2.18E-05 | 1514 | 73937472 | 2.05E-05 | 1534 | 68826807 | 2.23E-05 | 0.979398308 | 0.918741936 |
| ERVB4_2B-LTR_MM | 781 | 23088911 | 3.38E-05 | 2532 | 73937472 | 3.42E-05 | 2549 | 68826807 | 3.70E-05 | 0.913346084 | 0.924670262 |
| ERVB4_3-I_MM | 227 | 23088911 | 9.83E-06 | 753 | 73937472 | 1.02E-05 | 820 | 68826807 | 1.19E-05 | 0.825213221 | 0.854818965 |
| ERVB4_3-LTR_MM | 13 | 23088911 | 5.63E-07 | 45 | 73937472 | 6.09E-07 | 63 | 68826807 | 9.15E-07 | 0.615115932 | 0.664913253 |
| ERVB5_1-LTR_MM | 4 | 23088911 | 1.73E-07 | 11 | 73937472 | 1.49E-07 | 17 | 68826807 | 2.47E-07 | 0.701399162 | 0.602333182 |
| ERVB5_2-I_MM | 292 | 23088911 | 1.26E-05 | 909 | 73937472 | 1.23E-05 | 1009 | 68826807 | 1.47E-05 | 0.862672309 | 0.838621016 |
| ERVB5_2-LTR_MM | 59 | 23088911 | 2.56E-06 | 255 | 73937472 | 3.45E-06 | 228 | 68826807 | 3.31E-06 | 0.771385262 | 1.041114172 |
| ERVB7_1-I_MM | 2382 | 23088911 | 0.000103166 | 7781 | 73937472 | 0.000105 | 7990 | 68826807 | 0.000116088 | 0.888687661 | 0.906528914 |
| ERVB7_1-LTR_MM | 480 | 23088911 | 2.08E-05 | 1494 | 73937472 | 2.02E-05 | 1595 | 68826807 | 2.32E-05 | 0.897087329 | 0.871932639 |
| ERVB7_2-LTR_MM | 90 | 23088911 | 3.90E-06 | 261 | 73937472 | 3.53E-06 | 259 | 68826807 | 3.76E-06 | 1.035850114 | 0.938066805 |
| ERVB7_2B-LTR_MM | 63 | 23088911 | 2.73E-06 | 222 | 73937472 | 3.00E-06 | 236 | 68826807 | 3.43E-06 | 0.795761125 | 0.875656945 |
| ERVB7_3-LTR_MM | 79 | 23088911 | 3.42E-06 | 265 | 73937472 | 3.58E-06 | 293 | 68826807 | 4.26E-06 | 0.803736412 | 0.841920876 |
| ETNERV | 327 | 23088911 | 1.42E-05 | 1138 | 73937472 | 1.54E-05 | 1118 | 68826807 | 1.62E-05 | 0.871886838 | 0.947531122 |

| | WT | | | KO | | | Input | | | %WT/%Input | %KO/%Input |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GSAT_MM | 222170 | 23088911 | 0.009622368 | 973928 | 73937472 | 0.013172 | 987759 | 68826807 | 0.01435137 | 0.670484268 | 0.917844017 |
| IAP1-MM_I | 1471 | 23088911 | 6.37E-05 | 4842 | 73937472 | 6.55E-05 | 4124 | 68826807 | 5.99E-05 | 1.063281331 | 1.092947128 |
| IAP1-MM_LTR | 392 | 23088911 | 1.70E-05 | 1166 | 73937472 | 1.58E-05 | 1111 | 68826807 | 1.61E-05 | 1.051783082 | 0.97696165 |
| IAPA_MM | 507 | 23088911 | 2.20E-05 | 1672 | 73937472 | 2.26E-05 | 1851 | 68826807 | 2.69E-05 | 0.816499105 | 0.840858423 |
| IAPEY2_LTR | 1635 | 23088911 | 7.08E-05 | 5059 | 73937472 | 6.84E-05 | 4066 | 68826807 | 5.91E-05 | 1.198683577 | 1.158218053 |
| IAPEY3C_LTR | 1396 | 23088911 | 6.05E-05 | 4671 | 73937472 | 6.32E-05 | 3743 | 68826807 | 5.44E-05 | 1.111782321 | 1.161670779 |
| IAPEY3_I | 2513 | 23088911 | 0.00010884 | 9337 | 73937472 | 0.000126 | 8186 | 68826807 | 0.000118936 | 0.915113413 | 1.061765582 |
| IAPEY3_LTR | 598 | 23088911 | 2.59E-05 | 1744 | 73937472 | 2.36E-05 | 1411 | 68826807 | 2.05E-05 | 1.263363551 | 1.150568531 |
| IAPEY4_I | 2201 | 23088911 | 9.53E-05 | 7255 | 73937472 | 9.81E-05 | 6880 | 68826807 | 1.00E-04 | 0.953642894 | 0.981616847 |
| IAPEY4_LTR | 174 | 23088911 | 7.54E-06 | 532 | 73937472 | 7.20E-06 | 713 | 68826807 | 1.04E-05 | 0.727467995 | 0.69456857 |
| IAPEY5_I | 3725 | 23088911 | 0.000161333 | 12085 | 73937472 | 0.000163 | 9630 | 68826807 | 0.000139916 | 1.15306599 | 1.168189753 |
| IAPEY5_LTR | 283 | 23088911 | 1.23E-05 | 962 | 73937472 | 1.30E-05 | 840 | 68826807 | 1.22E-05 | 1.00429505 | 1.066077582 |
| IAPEYI | 1311 | 23088911 | 5.68E-05 | 15908 | 73937472 | 0.000215 | 6443 | 68826807 | 9.36E-05 | 0.606552969 | 2.298372813 |
| IAPEZI | 15704 | 23088911 | 0.000680153 | 51154 | 73937472 | 0.000692 | 48773 | 68826807 | 0.000708634 | 0.959809379 | 0.976322177 |
| IAPLTR1a_I_MM | 18093 | 23088911 | 0.000783623 | 59347 | 73937472 | 0.000803 | 57481 | 68826807 | 0.000835154 | 0.938297244 | 0.961097572 |
| IAPLTR1a_MM | 5770 | 23088911 | 0.000249904 | 17535 | 73937472 | 0.000237 | 13761 | 68826807 | 0.000199937 | 1.249913593 | 1.186175092 |
| IAPLTR3 | 265 | 23088911 | 1.15E-05 | 858 | 73937472 | 1.16E-05 | 800 | 68826807 | 1.16E-05 | 0.987438508 | 0.998367249 |
| IAPLTR3_I | 1241 | 23088911 | 5.37E-05 | 5364 | 73937472 | 7.25E-05 | 3927 | 68826807 | 5.71E-05 | 0.942030692 | 1.271513257 |
| IAPLTR4 | 320 | 23088911 | 1.39E-05 | 991 | 73937472 | 1.34E-05 | 895 | 68826807 | 1.30E-05 | 1.065813252 | 1.03072698 |
| IAPLTR4_I | 2124 | 23088911 | 9.20E-05 | 7052 | 73937472 | 9.54E-05 | 7313 | 68826807 | 0.000106252 | 0.865791089 | 0.897655622 |
| L1B_MM | 68621 | 23088911 | 0.002972033 | 235902 | 73937472 | 0.003191 | 263115 | 68826807 | 0.003822856 | 0.777437719 | 0.834601268 |
| L1MM_F | 464 | 23088911 | 2.01E-05 | 1467 | 73937472 | 1.98E-05 | 2234 | 68826807 | 3.25E-05 | 0.619140173 | 0.611279695 |
| L1VL1_5 | 652 | 23088911 | 2.82E-05 | 1962 | 73937472 | 2.65E-05 | 2383 | 68826807 | 3.46E-05 | 0.815600956 | 0.76642204 |
| L1VL2_5 | 167 | 23088911 | 7.23E-06 | 531 | 73937472 | 7.18E-06 | 734 | 68826807 | 1.07E-05 | 0.678226233 | 0.67342849 |
| L1_MM | 306622 | 23088911 | 0.013280055 | 1038253 | 73937472 | 0.014042 | 1105291 | 68826807 | 0.016059019 | 0.826953046 | 0.874419 |
| L32 | 36 | 23088911 | 1.56E-06 | 74 | 73937472 | 1.00E-06 | 98 | 68826807 | 1.42E-06 | 1.095041549 | 0.702908296 |
| LTRIAPEY | 345 | 23088911 | 1.49E-05 | 2294 | 73937472 | 3.10E-05 | 1190 | 68826807 | 1.73E-05 | 0.864223967 | 1.794483531 |
| LTRIS | 839 | 23088911 | 3.63E-05 | 2809 | 73937472 | 3.80E-05 | 2390 | 68826807 | 3.47E-05 | 1.046449398 | 1.094074417 |
| LTRIS2 | 738 | 23088911 | 3.20E-05 | 2473 | 73937472 | 3.34E-05 | 2777 | 68826807 | 4.03E-05 | 0.792199666 | 0.828974672 |
| LTRIS3 | 584 | 23088911 | 2.53E-05 | 1932 | 73937472 | 2.61E-05 | 1995 | 68826807 | 2.90E-05 | 0.872617905 | 0.901482389 |
| LTRIS4 | 18 | 23088911 | 7.80E-07 | 67 | 73937472 | 9.06E-07 | 75 | 68826807 | 1.09E-06 | 0.715427145 | 0.831584841 |
| LTRIS4A | 83 | 23088911 | 3.59E-06 | 295 | 73937472 | 3.99E-06 | 298 | 68826807 | 4.33E-06 | 0.830263605 | 0.921507293 |
| LTRIS4B | 67 | 23088911 | 2.90E-06 | 171 | 73937472 | 2.31E-06 | 173 | 68826807 | 2.51E-06 | 1.154470586 | 0.920116952 |
| LTRIS5 | 41 | 23088911 | 1.78E-06 | 123 | 73937472 | 1.66E-06 | 108 | 68826807 | 1.57E-06 | 1.131655592 | 1.060167242 |

| | WT | | | KO | | | Input | | | %WT/%Input | %KO/%Input |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LTRIS6 | 82 | 23088911 | 3.55E-06 | 251 | 73937472 | 3.39E-06 | 241 | 68826807 | 3.50E-06 | 1.014263933 | 0.96950422 |
| LTRIS_Mus | 1003 | 23088911 | 4.34E-05 | 3547 | 73937472 | 4.80E-05 | 3576 | 68826807 | 5.20E-05 | 0.836098791 | 0.923329483 |
| LTRVL30 | 13 | 23088911 | 5.63E-07 | 43 | 73937472 | 5.82E-07 | 32 | 68826807 | 4.65E-07 | 1.21100949 | 1.250868057 |
| LX10 | 0 | 23088911 | 0.00E+00 | 1 | 73937472 | 1.35E-08 | 17 | 68826807 | 2.47E-07 | 0 | 0.054757562 |
| LX2B | 1108 | 23088911 | 4.80E-05 | 3508 | 73937472 | 4.74E-05 | 4372 | 68826807 | 6.35E-05 | 0.75546401 | 0.746917193 |
| LX4 | 59 | 23088911 | 2.56E-06 | 206 | 73937472 | 2.79E-06 | 269 | 68826807 | 3.91E-06 | 0.653813531 | 0.712866104 |
| LX5 | 22 | 23088911 | 9.53E-07 | 51 | 73937472 | 6.90E-07 | 144 | 68826807 | 2.09E-06 | 0.455422372 | 0.329686154 |
| LX5b | 191 | 23088911 | 8.27E-06 | 609 | 73937472 | 8.24E-06 | 825 | 68826807 | 1.20E-05 | 0.690134266 | 0.687157623 |
| LX5c | 32 | 23088911 | 1.39E-06 | 88 | 73937472 | 1.19E-06 | 203 | 68826807 | 2.95E-06 | 0.469902887 | 0.40353356 |
| LX6 | 2 | 23088911 | 8.66E-08 | 6 | 73937472 | 8.11E-08 | 27 | 68826807 | 3.92E-07 | 0.220810847 | 0.206861901 |
| LX7 | 2 | 23088911 | 8.66E-08 | 6 | 73937472 | 8.11E-08 | 25 | 68826807 | 3.63E-07 | 0.238475715 | 0.223410853 |
| LX8 | | 23088911 | 0.00E+00 | | 73937472 | 2.70E-08 | 25 | 68826807 | 3.63E-07 | 0 | 0.074470284 |
| LX8B | 2 | 23088911 | 8.66E-08 | 2 | 73937472 | 1.22E-07 | 21 | 68826807 | 3.05E-07 | 0.283899661 | 0.398947952 |
| LX9 | 15 | 23088911 | 6.50E-07 | 9 | 73937472 | 8.11E-07 | 117 | 68826807 | 1.70E-06 | 0.38217262 | 0.477373617 |
| MERV1_I | 255 | 23088911 | 1.10E-05 | 60 | 73937472 | 1.14E-05 | 878 | 68826807 | 1.28E-05 | 0.865764626 | 0.893770639 |
| MERV1_LTR | 113 | 23088911 | 4.89E-06 | 843 | 73937472 | 5.36E-06 | 398 | 68826807 | 5.78E-06 | 0.846349114 | 0.926200772 |
| MERVL | 39136 | 23088911 | 0.001695013 | 396 | 73937472 | 0.001707 | 109632 | 68826807 | 0.001592868 | 1.064126531 | 1.071624348 |
| MERVL_LTR | 5845 | 23088911 | 0.000253152 | 126208 | 73937472 | 0.000257 | 19495 | 68826807 | 0.000283247 | 0.893748752 | 0.907576746 |
| MLTR18A_MM | 4 | 23088911 | 1.73E-07 | 19007 | 73937472 | 9.47E-08 | 9 | 68826807 | 1.31E-07 | 1.324865084 | 0.724016653 |
| MLTR18B_MM | 0 | 23088911 | 0.00E+00 | 7 | 73937472 | 1.35E-08 | 5 | 68826807 | 7.26E-08 | 0 | 0.186175711 |
| MLTR18C_MM | 2 | 23088911 | 8.66E-08 | 1 | 73937472 | 9.47E-08 | 26 | 68826807 | 3.78E-07 | 0.229303572 | 0.250621149 |
| MLTR18D_MM | 1 | 23088911 | 4.33E-08 | 7 | 73937472 | 2.16E-07 | 29 | 68826807 | 4.21E-07 | 0.102791256 | 0.513588168 |
| MLTR18_MM | 0 | 23088911 | 0.00E+00 | 16 | 73937472 | 1.35E-08 | 3 | 68826807 | 4.36E-08 | 0 | 0.310292851 |
| MLTR25A | 33 | 23088911 | 1.43E-06 | 1 | 73937472 | 1.58E-06 | 187 | 68826807 | 2.72E-06 | 0.526049371 | 0.582421341 |
| MLTR25B | 24 | 23088911 | 1.04E-06 | 117 | 73937472 | 1.20E-06 | 165 | 68826807 | 2.40E-06 | 0.433592209 | 0.50211025 |
| MLTR25C | 31 | 23088911 | 1.34E-06 | 89 | 73937472 | 1.23E-06 | 110 | 68826807 | 1.60E-06 | 0.840084905 | 0.77009044 |
| MLTR31A_MM | 23 | 23088911 | 9.96E-07 | 91 | 73937472 | 9.74E-07 | 73 | 68826807 | 1.06E-06 | 0.939202302 | 0.918126793 |
| MLTR31B_MM | 45 | 23088911 | 1.95E-06 | 72 | 73937472 | 2.66E-06 | 174 | 68826807 | 2.53E-06 | 0.770934424 | 1.053925719 |
| MLTR31C_MM | 5 | 23088911 | 2.17E-07 | 22 | 73937472 | 2.98E-07 | 14 | 68826807 | 2.03E-07 | 1.064623728 | 1.462809156 |
| MLTR31D_MM | 41 | 23088911 | 1.78E-06 | 109 | 73937472 | 1.47E-06 | 139 | 68826807 | 2.02E-06 | 0.879271971 | 0.729969513 |
| MLTR31E_MM | 17 | 23088911 | 7.36E-07 | 62 | 73937472 | 8.39E-07 | 80 | 68826807 | 1.16E-06 | 0.633451118 | 0.721430879 |
| MLTR31FA_MM | 25 | 23088911 | 1.08E-06 | 107 | 73937472 | 1.45E-06 | 128 | 68826807 | 1.86E-06 | 0.582216101 | 0.778156291 |
| MLTR31F_MM | 25 | 23088911 | 1.08E-06 | 116 | 73937472 | 1.57E-06 | 132 | 68826807 | 1.92E-06 | 0.564573189 | 0.81804479 |
| MLTR32C_MM | 0 | 23088911 | 0.00E+00 | 2 | 73937472 | 2.70E-08 | 4 | 68826807 | 5.81E-08 | 0 | 0.465439277 |

| | WT | | | KO | | | Input | | | %WT/%Input | %KO/%Input |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MMERVK10C | 8718 | 23088911 | 0.000377584 | 28579 | 73937472 | 0.000387 | 26071 | 68826807 | 0.000378791 | 0.996812207 | 1.020427992 |
| MMERVK10D3_I | 796 | 23088911 | 3.45E-05 | 2678 | 73937472 | 3.62E-05 | 2515 | 68826807 | 3.65E-05 | 0.943472511 | 0.991209848 |
| MMERVK10D3_LTR | 309 | 23088911 | 1.34E-05 | 848 | 73937472 | 1.15E-05 | 798 | 68826807 | 1.16E-05 | 1.154276252 | 0.989204278 |
| MMERVK9C_I | 484 | 23088911 | 2.10E-05 | 1495 | 73937472 | 2.02E-05 | 1917 | 68826807 | 2.79E-05 | 0.752622888 | 0.725959018 |
| MMERVK9E_I | 928 | 23088911 | 4.02E-05 | 3278 | 73937472 | 4.43E-05 | 3649 | 68826807 | 5.30E-05 | 0.758103123 | 0.836234557 |
| MMETN | 1750 | 23088911 | 7.58E-05 | 6589 | 73937472 | 8.91E-05 | 7613 | 68826807 | 0.000110611 | 0.685230036 | 0.805669091 |
| MMVL30 | 1643 | 23088911 | 7.12E-05 | 5446 | 73937472 | 7.37E-05 | 5616 | 68826807 | 8.16E-05 | 0.872096688 | 0.90270025 |
| MURVY_LTR | 271 | 23088911 | 1.17E-05 | 907 | 73937472 | 1.23E-05 | 964 | 68826807 | 1.40E-05 | 0.838004652 | 0.875836979 |
| MYSERV | 5 | 23088911 | 2.17E-07 | 18 | 73937472 | 2.43E-07 | 21 | 68826807 | 3.05E-07 | 0.709749152 | 0.797895903 |
| MYSERV16_I | 5 | 23088911 | 2.17E-07 | 33 | 73937472 | 4.46E-07 | 46 | 68826807 | 6.68E-07 | 0.324015917 | 0.66780418 |
| ORR1A0 | 8775 | 23088911 | 0.000380053 | 28105 | 73937472 | 0.00038 | 26798 | 68826807 | 0.000389354 | 0.976110344 | 0.976279638 |
| ORR1B1 | 617 | 23088911 | 2.67E-05 | 2092 | 73937472 | 2.83E-05 | 2380 | 68826807 | 3.46E-05 | 0.772791577 | 0.818234426 |
| ORR1B2 | 120 | 23088911 | 5.20E-06 | 425 | 73937472 | 5.75E-06 | 558 | 68826807 | 8.11E-06 | 0.64106375 | 0.709002483 |
| ORR1C1 | 12 | 23088911 | 5.20E-07 | 74 | 73937472 | 1.00E-06 | 185 | 68826807 | 2.69E-06 | 0.193358688 | 0.372351421 |
| ORR1C2 | 13 | 23088911 | 5.63E-07 | 37 | 73937472 | 5.00E-07 | 71 | 68826807 | 1.03E-06 | 0.545807094 | 0.485105725 |
| PMR66 | 34 | 23088911 | 1.47E-06 | 7 | 73937472 | 9.47E-08 | 101 | 68826807 | 1.47E-06 | 1.00348692 | 0.064516335 |
| PMR89 | 24 | 23088911 | 1.04E-06 | 40 | 73937472 | 5.41E-07 | 52 | 68826807 | 7.56E-07 | 1.375821433 | 0.716060426 |
| RLTR10B2 | 1726 | 23088911 | 7.48E-05 | 5272 | 73937472 | 7.13E-05 | 5059 | 68826807 | 7.35E-05 | 1.017021852 | 0.970071503 |
| RLTR10D2 | 251 | 23088911 | 1.09E-05 | 737 | 73937472 | 9.97E-06 | 649 | 68826807 | 9.43E-06 | 1.15287759 | 1.057099375 |
| RLTR10E | 87 | 23088911 | 3.77E-06 | 343 | 73937472 | 4.64E-06 | 315 | 68826807 | 4.58E-06 | 0.823309016 | 1.013623314 |
| RLTR10F | 717 | 23088911 | 3.11E-05 | 2095 | 73937472 | 2.83E-05 | 1729 | 68826807 | 2.51E-05 | 1.236170385 | 1.127929769 |
| RLTR11A | 21 | 23088911 | 9.10E-07 | 58 | 73937472 | 7.84E-07 | 89 | 68826807 | 1.29E-06 | 0.703369384 | 0.606639956 |
| RLTR11B | 16 | 23088911 | 6.93E-07 | 67 | 73937472 | 9.06E-07 | 70 | 68826807 | 1.02E-06 | 0.681359186 | 0.890983759 |
| RLTR11C_MM | 2 | 23088911 | 8.66E-08 | 4 | 73937472 | 5.41E-08 | 18 | 68826807 | 2.62E-07 | 0.331216271 | 0.206861901 |
| RLTR12A | 21 | 23088911 | 9.10E-07 | 63 | 73937472 | 8.52E-07 | 76 | 68826807 | 1.10E-06 | 0.823682568 | 0.771649327 |
| RLTR12B | 5 | 23088911 | 2.17E-07 | 20 | 73937472 | 2.70E-07 | 19 | 68826807 | 2.76E-07 | 0.784459589 | 0.979872162 |
| RLTR12B2 | 0 | 23088911 | 0.00E+00 | 3 | 73937472 | 4.06E-08 | 3 | 68826807 | 4.36E-08 | 0 | 0.930878554 |
| RLTR12D | 0 | 23088911 | 0.00E+00 | 4 | 73937472 | 5.41E-08 | 4 | 68826807 | 5.81E-08 | 0 | 0.930878554 |
| RLTR12E | 12 | 23088911 | 5.20E-07 | 15 | 73937472 | 2.03E-07 | 22 | 68826807 | 3.20E-07 | 1.625970784 | 0.634689923 |
| RLTR12F | 0 | 23088911 | 0.00E+00 | 3 | 73937472 | 4.06E-08 | 2 | 68826807 | 2.91E-08 | 0 | 1.396317831 |
| RLTR12G | 5 | 23088911 | 2.17E-07 | 15 | 73937472 | 2.03E-07 | 17 | 68826807 | 2.47E-07 | 0.876748952 | 0.82136343 |
| RLTR12H | 2 | 23088911 | 8.66E-08 | 5 | 73937472 | 6.76E-08 | 14 | 68826807 | 2.03E-07 | 0.425849491 | 0.332456626 |
| RLTR13A | 187 | 23088911 | 8.10E-06 | 598 | 73937472 | 8.09E-06 | 613 | 68826807 | 8.91E-06 | 0.909358864 | 0.908100123 |
| RLTR13A1 | 153 | 23088911 | 6.63E-06 | 585 | 73937472 | 7.91E-06 | 573 | 68826807 | 8.33E-06 | 0.79595952 | 0.950373393 |

| | WT | | | KO | | | Input | | | %WT/%Input | %KO/%Input |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RLTR13A2 | 231 | 23088911 | 1.00E-05 | 867 | 73937472 | 1.17E-05 | 811 | 68826807 | 1.18E-05 | 0.849073523 | 0.995156234 |
| RLTR13A3 | 266 | 23088911 | 1.15E-05 | 825 | 73937472 | 1.12E-05 | 848 | 68826807 | 1.23E-05 | 0.935061029 | 0.905630668 |
| RLTR13B | 55 | 23088911 | 2.38E-06 | 154 | 73937472 | 2.08E-06 | 165 | 68826807 | 2.40E-06 | 0.993648813 | 0.868819983 |
| RLTR13B1 | 753 | 23088911 | 3.26E-05 | 2992 | 73937472 | 4.05E-05 | 3146 | 68826807 | 4.57E-05 | 0.713494173 | 0.885311072 |
| RLTR13B2 | 829 | 23088911 | 3.59E-05 | 2943 | 73937472 | 3.98E-05 | 3216 | 68826807 | 4.67E-05 | 0.76840939 | 0.851858079 |
| RLTR13B3 | 300 | 23088911 | 1.30E-05 | 1073 | 73937472 | 1.45E-05 | 1148 | 68826807 | 1.67E-05 | 0.778992972 | 0.870063317 |
| RLTR13B4 | 353 | 23088911 | 1.53E-05 | 1212 | 73937472 | 1.64E-05 | 1241 | 68826807 | 1.80E-05 | 0.847924329 | 0.90912555 |
| RLTR13C1 | 130 | 23088911 | 5.63E-06 | 486 | 73937472 | 6.57E-06 | 522 | 68826807 | 7.58E-06 | 0.742381297 | 0.866680033 |
| RLTR13C2 | 852 | 23088911 | 3.69E-05 | 2763 | 73937472 | 3.74E-05 | 2743 | 68826807 | 3.99E-05 | 0.925908263 | 0.937665856 |
| RLTR13C3 | 113 | 23088911 | 4.89E-06 | 344 | 73937472 | 4.65E-06 | 404 | 68826807 | 5.87E-06 | 0.833779573 | 0.792629264 |
| RLTR13D | 82 | 23088911 | 3.55E-06 | 291 | 73937472 | 3.94E-06 | 257 | 68826807 | 3.73E-06 | 0.951119097 | 1.054029802 |
| RLTR13D1 | 245 | 23088911 | 1.06E-05 | 974 | 73937472 | 1.32E-05 | 995 | 68826807 | 1.45E-05 | 0.734001887 | 0.911231871 |
| RLTR13D2 | 96 | 23088911 | 4.16E-06 | 332 | 73937472 | 4.49E-06 | 363 | 68826807 | 5.27E-06 | 0.788349471 | 0.851382038 |
| RLTR13D3 | 79 | 23088911 | 3.42E-06 | 269 | 73937472 | 3.64E-06 | 269 | 68826807 | 3.91E-06 | 0.875445236 | 0.930878554 |
| RLTR13D3A | 83 | 23088911 | 3.59E-06 | 304 | 73937472 | 4.11E-06 | 267 | 68826807 | 3.88E-06 | 0.926661252 | 1.059876705 |
| RLTR13D3A1 | 183 | 23088911 | 7.93E-06 | 664 | 73937472 | 8.98E-06 | 752 | 68826807 | 1.09E-05 | 0.725416487 | 0.821945957 |
| RLTR13D4 | 78 | 23088911 | 3.38E-06 | 348 | 73937472 | 4.71E-06 | 320 | 68826807 | 4.65E-06 | 0.726605694 | 1.012330427 |
| RLTR13D5 | 415 | 23088911 | 1.80E-05 | 1519 | 73937472 | 2.05E-05 | 1527 | 68826807 | 2.22E-05 | 0.810145888 | 0.926001652 |
| RLTR13D6 | 154 | 23088911 | 6.67E-06 | 559 | 73937472 | 7.56E-06 | 703 | 68826807 | 1.02E-05 | 0.653009604 | 0.740200728 |
| RLTR13E | 319 | 23088911 | 1.38E-05 | 1083 | 73937472 | 1.46E-05 | 1106 | 68826807 | 1.61E-05 | 0.859784732 | 0.91152032 |
| RLTR13F | 160 | 23088911 | 6.93E-06 | 652 | 73937472 | 8.82E-06 | 666 | 68826807 | 9.68E-06 | 0.716143288 | 0.911310536 |
| RLTR13G | 108 | 23088911 | 4.68E-06 | 416 | 73937472 | 5.63E-06 | 553 | 68826807 | 8.03E-06 | 0.582173988 | 0.700263071 |
| RLTR14 | 33 | 23088911 | 1.43E-06 | 84 | 73937472 | 1.14E-06 | 122 | 68826807 | 1.77E-06 | 0.806321577 | 0.640932775 |
| RLTR16 | 5 | 23088911 | 2.17E-07 | 25 | 73937472 | 3.38E-07 | 45 | 68826807 | 6.54E-07 | 0.331216271 | 0.517154752 |
| RLTR16B_MM | 61 | 23088911 | 2.64E-06 | 180 | 73937472 | 2.43E-06 | 227 | 68826807 | 3.30E-06 | 0.801047281 | 0.738141584 |
| RLTR16C_MM | 0 | 23088911 | 0.00E+00 | 1 | 73937472 | 1.35E-08 | 5 | 68826807 | 7.26E-08 | 0 | 0.186175711 |
| RLTR17B_Mm | 6 | 23088911 | 0.00E+00 | 24 | 73937472 | 3.25E-07 | 37 | 68826807 | 5.38E-07 | 0 | 0.603813116 |
| RLTR17C_Mm | 11 | 23088911 | 0.00E+00 | 35 | 73937472 | 4.73E-07 | 40 | 68826807 | 5.81E-07 | 0 | 0.814518734 |
| RLTR17D_Mm | 6 | 23088911 | 0.00E+00 | 32 | 73937472 | 4.33E-07 | 24 | 68826807 | 3.49E-07 | 0 | 1.241171405 |
| RLTR17_MM | 248 | 23088911 | 0.00E+00 | 973 | 73937472 | 1.32E-05 | 1216 | 68826807 | 1.77E-05 | 0 | 0.744855948 |
| RLTR18_MM | 0 | 23088911 | 0.00E+00 | 9 | 73937472 | 1.22E-07 | 5 | 68826807 | 7.26E-08 | 0 | 1.675581397 |
| RLTR19A2 | 0 | 23088911 | 0.00E+00 | 8 | 73937472 | 1.08E-07 | 12 | 68826807 | 1.74E-07 | 0 | 0.620585702 |
| RLTR19A_MM | 9 | 23088911 | 0.00E+00 | 27 | 73937472 | 3.65E-07 | 34 | 68826807 | 4.94E-07 | 0 | 0.739227087 |
| RLTR19B_MM | 29 | 23088911 | 0.00E+00 | 106 | 73937472 | 1.43E-06 | 92 | 68826807 | 1.34E-06 | 0 | 1.072533986 |

| | WT | | | KO | | | Input | | | %WT/%Input | %KO/%Input |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RLTR19C_MM | 5 | 23088911 | 0.00E+00 | 33 | 73937472 | 4.46E-07 | 34 | 68826807 | 4.94E-07 | 0 | 0.903499773 |
| RLTR19D | 1 | 23088911 | 0.00E+00 | 3 | 73937472 | 4.06E-08 | 2 | 68826807 | 2.91E-08 | 0 | 1.396317831 |
| RLTR1A2_MM | 99 | 23088911 | 0.00E+00 | 232 | 73937472 | 3.14E-06 | 307 | 68826807 | 4.46E-06 | 0 | 0.703465226 |
| RLTR1B_MM | 1003 | 23088911 | 0.00E+00 | 3330 | 73937472 | 4.50E-05 | 3397 | 68826807 | 4.94E-05 | 0 | 0.91251857 |
| RLTR1C_MM | 901 | 23088911 | 0.00E+00 | 2817 | 73937472 | 3.81E-05 | 2535 | 68826807 | 3.68E-05 | 0 | 1.034431908 |
| RLTR1D2_MM | 114 | 23088911 | 0.00E+00 | 373 | 73937472 | 5.04E-06 | 334 | 68826807 | 4.85E-06 | 0 | 1.039573954 |
| RLTR1D_MM | 93 | 23088911 | 0.00E+00 | 384 | 73937472 | 5.19E-06 | 332 | 68826807 | 4.82E-06 | 0 | 1.076678809 |
| RLTR1E_MM | 11 | 23088911 | 0.00E+00 | 49 | 73937472 | 6.63E-07 | 49 | 68826807 | 7.12E-07 | 0 | 0.930878554 |
| RLTR1F_Mm | 73 | 23088911 | 0.00E+00 | 230 | 73937472 | 3.11E-06 | 213 | 68826807 | 3.09E-06 | 0 | 1.005174025 |
| RLTR1IAP_MM | 2636 | 23088911 | 0.00E+00 | 8303 | 73937472 | 0.000112 | 6853 | 68826807 | 9.96E-05 | 0 | 1.127839578 |
| RLTR1_MM | 9 | 23088911 | 0.00E+00 | 33 | 73937472 | 4.46E-07 | 42 | 68826807 | 6.10E-07 | 0 | 0.731404578 |
| RLTR20A1_MM | 7 | 23088911 | 0.00E+00 | 31 | 73937472 | 4.19E-07 | 31 | 68826807 | 4.50E-07 | 0 | 0.930878554 |
| RLTR20A2B_MM | 3 | 23088911 | 0.00E+00 | 21 | 73937472 | 2.84E-07 | 22 | 68826807 | 3.20E-07 | 0 | 0.888565892 |
| RLTR20A2_MM | 11 | 23088911 | 0.00E+00 | 26 | 73937472 | 3.52E-07 | 37 | 68826807 | 5.38E-07 | 0 | 0.654130876 |
| RLTR20A3_MM | 17 | 23088911 | 0.00E+00 | 43 | 73937472 | 5.82E-07 | 458 | 68826807 | 6.65E-06 | 0 | 0.087396895 |
| RLTR20B1_MM | 7 | 23088911 | 0.00E+00 | 14 | 73937472 | 1.89E-07 | 16 | 68826807 | 2.32E-07 | 0 | 0.814518734 |
| RLTR20B2_MM | 10 | 23088911 | 0.00E+00 | 31 | 73937472 | 4.19E-07 | 60 | 68826807 | 8.72E-07 | 0 | 0.480953919 |
| RLTR20B3A_MM | 2 | 23088911 | 0.00E+00 | 6 | 73937472 | 8.11E-08 | 33 | 68826807 | 4.79E-07 | 0 | 0.169250646 |
| RLTR20B4_MM | 4 | 23088911 | 0.00E+00 | 17 | 73937472 | 2.30E-07 | 19 | 68826807 | 2.76E-07 | 0 | 0.832891338 |
| RLTR20B5_MM | 14 | 23088911 | 0.00E+00 | 42 | 73937472 | 5.68E-07 | 86 | 68826807 | 1.25E-06 | 0 | 0.454615108 |
| RLTR20C1_MM | 575 | 23088911 | 0.00E+00 | 1695 | 73937472 | 2.29E-05 | 1533 | 68826807 | 2.23E-05 | 0 | 1.029249281 |
| RLTR20C2_MM | 28 | 23088911 | 0.00E+00 | 85 | 73937472 | 1.15E-06 | 94 | 68826807 | 1.37E-06 | 0 | 0.841751884 |
| RLTR20C_MM | 6 | 23088911 | 0.00E+00 | 18 | 73937472 | 2.43E-07 | 28 | 68826807 | 4.07E-07 | 0 | 0.598421927 |
| RLTR20D_MM | 11 | 23088911 | 0.00E+00 | 47 | 73937472 | 6.36E-07 | 64 | 68826807 | 9.30E-07 | 0 | 0.683613938 |
| RLTR21_MM | 0 | 23088911 | 0.00E+00 | 3 | 73937472 | 4.06E-08 | 4 | 68826807 | 5.81E-08 | 0 | 0.698158915 |
| RLTR23_MM | 3671 | 23088911 | 0.00E+00 | 9183 | 73937472 | 0.000124 | 5184 | 68826807 | 7.53E-05 | 0 | 1.648969475 |
| RLTR24_MM | 15 | 23088911 | 0.00E+00 | 51 | 73937472 | 6.90E-07 | 58 | 68826807 | 8.43E-07 | 0 | 0.818531142 |
| RLTR25_MM | 41 | 23088911 | 0.00E+00 | 176 | 73937472 | 2.38E-06 | 300 | 68826807 | 4.36E-06 | 0 | 0.546115418 |
| RLTR26B_MM | 12 | 23088911 | 0.00E+00 | 31 | 73937472 | 4.19E-07 | 51 | 68826807 | 7.41E-07 | 0 | 0.56582814 |
| RLTR26C_MM | 7 | 23088911 | 0.00E+00 | 27 | 73937472 | 3.65E-07 | 23 | 68826807 | 3.34E-07 | 0 | 1.092770476 |
| RLTR26D_MM | 7 | 23088911 | 0.00E+00 | 23 | 73937472 | 3.11E-07 | 24 | 68826807 | 3.49E-07 | 0 | 0.892091947 |
| RLTR27_MM | 158 | 23088911 | 0.00E+00 | 519 | 73937472 | 7.02E-06 | 461 | 68826807 | 6.70E-06 | 0 | 1.047995595 |
| RLTR28_MM | 3 | 23088911 | 0.00E+00 | 20 | 73937472 | 2.70E-07 | 37 | 68826807 | 5.38E-07 | 0 | 0.503177597 |
| RLTR29_MM | 27 | 23088911 | 0.00E+00 | 134 | 73937472 | 1.81E-06 | 122 | 68826807 | 1.77E-06 | 0 | 1.022440379 |

| | WT | | | KO | | | Input | | | %WT/%Input | %KO/%Input |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RLTR2IAP_MM | 1105 | 23088911 | 0.00E+00 | 3331 | 73937472 | 4.51E-05 | 2733 | 68826807 | 3.97E-05 | 0 | 1.134561457 |
| RLTR2aIAP_MM | 3534 | 23088911 | 0.00015306 | 11006 | 73937472 | 0.000149 | 9186 | 68826807 | 0.000133465 | 1.146817408 | 1.115311274 |
| RLTR30B_MM | 15 | 23088911 | 6.50E-07 | 28 | 73937472 | 3.79E-07 | 45 | 68826807 | 6.54E-07 | 0.993648813 | 0.579213322 |
| RLTR30C_MM | 17 | 23088911 | 7.36E-07 | 32 | 73937472 | 4.33E-07 | 59 | 68826807 | 8.57E-07 | 0.85891677 | 0.504883283 |
| RLTR30D2_MM | 11 | 23088911 | 4.76E-07 | 54 | 73937472 | 7.30E-07 | 60 | 68826807 | 8.72E-07 | 0.546506847 | 0.837790698 |
| RLTR30D_MM | 7 | 23088911 | 3.03E-07 | 25 | 73937472 | 3.38E-07 | 35 | 68826807 | 5.09E-07 | 0.596189288 | 0.664913253 |
| RLTR30E_MM | 9 | 23088911 | 3.90E-07 | 22 | 73937472 | 2.98E-07 | 36 | 68826807 | 5.23E-07 | 0.745236609 | 0.568870227 |
| RLTR30_MM | 73 | 23088911 | 3.16E-06 | 306 | 73937472 | 4.14E-06 | 355 | 68826807 | 5.16E-06 | 0.612983352 | 0.802391091 |
| RLTR31B_MM | 171 | 23088911 | 7.41E-06 | 622 | 73937472 | 8.41E-06 | 882 | 68826807 | 1.28E-05 | 0.577938595 | 0.65646991 |
| RLTR31C_MM | 11 | 23088911 | 4.76E-07 | 39 | 73937472 | 5.27E-07 | 45 | 68826807 | 6.54E-07 | 0.728675796 | 0.806761413 |
| RLTR31_MM | 12 | 23088911 | 5.20E-07 | 64 | 73937472 | 8.66E-07 | 85 | 68826807 | 1.23E-06 | 0.420839497 | 0.700896793 |
| RLTR33_MM | 82 | 23088911 | 3.55E-06 | 303 | 73937472 | 4.10E-06 | 330 | 68826807 | 4.79E-06 | 0.740720024 | 0.854715763 |
| RLTR34B_MM | 4 | 23088911 | 1.73E-07 | 7 | 73937472 | 9.47E-08 | 11 | 68826807 | 1.60E-07 | 1.083980523 | 0.592377261 |
| RLTR34C_MM | 1 | 23088911 | 4.33E-08 | 17 | 73937472 | 2.30E-07 | 11 | 68826807 | 1.60E-07 | 0.270995131 | 1.438630492 |
| RLTR34D_MM | 3 | 23088911 | 1.30E-07 | 8 | 73937472 | 1.08E-07 | 8 | 68826807 | 1.16E-07 | 1.117854914 | 0.930878554 |
| RLTR34_MM | 350 | 23088911 | 1.52E-05 | 987 | 73937472 | 1.33E-05 | 1413 | 68826807 | 2.05E-05 | 0.738380222 | 0.650231516 |
| RLTR35B_MM | 2 | 23088911 | 8.66E-08 | 9 | 73937472 | 1.22E-07 | 11 | 68826807 | 1.60E-07 | 0.541990261 | 0.761627908 |
| RLTR35_MM | 82 | 23088911 | 3.55E-06 | 266 | 73937472 | 3.60E-06 | 273 | 68826807 | 3.97E-06 | 0.895375853 | 0.907009873 |
| RLTR3_MM | 5 | 23088911 | 2.17E-07 | 17 | 73937472 | 2.30E-07 | 26 | 68826807 | 3.78E-07 | 0.57325893 | 0.608651362 |
| RLTR40_MM | 1 | 23088911 | 4.33E-08 | 9 | 73937472 | 1.22E-07 | 24 | 68826807 | 3.49E-07 | 0.124206102 | 0.349079458 |
| RLTR41A2 | 93 | 23088911 | 4.03E-06 | 229 | 73937472 | 3.10E-06 | 260 | 68826807 | 3.78E-06 | 1.06626161 | 0.819889188 |
| RLTR41B | 23 | 23088911 | 9.96E-07 | 55 | 73937472 | 7.44E-07 | 80 | 68826807 | 1.16E-06 | 0.857022101 | 0.639979006 |
| RLTR41C | 38 | 23088911 | 1.65E-06 | 81 | 73937472 | 1.10E-06 | 119 | 68826807 | 1.73E-06 | 0.951898863 | 0.633623217 |
| RLTR43C | 86 | 23088911 | 3.72E-06 | 275 | 73937472 | 3.72E-06 | 272 | 68826807 | 3.95E-06 | 0.942505124 | 0.941145597 |
| RLTR46A | 18 | 23088911 | 7.80E-07 | 56 | 73937472 | 7.57E-07 | 52 | 68826807 | 7.56E-07 | 1.031866075 | 1.002484596 |
| RLTR46A2 | 2 | 23088911 | 8.66E-08 | 15 | 73937472 | 2.03E-07 | 14 | 68826807 | 2.03E-07 | 0.425849491 | 0.997369879 |
| RLTR46B | 70 | 23088911 | 3.03E-06 | 231 | 73937472 | 3.12E-06 | 202 | 68826807 | 2.93E-06 | 1.033001241 | 1.064519534 |
| RLTR47 | 4 | 23088911 | 1.73E-07 | 28 | 73937472 | 3.79E-07 | 12 | 68826807 | 1.74E-07 | 0.993648813 | 2.172049959 |
| RLTR48A | 251 | 23088911 | 1.09E-05 | 903 | 73937472 | 1.22E-05 | 809 | 68826807 | 1.18E-05 | 0.924867189 | 1.039039968 |
| RLTR48B | 9 | 23088911 | 3.90E-07 | 42 | 73937472 | 5.68E-07 | 35 | 68826807 | 5.09E-07 | 0.766529084 | 1.117054264 |
| RLTR48C | 53 | 23088911 | 2.30E-06 | 192 | 73937472 | 2.60E-06 | 208 | 68826807 | 3.02E-06 | 0.759568083 | 0.859272511 |
| RLTR49 | 14 | 23088911 | 6.06E-07 | 52 | 73937472 | 7.03E-07 | 72 | 68826807 | 1.05E-06 | 0.579628474 | 0.672301178 |
| RLTR4I_MM | 3328 | 23088911 | 0.000144138 | 11806 | 73937472 | 0.00016 | 8747 | 68826807 | 0.000127087 | 1.134170544 | 1.256425312 |
| RLTR4_MM | 246 | 23088911 | 1.07E-05 | 1090 | 73937472 | 1.47E-05 | 828 | 68826807 | 1.20E-05 | 0.885643507 | 1.225431912 |

| | WT | | | KO | | | Input | | | %WT/%Input | %KO/%Input |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RLTR50A | 11 | 23088911 | 4.76E-07 | 57 | 73937472 | 7.71E-07 | 46 | 68826807 | 6.68E-07 | 0.712835018 | 1.153479947 |
| RLTR50B | 195 | 23088911 | 8.45E-06 | 714 | 73937472 | 9.66E-06 | 614 | 68826807 | 8.92E-06 | 0.946717517 | 1.082487439 |
| RLTR51A_Mm | 528 | 23088911 | 2.29E-05 | 1655 | 73937472 | 2.24E-05 | 2028 | 68826807 | 2.95E-05 | 0.776104398 | 0.75966667 |
| RLTR51B_Mm | 4 | 23088911 | 1.73E-07 | 12 | 73937472 | 1.62E-07 | 22 | 68826807 | 3.20E-07 | 0.541990261 | 0.507751938 |
| RLTR53_Mm | 8 | 23088911 | 3.46E-07 | 31 | 73937472 | 4.19E-07 | 29 | 68826807 | 4.21E-07 | 0.822330052 | 0.995077075 |
| RLTR5_MM | 39 | 23088911 | 1.69E-06 | 1441 | 73937472 | 1.95E-05 | 395 | 68826807 | 5.74E-06 | 0.294321294 | 3.39593923 |
| RLTR6B_Mm | 879 | 23088911 | 3.81E-05 | 2836 | 73937472 | 3.84E-05 | 3128 | 68826807 | 4.54E-05 | 0.837676445 | 0.843980684 |
| RLTR6C_Mm | 87 | 23088911 | 3.77E-06 | 267 | 73937472 | 3.61E-06 | 316 | 68826807 | 4.59E-06 | 0.820703608 | 0.786533462 |
| RLTR6I_MM | 9553 | 23088911 | 0.000413748 | 32180 | 73937472 | 0.000435 | 29259 | 68826807 | 0.000425111 | 0.973272543 | 1.023810515 |
| RLTR6_MM | 376 | 23088911 | 1.63E-05 | 1299 | 73937472 | 1.76E-05 | 1532 | 68826807 | 2.23E-05 | 0.731616097 | 0.789302377 |
| RLTR8 | 127 | 23088911 | 5.50E-06 | 373 | 73937472 | 5.04E-06 | 317 | 68826807 | 4.61E-06 | 1.194259298 | 1.095323976 |
| RLTR9A | 6 | 23088911 | 2.60E-07 | 16 | 73937472 | 2.16E-07 | 33 | 68826807 | 4.79E-07 | 0.541990261 | 0.451335056 |
| RLTR9A2 | 16 | 23088911 | 6.93E-07 | 77 | 73937472 | 1.04E-06 | 96 | 68826807 | 1.39E-06 | 0.496824406 | 0.746642173 |
| RLTR9A3 | 26 | 23088911 | 1.13E-06 | 107 | 73937472 | 1.45E-06 | 141 | 68826807 | 2.05E-06 | 0.549678067 | 0.706411385 |
| RLTR9A3A | 4 | 23088911 | 1.73E-07 | 6 | 73937472 | 8.11E-08 | 5 | 68826807 | 7.26E-08 | 2.38475715 | 1.117054264 |
| RLTR9A3B | 6 | 23088911 | 2.60E-07 | 19 | 73937472 | 2.57E-07 | 35 | 68826807 | 5.09E-07 | 0.511019389 | 0.505334072 |
| RLTR9A4 | 2 | 23088911 | 8.66E-08 | 12 | 73937472 | 1.62E-07 | 18 | 68826807 | 2.62E-07 | 0.331216271 | 0.620585702 |
| RLTR9B | 4 | 23088911 | 1.73E-07 | 15 | 73937472 | 2.03E-07 | 24 | 68826807 | 3.49E-07 | 0.496824406 | 0.581799096 |
| RLTR9B2 | 44 | 23088911 | 1.91E-06 | 153 | 73937472 | 2.07E-06 | 178 | 68826807 | 2.59E-06 | 0.736863164 | 0.800137184 |
| RLTR9D | 90 | 23088911 | 3.90E-06 | 278 | 73937472 | 3.76E-06 | 391 | 68826807 | 5.68E-06 | 0.686151354 | 0.661852271 |
| RLTR9D2 | 54 | 23088911 | 2.34E-06 | 219 | 73937472 | 2.96E-06 | 250 | 68826807 | 3.63E-06 | 0.643884431 | 0.815449613 |
| RLTR9E | 319 | 23088911 | 1.38E-05 | 886 | 73937472 | 1.20E-05 | 1266 | 68826807 | 1.84E-05 | 0.751123155 | 0.651467929 |
| RLTR9F | 40 | 23088911 | 1.73E-06 | 94 | 73937472 | 1.27E-06 | 146 | 68826807 | 2.12E-06 | 0.816697654 | 0.599332767 |
| RLTRETN_MM | 1055 | 23088911 | 4.57E-05 | 3620 | 73937472 | 4.90E-05 | 3408 | 68826807 | 4.95E-05 | 0.922798853 | 0.988785318 |
| RMER10A | 10 | 23088911 | 4.33E-07 | 26 | 73937472 | 3.52E-07 | 81 | 68826807 | 1.18E-06 | 0.368018079 | 0.298800523 |
| RMER15 | 18 | 23088911 | 7.80E-07 | 53 | 73937472 | 7.17E-07 | 73 | 68826807 | 1.06E-06 | 0.735027889 | 0.675843334 |
| RMER17A | 105 | 23088911 | 4.55E-06 | 374 | 73937472 | 5.06E-06 | 395 | 68826807 | 5.74E-06 | 0.792403483 | 0.881388808 |
| RMER17B | 172 | 23088911 | 7.45E-06 | 490 | 73937472 | 6.63E-06 | 575 | 68826807 | 8.35E-06 | 0.891691804 | 0.79327042 |
| RMER17B2 | 123 | 23088911 | 5.33E-06 | 363 | 73937472 | 4.91E-06 | 505 | 68826807 | 7.34E-06 | 0.726052301 | 0.669126564 |
| RMER17C | 134 | 23088911 | 5.80E-06 | 432 | 73937472 | 5.84E-06 | 660 | 68826807 | 9.59E-06 | 0.605222459 | 0.609302326 |
| RMER17D2_MM | 249 | 23088911 | 1.08E-05 | 756 | 73937472 | 1.02E-05 | 986 | 68826807 | 1.43E-05 | 0.75279479 | 0.713736498 |
| RMER17D_MM | 128 | 23088911 | 5.54E-06 | 453 | 73937472 | 6.13E-06 | 603 | 68826807 | 8.76E-06 | 0.632771383 | 0.699316724 |
| RMER19 | 27 | 23088911 | 1.17E-06 | 117 | 73937472 | 1.58E-06 | 154 | 68826807 | 2.24E-06 | 0.522633466 | 0.707225914 |
| RMER2 | 127 | 23088911 | 5.50E-06 | 345 | 73937472 | 4.67E-06 | 411 | 68826807 | 5.97E-06 | 0.921119702 | 0.781394406 |

| | WT | | | KO | | | Input | | | %WT/%Input | %KO/%Input |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RMER20 | 2 | 23088911 | 8.66E-08 | 4 | 73937472 | 5.41E-08 | 10 | 68826807 | 1.45E-07 | 0.596189288 | 0.372351421 |
| RMER20A | 9 | 23088911 | 3.90E-07 | 18 | 73937472 | 2.43E-07 | 42 | 68826807 | 6.10E-07 | 0.638774237 | 0.398947952 |
| RMER21B | 10 | 23088911 | 4.33E-07 | 35 | 73937472 | 4.73E-07 | 65 | 68826807 | 9.44E-07 | 0.458607144 | 0.501242298 |
| RMER6C | 317 | 23088911 | 1.37E-05 | 918 | 73937472 | 1.24E-05 | 1172 | 68826807 | 1.70E-05 | 0.806279881 | 0.729135249 |
| RMER6D | 76 | 23088911 | 3.29E-06 | 250 | 73937472 | 3.38E-06 | 334 | 68826807 | 4.85E-06 | 0.678299189 | 0.696765385 |
| SATMIN | 14462 | 23088911 | 0.000626361 | 52714 | 73937472 | 0.000713 | 46136 | 68826807 | 0.00067032 | 0.934421003 | 1.063601788 |
| SPIN_NA_10_Rodent | 161 | 23088911 | 6.97E-06 | 518 | 73937472 | 7.01E-06 | 586 | 68826807 | 8.51E-06 | 0.81899723 | 0.822858517 |
| SQR1_MM | 1 | 23088911 | 4.33E-08 | 7 | 73937472 | 9.47E-08 | 47 | 68826807 | 6.83E-07 | 0.063424392 | 0.138641487 |
| SQR2_MM | 6 | 23088911 | 2.60E-07 | 12 | 73937472 | 1.62E-07 | 218 | 68826807 | 3.17E-06 | 0.082044397 | 0.051241021 |
| SQR4_MM | 18 | 23088911 | 7.80E-07 | 65 | 73937472 | 8.79E-07 | 78 | 68826807 | 1.13E-06 | 0.687910716 | 0.775732128 |
| YREP | 8 | 23088911 | 3.46E-07 | 764 | 73937472 | 1.03E-05 | 326 | 68826807 | 4.74E-06 | 0.07315206 | 2.181568144 |
| ZP3AR_MM | 178 | 23088911 | 7.71E-06 | 605 | 73937472 | 8.18E-06 | 714 | 68826807 | 1.04E-05 | 0.743149112 | 0.788769643 |
| BIR containing sequences | 392430 | 23088911 | 0.016995951 | 1313705 | 73937472 | 0.017768 | 1392733 | 68826807 | 0.02023533 | 0.839914694 | 0.878057338 |

Appendix 1 – Repeats analysis.

## Appendix 2

| Name | Forward (5'-3') | Name | Forward (5'-3') |
|---|---|---|---|
| P1 seq L32R | GTCTGCTTCTGGGTGAAACG | P1 seq L79F | TTTGTTTGTTTTTCGAGACAGG |
| P1 seq L33R | CCTCACTTGATGCCCTGTCT | P1 seq L80F | ACCTCCCAATAAGGGGACAC |
| P1 seq L34F | TTTGTCCCTGTGCCCTTTAG | P1 seq L81F | GACCCTGTGCTCAGTCCAAT |
| P1 seq L35R | ATTCCGGTGTTTCATCCTTG | P1 seq L82F | TGGCTTCTGTCGGAGTTACC |
| P1 seq L36F | CGGAGCCACCCATTAAACTA | P1 seq L83F | GTTGTTGGGCTGTCATGTTG |
| P1 seq L37R | CCACTCCTGGAACTCAGCAT | P1 seq L84F | ACGTGCCTCACTTTTGAAGG |
| P1 seq L38R | TGATGAGTGTGAAGCAAATGG | P1 seq L85F | AAACATCCCTCGGCACTAGA |
| P1 seq L39R | TGCATCTCCAGCAGTTGTTC | P1 seq L86F | TTTTTATACAGAAACAAAATCACCATT |
| P1 seq L40R | CCAAACGATCAACTGGCTTT | P1 seq L87F | GGCTTAAGTGGGTGTCTCCA |
| P1 seq L41R | AGCATCAGGCAGACCTGAAC | P1 seq L88F | CAGGCCTTCCAAAACTTGAA |
| P1 seq L42R | GCACAGGTCTGCATTGGTTT | P1 seq L89F | CAGTGTGGCTCGGCTTATTT |
| P1 seq L43F | GGTCAGCAGGGATGACATAGA | P1 seq L90F | CTTTTGTGGCCACTGTACCA |
| P1 seq L44F | TCAACAGGAGAGGAGGAGGA | P1 seq L91F | TGCCTCAATGGAGACTGAAA |
| P1 seq L45F | TGTTGCAGTTTGAACCCAGA | P1 seq L92F | TGAATTTGCAAGTTTGAGGAGA |
| P1 seq L46F | CAGTCTCCTTCTGGCCTTTG | P1 seq L93F | CAGAGAGAAAAAGGGAAGTTGG |
| P1 seq L47F | CCCCAAACTCTCACTGAAGC | P1 seq L94F | TGGCCCCATTAGTTACTGAAA |
| P1 seq L48F | GTCACACCTTCCCACATCCT | P1 seq L95F | CCCGATAGCAAGCACTAGGA |
| P1 seq L49F | CAGTTGTCTCTGCCTGTGGA | P1 seq L96F | ACGCCTGGGCAAATTGTAT |
| P1 seq L50F | TGGTAAAATGGGAGCCTGAG | P1 seq L97F | GGGGATAGCCTGGGTCTCTA |
| P1 seq L51F | CATCATTGGGCAGTTGGTAA | P1 seq L98F | CTGACCTCCAATCACAGCAA |
| P1 seq L52F | TCCTCAGTGACGCTACATCAG | P1 seq L99F | TCTCCCAGTTCCCAGAGATG |
| P1 seq L53F | AACCAAAGCCCAGAGACTGA | P1 seq L100F | CGGTGTGTTTGCTTTCTTCC |
| P1 seq L54F | CACAGAGTGAGGGATCTTGGA | P1 seq L101F | GTACCTCTCCAGGGTGGTCA |
| P1 seq L55F | AGGGATCCACTGTCTGCTGT | P1 seq L102F | CTGCGCTACCACAGTTTGAA |
| P1 seq L56F | CATTTGGATTTTCTTTGCCACT | P1 seq L103R | GCTCCGTTAGTGTGGGCTAT |
| P1 seq L57F | CCCTGGGGAAGGTGATTATT | P1 seq L104R | ATTGGCAAGTGCTTCATTCC |
| P1 seq L58F | TCTCCGTGGAAAACTTCTGG | P1 seq L105F | TGGTTGATTTCACCCCTGAG |
| P1 seq L59F | ACTGGACAGAGGTGCCATGT | P1 seq L106R | CACCTTCCTGGTGAGAGCAC |
| P1 seq L60F | TGATGCATTGGTGAATTGTCA | P1 seq L107R | ACATAGTGCCAGCCGTTTCT |
| P1 seq L61F | TCCAACTAGTGCTGCCCATA | P1 seq L108F | GGGGCATCTAGGTTCTTTCC |
| P1 seq L62F | TTTGTCCCTGTGCCCTTTAG | P1 seq L109F | CAGGTCCATCCATTTGGCTA |
| P1 seq L63F | GATGCACGCAAAGCATGTAG | P1 seq L110F | ACACCTGGCTCAACCAAAAG |
| P1 seq L63R | TCTATGGGCTTCACTTGTTGG | P1 seq L111F | AGTGTGCTAGCAGGGGTCAC |
| P1 seq L64F | TTGTGGAATGTGGAACGTGT | P1 seq L112F | GTGCGGAAGCTAAGAAATGG |
| P1 seq L65F | CACCCCTTAACCCCTCAGAT | P1 seq L113F | GCGAGGATGTGGAGAAAGAG |
| P1 seq L66F | TTGGAGCGTTTGGGCTAATA | P1 seq L114F | CAGTGCCTGGCAAACACA |
| P1 seq L67F | TTGTGGCTTCACAGAATGAA | P1 seq L115F | TCAATCCCTTTAAAATGTCCAA |
| P1 seq L68F | CATTGATGAGAGTGGGGTGT | P1 seq L116F | CACAACCAAACAGGTAGAAGTCC |
| P1 seq L69F | AAGGGGTTTTTGTGTTTCCT | P1 seq L117R | AAGGACTTCTACCTGTTTGGTTG |
| P1 seq L70F | CTGAAGAGGACTCAAGCTGAAA | P1 seq L118F | ACGCCTGCTGATAGAGGGTA |
| P1 seq L71F | TGGCATTTTATTGCCACTGA | P1 seq L119R | CGTGGGAGACAAGCTCTCTT |
| P1 seq L72F | TGACAGACTTTTCCTGGGGTA | P1 seq L120R | CGTGGGAGACAAGCTCTCTT |
| P1 seq L73F | CCGTCTCTTTAGGGTCACCA | P1 seq L121F | ACGCCTGCTGATAGAGGGTA |
| P1 seq L74F | TCAGATCAGGTGCATTCAGG | P1 seq L122R | CGTGGGAGACAAGCTCTCTT |
| P1 seq L75F | ATTCATTCTCCTGGCCCTCT | P1 seq L123F | ACGCCTGCTGATAGAGGGTA |
| P1 seq L76F | GCCTTGCAGAAACTTTGGAG | P1 seq L124F | GCCGGGTATAGTACTGGCATT |
| P1 seq L77F | TCCAGCTACTTCACCGAAGC | P1 seq L125F | GAACACAAATGCCAGTACTATACCC |
| P1 seq L78R | CCAGGACCAGATGGGTTTAG | | |

Appendix 2 – Primers used for PAC sequencing.