Characterisation of AEBP2: a Polycomb Repressive Complex 2 component

Anne Grijzenhout

Lincoln College

Thesis submitted for the degree of Doctor of Philosophy

October 2013

Abstract

Characterisation of AEBP2: a Polycomb Repressive Complex 2 component

Anne Grijzenhout
Lincoln College
Submitted for the degree of Doctor of Philosophy, October 2013

Polycomb complexes act as repressors of transcription at many developmentally important genes. Repression by polycomb-repressive complex 2 (PRC2) is mediated by the deposition of a tri-methylation mark on lysine 27 of the histone H3 tail (H3K27me3). The PRC2 complex consists of three core proteins: EED, SUZ12, and one of either EZH2 or EZH1, which contain the enzymatic activity. The zinc finger containing protein AEBP2 (AE1 binding protein 2) has been co-purified with the PRC2 complex. The aim of this thesis was to study the function of AEBP2. Chromatin immunoprecipitation experiments indicate that AEBP2 is enriched at the same genomic locations as H3K27me3 genomewide. Using an X inactivation model, I found that AEBP2 and H3K27me3 also co-localise by immunofluorescence upon *Xist* induction.

Large scale immunoprecipitation followed by mass spectrometry confirmed the interaction of AEBP2 with PRC2 complex members. Interestingly, I did not detect any peptides of the Polycomblike (PCL) family proteins. These proteins have previously been shown to interact with the PRC2 complex. Their absence in the AEBP2 immunoprecipitation, together with published work demonstrating an absence of AEBP2 in PCL purifications, suggests that AEBP2 and PCL family proteins might be mutually exclusive in PRC2 complexes.

I have created an embryonic stem cell line in which both alleles of *Aebp2* contain a gene-trap. This cell line was used to address the suggestion that AEBP2 might use its zinc finger domain to act as a sequence-specific PRC2 recruitment factor. AEBP2-depleted cells do not show diminished recruitment of PRC2. Additionally, electromobility shift assays failed to show that AEBP2 could bind DNA. Furthermore, recruitment of AEBP2 to genomic targets depended on the presence of the core PRC2 protein EED. Together, this makes it unlikely that AEBP2 acts as a sequence-specific PRC2 recruitment factor.

In keeping with the lack of effect on PRC2 enrichment, preliminary experiments have shown that AEBP2 depletion does not greatly affect levels of H3K27me3, or the expression of PRC2 target genes. The *Aebp2* gene-trapped cell line will be used to investigate further functions of AEBP2. Additionally, these cells have been used for blastocyst injection to create an *Aebp2* gene-trapped mouse. Future work will show the effect of AEBP2 depletion on development.

Acknowledgements

Many people have contributed to the creation of this thesis, supporting me both academically and socially.

Of course, I thank Sarah for taking me under her wings, guiding my development as a scientist and perking me up when I was feeling low. I know I have needed your help at times when you were already stressed yourself, but I hope that in the end this journey has been as worthwhile to you as it has been to me.

I thank Neil for the good supervisor-student relationship we have built. I feel this relationship is marked by mutual respect and this has been very important to me. Of course, I also thank Neil for guidance on my project and for help during writing the thesis. Last, I thank Neil for giving me the opportunity to do my PhD in an environment that was both supportive and challenging.

I further thank the Brockdorff and Klose labs as a whole for making me feel welcome and contributing to this supportive and challenging environment. Thank you to each of you. I have learnt from you and I have laughed with you. Special thanks go to Rob, Anca and Hannah for scientific support and allowing me to use quite a few of their reagents. I also thank Jonathan for all the help with the mice. And extra special thanks go to Tatyana, not only for scientific support, but also for basically being the rock upon which

we have built our lab practice, the rock upon which we can lean when things are not going well and for being a lot of fun to be around.

Furthermore, I acknowledge scientific support from B. Thomas, J. McGouran, B. Kessler, L. Gregory, S. Lamble and I. Stancheva. I have also made use of reagents and cell lines received from H. Koseki, A. Otte and A. Wutz. I thank Neil, Sarah, Tatyana, Nick, Dave and Ron for useful discussion of this thesis. I also thank Louis Mahadevan and Liz Bikoff for being a tough but dedicated thesis committee.

Then there are many people outside of the lab that deserve my gratitude for support. Thank you to all the people I have worked with at Lincoln College boat club for making some great memories and keeping me sane. Especially W1 TT 2011 and W1 HT 2012 will always have my love. Thank you to my housemates Ron, Kristin and Ainhoa for making my life easy and being good listeners. And of course Oxford would not have been the same without the banter crew. Thanks guys!

I thank Jerome for making especially the final year of my PhD a lot more fun! I have learnt a lot from you and it has been wonderful having someone I can so easily talk to and relate to. You've made my life better.

Natuurlijk bedank ik ook mijn lieve vriendinnen uit Nederland, mijn allerliefste zusje en mijn ouders, die altijd voor mij klaarstaan. Jullie maakten het verschil.

Declaration of authorship

The thesis submitted is entirely my own work except where otherwise indicated, and has been completed between 1 April 2010 and 11 October 2013. I have clearly indicated the presence of material I have quoted from other sources. This thesis has not been submitted, either partially or in full, for another qualification of this University or for a qualification at any other institution.

Abbreviations

2i	Embryonic Stem Cell medium containing two chemical inhibitors directed against MEK and GSK3 kinase
3C	Chromatin Conformation Capture
AEBP2	AE1 binding protein 2
AEBP2-FL	Full length AEBP2 (amino acids 1-496)
AEBP2-ZF	recombinant AEBP2 containing the zinc finger domain (amino acids 223-348)
ATRA	all-trans retinoic acid
BAC	Bacterial Artificial Chromosome
bp	base pairs
Bio-CAP	biotinylated CXXC Affinity Purification
BSA	Bovine Serum Albumin
ChIP	Chromatin Immunoprecipitation
ChIP-chip	Chromatin Immunoprecipitation, after which bound DNA sequences are analysed by microarray.
ChIP-qPCR	Chromatin Immunoprecipitation, after which bound DNA sequences are analysed by qPCR (see qPCR)
ChIP-seq	Chromatin Immunoprecipitation, after which bound DNA sequences are analysed by massively parallel sequencing
CNS	Central Nervous System
CXC	Cysteine Rich domain present in EZH2 and EZH1
DMSO	Dimethyl sulfoxide
DNMT	DNA methyltransferase

E	Embryonic day (indication of progression of mouse development)
EED	Embryonic Ectoderm Development
EGS	ethylene glycol bis[succinimidylsuccinate]
EM	Electron Microscopy
EMSA	Electromobility Shift Assay
ENU	ethyl-*N*-nitrosurea
ES cells	Embryonic Stem cells
Esc	extra sex combs
EZH1/2	enhancer of zeste homologue 1 or 2
FISH	Fluorescence *in situ* hybridisation
FS2	Flag+2xStrep2 tag
GFP	Green Fluorescent Protein
GO	Gene Ontology
GST	Glutathione S-Transferase
H2AK119Ub	ubiquitination of lysine 119 on histone H2A
H3K27me3	trimethylation on lysine 27 on histone H3
HAT	Histone Acetyl Transferase
HDAC	Histone Deacetylase
HEK 293T	Human Embryonic Kidney 293T
HMT	Histone methyl transferase
iBAQ	intensity-Based Absolute Quantification Index
IF	Immunofluorescence
IMAC	Immobilised Metal ion Affinity Chromatography
kDa	kilo Dalton
LIF	Leukaemia Inhibitory Factor
MACS	Model-based analysis of ChIP-seq

NCE	Nuclear Cell Extract
NEB	New England Biolabs
NuRD	Nucleosome remodelling and deacetylase complex
PBS	Phosphate buffered saline
PcG	Polycomb Group
PCL1	Polycomblike 1, also known as PHF1
PCL2	Polycomblike 2, also known as MTF2
PCL3	Polycomblike 3, also known as PHF19
PEF	Primary Embryonic Fibroblast
PHD	Plant Homeodomain
PNK	Polynucleotide Kinase
PRC2	Polycomb Repressive Complex 2
qPCR	quantitative real-time Polymerase Chain Reaction
REST	RE-1 Silencing Transcription Factor, also known as Neuron-Restrictive Silencer Factor (NRSF)
SELEX	Systematic Evolution of Ligands by Exponential enrichment
SET domain	Su(var)3-9, E(z) and Trithorax domain
SUZ12	Suppressor of zeste 12 homologue
TET	Ten Eleven Translocation protein
TrxG	Trithorax Group
TSS	Transcription Start Site
U	Units (for enzymes)
UTR	Untranslated region
WCE	Whole Cell Extract
WT	Wild type
XCI	X chromosome inactivation
Xist	X inactive specific transcript

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Chromatin & epigenetics

1.1.1 History of epigenetics

Conrad Waddington first coined the term “epigenesis” in 1942 (reprinted in Waddington, 2012). By epigenesis, Waddington meant the process of one genotype giving rise to various phenotypes during development. Later, he used the phrase “epigenetic landscape” in which a marble rolls down to the point of lowest local elevation as a metaphor for how totipotent cells differentiate into specific cell types (Waddington, 1957). However, in later times, epigenetics has been defined as “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” (Russo et al., 1996) rather than the study of development as proposed by Waddington. The term epigenetics is often more generically applied to the study of the effect that chromatin, the complex of nucleic acids and proteins, has on gene transcription, without rigorously testing whether these effects are inherited across cell division. A more recent definition of epigenetics stated that epigenetics may be defined as “the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states” (Bird, 2007).

Many observations originally termed epigenetic, such as position effect variegation in the fruit fly *Drosophila melanogaster*, have indeed later been found to be caused by changes in chromatin structure. The importance of epigenetics was further highlighted by

observations that mutations in histones phenocopy mutations in the proteins that modify these histones (for example, Adhvaryu et al., 2005). Taken together, it has been shown that epigenetic phenomena widely impact on eukaryotic biology.

1.1.2 Histones and nucleosomes

Chromatin can be viewed as a series of compaction events, and regulation at each level of compaction can influence transcription. The basic unit of chromatin is the nucleosome. A nucleosome consists of approximately 147 base pairs (bp) DNA wrapped around a complex of 8 histones (called an octamer). The octamer is organised into a heterotetramer of two copies of core histones H3 and H4, with heterodimers of core histones H2A and H2B binding the heterotetramer on each side (Luger et al., 1997). The histones consist largely of globular domains but have N-terminal tails that protrude from the nucleosome. Nucleosomes are connected to each other through linker DNA, the length of which can vary from 20 to 80 bp (Van Holde, 1989).

Additionally, a single copy of the linker histone H1 can bind nucleosomes at the point where DNA enters or exits the nucleosome, but H1 is not present at all nucleosomes, and the ratio of H1 to nucleosomes can depend on cell type (Fan et al., 2003; Fan et al., 2005). H1 is thought to promote packaging of nucleosomes by shielding the negative charges on linker DNA (Thomas, 1999; Harvey and Downs, 2004). Chromatin compaction is thought to decrease accessibility of transcription factors and repress transcription (see 1.1.7).

1.1.3 Histone variants

The histone proteins themselves have diversified during evolution, such that now different subtypes of the core histones exist. One subtype is the H3 subtype Cenp-A, which is only 46% identical to H3 (Palmer et al., 1991), and is incorporated into nucleosomes present at the centromere that form the base on which the kinetochore assembles (Palmer et al., 1990). A centromeric subtype of H3 is found in all eukaryotes (Malik and Henikoff, 2003).

Histone H2A has the most known subtypes in mammals. Examples include H2AX, H2AZ and macroH2A. These subtypes are believed to play roles in transcription regulation, DNA repair, and other processes (reviewed in Skene and Henikoff, 2013).

1.1.4 Histone modifications

The amino acids of histone proteins can be modified by the covalent addition of many different moieties. The location and type of moiety determine the functional outcome.

Some histone modifications disrupt internucleosomal contacts and thus decrease chromatin compaction. An example of this type of histone modification is the addition of an acetyl group on lysine 16 on the histone tail of H4 (H4K16ac) (Shogren-Knaak et al., 2006). It is thought that acetylation disrupts interaction of histone H4 with an acidic patch on the H2A-H2B dimer on a neighbouring nucleosome (Robinson et al., 2008; Allahverdi et al., 2011).

Other histone modifications may not affect chromatin structure, but instead recruit, or block recruitment of transcription factors, other histone modifiers, or proteins involved in DNA replication and repair (reviewed in Rossetto et al., 2012). For example, phosphorylation of histone subtype H2A.X plays an important role in DNA repair by recruiting many factors involved in the DNA damage response (reviewed in Lukas et al., 2011).

As the field has matured, the enzymes that place and remove these marks, which are called "writers" and "erasers" respectively, have been discovered. Interestingly, many protein domains have been identified that can interact selectively with a modification and thereby recruit proteins to particular regions of chromatin. These proteins are called "readers". An overview of the most well-known domains is listed in Table 1. In 2001, Jenuwein and Allis hypothesized that the intricate combinatorial nature of histone modifications, together with the proteins that can "read" them, extends the information potential of the genetic code in the regulation of gene transcription. This theory became known as the histone code hypothesis (Jenuwein and Allis, 2001).

"Reading" domain	**Histone modification**	**Example of protein**
Bromodomain	H3Kac, H4Kac, H2AKac, H2BKac	TAF1
Chromodomain	H3K9me2/3, H3K27me2/3	CBX family; HP1
Tudor	H3K36me3	KDM4
PWWP	H3K36me3, H3K79me3, H4K20me3	DNMT3, LEDGF
PHD	H3K4me2/3, H3K9me3, H3un	CFP1, PCL1, PCL2, PCL3
WD40	H3K27me3, H3K9me3, H3un	EED
MBT	H3K4me1/2, H4Kme1/2	SFMBT1
14-3-3	H3S10P; H3S28P	14-3-3 family

Table 1: Well-known domains that can "read" histone modifications. Domains in proteins that can act as "readers" of the chromatin state are listed, with the types of histone modifications thet can recognize and an example of a protein containing such a domain. The list is not exhaustive. Ac=acetylation, me=methylation, me2=dimethylation, me3=trimethylation, P=phosphorylation, un=unmodified

A few prevalent histone modifications are discussed here. Methylation of lysine 27 on histone 3 (H3K27me) is discussed in 1.3.3.1.

1.1.4.1 Acetylation

The acetyl mark is deposited by histone acetyl transferases (HATs) by transferring an acetyl group from acetyl-coenzyme A to the ε-amino group of a lysine (Sterner and Berger, 2000) (Figure 1.1A). Conversely, acetylation is removed by histone deacetylases (HDACs). Levels of histone acetylation correlate with transcription (reviewed in Choi and Howe, 2009). The majority of known acetylation sites are on the tails of histone H3 and H4. However, some acetylation sites have also been detected in the globular domains of histones (Tropberger et al., 2013; Watanabe et al., 2013).

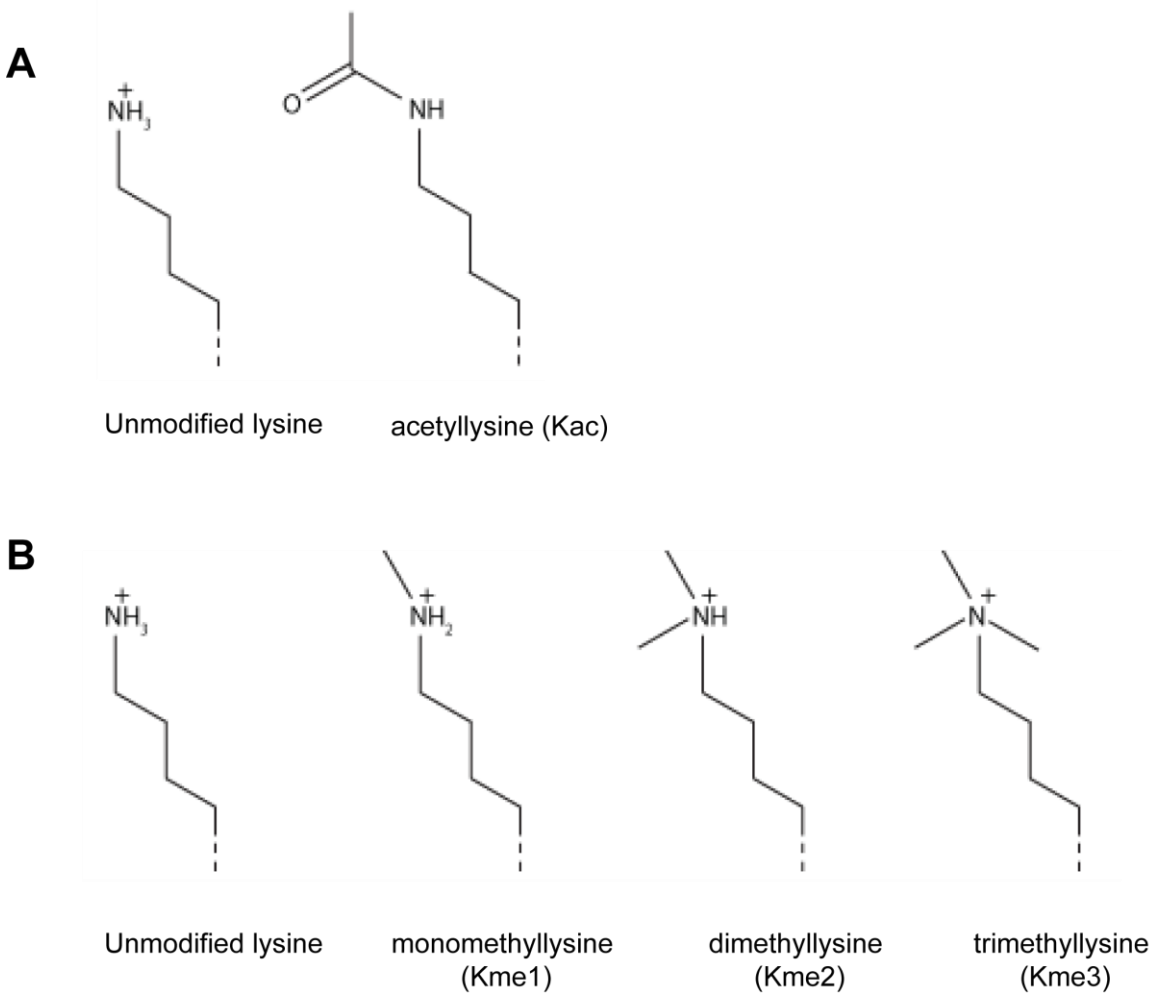


Figure 1.1: Common histone modifications on lysine tails. (A) Lysine acetylation. (B) Lysine methylation.

The acetylation of a lysine results in the neutralisation of lysine's positive charge, and as such is expected to reduce the interaction between histone and the negatively charged DNA. Additionally, many transcriptional activators contain a bromodomain which can bind to acetylated lysines (see Table 1). It is not clear to what extent acetylation's positive effect on transcription is caused by disrupting DNA-nucleosome interactions versus recruitment of activators. Specifically, acetylation of H3K27 has been found to be enriched at enhancer sequences (reviewed in Hon et al., 2009). It has been proposed that the function of acetylation at this residue may solely be to prevent methylation of it, as lysines can be either acetylated or methylated and H3K27 methylation is associated with

repression (see 1.3.3.1). Acetylation has also been implicated in regulation of DNA replication and repair (reviewed in Kouzarides, 2007).

1.1.4.2 Methylation

Methylation can occur on arginines or on lysines. Lysine methylation is the most studied histone modification and involves addition of one methyl group (monomethylation) or two or three (di- and tri-) to the ε-amino group of a lysine residue. Methylation of H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20 has been especially well studied. Methylation of H3K27 will be discussed in 1.3. Here I describe briefly findings related to H3K4, H3K9, and H3K36 methylation.

1.1.4.2.1 H3K4 methylation

Chromatin immunoprecipitation (ChIP) analysis has shown that H3K4me3 is enriched at transcription start sites of active genes (Bernstein et al., 2005). Further research in embryonic stem cells (ES cells) showed that H3K4me3 could be found at virtually all promoters that contained a genomic element known as a CpG island (discussed in 1.1.5), regardless of their expression status (Guenther et al., 2007; Mikkelsen et al., 2007). There are believed to be at least ten H3K4 methyltransferases in mammals (Hublitz et al., 2009), including members of the COMPASS family of H3K4 methyltransferases. These transferases are related to *Saccharomyces cerevisiae* Set1p, the only H3K4 methyltransferase in budding yeast and thus responsible for catalysing the mono-, di-, and trimethylation state of H3K4. The demethylases identified for H3K4me3 are members of the JARID1 family, whilst LSD1 can demethylate H3K4me2 and H3K4me1 (Shi et al., 2004). It is unclear to which extent H3K4me3 is required for and causative of transcription and to which extent it is a read-out of transcription. Arguments can be made for each case as RNA pol II can recruit H3K4 methyltransferases, and as such H3K4me3

could be a consequence of transcription, whilst the H3K4me3 modification might recruit chromatin remodellers associated with gene activation (Sims and Reinberg, 2006), and as such could increase transcription levels. It is known that a *SET1* deletion in yeast results in reduced gene expression of many genes, but the effect is moderate and cells are viable (Boa et al., 2003).

H3K4me2, like H3K4me3, is associated with transcription, but also with enhancers. H3K4me1 is also enriched at enhancer sequences (Heintzman et al., 2007; Hon et al., 2009).

1.1.4.2.2 H3K9 methylation

H3K9me2 and H3K9me3 are associated with gene repression. Large blocks of chromatin near centromeres (called pericentric heterochromatin) are marked with this modification as are several repetitive sequences and transposable elements (Rosenfeld et al., 2009). It is thought that this mark stabilises its own presence through a feed-forward loop: the chromodomain of HP1 can bind to H3K9me2/me2 and then recruit the H3K9 methyltransferase SUV39H1 (Bannister et al., 2001; Lachner et al., 2001; Nakayama et al., 2001).

Typically, H3K9me2/3 is associated with regions that are constitutively silenced, and correlate highly with other markers of repressed regions, such as DNA methylation (see 1.1.5), association with the nuclear lamina and reduced accessibility (see 1.1.7) (Zhu et al., 2013).

1.1.4.2.3 H3K36 methylation

H3K36me3 is highly enriched on the coding region of actively transcribed genes, with highest levels of enrichment towards the 3' end. It is deposited by SETD2, which is recruited by the elongating form of RNA polymerase II. In *S. cerevisiae* H3K36me3, placed by Set2p, is important for suppressing intragenic initiation of transcription by recruiting the histone deacetylase Rpd3S (Carrozza et al., 2005; Keogh et al., 2005).

H3K36me2 is found across the genome but is depleted at CpG islands (Blackledge et al., 2010). This is due to the demethylases KDM2A and KDM2B which are specifically recruited to CpG islands and remove H3K26me2 there (Tsukada et al., 2006; Blackledge et al., 2010; Farcas et al., 2012).

1.1.4.3 Other histone modifications

In addition to acetylation and methylation, there are many other histone modifications. An example is histone ubiquitination, which will be discussed in 1.3.2.2. Other examples are phosphorylation, proline isomerisation and sumoylation (reviewed in Kouzarides, 2007).

1.1.5 DNA methylation

The DNA of mammals is pervasively methylated on cytosine bases in the context of a CpG sequence (where the p indicates the phosphate group between a cytosine and guanosine base). When a methylated cytosine is spontaneously deaminated it is converted into a thymidine base, resulting in a C-T transition. It is believed the genome is largely depleted of CpG sequences in order to protect against these mutations (Bird, 1980). However, there are regions called CpG islands which are CpG and GC rich and

are on average 1000 bp in length. Despite genomewide pervasive DNA methylation, CpG islands are predominantly unmethylated. More than half of transcriptional start sites have a CpG island in humans and mice. It is believed that unmethylated CpG islands typically contain RNA polymerase II, even if repressed, and that transcriptional regulation of CpG island associated genes is more dependent on transcriptional elongation (reviewed in Deaton and Bird, 2011). Some CpG islands are methylated and this is associated with stable repression (Mohn et al., 2008). Recent evidence suggests that lower vertebrates also use regions of non-methylation in transcriptional regulation (Long et al., 2013), whilst this is not conserved in invertebrates such as *Drosophila melanogaster* or *Caenorhabditis elegans*.

DNA methylation is placed by DNA methyl transferases (DNMTs), which can either copy existing DNA methylation from the parental strand to the daughter strand during DNA replication (DNMT1) or catalyse methylation at new locations (DNMT3A and DNMT3B). DNA methylation can be removed passively (by dilution of the mark over multiple rounds of DNA replication and cell division) or actively. Which enzymes catalyse removal is the topic of much current research. The ten eleven translocation (TET) enzymes have shown to possess the capacity to convert methylated cytosines to hydroxymethylated cytosines and are thought to be play a role in DNA demethylation (Tahiliani et al., 2009; Inoue and Zhang, 2011; Pastor et al., 2013).

CpG islands regulate transcription through attraction and exclusion of other chromatin modifying proteins and transcriptional activators and repressors. For example, the CXXC domain binds exclusively to non-methylated DNA (Voo et al., 2000). CXXC domains are

found in proteins such as CFP1, which requires the CXXC domain to recruit SET1A and SET1B histone methyltransferases (HMTs) to place H3K4me3 at appropriate locations (Clouaire et al., 2012). Similarly the H3K4 histone methyltransferases MLL1 and MLL2 contain a CXXC domain, as do the H3K36me2 histone demethylases KDM2A and KDM2B. Conversely, MBD (methyl CpG Binding Domain) containing proteins bind exclusively to methylated CpG DNA and typically recruit further transcriptional repressors. Additionally, transcription factors may be prevented from binding by the presence of DNA methylation (reviewed in Klose and Bird, 2006).

It has also been reported that cytosine methylation occurs rarely in a non-CpG context (Yan et al., 2011; Arand et al., 2012). The function of this methylation remains unclear.

In summary, DNA methylation is an important epigenetic mechanism. Methylated cytosines at promoters typically antagonise transcription, whilst cytosines in CpG islands often remain unmethylated and attract other chromatin modifying proteins.

1.1.6 Cross-talk

In addition to regulation of individual histone and DNA modifications, there is increasing evidence for co-ordinate regulation of and by these modifications. One such instance of "cross-talk" was already described in 1.1.4.2.3: Rpd3S, which itself modifies chromatin by deacetylating histones, is recruited by the presence of another chromatin mark, namely H3K36me3. Furthermore, there are many interactions known between the H3K9 methylation machinery and DNA methylation (Lehnertz et al., 2003; reviewed in Denis et al., 2011). Many other examples have been reported (see also 1.3.4, 1.3.6). Taken

together, a picture emerges in which modifications are placed and interpreted co-ordinately.

1.1.7 Further mechanisms of chromatin regulation

Chemical modification of DNA and histones is one way in which chromatin influences gene expression. However, there are further layers of chromatin regulation that can impact on gene expression.

Initially, chromatin was divided into euchromatin and heterochromatin based on cytological criteria. Heterochromatin consists of densely packed DNA and therefore stained more intensely than euchromatin, which is less compacted DNA. Now it is known that these cytological observations correlate with functional readouts of gene expression. Overall, packaged chromatin is less accessible and correlates with decreased gene expression. This packaging occurs at several levels.

Firstly, the presence of nucleosomes is thought to present a barrier to transcription. It has been shown that the basic RNA polymerase II transcription machinery is inhibited by nucleosomes (reviewed in Workman and Kingston, 1998) whilst nucleosomes may also block access to transcriptional activators (reviewed in Zaret and Carroll, 2011). Therefore, there are ATP-dependent chromatin remodelling complexes that can alter nucleosome positions in a non-covalent manner. The presence of particular histone variants can also contribute to the stability of nucleosomes, either by affecting interactions with remodellers or the intrinsic stability of nucleosomes (reviewed in Li and Reinberg, 2011).

Secondly, the DNA wrapped around octamers is compacted further (Figure 1.2). Exactly how this packaging is organised is unknown (reviewed in Li and Reinberg, 2011).

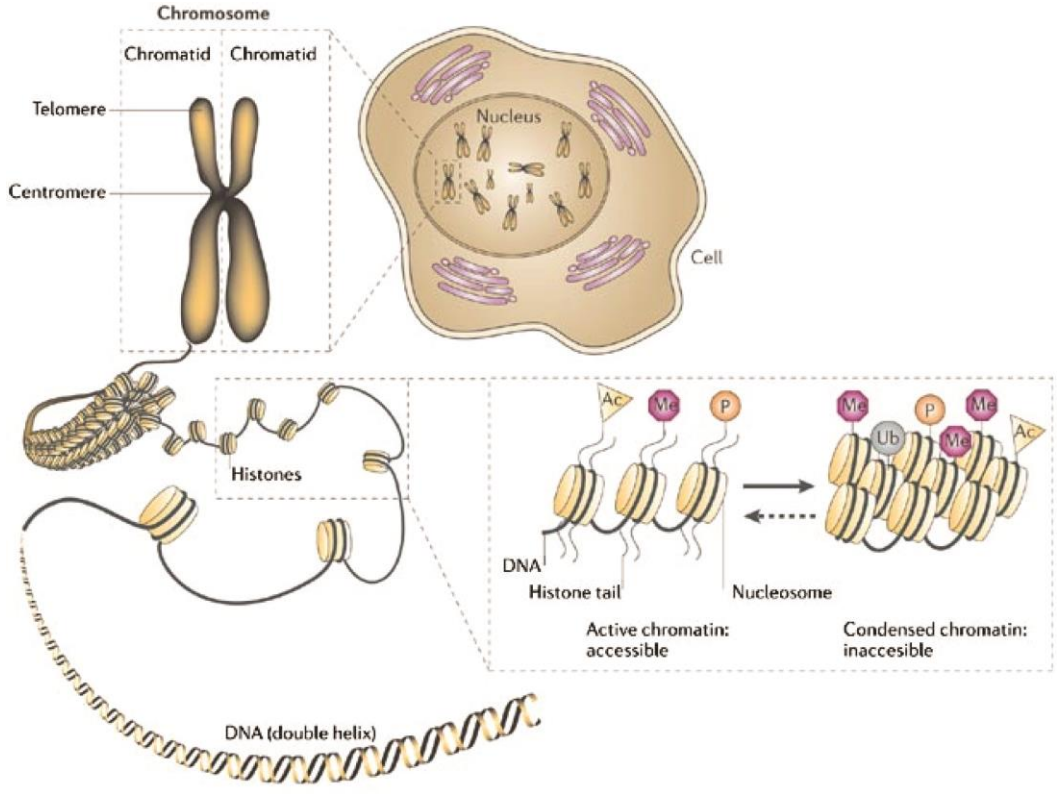


Figure 1.2: Chromatin compaction occurs at several levels. Figure reproduced with permission from Sparmann and van Lohuizen (2006). DNA is wrapped around nucleosome-forming histones. Nucleosomal DNA is further compacted by packing of nucleosomes. Further levels of higher order compaction result in highly compacted chromosomes. Histone modification can influence compaction.

Compaction can correlate with functionality. For example, heterochromatin is present at the centromeres and telomeres, which are gene-poor. Furthermore, there are regions on some chromosomes that are consistently densely packed. These regions are known as constitutive heterochromatin, and display heterochromatic packing in all cell types of a species. In contrast, regions which are heterochromatic in some cell types but not in

others are known as facultative heterochromatin. Euchromatin on the other hand, usually contains active genes or genes that are primed for activation.

Additionally, interactions between sequences, which are not close to each other in *cis*, in three-dimensional space can be important for the regulation of specific genes. Using DNA cross-linking approaches such as those based on 3C (Chromosome Conformation Capture) complemented by information from fluorescence *in situ* hybridisation (FISH), it has become clear that there are interaction domains, in which sequences interact more with other sequences within the domain, and less with those outside. Most interactions are intra-chromosomal, whilst it seems that the centromere forms some kind of barrier for intra-chromosomal interactions (reviewed in Bickmore and van Steensel, 2013). These domains often correlate well with blocks carrying similar histone modifications. Furthermore, these domains can correlate with replication timing domains (Ryba et al., 2010). Replication timing domains are domains that are replicated consistently at a particular point during S phase, measuring about 100 kb to several Mb in size. Typically, regions containing actively transcribed genes are replicated early, and repressed regions are replicated late (Hiratani et al., 2008; Schwaiger et al., 2009). In some cases, specific long-range interactions between enhancer or control elements and genes, or between transcription start and termination sites have been detected (Tan-Wong et al., 2008; Saramaki et al., 2009). It is thought that at least some of these long range interactions may be coordinated by insulator sequences and CTCF, currently the only protein known to bind to insulator sequences (reviewed in Bickmore and van Steensel, 2013).

Taken together, it is clear that more densely packed regions correlate with less gene expression. It remains unclear what the causative relationship between these observations is.

1.1.8 Heritability of chromatin marks

As discussed in 1.1.1, some phenomena have been classified as epigenetic because they were inherited across cell division in a way that seemed to deviate from classical Mendelian inheritance. For example, it has been shown that the DNA methylation status at the *Agouti* locus in mouse can be inherited and affect gene expression (Morgan et al., 1999). However, phenomena that are now routinely defined as epigenetic have often been shown not to be inherited. Whether chromatin marks can be inherited is a conceptually important question as it interrogates how cellular memory arises. More dramatically, it probes whether changes made in chromatin that could be adaptive to the environment can be passed on to offspring.

DNA methylation can be maintained by DNMT1, although not as faithfully as the DNA sequence itself. Here I will discuss whether histone marks can be inherited.

A few groups have tested whether histone marks can be inherited across mitosis. For example, it was found that artificial recruitment of HP1α resulted in a heterochromatic H3K9me3 enriched domain and gene silencing, and the heterochromatic domain was inherited after removal of the stimulus recruiting HP1α (Hathaway et al., 2012). A similar experiment showed that H3K27me3 could be maintained after artificial tethering

the EZH2 histone methyltransferase (see 1.3) and removing the tethering signal (Hansen et al., 2008). This shows that some histone marks can be maintained across cell division.

Exactly how histone mark inheritance occurs is unclear. One mechanism that has received some support is the reincorporation after DNA replication of histones at or near their original location and subsequent copying of histone marks onto neighbouring nucleosomes (Dodd et al., 2007; Francis et al., 2009). Another mechanism supported by some experimental evidence is the maintenance of localisation of writers or proteins instructively involved in chromatin regulation (Hansen et al., 2008; Rowbotham et al., 2011; Petruk et al., 2012; Steffen et al., 2013). These two mechanisms could singly or together provide the basis for a self-templating mechanism that would perpetuate the presence of histone marks without direction by the underlying DNA sequence. However, it is also conceivable that histone marks are reassembled after each cell division, instructed by the presence and absence of transcription factors and the DNA methylation system.

1.2 Embryonic stem cell biology

1.2.1 Establishing ES cells and towards mouse reverse genetics

Embryonic stem cells are pluripotent cells derived from the inner cell mass of an embryo. The study of embryonic stem cells began with the study of teratomas: a special type of neoplasm that contains differentiated cell types of all three germ layers as well as undifferentiated embryonic-like cells (Pierce, 1967; Stevens, 1967). Taking lessons from culturing teratomas cells (Evans, 1972), embryonic stem cells, derived from the inner cell

mass of blastocysts, were established in culture (Evans and Kaufman, 1981; Martin, 1981). Further progress was made when it was discovered that embryonic stem cells could be maintained in their pluripotent state robustly when grown in the presence of leukaemia inhibitory factor (LIF) (Smith et al., 1988; Smith et al., 1992), which helps maintain pluripotency by activation of the JAK/STAT3 pathway (Niwa et al., 1998). Research showed that ES cells can contribute to mouse development after blastocyst injection, and could contribute to the germ cell lineage {Bradley, 1984 #184}. Thus, it was also shown that genetic modifications in ES cells could be introduced into mice (Bradley et al., 1984; Robertson et al., 1986; Kuehn et al., 1987). With improvements made in gene-targeting to make specific mutations (Doetschman et al., 1987; Thomas and Capecchi, 1987) and the use of the bacteriophage enzyme Cre to make mutations conditional (Gu et al., 1994), ES cells and blastocyst injection present the opportunity to perform reverse genetics in mice (Figure 1.3). Additionally, ES cells provide an *in vitro* model system to understand in molecular detail the events occurring during early embryonic development.

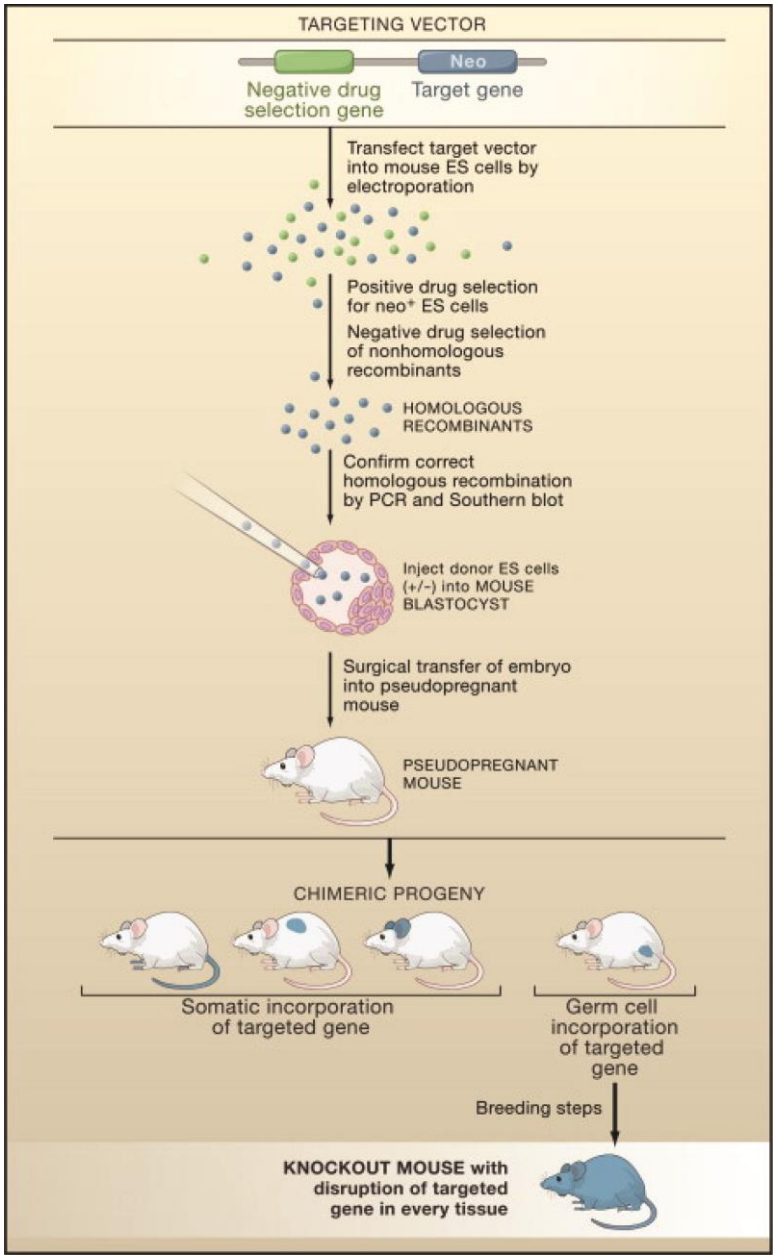


Figure 1.3: The creation of knockout mice using genetically modified embryonic stem cells. Reproduced, with permission, from Mak (2007).

1.2.2 The pluripotency core transcription factors and maintaining the pluripotent state

Once ES cells could be maintained in their pluripotent state in culture, the molecular factors controlling the ES cell state were established (reviewed in Young, 2011). From such studies, it became apparent that a trio of "core" pluripotency transcription factors,

OCT4, SOX2 and NANOG, play an important role. Importantly, these transcription factors positively regulate their own and each other's expression, so that the pluripotent state is highly stable, but affecting expression levels of one of these factors will stimulate differentiation (Boyer et al., 2005; Loh et al., 2006).

In addition to the three core transcription factors, other proteins contribute to the maintenance of pluripotency, such as REX1 and KLF4 (Thompson and Gudas, 2002; Jiang et al., 2008). Furthermore, CMYC has been shown to be important for regulation of the pluripotency network. This may be in part by controlling gene expression of lineage specific genes and pluripotency genes (Kim et al., 2008; Smith et al., 2010) but also through affecting proliferation (Rahl et al., 2010). The importance of these factors is illustrated by the fact that OCT4, SOX2, KLF4 and CMYC constitute the original transcription factor overexpression "cocktail" used to reprogram somatic cells into induced pluripotent stem cells, discussed further in 1.2.5 (Takahashi and Yamanaka, 2006).

Interestingly, it has been shown that some of these factors, including NANOG and REX1, are heterogeneously expressed within an ES cell population (Chambers et al., 2007; Toyooka et al., 2008). It has been suggested that this is due to ES cells being in a meta-stable state, where they are oscillating between a ground state of pluripotency, and a state in which they are primed to differentiate (Cahan and Daley, 2013). It was reported that this heterogeneity is not present when cells are cultured in the presence of two chemical inhibitors which block the Mek and GSK3 kinases ("2i" culturing conditions), rather than in the presence of LIF and serum (Marks et al., 2012).

1.2.3 Chromatin in ES cells

Many genes encoding chromatin modifying factors are direct targets of the core pluripotency factors, illustrating the importance of chromatin for pluripotency (reviewed in Shen et al., 2008; Orkin and Hochedlinger, 2011). These genes include proteins of the Polycomb and Trithorax groups and are discussed in more detail below. Additionally, it has been shown that chromatin modifying factors can influence the expression of pluripotency factors (Kidder et al., 2009). Typically, the chromatin state in ES cells is thought to be more "open" (reviewed in Fisher and Fisher, 2011). This hypothesis is supported by findings that there appear to be fewer heterochromatic regions in ES cells and that some structural chromatin proteins bind less tightly in ES cells (Meshorer et al., 2006; Bartova et al., 2008; Bhattacharya et al., 2009). It has been found that knockdown of the chromatin remodeller *Chd1* led to heterochromatin accumulation and a skewed differentiation, showing a functional role of chromatin regulation (Gaspar-Maia et al., 2009). Furthermore, levels of active histone marks are increased in ES cells, whilst levels of repressive marks are decreased (Meshorer et al., 2006; Efroni et al., 2008; Krejci et al., 2009; Wen et al., 2009). This more open chromatin state might explain why low-level stochastic transcription of repressed genes has been recorded in ES cells (Efroni et al., 2008).

1.2.4 Differentiating ES cells

ES cells are a model for understanding pluripotency, but they also provide the opportunity to understand how cell fate is established during differentiation. ES cells can be induced to differentiate into the three germ layers by the withdrawal of LIF and resultant embryo body formation (Doetschman et al., 1985; Abe et al., 1996; Leahy et al., 1999), by culturing ES cells as monolayers on extracellular matrix proteins (Nishikawa et

al., 1998) or by culturing them on a supportive layer of stromal cells (Nakano et al., 1994). Signalling molecules known to be important in the developing embryo usually affect ES cell differentiation (reviewed in Murry and Keller, 2008), suggesting that cell fate decisions occur in a similar manner. It is thought that cell fate decisions may be stabilised by chromatin modifying factors (Wutz, 2013). The ability of ES cells to differentiate makes them a useful tool for studying the role of chromatin modifying factors in differentiation.

1.2.5 Importance of the study of ES cells

The study of ES cells is very important for several reasons. Firstly, as described in 1.2.1, the study of ES cells led to the methodology that allows mutations to be introduced into mice and their phenotypes to be studied. As the mouse is an animal model that is closely related to humans, both genetically and physiologically, this has led the way for indirectly studying human disease.

Also, study of ES cell biology has led to insights into reprogramming of somatic cells into induced pluripotent stem cells (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007). Stem cells can be used to treat certain diseases (reviewed in Robinton and Daley, 2012; Ramsden et al., 2013) and it is hoped that one day iPS cells may be used for this purpose. Furthermore, deriving iPS cells from patients that have particular genetic defects presents the opportunity to research the consequences of these genetic defects and to screen drugs *in vitro*.

There are limitations to using mouse ES cells as a model for human ES cells and human cell biology, development and disease. For example, there are differences between human ES cells and mouse ES cells (Brons et al., 2007; Tesar et al., 2007) and there is variability between human ES lines, the cause of which is currently not well understood (Bock et al., 2011; reviewed in Cahan and Daley, 2013). Additionally, culturing conditions may affect experimental outcomes, as it has been suggested that cells grown in "2i" conditions represent a more naïve pluripotent state, whereas those grown in the more conventional serum + LIF condition are on the cusp of differentiation and are thought to be in the "primed" state.

Whilst the relation between mouse ES cells and human biology is imperfectly understood, research on mouse ES cells has provided and will continue to provide many insights and potential applications.

1.3 Polycomb group proteins

1.3.1 The discovery of Polycomb group proteins

In all metazoans, a cluster of genes called the *Hox* genes specify the anterior-posterior axis. In *Drosophila*, a segment-specific pattern of *Hox* gene expression is set up during early embryogenesis by the combinatorial localised expression of transcription factors (Akam, 1987).

Polycomb was identified as a *trans*-acting repressor of the *Hox* genes (Lewis, 1978). Whilst homozygous mutants are embryonic lethal, hypomorphic mutations or heterozygous mutants show posterior transformations: in the case of *Polycomb*, heterozygous males have additional sex combs on their second and third leg, whilst their wild type counterparts have sex combs only on the first leg (Figure 1.4). A mutation giving rise to this kind of phenotype, in which a part of the body is transformed to resemble another part, is called a homeotic mutation. Later, further genes were identified and classed into two antagonistic groups: the Polycomb Group (PcG) and Trithorax Group (TrxG). Studies led to the deduction that both groups were responsible for maintaining the correct pattern of *Hox* gene expression once the original transcription factors that directed the pattern are no longer present. This is therefore a prime example of cellular memory (as discussed in 1.1.8).

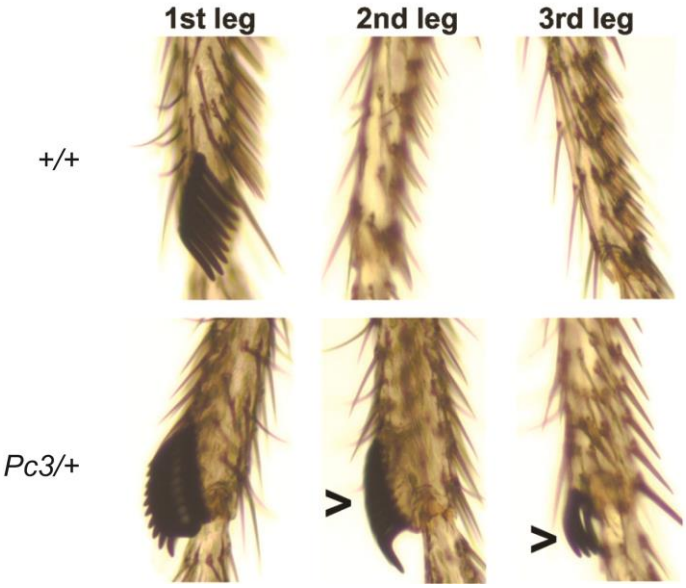


Figure 1.4: Heterozygous inheritance of the *Pc3* allele, which is a loss-of-function allele, results in extra sex combs on the second and third leg of males in *Drosophila*. Reproduced under creative commons licence CC BY-NC from Parrish et al. (2007).

Subsequently, homologues of *Drosophila* PcG genes were isolated in vertebrates. It was found that mutations in these genes caused homeotic transformations of the vertebrae, indicating that the function is conserved (van der Lugt et al., 1994; Coré et al., 1997).

1.3.2 Polycomb complexes in *Drosophila melanogaster* and mammals

Full name in *Drosophila melanogaster*	abbreviation	Homologue(s) in *Mus musculus*
PRC2		
suppressor of zeste 12 homologue	*Suz12*	*Suz12*
Extra Sex Combs	*Esc*	*Eed*
Extra Sex Combslike	*Escl*	-
Enhancer of zeste	*E(z)*	*Ezh1, Ezh2*
Nucleosome remodelling factor 55	*p55 (Caf-1 or Nurf55)*	*Rbbp4 (Rbap48), Rbbp7 (Rbap46)*
PRC1		
Ring1/Sex combs extra	*dRing1/Sce*	*Ring1A, Ring1B (Rnf2)*
Polyhomeotic proximal and Polyhomeotic distal	*Ph-p, Ph-d**	*Phc1 (Rae28), Phc2, Phc3*
*Posterior sex combs***	*Psc*	*Bmi1 (Pcgf4), Mel18 (Pcgf2), Nspc1 (Pcgf1), Pcgf3, Pcgf5, Pcgf6*
Polycomb	*Pc*	*Cbx2 (Pc1/M33), Cbx4 (Pc2), Cbx6, Cbx7, Cbx8 (Pc3)*

Table 2: Components and homologues of the PRC1 and PRC2 complexes. *The *Drosophila* polyhomeotic locus consists of a duplication of the polyhomeotic gene. It is thought the genes function in very similar ways (Dura et al., 1987). ** In addition, *Drosophila* contains the lesser known functional homologue of *Psc* called *Suppressor of zeste homologue 2 (Su(z)2)* (Lo et al., 2009).

After discovery of the PcG genes, subsequent biochemical analysis has identified two major polycomb complexes: Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2). An overview of the *Drosophila melanogaster* PRC1 and PRC2 components is presented in Table 2. Additional polycomb complexes have also been described in *Drosophila melanogaster* (Klymenko et al., 2006; Scheuermann et al., 2010).

1.3.2.1 The PRC2 complex

Drosophila PRC2 consists of four components: enhancer of zeste (E(z)), suppressor of zeste 12 homologue (Su(z)12), extra sex combs (Esc) and nucleosome remodelling factor 55 (p55 or Nurf55 or Caf-1) (Figure 1.5). This complex has not greatly diversified in mammals (Table 2).

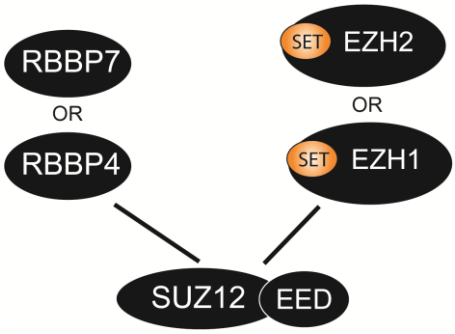


Figure 1.5: The PRC2 core complex in mammals. In addition to the different homologues for the E(z) component and the p55 component, further diversity is introduced by different translational isoforms of EED.

At the heart of the PRC2 complex lies the catalytic component enhancer of zeste, which corresponds to EZH1 or EZH2 in mouse. The catalytic activity is contained in the SET domain, which is responsible for H3K27 methylation. In addition EZH2 and EZH1 contain two SANT domains and one cysteine rich domain (CXC, not to be confused with CXXC). The function of these additional domains is not known, but it has been speculated that the CXC domain stabilises the SET domain, whilst the SANT domains may be involved in transmitting signals from associating proteins to the catalytic centre (Tschiersch et al., 1994; Boyer et al., 2004; Ciferri et al., 2012).

EZH1/2 are insufficient for H3K27 methylation activity on their own (Cao and Zhang, 2004; Pasini et al., 2004; Ketel et al., 2005; Nekrasov et al., 2005). PRC2 also contains

suppressor of zeste 12, SUZ12. SUZ12 has one zinc finger and a VEFS domain. It is thought that these domains bind to EZH1/2 (Yamamoto et al., 2004; Ketel et al., 2005), whilst the VEFS domain may also play a role in H3K4me3-mediated inhibition of PRC2 (see 1.3.5) (Schmitges et al., 2011).

In addition, PRC2 contains extra sex combs (Esc) or extra sex combs like (Escl), the mammalian orthologue of which is termed embryonic ectoderm development (EED). In *Drosophila*, *escl* seems to act redundantly with *esc*, and makes little contribution to total function both developmentally and enzymatically as judged by methyltransferase activity (Ohno et al., 2008). Esc and EED are largely comprised of 7 WD40 repeats that form a β propeller, which has been shown to have a preference for binding to the H3 tail when trimethylated at lysine 27 (Tie et al., 2007; Margueron et al., 2009). In mouse, there are four translational isoforms of EED, all with small truncations from the N-terminus of the protein. None of these truncations affect the WD40 domain, nor do the truncations affect the ability of EED to stimulate methyltransferase activity (Montgomery et al., 2007).

PRC2 also contains Nucleosome remodelling factor 55 (p55 or NURF55 or CAF1). The mammalian homologues of this protein are retinoblastoma binding proteins 4 and 7 (RBBP4 and RBBP7, also RbAp48 and RbAp46 respectively). RBBP4 and RBBP7 are not essential for methyltransferase activity but they do stimulate it. Like EED, they contain a β propeller consisting of WD40 repeats, which is thought to bind to a histone H3 tail. It should be noted that RBBP4 and RBBP7 can also participate in other chromatin modifying complexes, including the nucleosome remodelling and deacetylase (NuRD) complex (reviewed in Philpott et al., 2000). In contrast, as far as is known, the

other components are exclusive to PRC2. Recent research has reported that at least in the human transformed cell line HeLa, RBBP4/7 may be present in only 60% of PRC2 complexes, suggesting it may not be a core component (Smits et al., 2013).

It is clear that PRC2 components show a high degree of conservation between mice and humans, but also that they are well conserved even between *Drosophila* and mammals. Furthermore, homologues have been reported in the nematode *Caenorhabditis elegans* and also in the small flowering plant *Arabidopsis thaliana*, where, unlike in mammals, the PRC2 components have diversified and there are many homologues.

In addition to the core subunits, a number of other proteins have been found to associate with PRC2. These are discussed in 1.3.4.

1.3.2.2 The canonical PRC1 complex and non-canonical PRC1 complexes

Drosophila PRC1 consists of Polycomb (Pc), Polyhomeotic (Ph), Posterior Sex Combs (Psc) and Sex Combs Extra (Sce)/dRing. There are many homologues present in mammals (Table 2), whilst only a few poorly characterised homologues have been reported in *C. elegans* and *A. thaliana* (Karakuzu et al., 2009; Molitor and Shen, 2013).

The catalytic core of PRC1 is formed by Ring1 (RING1A or RING1B in mammals) and Psc (PCGF homologues in mammals). Both proteins contain a RING-finger domain. RING1B acts as an E3 ubiquitin ligase for lysine 119 on histone H2A by interacting with the E2 Ubiquitin-conjugase enzyme (Wang et al., 2004a; Cao et al., 2005), but in

addition the PCGF proteins are required for full activity and specificity (Cao et al., 2005; Elderkin et al., 2007). *Drosophila* Psc and the homologous mammalian CBX proteins contain a chromodomain which can bind to H3K27me3 and is believed to recruit PRC1 to locations already marked by PRC2 (Fischle et al., 2003; Min et al., 2003) (see 1.3.5). The functions of the other PRC1 proteins remain unclear at this time.

Recently it has been found that instead of one canonical complex, there appear to be many variants of the PRC1 complex in mammals with extra subunits found (reviewed in Luis et al., 2012). Importantly, they all contain the enzymatic subunit RING1A/B (Figure 1.6) and can associate with either RYBP or YAF2. Gao and colleagues divided the PRC1 subclasses based on which PCGF (*Drosophila* Psc) variant they contain (Gao et al., 2012). In addition, it has been shown that the presence of CBXs (*Drosophila* Pc) and RYBP or YAF2 is mutually exclusive (Tavares et al., 2012) in the "PRC1.2" and "PRC1.4" complexes. Within these two complexes, further diversity is introduced by the presence of CBX variants, which may have different genomic localisations, lineage expression patterns and interaction partners (Morey et al., 2012; O'Loghlen et al., 2012; Morey et al., 2013). Some of the newly discovered non-canonical PRC1 components have been shown to have important roles in development (Pirity et al., 2005; Qin et al., 2012).

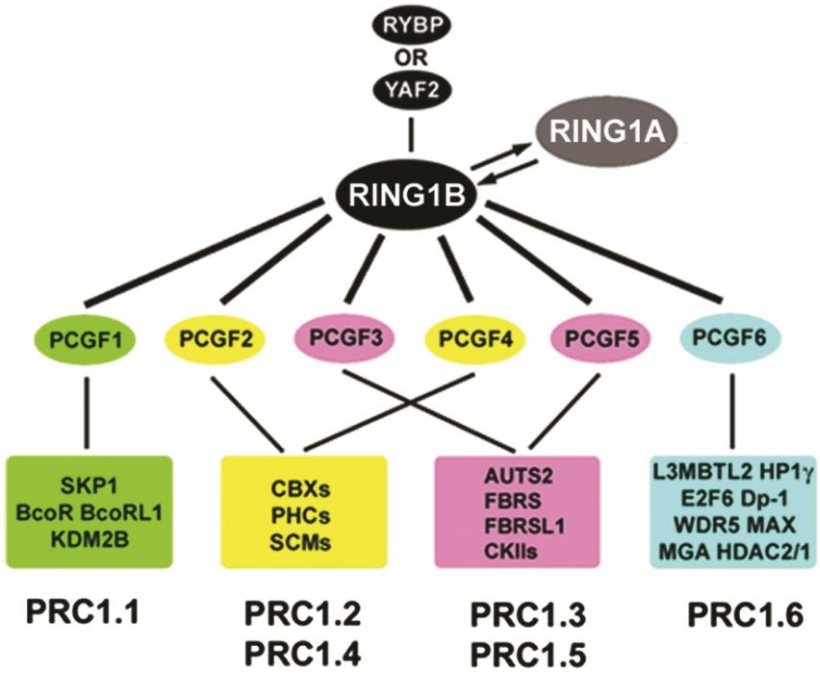


Figure 1.6: Variant PRC1 complexes. Every PRC1 complex contains RING1A or RING1B. PRC1.2 and PRC1.4 represent the canonical PRC1 complex. Figure reproduced with permission from Gao et al. (2012).

1.3.2.3 Further PcG complexes

In addition to PRC1 and PRC2, two further PcG complexes have been described in *Drosophila*: PR-DUB and Pho-RC. Possible roles of these complexes are discussed below.

1.3.3 Roles of PcG complexes

PcG proteins were discovered as repressors of the *Hox* genes. The identification of mouse homologues, together with the development of gene targeting, allowed researchers to assess the effects of genetic removal of PcG proteins in mice. Typically, it was found that homozygously deleting PRC2 components resulted in embryonic lethality at around embryonic day (E) 7.5 (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004), whereas mice homozygous knockout for several of the canonical PRC1 complex

survived to term with homeotic transformations and growth retardation. PRC1 knockout mice were either viable or died in the first 4 weeks after birth (van der Lugt et al., 1994; Akasaka et al., 1996; Coré et al., 1997; Takihara et al., 1997; del Mar Lorente et al., 2000). The exception to this was the *Ring1b* knockout, which led to embryonic lethality at E10.5 (Voncken et al., 2003). Relative viability of PRC1 knockout mice could reflect the fact that most PRC1 components have many homologues, whereas PRC2 components do not (Akasaka et al., 1996). In PRC2 mutants, embryonic lethality was due to a failure to complete gastrulation (discussed further in Chapter 7). Hypomorphic mutations led to posterior transformations (Faust et al., 1995; Suzuki et al., 2002).

Chromatin immunoprecipitation studies showed that PRC2 and PRC1, and H3K27me3 and H2AK119Ub, were present not only at the *Hox* genes, but also at a large number of other developmentally important genes in *Drosophila* and in mammalian ES cells (Boyer et al., 2006; Lee et al., 2006b; Schwartz et al., 2006; Ku et al., 2008; Kallin et al., 2009; Endoh et al., 2012). These observations establish an important role for PcG complexes during development. However, how this role is fulfilled remains an area of active research.

1.3.3.1 Role of PRC2 in gene repression

It was found that the *in vitro* reconstituted PRC2 complex can methylate histone H3 (Rea et al., 2000; Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002). It is thought that H3K27me3 is the functionally most important output for PRC2 because it is this mark that co-localises with the PRC2 components. H3K27me3 is present at 15% of nucleosomes (as opposed to H3K27me2, which is present at 50% of nucleosomes in ES cells) (Sarma et al., 2008), and these locations typically correlate to

locations of genes repressed by PcG proteins. It is believed that PRC2 complexes containing EZH2, rather than EZH1, are responsible for bulk H3K27me3, as they are more enzymatically active (Margueron et al., 2008; Shen et al., 2008). That PRC2 mediates H3K27 methylation *in vivo* is evidenced by the observation that H3K27me3 and H3K27me2 are undetectable in PRC2 mutant cells (Montgomery et al., 2005; Pasini et al., 2007; Chamberlain et al., 2008; Shen et al., 2008). Early reports suggested that H3K27me1 may persist in PRC2 mutants. However, later it was found that some antibodies against H3K27me1 also recognize H3K9me1. Importantly, H3K27me1 is undetectable in a cell line with *Ezh2* knockout combined with *Ezh1* knockdown, suggesting that these are the only histone methyltransferases for H3K27 (Shen et al., 2008).

Due to these observations, it has long been hypothesized that H3K27me3 is required for PRC2 function and gene repression. This hypothesis is supported by recent data from Müller and colleagues (Pengelly et al., 2013). In *Drosophila*, there are multiple copies of the histone genes present in 23 gene clusters. The authors deleted all of the 23 histone gene clusters and replaced them with 12 histone clusters, which were intact except for the mutation of H3K27 into an arginine, which cannot be methylated by PRC2. It was found that this genetic modification replicated the phenotype of PRC2 mutants, suggesting that methylation of H3K27 is the crucial functional output of PRC2 (Pengelly et al., 2013).

Exactly how H3K27me3 mediates gene silencing is unclear. It has been suggested that one of the functions may be the recruitment of PRC1 by generating a binding site for CBX-containing PRC1 (see 1.3.5), but recent reports have argued against the idea that

CBX-mediated recognition of H3K27me3 is essential for PRC1 recruitment (Tavares et al., 2012; Morey et al., 2013). It is possible that recognition of the H3K27me3 mark could recruit other silencing proteins but so far none have been reported. It could also be possible that the biophysical properties of octamers modified by H3K27me3 are altered so as to reduce transcription, for example, by increasing inter-nucleosomal packing or by increasing the interactions between DNA and histones. However, no evidence for such a mechanism has been found.

Another possibility is that the presence of H3K27me3 may block the recruitment of transcriptional activators or the general transcription machinery. One study found that in PRC2 mutant cells in *Drosophila* embryos, 25% of genes showed increased expression and concomitant increased levels of 5' paused RNA polymerase II (Chopra et al., 2011). This could suggest that H3K27me3 may inhibit recruitment of RNA polymerase II. However, some paused polymerase II, marked by phosphorylation on serine 5 on the heptad repeat of the C-Terminal Domain, could also be detected at many genes in wild type embryos, consistent with other studies (Stock et al., 2007; Min et al., 2011; Brookes et al., 2012) and with the detection of low transcript levels of PcG target genes (Zhao et al., 2007; Chopra et al., 2009; Kanhere et al., 2010), suggesting that H3K27me3 may instead act by inhibiting elongation. Given that H3K27me3 can be enriched over very large domains, including enhancers, promoters and gene bodies, it is possible that it may act at multiple points in the transcription cycle (Simon and Kingston, 2013).

It has also been suggested that PcG bound sequences may cluster in 3D space and form "polycomb bodies" (Tolhuis et al., 2011; reviewed in Pirrotta and Li, 2012). How these

bodies are formed is unclear but may depend on insulator sequences (Li et al., 2011a; Li et al., 2013). These nuclear bodies are thought to contribute to silencing.

Additionally, it has been suggested that PRC2 can silence genes in an H3K27me3-independent way. It has been shown that PRC2 can also methylate H1K26, as well as GATA4 and JARID2 (Kuzmichev et al., 2004; He et al., 2012; R. Margueron, personal communication) which may contribute to PRC2 silencing. Furthermore, it was found that EZH1-PRC2, but not EZH2-PRC2, could compact chromatin in a methylation-independent way and repressed *in vitro* transcription (Margueron et al., 2008). How this effect was mediated remains unclear, though the authors did find repression only occurred prior to the formation of the pre-initiation complex and required the presence of histone tails. Also, it has been shown that PRC2 can target RBP2 H3K4 demethylase (also known as JARID1A) and thus remove activating H3K4 methylation (Pasini et al., 2008).

Taken together, it seems likely that PRC2 represses transcription largely through the deposition of H3K27me3. However, how this mark leads to repression remains unclear and may involve several mechanisms.

1.3.3.2 Role of PRC1 in gene repression

RING1A and RING1B, and the *Drosophila* homologue dRing/Sce, are E3 mono-ubiquitin ligases (de Napoles et al., 2004; Wang et al., 2004a; Cao et al., 2005). As such they provide the specificity with which an E2 ubiquitin-conjugating enzyme conjugates the ubiquitin group. It has been shown that association with the Psc homologue BMI-1

(Wang et al., 2004a; Cao et al., 2005) or MEL18 (Elderkin et al., 2007) greatly enhances the ubiquitin ligase activity of RING1B *in vitro*.

The study of the importance of the ubiquilylating activity of RING1A and RING1B has been complicated by two factors. Firstly, some studies have studied the removal of RING1B without assessing what the compensatory role of RING1A might be. Secondly, other studies have used catalytic mutants (*Ring1b*^{I53S} and *Ring1b*^{I53A}) of which it is not completely clear to which extent they remove catalytic activity. Mutations at this site had previously been shown to almost fully deplete ubiquitin ligase activity by interfering with the binding of the E2 enzyme UbcH5c (Ben-Saadon et al., 2006; Buchwald et al., 2006). However, a later study with the structure of UbcH5c in complex with RING1B suggested that contacts between the interfaces involve more amino acids in addition to isoleucine-53 of RING1B so that this mutation would be expected to not be entirely catalytically inactive.

In the absence of *Ring1b*, numerous PcG target genes become derepressed (Leeb and Wutz, 2007; van der Stoop et al., 2008). Additionally, other members of PRC1, including RYBP, MPH1/2, MEL18 and BMI1 complex become destabilised, whilst the PRC2 complex is not affected (Leeb and Wutz, 2007). However, not all PRC1-target genes become derepressed in a *Ring1b* knockout cell line (van der Stoop et al., 2008). *Ring1b/Ring1a* double knockout leads to greater derepression of PRC1-target genes and these cells eventually cease to proliferate (Endoh et al., 2008). In a later study performed by Endoh and colleagues, cells were null for *Ring1a* and *Ring1b* and then either rescued with wild type *Ring1b* or the mutated version (Endoh et al., 2012). In the absence of

Ring1a and *Ring1b*, many genes became derepressed. The authors found that *Ring1b*^{I53S} and *Ring1b*^{I53A} rescue could not completely restore gene repression, where the wild type version of *Ring1b* could. However, these genes were more repressed than in the complete absence of *Ring1b*, suggesting that *Ring1b* with severely depleted catalytic activity can still contribute to repression. The authors found that some genes seemed more dependent on the catalytic activity than others.

These findings were consistent with those by Eskeland and colleagues, who found that a catalytically compromised RING1B could compact the *Hoxb* locus and thus maintain gene repression (Eskeland et al., 2010). It was previously described that PRC1 can compact nucleosomal arrays *in vitro*, independent of histone tails (Francis et al., 2004). However, due to the caveats regarding the compensatory role of RING1A and the exact effect of the RING1B mutation, data from these studies should be interpreted with caution. It could be that a small amount of H2AK119ub would be sufficient to initiate repression, and this repression might be stabilised by other silencers.

In a study performed by Gutierrez and colleagues, the entire *dRing/Sce* locus was removed in *Drosophila melanogaster* and it was found that whilst H2AK119ub was lost, the majority of target genes remained repressed (Gutierrez et al., 2012). Furthermore, embryos that show lack of both Psc and its paralogue Su(z)2 display more severe defects in development than those with genetic removal of *sce* (Gutierrez et al., 2012). This shows that PRC1 components can act in an H2AK119ub-independent manner in *Drosophila*.

Together, these studies suggest that PRC1 has both H2AK119ub-dependent and independent mechanisms of silencing. How each mechanism achieves silencing is not fully elucidated. The H2AK119ub-independent mechanism may depend at least in part on compacting activity.

One study that investigated how H2AK119ub might affect silencing reported that deletion of *Ring1a* and *Ring1b* is associated with a release of paused polymerase (Stock et al., 2007). These findings are consistent with the report that the presence of PRC1 does not preclude the presence of general transcription factors (Breiling et al., 2001), whilst more recent work suggests that genes bound by PRC1 and PRC2 are less likely to contain productively elongating polymerase as compared to those bound only by PRC2 (Min et al., 2011). The presence of H2AK119ub has been linked to inhibit elongation by preventing binding of the elongation factor FACT, although ubiquitination in this study was mediated by Dzip3 and not PRC1 (Zhou et al., 2008).

Other studies did not dissect whether silencing was H2AK119ub dependent or independent, but nonetheless provide insight into how PRC1 might silence genes. It has been reported that the presence of PRC1 prevents nucleosome remodelling by SWI/SNF in *Drosophila* (Shao et al., 1999) and nucleosome turnover is lower at PcG target sites (Deal et al., 2010). Another model suggests that PRC1 may wrap DNA around itself which could affect silencing (Mohd-Sarip et al., 2006). It has furthermore been suggested that the presence of PRC1 may block association of the transcriptional activator Mediator

(Lehmann et al., 2012). In summary, PRC1 may silence genes by altering the chromatin state or by interacting directly with the transcription machinery.

1.3.3.3 PR-DUB

In *Drosophila*, an additional complex has been described: Polycomb repressive deubiquitinase (PR-DUB) (Scheuermann et al., 2010). PR-DUB comprises the deubiquitinase Calypso (dBap1) and Additional sex combs (Asx). This complex contains deubiquitinase activity against ubiquitinated H2A, as does the homologous mammalian complex, consisting of BAP1 and ASXL1. *Drosophila* PR-DUB is required for the silencing of a subset of genes that also require Sce for silencing (Milne et al., 1999; Gaytan de Ayala Alonso et al., 2007; Scheuermann et al., 2010; Scheuermann et al., 2012). Why a deubiquitinase, as well as an ubiquitin ligase, are required for silencing is unknown. The presence of further H2AK119ub deubiquitinases has also been reported (Weake and Workman, 2008; Atanassov et al., 2011).

1.3.3.4 Pho-RC

Another complex identified in *Drosophila* is the Pho Repressive Complex (Pho-RC). This complex consists of Pho and dSfmbt (Klymenko et al., 2006). Pho is a DNA sequence-specific binding protein and Pho binding sequences can be found at many PcG target sites in *Drosophila* (Brown et al., 1998). Both Pho and dSfmbt are important for *Hox* silencing and the Pho-RC has been proposed to be important for PcG recruitment (see 1.3.5). However, it seems that the mammalian Pho homologue YY1 does not play this role, as evidenced by limited overlap with PcG binding sites and the fact that many PcG enriched sites do not contain YY1 binding sites (Squazzo et al., 2006; Margueron and Reinberg, 2011).

1.3.3.5 PcG complexes are important for lineage commitment

Genetic removal of PcG components has established a role for PcG complexes in gene repression. In ES cells, PcG complexes bind to many lineage specific developmental regulators (Boyer et al., 2006; Lee et al., 2006b; Ku et al., 2008), some of which become upregulated upon PcG depletion (Breiling et al., 2001; Azuara et al., 2006; Boyer et al., 2006; Jorgensen et al., 2006; Leeb and Wutz, 2007; Pasini et al., 2007; Chamberlain et al., 2008; Shen et al., 2008). Perhaps surprisingly, ES cells are viable in the absence of PRC2. However, defects are seen when cells are forced to differentiate and these defects are very much exacerbated in cells depleted for both PRC1 and PRC2 by genetic removal of *Eed* and *Ring1b* (Leeb et al., 2010). Levels of PRC2 and PRC1 components decrease during differentiation *in vitro* and *in vivo* (Mak et al., 2002; Plath et al., 2003; Silva et al., 2003). The defects in differentiating ES cells fit with the phenotypes seen for complete removal of PcG components: development is severely compromised at the point at which epiblast cells (the epithelial tissue that develops from the inner cell mass) attempt to differentiate into the three germ layers. It may also explain why mutations in PcG components have been implicated in many cancers, as cancer cells undergo changes in cell identity (reviewed in Sparmann and van Lohuizen, 2006). Similarly, it has also been shown that the PRC2 complex is required for reprogramming (Pereira et al., 2010). Therefore, polycomb complexes are important for the cell fate transitions during differentiation and reprogramming.

1.3.4 PRC2 associating factors

Over recent years, it has become clear that in addition to the core components of PRC2, there are several substoichiometric subunits that can also associate with PRC2 (Cao et al., 2002; Cao and Zhang, 2004; Cao et al., 2008; Sarma et al., 2008; Peng et al., 2009; Shen

et al., 2009; Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010; Walker et al., 2010; Casanova et al., 2011; Zhang et al., 2011; Ballaré et al., 2012; Brien et al., 2012; Hunkapiller et al., 2012; Cai et al., 2013). An overview of these subunits is shown in Figure 1.7.

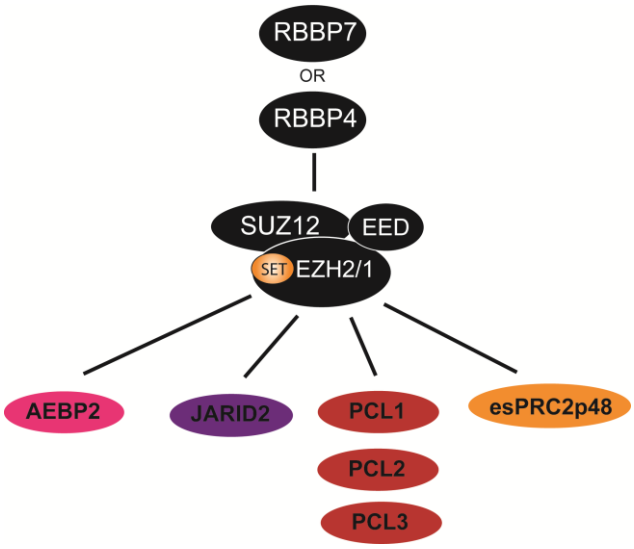


Figure 1.7: Associating factors of the PRC2 complex in mouse.

AEBP2 (AE1 binding protein 2), the subject of this thesis, is one of these associating factors. It has been proposed that AEBP2 may enhance H3K27me3 deposition (Cao and Zhang, 2004), or could act as a PRC2 targeting factor (Kim et al., 2009a). AEBP2 will be discussed in more detail in section 1.4.

Another PRC2 associating subunit is JARID2 (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010). JARID2 contains an N-terminal transrepression domain, which was shown to be required for binding to PRC2 (Pasini et al., 2010), whilst others posited that a more C-terminally located region was required for binding to SUZ12 in *in vitro* immunoprecipitations (Peng et al., 2009). JARID2 also

contains a jumonji C domain implicated in lysine demethylation, but it lacks the amino acids absolutely required for this activity and is therefore thought to be catalytically inactive (Takeuchi et al., 2006). It is likely that this domain nonetheless contributes in some other way to JARID2 function as the amino acid sequence is highly conserved in vertebrates (Landeira and Fisher, 2011). Additionally, it contains a jumonji N domain, an AT-rich interaction domain (ARID) and a zinc finger. Some of these may contribute to DNA binding, but no specific sequence has been reported (Shen et al., 2009; Li et al., 2010). This is interesting because none of the core PRC2 components have been reported to bind DNA.

It was shown that the recruitment of JARID2 and PRC2 were interdependent. In the absence of JARID2, PcG genes were typically not misexpressed in ES cells, but were poorly induced during differentiation. The effect on H3K27me3 is still under debate with some groups reporting an increase or no change (Peng et al., 2009; Shen et al., 2009), and others a decrease (Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010) in the absence of JARID2. Interestingly, it has also been shown that *Jarid2*$^{-/-}$ cells have lower levels of paused polymerase, suggesting that JARID2 may be important for establishing the "poised" state (Landeira et al., 2010) (see 1.3.6). The role of JARID2 in ES cells is discussed in more detail in Chapter 7.

Jarid2 mutant embryos die between E10.5-E15.5 (Takeuchi et al., 1995) and JARID2 has been previously implicated in heart development, haematopoiesis and cell cycle regulation (Kitajima et al., 1999; Ohno et al., 2004; Jung et al., 2005). It has also been reported that JARID2 can bind to several other proteins (Kim et al., 2004; Kim et al.,

2005; Mysliwiec et al., 2007; Shirato et al., 2009), but whether these activities occur in the context of the PRC2 complex or are PRC2-independent remains unknown. There is a *Drosophila* homologue of JARID2, which has been reported to interact with the PRC2 complex, but has also been described to have PRC2-independent roles (Herz et al., 2012a).

It has also been found that *Drosophila* polycomblike and related mammalian proteins, PCL1/PHF1, PCL2/MTF2, PCL3/PHF19, can associate with PRC2 and stimulate H3K27me3 deposition (O'Connell et al., 2001; Nekrasov et al., 2007; Tie et al., 2007; Cao et al., 2008; Sarma et al., 2008; Savla et al., 2008; Casanova et al., 2011; Ballaré et al., 2012; Brien et al., 2012; Hunkapiller et al., 2012; Cai et al., 2013). Some studies report that PCL proteins may be especially important for the conversion of H3K27me2 into H3K27me3 (Nekrasov et al., 2007; Sarma et al., 2008). It has also been suggested that the PCL proteins may contribute to PRC2 recruitment, though this effect is not as pronounced as that of JARID2 (Li et al., 2010; Casanova et al., 2011).

The PCL proteins contain plant homeodomain (PHD) fingers known to mediate the interaction with the PRC2 complex (Casanova et al., 2011; Abed and Jones, 2012). Additionally, they contain a Tudor domain. A series of papers recently reported that this domain binds to H3K36me3 (Ballaré et al., 2012; Brien et al., 2012; Musselman et al., 2012; Cai et al., 2013). This was surprising as it had previously been shown that the presence of H3K36me3 antagonises PRC2 activity (Schmitges et al., 2011; Yuan et al., 2011) and H3K36me3 and H3K27me3 do not typically colocalise. A possible explanation for this observation is that the PCL proteins may be involved in switching

the state of a gene from actively transcribed (and therefore marked by H3K36me3) to PRC2 repressed. This explanation is supported by the observation that the H3K36me demethylase NO66 is recruited by PCL3 (Brien et al., 2012). However, Musselman and colleagues found that PCL1 binding to H3K36me3 inhibited PRC2 activity, which would suggest that PCL1 may function to prevent H3K27me3 accumulation in active regions. Further research is required to probe the role of H3K36me3 recognition by PCL Tudor domains, and whether it may differ for different PCL members. Interestingly, *Drosophila* polycomblike Tudor domain does not bind to any methylated histones (Friberg et al., 2010), suggesting that this function may be specific to vertebrates.

Lastly, the protein C17orf96, dubbed esPRC2p48 and encoded by the *E130012A19Rik* gene, has been reported to associate with the PRC2 complex and be important in PRC2 function and neuronal differentiation (Shen et al., 2009; Zhang et al., 2011; Hunkapiller et al., 2012; De Cegli et al., 2013; Smits et al., 2013). This protein contains no domains to which a putative function can be described and its role in PRC2 biology remains largely uncharacterised.

1.3.5 Targeting polycomb complexes and their activities

PcG complexes, and the marks they deposit, are found at the *Hox* genes, but also at a large number of genes that encode for developmental regulators and genes expressed during development, in both mammals and *Drosophila* (Azuara et al., 2006; Bernstein et al., 2006; Boyer et al., 2006; Lee et al., 2006b; Schwartz et al., 2006; Mikkelsen et al., 2007). PcG complexes are present at at least 5-10% of genes in ES cells (Mohn et al., 2008).

1.3.5.1 PcG recruitment in *Drosophila melanogaster*

In *Drosophila*, sites at *Hox* genes to which PcG complexes are recruited, were termed Polycomb Responsive Elements (PREs). These sequences are usually located a few kilobases away from the promoter of the gene they regulate and can ectopically recruit PcG complexes if introduced to an exogenous site. The PRE concept has also been extended to other PcG target loci in *Drosophila*. Inspecting the collection of PREs, it is apparent there is no simple consensus sequence that constitutes a PRE, and the PRC2 core components have not been shown to contain sequence-specific DNA binding activity. However, it is thought that the PcG complexes are recruited by sequence-specific factors, most notably Pho and the pholike protein Phol (see also 1.3.3.4) (Brown et al., 1998; Fritsch et al., 1999; Brown et al., 2003; Oktaba et al., 2008; reviewed in Schuettengruber and Cavalli, 2009). It has been proposed that PRC2 is the initiator of PcG mediated silencing, and recruits canonical PRC1 by CBX-mediated binding to the H3K27me3 mark (Fischle et al., 2003; Min et al., 2003; Wang et al., 2004b). However, current evidence renders it unlikely this is the only method of PRC1 targeting (Simon and Kingston, 2009, see also below), and it has been suggested H3K27me3 binding may simply be a mechanism to stabilise already recruited PRC1 onto chromatin (Papp and Muller, 2006). PRC1 recruitment appears to be partially dependent on PRC2 activity in mammals too, whilst there is also evidence of PRC1 stabilising the presence of PRC2 (Leeb and Wutz, 2007; Endoh et al., 2008; Tavares et al., 2012), although the mechanism by which this occurs is unknown.

1.3.5.2 PcG recruitment in mammals by CpG islands

The sequence-specific factors implicated in recruitment of PcG proteins in *Drosophila* are typically not conserved in mammals (with the exception of Pho, see below), raising the question how PcG complexes are recruited in mammals. A few mammalian PREs

have been reported, but how they function remains unclear (Sing et al., 2009; Woo et al., 2010; Cuddapah et al., 2012; Woo et al., 2013). It has been observed that PcG recruitment co-incides with CpG islands (Bernstein et al., 2006; Boyer et al., 2006; Mikkelsen et al., 2007) and that these CpG islands typically lack sequences that recruit activators which are present in ES cells, though they may contain sequences that recruit activators in differentiated cell types (Ku et al., 2008). It has also been found that these regions are typically unmethylated (Mohn et al., 2008). Supporting an instructive role for CpG islands in PcG recruitment, it has been reported that even CpG-rich DNA from bacteria, depleted of possible vertebrate activating sequences, could recruit PRC2 when integrated into the ES cell genome (Mendenhall et al., 2010). Similarly, a human CpG island at the α globin locus that was enriched for PcG, could recruit PcG proteins in mouse ES cells when integrated into mouse locus (Lynch et al., 2012). Also, when the DNA methyltransferases are depleted from ES cells, PcG proteins were found to redistribute, suggesting that the absence of DNA methylation created new binding sites for PcG complexes (Lynch et al., 2012). A rationale for how PRC1 may be recruited to unmethylated CpG islands comes from the finding that the non-canonical PRC1 component KDM2B contains a CXXC domain and can recruit RING1B (Farcas et al., 2012; He et al., 2013; Wu et al., 2013). This also illustrates why PRC1 is unlikely to be entirely dependent on PRC2, as illustrated by PRC2/H3K27me3-independent recruitment of PRC1 in a number of instances (Schoeftner et al., 2006; Nekrasov et al., 2007; Pasini et al., 2007; Tavares et al., 2012). Specifically, RYBP-containing PRC1 complexes were independent of PRC2, whereas CBX-containing PRC1 complexes showed dependency on PRC2, explaining why the presence of RING1B is only partially depleted in the absence of PRC2 (Tavares et al., 2012; Morey et al., 2013).

Although CpG islands appear to be important in PcG recruitment, PcG complexes are not found at all CpG islands. They may be specifically excluded from CpG islands when they are the site of active transcription or contain sequences that recruit activators. An example through which this antagonism might occur was illustrated by the finding that the presence of H3K36me3 and H3K4me3 could inhibit PRC2-mediated H3K27me3 deposition onto the same histone tail through interactions with RBBP4/7 and the C-terminus of SUZ12, including the VEFS domain (Schmitges et al., 2011; Yuan et al., 2011). However, as described in 1.3.4, sometimes the presence of H3K36me3 may recruit PRC2 via the PCL subunit. Interestingly, the COMPASS component CFP1 may be excluded from CpG islands harbouring PcG proteins, suggesting that the antagonism might work in both directions (Vernimmen et al., 2011). Furthermore, recruitment could also be regulated by the level of CpG methylation, which explains why the methylcytosine dioxygenase TET1 is required for full recruitment of PRC2 (Wu et al., 2011).

There are further examples of chromatin marks that could instruct PRC2 recruitment to specific islands. As discussed, the PRC2 subunit EED contains a β propeller which can bind H3K27me3 (Tie et al., 2007; Margueron et al., 2009). This binding might stabilise PRC2 onto chromatin and the presence of H3K27me3 has been shown to stimulate EZH2's enzymatic activity. Additionally it has been suggested that the NuRD complex might aid PRC2 recruitment by deacetylating H3K27ac (Reynolds et al., 2012). Taken together, it appears that PRC2 targeting to specific CpG islands might occur, at least in part, by integrating cues from the chromatin environment. However, other mechanisms of PcG recruitment have been proposed. These mechanisms may explain in part why the

correlation between CpG islands and H3K27me3 appears weaker in differentiated cells (Orlando et al., 2012; Voigt et al., 2013).

1.3.5.3 PRC2 recruitment in mammals by associating factors

Some of the substoichiometric proteins that associate with PRC2 have been proposed to have a recruitment function. For the PCL proteins, this function may lie in the interaction with chromatin mediated by their tudor domain. It is evident that JARID2 is important for PcG recruitment, but how it achieves this remains unclear. JARID2's proposed DNA binding activity may contribute (Li et al., 2010), but more experimental evidence is required to support or refute this hypothesis. Furthermore, the PRC2 associating factor AEBP2 has been proposed to act as a recruiter, which will be discussed in 1.4 (Kim et al., 2009a).

1.3.5.4 PcG recruitment in mammals by sequence-specific factors

A few mammalian proteins have been suggested to act as direct recruiters of PcG complexes. One of these is YY1. As already mentioned, ChIP-seq studies have indicated that YY1-based recruitment is unlikely to underlie all PcG recruitment, but nevertheless the YY1 binding sequence has been found in some of the proposed vertebrate PREs and contibuted to PcG enrichment (Sing et al., 2009; Woo et al., 2010, 2013). Knockdown of *Yy1* has also been reported to reduce PcG recruitment at specific targets (Caretti et al., 2004).

Other transcription factors implicated in PcG recruitment are RE-1 Silencing Transcription Factor (REST) (Ren and Kerppola, 2011; Dietrich et al., 2012; Arnold et al., 2013), RUNX1/CBFβ (Yu et al., 2012), Cux1 (Ku et al., 2008) and SNAIL (Arnold et

al., 2013). However, like YY1, their binding sites overlap only with a subset of PcG target sites. Therefore, sequence-specific binding factors may contribute to PcG recruitment in at least some cases by directly binding to PcG proteins. However, it is also possible they contribute to recruitment by setting up a permissible chromatin environment that facilitates recruitment (see 1.3.5.6).

1.3.5.5 PcG recruitment in mammals by long noncoding RNAs

Several studies have suggested that long noncoding RNAs (lncRNAs) could play a role in recruiting PcG complexes. One example is the human HOTAIR lncRNA, which is transcribed from the *HOXC* locus and proposed to aid PcG recruitment at the *HOXD* locus (Rinn et al., 2007). However, the importance of this RNA is uncertain as HOTAIR-dependent repression by PcG proteins is not observed in mouse (Schorderet and Duboule, 2011). Nonetheless, many lncRNAs have been implicated in *trans* regulation of proteins involved in pluripotency based on knockdown studies in ES cells (Guttman et al., 2011; Ng et al., 2012).

Other examples of lncRNA based recruitment are thought to operate in *cis*. PcG recruitment to the inactive X (discussed in 1.3.8) requires the expression of the lncRNA *Xist* (Plath et al., 2003; Silva et al., 2003; Schoeftner et al., 2006). It has been suggested that *Xist* can directly recruit PcG complexes (Zhao et al., 2008; Kaneko et al., 2010; Zhao et al., 2010), although this has not been formally proven (Brockdorff, 2011). Other examples include the ANRIL lncRNA at the *INK4A/ARF* locus and an interaction between short promoter proximal RNAs and PRC2 (Kanhere et al., 2010; Orlando et al., 2012).

Problematically, none of the PRC2 core components contain a RNA binding domain. Furthermore, *in vitro* interaction experiments between proteins and RNAs are difficult due to many non-specific interacting RNAs. It is therefore difficult to rigorously test the importance of RNAs in PcG recruitment (Brockdorff, 2013).

1.3.5.6 A responsive model for PcG recruitment

Taken together, it appears that there are many different mechanisms that might contribute to PcG complex recruitment. An alternative perspective on this issue was recently presented by Klose and colleagues (Klose et al., 2013). They suggested that rather than being recruited by any direct mechanism, PcG complexes sample all CpG islands and are either stabilised there or only interact transiently. Stability of association is based on the level of transcription or activating versus repressing chromatin environment. This sampling hypothesis is supported by the finding that low levels of RING1B can be found at all CpG islands, whilst high levels are only at known PRC1 targets (Farcas et al., 2012). It has also been shown that transcription through a PRE causes PcG displacement in *Drosophila* (Schmitt et al., 2005). Furthermore, as discussed in 1.3.5.2, it is clear that PcG complexes take many cues from the chromatin environment and can stabilise their own presence.

In this context, it could be suggested that the presence of particular transcription factors may be permissive to PcG stabilisation whereas others are not. Alternatively, it could be that proposed recruiters in fact simply either remove non-permissive marks or place stabilising marks. Similarly, lncRNAs could appear to act as recruiters, when they are in fact blocking transcription; lack of transcription may subsequently permit PcG stabilisation.

This hypothesis therefore proposes that PcG recruitment is responsive to the expression status of a gene, rather than instructive, and acts to stabilise this status. The authors suggest that the observation that most PcG-target genes are *not* upregulated upon PcG depletion (as discussed in 1.3.3) fits with the hypothesis that these genes are not upregulated because the required transcription factors are not present in that cell type in the first place.

PRC1 and PRC2 complexes then act to mark chromatin as silenced, whilst stabilising their own presence through positive feedback (Leeb and Wutz, 2007; Endoh et al., 2008; Margueron et al., 2009; Tavares et al., 2012). However, strong cues for switching in transcription status such as the new expression of master transcriptional regulators as happens during differentiation are suggested to be able to overcome PcG mediated repression. Perhaps the PcG repressors are required to keep genes in a stably silenced state that nonetheless remains responsive to changes occurring during differentiation (Voigt et al., 2013).

1.3.6 Bivalent domains

ChIP analysis has shown that, in ES cells, H3K27me3 was frequently co-enriched with H3K4me3 (Azuara et al., 2006; Bernstein et al., 2006). This was very surprising because these marks are thought to have opposing effects, and as such these regions were called "bivalent" domains. By performing re-ChIP, whereby the simultaneous presence of two marks or proteins is assayed, it was shown that these modifications do occur together in the same region, rather than on different alleles or within different cells of the population.

Later, it has been shown that these modifications can occur on the same nucleosome, but not on the same histone tail (Young et al., 2009; Schmitges et al., 2011; Yuan et al., 2011; Voigt et al., 2012). Despite containing H3K4me3 and replicating early during S phase, these domains were usually transcribed at very low levels (Azuara et al., 2006; Bernstein et al., 2006).

Interestingly, it was described that bivalent domains often resolve to monovalency, i.e. H3K27me3 only or H3K4me3 only, during differentiation. Resolution to H3K4me3 only correlated with active transcription, whilst H3K27me3 only correlated with repression and usually occurred at genes that were not expressed in the lineage that the cells had differentiated into. Intriguingly, some genes retained their bivalency and these genes were often expressed in related lineages (Bernstein et al., 2006; Mikkelsen et al., 2007). It has also been reported that some genes gain bivalent domains during differentiation (Mohn et al., 2008).

Subsequently, bivalent domains have also been described in humans (Pan et al., 2007; Zhao et al., 2007) and in iPS cells (Maherali et al., 2007; Mikkelsen et al., 2007; Guenther et al., 2010). Importantly, bivalent domains have been reported in the developing embryo (Alder et al., 2010; Dahl et al., 2010; Rugg-Gunn et al., 2010). The presence of bivalent domains in lower vertebrates is unclear, as they have been reported to be present in the developing zebrafish (Vastenhouw et al., 2010; Lindeman et al., 2011), but not in the Western clawed frog *Xenopus tropicalis* (Akkers et al., 2009). It has been shown that *Drosophila* embryos do not contain bivalent domains (Schuettengruber

et al., 2009; Gan et al., 2010). In summary, bivalency is a biological phenomenon that has been observed by several techniques and in several vertebrates.

The precise mechanism by which bivalency is established is not known, nor is the biological role of bivalent domains. It has been suggested that the co-existence of activating and repressing histone modifications at bivalent domains actively keeps genes "poised" for expression. However, an alternative interpretation, based on the responsive model discussed in 1.3.5.6, is that bivalency is a read-out of the simultaneous recruitment of PcG and TrxG (see 1.3.7) complexes simply because associated genes are poised for expression.

1.3.7 Trithorax complexes and function

After the discovery of the first PcG proteins, the TrxG founding member Trithorax was characterised as a positive regulator of the *Hox* genes (Ingham, 1985). Some genes were identified in genetic screens for their ability to repress PcG-induced phenotypes. In contrast to mutations in PcG members, mutations in TrxG members caused anterior homeotic transformations. Some TrxG proteins have been shown to regulate global transcription (Armstrong et al., 2002; Srinivasan et al., 2008). It remains therefore unclear to what extent particular TrxG proteins are general activators as opposed to specifically antagonising PcG complexes.

TrxG proteins have found to perform various activities. Some of the best characterised TrxG proteins are found in the COMPASS family complexes, including Trithorax itself. Each COMPASS complex contains an H3K4me transferase related to *S. cerevisiae* Set1.

In *Drosophila melanogaster* there are three COMPASS complexes based on the unique HMT and each HMT has 2 homologues in mammals (in brackets): dSet1 (SET1A/B), Trithorax (MLL1/2), and Trithorax related (MLL3/4) (reviewed in Shilatifard, 2012). It appears that SET1A/B mediate bulk H3K4me3, whereas MLL1/2 is required for H3K4me3 at only a small subset of genes, including the *Hox* genes (Wu et al., 2008).

Interestingly, MLL3/4 are not required for H3K4me at *Hox* genes (Goo et al., 2003; Lee et al., 2006a), but do contain the H3K27me3 demethylase UTX. UTX and the other H3K27me3 demethylase JMJD3A are found rarely at bivalent domains in ES cells, but are recruited during and required for differentiation (Agger et al., 2007; Lee et al., 2007; Burgold et al., 2008). However, it has also been proposed that the main function of MLL3/4 is to allow correct chromatin formation at enhancers by demethylating H3K27me3 to allow H3K27ac accumulation and placing H3K4me1 (Herz et al., 2012b; Voigt et al., 2013). Nonetheless, depletion of both the maternal and zygotic pool of Utx in *Drosophila melanogaster* leads to death at the larval stage and loss of expression of *Hox* genes (Copur and Muller, 2013), confirming its identity as a TrxG gene.

Another HMT that acts a TrxG member is the protein Absent small and homeotic discs 1 (Ash1) in *Drosophila* and its mammalian homologue ASH1L. It has been proposed that this protein acts as an H3K36me2 histone transferase (Tanaka et al., 2007; An et al., 2011; Tanaka et al., 2011; Yuan et al., 2011). Taken together, this indicates that some TrxG proteins act by placing histone modifications known to inhibit PRC2 activity.

Other TrxG proteins form part of ATP-dependent chromatin remodelling complexes, such as *Drosophila* Brahma and mammalian BRM and BRG1 that are part of the SWI/SNF complex (reviewed in Schuettengruber et al., 2011). These proteins may counteract PcG mediated compaction or may aid the recruitment of transcriptional activators which inhibit PcG mediated silencing.

How TrxG proteins are recruited to their targets is currently unclear. It looks likely that recognition of chromatin marks may play a role here again, as the MLL1/2 proteins contain a CXXC domain that could target them to unmethylated CpG islands, whilst MLL proteins also contain PHD domains that can bind H3K4me2 and H3K4me3 (Milne et al., 2010). Furthermore COMPASS HMT activity is stimulated by H2B ubiquitination (Kim et al., 2009b). Other targeting mechanisms have also been proposed, including direct interactions with the polymerase-associated factor 1 (PAF1) transcriptional elongation complex (Muntean et al., 2010), recruitment by sequence specific binding factors, including some also implicated in *Drosophila* PREs, and non-coding RNAs (reviewed in Schuettengruber et al., 2011).

Therefore, TrxG proteins exert their PcG antagonistic functions by altering the chromatin environment and stimulating transcription. The extent to which they are targeted to specifically antagonise PcG-mediated silencing versus simply counteracting silencing in general remains an open question.

1.3.8 PcG proteins in X chromosome inactivation

Female mammals inactivate one of their X chromosomes, with the expression of many X-linked genes repressed, in order to ensure correct dosage of gene expression from the X chromosome between males and females, (Lyon, 1961). The inactivated X is inherited faithfully over many generations (reviewed in Wutz, 2011). How X chromosome inactivation (XCI) is initiated and maintained is a topic of great interest in epigenetic research. Because PcG proteins are recruited to the future inactive X during XCI, this represents a model system for polycomb recruitment.

1.3.8.1 Regulation of X chromosome inactivation

X chromosome inactivation occurs in two different ways in female zygotes: imprinted XCI and random XCI. The paternally inherited X chromosome is inactivated from the 4 cell stage onwards. This imprinted inactivated state is maintained in the extraembryonic lineages, whilst cells of the inner cell mass reactivate the paternal X chromosome at E3.5 (Okamoto et al., 2004). Subsequently, either the maternally or paternally inherited X chromosome is inactivated from around E5.5. This process of random XCI is recapitulated *in vitro* by differentiating ES cells (Martin et al., 1978; Takagi and Martin, 1984; Lee et al., 1996; Panning and Jaenisch, 1996; Penny et al., 1996).

In mammals, the *X-inactivation specific transcript* (*Xist*) is a spliced lncRNA greatly upregulated on the future inactive X (Borsani et al., 1991; Brockdorff et al., 1991; Brown et al., 1991). The *Xist* gene is required for the initiation of XCI (Penny et al., 1996).

After upregulation of *Xist* from the prospective inactive X, the *Xist* RNA coats the X chromosome and forms a characteristic *Xist* cloud which can be visualised by FISH 24-48 hours after *Xist* induction by differentiation (Brown et al., 1992).

It has been found that the inactive X forms a silencing compartment, with silent regions in the core, whilst actively transcribed regions are on the outer edge of the X chromosome territory (Chaumeil et al., 2006; Clemson et al., 2006). After *Xist* coating of the X chromosome, the subsequent observable event is the depletion of RNA polymerase II and splicing factors (Okamoto et al., 2004; Chaumeil et al., 2006), as well as the depletion of active marks such as acetylation (Heard et al., 2001). How this depletion is achieved is unclear. It has been suggested that lack of transcription itself may be sufficient to cause partitioning of the chromatin for compaction, as the active X was severely reduced in size in actinomycin D-treated cells (Naughton et al., 2010). Subsequently, PRC1 and PRC2 are recruited and the inactive X becomes enriched for H3K27me3 and H2AK119ub (see also discussion below) (Mak et al., 2002; Plath et al., 2003; Silva et al., 2003; de Napoles et al., 2004; Wang et al., 2004a). It is believed that these events constitute the initiation phase of XCI. It has been shown that *Xist* cannot initiate XCI in differentiated cells, showing that XCI is tightly regulated by development (Wutz and Jaenisch, 2000; Kohlmaier et al., 2004).

Later events are associated with the maintenance, rather than the initiation, of XCI. These include the adoption of a later replication timing profile and the enrichment of the histone variant macroH2A (Costanzi and Pehrson, 1998; Csankovszki et al., 1999; Mermoud et al., 1999). Several days later, CpG islands at genes that have been silenced become

methylated, whilst gene-poor regions are hypomethylated. Eventually, female cells contain a highly stably repressed X chromosome, which can be visualised as the Barr body {Barr, 1949 #562}.

1.3.8.2 Role of PcG proteins in XCI initiation and maintenance

The PcG proteins have been studied widely in the context of XCI because of their early recruitment to the inactive X and their known role as repressors. As discussed in 1.3.5.5, it has been suggested that *Xist* itself can recruit the PcG proteins. Evidence for this comes from *in vitro* experiments using recombinant single subunits of PRC2 and from *in vivo* cross-linking studies (Zhao et al, 2008, Kaneko et al, 2010; Kanhere et al, 2010), and have implicated both SUZ12 and EZH2 as direct RNA binders. In addition to *in vitro* association studies between protein and RNA being prone to artefacts, no specific binding surfaces have been identified and rigorously tested thus far, so it is difficult to know whether the interaction is direct or indirect (Brockdorff, 2011). Furthermore, PRC1 can be recruited in the absence of PRC2 (Schoeftner et al., 2006; Tavares et al., 2012), and a direct recruitment model would therefore also have to provide evidence for a direct PRC1-*Xist* interaction, evidence which is lacking so far. Importantly, *Xist* with a deletion of a conserved region called repeat A (*XistΔrepA*) cannot silence the X chromosome, but can recruit PcG proteins, albeit at reduced levels (Plath et al., 2003; Kohlmaier et al., 2004; Schoeftner et al., 2006; Agger et al., 2007; Zhao et al., 2008). Therefore, it has been reasoned that PRC2 is directly responsive to *Xist* and not a downstream silencer: an argument for the direct recruitment hypothesis. However, it should be noted that despite failing to silence the X chromosome, *XistΔrepA* does cause RNA pol II depletion and some change in chromatin marks (Chaumeil et al., 2006; Pullirsch et al., 2010), which could be sufficient to recruit PcG proteins.

Importantly, disruption of components of either PRC1 or PRC2 does not lead to XCI defects in differentiating ES cells and the embryo proper (Kalantry and Magnuson, 2006; Schoeftner et al., 2006; Leeb and Wutz, 2007). In contrast, in female *Eed*$^{-/-}$ embryos, extraembryonic trophoblast tissue displayed defects in imprinted XCI maintenance (Wang et al., 2001). However, subsequent analysis revealed that whilst *Xist* localisation is lost, gene repression was maintained in trophoblast stem cells, and only aberrantly reactivated during differentiation of these trophoblast stem cells (Kalantry et al., 2006) suggesting its role in XCI maintenance might be less important than previously thought.

Taken together, it has been conclusively demonstrated that PcG proteins are recruited to the X chromosome during XCI, although the mechanism of recruitment is still under debate. It appears that PcG recruitment is neither sufficient nor necessary for gene repression. Rather, it probably works cooperatively with other silencing mechanisms to initiate and maintain X chromosome inactivation.

1.4 AEBP2: a PRC2 associating factor

1.4.1 Identification of AEBP2

Aebp2 was identified based on its ability to bind to a regulatory DNA sequence known as *Ae-1* upstream of the *aP2* gene, which is important for adipocyte differentiation (He et al., 1999). It contains a putative nuclear localisation signal and does indeed localise to the nucleus. Additionally, it contains the consensus phosphorylation site for Cdk1 which suggests it might be regulated in a cell cycle dependent manner. The ability of AEBP2 to

bind to the *Ae-1* sequence was verified in an electromobility shift assay (EMSA). Also, it was shown that overexpressed AEBP2 could act as a repressor by reducing transcription from a luciferase expressing plasmid with three copies of the *Ae-1* sequence upstream of the promoter. Furthermore, if an AEBP2-Gal4 fusion was transfected and targeted to Gal4 binding sites, it could repress transcription even more effectively. Together, this suggested that the AEBP2 was a DNA-binding repressor.

Subsequently, AEBP2 was identified as part of the PRC2 complex (Cao et al., 2002). Later this interaction was confirmed in many studies, using mass spectrometry on epitope-tagged members of the PRC2 complex (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li et al., 2010; Walker et al., 2010; Hunkapiller et al., 2012). Here I will summarise what is currently known about AEBP2.

1.4.2 Domain architecture and expression of AEBP2

Initially, a 303 amino acid protein was identified, which contained the S-rich domain, the zinc finger domain and the basic domain (see Figure 1.8). Larger isoforms, with amino acid coding exons 5' to the originally identified start site, were identified and annotated later, most notably a 274 amino acid isoform and a 496 amino acid isoform. Further, the addition of very small exons was observed at the 3' end of some transcripts (exon 9a or 9b, termed alternative C-terminal domains (ACTD) a or b). ACTDa is a conserved lysine rich region consisting of 14 amino acids, whilst ACTDb is not conserved between human and mouse (Kim et al., 2009a). Instead, human isoforms appear to have other short ACTDs present.

It was found that AEBP2 contains 3 consecutive zinc fingers from the Gli-Krüppel family (He et al., 1999). AEBP2's other domains however share no homology with other known domains (Figure 1.8). Interestingly, the zinc finger domain is 100% conserved between human and mouse, whilst the proteins containing exons 1b-8 share 84% amino acid identities, with most of the variation in the acidic and neutral domains.

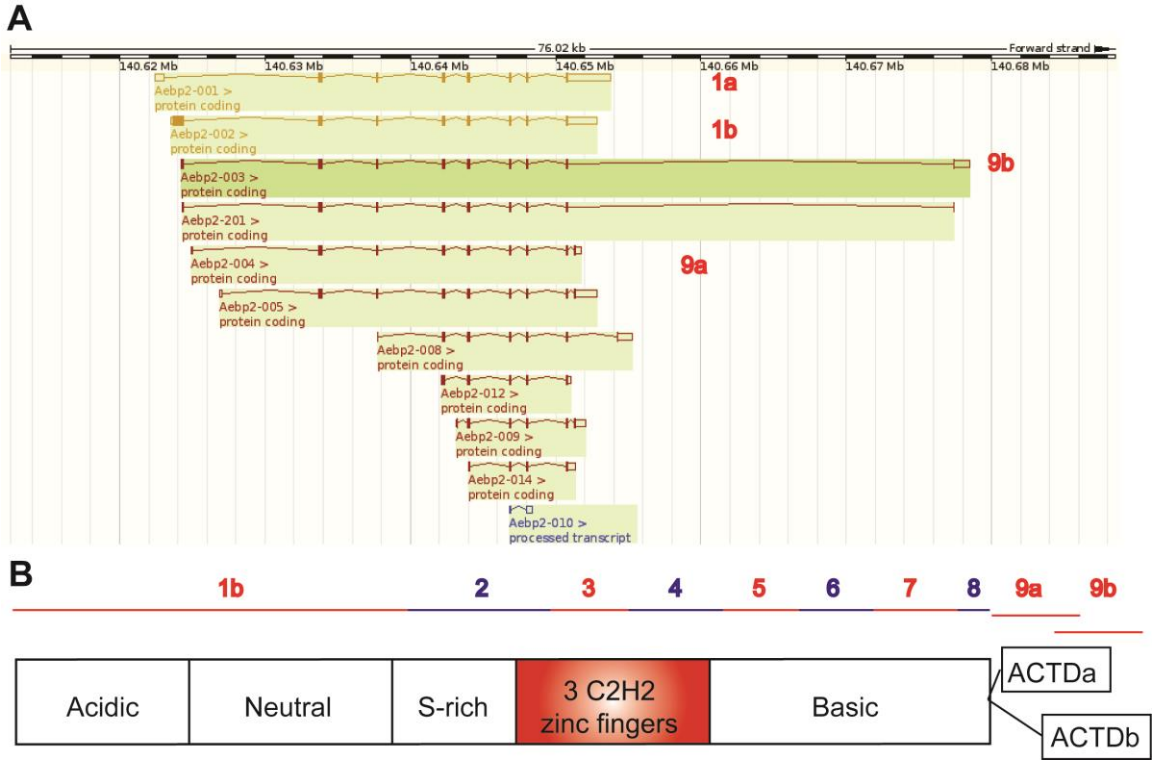


Figure 1.8: Different *Aebp2* isoforms. (A) Different transcriptional isoforms of *Aebp2* taken from the Ensembl 72 database. Golden colour indicates it has been annotated by both Ensembl and Havana. Open boxes indicate untranslated regions. (B) Protein domains of AEBP2 and how they relate to the different isoforms described in A. ACTD: Alternative C-terminal Domain, made up by exon 9a or 9b

Not much is known about the regulation of expression of *Aebp2*. A CpG island overlaps with exon 1b, whilst it has been reported that MAX and MYC bind to the *Aebp2* promoter in ES cells and stimulate transcription (Neri et al., 2012). Northern blotting showed that a ~3.5 kbp and a ~4.5 kbp transcript were detected ubiquitously expressed in

adult tissues, with the exception of the liver (He et al., 1999). Most *Aebp2* transcripts have a long 3' untranslated region (UTR) and sometimes a long 5' UTR. These UTRs may act as targets for micro RNAs. For example, the miRNA target predictor TargetScan (Lewis et al., 2003) predicts AEBP2 is a good target for miR-101, which has been shown to target and repress expression of EZH2 in humans (Varambally et al., 2008; Friedman et al., 2009).

How the different *Aebp2* isoform transcripts are regulated is unknown. The Ensembl database suggests that a 3727 bp transcript corresponds to the 496 amino acid isoform, whilst a 4460 bp transcript represents the 274 amino acid isoform. These sizes likely represent the ~3.5 kbp and ~4.5 kbp transcripts detected by He et al. (1999). From a survey of cDNAs, it is clear that both transcripts are detected in embryonic and adult tissues. This is consistent with the detection of peptides of the longer isoform in mouse ES cells (Peng et al., 2009; Walker et al., 2010). However, one study found that the shorter isoform was predominantly expressed in embryos, whilst the longer was more prevalent in adult tissues (Kim et al., 2009a).

The zinc finger domain has been proposed to be responsible for the reported DNA binding activity (He et al, 1999; Kim et al, 2009), and has been proposed to recruit the PRC2 complex to target genes (see also Chapter 2). No function has been proposed for any of the other domains. Interestingly, homologues have been identified in fish and insects (Kim et al., 2009a). In all homologues the zinc finger domain and the basic domain are highly conserved, whilst the serine rich domain is conserved between

mammals and fish. This domain is absent in insects, which instead contain large N-terminal uncharacterised domains.

1.4.3 Interaction with the PRC2 complex

After the initial description of repressor protein AEBP2, it was isolated in association with the PRC2 complex in HeLa cells (Cao et al., 2002). The authors had set out to purify the complex responsible for a histone methyltransferase activity that they observed and identified the PRC2 complex as a H3K27 methyltransferase. The identified complex also contained the larger form of AEBP2. In a subsequent study, it was found that whilst a trimeric complex of EED, SUZ12, and EZH2 was absolutely required for H3K27 methylation *in vitro*, addition of RBBP4 and the short form of AEBP2 further increased methylation levels (Cao and Zhang, 2004). This study also showed that AEBP2 could interact directly with EED, RBBP4 and SUZ12 in single subunit *in vitro* immunoprecipitation experiments. Results also indicated that AEBP2 could self-associate and that self-association was important for interaction with RBBP4, although the significance of this is not understood. Through deletion studies the SUZ12 subunit was suggested to interact mostly with AEBP2 through a 150 amino acid region in the C-terminus, whilst N-terminal sequences and the zinc finger appeared more important for RBBP4 interaction.

Further insight was brought by a study performed by Ciferri and colleagues (Ciferri et al., 2012). The authors used electron microscopy (EM) to determine the overall shape of the PRC2 complex. It was found that clear structural details could not be obtained when using a tetrameric complex of EED, EZH2, SUZ12 and RBBP4, likely due to conformational flexibility of the complex. However, when the short isoform of AEBP2

was incorporated and followed by mild cross-linking, electron microscopy and subsequent class averaging showed the PRC2 complex adopted a four-lobed organisation. The authors then used cross-linking followed by mass spectrometry to identify regions of the proteins that are close to each other, as well as tagging of individual subunits to deduce where each of the subunits was within the four-lobed structure based on increased electron density from the tag. Thus it was shown that AEBP2 is in close proximity to all the other subunits, except RBBP4. Notably, many cross-links between the AEBP2 zinc finger domain and the EZH2 SET domain were observed. Furthermore, AEBP2 made many contacts to a central region of SUZ12, and together these proteins appeared to form an "arm" bridging the upper half (consisting largely of EED and EZH2) and the lower half (consisting largely of SUZ12 and RBBP4) of the structure (Figure 1.9).

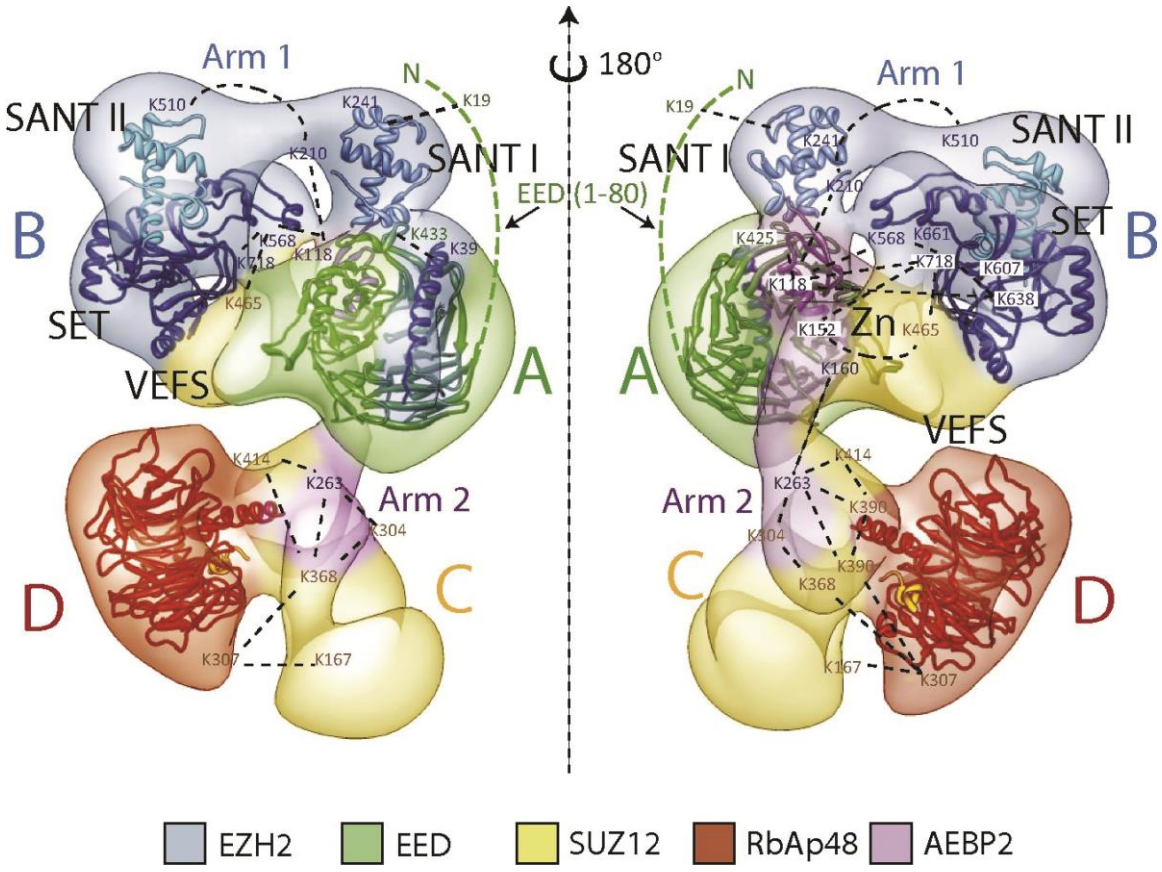


Figure 1.9: The overall structure of the PRC2 complex. Reproduced from Ciferri et al. (2012) under CC BY 3.0 license. The crystal structures of EED and the WD40 domain of RBBP4, as well as crystal structures of homologue SANT, SET and Zinc finger domains have been docked. Positions of lysine cross-links are also indicated. The structure shows AEBP2 and SUZ12 forming a bridge between upper half and the lower half.

It was somewhat surprising that AEBP2 appeared to have a central role in coordinating the PRC2 subunits as it is believed to be a substoichiometric subunit itself and therefore not all PRC2 complexes will contain AEBP2. However, it is possible that SUZ12 alone is sufficient to bridge the two parts of the PRC2 complex or that other associating factors could assist SUZ12. This research has nevertheless placed AEBP2 at the heart of the PRC2 complex.

1.4.4 A developmental role for *jing* and AEBP2

A *Drosophila melanogaster* homologue of *Aebp2* has been identified: *jing*. It has been shown that *jing* is important for the development at several stages and organs in the fruit fly, including the central nervous system and leg development (Sedaghat et al., 2002; Sedaghat and Sonnenfeld, 2002; Culi et al., 2006; McClure and Schubiger, 2008). However, a connection to PRC2 has been less well studied. *jing* has been reported to interact genetically with *Polycomb* (Culi et al., 2006) and a weak physical interaction with the PRC2 complex has been reported (Herz et al., 2012a). However, to what extent *jing* may perform similar functions to *Aebp2* or may act differently remains unclear, all the more because of the lack of homology in the N-terminal domain.

Because AEBP2 is part of the PRC2 complex, it is likely it is important for development in mammals. This was confirmed by a study in which embryos homozygous for a constitutive *Aebp2* gene-trap could not be detected at E10.5, suggesting that embryonic lethality occurs prior to this time point (Kim et al., 2011). However, a more detailed analysis of homozygous embryos has not been reported, nor the analysis of any knockout or knockdown ES cell lines. Instead, analysis of location of gene expression of *Aebp2* suggested that it was highly expressed in neural crest cells and its derivatives (Kim et al., 2011). Animals with a heterozygous constitutive *Aebp2* gene-trap displayed symptoms similar to the rare genetic disease Waardenburg syndrome. Together, these findings indicate that *Aebp2* plays a role in development, but the mechanism by which it does so remains to be elucidated.

1.5 Aims of this study

He and colleagues discovered that AEBP2 can act as a repressor, and this is likely to depend on its association with the PRC2 complex. However, AEBP2's role specifically within the PRC2 complex is less well described. In this study, I have aimed to elucidate the role that AEBP2 plays in PRC2 biology, by aiming to answer the following questions:

-Does AEBP2 bind to a specific DNA sequence?

-Is AEBP2 recruited to PcG target sites?

-Which proteins does AEBP2 associate with?

-What is the result of genetic depletion of *Aebp2* in mouse ES cells and in the mouse embryo?

2. Materials and methods

2.1 Primers for cloning

Primers used for cloning of overexpression constructs are presented in Table 3.

Name	**Sequence 5'→ 3'**
Aebp2LicF1	TACTTCCAATCCATGGCCGCCGCGCTCGCC
Aebp2LicStrep_Rev	TATCCACCTTTACTGTTACTTCTCGAACTGAGGGTGGCTCCATCCTCCCCTCTTCAACCTCTTCTG
Aebp2Lic 496R	TATCCACCTTTACTGTTACCTCTTCAACCTCTTCTG
NFlagFW	CTCGAGACCATGGACTACAAGGACGACGATGACAAGGGTGACTACAAGGACGACGATGACAAGGGAGGTACCCTCGGATCCATGGCCGCCGCGCTCGCCG
NFlagR	GCGGCCGCTTACCTCTTCAACCTCTTCTG
CFlagFW	CTCGAGACCATGGCCGCCGCGCTCGCCG
CFlagR	GCGGCCGCCTTACTTGTCATCGTCGTCCTTGTAGTCACCCTTGTCATCGTCGTCCTTGTAGTCGGATCCGAGGGTACCTCCCCTCTTCAACCTCTTCTG

Table 3: Primers used for cloning.

2.2 Polymerase Chain Reaction (PCR)

PCR mixtures were set up as follows:

	Taq	**Expand High Fidelity Plus**	**Phusion**	**Expand Long Template**
DNA template	1 µl	1 µl	2 µl	3 µl
Buffer	1x Taq Buffer supplemented with 1.5 mM MgCl	1x HiFi^{PLUS} buffer	1x Phusion® HF Buffer	1x Long Template buffer 1
Primers (each)	0.5 µM	0.4 µM	0.5 µM	0.3 µM
dNTP	0.2 mM	0.2 mM	0.2 mM	0.35 mM
Enzyme	1.25 U	2.5 U	0.5 U	3.75 U
Total reaction volume	25 µl	50 µl	50 µl	50 µl

Table 4: Conditions used for PCR reactions in this study.

The cycling parameters were as follows:

step	*Taq*	Expand High Fidelity Plus	Expand Long Template
1	94 °C 2 min	94 °C 2 min	94 °C 2 min
2	94 °C 30 s	94 °C 30 s	94 °C 10 s
3	55 °C 30 s	55 °C 30 s	53 °C 30 s
4	72 °C 1 min	72 °C 90 s	68 °C 2 min 30 s
5	Cycle to step 2 34 more times	Cycle to step 2 9 more times	Cycle to step 2 9 more times
6	72 °C 5 min	94 °C 30 s	94 °C 10 s
7		55 °C 30 s	53 °C 30 s
8		72 °C 90 s (increase by 10 s each cycle)	72 °C 2 min 30 s (increase by 20 s each cycle)
9		Cycle to step 6 14 more times	Cycle to step 6 19 more times
10		72 °C 7 min	68 °C 7 min

Table 5: Cycling parameters for PCR reactions in this study. Parameters for phusion polymerase are detailed in 2.16.4.

2.3 Ligation independent cloning (LIC)

5µg of vector plasmid was digested with the appropriate restriction enzyme in order to linearise the plasmid and to excise the SacB gene. The vector DNA was isolated by gel extraction using the QIAquick Gel Extraction Kit (Qiagen). Insert DNA was amplified by PCR as described and purified by gel extraction using the QIAquick Gel Extraction Kit. Vector and insert were processed by T4 DNA polymerase (ThermoScientific) in the presence of 0.04 mM dGTP (vector) and 0.04 mM dCTP (insert) for 30 min at 22 ºC. After heat-inactivation of the polymerase (20 min, 75 ºC), samples were purified using the QIAquick Gel Extraction Kit. Vector and insert were incubated using a range of ratios (3:1 – 1:6) in 2 µl total volume for 30 min at 25 ºC and introduced into bacteria.

2.4 Chemical transformation of DH5α and BL21 DE3 pLysS *E. coli*

Competent DH5α (*fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) and BL21 DE3 pLysS (*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS + λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*) cells were generated in-house using a protocol based on Inoue et al. (1990). Bacteria were thawed on ice and 1 µl DNA or ligation mixture where added. DNA was mixed with bacteria by gentle flicking of the tube and incubated on ice for 25-30 min. Bacteria were heat-shocked at 42 °C for 30 s and allowed to recover on ice for 2 min, before being spread on high salt (10 mg NaCl/ml) LB agar plates containing the appropriate antibiotic for selection.

2.5 Constructions of overexpression vectors and targeting vector

2.5.1 Bacterial overexpression *Aebp2* constructs

Aebp2 constructs consisting of the full length protein (496 amino acids; AEBP2-FL), an N-terminal truncation (amino acids 223-496) and a construct with an N-terminal and C-terminal truncation, the "zinc finger fragment" or AEBP2-ZF (amino acids 223-348) were provided in the pNIC28-Bsa4 vector by Sarah Cooper. They were created from image clone 3499299. These constructs were tagged with an N-terminal 6xHis tag. A C-terminal Strep-tag was fused to the full length protein by Expand High Fidelity PLUS (Roche) PCR amplification with Aebp2LicF1 and Aebp2LicStrep_Rev, the product of which was inserted into an empty pNIC28-Bsa4 vector by LIC cloning (see 2.3) and

introduced into DH5α bacteria by chemical transformation. Successful ligation was confirmed by sequencing (Geneservice from Source BioScience).

2.5.2 Mammalian overexpression constructs

To generate the insert for the pCAGIPuro-N-Flag-Aebp2, the AEBP2-FL cDNA was used as a template in a PCR with the forward primer and the reverse primer NFlagR. To generate the insert for the pCAGIPuro-Aebp2-C-Flag, a PCR was set up with the forward primer CFlagFW and the reverse primer CFlagR using the AEBP2-FL as template.

PCR products were first cloned into pGEM® T-easy vector (Promega) and then this vector and the pCAGIPuro vector were digested with XhoI and NotI restriction enzymes (New England Biolabs, NEB). The pCAGIPuro vector was treated with Alkaline phosphatase from calf intestine (CIAP, NEB) and inserts were ligated using T4 ligase (Promega). Ligations were introduced into DH5α bacteria by chemical transformation. Successful ligation was confirmed by sequencing (Geneservice from Source BioScience). Correct vector DNA was purified using QIAprep Spin Miniprep Kit (Qiagen).

The insert for the pLion-N-Flag-Aebp2 vector was generated by using the Aebp2LicF1 primer and Aebp2Lic496R using the AEBP2-FL cDNA as template. Subsequently the PCR product was cloned into the pLion vector (which contained an in-frame FLAG) using LIC cloning.

The insert for the FS2-AEBP2 vector was generated using the Aebp2LicF1 primer and Aebp2Lic496R using the AEBP2-FL cDNA as a template. Subsequently the PCR product was cloned into the pCAGIPuro_FS2LIC vector (which contained a LIC site and an in-frame Flag-2XStrepII tag) using LIC cloning.

2.5.3 Design of synthesis constructs for targeting vector

Please note: genomic coordinates are those as found on the UCSC browser using mm9.

Two constructs were synthesised by GeneArt® (Life Technologies): "pMA Aebp2 vector" and "pMA loxP STOP cassette".

"pMA Aebp2 vector" contained 428 bp homologous to a region 6745 bp upstream of the intended targeting location of *Aebp2* (140573572-140573999 on chromosome 6). This was followed by a SmaI/XmaI digestion site and then 202 bp homologous to a region 9741 bp downstream of the intended targeting location (140590400-140590601 on chromosome 6). This sequence was inserted into the pMA vector.

"pMA loxP STOP cassette" was constructed as follows: first 243 bp immediately upstream to the intended targeting location (140580074-140580316 on chromosome 6). This was followed by an FRT "up", F3 "up", loxP "up" and lox511 "up" site, where "up" and "down" mean opposing directionality of the site, such that sites that are in *cis* which are both "up" will cause a deletion, whereas sites which are "up" and "down" will cause an inversion. The splice acceptor was situated downstream (note that the first nucleotides included in the splice product are "CTCGCGGAGG" and that it is important this will

result in a sequence that is in-frame with the start of *GFP*), which was then fused to the *GFP* gene. The construct then contained a bacterial promoter, flanked by NotI restriction enzyme sites, for transcription of the aminoglycoside 3' phosphotransferase required for kanamycin and geneticin resistance. This was followed by FRT "down", F3 "down", loxP "down" and lox511 "down" site and then 231 bp immediately downstream to the intended targeting location (140580631-140580861 on chromosome 6). Therefore integration of the cassette into the genome eventually removed sequence 140580317-140580630 on chromosome 6. This sequence was inserted into the pMA vector with SfiI restriction enzyme recognition sites on either side. The sequence of the vectors is available upon request.

2.5.4 Construction of targeting vector by BAC recombineering and cloning

2.5.4.1 Primers for colony PCR

name	sequence
rec1_5prime fw	TTCACAGTAGTGGACATTTTG
rec1_3prime rev	TCAGCCGCACCAGAAGAGGG
rec2_300fw1	TAAATAACAAGGTGGAGGCC
rec2_300rev1	GTTAATGCAGCTGGCACGAC

Table 6: Primers used for colony PCR. See sections 2.5.4.4 and 2.5.4.5

2.5.4.2 BAC purification

100 ml overnight culture was pelleted and resuspended in 12.5 ml P1 buffer from the QIAprep Spin Miniprep Kit. Subsequently, 12.5 ml of P2 buffer was added and incubated at 25 °C for 5 min. Then this was mixed with 17.5 ml N3 buffer. Cell debris was removed by centrifugation (3700*g*, 25 °C, 5 min), after which the supernatant was centrifuged again. 28 ml isopropanol was added to the supernatant and then centrifuged

for 15 min at 4 °C, 20,400*g*. The pellet was washed in 10 ml 70% ethanol and centrifuged for 10 min at 4 °C, 17,400*g*, and then in 1 ml and centrifuged in a microfuge for 10 min at 4 °C, 16,100*g*. The pellet was dissolved in 600 µl TE-NaOH pH 7.6 and treated with 0.33 mg/ml RNase A for 30 min at 37 °C. BAC DNA was purified by two phenol:chloroform extraction steps and one chloroform extraction and then precipitated by the addition of 0.04 volume 5 M NaCl and 2.5 volume 100% ethanol. DNA was pelleted by centrifugation (16,100*g*, 10 min, 4 °C) and washed with 1 ml 70% ethanol. The pellet was air-dried and dissolved in 100 µl water and stored at 4 °C.

2.5.4.3 Introduction of the pRed/ET recombineering plasmid

DH5α bacteria containing BAC RP23-68-M10 (which contains the targeting region of *Aebp2*) were transformed with 34 ng plasmid pRed/ET (Gene bridges, a gift from I. Stancheva) by electroporation as follows. 30 µl of an overnight culture of DH5α bacteria containing BAC RP23-68-M10 was used to inoculate 1.4 ml SOC medium containing 15 µg/ml chloramphenicol. Cultures were incubated for 2.5 hrs in a Thermomixer (Eppendorf) whilst shaking at 1000 rpm. Cells were pelleted (11,200*g*, 30 s, 2 °C) and washed once in 1 ml ice-cold water. Subsequently cells were resuspended in 30 µl ice-cold water and pRed/ET was added and briefly mixed. Cell suspension was transferred to a pre-chilled 1 mm gap electroporation cuvette (Bio-Rad) and immediately electroporated (1350 V, 10 µF, 600 Ω, 5 ms) in a Gene Pulser Xcell™ (Bio-Rad). Cells were immediately transferred to 1 ml SOC and incubated at 30 °C for 70 min. As a control, cells without addition of the pRed/ET plasmid were used. After incubation, 100 µl was spread onto high salt LB agar plates containing 15 µg/ml chloramphenicol and 3 µg/ml tetracyclin. Plates were kept at 30 °C overnight, whilst wrapped in aluminium foil.

2.5.4.4 Integration of the STOP cassette into the BAC

"pMA loxP STOP cassette" was digested by AscI and AvrII and the insert was purified by QIAquick Gel Extraction Kit (Qiagen). This was subsequently digested by SfiI and purified in order to remove any uncut vector. 30 µl pRedET + BAC containing cells from an overnight culture (experimental) and 30 µl only BAC containing cells from an overnight culture (control) were used to inoculate 3x1 ml LB containing 15 µg/ml chloramphenicol and 3 µg/ml tetracyclin, and 3x1 ml LB containing only 15 µg/ml chloramphenicol respectively. Tubes were incubated at 30°C for 2 h with shaking at 1100 rpm until OD600 ~ 0.2-0.3 (one of the tubes was used to measure the OD600). Expression of Red/ET recombination proteins was induced by the addition of 50 µl 10% Arabinose (Sigma A-3256) into one culture, whilst one was kept as uninduced control and incubated for 50 min at 37 °C. Cells in all cultures were pelleted and washed in 1 ml ice-cold water and electroporated (as described in 2.5.4.3) with 200 ng of excised construct. Cells were resuspended in 1 ml SOC buffer and incubated at 37 °C for 120 min. The whole culture was then plated on High Salt LB agar plates containing 15 µg/ml chloramphenicol and 15 µg/ml kanamycin, which were incubated at 37 °C overnight.

Success was first tested by the ratio of colonies on the induced experimental plate versus the other plates (uninduced cells should give fewer colonies and control plates should have no colonies). Secondly, Taq colony PCR (bacteria are transferred straight to PCR reaction mixture) with rec1_5prime fw and rec1_3prime rev was used to show the insertion of sequence in the correct place. Lastly, purified BAC DNA was digested with restriction enzymes to show correct digestion pattern.

2.5.4.5 BAC subcloning

The “pMA Aebp2 vector” was linearised by SmaI digestion and purified by QIAquick gel extraction, after which the sample was treated with CIAP. The pRed/ET plasmid was re-introduced, as described in 2.5.4.3, into the bacteria that now contained a targeted BAC. These bacteria, and control bacteria containing only the targeted BAC but not pRed/ET, were grown in 30 °C in LB containing 15 µg/ml chloramphenicol, 15 µg/ml kanamycin and 3 µg/ml tetracyclin, and 15 µg/ml chloramphenicol, 15 µg/ml kanamycin respectively. 1 µl linearised vector (330 ng) was electroporated in the same fashion as the STOP cassette had been (2.5.4.4). Cultures were spread onto LB High Salt agar plates containing 100 µg/ml ampicillin and 50 µg/ml kanamycin.

Success was first tested by the ratio of colonies on the induced experimental plate versus the other plates. Secondly, colonies were re-streaked to test whether they were now sensitive to chloramphenicol. Thirdly, Taq colony PCR with rec2_300fw1 and rec2_300rev1 showed a correct PCR product. Fourthly, purified vector DNA was digested with restriction enzymes to show correct digestion pattern.

2.5.4.6 Removal of the bacterial promoter

Vector DNA was digested with NotI for 10 min, and re-ligated using the high concentrated T4 Ligase (NEB). One successful colony was shown to be sensitive to kanamycin again and show a correct digestion pattern. Subsequently, the whole vector (21792 bp) was sequenced. The sequence of the targeting vector is available upon request. This vector was then used for targeting.

2.6 Cell Culture

2.6.1 Cell lines

Cell lines used in this study are listed below.

Cell line	Medium	Gelatine	Feeders	Reference
PGK12.1	ES cell	+	-	Penny et al. (1996)
E14TG2A	ES cell	+	-	
Eed lines	ES cell KO	+	PEFs	Landeira et al. (2010)
Eed lines with *Xist* transgene	ES cell KO	+	-	Schoeftner et al. (2006)
129/1	ES cell	+	SNL	Kay et al. (1993)
HEK 293-T	EC10	-	-	

Table 7: Cell lines used in this study. PEFs: Primary Embryonic Fibroblasts; SNL: a mouse fibroblast STO cell line transformed with neomycin resistance and murine *Lif* genes.

2.6.2 Culturing conditions

ES cells were grown in ES cell medium, which consisted of Dulbecco's Modified Eagle Medium (DMEM, from Life Technologies) supplemented with 6.7% foetal calf serum (FCS, from Seralab), 2 mM L-glutamine, 1x non-essential amino acids, 50 µM 2-mercaptoethanol, 50 µg/ml penicillin/streptomycin (all from Life Technologies) and LIF-conditioned medium, made in-house, at a concentration equivalent to 1000 U/ml. Non-differentiating ES cells were grown on tissue culture dishes coated with Phosphate Buffered Saline (PBS, 160 mM NaCl, 3 mM KCl, 8 mM Na_2HPO_4, 1 mM KH_2PO_4, pH 7.3) +1% gelatine. *Eed* lines were grown in ES cell knockout (KO) medium, which was KnockOut™ DMEM (Life Technologies) with the same supplements as ES cell medium, whilst also containing 3.3% Knockout Serum Replacement (KnockOut™ SR from Life Technologies). EC10 medium consisted of DMEM, supplemented with 6.7% FCS, 2 mM L-glutamine, 1x non-essential amino acids, 50 µM 2-mercaptoethanol and 50 µg/ml penicillin/streptomycin. Cells were grown at 37°C in a humid atmosphere with 5% CO_2.

Cells were passaged using 0.05% trypsin-EDTA (Life Technologies) with 2% Chicken Serum (Life Technologies) and frozen in FCS + 10% dimethyl sulfoxide (DMSO, Sigma).

2.6.3 Preparation of feeders

To prepare feeders for ES cell culture, SNL cells were grown to confluence in EC10 medium and inactivated by treating cells with EC10 containing 10µg/ml of mitomycin C (Sigma) for 2h. SNL cells were used at $4x10^4/cm^2$. PEFs were derived from E 13.5 embryos, passaged once in culture and then inactivated as SNL-Ps. PEFs were used at $3x10^4/cm^2$. ES cells grown on a feeder cell layer were separated from feeders for experiments such as ChIP and immunoprecipitation by pre-plating for 25-30 min.

2.6.4 Doxycycline treatment

Eed lines with *Xist* transgene (Schoeftner et al., 2006) were grown on 25x75x1.0 mm SuperFrost® Plus slides (VWR) in square petri dishes and treated with 1 µg/ml doxycycline. After 24 hours these slides were used for imaging (see 2.13).

2.6.5 Differentiating PGK12.1 cells

ES cells were plated at a low density ($0.6 x10^6$ cells) in a 9 cm non-gelatinised tissue culture dish in EC10 medium. Medium was changed after 24 hours. On day 2, cells were transferred to square dishes containing 25x75x1.0 mm SuperFrost® Plus slides (VWR) which had been coated with poly-D-lysine hybromide (Sigma). On day 3, these slides were used for imaging (see 2.13).

2.6.6 Transfection of mammalian overexpression constructs

To generate stable ES cell lines, 4 µg expression constructs were transfected into E14 cells using Lipofectamine 2000 (Life Technologies). Stable integrants were selected

using 1.75 µg/mL puromycin and individual colonies were picked and tested for expression of the construct by immunoblotting.

2.6.7 Metaphase spread

Cells were grown until nearly confluent in T25 flasks and incubated at 37 °C with fresh medium containing 1.5 µg/ml ethidium bromide (Roche). After 1 hr and 20 min, colchicine (KaryoMAX® Colcemid, Life Technologies) was added to a final concentration of 0.1 µg/ml and incubated for a further 40 min. Then cells were washed in PBS and harvested by trypsinisation at room temperature. After inactivation of the trypsin by addition of medium cells were pelleted by centrifugation (400*g*, 3 min, 25 °C). A pellet of 1-2 mm thickness was carefully resuspended in 1 ml Hypotonic Solution (75 mM KCl) for no more than 5 min. Then 200 µl freshly prepared Fixative (4 °C, 75% methanol, 25 % Acetic Acid) was added drop wise. The tube was not agitated in any way, but carefully placed in the centrifuge for pelleting (400*g*, 3 min, 25 °C). Supernatant was removed leaving about 100 µl, which was used to resuspend the cells in by gentle flicking. 1 ml of Fixative was added to the resuspended cells and incubated overnight at 4 °C without agitation. The following day the cells were carefully resuspended in the same fixative and pelleted as before. The pellet was resuspended in 1.5 ml Fixative and pelleted again. This step was repeated twice more. Then cell suspension was dropped onto clean microscope slides and air dried. Once dry, slides were washed twice and PBS and DAPI stained and analysed using a 63x immersion oil objective on the Zeiss Axio Observer.Z1 microscope.

2.7 Antibodies

Antibodies used in this study are listed in Table 8.

Antibody	Source	Product number or reference	In ChIP	In IF	In IB	In IP
AEBP2-unpurified	This study	-	20 µl	1:100	-	-
AEBP2-Protein A purified	This study	-	-	-	1:200	-
AEBP2-antigen purified	This study	-	-	1:10	1:200	16 µl
IgG	Sigma	M7023	2 µl	-	-	4 µl
H3K27me3	Diagenode	pAb-069-050	5 µl	-	1:1000	-
H3	Abcam	1791	3 µl	-	1:10,000	-
SUZ12	Cell Signalling	3737	5 µl	1:800	1:1000	-
EZH2	Cell Signalling	5246	5 µl	1:100	1:1000	10 µl
EZH2	Diagenode	C15410039	-	-	1:1000	-
EED	Gift from A. Otte	Sewalt et al. (1998)	-	1:100	1:1000	-
RING1B	Gift from H. Koseki	Atsuta et al. (2001)	15 µl	1:500	-	-
H2AK119Ub	NEB	8240	-	1:500	-	-
YY1 (C-20)	Santa Cruz	Sc-281	-	-	1:200	-
FLAG M2	Sigma	F1804	-	1:500	1:1000	-
FS2	Gift from R. Klose	Farcas et al. (2012)	5 µl	-	1:1000	-
Anti-mouse Ig, HPR	Amersham	NA934V	-	-	1:2000	-
Anti-rabbit Ig, HRP	Amersham	NA931V	-	-	1:20,000	-
Alexa Fluor® 568 goat anti-mouse IgG (H+L) highly cross absorbed	Life Technologies	A11031	-	1:400	-	-
Alexa Fluor® 488 anti-rabbit IgG (H+L)	Life Technologies	A11008	-	1:400	-	-

Table 8: Antibodies used in this study. ChIP: Chromatin Immunoprecipitation; IF: immunofluorescence; IB: immunoblot; IP: immunoprecipitation.

2.8 Cell extracts

2.8.1 Whole cell extracts

Whole cell extracts were prepared by harvesting cells, washing the cells in PBS and resuspending them in PBS. The cell suspension was diluted twofold with Loading Buffer (2% SDS, 20% glycerol, 0.1 M Tris-HCl pH 8.0, 0.002% Bromophenol Blue, 1.4 M beta-mercaptoethanol (Sigma)) and snap-frozen on dry ice. After thawing, samples were boiled for 5 min at 95°C.

2.8.2 Nuclear cell extracts

Nuclear cell extracts were prepared by harvesting cells after which the cells were either processed immediately or snap-frozen on dry ice and stored at -80°C. The cell pellets were washed with PBS and resuspended in 10 packed cell volume (PCV) buffer A (10 mM HEPES-KOH pH 7.9, 1.5 mM MgCl2, 10 mM KCl, with 0.5 mM DTT, 0.5 mM PMSF, and complete protease inhibitors (Roche) added fresh). After a 10 min incubation at 4 ºC, cells were collected by centrifugation (1500*g*, 5 min, 4 ºC) and resuspended in 3 PCV of buffer A + 0.1% NP-40 (Sigma). After another 10 min incubation at 4 ºC, nuclei were collected by centrifugation (400*g*, 5 min, 4 ºC) and resuspended in 1 PCV buffer C (250 mM NaCl, 5 mM HEPES-KOH (pH 7.9), 26 % glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA-NaOH, pH 8.0 with complete protease inhibitors (Roche) + 0.5mM DTT added fresh). 5 M NaCl was added drop wise to bring the concentration to 350 mM and the mixture was incubated for 1 hr at 4 ºC with occasional agitation. After centrifugation (16,100*g*, 20 min, 4 ºC), the concentration of the supernatant was quantified using the Bio-Rad Bradford assay and stored at -80 ºC.

2.9 Immunoblotting

Samples were separated on a polyacrylamide and transferred onto a nitrocellulose membrane by wet transfer (for histones; 100V for 1 hr) or semi-dry transfer (for all other protein; 12 V for 50 min). Membranes were blocked by incubating them for 1 hour at room temperature in 10 ml Tris Buffered Saline (150 mM NaCl, 0.1 M Tris-HCl pH 8.0) containing 0.1% Tween (TBST) with 5% w/v Marvell milk powder. Blots were incubated overnight at 4°C with primary antibody, washed 4 times for 10 min with TBST and incubated for 40 min with secondary antibody conjugated to horseradish peroxidase. After 4 times 5 min washes with TBST bands were visualised using ECL (GE Healthcare).

2.9.1 Quantitation of intensity of bands

Bands on immunoblots were quantified using AIDA Image Analyzer v.4.22.

2.10 Co-immunoprecipitation

2.10.1 Co-immunoprecipitation of endogenous proteins

Cells from a confluent 140 mm plate were harvested and resuspended in 500 µl cell lysis buffer C300 (20 mM HEPES-KOH pH 7.9, 1.5 mM $MgCl_2$, 0.1 % NP40, 0.2 mM EDTA-NaOH pH8.0, 300 mM KCl with freshly added 0.5 mM DTT, cOmplete EDTA-free (Roche)). Cells were incubated on ice for 20 min and cell debris was removed by centrifugation (16,100*g*, 20 min, 4°C). Alternatively, 400-500 µg of NCE was used and salt concentration was adjusted to 300 mM NaCl in a 500 µl total reaction volume. Extracts were pre-cleared with 30 µl packed volume salmon sperm-blocked Protein A Agarose beads (Millipore) in the presence of 250 U Benzonase® Nuclease (Millipore) for 30 min at 4 ºC. Antibodies were added as indicated in Table 8 and incubated at 4 ºC

overnight. Subsequently, 30 µl packed volume Protein A Agarose beads were added and incubated for 1 hr at 4 ºC. The flow through was collected and beads were washed 5 times with 1 ml BC300 (300 mM KCl, 10% Glycerol, 50 mM HEPES-KOH pH 7.9, 0.5 mM EDTA-NaOH pH 8.0, with 0.5 mM DTT added fresh). Subsequently beads were boiled in 30 µl SMASH buffer (50mM Tris-HCl pH 6.8, 10% Glycerol, 2% SDS, 0.02% bromophenol blue, 1% beta-mercaptoethanol) for 5 min at 95 ºC. 5 µl of the supernatant was loaded as the immunoprecipitation (IP) sample.

2.10.2 Co-immunoprecipitation of epitope-tagged AEBP2 for mass spectrometry

E14 cells containing either the empty pCAGIPuro_FS2LIC vector or the FS2-AEBP2 vector were grown to confluency on 64 140 mm dishes and harvested into PBS (with freshly added 1x cOmplete protease inhibitor) by scraping at 4 °C. Subsequently nuclear cell extracts were prepared as described in 2.8.2. The salt concentration of 10 mg NCE was reduced to 150 mM NaCl by diluting with buffer C (see 2.8.2) without NaCl. Extracts were incubated with 200 µg Avidin (IBA) and 250 U Benzonase® Nuclease for 30 min at 4 ºC with rotation. Any precipitates were removed by centrifugation (16,100*g*, 5 min, 4 ºC) and the supernatant was incubated for 3 hrs at 4 ºC whilst rotating with 50 µl packed resin StrepTactin Superflow beads, which had been washed twice with buffer C. Subsequently, the beads were transferred to a Protein LoBind tube (Eppendorf) and washed eight times with wash buffer (20mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% NP-40, 5% glycerol with 1 mM DTT added fresh). Subsequently proteins were eluted with 4 x 65 µl elution buffer (10 mM Desthiobiotin, 20mM Tris-HCl pH 8.0, 150 mM NaCl, 0.2% NP-40, 5% glycerol with 1 mM DTT added fresh) by incubation for 15 min at 4 ºC without rotation). 6 µl per elution was loaded for immunoblotting.

2.11 Gel filtration

2.11.1 Size Exclusion Chromatography

To analyse complex size by gel filtration, ~1 mg nuclear cell extract in 450 µl was loaded onto a superose 6 10/300 GL column (GE healthcare) which had previously been pre-calibrated with Mix 2 (thyroglobulin 669 kDa, aldolase 158 kDa, ovalbumin43 kDa, both from GE Healthcare) in BC300 buffer with freshly added 0.5 mM DTT. Fractions were collected in 200 µl.

2.11.2 TCA extraction

0.4 volume 100% ice-cold Trichloric Acid (TCA) was added to eluted fractions, mixed and incubated for 30 min at 4 °C. Precipitates were pelleted (16,100*g*, 15 min, 4 °C) and resuspended in 1 ml ice-cold acetone. Precipitates were pelleted (16,100*g*, 15 min, 4 °C) and acetone removed. Protein was air-dried for 10-15 min and resuspended in 100 µl SMASH buffer supplemented with 15 mM Tris-HCl pH 8.0.

2.11.3 Analysis by immunoblot

Subsequently, 20% (i.e. 20 µl) of TCA extracted samples was analysed by immunoblot. For this, large 10% SDS-PAGE gels (16x16 cm from Bio-Rad) were run for 4 hrs with constant 50 mA and proteins were transferred by semi-dry transfer (constant 200 mA for 90 min). Subsequent protocol as described in 2.9.

2.12 Chromatin immunoprecipitation

2.12.1 ChIP protocol based on Stock et al. (2007)

The ChIP protocol for those experiments which included immunoprecipitation of AEBP2 was based on that described in Stock et al. (2007), with modifications. Cells were

harvested and counted. 5 x 10^7 cells were resuspended in 10 ml fresh medium with 1% formaldehyde and incubated for 10 min at 37 °C, whilst gently shaking. The reaction was stopped with addition of glycine to a final concentration of 0.125 M and incubated for 5 min at 25 ºC with gentle agitation. Cross-linked cells were washed three times with 10 ml ice-cold PBS and subsequently resuspended in Swelling Buffer (25 mM HEPES-KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% NP-40 with 1x cOmplete EDTA-free protease inhibitors added fresh). Cells were lysed for 10 min at 4 °C with occasional agitation. Subsequently cells were scraped from petri dishes and nuclei isolated by Dounce homogenisation (50 strokes, "Tight" pestle). Nuclei were pelleted by centrifugation (1,800*g*, 15 min, 4 °C) and resuspended in Sonication Buffer (50 mM HEPES-KOH pH 7.9, 140 mM NaCl, 1 mM EDTA-NaOH pH 8.0, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS with 1x cOmplete EDTA-free protease inhibitors added fresh) at 1 ml buffer per 1 x 10^7 cells as counted after harvesting. Subsequently samples were sonicated with 1 ml sample per 15 ml High Clarity Polypropylene Conical Tube (Scientific Laboratories Supplies) in a BioRuptor sonicator (Diagenode) for 30 min in cycles of 30 s on, 30 s off on the "high" setting, at 4°C to produce DNA fragments with a length of <1.6 kb. Chromatin was cleared from cell debris by centrifugation (16,100*g*, 15 min, 25 ºC) and approximate concentration was measured by measuring absorbance using the Nanodrop ND-1000 (ThermoScientific) of sonicated chromatin diluted five-fold by 0.1 M NaOH, using Sonication Buffer diluted five-fold by 0.1 M NaOH as a blank. 200 µg chromatin per IP was subsequently pre-cleared (1 hr, 4 ºC, with rotation) using 30 µl rProtein A Sepharose beads (GE Healthcare) that had been blocked for 1 hr at 4 °C with 1 mg/ml Bovine Serum Albumin (BSA) and 1 mg/ml yeast tRNA (Sigma). 20 µg chromatin was saved as a "10% input" sample. Pre-cleared samples incubated with antibodies (as indicated in Table 8) overnight at 4°C with rotation. Protein-antibody

complexes were isolated by incubating with 50 µl washed and blocked rProtein A Sepharose beads for 2 hrs at 4 °C with rotation. Complexes were washed with 1 ml wash buffer per wash for 5 minutes at 4 ºC rotating as follows: once with Sonication Buffer, twice with Wash Buffer A (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA-NaOH pH 8.0, 1% NP-40, 0.5% Na-deoxycholate), twice with Wash Buffer B (20 mM Tris-HCl pH 8.0, 1 mM EDTA-NaOH pH 8.0, 250 mM LiCl, 1% NP-40, 0.5% Na-deoxycholate) and finally twice with TE (10 mM Tris-HCl pH8.0). Complexes were eluted from beads by incubation with 250 µl Elution Buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA-NaOH pH 8.0, 1% SDS) first at 65 ºC for 5 min, and subsequently for 15 min at 25 ºC with fast rotation. The supernatant was collected and the elution procedure repeated. Saved "10% input" samples were brought up to 500 µl with TE and to these and to the 500 µl combined elutions from the immunoprecipitated samples NaCl was added to a final concentration of 160 mM. Cross-links were reversed for >16 hrs at 65 ºC in the presence of 50 µg/ml DNase-free RNase A (Roche). Subsequently, samples were treated with Proteinase K (Roche) (0.2 mg/ml in the presence of 4 mM EDTA-NaOH pH 8.0) for 2 hrs at 45 ºC. DNA was purified using ChIP DNA Clean & Concentration™ columns (Zymo Research) and eluted in 10 µl volume. For PCR, samples were diluted ten-fold with water and 2 µl was used as a template per quantitative real time PCR reaction (qPCR). DNA was amplified in the presence of 0.3 mM forward and reverse primer (see Table 9) and 1x Sybr-Green (Bio-Rad) over 40 cycles after Taq activation (95 ºC for 3 min) with the following conditions: 20 s at 95 ºC, 20 s at 65 at ºC, plate read.

Name	Sequence forward primer	Sequence reverse primer
Gata1	GTTGGATCCCCTTTGCCTTT	TCCAACCCACACATAGCCTG
Nanog	CCCAGGTTTCCCAATGTGAAG	AAAGAGTCAGACCTTGCTGCCA
Cdx2	ACCACCTTCTGCCTGAGAATGTAC	CCTCCAATCACAGGTTCAAAGACT
Dlx1	ATGTCTCCTTCTCCCATGTCC	ACTGCACGGAACTGATGTAGG
Nkx2.2 -3.5	TAACGCTTTGGAAGCAGACGTG	GCACTGGGCAGCTAGACTAGGA
Nkx2.2 -2.5	CAGAGAGCCCCAGCCTGAAA	TCCCTTTTCCCCTTGCAAGA
Nkx2.2 A	GTCGCTGACCAACACAAAGACG	TGTCGTAGAAAGGGCTCTTAAGGG
Nkx2.2 G	AAAGTATGCCAACTCGGTGCCA	GGAAGATAATCTTCTGGGCTCCCA

Table 9: Primers used for qPCR for ChIP.

2.12.1.1 Blocking experiment

20 µl antibody was blocked by either 30 µg BSA or 30 µg AEBP2-FL in a total reaction volume of 100 µl in BC100 (100 mM KCl, 10% Glycerol, 50 mM HEPES-KOH pH 7.9). The mixture was incubated for 1 hr at 25 ºC with rotation and subsequently added to the chromatin samples and incubated overnight as described in 2.12.1.

2.12.2 ChIP protocol based on Schmidt et al. (2009)

The ChIP protocol for those experiments which did not include immunoprecipitation of AEBP2, was based on that described in Schmidt et al. (2009), with modifications. Note that this protocol was used to prepare samples for FS2-AEBP2 ChIP-seq. Ethylene glycol bis[succinimidylsuccinate] (EGS, ThermoScientific) was first dissolved in DMSO (Sigma) and then added to 10 ml PBS in a final concentration of 2 µM. Cells were harvested and counted and 5 x 10^7 cells were resuspended in 10 ml PBS with EGS and incubated for 1 hr at 25 ºC with rotation. Subsequently formaldehyde was added to a final concentration of 1% and incubated for 15 min at 25 °C. The reaction was stopped with addition of glycine to a final concentration of 0.125 M and incubated for 3 min at 25 ºC with gentle agitation. After centrifugation (700*g*, 4 min, 4 ºC), cells were resuspended in

10 ml LB1 (50 mM HEPES-KOH, pH 7.9, 140 mM NaCl, 1 mM EDTA-NaOH pH 8.0, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100 with 1x cOmplete EDTA-free protease inhibitors added fresh) and incubated at 4 ºC for 10 min with rotation. Nuclei were pelleted (700*g*, 4 min, 4 ºC) and resuspended in 10 ml buffer LB2 (10 mM Tris-HCl, pH8.0, 200 mM NaCl, 1 mM EDTA-NaOH pH 8.0, 0.5 mM EGTA with 1x cOmplete EDTA-free protease inhibitors added fresh). After a 5 min incubation at 4 ºC with rotation, nuclei were pelleted and resuspended in 1 ml LB3 (10 mM Tris-HCl, pH 8, 100 mM NaCl, 1 mM EDTA-NaOH pH 8.0, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% *N*-lauroylsarcosine with 1x cOmplete EDTA-free protease inhibitors added fresh). Samples were aliquoted in 1 ml per 15 ml Polypropylene tube and sonicated in a BioRuptor sonicator (Diagenode) for 30 min in cycles of 30 s on, 30 s off on the "high" setting, at 4°C to produce DNA fragments with a length of <2 kb. After sonication, 100 µl of LB3 buffer also containing 10% Triton X-100 was added per 1 ml sample and chromatin was cleared from cell debris by centrifugation (16,100*g*, 15 min, 25 ºC). Chromatin was either frozen at -80 ºC or immediately used. 1 ml sample was diluted ten-fold with dilution buffer (1% Triton X-100, 2 mM EDTA-NaOH pH 8.0, 150 mM NaCl, 20 mM Tris-HCl pH 8.0 with 1x cOmplete EDTA-free protease inhibitors added fresh). 1 ml of diluted sample was used per immunoprecipitation whilst 100 µl was set aside as "10% input" sample. Samples were pre-cleared (1 hr, 4 ºC, with rotation) using blocked 30 µl rProtein A Sepharose beads (see 2.12.1). Pre-cleared samples were incubated with antibodies (as indicated in Table 8) overnight at 4°C with rotation. Protein-antibody complexes were isolated by incubating with 30 µl washed and blocked rProtein A Sepharose beads for 1 hr at 4 °C with rotation. Complexes were washed with 1 ml wash buffer for 4 minutes rotating at 4 ºC per wash as follows: once with Low Salt Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA-NaOH pH 8.0, 150 mM NaCl, 20 mM Tris-

HCl pH 8.0), once with High Salt Buffer (0.1% SDS, 1% Triton-X-100, 2 mM EDTA-NaOH pH 8.0, 150 mM NaCl, 20 mM Tris-HCl pH 8.0), once with LiCl Buffer (0.25 M LiCl, 1% NP-40, 1% Na-Deoxycholate, 1 mM EDTA-NaOH pH 8.0, 10 mM Tris-HCl pH 8.0) and twice with TE. Complexes were eluted from beads by incubation with 100 µl Elution Buffer (1% SDS, 0.1 M $NaHCO_3$) first at 65 ºC for 5 min, and subsequently for 30 min with vortexing at 25 ºC. The supernatant was collected and to all samples, including saved "10% input" samples, NaCl was added to a final concentration of 200 mM. Cross-links were reversed for >4 hrs at 65 ºC, after which 100 µg/ml DNase-free RNase A was added and incubated at 42 ºC for 90 min. Subsequently, samples were treated with Proteinase K (0.2 mg/ml) for 1 hr at 45 ºC. DNA was purified using ChIP DNA Clean & concentration™ columns and eluted in 10 µl volume. For PCR, samples were treated as described in 2.12.1.

2.12.3 ChIP-seq

For ChIP-seq, undiluted purified DNA was post-sonicated in a BioRuptor sonicator for 60 min in cycles of 30 s on, 30 s off, at 4°C. Concentration was quantified using the PicoGreen® dsDNA Quantitation Kit (Molecular Probes). 125-180 ng was sent to the Oxford Genomics Centre. 5 ng was used to prepare libraries using the NEB Next® DNA Library Prep Master Mix Set for Illumina®. Adaptor sequences were ligated at a concentration of 150 nM and sequences with an adaptor were amplified using a PCR of 18 cycles. This PCR also added the unique index tag. Concentration of the samples was analysed by Qubit. Sample sizes, as well as the presence of any adapter-adapter species, were analysed by running the libraries on an Agilent TapeStation. Libraries were normalised and pooled to create a balance for sequencing. The final pool was diluted to 10nM and a qPCR was carried out to determine how much of the library to run on the

sequencing flowcell. The library was sequenced on one lane of an Illumina HiSeq2000 50bp paired end run. Four samples were used per lane.

2.12.4 ChIP-seq analysis

Analysis was performed using the default settings of each programme unless otherwise mentioned.

2.12.4.1 Deriving Bed and bedgraph files, viewing in IGB

ChIP-seq data was analysed using the Galaxy platform (Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010). Quality of sequencing data was assessed using FastQC (version 1.0.0) (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). 51 bp single-end tags were mapped to the mouse genome (mm9) using Map with Bowtie for Illumina (version 1.1.2) (Langmead et al., 2009). By setting the "-m" parameter to 1, reads which could align to more than one place in the genome were excluded from further analysis. After mapping, the input sample contained 39,937,047 reads and the IP sample contained 40,136,209 reads. Reads were ordered by the "Convert SAM to BAM (version 2.0.0)" tool available on Galaxy. Subsequently, peaks were called using Model-based Analysis of ChIP-Seq (MACS2 version 2.0.10.2). Filtering took place in order to remove exact copies (which may have arisen during library preparation rather than being immunoprecipitated sequence) such that for input sample 38,239,249 and for IP sample 37,410,529 reads remained. Peaks were detected with a q-value cut-off at 0.05 (Zhang et al., 2008). MACS2 also generated a sequencing trace which was visualised in the Integrated Genome Browser (IGB) (Nicol et al., 2009).

To analyse EZH2 and SUZ12 datasets, SRR034760 (experiment SRX015830) and SRR034190 (SRX015831) datasets were downloaded, which represent the EZH2 and

SUZ12 IP datasets from Peng et al. (2009). The same procedure as for the FS2-AEBP2 ChIP-seq dataset was followed. 6,092,414 reads were obtained for EZH2 and 17,191,320 reads were obtained for SUZ12. After MACS2 filtering, 3,536,454 and 15,738,554 reads were used to call peaks for EZH2 and SUZ12 respectively.

2.12.4.2 Distribution over genomic elements

Peaks were used to quantify coverage of specific genomic elements using "CEAS: Enrichment on chromosome and annotation" (Shin et al., 2009) using the Cistrome platform (Liu et al., 2011). Visualisation of distribution over transcription start sites was performed using MakeTSSdist (version 1.0) available on Galaxy.

2.12.4.3 GO analysis

Nearest genes to peaks obtained in MACS2 were annotated using HOMER (Heinz et al., 2010). Extracted Ensembl identifiers were analysed using the DAVID gene functional classification tool (Huang da et al., 2009b, a).

2.12.4.4 Overlap between datasets and aggregation plots

Overlap or exclusivity of peaks was calculated using the intersect and subtract functions of Operate on genomic intervals available in Galaxy. Subsequently Bedgraphs as generated by MACS2 were plotted over obtained BED files using SitePro: Aggregation plot tool for signal profiling (version 1.0.0) (Shin et al., 2009).

2.12.4.5 Heatmaps

To generate heatmaps, BED files were ordered based on peak width and read intensity was plotted across the intervals using seqMINER (Ye et al., 2011).

2.13 Immunofluorescence microscopy

Cells were either grown on slides or cover slips (13 mmm diameter cover slips from VWR). Cells were washed twice in PBS and fixed in 2% formaldehyde for 15 min at 25 °C. Cells were then permeabilised in 0.4% Triton-X-100 for 5 min and washed three times in PBS. Cells were blocked in fish gelatin and incubated with the appropriate dilution of antibody in fish gelatine (128 mM NaCl, 2.4 mM KCl, 6.4 mM Na_2HPO_4, 0.8 mM KH_2PO_4, pH 7.3 with 1 g gelatin from cold water fish skin (Sigma) added fresh) + 5% normal goat serum for 2 hrs at 25 °C. After three fish gelatin washes cells were incubated with the appropriate dilution of secondary antibody conjugated to Alexa fluorophore for 1 hr at 25°C. Cells were washed twice in fish gelatin and twice in PBS. Subsequently DNA was stained using DAPI (Invitrogen), washed in PBS and mounted in mounting medium (Dako). Slides were visualised using a 63x oil immersion objective with the Zeiss Axio Observer.Z1 microscope. Images were taken as a single focal plane.

2.13.1 Blocking experiment

3 µl AEBP2 unpurified antibody was blocked with a range of amounts of AEBP2-FL in fish gelatine in a 50 µl reaction for 1 hr before being applied as primary antibody.

2.14 Mass spectrometry

2.14.1 For epitope-tagged AEBP2

To assess the quality of the sample, 15 µl of the first elution was run on a 10% SDS-PAGE gel and stained with SYPRO® Ruby (Bio-Rad). 50 µl of the first elution was subjected to trypsin digestion and used for Liquid Chromatography – Tandem Mass Spectrometry (LC-MS/MS) and subsequent analysis as described in Farcas et al. (2012),

performed by J. McGouran and B. Kessler at the Nuffield Department of Medicine, University of Oxford.

2.14.2 For SELEX contaminants and AEBP2 immunoprecipitation

Samples were run on a 4-12% NuPAGE Bis-Tris acrylamide gel (Invitrogen) and stained with InstantBlue (Triple Red) or the mass spectrometry-compatible SilverQuest silver staining kit (Invitrogen). Bands of interest were excised and, if silver-stained, washed according to manufacturer's instructions. Subsequently the dried gel pieces were subjected to trypsin digestion and LC-MS/MS by the Thermo Q Exactive Orbitrap LC-MS/MS mass spectrometer, performed by B. Thomas at the Oxford Central Proteomics Facility, Sir William Dunn School of Pathology, University of Oxford. For SELEX contaminants, the data were searched against an in-house made database of *E. coli* proteins with fixed modification set as carbamidomethyl cysteine, and the variable modification set as oxidised methionine. For the endogenous AEBP2 IP, the data were searched against the UniProtCP_mouse database and the UniProt_SwissProt database, with fixed modification set as carbamidomethyl cysteine, and the variable modification set as oxidised methionine.

2.15 Assessing RNA levels

2.15.1 RNA extraction

Cells from 90 mm dishes were harvested and washed twice in PBS after which the cell pellet was resuspended in 1-2 ml Trizol® reagent (Life Technologies). Samples were incubated for 2 min at 25 °C and then 0.2 volume chloroform was added to each and mixed by vigorous shaking for 15 s. Samples were incubated for 2 min at 25 °C and centrifuged at 12,000*g* for 15 min at 4 °C. The upper aqueous phase was transferred to a clean pre-chilled tube and an equal volume of isopropanol, pre-chilled to 4 °C, was

added. The sample was incubated at 4 °C for 10 min and RNA was pelleted at 12,000*g* at 4 °C for 30-60 min. Subsequently the pellet was washed twice in 75% ethanol and then air-dried in 3-5 min and resuspended in 100 µl RNase-free water.

2.15.2 DNase treatment and quantitative real time PCR

Contaminating DNA was removed using the Ambion® DNA-free™ DNase Treatment kit (Life Technologies) according to the manufacturer's instructions, except that 2 µl DNase I was used (instead of 1). cDNA was synthesised using the SuperScript™ III kit (Life Technologies) using Oligo dT primers (Life Technologies) according to the manufacturer's instructions. Quantitative real time PCR was performed using the same cycling conditions as those described in 2.12.1. A list of primers for testing RNA expression is shown in Table 10.

Primer pair	**Forward primer**	**Reverse primer**
Idh1	TGCCAGCTCGATCTACCACAAAAT	AGAAAATGTGGAAGAGCCCTAACG
Aebp2 exon 3-4	TGCAAGGTGTATAACACCCC	AACTTTTGAGGAGTTCTGCT
Aebp2 exon 4-5	CAATGCCAGCTTTGCTTCTC	ATCGAAGAAATCGTGTGGCC
Aebp2 exon 5-6	GGCCACACGATTTCTTCGAT	GCATCCAATGAAGCAAAAGC
Aebp2 exon 1a-Gfp	CTCCTCTGGAAGAGCGGTTA	GCCCTTGCTCACCATCAA
Rex1	GAGCTGAACTCCTAGCCGCCTAGATT	TTTGGTCAGTGGTATTTGGGGAGA
Eed	CAAATGTGTGAACAGCCCTCAAGGAA	GCATCAGCATCTACATAGGACTGCAA
Sox9	TATTAGAGACCCTGAGCTGGAAGT	CTGGACTGAAACTGGTAAAGTTGT
HoxA4	GATGAAGAAGATCCACGTGAGCG	CCAGTTCCAAGACTTGCTGCC
Mrg1	TATCACATAGTTCGTGGACCTTTTCT	CTGGTTTTTGAGAAGTGCGA

Primer pair	Forward primer	Reverse primer
Ebf1	CCAACGGAGTGGAAGCAGTATG	TAGAGGGCCAGGACGAAGTG
Pax3	CCCAGTGAGAGGGGAGAGAGCATA	CTTCGAACGCAGACAGCAGC
Irx3	GCCAACATGATCTTCTCGCCCT	GAGCTGCCCATCTTCCCACA
Gata6	GACTCCTACTTCCTCTTCTTCTAATTCAGA	ACCTGAATACTTGAGGTCACTGTTCTC

Table 10 List of primers used for expression analysis by qPCR.

2.16 Systematic evolution of ligands by exponential enrichment (SELEX)

2.16.1 AEBP2 purification

Plasmids containing the constructs were introduced into BL21 DE3 pLysS by chemical transformation. Cultures expanded from single colonies were grown in 2 L LB with kanamycin to an optical density at 600 nm of 0.5-0.7. 1 mM IPTG was added and cells were incubated at 30°C for 3 hours. Cells were harvested by centrifugation and stored at -80°C.

Subsequently, cells were lysed by resuspension in lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.1% NP40 and 1x EDTA-free cOmplete protease inhibitor cocktail (Roche)) and subsequent sonication (6x30 seconds at 60% using the microtip on the Vibra-Cell (Sonics)). The cell debris was pelleted by centrifugation after which the supernatant was incubated on a rotor at 4°C for 1 hr with 1 ml immobilised metal affinity chromatography (IMAC) Sepharose-6 Fast Flow beads (GE Healthcare) that had been charged with nickel ions. The mixture was poured onto a Econopac Chromatography Column (Bio-Rad) and the beads were washed with 30 ml wash buffer (50 mM

NaH_2PO4, 300 mM NaCl, 20 mM imidazole-NaOH pH 8.0). Protein was eluted in 1 ml fractions by the elution buffer (50 mM NaH2PO4, 300 mM NaCl, 250 mM imidazole-NaOH pH 8.0). Fractions were analysed by SDS-PAGE and protein-containing fractions combined.

The zinc finger fragment was dialysed into BC100 buffer (50 mM HEPES-KOH pH 7.9, 100 mM KCl, 10% glycerol, 0.5 mM DTT). It was subsequently subjected to both cation and anion exchange chromatography (HiTrap SP FF and HiTrap Q FF respectively from GE Healthcare). The flow-through in each case contained the AEBP2 zinc finger protein, which was quantified using the Bio-Rad Bradford assay and was stored at -80°C. Later, as shown in Figure 3.6, 1.2 mg purified AEBP-ZF was loaded onto a Superdex 75 10/300 GL column (GE Healthcare) and 250 µl fractions were collected. Fractions 39-42 were taken as the "contaminants" whilst fractions 48-55 were taken as "AEBP2-ZF GF".

After purification using its His-tag, the full length fragment was purified using its Strep tag. The eluate was applied to a column containing 250 µl StrepTactin Superflow beads (IBA). Subsequently the flow through was applied to the same column twice more. The beads were then washed in Strep wash buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl) and protein was eluted in 250 µl fractions in Strep elution buffer (100 mM Tris-Hcl pH 8.0, 150 mM NaCl. 2.5 mM Desthiobiotin). Fractions were analysed by SDS-PAGE and protein-containing fractions combined were quantified by the Bio-Rad Bradford assay and dialysed into BC100 buffer.

2.16.2 Preparing radiolabelled oligonucleotide probes for EMSA

5 μg primers containing the sequence of interest (AE1, T1, MNM, tcga and tcga mut in Table 11) flanked by M13F and M13R binding sites were incubated with an equimolar amount of M13F primer at 100°C, and slowly cooled down to 23°C. After annealing, the primer was extended by 6 U Klenow polymerase (NEB) in the presence of 2 mM dNTPs.

Probe name	Sequence
AE1	TGTAAAACGACGGCCAGTCCAGGGAGAACCAAAGTTGAGAA ATTTCTATTAAAAACGGTCATAGCTGTTTCCTG
T1	TGTAAAACGACGGCCAGTCTTTAGTTAACTTCCTCTCTGTCT GCAGGCCACTCCAGGGTCATAGCTGTTTCCTG
MNM	TGTAAAACGACGGCCAGTNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNGGTCATAGCTGTTTCCTG
tcga	TGTAAAACGACGGCCAGTGTCTAATGTTTCAAATGAAAGTTCGAG CTTACGATTACGGTCATAGCTGTTTCCTG
tcga mutant	TGTAAAACGACGGCCAGTGTCTAATGTATCAAATGTAAGTTCGAG CTTACGATTACGGTCATAGCTGTTTCCTG

Table 11: Sequences of probes used in EMSA.

The products were purified using the gel extraction protocol from Qiagen and analysed on a 10% Tris Boric acid (TB) polyacrylamide gel, alongside a low mass marker (Invitrogen) to estimate the concentration. Subsequently, 300 ng probe was mixed with 2 μl ^{32}P-γ-ATP and 30 U T4 polynucleotide kinase (PNK) (NEB) and incubated at 37°C for 1 hr.

2.16.3 EMSA

Proteins were mixed with 250 ng polydeoxyadenic polydeoxythymidinic acid (polydAdT) (Sigma) in binding buffer (16% Ficoll 400, 80 mM HEPES-KOH pH 7.9, 600 mM KCl, 2 mM DTT) also containing 1 μg BSA. This mixture was incubated at 23°C for 15 minutes. Subsequently a 5:3 radiolabelled probe:0.01% bromophenol blue mixture was added, mixed and incubated at 23°C for a further 25 minutes. The reactions were subsequently loaded on a 5% native 0.2x TB polyacrylamide gel (zinc finger

fragment) or a 1.2% native 0.2x TB agarose gel (full length protein). Reactions were run at 4°C until the dye front had migrated 9.5 and 7 cm into the gel respectively. The polyacrylamide gel was dried onto Whatman paper for 1 hour at 80°C *in vacuo*, whilst the agarose gel was dried onto DE81 anion exchange paper for 2 hours at 60°C *in vacuo*. The dried gels were then exposed to a Fuji MS-type imaging plate overnight and scanned on a FLA-7000 scanner (GE Healthcare).

2.16.4 Performing SELEX

The first round of SELEX was performed as a conventional EMSA, except that the gel was finally exposed to Biomax XAR film (Kodak) at -80°C. The film was then used as a guide to excise the bands shifted by AEBP2 binding. In the case of the full length fragment, the gel was carefully separated from the DE81 paper and the paper was boiled in 100 µl water at 100°C for 10 minutes, after which the DNA was isolated by ethanol precipitation and resuspended in 25 µl, 5 of which was used for a subsequent PCR reaction. In the case of the zinc finger fragment, the Whatman paper was boiled at 100°C for 10 minutes in 100 µl water, 5 of which was used for a subsequent reaction. Each round of PCR also included a water control to ensure that no contamination was present. The PCR reaction consisted of 95°C for 5 minutes, followed by 24 cycles of 95°C for 40 seconds, 60°C for 40 seconds and 72°C for 10 seconds. An 8 min extension period at 72°C preceded the addition of 4 µl 10 µM M13F and M13R, 2 µl 2 mM dNTPs and 0.5 U Phusion. A cycle for 95°C for 5 minutes, 64°C for 2 minutes and 72°C for 10 minutes followed. The PCR products were purified using the gel extraction protocol from Qiagen and analysed on a 10% TB polyacrylamide gel. It was found that the quantity of input DNA was very important, as very low amounts resulted in the production of slower migrating DNA bands that appeared to be multimers. Correctly sized products were labelled and used in a subsequent round of selection.

In total, six rounds of selection were performed. Finally, adenosine bases were added to PCR products by incubating 24 µl PCR product with 0.33 mM dATP and 5 U *Taq* polymerase (Invitrogen) in *Taq* polymerase buffer. Probes were subsequently cloned into the pGEM-T-Easy vector (Promega) according to the manufacturer's instructions. For each repeat of the SELEX experiment, a minimum of 50 clones were sequenced. As a control for PCR bias, 20 clones derived from probes that had been simply amplified during each round (without selection for binding activity) were also sequenced. Furthermore, probes from the input ("MNM") were also sequenced to ensure it was not biased. Sequences were analysed for enrichment using the MEME program (Bailey and Elkan, 1994).

2.17 Generating the AEBP2 antibody

2.17.1 Antibody production by Eurogentec

AEBP2-FL was purified as described in 2.16.1. Antigen was sent to Eurogentec, who used 0.75 mg per injection per rabbit (4 injections in total) in the 28-day "Speedy protocol". Bleeds were stored at -80°C.

2.17.2 Antigen-based purification of the AEBP2 antibody

500 µl well-resuspended Affigel-15 (Bio-Rad) was washed with 1 ml water and twice with 1ml Phosphate Buffered Saline (PBS, 160 mM NaCl, 3 mM KCl, 8 mM Na_2HPO_4, 1 mM KH_2PO_4, pH 7.3). 5 µg AEBP2-FL in 750 µl BC100 was added to the washed Affigel and incubated at 4ºC overnight whilst rotating. The next day the flow-through was collected and analysed on an SDS-PAGE gel to ensure that all antigen had bound to the Affigel. Simultaneously, Affigel was washed with 1 ml 0.2 M ethanolamine and

subsequently blocked with 1 ml 0.2 M ethanolamine at 4ºC for 2 hours. The mixture was loaded onto a 10 ml Econopac Chromatography Column (Bio-Rad) and washed with 25 ml ice-cold PBS.

Then the antigen-based column was incubated with 10 ml serum from the final bleed in the capped column and rotated at room temperature for 3 hours. The column was washed with 25 ml 0.5 M NaCl and 25 ml PBS. Antibodies were eluted with 500 µl 0.1 M glycine pH 3.0. Elutions were collected in eppendorf tubes containing 50 µl 1M Tris pH8.0 and mixed immediately to neutralise the pH. Antibody-containing elutions were combined and frozen in 10% glycerol, 0.5 M NaCl.

2.17.3 Protein A-based purification of the AEBP2 antibody

5 ml antiserum was thawed on ice and cleared of precipitate by centrifugation at 15,000*g* for 5 minutes. The supernatant was filtered through a 0.22 µm non-protein binding filter and 0.5 ml 1M Tris-HCl pH8.0 was added. This mixture was applied to a 15 ml Falcon Tube containing 1 ml rProtein A sepharose beads (GE Healthcare), which had been washed three times with 10 ml PBS. After incubation with rotation for 2 hours at room temperature, the mixture was loaded onto a 10 ml Econopac Chromatography Column (Bio-Rad). The column was washed twice with 5 ml ice-cold 100 mM Tris-HCl pH 8.0 and then twice with 5 ml ice-cold 10 mM Tris-HCl pH 8.0. Antibodies were eluted with twelve 500 µl elutions with 0.1 M glycine pH 3.0. Elutions were collected in eppendorf tubes containing 50 µl 1M Tris pH 8.0 and mixed immediately to neutralise the pH. Eluted samples were analysed by SDS-PAGE and antibody-containing fractions 1-8 were combined and diluted with 100 mM Tris-HCl pH 8.0 until the mixture became clear and stored at -20 ºC in 10% glycerol and 50 mM NaCl.

2.18 Targeting *Aebp2* alleles

2.18.1 Targeting the *Aebp2* locus

40 µg targeting vector (described in 2.5.4) was linearised by overnight SfiI digestion. DNA was purified by phenol-chloroform extraction and resuspended in 30 µl water. Medium was changed on cells growing on a minimum of two T25 plates 2 hours before they were harvested and counted. 1.5 x 10^7 were resuspended in 0.8 ml PBS to which 20 µg linearised vector was added. Cells and vector were mixed and transferred to a pre-chilled 0.4 cm electrode gap electroporation Gene Pulser cuvette (Bio-Rad). The mixture was incubated at 4 °C for 5 min and electroporated at 200 V, 950 µF. Electroporated cells were then incubated at 4 °C for a further 5 min and then transferred to 6.2 ml ES medium. The 7 ml mixture was then divided across 7 90 mm dishes. The following day, medium was changed, and the day after that selection by geneticin was initiated at 400 µg/ml. When geneticin-resistant colonies became of sufficient size (usually around 11 days after initiation of selection), colonies were picked under a microscope into geneticin-containing medium in 96-well plates. Lines were expanded and correct targeting tested by Southern Blot. When targeting the second allele, many more colonies appeared during targeting yet many died after picking, suggesting they were still somewhat resistant to geneticin. Of the colonies that survived, many fewer had the correct targeting. This could be due to the presence of the inverted cassette in the first targeted allele. Low level transcription across the inverted cassette containing the *neo* gene could generate a low level of geneticin resistance.

2.18.2 Extracting genomic DNA from ES cells

Cells from a confluent T75 were harvested and resuspended in 1-3 ml lysis buffer (10 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM EDTA-NaOH pH 8.0, 0.5% Sodium lauroyl

sarcosinate) and proteinase K was added to a final concentration of 200 µg/ml. Samples were incubated overnight at 55 °C. Then, equal volume of phenol:chloroform:isoamyl alcohol 25:24:1 (Sigma) was added and the mixture was incubated for 1 hr at 25 °C with agitation. After centrifugation (3600*g*, 10 min, 25 °C), the aqueous phase was transferred to a clean tube. 1/25 volume of 5 M NaCl and 2.5 volume of 100% ethanol was added. After mixing, a visible white cloud of DNA was extracted using a bent pipette tip and transferred to a clean tube containing 1 ml 70% ethanol. DNA was pelleted (16,100*g*, 5 min, 4 °C) and air dried. Subsequently the pellet was resuspended in 300-400 µl 10 mM Tris pH 8.5 and the concentration measured by Nanodrop.

2.18.3 Southern Blotting

20 µg DNA was digested overnight in a 50 µl reaction with 17.5 U MscI and 15 U KpnI (5' Southern Blot and GFP Southern Blot), 30 U BclI (3' Southern Blot), or 50 U SpeI (cassette inversion Southern blot). Digestions were loaded onto a 1x TAE 0.8% agarose gel containing 0.1 µg/ml ethidium bromide (Fisher Scientific). The gel was run at 80 V for a minimum of 6 hours. All gel treatments were performed at 25 °C with agitation. The gel was rinsed in water and then washed twice in Depurination Solution (0.125 M HCl) for 5 min and then twice with water for 4 min. Subsequently the gel was incubated for 45 min with Denaturation Buffer (0.5 M NaOH, 1.5 M NaCl). The gel was then washed twice with Neutralisation buffer (0.5 M Tris-HCl pH 7.5, 1.5 M NaCl) and rinsed in water. DNA was then transferred to Hybond-XL (GE Healthcare) by capillary transfer with 20x SSC (3 M NaCl, 0.3 M Sodium Citrate dehydrate (Fisher Scientific), pH 7.0) for a minimum of 16 hrs. Next, the membranes were washed in 2x SSC for 5 min, and then allowed to dry for 1 hr, after which DNA was cross-linked to the membrane using the "optimal crosslink" setting on Stratalinker (which is 120,000 microjoules/cm^2). Membranes were then pre-incubated with pre-heated Church Buffer (0.25 M sodium

phosphate buffer pH 7.2, 7% SDS, 1 mM EDTA-NaOH pH 8.0) at 65 °C with rotation. Probes (Table 12) were generated by PCR and purified by QIAquick gel extraction followed by QIAquick PCR purification (Qiagen). 30 ng probe was labelled with ^{32}P using the Prime-It II kit (Stratagene). Radiolabelled probes were added in fresh Church Buffer and incubated overnight at 65 °C with rotation. Subsequently blots were washed once at 65 °C with Low Stringency Wash Solution (2xSSC, 0.1% SDS) and once at 25 °C with Low Stringency Wash Solution, before being exposed to a Fuji MS-type imaging plate and imaged using a FLA-7000 scanner (GE Healthcare).

Probe	**Genomic coordinates (on chromosome 6)**
5' SB	140572805 - 140573069
3' SB	140590603 - 140590886
Cassette inversion (orientation probe)	140581942 - 140582206

Table 12: Coordinates of Southern Blot probes used in this study. For the GFP blot, a 271 bp probe was generated using the GFPfw primer 5'-GGTGAGCAAGGGCGAGGAGC-3' and the GFPrev primer 5'-TTCGGGCATGGCGGACTT-3'.

2.18.4 Nucleofection of the *FlpE* expressing plasmid

Targeted cells were grown in T25 flasks and medium-changed several hours before nucleofection. Nucleofection was performed according to instructions in the Mouse ES Cell Nucleofector® Kit (Lonza). Cells were harvested (note that feeder cells were not separated from ES cells), washed in PBS and counted. 4-5x10^6 cells were resuspended in 95 µl Mouse ES Cell Nucleofector® Solution which had been pre-warmed to room temperature. 5.7 µg pCAG-FlpE-puro plasmid in 5 µl was added and mixed. The mixture was transferred to a cuvette either from the kit or a BioRad Genepulser® cuvette 2 mm (catalogue number 165-2086) and nucleofected using programme A-024 on Amaxa® Nucleofector® II device (Lonza). Immediately after nucleofection, 500 µl of culture medium pre-warmed to 37 °C was added and all cells (except for a clump of dead cells) were added to a 90 mm dish. After 2 days, cells were counted and 400, 1,000 and 2,500

cells were plated onto 90 mm dishes. After 8 days colonies were picked under a microscope and transferred to 96-well plates. Cells were then tested for geneticin sensitivity and geneticin-sensitive clones (occurring at a rate of about 5%) were expanded and tested by Southern Blot.

2.18.5 Cre nucleofection

Cre nucleofection was performed as FlpE nucleofection with the following exceptions: 5.3 µg pTurboCre (pCAGSS plasmid with turbo Cre under the control of CMV enhancer and beta actin promoter, with an NLS) was added in 10 µl to 90 µl Mouse ES Cell Nucleofector® Solution. One day after nucleofection 1,000, 2,500 and 8,000 cells were plated on 90 mm dishes and the following day selection was initiated with 400 µg/ml geneticin. Successful inversion was tested by Southern Blot.

2.18.6 PCR to test for correct inversion and to test heterozygote mice

To test the orientation of the cassette, a combination of primers were used. Configurations are indicated in Figure 2.1. Primer sequences, and an indication with which PCR protocol they were used, are in Table 13. See also 2.19.

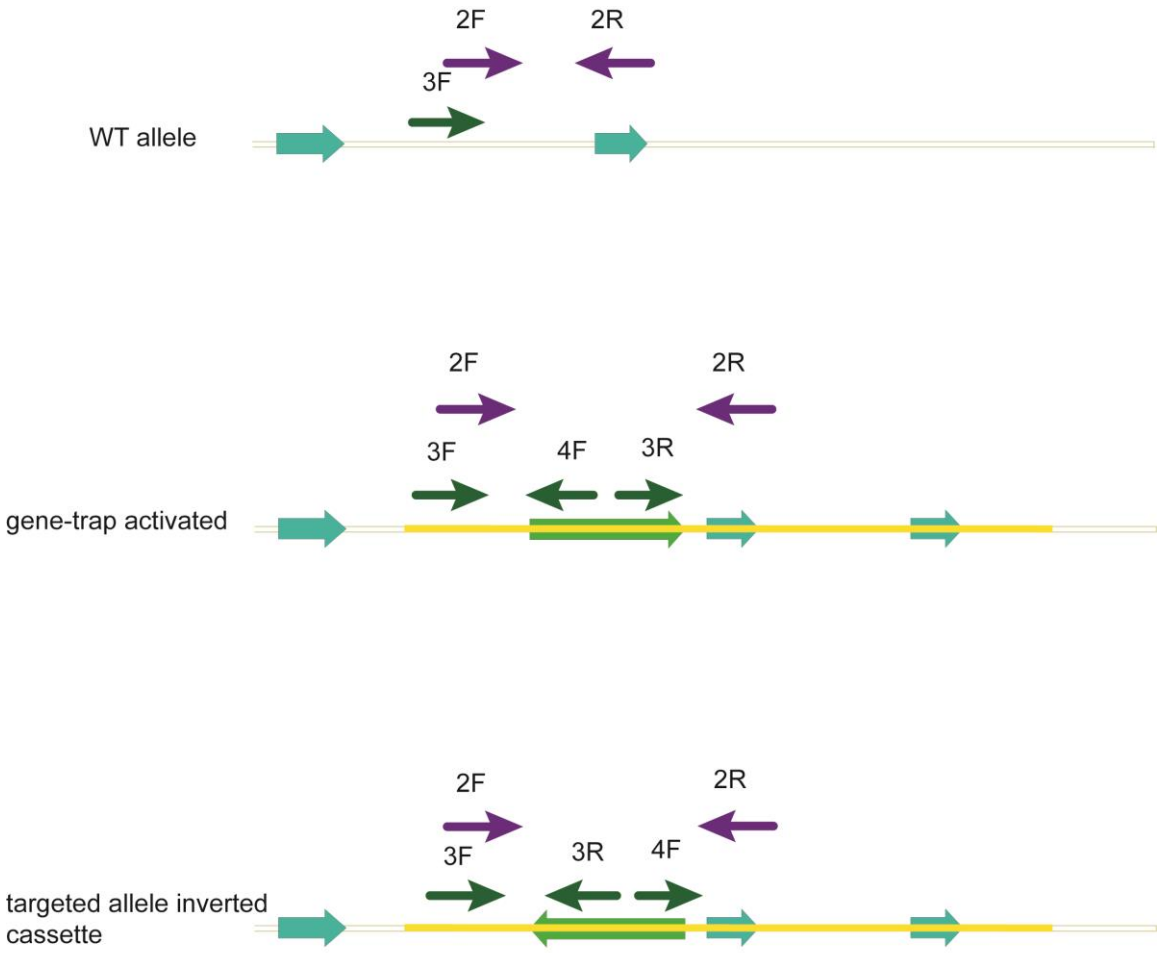


Figure 2.1: Locations of binding sites in the different configurations of the cassette in the *Aebp2* locus.

Primer name	Primer sequence	PCR protocol
casor_2F	5'-ATAAAAATGTGAGGAGCCAGATGTG-3'	Expand Long Template
casor_2R	5'-TGGCTTCTTAGGGACTTGCT-3'	Expand Long Template
casor_3F	5'-TCGTTCAAAGCTGCCCTCTCAAGC-3'	Expand High Fidelity
casor_3R	5'-CGCATCGCCTTCTATCGCCT-3'	Expand High Fidelity
casor_4F	5'-GTCGATCCCCACTGGAAAGA-3'	Expand High Fidelity

Table 13: Primers used to analyse presence and orientation of the gene-trapping cassette in the *Aebp2* locus.

2.19 Mouse work

All licensed animal procedures were carried out by Jonathan Godwin at the University of Oxford under the United Kingdom Home Office ASPA project, licence number 30/2800. 10-20 ES cells were injected into each cavitated blastocyst (x50) and allowed to recover (>3 hours) before unilateral implantation into the uterus of a B6CBF1 (C57Bl/6 female crossed to CBA male) hybrid foster mother by laparotomy. These foster mothers littered approximately 18 days after implantation and the resultant high percentage chimeric offspring (most 129 colour) were mated to C57BL/6 mice. The litters (129 colour pups) from these breeders were genotyped by removal of a small ear punch. DNA was isolated using the PureLink™ genomic DNA mini kit (Life Technologies). Purified DNA was used as a template for PCRs with primers casor_2F and casor_2R (Table 13) using the Expand Long Template protocol.

3. DNA-binding activity of AEBP2

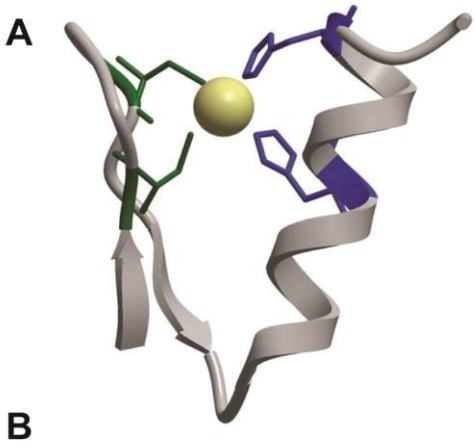
3.1 Introduction

The only conserved protein domain in AEBP2 is a zinc finger domain, which consists of 3 C2H2 zinc fingers. Because zinc finger domains can bind to DNA in a sequence specific manner (reviewed in Wolfe et al., 2000; Klug, 2010), it has been proposed that AEBP2 could be a sequence-specific PRC2 recruiter (Kim et al., 2009a). This hypothesis is supported by the fact that AEBP2 was originally identified as a DNA binder of the AE-1 sequence (He et al., 1999). Therefore I set out to identify any specific sequences that AEBP2 might bind to.

3.1.1 Zinc finger domains binding to DNA

Zinc fingers have a conserved structure consisting of two β sheets followed by an α helix. A zinc ion is used to stabilise the folding of the zinc finger (Figure 3.1). The zinc ion is coordinated by cysteine and histidine amino acids that are essential for a functional zinc finger. There are different families of zinc finger protein, using different consensus amino acid sequences to coordinate the zinc ion. The first zinc finger domain was discovered in TFIIIA, which contains Cys_2His_2 zinc fingers (Brown et al., 1985; Miller et al., 1985). The consensus for these zinc fingers is (F/Y)-X-C-X_{2-5}-C-X_3-(F/Y)-X_5-ψ-X_2-H-X_{3-5}-H, where X is any amino acid and ψ is a hydrophobic residue (reviewed in Wolfe et al., 2000). All three of AEBP2's zinc fingers conform to this consensus (indicated in red in Figure 3.1B), with the exception of the second zinc finger, which has an extended loop (indicated in blue in Figure 3.1B). AEBP2's zinc fingers are most closely related to

those of the Gli-Krüppel family. Whilst the cysteine and histidine and hydrophobic amino acids are required for structural stabilisation, the amino acids on the surface of the alpha helix can contact specific bases in the major groove of double stranded DNA, thus conferring sequence specificity. However, recent studies also suggest that zinc fingers can be used as a module for protein-protein interactions or for RNA binding (reviewed in Hall, 2005; Burdach et al., 2012).



Cys2His2 consensus	(F/Y)-X-C-XXXXX-C--X---XX-(F/Y)----XXXXX-ψ-XX-H-XXXXX-H
Gli-Krüppel consensus	(F/Y)-X-C-XXXG--C-(S/T)KX-(F/Y)----XXXXX-L-XX-H-XXXX--H
AEBP2 zinc finger 1	Y-----N-C-CWDQ--C--Q---AC--F-------NSSPD-L-AD-H-IRSI--H
AEBP2 zinc finger 2	F-----V-C-LWKG--C------KV--Y---NTPSTSQSW-L-QR-H-MLT---H
AEBP2 zinc finger 3	F-----K-C-VVGG--C--N---AS--F-------ASQGG-L-AR-H-VPT---H

Figure 3.1: Typical structure of Cys_2His_2 zinc fingers. (A) The structure of a typical zinc finger consists of two β sheets followed by an α helix, with cysteine and histidine amino acids coordinating the zinc. Reproduced with permission from Brayer et al. (2008). The cysteines are indicated in green and the histidines in blue. The zinc ion is represented by a yellow ball. (B) Overview of the consensus Cys_2His_2 zinc finger, those of the Gli-Krüppel family and the zinc fingers of AEBP2. Consensus amino acids are indicated in red.

3.1.2 Is AEBP2 a sequence specific DNA binding factor?

AEBP2 was identified as a binder of the AE-1 sequence (He et al., 1999). The AE-1 sequence is a 34 nucleotide sequence located 159 bp upstream of the adipose *P2* gene (*aP2*, also known as *Fabp4*), which is a metabolic gene thought to be important during adipocyte differentiation. This sequence had previously been shown to be bound by

CCAAT enhancer binding protein α (C/EBPα) (Christy et al., 1989; Herrera et al., 1989) and AEBP1 (He et al., 1995). AEBP2 was found by screening a library of (partial) cDNAs based on mRNAs expressed in 3T3-L1 preadipocytes. In this experiment, cDNAs were expressed in *Escherichia coli* and transferred to a protein filter and tested for the ability to bind to AE-1 concatamers. After identification of AEBP2, the authors showed that the shorter 34 kDa isoform of AEBP2 could bind to AE-1 probes by electromobility shift assays (EMSA), and that this interaction appeared to be specific as it could not be reduced by competition from probes carrying different sequences. Interestingly, they found that, after disrupting the second finger by mutating the second histidine and the following serine (H315 and S316 in full length AEBP2), binding was not abolished but that the mutated AEBP2 could no longer induce silencing of a reporter construct transfected into the mouse embryonic fibroblast line NIH 3T3.

The DNA binding abilities of AEBP2 were further investigated in a following study (Kim et al., 2009a). The authors expressed Glutathione S-transferase (GST) fusions of AEBP2 full length, the 34 kDa isoform, and a construct that contained amino acids 223-348 of the full length AEBP2 which contains the 32 C-terminal amino acids of the S-rich domain and the entire zinc finger domain (see Figure 1.8). They could not detect any DNA binding activity of full length AEBP2, but did detect binding for the other two smaller recombinant proteins. They then used the GST-fused 34 kDa isoform to characterise AEBP2 binding sequences. They found that this construct bound well to the T1 sequence (see also Figure 3.3). Subsequently, the authors made a series of mutations in the T1 sequence in order to elucidate which regions might be important and proposed that a sequence of CTT and GCC, separated by a 15-21 nucleotide spacer region, was

important for AEBP2 binding. The authors proposed that AEBP2 might bind to specific sequences and thus recruit PRC2 to target genes.

However, comparison of the AE-1 and T1 sequence, which were both reported to bind AEBP2, indicates little overlap, leaving it unclear what (if any) AEBP2's preferred binding sequence is. In *Drosophila melanogaster*, PRC2 appears to be recruited by sequence specific proteins, whilst no mammalian homologues have been found (as discussed in the Introduction). It could be that sequence-specific recruitment of PRC2 is mediated by AEBP2, as no other sequence specific binding activities have been reported for PRC2 core components. Therefore, I decided to analyse the ability of AEBP2 to bind to specific DNA sequences.

3.2 Purification of full length AEBP2 and Zinc Finger only proteins

Recombinant full length AEBP2, corresponding to exons 1b-8, was expressed with an N-terminal His-tag and a C-terminal Strep tag and purified from *E. coli* bacterial cells (Figure 3.2). Although the predicted molecular weight of this isoform of mouse AEBP2 is 54 kDa, the recombinant full length AEBP2 (AEBP2-FL) protein migrated with an apparent molecular weight of ~73 kDa. However, Cao et al. (2002), who identified AEBP2 as a PRC2 associating factor based on a biochemical purification of the endogenous PRC2 complex from HeLa cells, also found that AEBP2 migrated at higher molecular weight than expected, at a similar position as the recombinant full length AEBP2 I produced. I could not obtain soluble protein from the 34 kDa isoform used by Kim et al. (2009a) to identify AEBP2 binding sequences, even when fused to GST or

Maltose Binding Protein (MBP) (data not shown). I also used a zinc finger domain only (AEBP2-ZF) recombinant protein. This protein contained amino acids 223-348 of the full length AEBP2 which corresponds to the part of the AEBP2 protein that was used by Kim et al. (2009a). I used an N-terminal 6xHis tag to purify this protein, and subsequently purity was increased by passing the eluate over a cation and an anion exchange column, neither of which retained the AEBP2-ZF. The final eluate was assessed for purity and protein concentration by Coomassie staining (Figure 3.2).

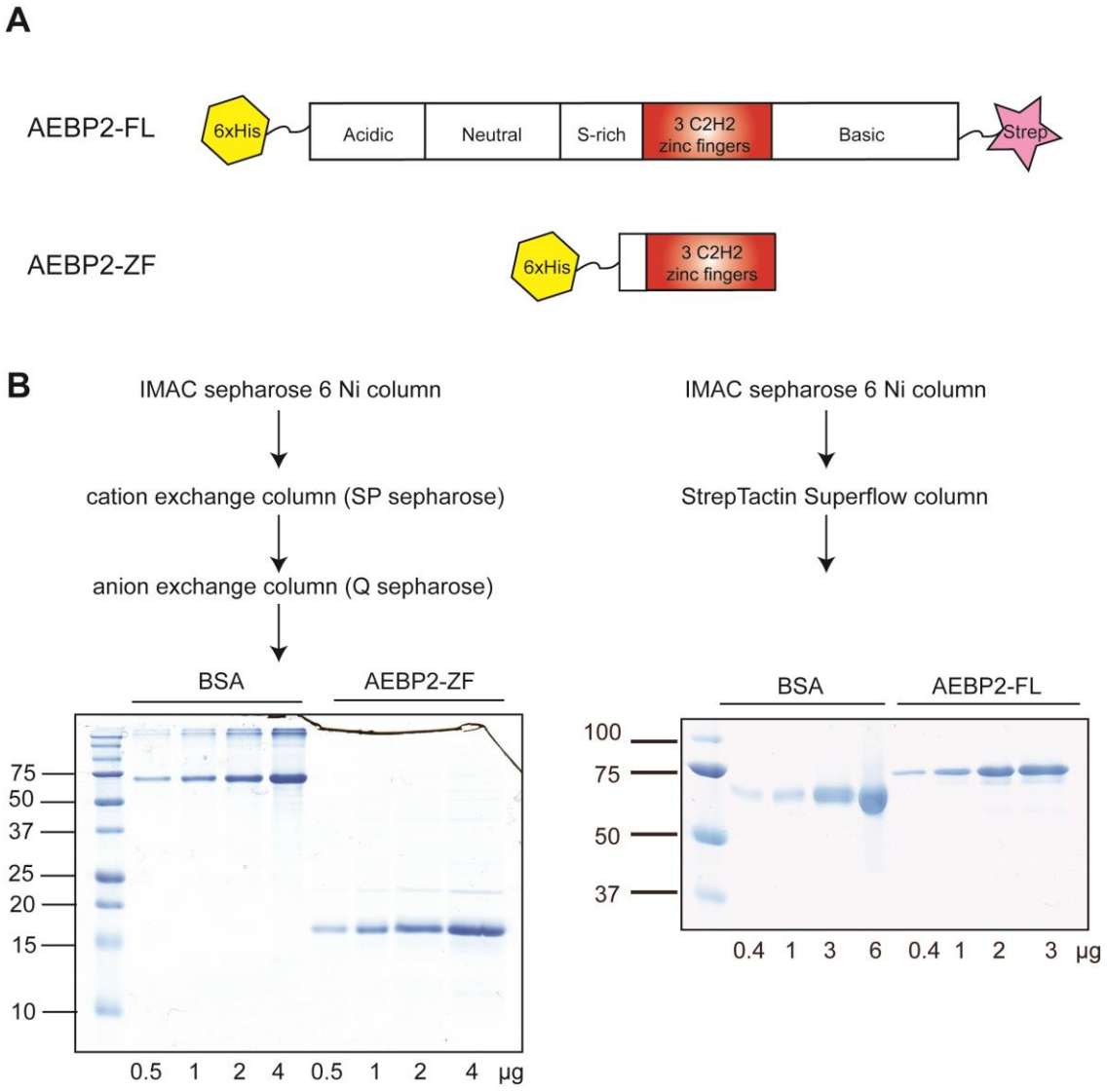


Figure 3.2 Purification of AEBP2-ZF. (A) Schematic of AEBP2-FL and AEBP2-ZF recombinant proteins. (B) Purification scheme and quantification of AEBP2-ZF and AEBP2-FL. Bovine Serum Albumin (BSA) used as a control to quantify protein amount. Molecular weight markers in kDa.

3.3 AEBP2-FL and AEBP2-ZF bind to the AE-1 and T1 sequences

I assessed whether the AEBP2-FL and AEBP2-ZF could bind to the AE1 and T1 DNA sequences using electromobility shift assays as previously published (He et al., 1999; Kim et al., 2009a) (Figure 3.3A). I found that the AEBP2-FL did not appear to migrate in a native polyacrylamide gel, and instead a strong radioactive signal from the radio-labelled DNA probe was found in the wells. This could be caused by aggregated probe and unfolded protein. In order to increase the availability of zinc which might be required for the stability of the zinc fingers and therefore the whole protein, I added $ZnCl_2$ to the binding reaction, which was found essential for success by the original report that the 34 kDa isoform of AEBP2 is a DNA binder (He et al., 1999). However, this addition did not overcome the problem, consistent with findings by Kim and colleagues who also did not find any binding activity by EMSA using polyacrylamide gels. Furthermore, analysis by gel filtration showed that AEBP2-FL eluted at defined fractions, and not in the void, suggesting that AEBP2-FL protein does not aggregate (data not shown). Subsequently, DNA-AEBP2-FL complexes were separated on an agarose gel, which can separate complexes of larger sizes than polyacrylamide gels. AEBP2-FL could bind to both the AE1 and T1 sequences (Figure 3.3B). AEBP2-ZF did migrate into a native polyacrylamide gel, and was shown to also bind both sequences.

A

Name	Sequence
AE1	CCAGGGAGAACCAAAGTTGAGAAATTTCTATTAAAAAC
T1	CTTTAGTTAACTTCCTCTCTGTCTGCAGGCCACTCCAG

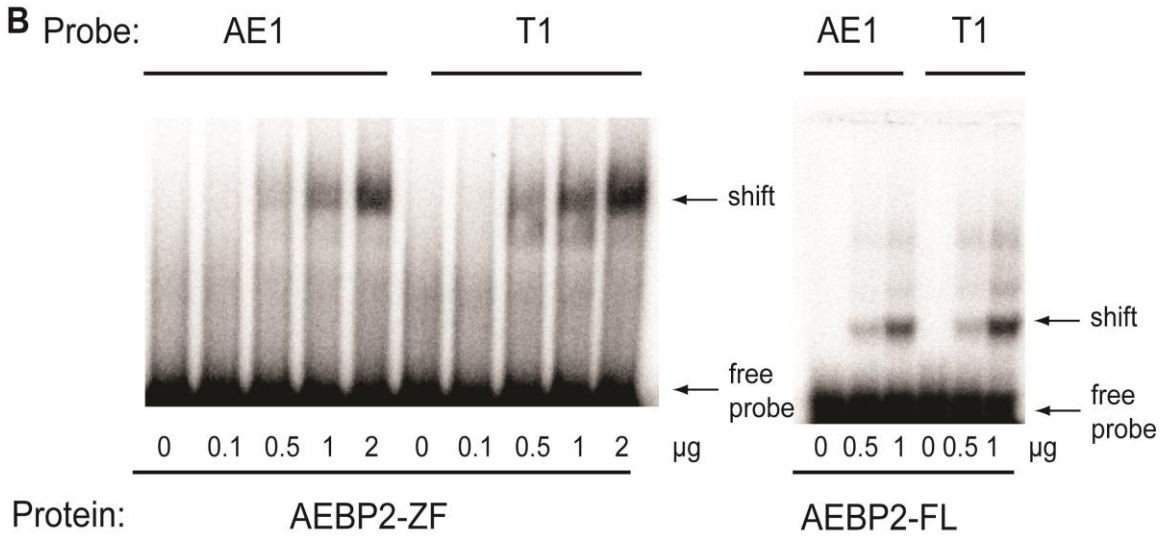


Figure 3.3: AEBP2-ZF and AEBP2-FL can bind to the AE1 and T1 DNA probes. (A) Sequences of the AE1 and T1 DNA probes. Each probe was flanked by M13F and M13R sequences, which are not included in the sequence presented here. (B) EMSAs for AEBP2-ZF and AEBP2-FL. Both proteins can bind to both probes.

3.4 AEBP2-ZF appears to bind a specific DNA sequence

As the AE1 and T1 sequences do not bear obvious similarities, I wondered whether there might be a DNA sequence that AEBP2 binds to with higher affinity, and may therefore represent the "true" AEBP2 binding sequence. To investigate this I initiated a Systematic Evolution of Ligands by Exponential enrichment (SELEX) experiment, in which AEBP2-binding sequences were purified from randomised DNA probe over six rounds of EMSAs (Figure 3.4A). Sequencing of the input for the first round confirmed that the sequences were random. It was found that the amount of probe that bound to the recombinant protein increased over the rounds, and by the 6th round also bound to more of the SELEX-generated DNA probe than to either AE1 or T1 (Figure 3.4B).

Furthermore, binding between AEBP2-ZF and 6th round probe was inhibited much less by competition from the random probe or probe from SELEX round 3, or the T1 and AE1 probes, suggesting that AEBP2-ZF bound 6th round probe with the highest affinity (Figure 3.4C).

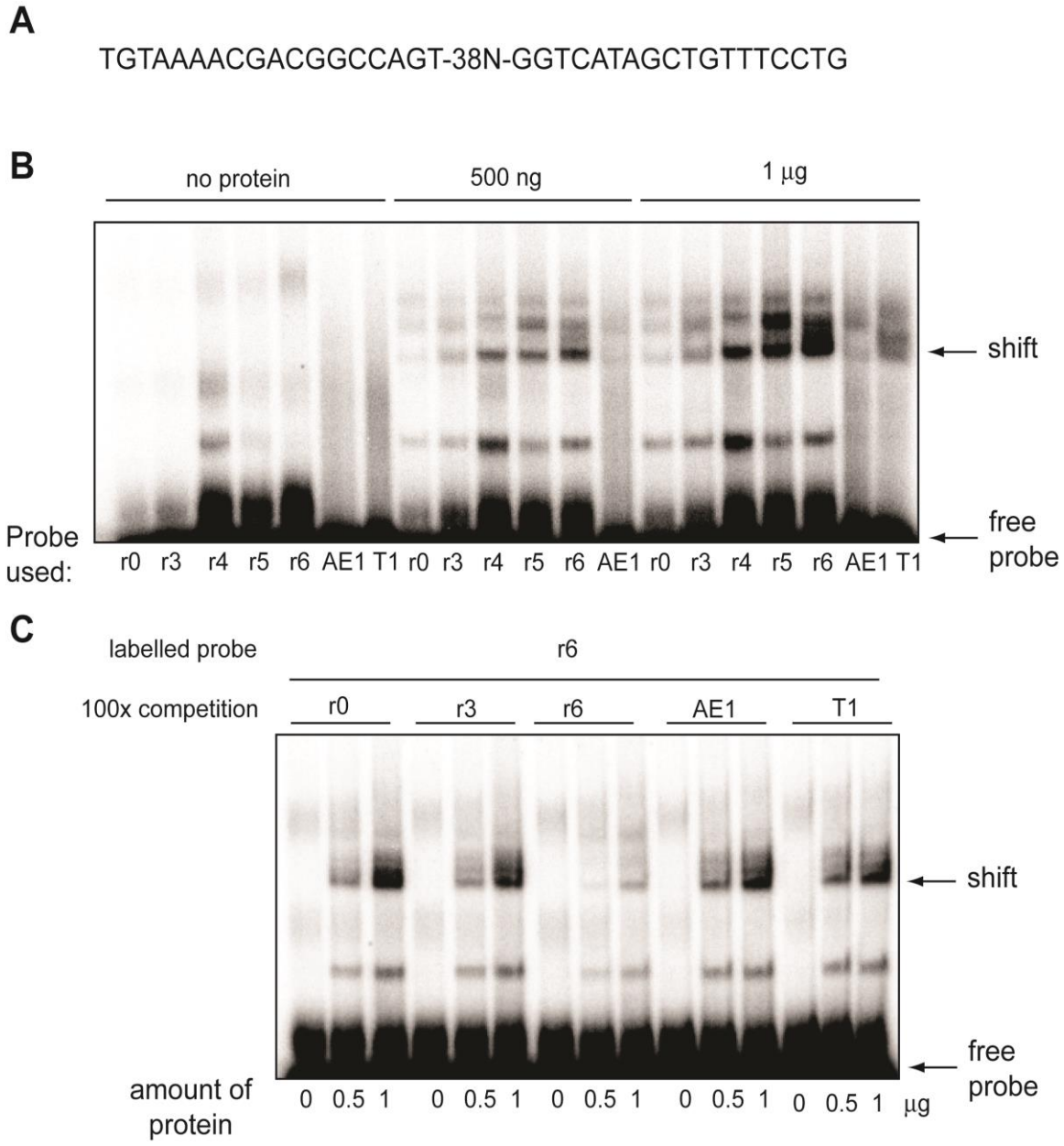


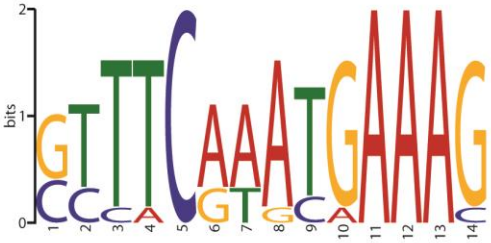
Figure 3.4: SELEX enriches for DNA sequences that AEBP2-ZF binds to. (A) The probe used as input for the SELEX experiment consisted of the M13F and M13R primer binding sides flanking 38 random nucleotides. (B) EMSAs for AEBP2-ZF and AEBP2-FL. r0 is random probes that were the input. r3 is probe from third round, etc. AEBP2-FL binds probe from 6th round with highest affinity (most probe shifted). (C) Competition experiment showing that probe from round 6 can only be outcompeted by itself, suggesting this probe has the highest affinity.

After six rounds, probes were cloned and sequenced and entered into the MEME (Multiple Em for Motif Elucidation) program for detection of sequence enrichment (Bailey and Elkan, 1994). As a control for PCR amplification bias I also amplified six times a randomised DNA probe that had not been subjected to EMSA.

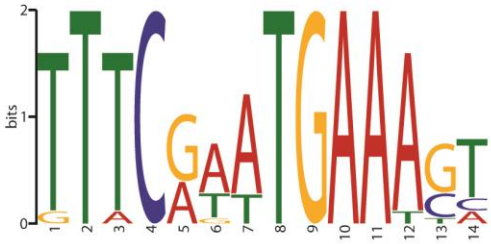
Analysis of the control probes showed a bias towards AT-rich sequences, but no consensus sequence. Analysis of the probes obtained from SELEX with the zinc finger fragment showed enrichment for a sequence GTTTCAAAGAAAG. I repeated the SELEX experiment with a fresh pool of randomised probe and freshly purified protein and obtained the same result (Figure 3.5A). I subsequently designed a probe which either contained the binding sequence embedded in a sequence not enriched during SELEX, or the same sequence but with 2 mutations (Figure 3.5B, mutations indicated in red). These mutations were chosen based on their information content as calculated by the MEME programme (height of letters). When I tested the ability of AEBP2 to bind to these probes I found that, as predicted, AEBP2 could bind to the probe containing the sequence, but not to the mutated probe (Figure 3.5C).

Binding did not increase over the course of the SELEX for full length AEBP2 (data not shown). Surprisingly, it appeared that 100x competition of unlabelled probe could not outcompete binding of labelled probe (data not shown), suggesting that the interaction between full length AEBP2 and DNA might be due to non-physiological aggregation, possibly consistent with the finding that complexes could not be separated on a polyacrylamide gel. These observations were further supported by the finding that there was no enrichment for specific DNA sequences.

A



Sequence logo from
experiment 1
e-value: 8.6e-17



Sequence logo from
experiment 2
e-value:3.0e-29

B

Name	Sequence
tcga	TGTAAAACGACGGCCAGTGTCTAATGTTTCAAATGAAAGTTCG AGCTTACGATTACGGTCATAGCTGTTTCCTG
tcga mut	TGTAAAACGACGGCCAGTGTCTAATGTATCAAATGTAAGTTCG AGCTTACGATTACGGTCATAGCTGTTTCCTG

C

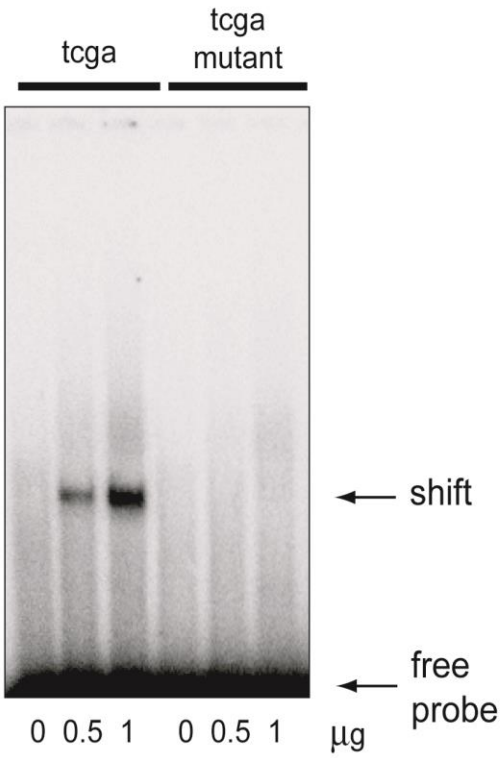


Figure 3.5: SELEX enriched for a specific DNA sequence. (A) Sequencing 50 probes from two repeat experiments produces very similar sequence logos. (B) DNA probes designed to either contain the consensus sequence (tcga) or a mutation in the consensus sequence (tcga mutant). The consensus sequence is shown in blue with mutations in red. (C) The AEBP2-ZF binds to the probe containing the consensus sequence, but not the mutant sequence.

3.5 Specificity of the AEBP2-ZF-DNA binding reaction

In an attempt to show that AEBP2 was the protein responsible for generating the shift of radio-labelled DNA probe in the EMSAs was AEBP2, I added AEBP2 antibody (see 4.2) or His-antibody to the reaction mixture. When an antibody binds to the protein-DNA complex, it usually causes the complex size to increase and a concomitant retardation in the native polyacrylamide gel known as a supershift. However, I did not observe such a shift using either antibody (data not shown). Furthermore, an AEBP2-ZF recombinant protein which was engineered to contain a C-terminal Strep tag, and was purified using the Ni-based IMAC sepharose and a StrepTactin column, failed to bind the DNA probes (data not shown). I attempted to purify AEBP2-ZF with mutations in the cysteines in the first zinc finger, which would be hypothesized to abolish DNA binding, but unfortunately this protein was not soluble and could therefore not be purified and used as a control. Further raising concern, the sequence found in SELEX appeared palindromic, which is a feature of sequences bound by Helix-Turn-Helix containing DNA binding proteins (Brennan and Matthews, 1989), and appeared too long to be bound by 3 zinc fingers.

Although it seemed unlikely that binding would be caused by a contaminating bacterial protein, as the protein preparation appeared quite pure, I did notice that there was still a faint, more slowly migrating band visualised in the Coomassie gel, most visible in the lane in which 4 µg was loaded (Figure 3.2B). This faint band could be contaminating bacterial proteins that might be DNA-binding. Therefore, I purified the AEBP2-ZF further by separating this presumably larger protein from AEBP2-ZF using gel filtration. I combined the fractions to generate eluates containing only the contaminating band or the AEBP2-ZF band, and assessed how much of the contaminating protein was usually

present in my AEBP2-ZF protein preparation that had not been purified by gel filtration (Figure 3.6A). From this, I estimated that 2.5 µl of my contaminants preparation contained roughly the same amount of contaminant protein as was present in a 1 µg preparation of AEBP2-ZF prior to gel filtration (input). I showed that the DNA binding activity in the original AEBP2-ZF protein preparation was due to the contaminant and not the AEBP2-ZF (Figure 3.6B). Using this further purified AEBP2-ZF protein preparation I could not detect any binding to the AE1 and T1 sequences (data not shown).

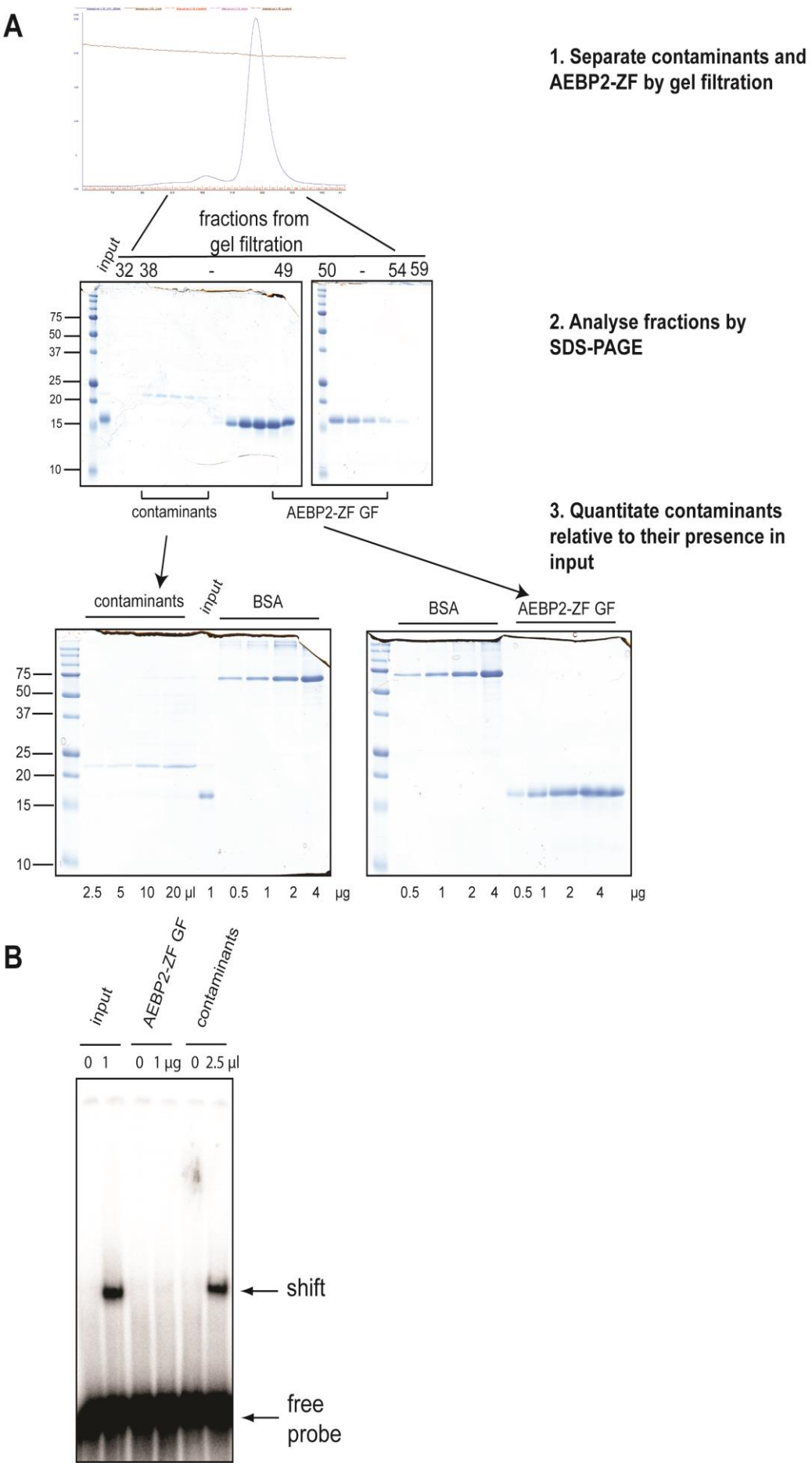

A
1. Separate contaminants and AEBP2-ZF by gel filtration
fractions from gel filtration
input
32 38 - 49 50 - 54 59
75
50
37
25
20
15
10
2. Analyse fractions by SDS-PAGE
contaminants
AEBP2-ZF GF
3. Quantitate contaminants relative to their presence in input
contaminants
input
BSA
BSA
AEBP2-ZF GF
2.5 5 10 20 µl 1 0.5 1 2 4 µg
0.5 1 2 4 0.5 1 2 4 µg
B
input
AEBP2-ZF GF
contaminants
0 1 0 1 µg 0 2.5 µl
shift
free probe

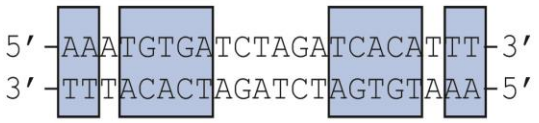
Figure 3.6: Further purification of the AEBP2-ZF protein preparation demonstrates that it is a contaminating protein, and not AEBP2-ZF, that binds to the tcga probe. (A) Further purification of AEBP2-ZF separated the elutions from the gel filtration column into AEBP2-ZF-GF and contaminants. Combined fractions were analysed by SDS-PAGE to estimate the quantity of contaminants present in input. Molecular weights in kDa. (B) Contaminants, but not the AEBP2-ZF, bind the tcga probe.

In an attempt to discover which bacterial protein was responsible for the contaminating DNA binding activity, I analysed the contaminants fraction by mass spectrometry. This fraction contained several proteins, including some known DNA binding proteins (marked with an arrow in Figure 3.7A). However, the published DNA binding sequences for these proteins did not match the sequence found in SELEX. It remains therefore unclear which bacterial protein caused the contaminating DNA binding activity.

A

Protein	Mascot Score
Catabolite gene activator	1770
Catabolite gene activator	1767
GTP cyclohydrolase 1	507
Cob(I)alamin adenolsyltransferase	287
Transcriptional regulator, TetR family	252
3-oxo-acyl-[acyl-carrier-protein] reductase	225
Chloramphenicol acetyltransferase	189
Alkyl hydroperoxide reductase subunit C	162
Thioredoxin	138
Phosphoheptose isomerase	106
FKBP-type peptidyl-prolyl cis-trans isomerase	106
Azoreductase OS=Escherichia coli	100
Superoxide dismutase	92
Spermidine N(1)-acetyltransferase	90
Uncharacterized protein yecA	78
FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase	72
CobW/P47K family protein	67
Uracil phosphoribosyltransferase	65
Glutaredoxin, GrxA family	64

B



C

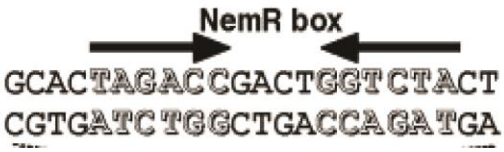


Figure 3.7: Mass spectrometry identifies potential contaminating proteins, but their published binding sequences do not match the consensus sequence found in SELEX. Highest-scoring hits from mass spectrometry analysis of contaminant fractions. (B) Published binding sequence of Catabolite gene activator. Adapted with permission from Lawson et al. (2004). Boxed regions indicate bases especially important for recognition by catabolite gene activator. (C) Published binding sequence of NemR (transcriptional regulator, TetR family in the table in A). Reproduced with permission from Umezawa et al. (2008).

3.6 Discussion

I set out to identify DNA sequences that AEBP2 might bind to. However, experiments showed that the binding of AEBP2-FL was either absent or non-physiological. Binding of AEBP2-ZF was initially observed, but later shown to be caused by bacterial contaminating proteins. Subsequently, no binding of AEBP2-ZF to DNA probes could be detected.

If AEBP2 is a DNA binding protein, it is surprising that no binding of AEBP2-FL could be detected, either in this study or elsewhere (Kim et al., 2009a). A possible explanation for this would be that AEBP2 does not adopt its normal conformation in the absence of PRC2, as it is in EMSAs. It may therefore be interesting to determine whether any DNA binding activity would be detected in the presence of the AEBP2-PRC2 complex. An alternative explanation would be that it is the physiological role of the N-terminal domains to prevent DNA binding. For example, the acidic domain could provide this role as has been shown for acidic domains in other proteins (Cross et al., 2011). Alternatively, the right experimental conditions which would allow AEBP-FL binding to DNA *in vitro* have not been established yet. Whether AEBP2-FL binds DNA therefore remains open for debate.

It is surprising that I did not detect binding of purified AEBP2-ZF to either the AE-1 or T1 sequence, given the results from EMSAs performed in other groups (He et al., 1999; Kim et al., 2009a). However, it should be noted that neither group conclusively showed that it was AEBP2, and not bacterial contaminants, that caused the DNA binding activity observed, as no supershift was shown, nor was there an indication of the purity of their

protein preparations. In the case of Kim et al, the authors describe simply taking the soluble fraction of lysed bacteria expressing GST-AEBP2, so it is almost certain this protein preparation would contain bacterial proteins and may be responsible for the DNA binding observed. Without appropriate controls being present, it is hard to judge the validity of their conclusions.

Despite the lack of controls in these published EMSAs, AEBP2 was originally identified by its DNA-binding capabilities (He et al., 1999), and zinc fingers often do function as DNA binding modules. It could be possible that removal of the C-terminal basic domain leads to a reduced ability of the zinc fingers to interact with DNA, as my protein construct lacked this domain, whilst this domain was present in the experiments performed by Kim and colleagues and He and colleagues. It is also possible I initially optimised the DNA binding conditions in the EMSAs for what later proved to be the bacterial contamination, and that re-optimisation would show DNA binding by AEBP2-ZF.

A final possibility is that the AEBP2 zinc fingers do not in fact bind DNA. As described in 3.1.1, not all zinc fingers bind DNA, and some can bind protein or RNA. This hypothesis is supported by my work, as well as by the observation that mutating the second zinc finger did not abolish observed DNA binding, but did abolish silencing of a reporter construct. The hypothesis that the AEBP2 zinc fingers, or at least not all three of them, are involved in DNA binding is further supported by a recent report that highlighted AEBP2 as a protein containing a tandem C2WH2 (tC2WH2) motif (Hatayama and Aruga, 2010). The name refers to a highly conserved tryptophan residue

two amino acids carboxy-terminal to the first cysteine in two consecutive zinc fingers as is present in zinc fingers 1 and 2 in AEBP2 (see Figure 3.1). These residues appear to stabilise inter-zinc finger interactions by creating a hydrophobic core between two adjacent zinc fingers. Further sequence features of tC2WH2 motifs are additional hydrophobic amino acids to stabilise the hydrophobic core. It was furthermore observed that tC2WH2 motifs often contain extra sequences, as is seen for the extended loop in AEBP2's second zinc finger (discussed in 3.1.1). Based on a survey of tC2WH2 containing proteins, the authors propose that the main function of tC2WH2 is probably not DNA binding, but instead suggest that it may control the relative positioning of other functional domains present in the protein, such as a C-terminally located zinc finger, or could act as a capping structure isolating the structural influence of an N-terminal flanking region.

A structural role for the tC2WH2 motif in AEBP2 is supported by data obtained by Ciferri et al. (2012). Their cross-linking experiments show that lysines located in a region encompassing the second and third zinc finger make many intramolecular contacts with AEBP2, as well as interacting with the SET domain of EZH2 and a single zinc finger in SUZ12. However, when the structure of the PRC2 complex was modelled in the presence of two nucleosomes, the zinc fingers were close to internucleosomal DNA, and the authors suggested that the zinc fingers could be involved in both protein-protein interactions as well as interacting with DNA. This simultaneous dual interaction has been observed for other proteins (Hoffmann et al., 2006). Furthermore, the linker sequence between zinc finger 2 and zinc finger 3 is SGDKP, which is similar to the canonical zinc finger linker sequence TGEKP. Whilst this sequence has been found in non-DNA binding proteins (Pavletich and Pabo, 1993), generally this is considered to be a predictor

for DNA binding, as it was shown that these amino acids contribute to the DNA binding of the associated zinc finger (Ryan and Darby, 1998; Laity et al., 2000).

Combining the results from my experiments with the results from published experiments and deductions based on the amino acid sequence of AEBP2, it remains unclear whether the zinc fingers of AEBP2 are involved in DNA binding. It seems likely that not all three are involved, especially given the observation that mutation of the second zinc finger did not disrupt DNA binding (He et al., 1999). If the zinc finger domain does bind DNA, it is unclear with what degree of sequence specificity it does so. Further mutational analysis and *in vitro* and *in vivo* binding and gene expression analysis will help to elucidate to what extent the zinc fingers are required for DNA binding and for interaction with the PRC2 complex.

4. Localisation of AEBP2 in the cell

4.1 Introduction

The localisation of a protein within a cell can give important clues to its function. Therefore, I assessed AEBP2 localisation by immunofluorescence (IF) and ChIP.

4.1.1 Localisation of PRC2 components by immunofluorescence

Early immunofluorescence studies showed that the PRC2 components EED, SUZ12 and EZH2 are localised to the nucleus in ES cells with a diffuse pattern (Plath et al., 2003; Silva et al., 2003; de la Cruz et al., 2005). Several studies employed *in vitro* differentiation methods to investigate the localisation of PRC2 components during differentiation from ES cells into different tissue types (Plath et al., 2003; Silva et al., 2003; de la Cruz et al., 2005). One method to differentiate ES cells is by LIF withdrawal (Smith et al., 1988; Williams et al., 1988). Cells, cultured in suspension in the absence of LIF, form embryoid bodies. Early (day 1-3) embryoid bodies are the equivalent of E5.5-E6 mouse embryos (Leahy et al., 1999). Between days 3-6, an outer endoderm layer encloses cells in the inside of the embryoid body, which are differentiating into the three germ layers, and contain cell types also found in the gastrulating embryo. Further culturing leads to the generation of more differentiated cells, such as cardiomyocytes. Upon differentiating male ES cells, PRC2 proteins remain nuclear, whilst the intensity of immunofluorescence signal may decrease somewhat as the abundance of some PRC2 proteins is reduced (Silva et al., 2003). However, PRC2 accumulates on the inactive X in *in vitro* differentiating female ES cells, resulting in a clear focus resembling the *Xist*

cloud described in the Introduction. During *in vitro* differentiation, one of the X chromosomes is randomly chosen to be inactivated, in keeping with what is happening in differentiating cells in embryos during development (Rastan and Robertson, 1985; Penny et al., 1996). *Xist* RNA accumulates on the inactivated chromosome, which is visible at day 2 of differentiation. It was found that EED, SUZ12 and EZH2 all transiently associated with the X chromosome, with the highest amount of accumulation around day 4 of differentiation. This accumulation was also detected in developing embryos at E6.5, with a gradual decrease over development to E11.5. H3K27me3 also becomes enriched on the inactive X with similar dynamics (Plath et al, 2003). However, in fully differentiated cells H3K27me3 often still marks the inactive X chromosome, whereas enrichment of PRC2 components is no longer visible (Gilbert et al., 2003; Plath et al., 2003; de la Cruz et al., 2005). Accumulation of PRC2 components and H3K27me3 also occurs during imprinted XCI (Mak et al., 2002; Okamoto et al., 2004).

Xist RNA accumulation does not appear to depend on PRC2 components or H3K27me3, as *Xist* still accumulates in *Eed*-/- cells, and X chromosomes are successfully transcriptionally silenced in the absence of a functional PRC2 complex. It has been suggested that EED is required for the maintenance of XCI in extraembryonic tissue, but this remains a controversial issue (Silva et al., 2003; Kalantry and Magnuson, 2006; Kalantry et al., 2006). Conversely, using a system with an inducible *Xist* cDNA transgene, several groups have shown that recruitment of PRC2 components depends on *Xist* accumulation (Silva et al., 2003; Schoeftner et al., 2006) and that even a RepA mutant *Xist* can still recruit PcG proteins (Plath et al., 2003). Therefore, PRC2 accumulation lies downstream of *Xist* accumulation in the process of X chromosome inactivation.

More recent studies have confirmed that PRC2-associated factors such as PCL2 also associate with the inactive X during differentiation (Casanova et al., 2011). Furthermore, PRC1 components are also recruited, and the inactive X is marked by H2AK119Ub (de Napoles et al., 2004; Schoeftner et al., 2006; Tavares et al., 2012).

4.1.2 Localisation of PRC2 components by ChIP

Research in *Drosophila melanogaster* established PcG proteins as repressors of the *Hox* genes. However, research from several groups indicated that in mouse and human embryonic stem cells, PcG proteins not only repress the *Hox* clusters but also many other genes important for mammalian development (Bernstein et al., 2006; Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006b; Ku et al., 2008). Both ChIP-chip, in which chromatin immunoprecipitation is followed by hybridisation to tiling arrays, such as arrays covering known promoters, and ChIP-seq approaches were employed and it was found that PRC2 binding sites (as defined by enrichment for one or more of the PRC2 components EZH2, EED and SUZ12) closely overlapped sites of H3K27me3 enrichment. These sites were typically found at promoters and genes associated with these promoters were found to be highly significantly enriched for genes grouped in Gene Ontology (GO) categories such as “Development”, “Regulation of transcription”, “Cell differentiation” and many other categories typically associated with embryonic development. Together, these studies in combination indicated that the promoters of about 2,000 genes are occupied by PRC2 in embryonic stem cells. These promoters contain H3K27me3 as well and most are bivalent (as discussed in the Introduction). It was also noticed that nearly all PRC2 targets were CpG islands and regions devoid of CpG methylation as detected by biotinylated CXXC Affinity Purification (Bio-CAP) (Ku

et al., 2008; Blackledge et al., 2012). Furthermore, many PRC2 enriched promoters are also bound by PRC1. A typical view of PRC2 and PRC1 binding to CpG islands of the *Lhx4* gene is presented in figure 1, taken from Klose et al. (2013).

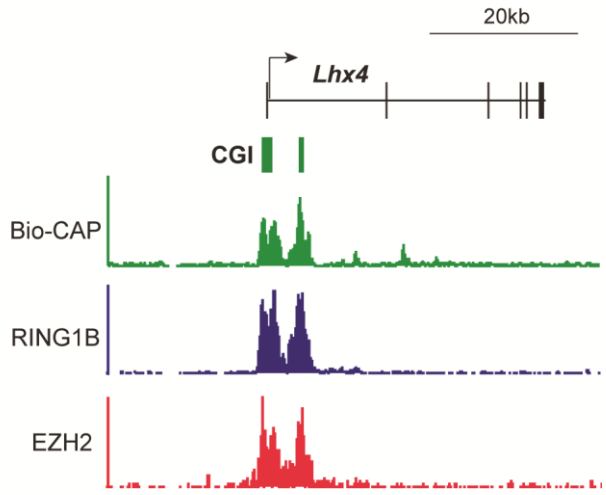


Figure 4.1: ChIP-seq traces for RING-1B and EZH2 and Bio-CAP trace show that PRC1 and PRC2 localise to an unmethylated CpG island at the *Lhx4* gene. Reproduced from Klose et al. (2013) under Creative Commons license CC-BY.

More recently, several proteins were found to associate with PRC2 in a substoichiometric manner and that they localise to PRC2 targets (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010; Casanova et al., 2011; Li et al., 2011b; Hunkapiller et al., 2012). Peng and co-workers found that JARID2 localises to virtually all PRC2 targets using ChIP-seq, whereas other groups found a more moderate but still high overlap of JARID2 targets and PRC2 targets using ChIP-chip approaches. Groups led by Danny Reinberg and by William Stanford both indicated that about 48-63% of PRC2 targets were also enriched for PCL2. Hunkapiller and colleagues found that a TAP-tagged version of PCL3 localised to 44% of SUZ12 targets, whilst Brien and co-workers reported a 16% overlap of endogenous PCL3 with PRC2 targets. PCL1 has

been shown to localise to some PRC2 targets but has thus far not been tested genomewide.

Although one might conclude from these observations that JARID2 appears to be at virtually all PRC2 targets whereas proteins from the polycomblike family are present at more limited locations, this conclusion may prove false. This is because overlap of peaks depends on peak calling, and the accuracy of peak calling is influenced by the signal-to-noise ratio, which is in turn influenced by technical aspects of the ChIP experiment, such as the conditions used and the affinity of the antibody for its epitope. Therefore, peak calling on ChIP material obtained from immunoprecipitation with an antibody with a poorer affinity will lead to fewer peaks being called due to a lower signal-to-noise ratio. This situation could give the impression that the protein with the poorer antibody binds fewer places when this is in fact not what is happening on a biological level. Indeed, when Brien and colleagues looked for PCL3 enrichment by ChIP followed by quantitative real-time PCR (qPCR) at places that were deemed PCL3 negative by ChIP-seq, they found that these areas were significantly enriched for PCL3 compared to background, but less so than the regions deemed PCL3 positive by ChIP-seq, thereby just falling below the detection limits of the peak calling algorithm. Therefore I posit that it has been shown that PRC2 associating factors associate to some and possibly all PRC2 targets examined so far.

Thus far, the localisation of AEBP2 has been investigated only by the group of Joomyeong Kim at the University of Louisiana. In their first study, Kim and colleagues performed a modified ChIP protocol on brain tissue from a 1-month old mouse (Kim et

al., 2009a). In this protocol, immunoprecipitated DNA was subjected to restriction enzyme digestion and subsequently cloned to investigate targets. Using this method, 126 unique sequences were identified, 53 of which (42%) corresponded to sites near genes known to be PcG targets. This approach gives an indication of association but does not supply information about enrichment due to the cloning step. In a subsequent study, Kim and colleagues used E14.5 embryos to test the association of AEBP2 with a number of loci indicated to be susceptibility genes associated with Waardenburg Syndrome, because they predicted that AEBP2 plays a role in this disease based on the phenotype of a heterozygote knockout (as discussed in Chapter 7) (Kim et al., 2011). Generally, there was poor enrichment of signal to background at proposed target loci, although at one locus, the *Sox10* locus, above-background signal was obtained, for both AEBP2 and EZH2. However, not much enrichment for H3K27me3 was detected at this site.

I wanted to obtain a better and more encompassing view of where AEBP2 was localised in the genome. Therefore, in addition to analysing AEBP2 recruitment to the inactive X during X chromosome inactivation, I set out to analyse AEBP2 recruitment genomewide.

4.2 Raising a rabbit polyclonal antibody against the mouse AEBP2 protein

In order to assess the localisation of endogenous AEBP2, I raised an antibody, using the 496 amino acid length protein as antigen. This recombinant protein covers all of the exons that are translated, except for exons 9a and 9b, which are 14 and 8 amino acids long respectively (see Figure 1.8). The antigen was purified as described in Chapter 2.

Sera from the final bleed from two immunised rabbits were obtained from Eurogentec. The final sera of both rabbits showed increased reactivity towards the antigen, as compared to the pre-immunised sera from the same rabbits. The serum of one of the rabbits also reacted with a species of similar molecular weight in both whole cell extracts (WCE) and nuclear cell (NCE) extracts from several ES cell lines commonly used in our lab (Figure 4.2).

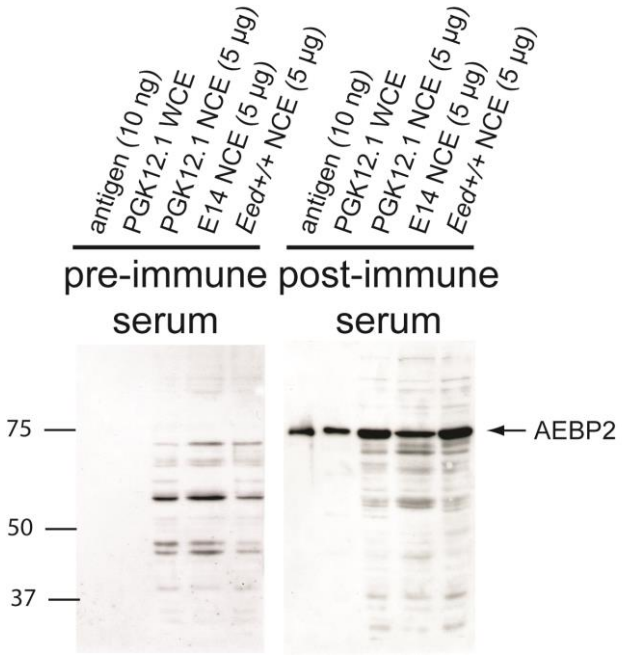


Figure 4.2: Post-immune, but not pre-immune, serum reacts with the antigen and a band of similar size in whole cell and nuclear cell extracts. WCE: Whole Cell Extract, NCE: Nuclear Cell Extract. PGK12.1 is a female ES cell line. E14 is a male ES cell line. *Eed*$^{+/+}$ is an ES cell line used as control for *Eed*$^{-/-}$ cell line used later in this study.

To further validate the reactivity of the antibody, I produced 3 different constructs which caused overexpression of N-terminally or C-terminally FLAG-tagged AEBP2 in mammalian cells after transfecting 4 or 6 µg of DNA. Human Embryonic Kidney 293T (HEK 293T) cells were used for ease of transfection and whole cell extracts were

analysed for overexpression by immunoblotting (Figure 4.3A). The different sizes of the overexpressed proteins correlate with different sizes of the protein construct due to different linker length. Surprisingly, the protein appears as a doublet when overexpressed; the reason for this is currently unknown, but has also been observed by others (Cao et al., 2008). In addition, I analysed the ability of the AEBP2 antibody to recognize overexpressed AEBP2 by immunofluorescence. Overexpressed AEBP2 was nuclear, as was expected based on its proposed role in the PRC2 complex and the presence of putative nuclear localisation signals in the C-terminal basic domain (LKNKRRRSLP starting at amino acid 376). The signal was reduced by increasing competition of the antigen used to raise the antibody against (Figure 4.3B). In non-transfected cells, the AEBP2 signal was predominantly from the cytoplasm, suggesting that endogenous AEBP2 could be mostly in the cytoplasm or that the serum from the final bleed is also cross-reactive with other proteins.

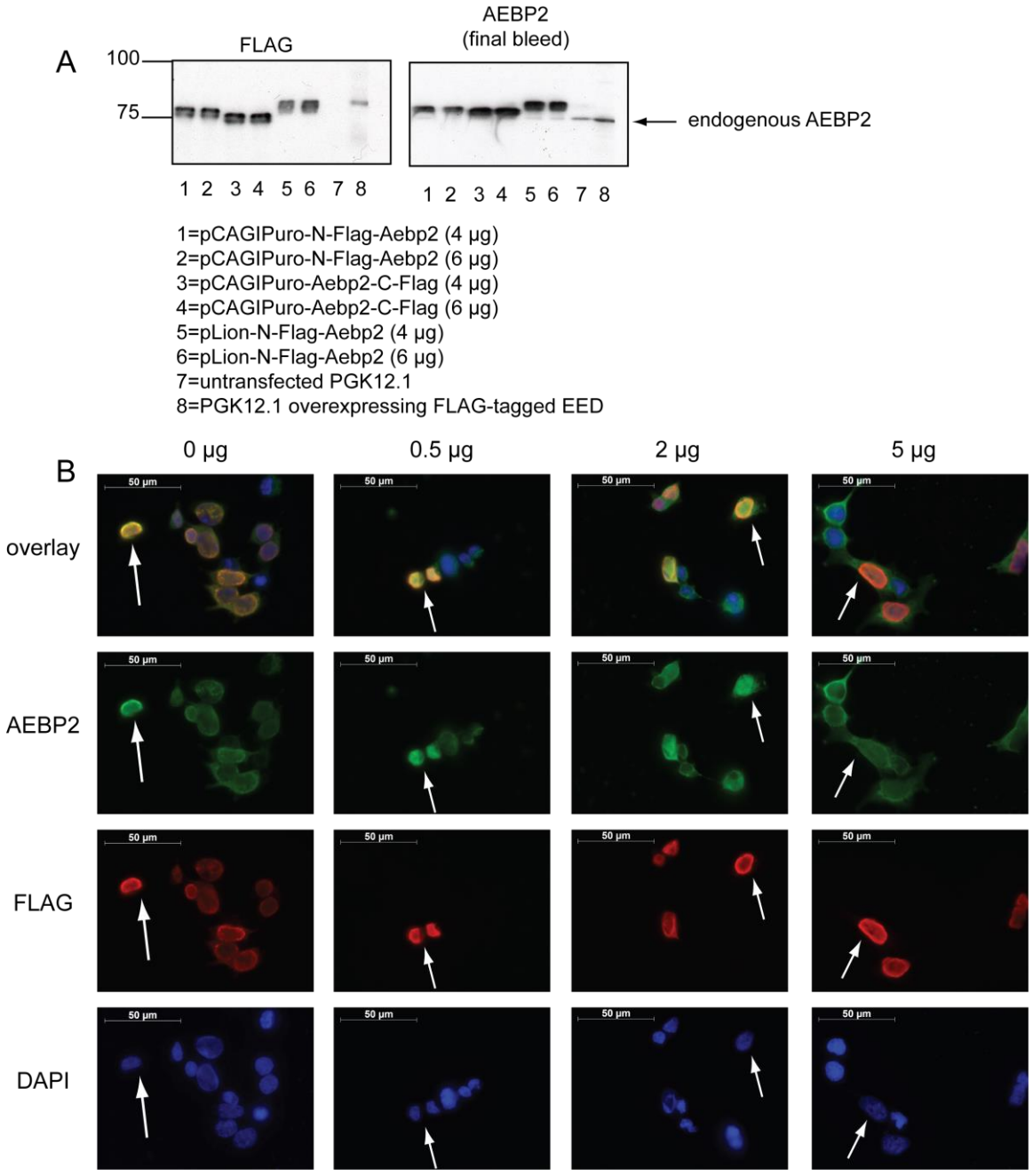


Figure 4.3: The AEBP2-antibody reacts with AEBP2. (A) AEBP2 antibody reacts with overexpressed FLAG-tagged AEBP2 by immunoblot. As a negative control untransfected PGK12.1 sample is loaded in lane 7. As a positive control FLAG-EED sample is loaded in lane 8. In the AEBP2 blot, control samples show endogenous AEBP2, but not overexpressed. In brackets are the amounts of vector that was used to transiently transfect HEK 293T cells. (B) AEBP2 antibody reacts with overexpressed FLAG-tagged AEBP2 by immunofluorescence, and signal is reduced by antigen competition. Overlap between FLAG and AEBP2 is indicated with arrows. When the unpurified AEBP2 antibody was blocked with 5 µg antigen, AEBP2 and FLAG no longer overlap in the nucleus (red instead of yellow cells), indicating that the AEBP2 antibody is blocked by the antigen. In untransfected cells, unpurified AEBP2 antibody primarily cross-reacts with a protein present in the cytoplasm.

In order to reduce any cross-reactivity I purified AEBP2 reactive antibody from serum. An antigen-based column purified only a small fraction from the total amount of antibody present (very little antibody purified relative to rabbit serum albumin) and did not increase reactivity towards the band proposed to be AEBP2 by immunoblotting (Figure 4.4) or reduce signal from the flow through. However, it did improve signal from immunofluorescence (which showed a nuclear signal with purified antibody) and therefore the purified antibody was used for this application.

Because of the low yield from the antigen column purification, I purified bulk antibody from the serum by using a protein-A based column. After quantification, purified antibody retained the ability to recognize AEBP2 whilst reducing background (as compared to post-immune serum in Figure 4.2) (Figure 4.5). The protein-A purified antibody was used for immunoblotting.

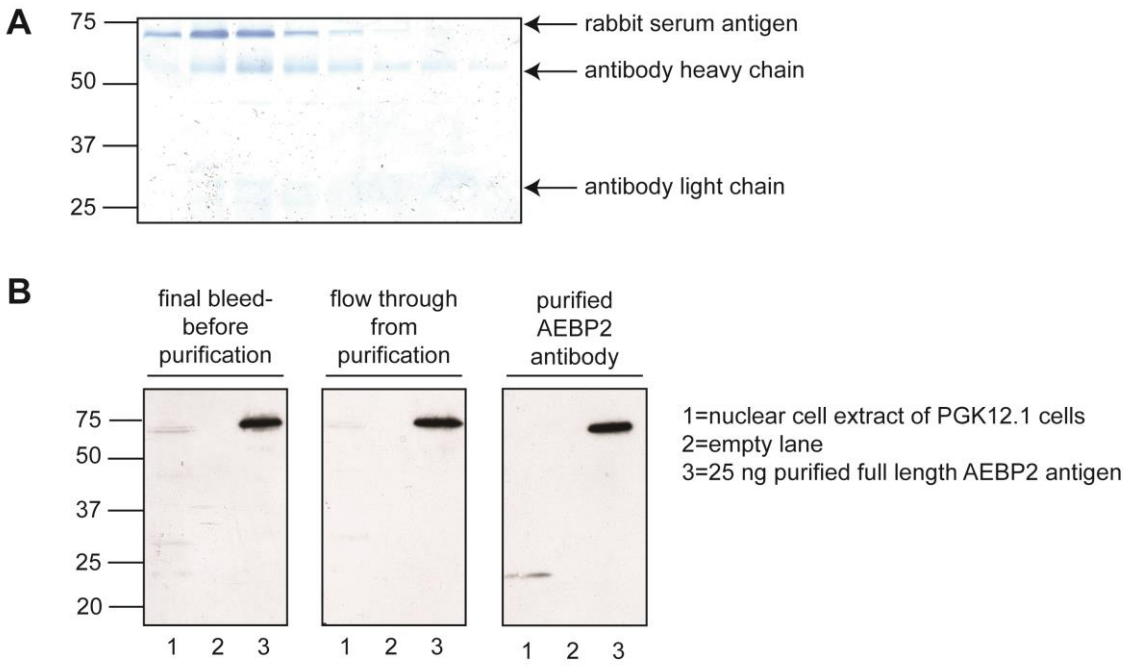


Figure 4.4: AEBP2 antibody purification from an antigen-based column has a poor yield. (A) A low amount of antibody is eluted from the antigen-based column. (B) Less signal at ~73 kDa using the purified AEBP2 antibody to immunoblot nuclear cell extract. This band is proposed to be endogenous full length AEBP2. Therefore, purified AEBP2 antibody is less able to react with AEBP2 than the serum from the final bleed.

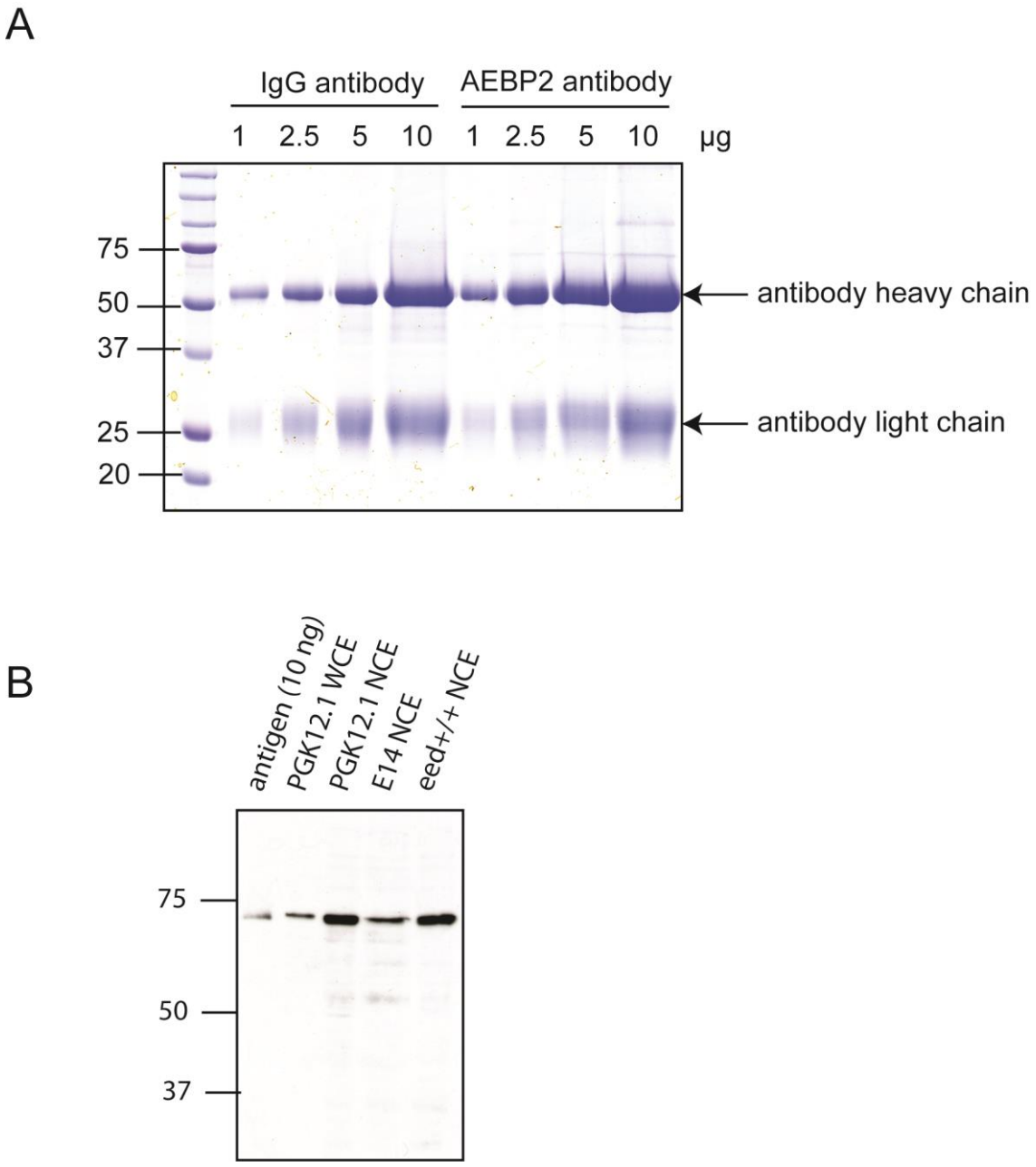
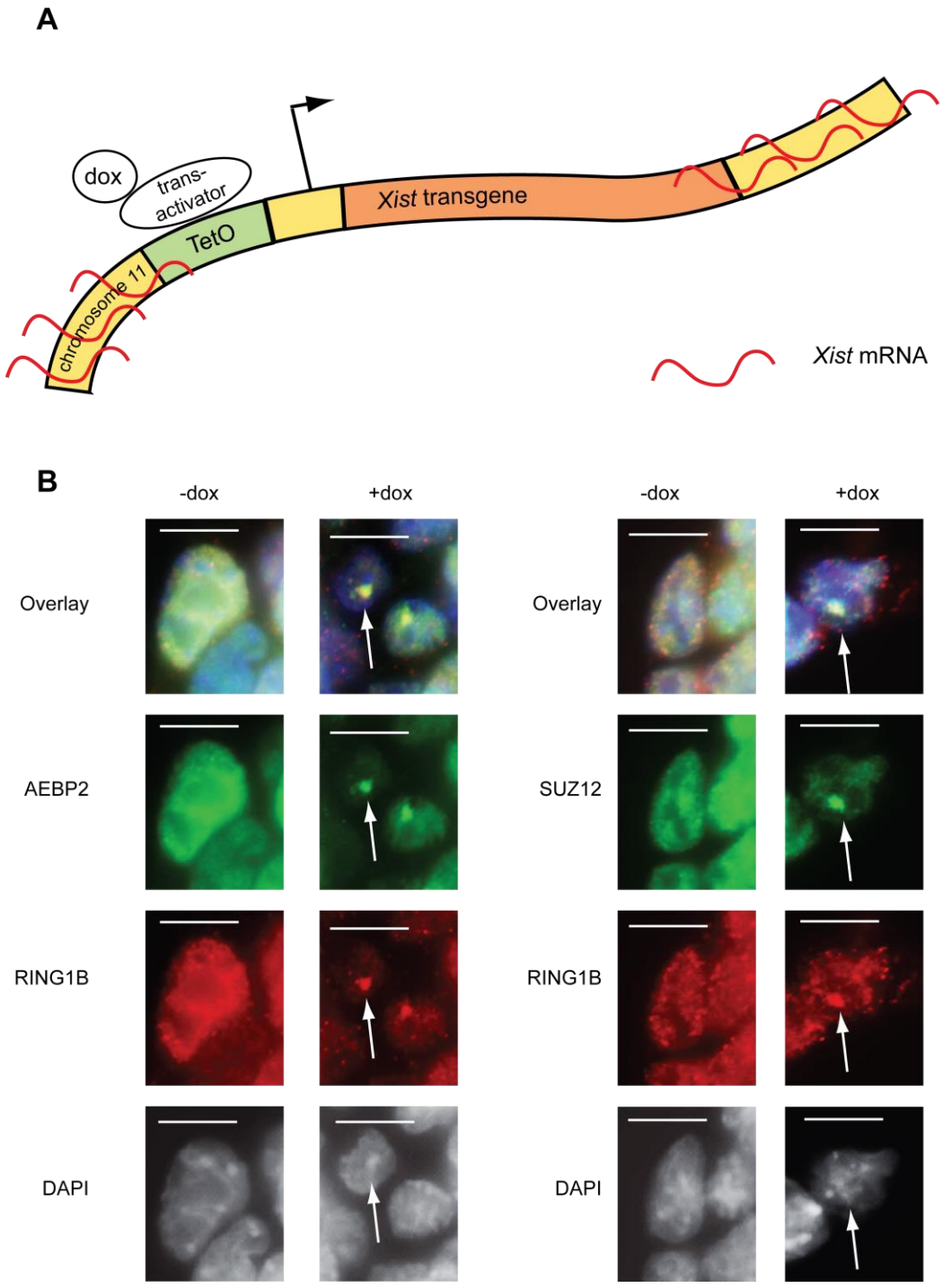


Figure 4.5: Purification of antibody from serum using a protein A-based column results in a high yield. (A) Quantification of eluted antibody. (B) Eluted antibody works well in immunoblotting.

4.3 AEBP2 localises to the inactive X

In order to assess whether AEBP2 localises to the inactive X in a fashion similar to other PRC2 components, I employed a XCI model system set up by Anton Wutz and Rudolph Jaenisch (Wutz and Jaenisch, 2000). In this system, a 15 kb *Xist* cDNA, containing the full length somatic transcript, was randomly inserted in chromosome 11 in J1 ES cells, a male 129 mouse strain. The *Xist* transgene is under control of a promoter that requires the binding of nls-rtTA to the Tet operator for activation. The cells also have a cDNA in the *ROSA26* locus encoding the nls-rtTA doxycycline-inducible activator. Therefore addition of doxycycline leads to the association of the activator with the Tet operator, leading to expression of *Xist* and coating of chromosome 11 by *Xist* RNA (Figure 4.6A). Subsequently PRC2 is recruited to the inactivated chromosome and the chromosome becomes marked with H3K27me3 and H2AK119ub as visualised by immunofluorescence (Schoeftner et al., 2006).

48 hours of doxycycline treatment resulted in RING1B foci that co-localised with SUZ12 and EZH2, and H2A119ub foci that co-localised with EED (Figure 4.6B, C). Purified AEBP2 antibody was used to visualise AEBP2 foci that also co-localised with RING1B, and 56.4% of cells were found to have an inactive chromosome like focus (Figure 4.6D).


A
dox
trans-
activator
TetO
Xist transgene
chromosome 11
Xist mRNA
B
-dox
+dox
-dox
+dox
Overlay
AEBP2
RING1B
DAPI
Overlay
SUZ12
RING1B
DAPI

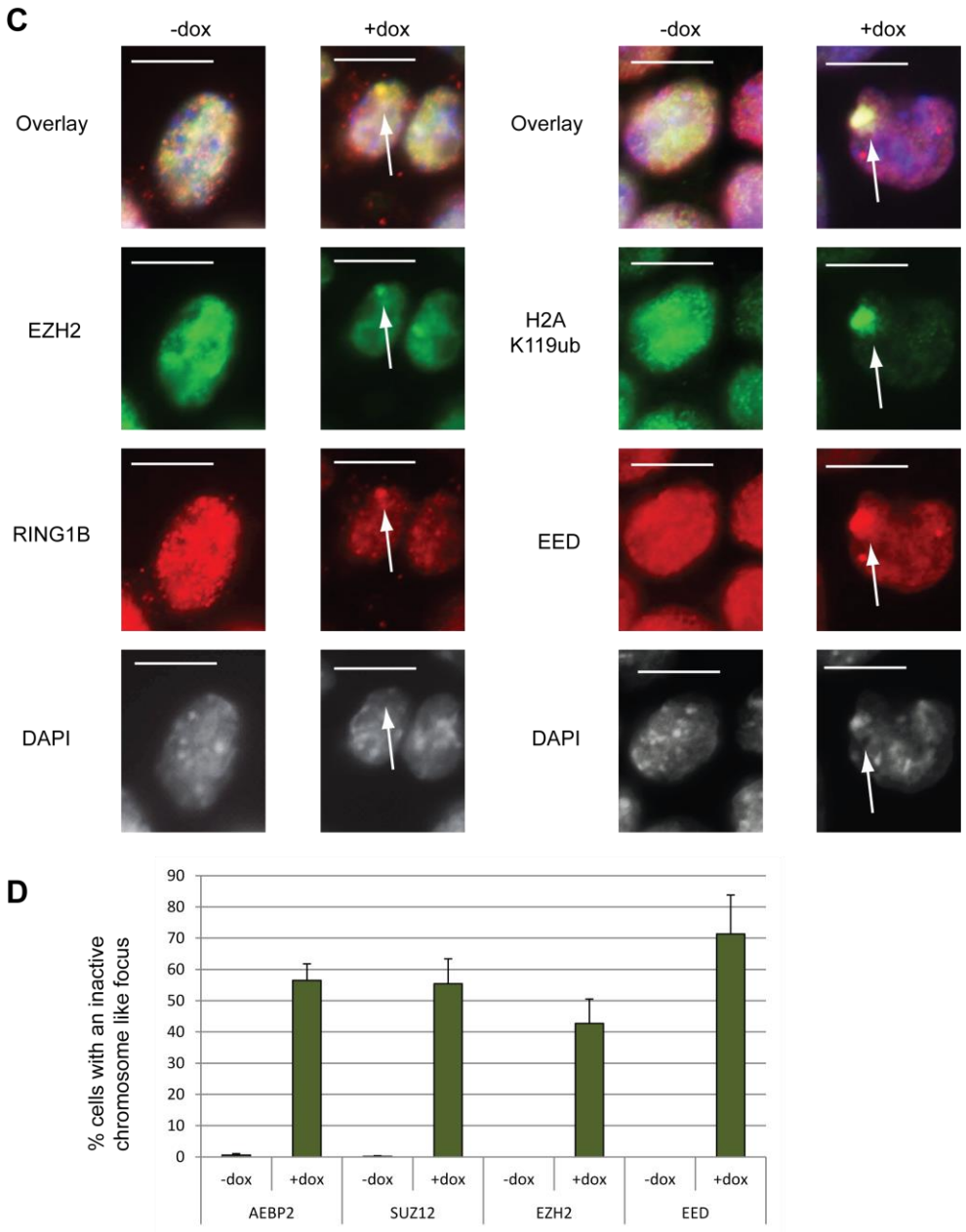


Figure 4.6: AEBP2 is recruited to an inactivated chromosome. (A) Doxycycline induces transcription of an *Xist* transgene from chromosome 11 in the male J1 ES cell line. (B) AEBP2, SUZ12 and Ring1B accumulate on the inactivated chromosome after doxycycline treatment as observed by immunofluorescence. (C) EZH2, EED and H2AK119Ub accumulate on the inactivated chromosome after doxycycline treatment as observed by immunofluorescence. Scale bar is 10 µm. Arrows indicate the location of an "inactivated chromosome like" focus. Isolated DAPI staining is shown in greyscale for enhanced contrast. (D) Quantification of percentage of cells with an "inactivated chromosome like" focus. Error bars represent standard deviation of 3 biological repeats. A minimum of 117 cells was counted for each condition in each experiment.

It was confirmed that the "inactivated chromosome like" signal from the SUZ12 and AEBP2 antibodies was not due to cross-reactivity of the anti-rabbit secondary antibody with the mouse RING1B primary antibody, because no foci were observed in the absence of a rabbit antibody (Figure 4.7A). Additionally, in the absence of RING1B antibody, foci from AEBP2 and SUZ12 were still observed (Figure 4.7A, B, C), confirming that there is no cross-reactivity between the RING1B primary antibody and the AEBP2 or SUZ12 antibodies.

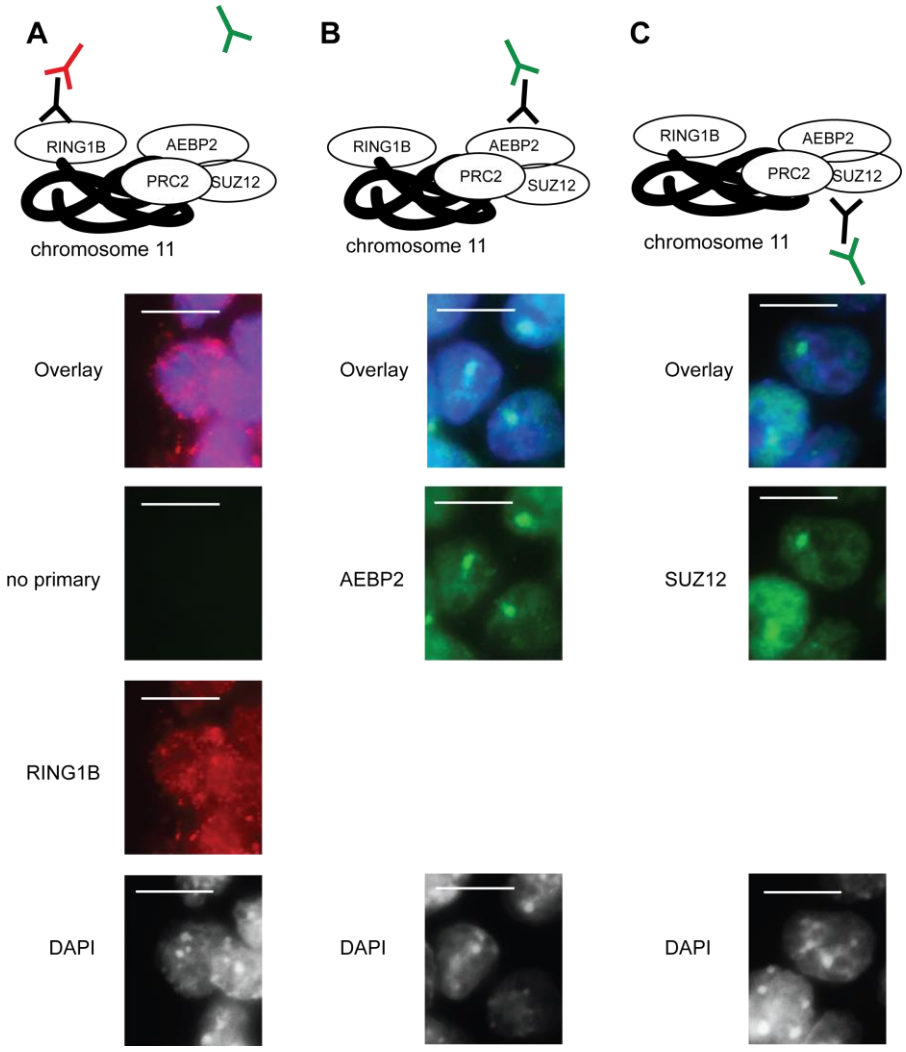


Figure 4.7: Signal from foci in immunofluorescence is not due to cross-reactivity. (A) The GFP-conjugated secondary antibody does not recognize the RING1B primary or mCherry-conjugated secondary antibody. (B) There are still AEBP2 foci in the absence of RING1B primary or mCherry conjugated secondary antibody. (C) There are still SUZ12 foci in the absence of RING1B primary or mCherry conjugated secondary antibody. Arrows indicate "inactivated chromosome like" foci in the male J1 ES cell line, carrying an inducible *Xist* transgene on chromosome 11. Scale bar is 10 µm. Isolated DAPI images are shown in greyscale for enhanced contrast.

Subsequently, female PGK12.1 ES cells (Norris et al., 1994; Penny et al., 1996) were differentiated by LIF withdrawal and on day 3 the localisation of AEBP2 was analysed by immunofluorescence. Co-localisation between AEBP2 and H3K27me3, as well as AEBP2 and EED, was observed in three biological repeats (Figure 4.8). However, not many cells could be counted due to a low frequency of successful staining. Nonetheless, the findings from the XCI model system and endogenous staining suggest that AEBP2 localises with the PRC2 complex to the inactive X.

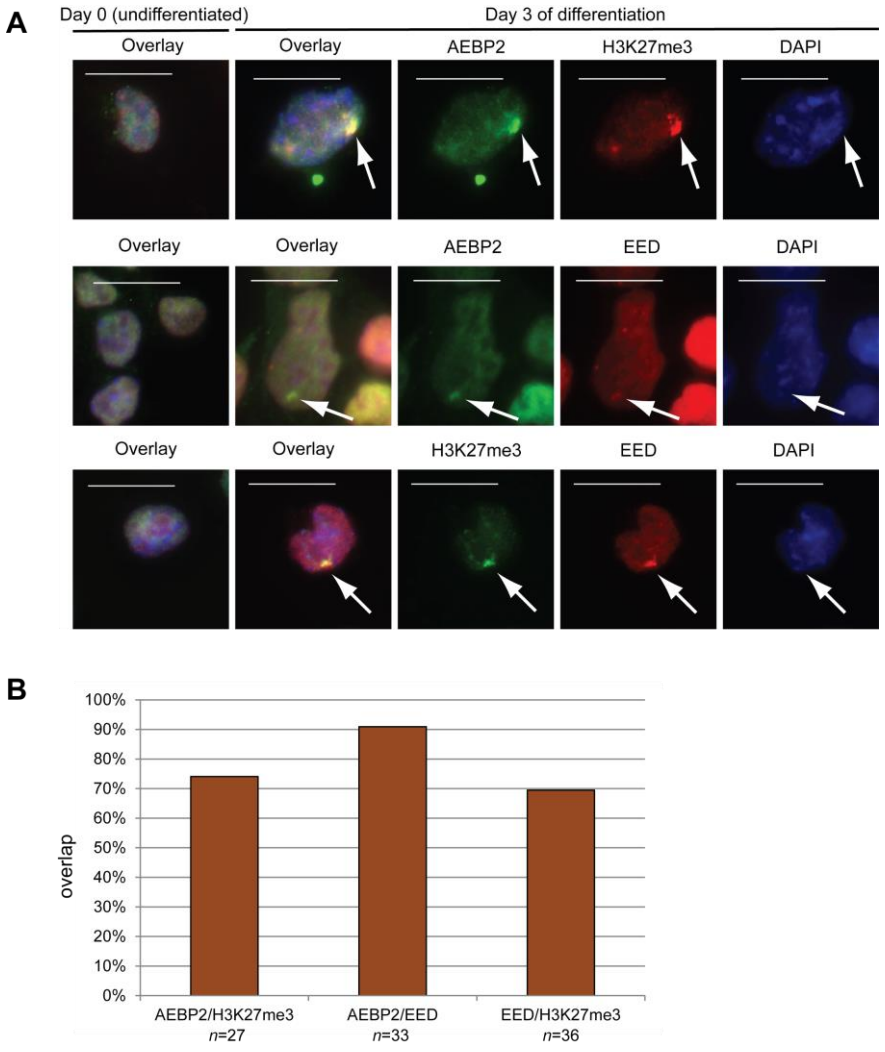


Figure 4.8: AEBP2 overlaps with EED and H3K27me3 in differentiating female ES cells. (A) Immunofluorescence shows that AEBP2, EED and H3K27me3 foci overlap. (B) Quantification of overlap. Total number of cells that had H3K27me3 or EED foci is indicated, and overlap is quantified as number of cells that had e.g. an AEBP2 focus overlapping the H3K27me3 focus. Scale bar is 20 µm.

4.4 AEBP2 localises to PRC2 targets genomewide

I also assessed whether AEBP2 localises to PRC2 targets by ChIP-qPCR. Previous studies had indicated that AEBP2 localised to at least some PRC2 targets, but this had not been assessed in ES cells, or in a genomewide fashion. Several primer sets against the *Nkx2.2* gene were used, which give a characteristic enrichment pattern, with increased enrichment at the 5' end compared to the 3' end of the gene (Figure 4.9A). I used my post-immune serum to test enrichment of endogenous of AEBP2 at genes known to be targets of PRC2, as evidenced by their enrichment for H3K27me3 (Figure 4.9B, C). It was demonstrated that AEBP2 localises to PRC2 targets.

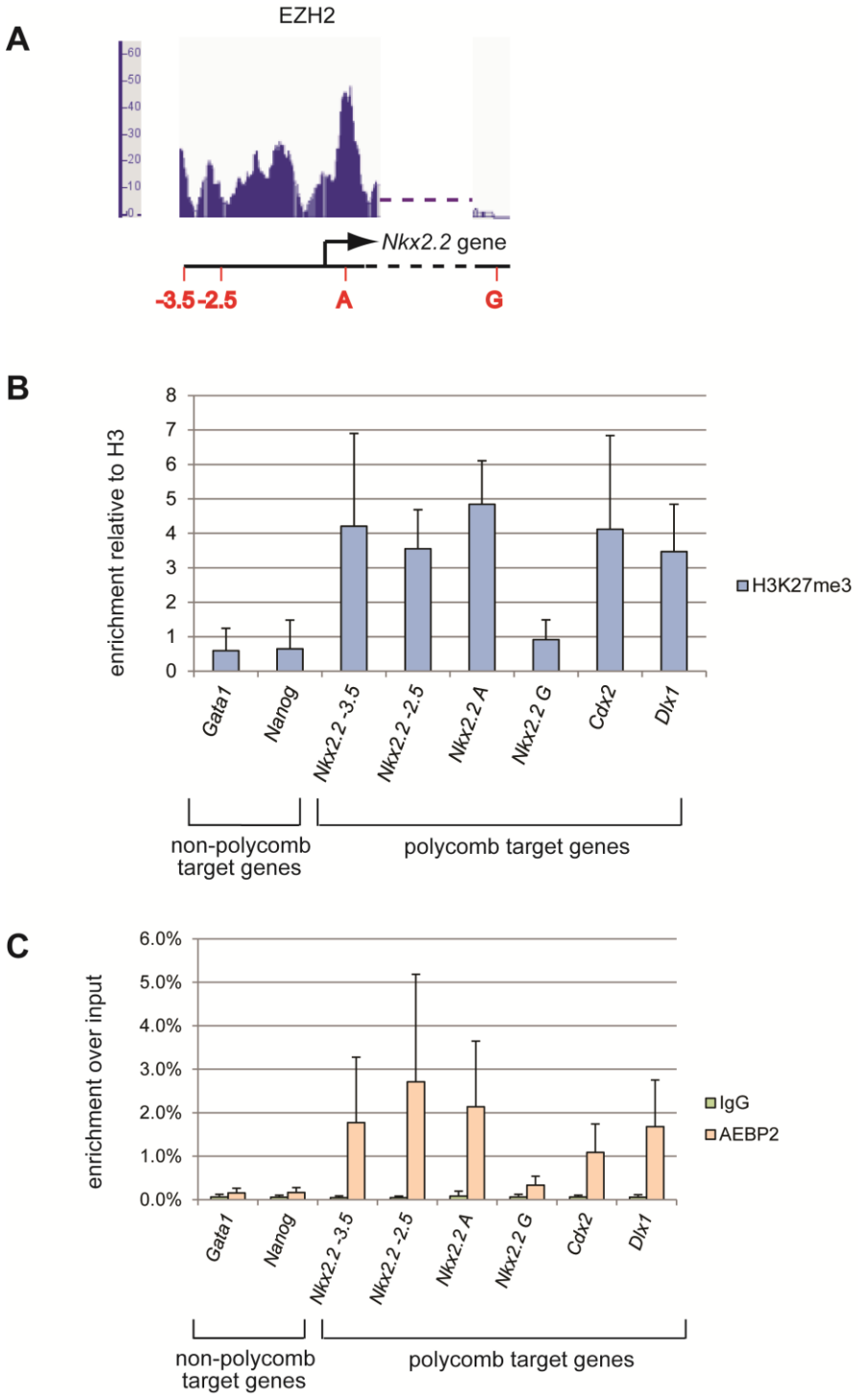


Figure 4.9: AEBP2 localises to genomic locations also enriched for H3K27me3 in E14 ES cells. (A) Primers across the *Nkx2.2* gene. Figure adapted from Stock et al. (2007) with permission. EZH2 enrichment across the *Nkx2.2* gene shows a characteristic pattern of enrichment. EZH2 profile from EZH2 ChIP-seq from Peng et al. (2009). (B) ChIP-qPCR for H3K27me3. (C) ChIP-qPCR for AEBP2 and IgG control. Error bars represent standard deviation of 3 biological repeats.

In a single blocking experiment, I showed that the amount of DNA immunoprecipitated by the AEBP2 antibody was strongly reduced when the antibody was blocked with full length recombinant AEBP2 (rAEBP2) but not when blocked with an equivalent amount of Bovine Serum Albumin (BSA).

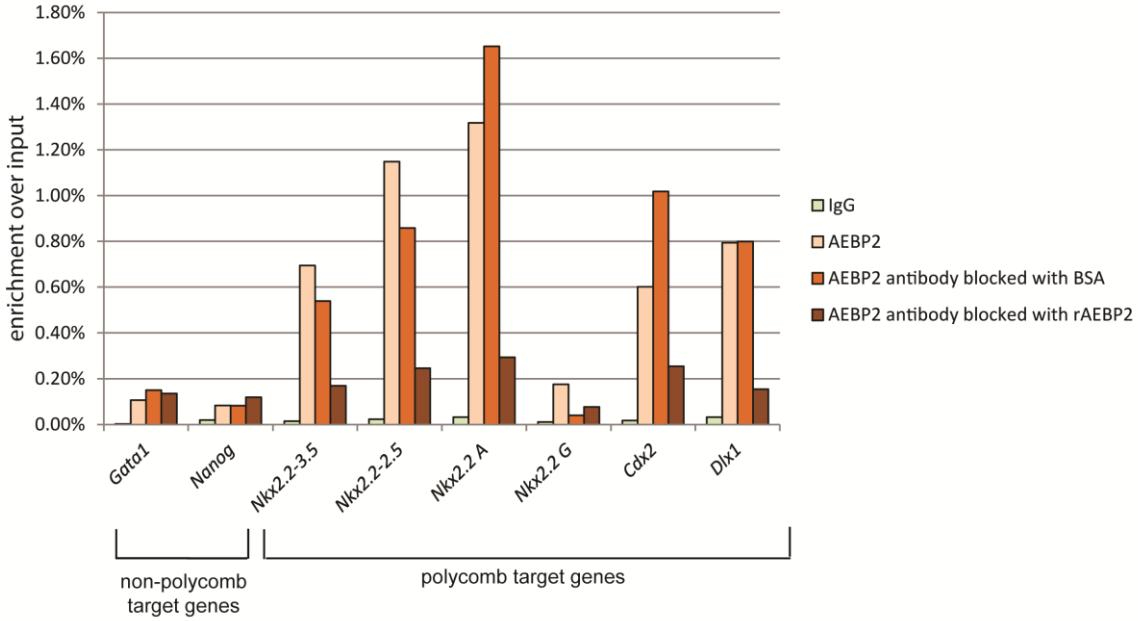


Figure 4.10 Full length recombinant AEBP2 blocked antibody, but not BSA-blocked AEBP2 antibody immunoprecipitates less DNA from polycomb target genes. This experiment was done once.

I then assessed the localisation of AEBP2 genomewide by performing ChIP-seq. Unfortunately the use of the endogenous AEBP2 antibody has proven unsuccessful so far. Therefore I decided to use a tagged version of AEBP2. I used cDNA of full length AEBP2 as I had detected this isoform in whole cell extracts and nuclear cell extracts from several mouse ES cell lines (Figure 4.2). Although it has been reported that the main form of AEBP2 is ES cells is the shorter isoform, lacking the acidic and neutral domains (Kim et al., 2009a), studies that identified AEBP2 as a PRC2-associating factor by mass spectrometry detected peptides corresponding to the full length isoform, indicating that it must be present in ES cells (see Table 14 in Chapter 5). Therefore, full

length AEBP2 was N-terminally tagged with a Flag+2xStrep2 (FS2) tag and expressed in E14 cells at levels similar to the endogenous protein (Figure 4.11A, however see also Discussion of Chapter 6). ChIP-qPCR confirmed that tagged AEBP2 localised to PRC2 targets (Figure 4.11B, C).

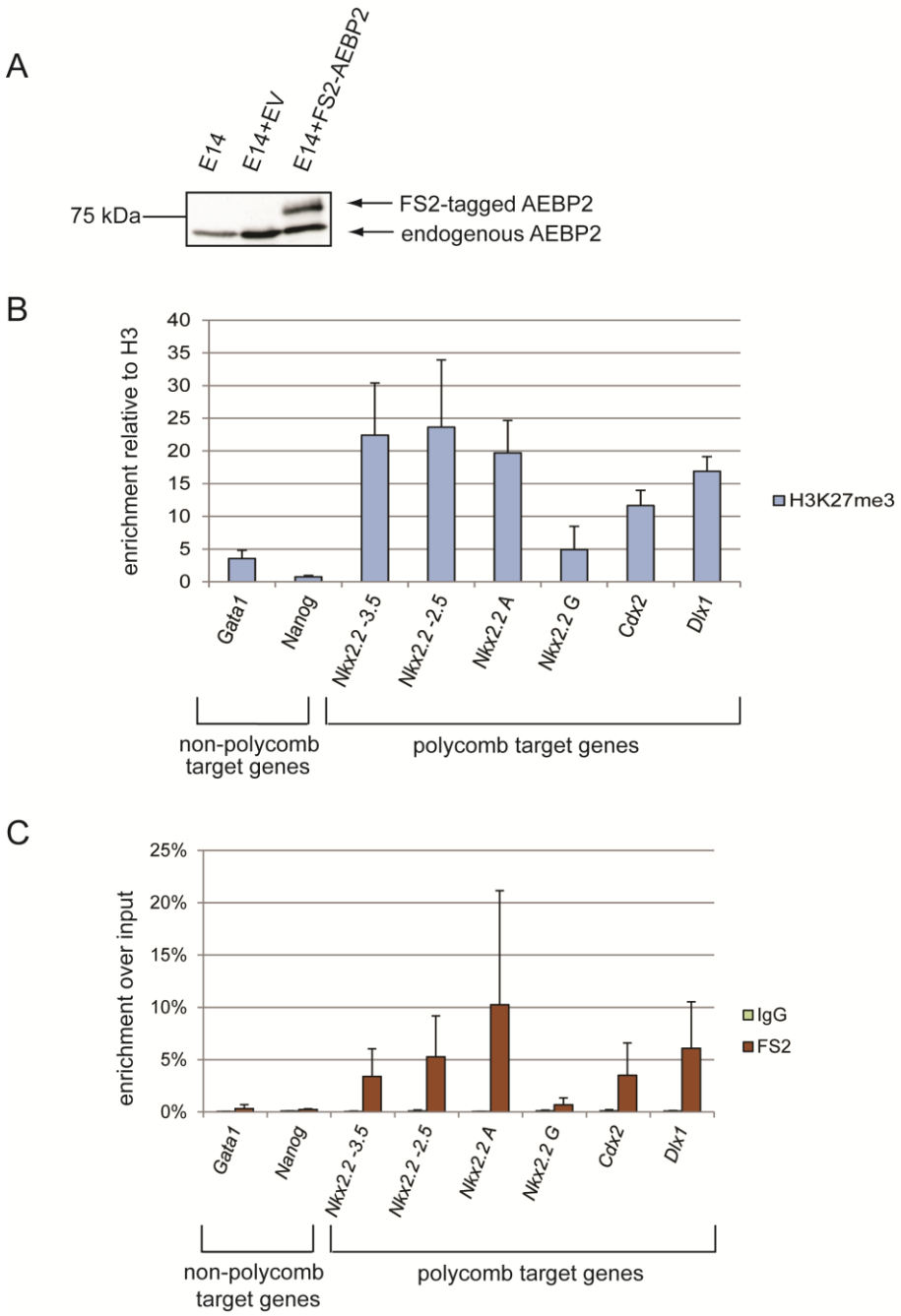


Figure 4.11: FS2-AEBP2 localises to PRC2 targets. (A) Overexpression of FS2-AEBP2 in E14 mouse ES cells as detected by immunoblotting against AEBP2. EV=empty vector. (B) ChIP-qPCR for H3K27me3. (C) ChIP-qPCR for FS2 and IgG control. Error bars represent standard deviation of 3 biological repeats.

Subsequently, ChIP-seq analysis was carried out to determine the binding profile of FS2-AEBP2 genomewide. I performed an analysis of the data using the Galaxy platform.

After aligning the reads, it was established that AEBP2 binding was greatly enriched near transcription start sites (TSS) when an aggregate plot was assembled from all the reads combined (Figure 4.12A). This is what you would expect from a PRC2 associating protein. The input also slightly enriched for TSSs. However, during the library preparation step and amplification step in the ChIP-seq protocol there is a bias towards GC rich regions (Park, 2009), and given that many TSSs contain a GC rich promoter (reviewed in Deaton & Bird, 2011), the apparent enrichment may be explained thus. In peak based analysis, signal from input was subtracted from the immunoprecipitation (IP) signal.

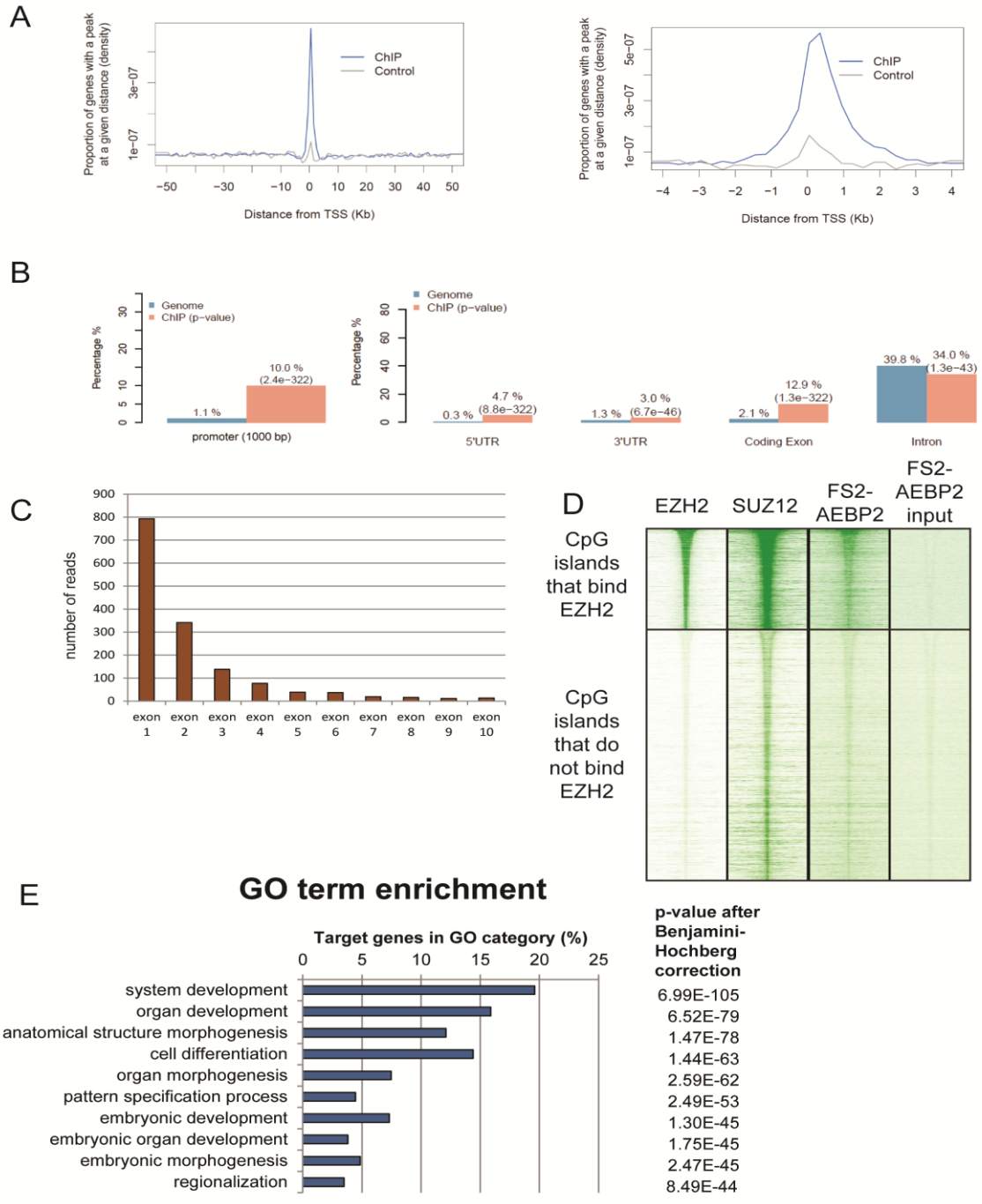


Figure 4.12: Preliminary analysis of FS2-AEBP2 ChIP-seq. (A) FS2-AEBP2 is enriched at promoters. (B) FS2-AEBP2 is enriched at 5' and 3' UTRs and exons, whilst being significantly depleted from introns. (C) Most reads marked as exonic align to exons 1 and 2. (D) Heatmap showing enrichment of EZH2, SUZ12, FS2-AEBP2, and input for FS2-AEBP2 ChIP at CpG islands +- 5 kb. All CpG islands were divided by whether their coordinates overlap with EZH2 peaks or not and then arranged by CpG island size. FS2-AEBP2 binds to the same CpG islands as EZH2 and SUZ12. (E) Genes near FS2-AEBP2 peaks are associated with GO categories related to development.

I then analysed the enrichment of FS2-AEBP2 binding relative to other genomic elements (Figure 4.12B). As expected, binding to promoters was highly enriched. There was also a high enrichment for regions associated with 5' and 3' untranslated regions (UTRs). Furthermore, whilst binding significantly less to introns than expected from a random distribution, FS2-AEBP2 was found to be enriched at exons. However, this may be due to the fact that PRC2 is enriched at CpG islands (Ku et al., 2008) and that CpG islands at promoters often extend into the first exon (Choi, 2010). This proposed correlation is supported by the fact that most of the peaks marked as exonic are in exons 1 and 2 (Figure 4.12C), and as you would expect from Figure 4.12A, most of those reads align close to the TSS. This correlation between start of the gene and increased CpG content may also explain the enrichment for 5' UTRs.

In order to analyse enrichment at CpG islands, all annotated CpG island elements were analysed for overlap with an EZH2 dataset obtained by the Wysocka lab (Peng et al., 2009), which I aligned using the same parameters as the alignment of FS2-AEBP2 and input (Figure 4.12D). From this, it is clear that FS2-AEBP2, as well as SUZ12 (dataset also obtained from the Wysocka lab), are enriched at the same CpG islands as EZH2.

Subsequently, I analysed what biological processes were associated with the genes that were closest to genomic coordinates of FS2-AEBP2 enrichment. Similar to previous findings for core members of the PRC2 complex and other associating factors, it was found that FS2-AEBP2 enrichment was near genes with Gene Ontology (GO) terms associated with development (Figure 4.12E).

Next, I wanted to know what the overlap was between FS2-AEBP2 enriched regions and PRC2-enriched regions. In order to do this, I compared SUZ12 and EZH2 peaks overlap with FS2-AEBP2 peaks. Just under half of AEBP2-enriched locations were also bound by SUZ12, most of which were also bound by EZH2 (Figure 4.13).

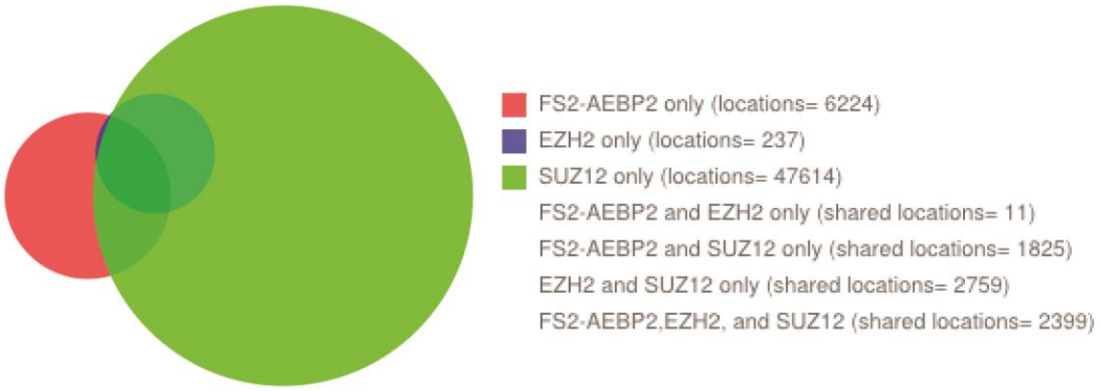


Figure 4.13: FS2-AEBP2 peaks overlap with some, but not all, EZH2 and SUZ12 peaks.

Surprisingly, there were quite a number of peaks that seemed to be bound by FS2-AEBP2 only. However, when visualising aligned reads and called peaks, it became clear that many of these peaks are near EZH2 and SUZ12 peaks, and are likely to arise from different experimental settings (such as antibody, protocol and DNA fragment size) leading to different outcomes from the peak calling algorithms (Figure 4.14A, B). To show this in a different way, I generated a heatmap showing enrichment of SUZ12 and EZH2 at FS2-AEBP2 peaks (Figure 4.14C). This demonstrated that generally, in areas where FS2-AEBP2 is enriched, SUZ12 and EZH2 are also enriched nearby. Furthermore, when I plotted SUZ12 read alignments over regions that were all called as SUZ12 negative, EZH2 negative, FS2-AEBP2 positive (SUZ12-EZH2-FS2-AEBP2+), it seemed that SUZ12 did indeed show some enrichment, but that peaks were not being called

because there was more enrichment nearby. Furthermore, EZH2 enrichment was also somewhat increased in nearby regions. I therefore propose that most "AEBP2-only" loci are in fact also bound by SUZ12 and EZH2. If there are indeed regions that are truly SUZ12 negative, EZH2 negative, FS2-AEBP2 positive, then this could indicate that AEBP2 could localise there independently of the PRC2 complex, or that this could be a non-physiological effect of overexpressing FS2-AEBP2.

Additionally, there were regions that were bound only by SUZ12 and EZH2, but not by FS2-AEBP2. Comparison of SUZ12+EZH2+FS2-AEBP2+ with SUZ12+EZH2+FS2-AEBP2- shows that peaks associated with FS2-AEBP2 are generally broader and also show higher enrichment levels of EZH2 and SUZ12. FS2-AEBP2 signal at regions called SUZ12+EZH2+FS2-AEBP2- does indeed show barely any enrichment. However, when EZH2 peaks are sorted by width and read density of SUZ12 and FS2-AEBP2 is plotted across those peaks, it is clear that in most cases that most, if not all, EZH2 containing regions also are somewhat enriched for FS2-AEBP2. Taken together, the ChIP-seq data shows that FS2-AEBP2 is enriched at PRC2 targets.

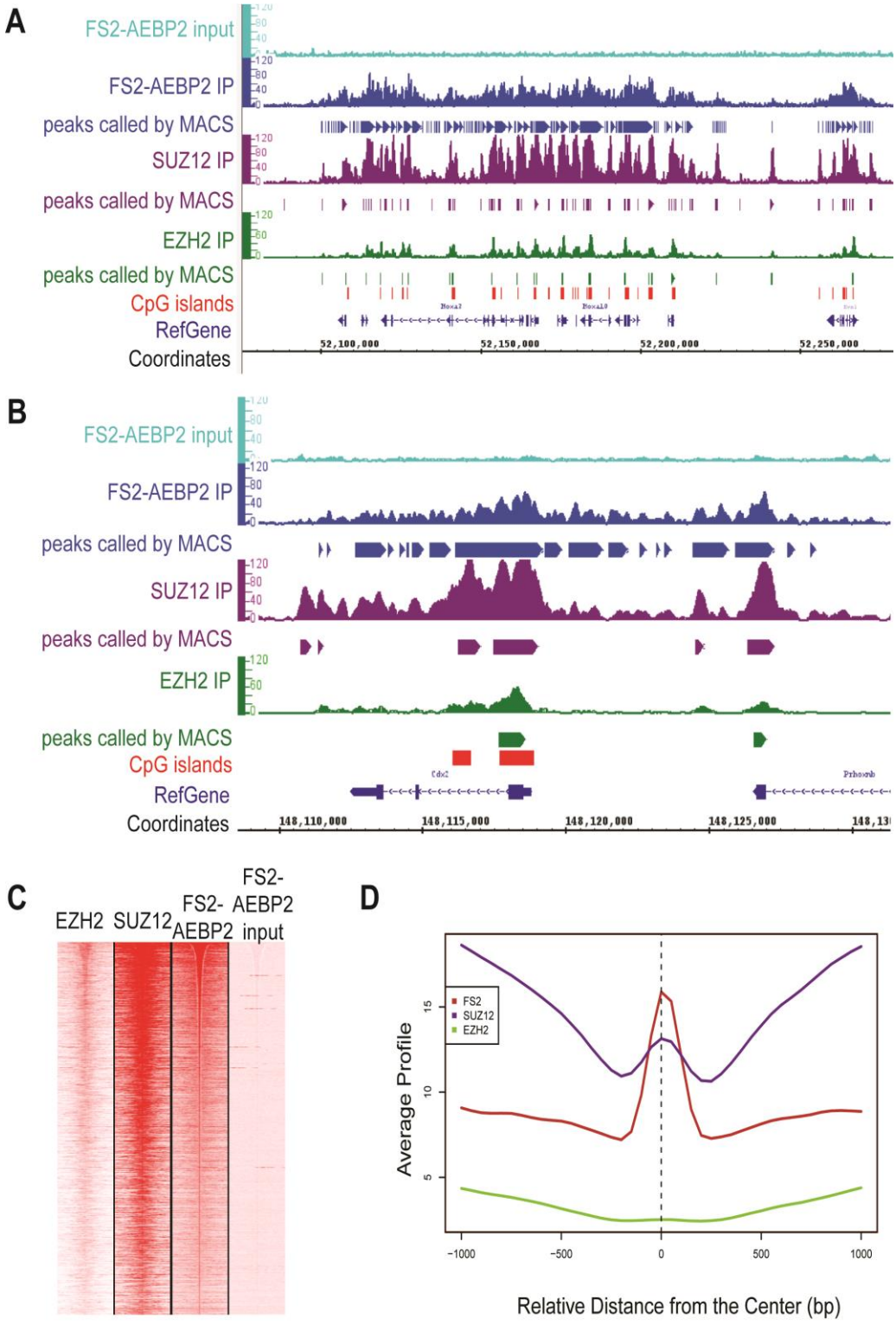


Figure 4.14: SUZ12 and EZH2 are also enriched at "AEBP2-only" peaks. (A) View of the *HoxA* cluster on chromosome 6 with read pile-up and peaks called for FS2-AEBP2 (this study) and EZH2 and SUZ12 (Peng et al., 2009). (B) View of the *Cdx2* gene on chromosome 5. (C) Heatmap showing the binding profile +- 5kb from FS2-AEBP2 peaks for EZH2, SUZ12, FS2-AEBP2 and the input for FS2-AEBP2. (D) Aggregate plot of FS2-AEBP2, SUZ12 and EZH2 at "AEBP2-only" peaks.

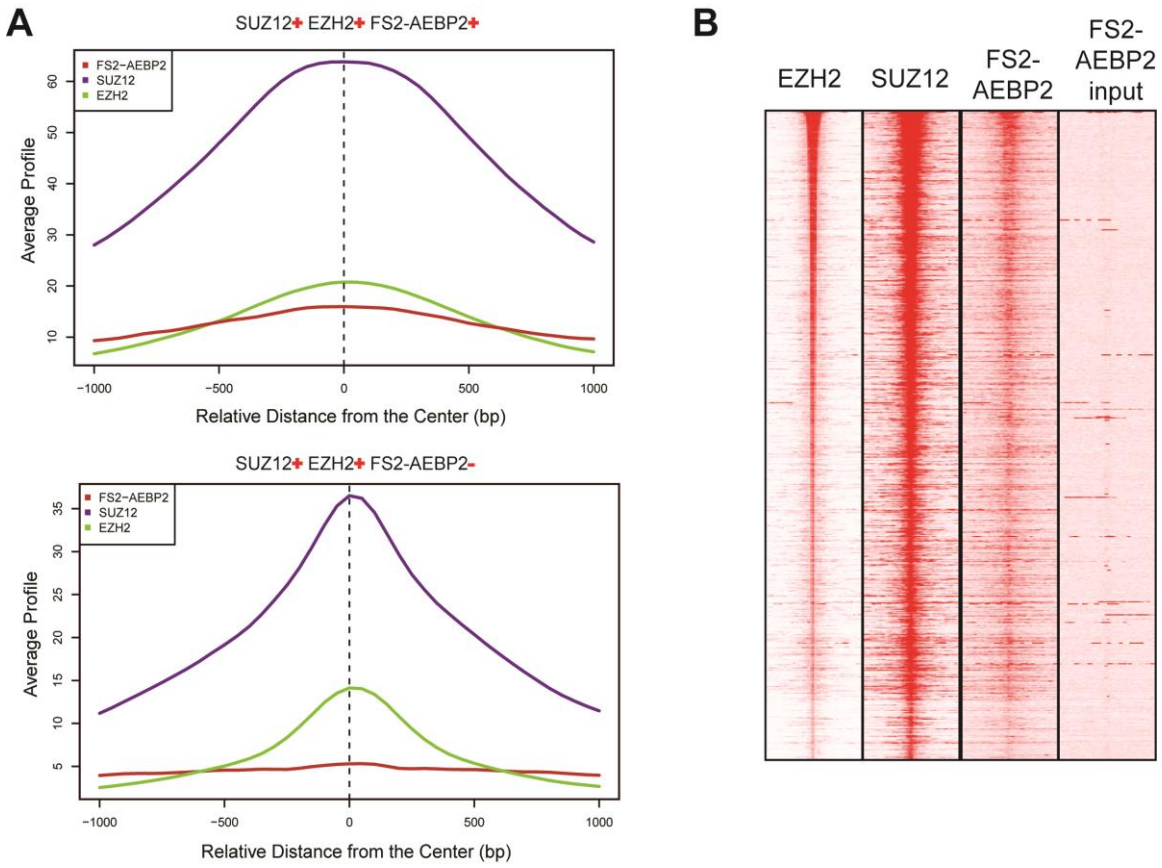


Figure 4.15: "SUZ12 positive, EZH2 positive, FS2-AEBP2 negative" regions still show some FS2-AEB2 enrichment. (A) SUZ12 and EZH2 peaks associated with FS2-AEBP2 are generally broader and show higher enrichment levels of EZH2 and SUZ12 compared to SUZ12 and EZH2 peaks at sites not enriched for FS2-AEBP2. (B) Heatmap showing the binding profile +- 5kb from EZH2 peaks for EZH2, SUZ12, FS2-AEBP2 and input for FS2-AEBP2.

4.5 Discussion

I have shown that AEBP2 becomes enriched on a chromosome that is undergoing *Xist* transgene mediated inactivation. Furthermore, I have observed co-localisation between AEBP2, EED and H3K27me3 to the presumed inactive X during female ES cell differentiation. Taken together, this provides evidence further to already published data that AEBP2 is a *bona fide* partner of PRC2 and likely has a role in ES cells during XCI.

Further analysis of these processes could provide more insight into AEBP2 function. It would be interesting to assess the exact timing of AEBP2 association with the presumed inactive X, so that this can be matched against the timing of PRC2 recruitment. Importantly, analysis of an AEBP2-depleted cell line would provide insight into whether AEBP2 is required or instructive in PRC2 recruitment to the inactive X. Conversely, it would be interesting to know whether the recruitment of AEBP2 is in fact dependent on PRC2 components (discussed more in Chapters 5 and 7). Additionally, I did not analyse AEBP2 localisation in embryos. The female ES cell differentiation only captures random XCI, and therefore validation of my findings in the developing embryo at time points when imprinted XCI takes place and at time points when random XCI takes place would be useful.

I have also shown that endogenous AEBP2 is targeted to PRC2-enriched genes in ES cells. Using an FS2-tagged AEBP2, I have furthermore shown that AEBP2 is enriched at promoters genomewide and is enriched near genes that play roles in development. These sites of enrichment overlap with PRC2 targets. From my analysis, it seems likely that AEBP2 is not enriched at any other locations. Whether there are any PRC2 targets that do not show enrichment for FS2-AEBP2 is unclear at this point. Comparing the traces from my experiment and those performed by Peng et al. (2009) (Figure 4.14), it is clear that the FS2-peaks are generally less well defined, and this results in quite different peak calling behaviour by the MACS2 algorithm. It may be possible to better optimise the parameter settings to improve peak calling, which would likely result in a higher degree of overlap. It may also be that the current dataset is of insufficient quality to confidently make statements about whether there are any PRC2 targets that do not show enrichment for FS2-AEBP2. A biological repeat of this experiment will increase the confidence with

which I might call regions FS2-AEBP2 negative. In the future it may be possible, with further optimisation, to use endogenous AEBP2-associated DNA immunoprecipitated with the antibody described in this study for ChIP-seq analysis.

I have shown that AEBP2 localises, together with PRC2 components, to an inactivated X chromosome in a model of X chromosome inactivation. Furthermore, I have shown it is probable that AEBP2 is recruited to the inactive X during endogenous X chromosome inactivation, but further analysis is required to confirm these observations. I have also shown that AEBP2 localises to PRC2 targets by ChIP. Together, these observations suggest that AEBP2 may have a role in PRC2 biology.

5. Composition of PRC2 complexes

5.1 Introduction

In recent years, a number of PRC2 associating factors have been identified, including AEBP2. I set out to validate these findings and to ask whether AEBP2 might also associate with other proteins. Furthermore, it has been reported that in the absence of EED, the PRC2 complex becomes destabilised and some of the components are degraded. I therefore assessed AEBP2 stability and localisation in the absence of EED.

5.1.1 Association of AEBP2 with PRC2

Typically, the PRC2 complex is described as consisting of four core components (*Drosophila melanogaster* counterparts in brackets): EZH2/EZH1 (E(z)), SUZ12 (su(z)12), EED (esc and escl) and RBBP4/RBBP7 (p55). In addition, there are four different translational isoforms of EED (Montgomery et al., 2007). Therefore, there will be some variation in which E(z) homologue, which p55 homologue and which EED isoform is incorporated in any particular PRC2 complex and thus it is conceivable that different subclasses of PRC2 could exist with specialised roles. However, so far different roles have only been described for EZH2 and EZH1, where EZH2 expression is associated with embryonic stem cells and proliferative tissue and is down-regulated after birth, whereas EZH1 levels remain constant (Laible et al., 1997; Margueron et al., 2008). EZH2 has been reported to have higher catalytic activity for H3K27me3 deposition, whereas it has been proposed that EZH1 can also repress transcription by chromatin compaction (Margueron et al., 2008; Shen et al., 2008). However, no differentiating roles

for the other PRC2 homologues have been shown. It is therefore unclear how many different subclasses of PRC2 exist, and whether there are PRC2 components that exclusively associate with one PRC2 component and not others, and whether any subclasses may have different roles.

In addition to differences in the presence of homologues of the four perceived core components, there may also be differences in the association of co-factors of the PRC2 complex. A series of papers identified AEBP2 as a PRC2 associating factor (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li et al., 2010; Walker et al., 2010; Hunkapiller et al., 2012). An overview of these studies is presented in Table 14. All studies that used PRC2 core component as bait identified AEBP2 as an associating factor, except for one study by Zhang and colleagues. Similarly, all studies that co-immunoprecipitated tagged JARID2 also found AEBP2, yet they rarely found a polycomblike family protein. Noticeably, studies that used a member of the polycomblike family as bait never co-immunoprecipitated AEBP2. Furthermore, the PRC2-associating factor esPRC2p48 was only detected in immunoprecipitations on the core PRC2 components or the polycomblike proteins. Together, this suggests that there may be proteins that can associate with PRC2 in a mutually exclusive way and indicates there may be subclasses of PRC2 complex.

Study	Cell line	Bait	Identify peptides of AEBP2?	Identify peptides of AEBP2 exon 1b?	Also identify
Peng et al. (2009)	Mouse ES cells	Flag-EED	Yes	Yes	JARID2
Shen et al. (2009)	Mouse ES cells	Biotin-Flag-EED	Yes	No/Not published	JARID2, esPRC2p48, PCL2
Shen et al. (2009)	Mouse ES cells	Biotin-Flag-EZH2	Yes	No/Not published	JARID2, esPRC2p48, PCL2
Shen et al. (2009)	Mouse ES cells	Biotin-Flag-EZH1	Yes	No/Not published	JARID2, esPRC2p48, PCL2
Shen et al. (2009)	Mouse ES cells	Biotin-Flag-JARID2	Yes	No/Not published	PCL2
Shen et al. (2009)	Mouse ES cells	Biotin-Flag-PCL2	No	N/A	esPRC2p48
Li et al. (2010)	Human 293F cells	Flag-Ezh2	Yes	Not published	Jarid2, Pcl1, Pcl2
Pasini et al. (2010)	Human HeLa	Flag-HA-Suz12	Yes	Not published	Jarid2, Hdac2
Pasini et al. (2010)	Human 293T	Flag-HA-Jarid2	Yes	Not published	Hdac2
Landeira et al. (2010)	Mouse ES cells	Flag-EED (isoform 3)	Yes	Not published	JARID2, PCL2
Landeira et al. (2010)	Mouse ES cells	Flag-JARID2	Yes	Not published	PCL2, no EZH1
Zhang et al. (2011)	Mouse ES cells	Purified PRC2 and then used endogenous SUZ12 as bait	No	N/A	JARID2, MTF2, esPRC2p48
Walker et al. (2010)	Mouse ES cells	6xHis3xFlag-SUZ12	Yes	Yes	PCL2
Brien et al. (2012)	Human 293T and HMEC cells	Flag-HA-Phf19 (PCL3)	No	N/A	Amongst others, No66
Ballaré et al. (2012)	Human 293T cells	Tap-Phf19 (PCL3)	No	N/A	Not published
Cai et al. (2013)	Human 293	Phf1 (PCL1)	No	N/A	Hdac1
Casanova et al. (2011)	Mouse ES cells	PCL2-FLAG	No	N/A	Amongst others, esPRC2p48, SALL4
Hunkapiller et al. (2012)	Mouse ES cells	TAP-SUZ12 (endogenous)	Yes	Not published	JARID2, PCL2, PCL3, esPRC2p48

Table 14: Mass spectrometry analysis of immunoprecipitations in published studies. In some cases not all identified proteins are listed. If peptides were published, the dataset was inspected to see whether peptides from the 1b exon of AEBP2 were detected. In the study published by Shen et al. (2009), it is not clear whether the published peptides solely align to exons 2-8, or whether the 1b exon sequence was simply not included in the alignments.

A study performed in the Vermeulen lab (Smits et al., 2013) set out to quantify the stoichiometry of PRC2 components. Using a GFP-tagged EED in HeLa cells, they used a combination of intensity-based absolute quantification (iBAQ) and label free quantification of single affinity enrichments to determine that the core components EED, SUZ12 and EZH2/1 were present in a 1:1:1 ratio. The 'core component' RBBP4/7 was present in 60% of the complexes, suggesting that it might not be a core component in the strictest sense. Furthermore they found the above-mentioned co-factors to be present: PCL1/2/3 (35%), esPRC2p48 (47%), JARID2 (12%) and AEBP2 (20%). This shows that certainly not all co-factors are present in every PRC2 complex, but this experiment does not show the stoichiometry of subclasses of PRC2 complexes. Also this experiment was performed in HeLa cells, and PRC2 composition might be different in ES cells.

5.1.2 The PRC2 complex in the absence of EED

Because the proteins EZH2/1, SUZ12 and EED make up the core of the PRC2 complex and are not part of any other complexes as far is known, many studies have been performed on the effects of the genetic removal of one of these components. Especially, much work has been done on *Eed*$^{-/-}$ cells. EED was identified as the murine homologue of *Drosophila* Esc by Magnuson and colleagues (Schumacher et al., 1996). The authors made use of mutants generated using ethyl-*N*-nitrosurea (ENU). A mutant denoted as *l7Rn5*^{3354SB}, which gave an early embryonic lethality phenotype, was shown to contain a mutation at position 1040 of the *Eed* gene, causing a transition from thymine to cytosine. This nucleotide change results in a mutation of amino acid 196 in the EED protein from a

leucine to a proline. This amino acid is located in the second of seven WD repeats. It has been shown that the WD repeats, which are formed from the C-terminal 91-441 amino acids of EED, fold into a seven-bladed β-propeller domain (Margueron et al., 2009). Presumably the L196P mutation causes destabilisation and subsequent proteolytic degradation of EED, as it has been shown that EED protein is undetectable in protein extracts in the *l7Rn5*^{3354SB} cells whilst levels of *Eed* mRNA are unchanged (Montgomery et al., 2005). Therefore these cells are known as *Eed*$^{-/-}$ cells.

In *Eed*$^{-/-}$ cells, H3K27 tri- and dimethylation is no longer detected, whilst monomethylation is still present, at least in early passage cells (Silva et al., 2003; Montgomery et al., 2005; Chamberlain et al., 2008). It has also been shown that EZH2 requires EED, as well as SUZ12, for PRC2-mediated H3K27 methylation *in vitro* (Cao and Zhang, 2004). Interestingly, *Eed*$^{-/-}$ cells also showed a strong reduction in EZH2 protein and not in *Ezh2* mRNA, and were also reported to have reduced SUZ12 protein (Montgomery et al., 2005). *Eed*$^{-/-}$ cells, however, are viable and retain pluripotency (Chamberlain et al., 2008). Due to absence of H3K27me3 and H3K27me2 and the PRC2 core subunits, it has generally been assumed that *Eed*$^{-/-}$ cells are effectively PRC2-depleted. Because AEBP2 had been proposed to function as a targeting factor for PRC2, I set out to determine whether AEBP2 could still localise to its normal targets in the absence of PRC2.

5.2 Association of AEBP2 with PRC2

As a first step to establish the association of AEBP2 with the PRC2 complex, nuclear cell extracts from E14 mouse embryonic stem cells were analysed by gel filtration using a Superose 6 gel filtration column. It was found that AEBP2 eluted at similar fractions to

other the components, whilst also eluting at further fractions representing smaller complexes (Figure 5.1). This indicated that a proportion of AEBP2 was in complexes of similar size to PRC2, whilst another proportion of AEBP2 in the cell might be in smaller complexes or not associated with other proteins.

In subsequent work, I found that in homozygous *Aebp2* gene-trapped cells, the band I presumed to be AEBP2 was still present. However, the slightly faster migrating band immediately below it, which also co-elutes with the PRC2 complex, was absent (see also Chapter 6). With this in mind, I conclude that AEBP2 is present only in complexes of similar size to PRC2.

Subsequently, I immunoprecipitated AEBP2 from PGK12.1 cells and analysed the immunoprecipitate for co-immunoprecipitation of EZH2, a PRC2 component, and YY1, a component known not to associate with the PRC2 complex (Li et al., 2010; Casanova et al., 2011). As expected, AEBP2 co-immunoprecipitates EZH2, also in the presence of a nuclease demonstrating that this interaction is not DNA-dependent (Figure 5.2). AEBP2 did not co-immunoprecipitate the negative control YY1.

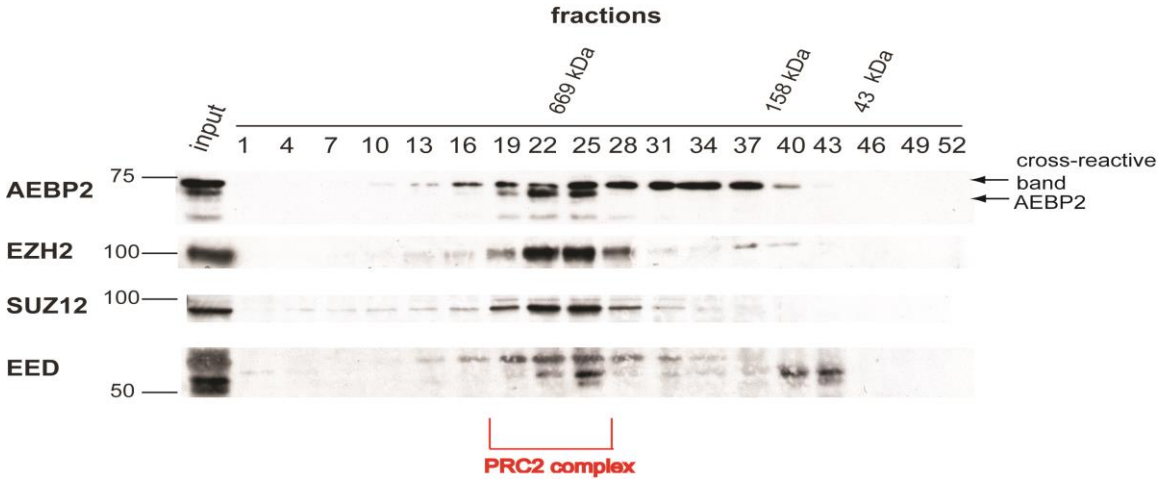


Figure 5.1: AEBP2 co-elutes with PRC2 components from a Superose 6 Gel Filtration column. Nuclear Cell Extract was prepared from E14 ES cells.

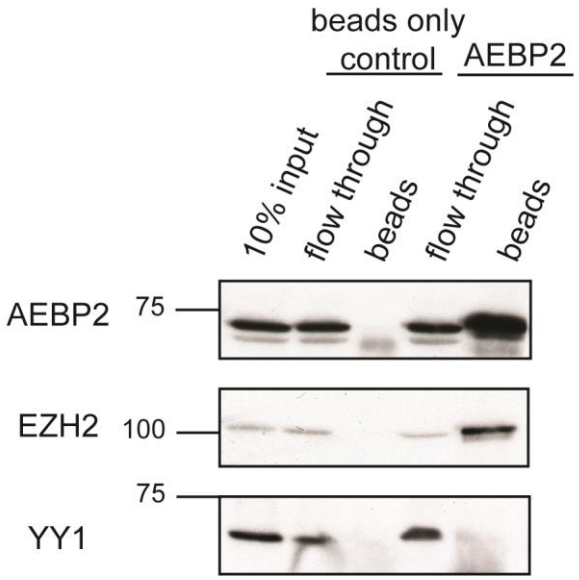


Figure 5.2: AEBP2 co-immunoprecipitates EZH2, but not YY1 from whole cell extract from PGK12.1 ES cells.

Subsequently, I decided to use the FS2-tagged AEBP2 described in 4.4 to identify AEBP2-interacting proteins in a more unbiased manner. Nuclear cell extract from overexpressing FS2-AEBP2 E14 ES cells or FS2-only (empty vector) control cells were passed over a StrepTactin superflow high-capacity resin and the elution was analysed by

Sypro-Ruby staining (Figure 5.3A). In the FS2-AEBP2 sample, several bands of high molecular weight were visible that were not detected in the empty vector sample. Immunoblotting against the strep tag showed retention of a doublet in the FS2-AEBP2 sample, with some degradation products too (Figure 5.3B). A band of similar molecular weight was also seen in the empty vector, which is a background band also seen in other experiments in our and Rob Klose's laboratory performed using the same tag system. Importantly, it was shown that the FS2-AEBP2 elutions contained EZH2, whereas the empty vector elutions did not (Figure 5.3C).

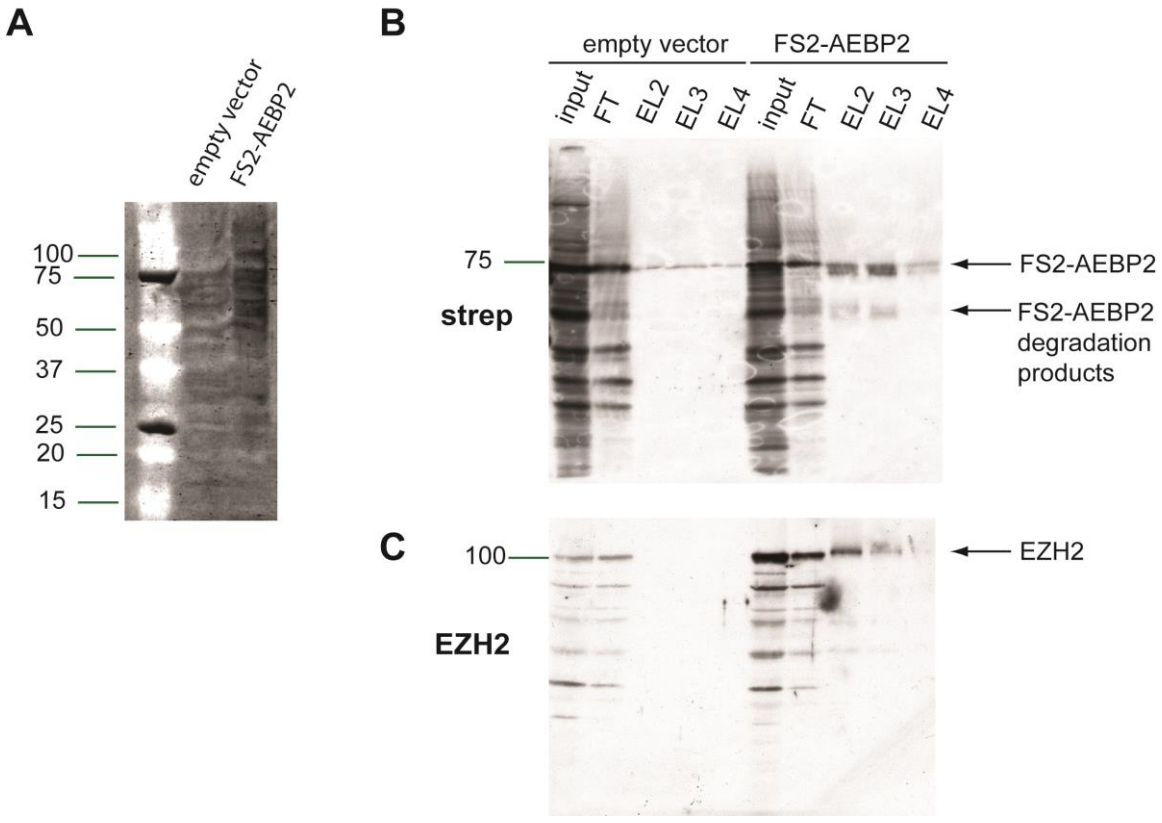


Figure 5.3: FS2-tagged AEBP2 associates with EZH2 and other high molecular weight proteins. (A) Sypro-Ruby gel showing proteins associated with FS2 only and FS2-AEBP2 in E14 ES cells in an elution from a StrepTactin resin. (B, C) Elutions (EL2-4) from the StrepTactin high capacity resin contain FS2-tagged AEBP2 and EZH2. Molecular weights indicated in kDa.

The elutions were analysed by mass spectrometry to identify associated proteins. Two biological repeats identified all the core components of the PRC2 complex (Figure 5.4). In addition, JARID2 was identified. Although exponentially modified Protein Abundance Index (emPAI) is not an absolute quantification, the score for JARID2 suggests that not all AEBP2 complexes contain JARID2, i.e. JARID is not an obligate partner in AEBP2-PRC2 complexes (Ishihama et al., 2005). The emPAI for RBBP4 and RBBP7 is also lower than what you would expect for a core component, in line with the observation made by Smits et al. (2013) that RBBP4/7 is present in only 60% of PRC2 complexes in HeLa cells.

Strikingly, no peptides of any of the polycomblike family proteins were identified, nor were there any peptides of esPRC2p48. This therefore indicates that there are likely different PRC2 complexes, one subclass containing proteins of the polycomblike family and/or esPRC2p48, and one subclass containing JARID2 and/or AEBP2.

In addition, no peptides of any other proteins were identified with high confidence. It therefore seems unlikely that AEBP2 participates in any other complexes that are stable under high salt conditions. However, it is possible that AEBP2 might form more transient interactions with other proteins (see discussion).

A

Experiment 1

Protein	Mascot Score	Peptides	Mass (kDa)	emPAI
SUZ12	1822	73	83.7	3.83
AEBP2	**1686**	**56**	**53.5**	**2.51**
EZH2	1129	55	87.2	2.25
EZH1	292	17	87.2	0.51
EED	692	47	50.9	3.50
JARID2	640	38	139	0.29
RBBP4	411	27	47.9	1.71
RBBP7	184	12	48.1	0.49

Experiment 2

Protein	Mascot Score	Peptides	Mass (kDa)	emPAI
AEBP2	**1694**	**59**	**53.5**	**4.02**
SUZ12	1393	69	83.7	3.14
EZH2	826	43	87.2	1.42
EZH1	147	13	87.2	0.34
EED	606	43	50.9	3.23
JARID2	532	24	139	0.35
RBBP4	172	18	47.9	0.59
RBBP7	138	9	48.1	0.39

B

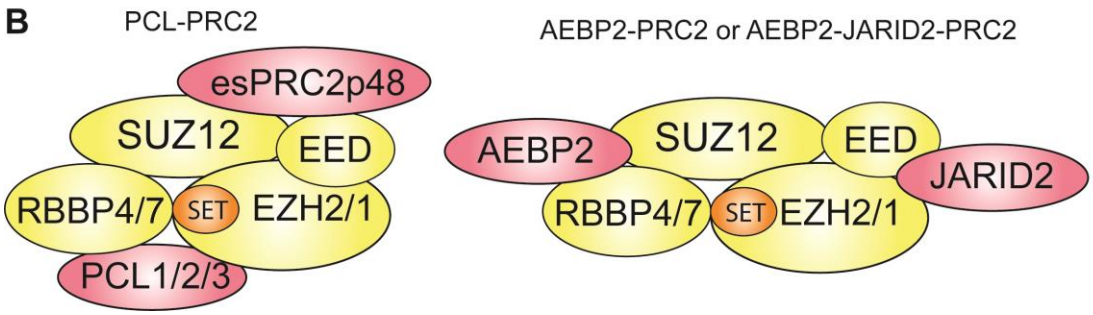


Figure 5.4: AEBP2 associates with the PRC2 complex but excludes proteins from the polycomblike family. (A) Proteins identified in FS2-AEBP2 but not in vector only in two biological repeats. (B) Proposed model of subclasses of PRC2.

5.3 The PRC2 complex in the absence of EED

I next analysed the effects of EED depletion on PRC2 and AEBP2. Firstly, I analysed protein levels of H3K27me3 and EZH2, SUZ12 and EED in *Eed*$^{+/+}$ and *Eed*$^{-/-}$ lines that our laboratory had received from Anton Wutz's laboratory (Schoeftner et al., 2006) (Figure 5.5). It was confirmed that H3K27me3 levels are virtually undetectable. Surprisingly, I found that whilst EZH2 was reduced, it was not completely absent, as some previous reports had indicated. This observation was confirmed by using a different antibody, also raised against EZH2. Furthermore, EZH2 was also still present in an *Eed*$^{-/-}$ cell line derived in our laboratory (data not shown). It was also observed that SUZ12 levels were not reduced.

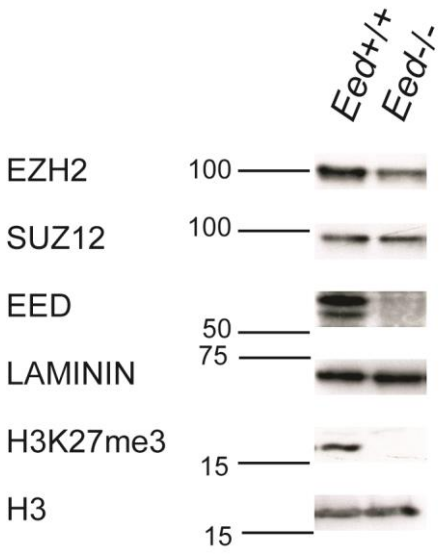
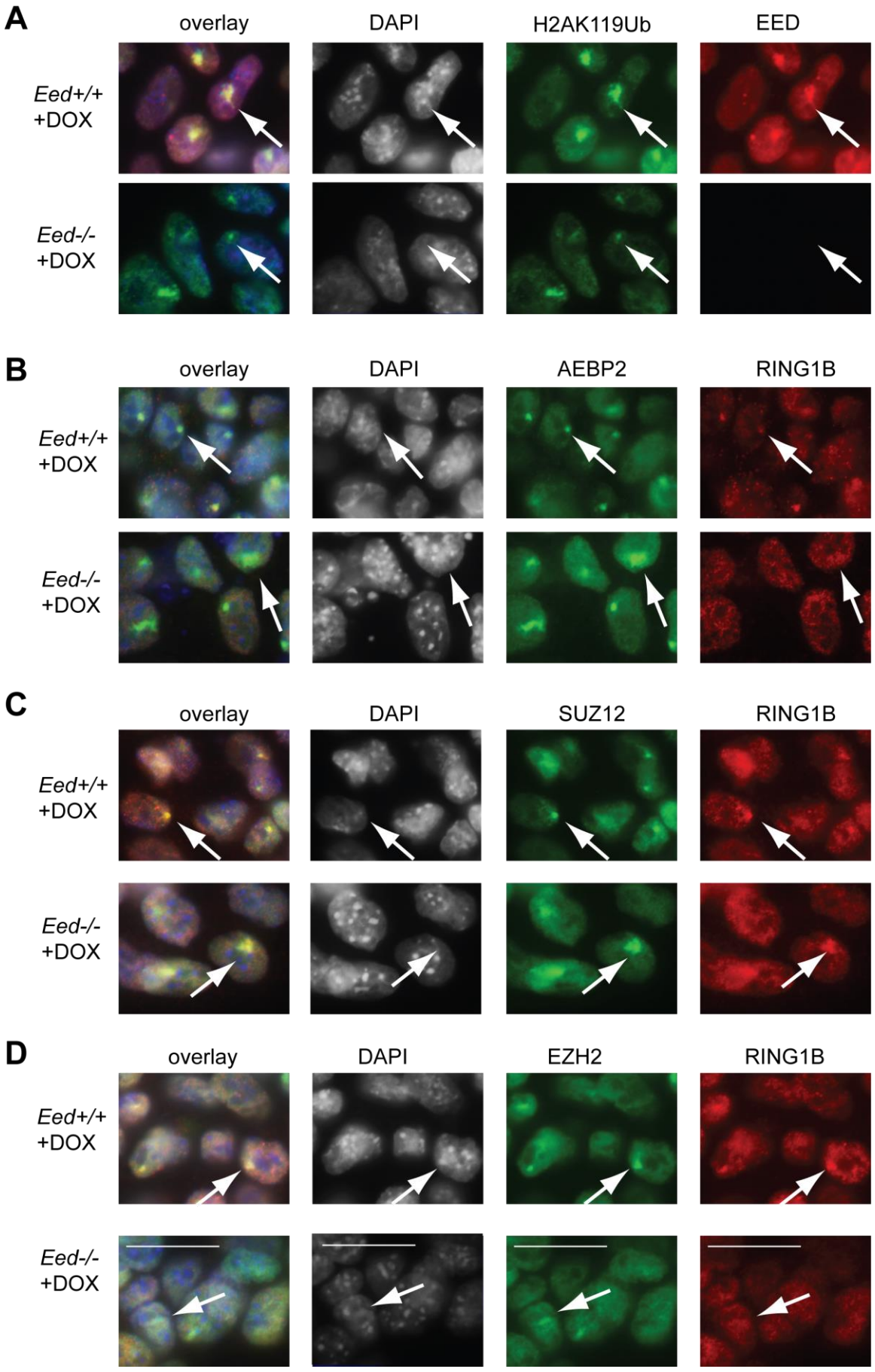


Figure 5.5: Levels of PRC2 components and H3K27me3 in *Eed*$^{+/+}$ and *Eed*$^{-/-}$ cells. H3 and laminin were used as a loading control.

The *Eed*$^{-/-}$ cell line from Anton Wutz, and its control line *Eed*$^{+/+}$, have an inducible *Xist* transgene on chromosome 11 as shown in Figure 4.6. I therefore decided to use this

system to assess what happens to PRC2 recruitment to the inactivated chromosome in the absence of EED. Surprisingly, I found that upon 48 hours doxycycline treatment of *Eed*$^{-/-}$ cells, SUZ12 and AEBP2 were consistently recruited to inactivated chromosome like foci, albeit at somewhat reduced levels (Figure 5.6). Occasionally, accumulation of EZH2 could also be seen in the *Eed*$^{-/-}$ cells. EED protein, as expected, was not seen to accumulate in the *Eed*$^{-/-}$ cells. This suggests that PRC2 components in the absence of EED can still localise to the inactivated chromosome.


A
overlay
DAPI
H2AK119Ub
EED
Eed+/+
+DOX
Eed-/-
+DOX
B
overlay
DAPI
AEBP2
RING1B
Eed+/+
+DOX
Eed-/-
+DOX
C
overlay
DAPI
SUZ12
RING1B
Eed+/+
+DOX
Eed-/-
+DOX
D
overlay
DAPI
EZH2
RING1B
Eed+/+
+DOX
Eed-/-
+DOX

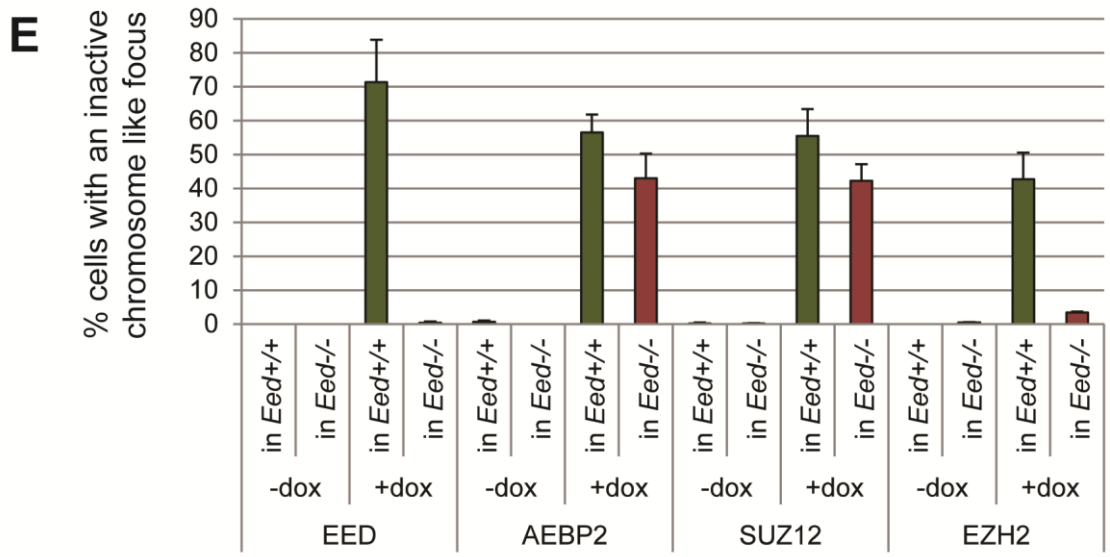


Figure 5.6: SUZ12 and AEBP2 are still recruited to an inactivated chromosome in the absence of EED. Immunofluorescence of male J1 ES cells WT for *Eed* (*Eed*$^{+/+}$) or knockout (*Eed*$^{-/-}$) stained for EED and H2AK119Ub (A), RING1B and AEBP2 (B), RING1B and SUZ12 (C), and RING1B and EZH2 (D). Scale bar is 20 µm. Arrows indicate the location of an "inactivated chromosome like" focus. Isolated DAPI staining is shown in greyscale for enhanced contrast. (E) Quantification of percentage of cells with an "inactivated chromosome like" focus. Error bars represent standard deviation of 3 biological repeats. A minimum of 117 cells was counted for each condition in each experiment.

I next analysed AEBP2 localisation in *Eed*$^{-/-}$ cells using ChIP-qPCR. Using the *Eed*$^{-/-}$ ESCs generated in our laboratory, I assessed what happens to endogenous AEBP2 and H3K27me3 (Figure 5.7). As expected, H3K27me3 was undetectable at any of the PRC2 targets. AEBP2 recruitment to PRC2 target genes was also severely diminished. It is difficult to judge whether it is completely absent, as a slight enrichment is still observed. However, experimental variability and a relatively low signal-to-noise ratio, combined with the fact that these cells grow on a layer of feeder cells, which could have been incompletely removed and result in residual signal, make it impossible to determine whether this slight enrichment is biologically relevant. This therefore suggests that AEBP2 localisation to PRC2 targets is severely diminished, if not abolished, in the absence of EED.

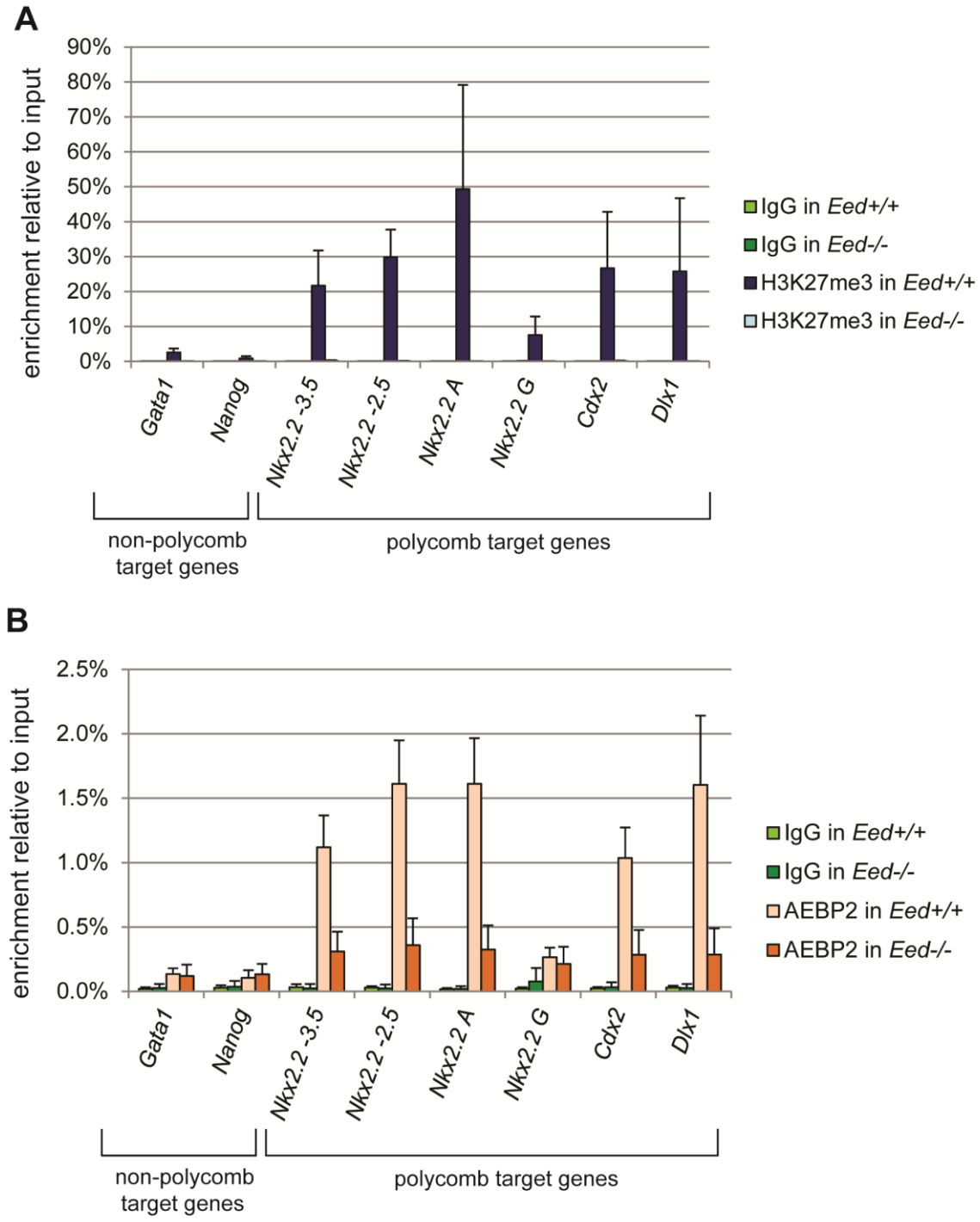


Figure 5.7: AEBP2 recruitment to polycomb target genes is severely diminished in the absence of EED. (A) ChIP-qPCR shows an absence of H3K27me3 in *Eed*$^{-/-}$ cells. (B) ChIP-qPCR shows reduced enrichment of AEBP2 at polycomb targets in *Eed*$^{-/-}$ cells. Error bars represent standard deviation of 3 biological repeats.

5.4 Discussion

I have shown that AEBP2 is a PRC2 associating factor, and can be found in PRC2 complexes also containing JARID2. Interestingly, no peptides of the polycomblike family or esPRC2p48 were found, and taking this with published results, it seems that there are different subclasses of PRC2. The exact composition of these proposed subclasses remains unclear. My data preliminarily indicates that not all AEBP2-containing complexes also have JARID2. However, based on results by Smits et al. (2013), it would seem that a smaller portion of PRC2 contains JARID2, so it would be interesting to know if all JARID2 complexes contain AEBP2. Conversely the quantitative relationship between polycomblike family members and esPRC2p48 would be interesting to study. Additionally, performing mass spectrometry on esPRC2p48 containing PRC2 complexes would be useful to confirm the hypothesized exclusion of JARID2 and AEBP2.

More structural information on how PRC2 associating factors interact with PRC2 components will enhance our understanding of PRC2 subclasses. The electron microscopy structure of PRC2 containing AEBP2 showed that AEBP2, together with SUZ12, forms a narrow arm connecting the upper half (consisting of EED, EZH2 and the C-terminal region of SUZ12) and the lower half (consisting of RBBP4 and the N-terminal region of SUZ12) (Ciferri et al., 2012). Further extensive contacts were seen between the SET domain of EZH2 and a loop between the second and third zinc finger of AEBP2. It is currently unknown where the other PRC2-associating proteins bind the PRC2 complex. JARID2 has been suggested to bind to EZH2 based on *in vitro* co-immunoprecipitation experiments and specifically to its SET domain based on deletions

experiments followed by co-immunoprecipitation (Shen et al, 2009). However, another group proposed that JARID2 binds to SUZ12 (Peng et al, 2009). It has been shown that *Drosophila melanogaster* Pcl binds E(z) directly (O'Connell et al., 2001; Nekrasov et al., 2007), although it may also interact with Su(z)12 and Nurf55, the RBBP4/7 homologue. In mammals, it has been shown that the PCL1 PHD fingers, but not PCL2 PHD fingers, interact with EZH2 in a yeast-two-hybrid assay (O'Connell et al., 2001), leaving unclear how PCL2 and PCL3 may bind PRC2 if not in the same place. There are no published reports with regards to where esPRC2p48 might bind the PRC2 complex. Structural information could provide more insight into the different subclasses; for example it could indicate whether associating factors are competing for particular binding surfaces on PRC2 or conversely, whether associating factors might bind cooperatively.

It would be interesting to see whether different EED translational isoforms partake in different PRC2 complexes. Landeira and colleagues (2010) reported that they enriched for EED-3 upon immunoprecipitation of JARID2. In ES cells, three of the four translational isoforms are expressed: EED-1, which contains an additional N-terminal 59 amino acids and is translated from a non-consensus GUG (Leucine) start site, EED-3, which is the 441 amino acid isoform, and EED-4, which misses the first 14 N-terminal amino acids (Montgomery et al., 2007). No peptides corresponding to EED-1 or EED-2 were detected in the FS2-AEBP2 immunoprecipitate.

What could be the function of different subclasses of PRC2? JARID2 has been shown to have a role in PRC2 recruitment, as JARID2-depleted cells have reduced occupancy of PRC2 at PRC2 target genes (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li

et al., 2010; Pasini et al., 2010). Polycomblike proteins have also been shown to recruit PRC2 and have been especially linked to recruiting PRC2 to regions enriched for H3K36me3 because the tudor domains of PCL1 and PCL3 have been shown to bind to H3K36me3 (Brien et al., 2012; Cai et al., 2013). It has also been suggested that PCL proteins are especially important for the conversion of H3K27me2 into H3K27me3 (Nekrasov et al., 2007; Sarma et al., 2008). No function has been described for esPRC2p48 other than a possible stimulation of H3K27me3 deposition (Zhang et al., 2011). With respect to the different isoforms of EED, it has been shown that the N-terminal region (including the region that distinguishes the different isoforms) is not required for H3K27 methylation, but instead the WD40 motifs are essential (Montgomery et al., 2007). Taken together, it is difficult to speculate on the potential different functions of subclasses of PRC2. Given that the associating factors appear to be present at all PRC2 targets (see 4.1.2) it seems unlikely that the different complexes are responsible for targeting PRC2 to different regions of the genome. Instead, they may be responsible for recognizing different features of chromatin that mark locations for PRC2-mediated silencing. Alternatively, one subclass may be involved in recruitment, whilst another class may represent a more enzymatically active version of PRC2. It is also possible that different associating factors may direct PRC2's methylating activity to different (non-histone) substrates. For example, it has been shown that PRC2 can methylate GATA4 (He et al., 2012) and JARID2 (R. Margueron, personal communication). Further research is required to establish the roles of different PRC2 subclasses.

Remarkably, no proteins other than listed in Figure 5.4 were detected with high confidence. This indicates that the PRC2 complex is probably the only complex AEBP2

stably interacts with under these conditions. Some proteins have been reported to have lower affinity interactions with the PRC2 complex (Brien et al., 2012; Dietrich et al., 2012; Wang et al., 2013) and these interactions might be captured experimentally by reducing the stringency of several steps in the immunoprecipitation protocol, such as reducing the salt concentration during protein extraction from cell nuclei, or by reducing the number of washing steps. This may provide insight into whether AEBP2 is involved in recruiting other transiently associating proteins to the PRC2 complex.

Having established that AEBP2 interacts with the PRC2 complex, I assessed the effects of EED depletion. It has been reported that EED depletion leads to almost complete absence of EZH2 and a reduction in SUZ12 protein levels. However, I found that EZH2 levels were still detectable, but were decreased, whilst SUZ12 levels seem to be equal in the wild type and *Eed*$^{-/-}$ cells. This was surprising given reports from the literature. However, upon closer examination of other published reports, many examples could be found which supported the presence of EZH2 and SUZ12 in *Eed*$^{-/-}$ cells. For example, Boyer et al. (2006) also report a reduction of EZH2 and a stable level of SUZ12, as did Shen et al. (2008). Another report comes from Dietrich et al. (2012) who showed that in *Eed*$^{-/-}$ cells, EZH2 and SUZ12 levels are reduced, but still present, and furthermore that these proteins still exist in high molecular weight complexes, suggesting that a partial PRC2 complex is still formed. Also, Orkin and colleagues demonstrated that SUZ12 can co-immunoprecipitate EZH2 in Eed-/- cells (Shen et al., 2008) and it was shown that in the absence of EED, JARID2 can still co-immunoprecipitate EZH2 and SUZ12, further supporting the idea that a partial PRC2 complex can exist (Shen et al., 2009).

These reports are consistent with my observation that in *Eed*$^{-/-}$ cells, SUZ12, AEBP2 and, to a lesser extent EZH2, can still be recruited to an inactivated chromosome by *Xist* transgene expression. The recruitment of SUZ12 and AEBP2 was somewhat reduced in *Eed*$^{-/-}$ cells, suggesting that EED is still important for the recruitment of SUZ12 and AEBP2, whilst EZH2 recruitment was severely reduced. Severe reduction in EZH2 recruitment is likely at least in part due to EZH2 depletion at the protein level. Furthermore, antibody sensitivity might determine a threshold of signal-to-noise below which foci can no longer be detected, and differences between different antibodies may lead to contrasting apparent levels of recruitment. For example, originally it was reported that a very low level of SUZ12 recruitment could still be seen in the *Eed*$^{-/-}$ cells (Schoeftner et al., 2006). I believe that an improved antibody has allowed me to observe a higher level of SUZ12 recruitment. An illustration of the fact that antibody sensitivity may be important is given by the fact that EZH2 foci in *Eed*$^{+/+}$ cells are less well defined than those of SUZ12 and EED (Figure 4.6C), even though they are presumed to be present at equal levels.

The immunofluorescence data indicates that SUZ12 and AEBP2, together or independently, can still be recruited to their expected location. This therefore suggests that SUZ12 and AEBP2 recruitment is at least in part independent of EED. The fact that I see occasional EZH2 recruitment also indicates that EZH2 retains some of its expected behaviour in the absence of EED. However, from my experiments and published reports, it is clear that there is virtually no H3K27me3. Therefore, it remains to be determined whether this recruitment and potentially partially forming PRC2 complex has any functional significance or not. It has been reported that EZH1 can compact chromatin independent of a functional SET domain. It is therefore possible that recruitment of an

EZH1-SUZ12 (and possibly AEBP2 containing) complex could still contribute to silencing. It would therefore be interesting to test whether EZH1 and also other PRC2-associating factors still localise to the inactivated chromosome.

Given my observation that AEBP2 can still be recruited to an inactivated chromosome in the absence of EED, it was somewhat surprising to see a severe reduction, if not a complete abolition, of AEBP2 enrichment at PRC2 targets by ChIP-qPCR. One explanation for this, as discussed, might be that when there is very little AEBP2 present, the antibody might not be sensitive enough to distinguish signal from background. It would be interesting to also test SUZ12 and EZH2 enrichment at these locations by ChIP. It could also be possible that in the absence of EED, PRC2 components are still recruited to the chromatin but are not stabilised there, which could render them more refractory to the cross-linking conditions used for ChIP experiments. An alternative explanation is that the inactivated chromosome presents a unique environment that does not reflect PcG recruitment at other genic loci. Experiments with a similar system suggest that H3K27me3 levels are slightly reduced at the endogenous *HoxC* locus when *Xist* transgene expression is induced at another chromosome in an $Eed^{+/+}$ background (A. Cerase, personal communication). This might suggest that the *Xist* transgene expression recruits the PRC2 complex so effectively it acts as a sink, which would in turn suggest that the PRC2 recruitment levels are above what would be considered physiological. To circumvent this possible issue, it would be interesting to look at SUZ12, EZH2 and AEBP2 recruitment to the inactive X in differentiating $Eed^{-/-}$ XX ES cells.

It therefore remains a challenge to understand what is happening to the PRC2 complex in the absence of EED. Insight into this issue is important because the use of *Eed*$^{-/-}$ cells as a system in which there is presumed to be no PRC2 functionality is widespread. Knowing that PRC2 protein levels are maintained (to an extent), that they are likely still interacting (Shen et al., 2008; Dietrich et al., 2012) and may still be recruited to at least some appropriate locations, may alter the interpretation of some of these experiments. Additionally, understanding the nature of the PRC2 complex in the absence of EED might shed light on H3K27me3-independent functions. Further experiments could include making double knockouts of PRC2 components and observing whether the cells display a more severe phenotype.

I have shown that AEBP2 interacts with the PRC2 complex, but appears to exclude components of the polycomblike family and esPRC2p48. This observation, combined with published observations, establishes that there are PRC2 subclasses, which leaves the question of whether these subclasses may have different functions. Additionally, I have shown that by immunofluorescence AEBP2, as well as SUZ12 and to a small extent EZH2, can be recruited to an inactivated chromosome in the absence of EED. However, by ChIP it was observed that AEBP2 recruitment to endogenous PRC2 targets is negligible. More experiments are needed to establish what happens to the PRC2 complex in the absence of EED.

6. Generating an *Aebp2* conditional knockout ES cell line and mouse

6.1 Introduction

A common approach to understand the role of a protein is to genetically deplete it from the cell. I therefore decided to generate an *Aebp2* conditional knockout ES cell line and mouse. During this process, a report was published which used a publically available constitutive *Aebp2* gene-trap mouse (Kim et al., 2011). It was reported that no live homozygous gene-trap mice were born, and that at E10.5 and E14.5 no homozygote embryos were harvested, and from this the authors concluded that *Aebp2* depletion is embryonic lethal. The study did not describe the defects in *Aebp2* homozygote gene-trap embryos at earlier developmental stages, nor did the authors report the derivation of an embryonic stem cell line from the *Aebp2* homozygote gene-trap embryos. Therefore, I continued my studies to further investigate the effects of the genetic removal of *Aebp2*. Setting up a conditional system would allow me to investigate the role of AEBP2 after the developmental time point at which it causes embryonic lethality by inducing the removal of AEBP later in development. Therefore I generated a conditional knockout ES cell line and I am in the process of generating a conditional knockout mouse.

6.2 Targeting strategy

In order to construct a targeting vector, I made use of a gene-trap system that can be activated by induction of *Cre* (Figure 6.1). This system was set up in the Von Melchner

laboratory (Schnutgen et al., 2005) and has led to a number of publications (for example, Llano et al., 2008; Kohoutek et al., 2009; Remeseiro et al., 2012). In this system, the gene-trapping cassette is targeted in-frame with the first exon(s) of a gene. In the case of *Aebp2*, I chose to insert the cassette immediately 5' to exon 2. Because the cassette contains a very strong splice acceptor, the splicing machinery includes the genetic information in the cassette in the mature mRNA. The cassette consists of the *GFP* gene, fused to a *neo* gene, which is an aminoglycoside 3' phosphotransferase conferring resistance to kanamycin and geneticin (G418). The *neo* gene contains a stop codon and is followed by a transcription termination site and a poly-adenylation site, so that when the cassette exon is included, the transcript is truncated, and when this transcript is translated, translation stops after the *neo* gene (Figure 6.1A).

The cassette is flanked by frt and f3 sites, which are subject to FlpE mediated inversion/deletion, and loxP and lox5171 sites, which are subject to Cre mediated inversion/deletion (Figure 6.1B). Each enzyme can act on either type of site, but it must use the same sites in any recombination event, e.g. FlpE will act on a pair of frt sites or a pair of f3 sites, but not on a set of one frt and one f3 site (Schnutgen et al., 2003).

The orientation of the cassette controls whether the *Aebp2* transcript is being trapped. Therefore control over the orientation by FlpE and Cre allows for the generation of a conditional gene trap, as explained in detail below.

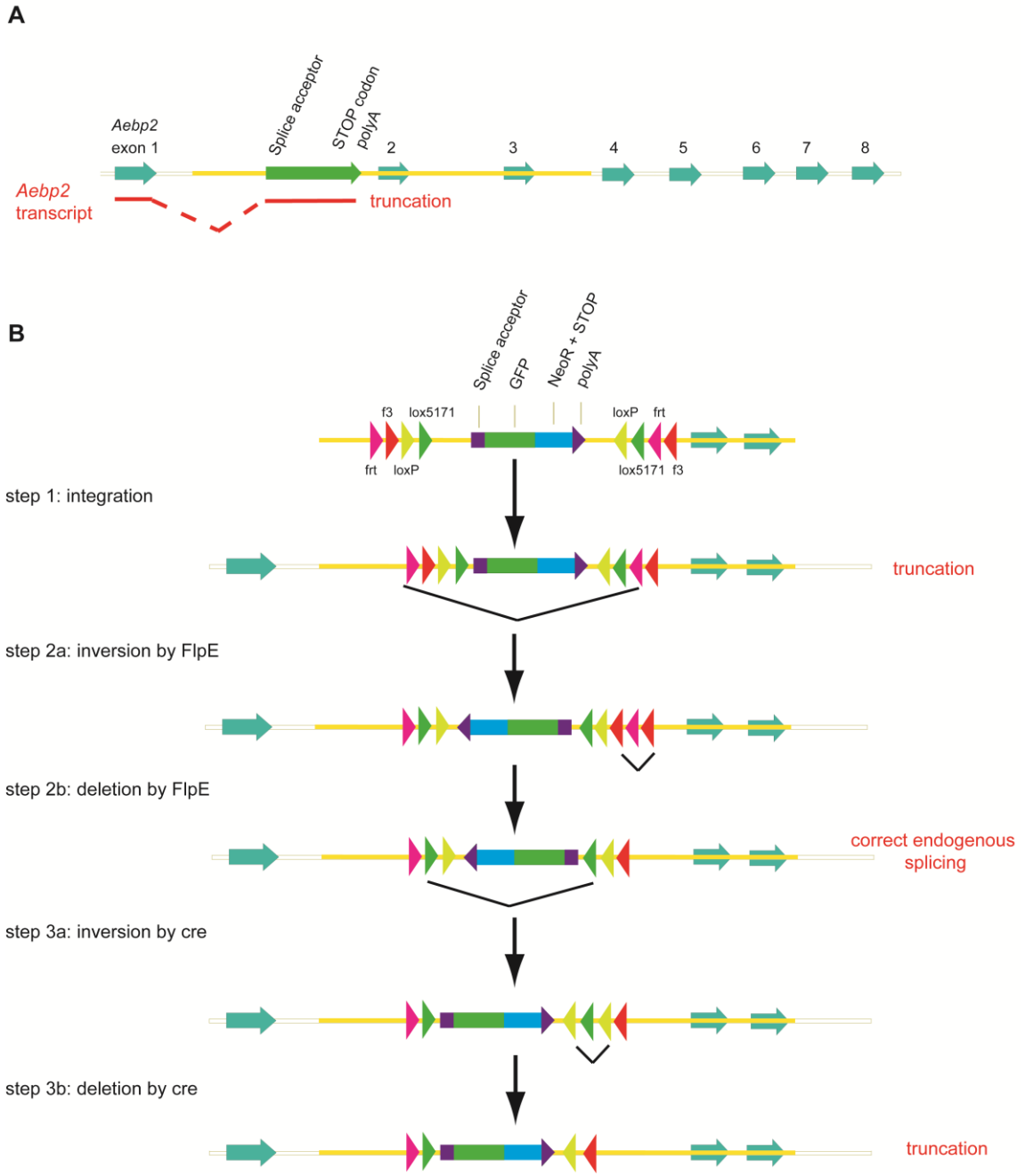


Figure 6.1: Targeting strategy to create *Aebp2* conditional knockout setup. (A) The gene-trapping cassette contains a splice acceptor that traps the *Aebp2* transcript after exon 1 and causes truncation of both the transcript and the protein. The yellow line indicates the arms of homology used for recombination. The turquoise arrows indicate exons of the *Aebp2* gene. The large green arrow is the trapping cassette. (B) After insertion of the gene-trapping cassette into the *Aebp2* locus, its orientation is controlled by the addition of FlpE and Cre. In red is indicated what is happening to the *Aebp2* transcript at each stage, depending on the cassette orientation. polyA: polyadenylation site. Black lines indicate inversion or deletion events.

In the first step the cassette is targeted to the correct genomic location. Subsequently, the cassette is inverted by FlpE using either the frt or f3 sites. If frt sites were used, the resulting configuration will contain two f3 sites in the same orientation flanking an frt site, which will lead to FlpE mediated deletion of sequence between f3 sites. Similarly, if f3 sites are used in the first inversion step by FlpE, the frt sites will be used for deletion. After this event, the cassette is effectively locked in an inverted orientation. Because the splice acceptor is now also in an inverted orientation, it no longer traps the transcript, and the *Aebp2* transcript is processed as normal. In this configuration the cells are also no longer geneticin resistant, which allows the second allele to be targeted and selected for by geneticin. Subsequent inversion of the cassette located in the second allele generates cells in which both *Aebp2* alleles contain an inverted cassette.

In a subsequent step, which can be in culture or *in vitro*, expression of *Cre* leads to the re-inversion of the cassette, which traps the transcript again. Therefore cells that contain the FlpE-inverted cassette in both alleles will become *Aebp2* knockout when *Cre* is expressed. As such the cassette creates a conditional knockout set-up.

6.3 Constructing a targeting vector

A targeting vector was constructed by Bacterial Artificial Chromosome (BAC) recombineering. An in-frame gene-trapping cassette was synthesised by the GeneArt® service of Invitrogen. The trapping cassette was flanked by short regions with homology to the intended insertion site in the *Aebp2* gene.

The selected site is 1.9 kb upstream of the second exon. The location of insertion site was chosen for several reasons. I wanted to ensure that trapping occurred before the zinc finger domain, which is encoded in exons 2-4. Another consideration was that if there is a transcript initiating 3' to the cassette then this transcript will not be trapped by the cassette. Because there are a number of different annotated start sites and different first exons, the cassette was designed to insert just before the second exon. The precise location was chosen because it contains non-repeat sequence so that the construct would recombine with unique sequence and not be inserted into off-target sites.

Targeting of the cassette to the BAC was selected for by kanamycin resistance and correct insertion was confirmed by restriction enzyme digestion which generated banding patterns as expected (Figure 6.2B). "BAC target 1" still contained some of the undigested vector with the trapping cassette (marked by an asterisk) and therefore I continued with "BAC target 2".

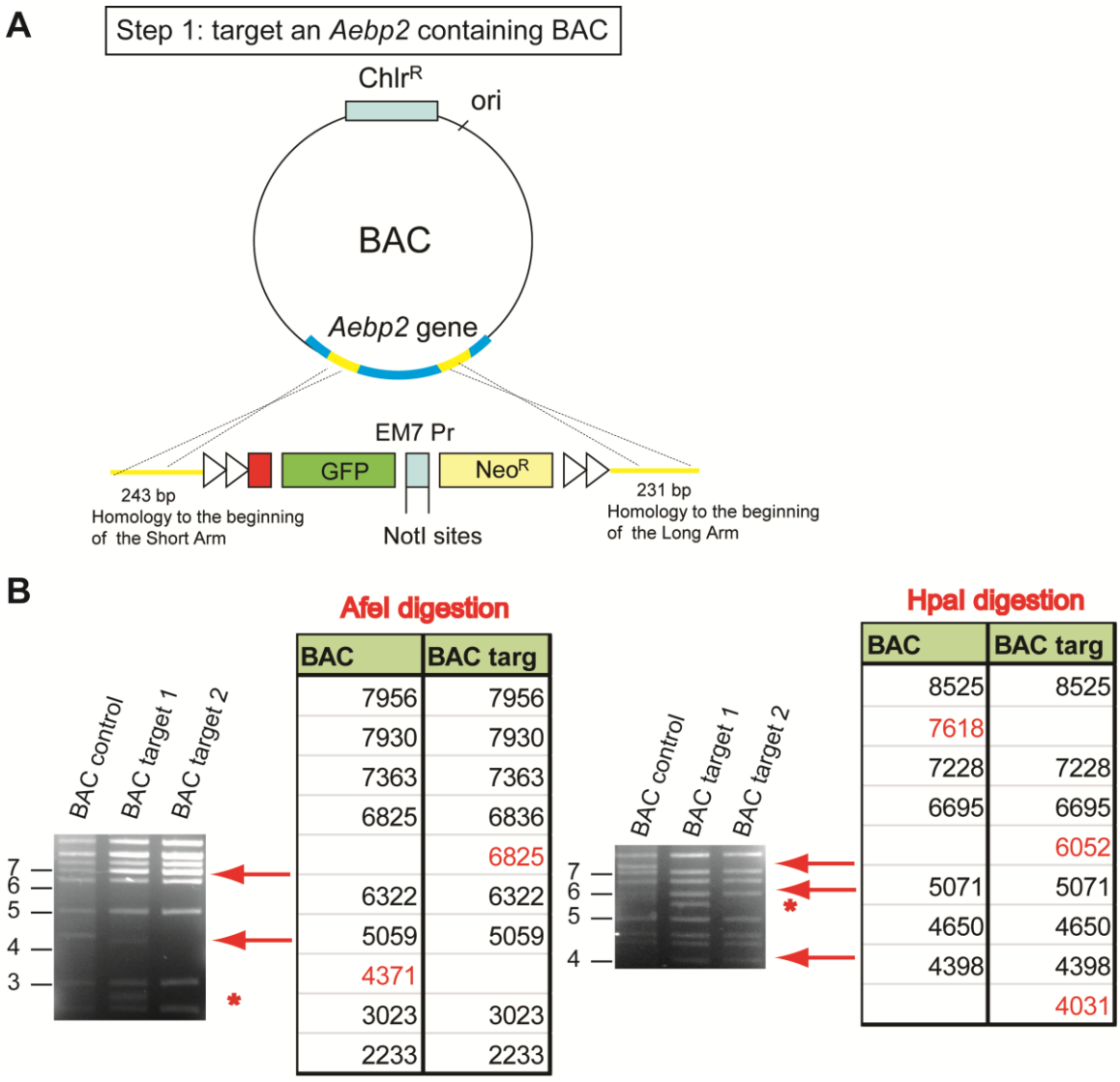


Figure 6.2: Targeting an *Aebp2* containing BAC with the trapping cassette. (A) Graphical representation of targeting step. A synthesised cassette containing a splice acceptor site, followed by *GFP*, a bacterial promoter flanked by NotI sites and a gene conferring resistance to kanamycin/geneticin, is inserted into a BAC containing the *Aebp2* gene using homologous recombination. (B) Digestions with AfeI and HpaI show that the BAC has been targeted correctly. A subset of expected band sizes and the corresponding part of the gel is shown. Sizes are shown in kb. Asterisks indicate digestion products from the intact vector containing the trapping cassette that contaminate the targeted BAC preparation. Sizes next to gel indicated in kbp, sizes in predicted digestion pattern (in table) in bp.

After targeting of the cassette to the BAC, the BAC was subcloned into the pMA vector from GeneArt®. This vector had also been synthesised with small arms of homology to the *Aebp2* gene 6.7 kb from the 5' end of the cassette and 10 kb from the 3' end of the cassette. The resulting vector therefore contained the cassette with very long arms of homology and was now ampicillin resistant whilst no longer being chloramphenicol

resistant. Again, correct insertion was confirmed by restriction enzyme digestion (Figure 6.3).

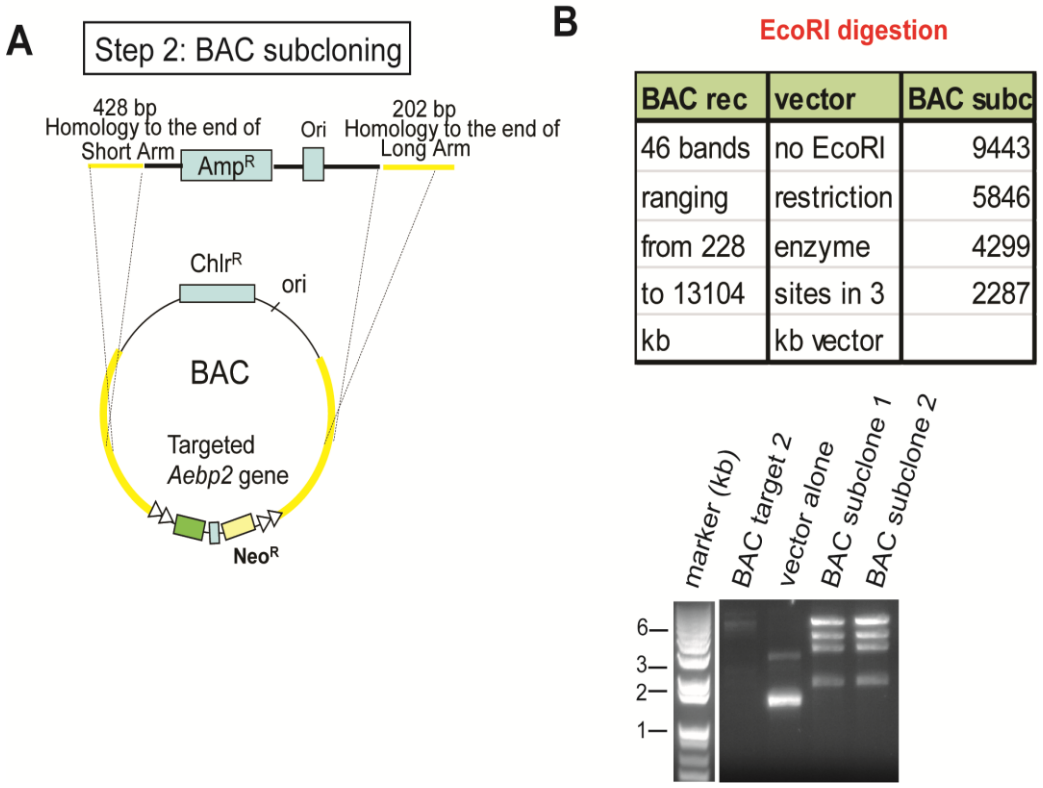


Figure 6.3: Subcloning the targeted *Aebp2*-containing BAC. (A) Graphical representation of the subcloning step. The BAC, containing the cassette inserted into the *Aebp2* gene, is subcloned into the pMA vector using homologous recombineering. (B) Digestion with EcoRI shows correct subcloning. The digestion of BAC target 2 has not completed. BAC subclones 1 and 2 show correct sizes. Sizes in table with predicted digestion pattern are in bp.

Because the *neo* gene had been used as a selectable marker in the recombineering strategy, it contained a bacterial promoter. The promoter sequence was removed by NotI digestion and re-ligation, generating a trapping cassette with an in-frame *GFP* and *neo* gene. Correct removal and re-ligation was confirmed by restriction enzyme digestion (Figure 6.4) and correct construction of the final 21.8 kb vector was verified by sequencing.

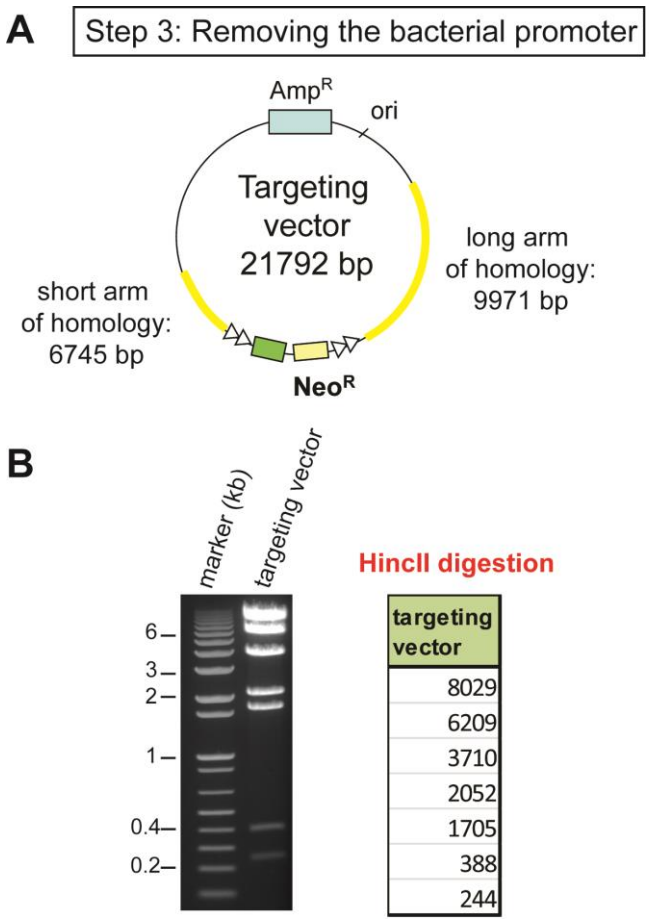


Figure 6.4: Removal of the bacterial promoter to generate final targeting vector. (A) Graphical representation of the final targeting vector after the bacterial promoter has been removed by NotI digestion. (B) HincII digestion confirms correct removal and re-ligation.

6.4 Targeting the first *Aebp2* allele

The first *Aebp2* allele was targeted through introduction of the SfiI-digested targeting vector into early passage 129/1 ES cells (Kay et al., 1993) by electroporation. Cells were plated at low density and subjected to geneticin selection for ~11 days until colonies were visible. Picked colonies were tested for correct integration by Southern Blot (Figure 6.5). It was found that approximately 46% of geneticin-resistant colonies contained a correctly targeted cassette. This high rate of efficiency is likely due to the long arms of homology and the use of a gene-trap method.

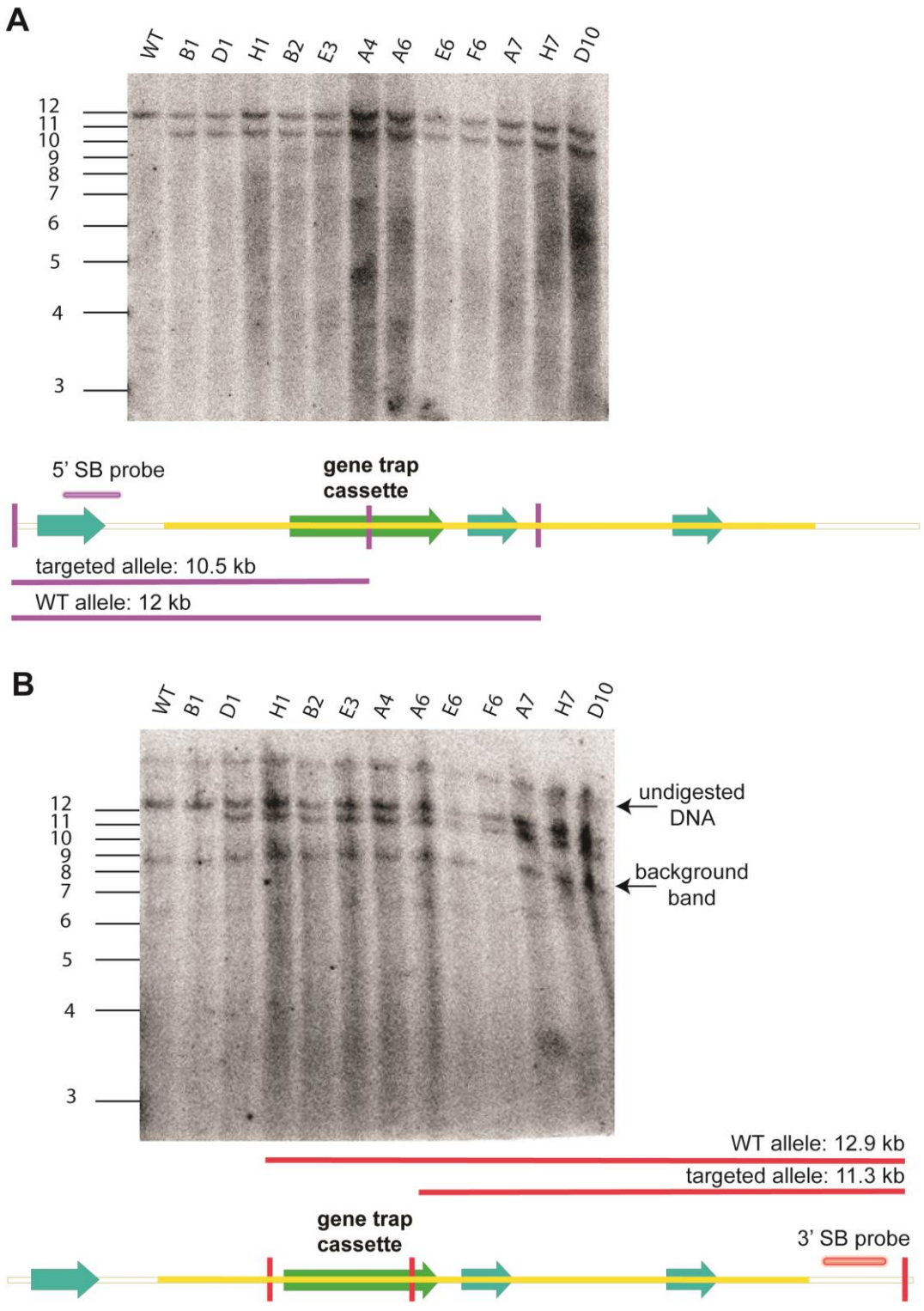


Figure 6.5: The first *Aebp2* allele is correctly targeted. (A) Digestion of genomic DNA by MscI and KpnI shows bands representing the wild type (WT) allele and the targeted allele. In the diagram, the cassette is indicated by the green arrow with the yellow line representing the arms of homology. The probe is shown to bind outside the arms of homology. Restriction enzyme sites are indicated by vertical lines. The length of digestion product of the WT allele and the targeted allele are shown. Not to scale. (B) Digestion of genomic DNA by BclI shows that all targeted clones, except B1, also have correct targeting at the 3'.

To verify that the cassette had integrated into the *Aebp2* locus only, and nowhere else, a further Southern Blot was performed using a probe that recognizes the *GFP* gene (Figure 6.6).

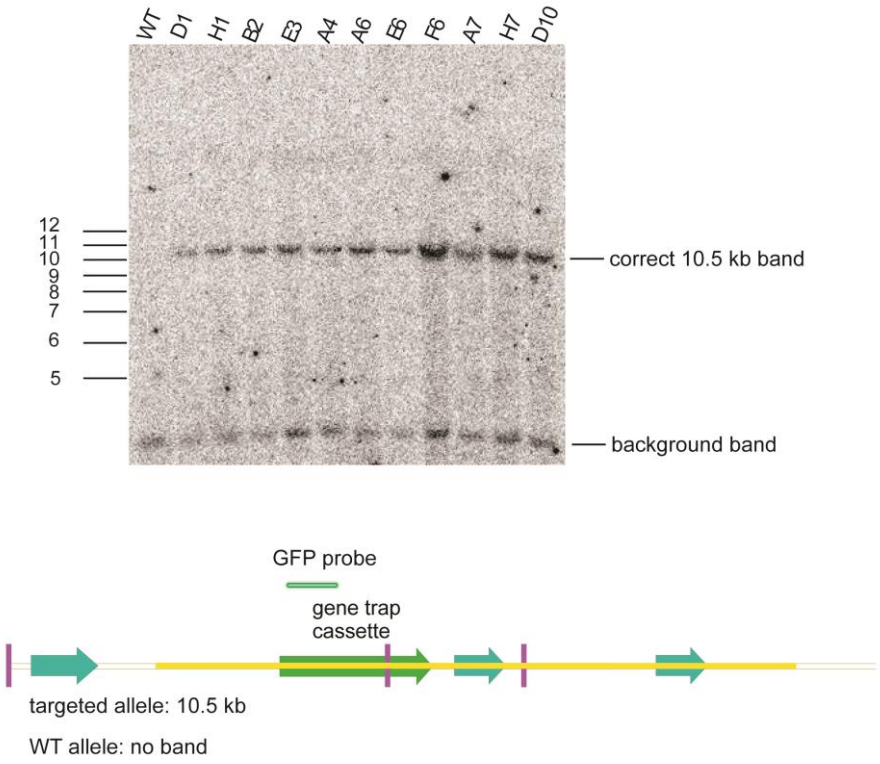


Figure 6.6: Digestion of genomic DNA by MscI and KpnI restriction enzymes shows single integration of the cassette into the *Aebp2* locus. Diagram shows restriction enzyme sites (vertical lines) and the GFP probe binding to the gene trap cassette DNA. DNA sizes indicated in kbp.

6.5 Inverting the cassette at the targeted *Aebp2* allele

Subsequently a plasmid containing the *FlpE* gene was introduced into the targeted ES cells by nucleofection. Cells were plated at a range of very low densities and grown for 8 days before picking. Picked colonies were tested for geneticin sensitivity, and those that were sensitive were tested for inversion by PCR (Figure 6.7). Subsequently, a Southern Blot confirmed correct inversion by FlpE (Figure 6.8).

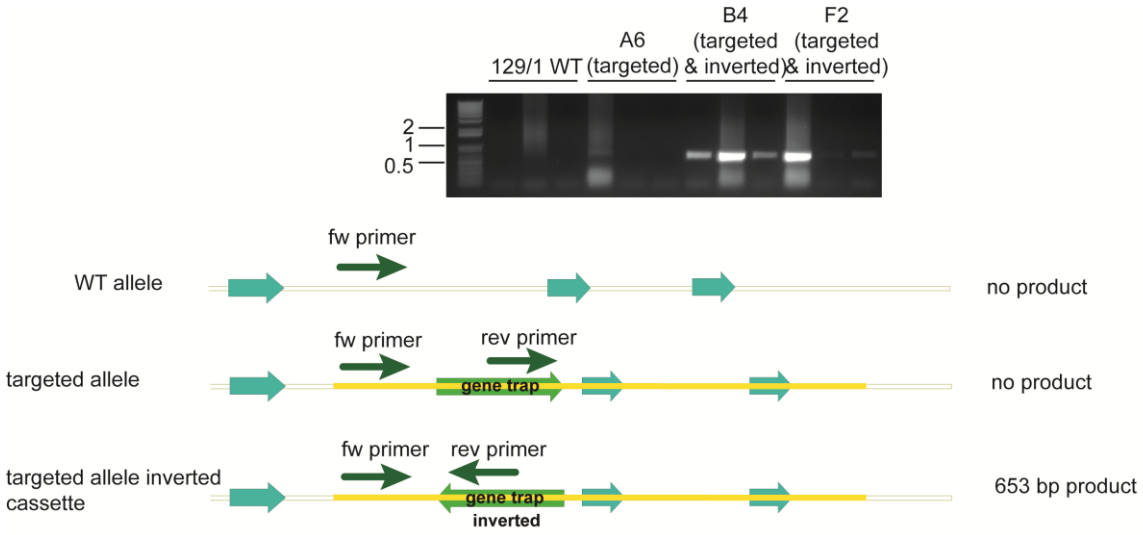


Figure 6.7: PCR for targeted ES cells and descendants with inverted cassette indicate that inversion has occurred. Three PCRs using different conditions were used for each DNA preparation (and therefore there are three lanes each). Sizes are indicated in kbp. The orientation of the primers (in dark green) is indicated (fw: forward; rev: reverse). A PCR product can only be produced when the template contains a cassette that has been targeted to the Aebp2 gene and then been inverted.

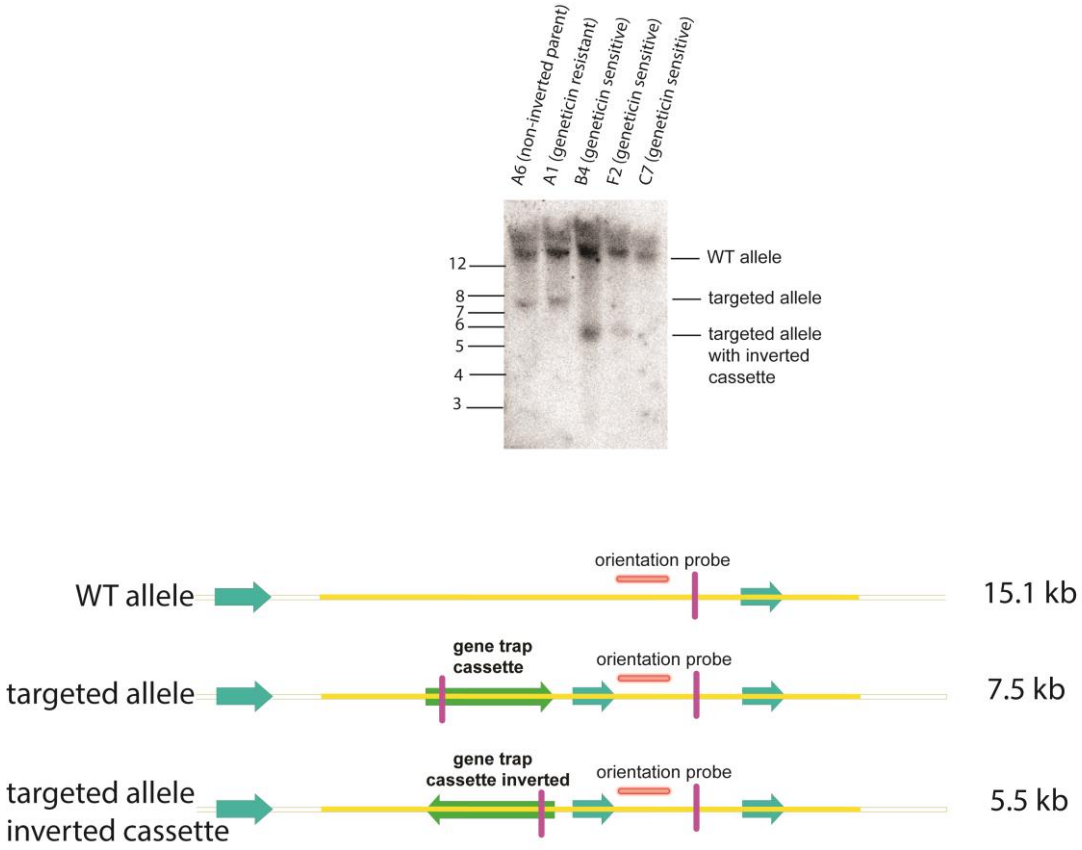
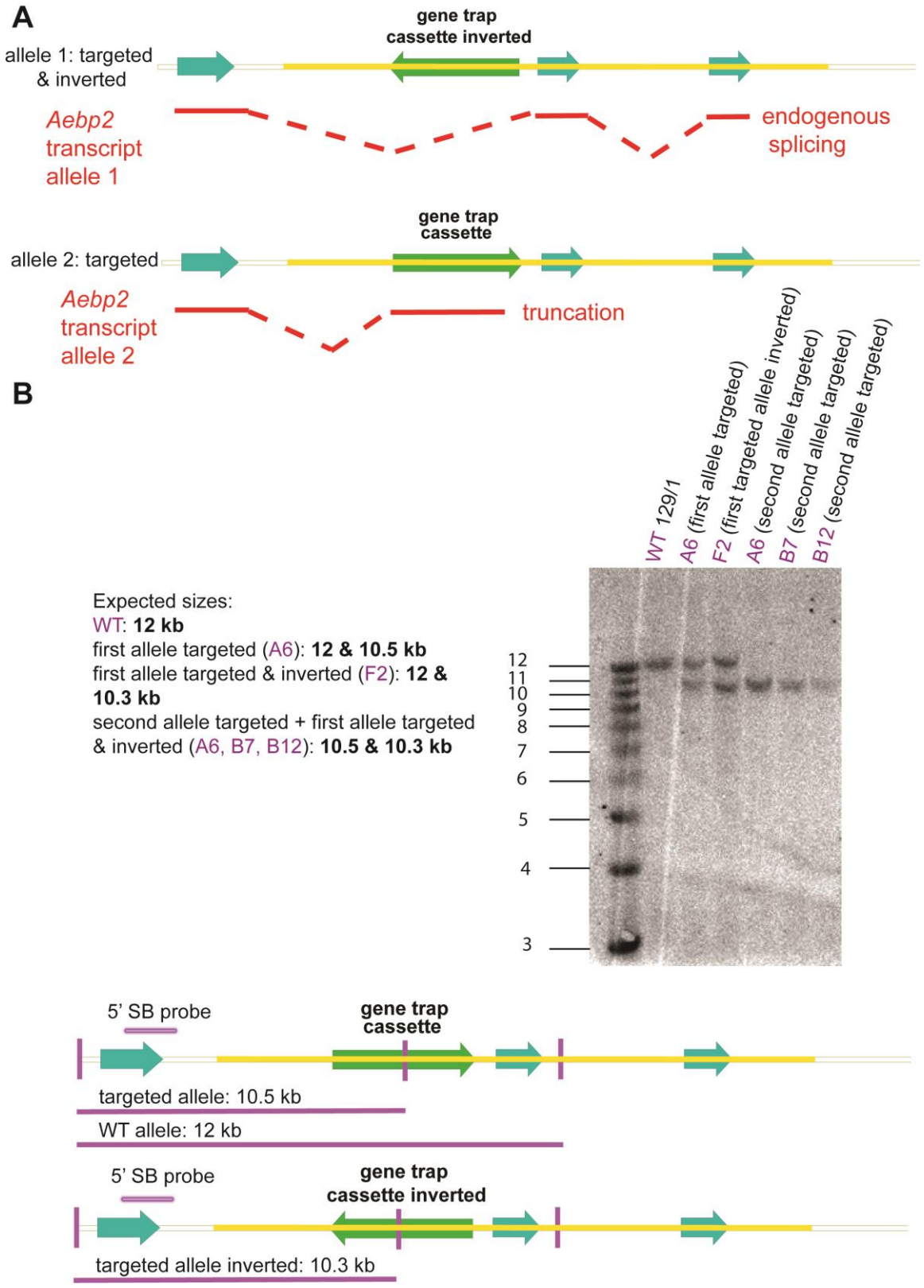


Figure 6.8: Digestion of genomic DNA by SpeI confirms correct inversion of geneticin-sensitive clones.

6.6 Targeting the second *Aebp2* allele

The second *Aebp2* allele was targeted using the same approach as for the first allele. Again, I verified correct targeting by Southern Blot (Figure 6.9).

For reasons that are unknown, I found a much lower efficiency of targeting (~3%) whilst targeting the second allele. Again, singular integration of the construct was tested by Southern Blot (Figure 6.10).


A
gene trap cassette inverted
allele 1: targeted & inverted
Aebp2 transcript allele 1
endogenous splicing
gene trap cassette
allele 2: targeted
Aebp2 transcript allele 2
truncation
B
WT 129/1
A6 (first allele targeted)
F2 (first targeted allele inverted)
A6 (second allele targeted)
B7 (second allele targeted)
B12 (second allele targeted)
Expected sizes:
WT: 12 kb
first allele targeted (A6): 12 & 10.5 kb
first allele targeted & inverted (F2): 12 & 10.3 kb
second allele targeted + first allele targeted & inverted (A6, B7, B12): 10.5 & 10.3 kb
12
11
10
9
8
7
6
5
4
3
5' SB probe
gene trap cassette
targeted allele: 10.5 kb
WT allele: 12 kb
5' SB probe
gene trap cassette inverted
targeted allele inverted: 10.3 kb

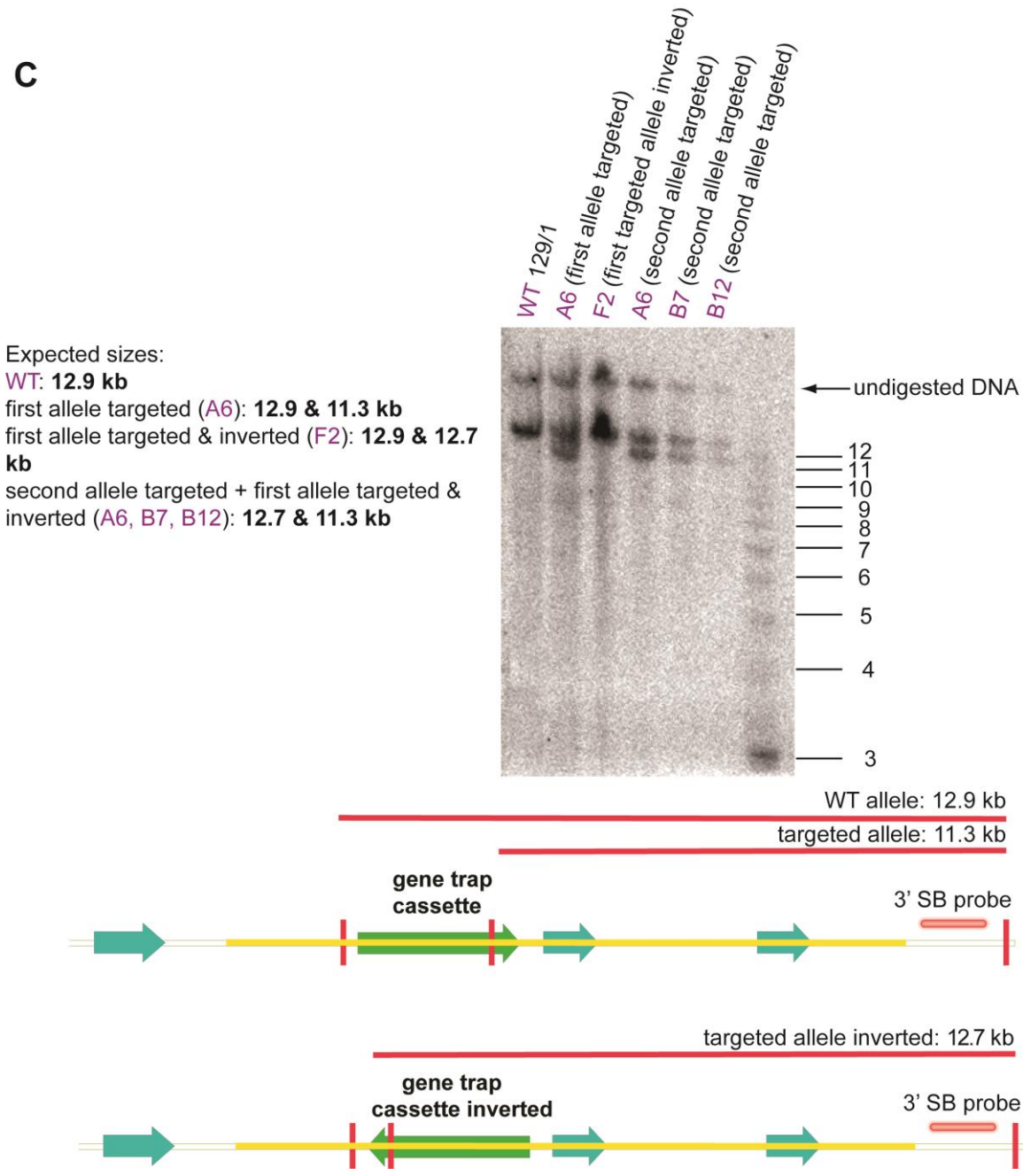


Figure 6.9: Correct targeting of the second *Aebp2* allele. (A) Graphical representation of the intended genetic configuration after this step. Digestion of genomic DNA by MscI and KpnI (B) and BclI (C) show correct targeting for the second allele. The possible configurations are indicated with the resultant expected sizes of digestion. Markers are indicated in kbp.

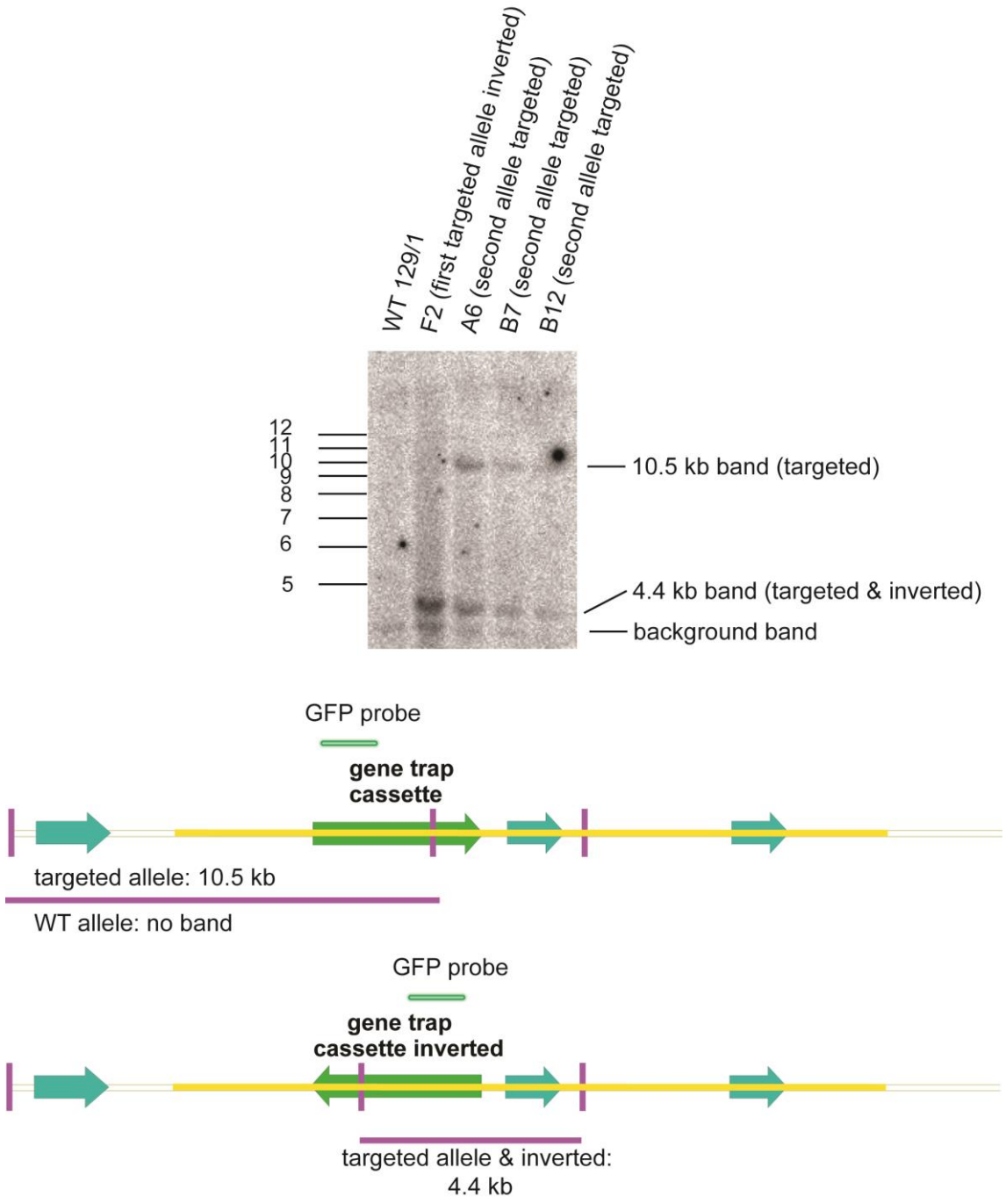
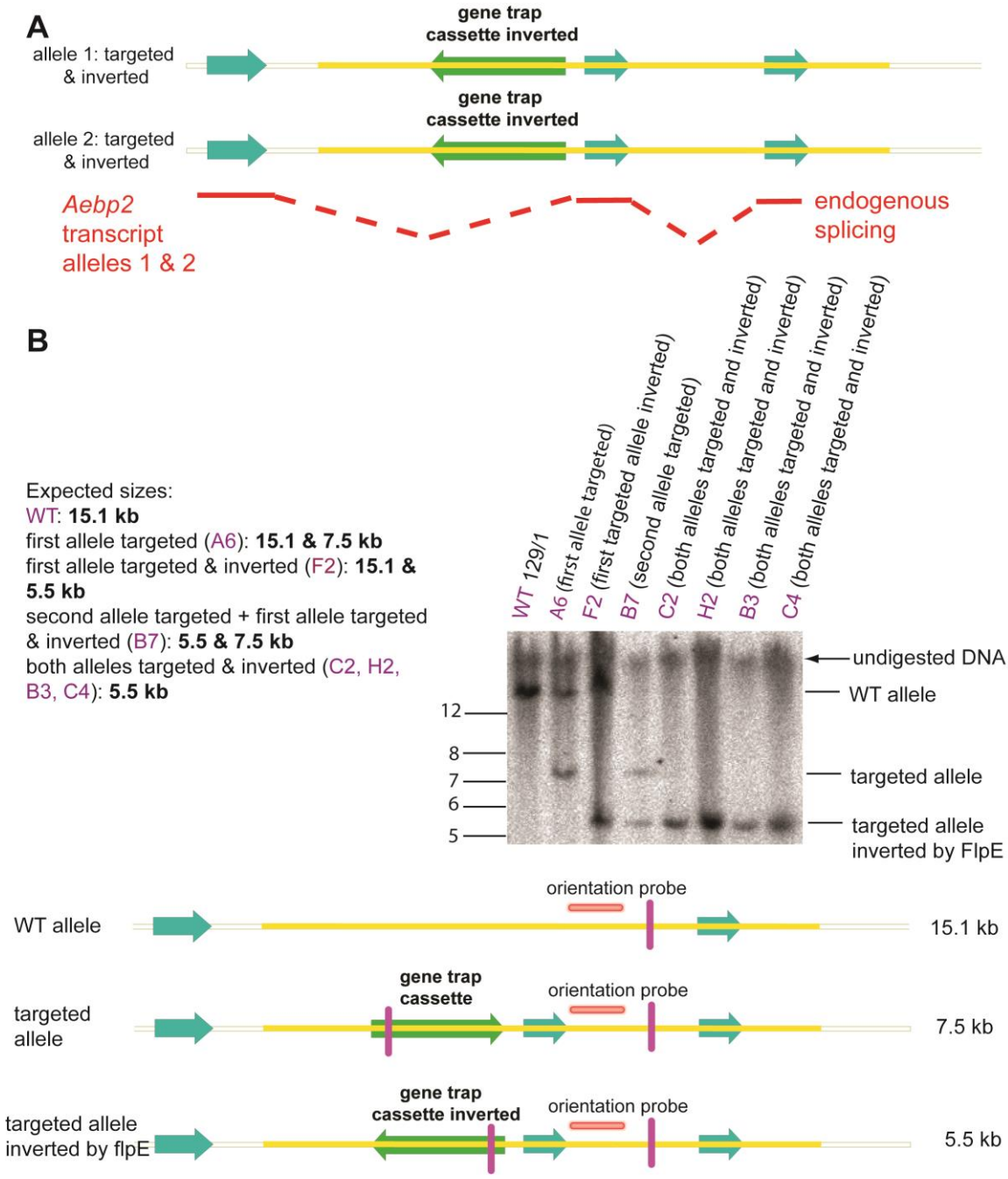


Figure 6.10: A single gene trap cassette has integrated in the correct location during the second targeting step as shown by digestion of genomic DNA by MscI and KpnI.

6.7 Inverting the cassette in the second *Aebp2* allele

The cassette inserted into the second allele was again inverted by the nucleofection of the *FlpE* containing plasmid. Correct inversion was confirmed by Southern Blot (Figure 6.11). The resulting ES cells had an inverted gene-trap inserted into both *Aebp2* loci, but these insertions should not interfere with transcription due to being on the strand opposite to the strand coding for the *Aebp2* transcript. These cells contained *Aebp2* mRNA at levels similar to wild type (WT) (see section 6.9). Therefore these cells are the control wild type cells used in subsequent experiments (see Chapter 7).


A
allele 1: targeted & inverted
gene trap cassette inverted
allele 2: targeted & inverted
gene trap cassette inverted
Aebp2 transcript alleles 1 & 2
endogenous splicing
B
Expected sizes:
WT: 15.1 kb
first allele targeted (A6): 15.1 & 7.5 kb
first allele targeted & inverted (F2): 15.1 & 5.5 kb
second allele targeted + first allele targeted & inverted (B7): 5.5 & 7.5 kb
both alleles targeted & inverted (C2, H2, B3, C4): 5.5 kb
WT 129/1
A6 (first allele targeted)
F2 (first targeted allele inverted)
B7 (second allele targeted)
C2 (both alleles targeted and inverted)
H2 (both alleles targeted and inverted)
B3 (both alleles targeted and inverted)
C4 (both alleles targeted and inverted)
12
8
7
6
5
undigested DNA
WT allele
targeted allele
targeted allele inverted by FlpE
orientation probe
WT allele
15.1 kb
gene trap cassette
orientation probe
targeted allele
7.5 kb
gene trap cassette inverted
orientation probe
targeted allele inverted by flpE
5.5 kb

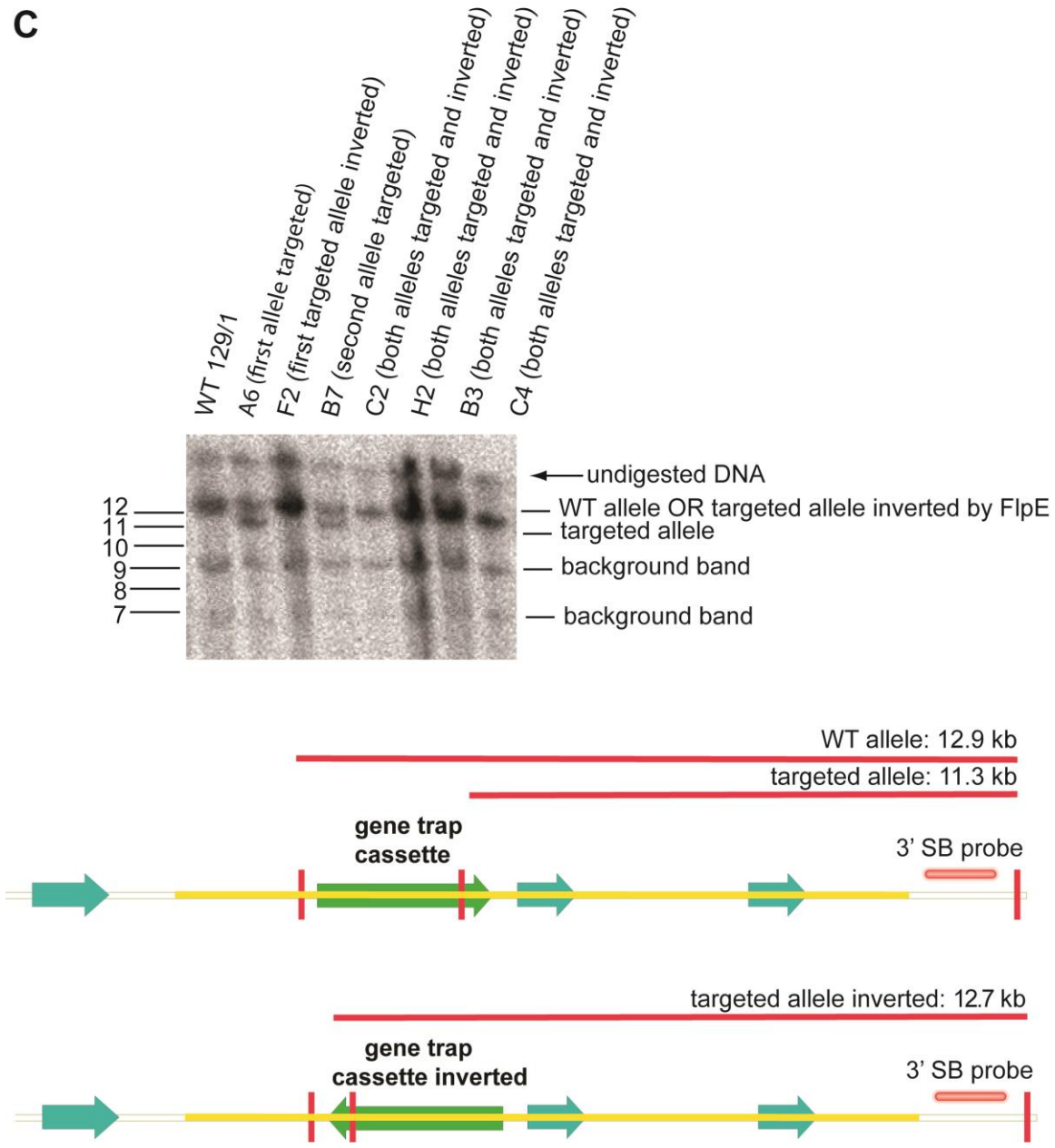
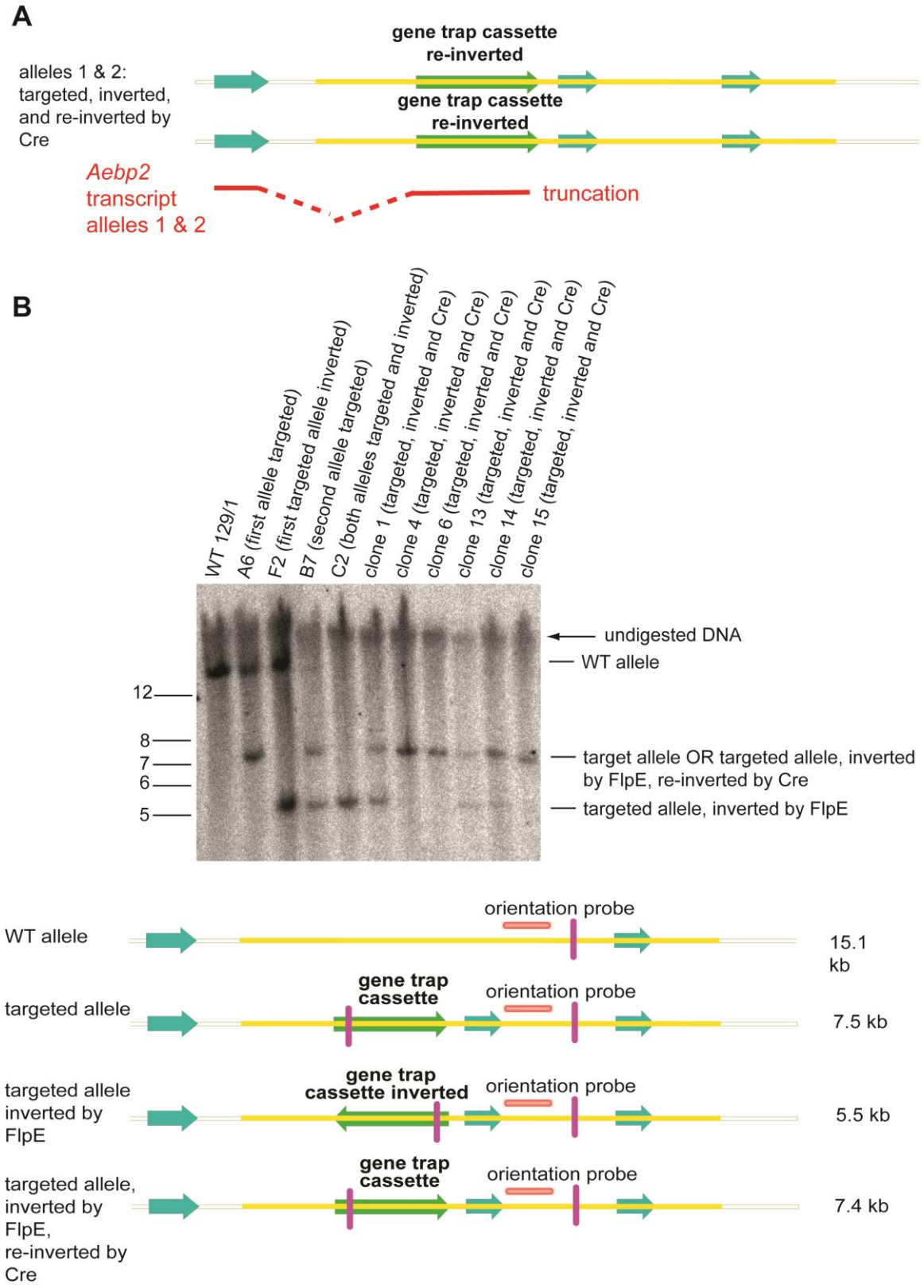


Figure 6.11: Successful inversion of the second targeted *Aebp2* allele. (A) Graphical representation of the intended genetic configuration after this step. When both cassettes are in the inverted orientation the *Aebp2* transcript is produced as normal, without the inclusion of the cassette. Digestion of genomic DNA by SpeI (B) and BclI (C) shows correct inversion of the cassette in the second allele.

6.8 Re-inverting the cassette to generate homozygous *Aebp2* gene-trapped ES cells

As a final step, both the cassettes were re-inverted by the introduction of a plasmid containing the *Cre* recombinase gene. Colonies were selected by geneticin resistance and verified by Southern Blot (Figure 6.12).

I found that all colonies tested had at least one allele inverted. Whilst some were heterozygotes (clones 1 and 13 in Figure 6.12B, C) as judged by the fact that they contained bands of equal intensity, others appeared to be a mixed population, in which some cells contained one inverted allele and other cells contained two inverted alleles (clones 14 and 15 in Figure 6.12B, C). A possible explanation is that at the time of plating at low density, a single cell with just one allele inverted still contained the plasmid with the Cre recombinase. Subsequently the Cre plasmid was inherited by some daughter cells, and that Cre acted on the second allele in those cells, but was not inherited by others. Encouragingly, I found two colonies which appeared to have both alleles inverted (clones 4 and 6 in Figure 6.12B, C).


A
gene trap cassette re-inverted
alleles 1 & 2: targeted, inverted, and re-inverted by Cre
gene trap cassette re-inverted
Aebp2 transcript alleles 1 & 2
truncation
B
WT 129/1
A6 (first allele targeted)
F2 (first targeted allele inverted)
B7 (second allele inverted)
C2 (both alleles targeted)
clone 1 (targeted, inverted and inverted)
clone 4 (targeted, inverted and Cre)
clone 6 (targeted, inverted and Cre)
clone 13 (targeted, inverted and Cre)
clone 14 (targeted, inverted and Cre)
clone 15 (targeted, inverted and Cre)
12
8
7
6
5
undigested DNA
WT allele
target allele OR targeted allele, inverted by FlpE, re-inverted by Cre
targeted allele, inverted by FlpE
WT allele
orientation probe
15.1 kb
targeted allele
gene trap cassette
orientation probe
7.5 kb
targeted allele inverted by FlpE
gene trap cassette inverted
orientation probe
5.5 kb
targeted allele, inverted by FlpE, re-inverted by Cre
gene trap cassette
orientation probe
7.4 kb

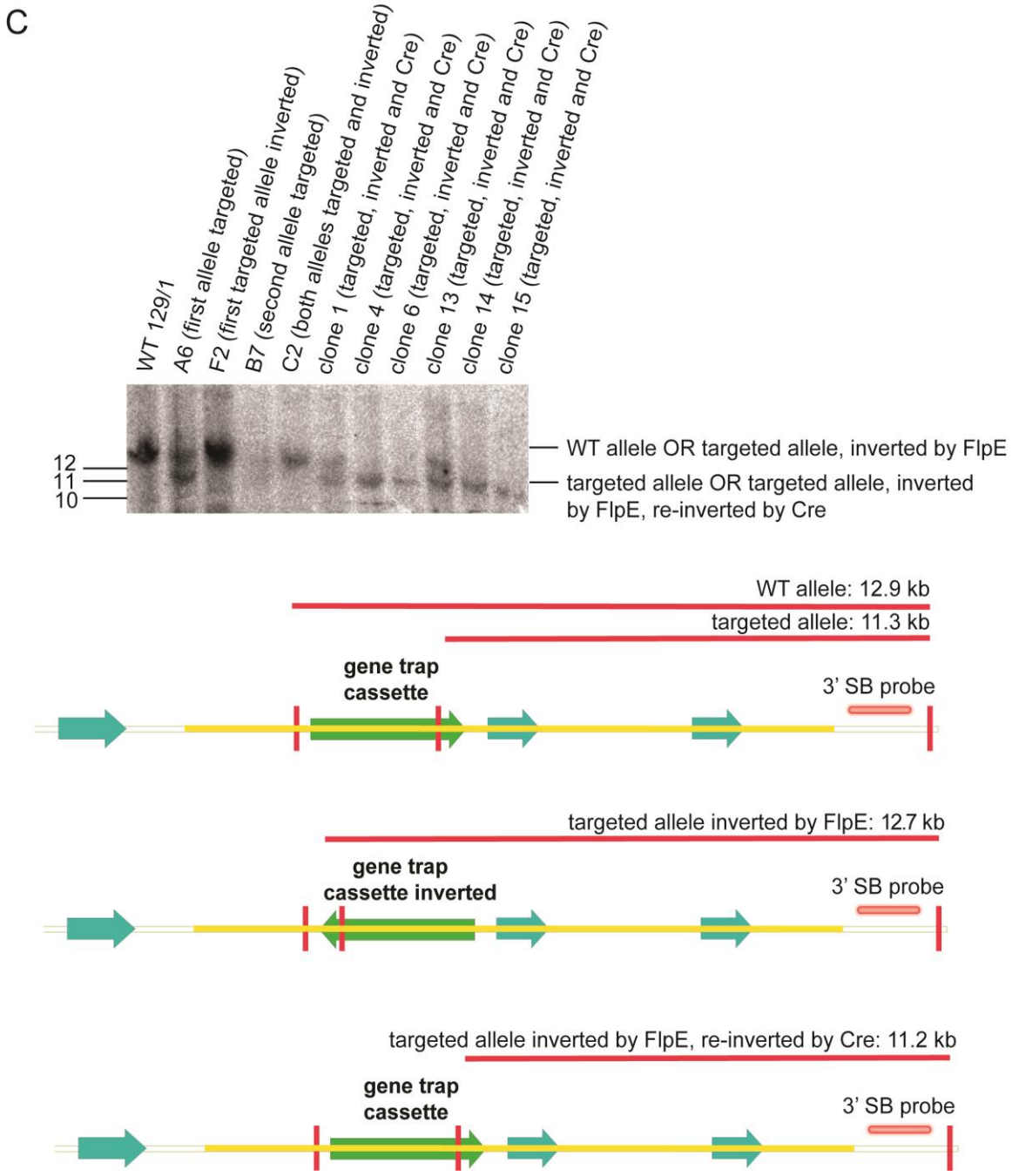
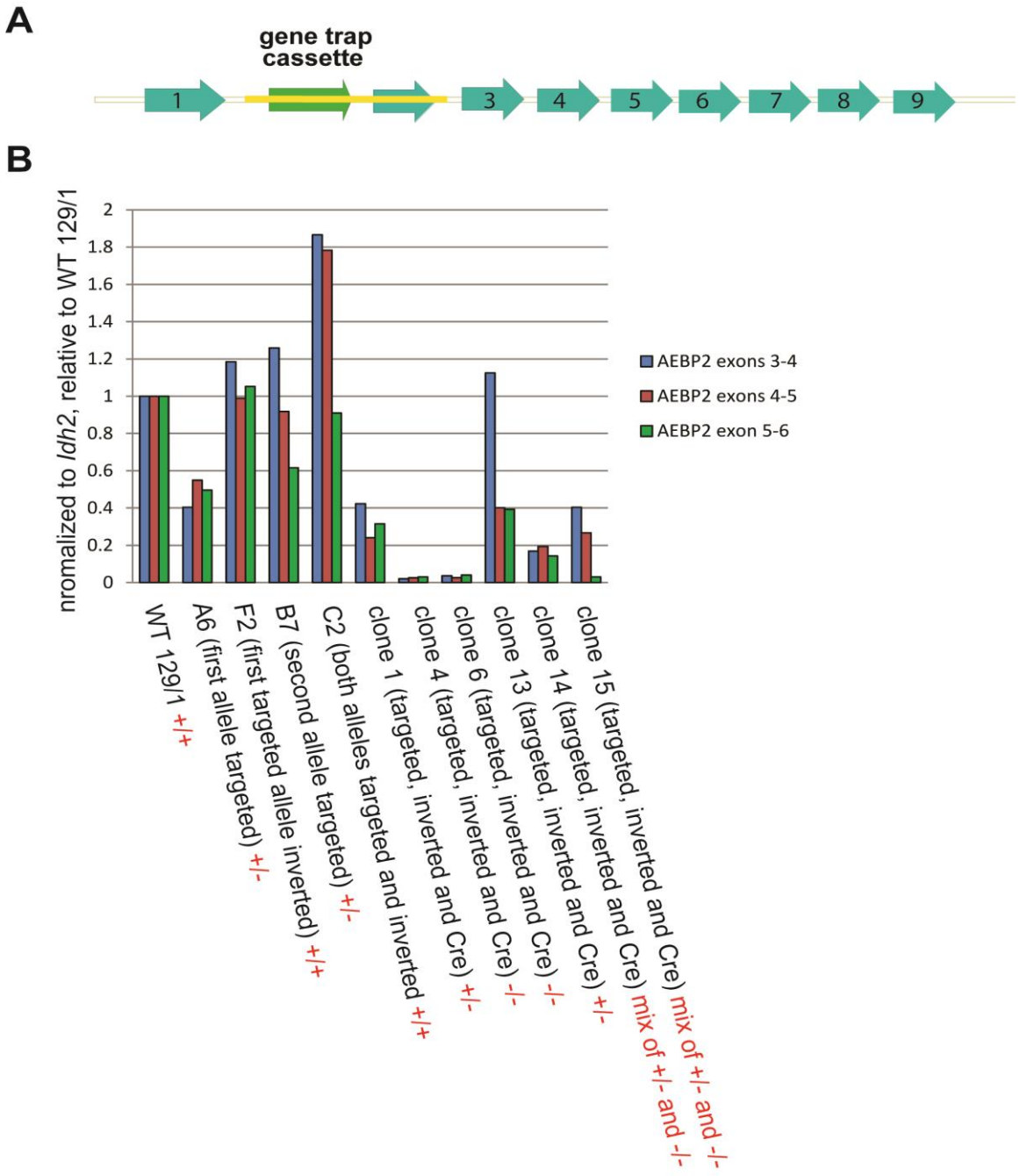


Figure 6.12: Re-inversion of the cassettes to generate homozygous *Aebp2* gene-trapped ES cells. (A) Graphical representation of the intended genetic configuration after this step. Both alleles should be in the active gene trap orientation so that no full length *Aebp2* transcript is produced. Digestion of genomic DNA by SpeI (B) and BclI (C) shows correct re-inversion of alleles. Sizes are indicated in kb.

6.9 Validation of homozygous *Aebp2* gene-trapped ES cells

I next performed qPCR using several primer pairs to assess *Aebp2* mRNA levels in the homozygous *Aebp2* gene-trapped ES cells, as well as the heterozygote and mixed population colonies. It was found that the amount of transcript generally corresponded well with the genetic status of the *Aebp2* loci as assessed by Southern Blot (Figure 6.13). Clones 4 and 6 had the lowest level of transcript, at about 4-5% levels compared to the 129/1 parent, whereas clones 1 and 13, which are heterozygotes, also had transcript levels that compared to those cell lines in which only one allele was gene-trapped. Clones 14 and 15 displayed intermediate levels of *Aebp2* mRNA transcript, as expected. The qPCR experiments were repeated for a number of cell lines, using primers for exons 3-4, 4-5 and 5-6 (Figure 6.13C). I also used a primer set to amplify a product across exon 1a to the cassette (Figure 6.13D). As predicted, this primer set only generated a product in those cell lines which had at least one allele trapped.


A
gene trap cassette
1
3
4
5
6
7
8
9
B
nromalized to Idh2, relative to WT 129/1
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
AEBP2 exons 3-4
AEBP2 exons 4-5
AEBP2 exon 5-6
WT 129/1 +/+
A6 (first allele targeted) +/-
F2 (first targeted allele inverted) +/+
B7 (second allele targeted) +/-
C2 (both alleles targeted and inverted +/+
clone 1 (targeted, inverted and Cre) +/-
clone 4 (targeted, inverted and Cre) -/-
clone 6 (targeted, inverted and Cre) -/-
clone 13 (targeted, inverted and Cre) +/-
clone 14 (targeted, inverted and Cre) mix of +/- and -/-
clone 15 (targeted, inverted and Cre) mix of +/- and -/-

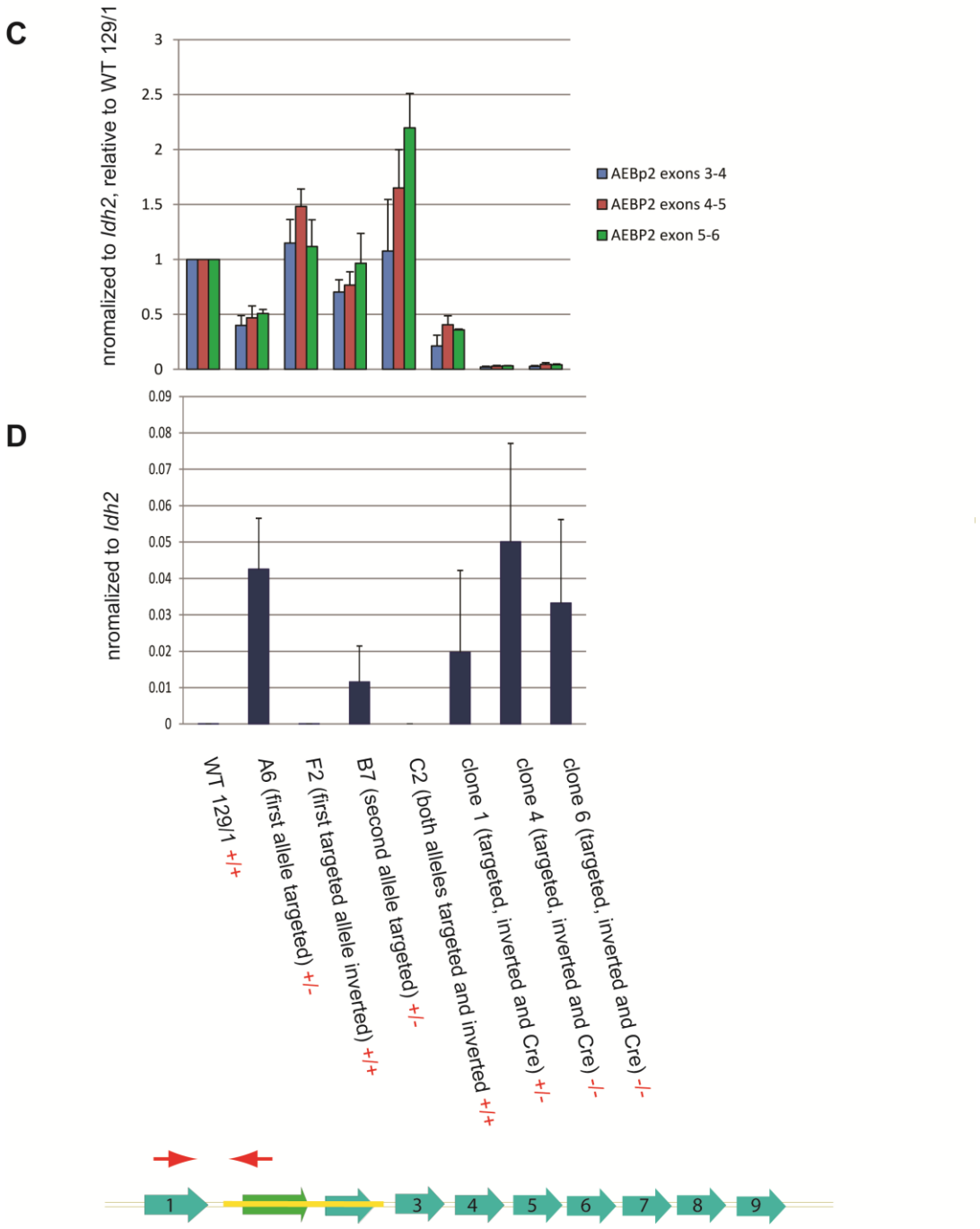


Figure 6.13: qPCR shows that the *Aebp2* transcript is successfully trapped. (A) Genetic configuration in *Aebp2* gene-trapped allele showing all exons. The cassette should trap transcripts that initiate upstream of the cassette so that there are no transcripts from downstream exons. (B) Single qPCR experiment testing the levels of the *Aebp2* transcript in several cell lines using 3 primer pairs. (C) qPCR experiment testing the levels of the *Aebp2* transcript in several cell lines using 3 primer pairs. Error bars are standard deviation of at least 3 biological repeat experiments. (D) qPCR experiment testing the levels of the *Aebp2-GFP* fusion transcript trapped by the trapping cassette in several cell lines. A graphical representation of where the primers (in red) bind to exon 1a and the *GFP* gene is presented. Error bars are standard deviation of 3 biological repeat experiments.

In order to be absolutely confident that the homozygous *Aebp2* gene-trapped cell line was clonal and did not contain any cells that had one *Aebp2* allele in which the cassette had not been inverted, I subcloned line 6 by plating at low density and picking colonies. Descendants were verified to be homozygous *Aebp2* gene-trapped by Southern Blot and qPCR.

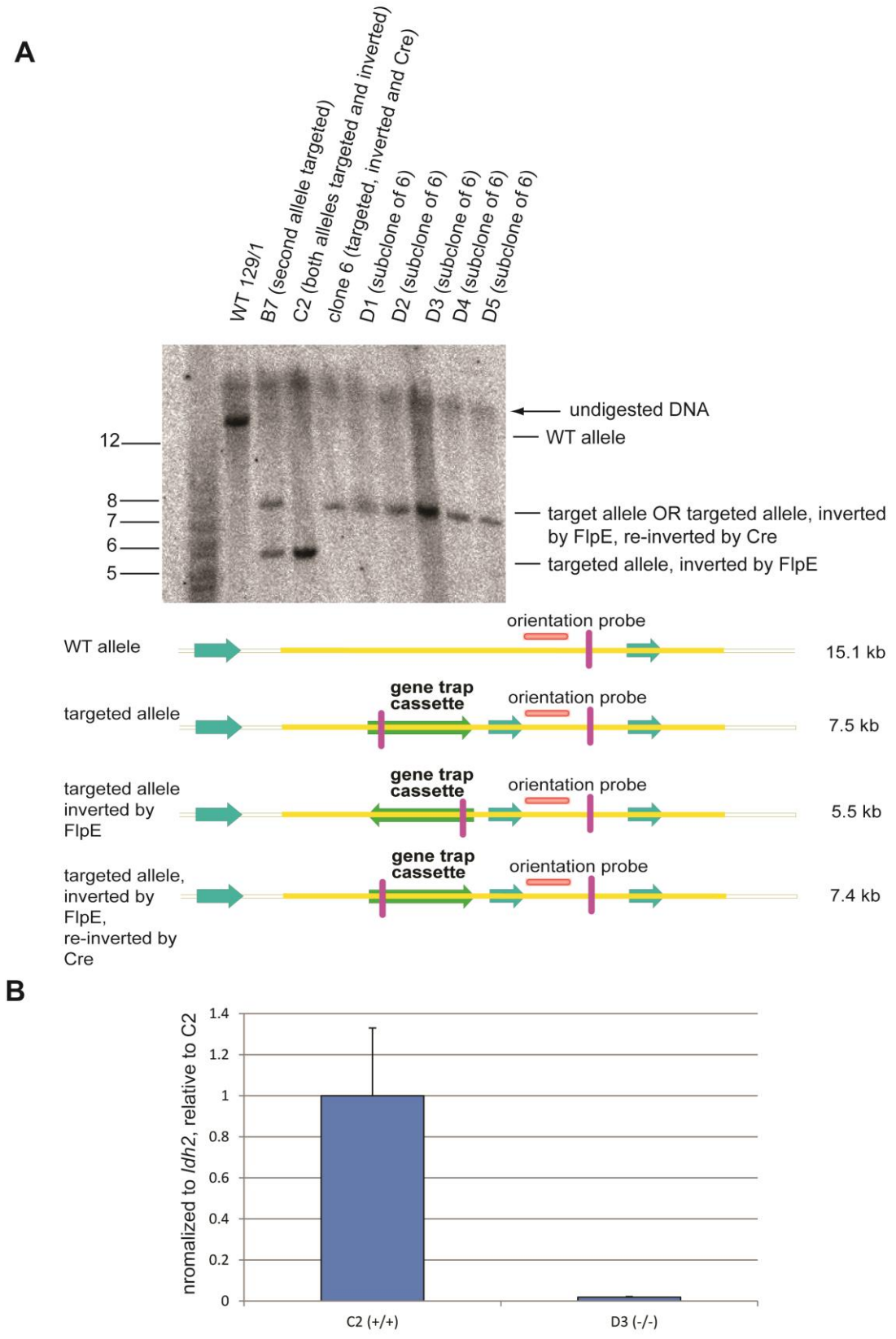


Figure 6.14: Subcloning to generate a definite homozygous *Aebp2* gene-trapped cell line. (A) Restriction enzyme digestion of genomic DNA by SpeI confirms that the descendants of clone 6 contain the re-inverted gene-trap in both *Aebp2* alleles. (B) *Aebp2* transcript levels are 2% of the control cell line in D3, a homozygous *Aebp2* gene-trapped cell line.

Subsequently, I analysed AEBP2 protein levels in the D3 homozygous *Aebp2* gene-trapped cell line. In an initial experiment, I found that protein levels were equal in the WT and gene-trapped cells (data not shown). This was also the case when I used the antibody that had been purified by an antigen based column (see 4.2). I furthermore tried several commercially available antibodies, but these antibodies did not recognize overexpressed AEBP2 (see 4.2 and 4.4) and were therefore not useful (data not shown). I reasoned that the most likely explanation for this observation was that the antibody in fact cross-reacts with another band that lies over the genuine AEBP2-representing band when separated on a standard polyacrylamide gel. This could be consistent with the smaller-sized band seen in the gel filtration in Figure 5.1 representing AEBP2 and co-eluting tightly with the other PRC2 complex members, whilst the band of larger apparent molecular weight representing a protein that is also recognized by the antibody. Therefore, I analysed the fractions of the gel filtration of nuclear cell extracts made from the C2 WT cells (indicated as $Aebp2^{+/+}$) and the D3 gene-trapped line (indicated as $Aebp2^{-/-}$) (Figure 6.15). As expected, the smaller band was absent from the D3 gene-trapped cell line, suggesting that this band represents AEBP2 and is lost in the homozygous *Aebp2* gene-trapped cell line.

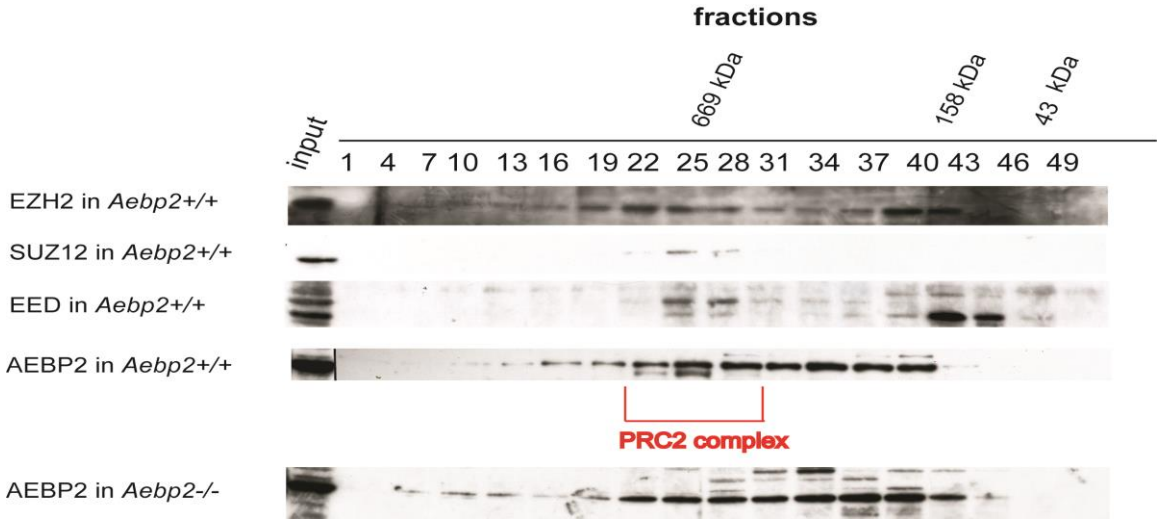
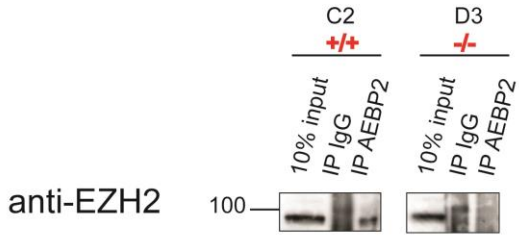


Figure 6.15: Fractions from gel filtration of *Aebp2*$^{+/+}$ show a band not present in the *Aebp2*$^{-/-}$ fractions, likely representing AEBP2. The fractions at which the canonical PRC2 complex elutes are indicated.

Additionally, I performed an immunoprecipitation experiment using the AEBP2 antibody in C2 *Aebp2*$^{+/+}$ and D3 *Aebp2*$^{-/-}$ cells. Whilst I found that this still immunoprecipitated a band which I had previously assumed to be AEBP2, EZH2 was no longer co-immunoprecipitated, consistent with my assumption that AEBP2 protein levels are severely depleted/absent in the D3 *Aebp2*$^{-/-}$ cells (Figure 6.16A).

A



B

Protein	molecular weight	Mascot score in IgG	Mascot score in C2	Mascot score in D3
SUZ12	83715	ND	134	26
AEBP2	52498	17	91	238
EED	56952	ND	62	ND
EZH2	82884	ND	46	ND
Aminoglycoside 3'-phosphotransferase	29315	ND	ND	105
GFP	26953	ND	ND	52

C

Peptides matching AEBP2 (total 6) in C2

```
  1 MAAALADMAD LEELSRLSPL SPGSPGPAAR GRAEPPEEEE EEDDEEAEAE
 51 AVAALLLNGG AGGGAGGGEA ETMSEPSPES ASQAGGDEDE DEEDDEDEGS
101 SSGGAEEESS AESLVGSSSG GCSGDETRSL SPGAASSSSG DGDGKEGLEE
151 PKGPRGGPGG PGSSGGGSSS SSVVSSGGDE GYGTGGGGSS ATSGGRRGSL
201 EMSSDGEPLS RMDSED|SISS TLMDIDSTIS SGRSTPAMMN GQGSTTASSK
251 HIAYNCCWDQ CQACFNSSPD LADHIRSIHV DGQRGGVFVC LWKGCKVYNT
301 PSTSQSWLQR HMLTHSGDKP FKCVVGGCNA SFASQGGLAR HVPTHFSQQN
351 SSKVSSQPKA KEESPSKAGM NKRRKLKNKR RRSLPRPHDF FDAQTLDAIR
401 HRAICFNLSA HIESLGKGHS VVFHSTVIAK RKEESGKIKL LLHWMPEDIL
451 PDVWVNESER HQLKTKVVHL SKLPKDTALL LDPNIYRTMP QKRLKR
```

Peptides matching AEBP2 (total 10) in D3

```
  1 MAAALADMAD LEELSRLSPL SPGSPGPAAR GRAEPPEEEE EEDDEEAEAE
 51 AVAALLLNGG AGGGAGGGEA ETMSEPSPES ASQAGGDEDE DEEDDEDEGS
101 SSGGAEEESS AESLVGSSSG GCSGDETRSL SPGAASSSSG DGDGKEGLEE
151 PKGPRGGPGG PGSSGGGSSS SSVVSSGGDE GYGTGGGGSS ATSGGRRGSL
201 EMSSDGEPLS RMDSED|SISS TLMDIDSTIS SGRSTPAMMN GQGSTTASSK
251 HIAYNCCWDQ CQACFNSSPD LADHIRSIHV DGQRGGVFVC LWKGCKVYNT
301 PSTSQSWLQR HMLTHSGDKP FKCVVGGCNA SFASQGGLAR HVPTHFSQQN
351 SSKVSSQPKA KEESPSKAGM NKRRKLKNKR RRSLPRPHDF FDAQTLDAIR
401 HRAICFNLSA HIESLGKGHS VVFHSTVIAK RKEESGKIKL LLHWMPEDIL
451 PDVWVNESER HQLKTKVVHL SKLPKDTALL LDPNIYRTMP QKRLKR
```

Figure 6.16: Co-immunoprecipitation of EZH2 in C2 *Aebp2*$^{+/+}$ but not in D3 *Aebp2*$^{-/-}$ cells by the AEBP2 antibody. (A) Immunoblot showing that EZH2 is no longer co-immunoprecipitated from nuclear cell extract from D3 *Aebp2*$^{-/-}$ cells by the AEBP2 antibody. (B) Mascot scores for proteins identified in immunoprecipitations. (C) Location of peptides for AEBP2. The junction between exon 1b and exon 2 is indicated by a vertical line.

The immunoprecipitations were analysed by mass spectrometry to confirm the findings (Figure 6.16B). This showed that the AEBP2 antibody does indeed immunoprecipitate AEBP2 and the PRC2 complex, as SUZ12, EED and EZH2 are all identified in the *Aebp2*$^{+/+}$ cells. In the *Aebp2*$^{-/-}$ cells, peptides matching to AEBP2 are still immunoprecipitated. This is unsurprising as the first *Aebp2* exon 1b, comprising of amino acids 1-216, is still present and fused to the GFP and aminoglycoside 3' phosphotransferase. The existence of an AEBP2-GFP-amigoglycoside 3' phosphotransferase fusion protein is corroborated by the finding that GFP and aminoglycoside 3' phosphotransferase are co-immunoprecipitated in D3 *Aebp2*$^{-/-}$ cells, but not in C2 *Aebp2*$^{+/+}$ cells. Nine out of the ten AEBP2 peptides match to the 1b exon, whilst one matches to the C-terminus of the protein, which was considered background as one peptide was also detected in the IgG immunoprecipitation. Two peptides matching SUZ12 were also identified in the *Aebp2*$^{-/-}$ cells. This could be due to SUZ12 interacting with the part of AEBP2 that is still present.

In an attempt to discover which protein might be the one that the AEBP2 antibody is cross-reacting with, I also analysed the other immunoprecipitated proteins that were found in immunoprecipitation using the AEBP2 antibody from both C2 *Aebp2*$^{+/+}$ cells and D3 *Aebp*$^{-/-}$ cells, but not in an immunoprecipitation using the IgG antibody from C2 *Aebp2*$^{+/+}$ cells. An overview of these proteins is presented in Figure 6.17A.

A

Protein	cellular localisation	molecular weight	Mascot score in C2	Mascot score in D3
Ankycorbin	Cytoplasm/cytoskeleton/nucleus	109470	193	33
Striatin-4	Cytoplasm/membrane	81609	93	219
Rpn1	Membrane of the rough ER	68657	57	91
CTTNBP2 N-terminal-like protein	Cytoplasm/cytoskeleton	70255	56	57
Mprip	Cytoplasm/cytoskeleton	118748	48	82
Pou2f1	Nucleus	72471	49	122
RBBP5	Nucleus	41024	42	24
Mlh1	Nucleus	85301	41	44
MTA2	Nucleus	75724	40	82

B

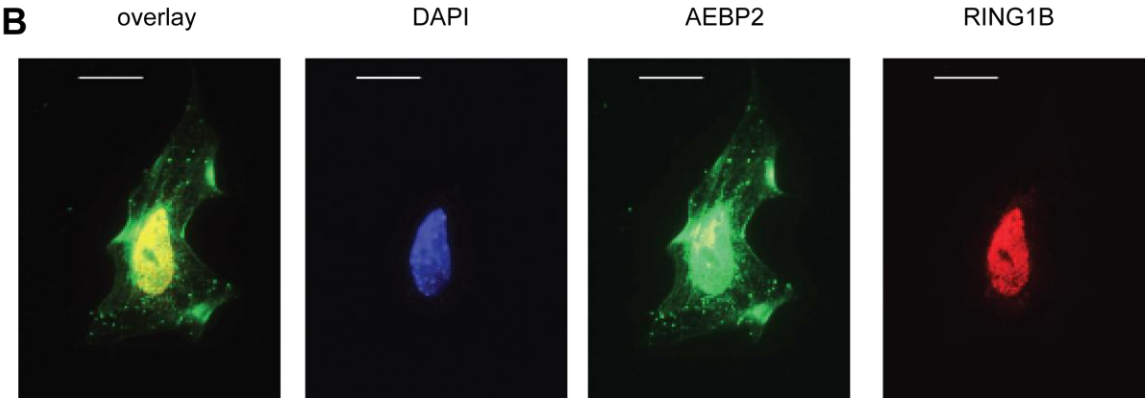


Figure 6.17: Potential candidates for the protein cross-reacting with the AEBP2 antibody. (A) Proteins which were identified in both C2 *Aebp2*$^{+/+}$ and D3 *Aebp2*$^{-/-}$ cells, but not in the IgG immunoprecipitation, with a Mascot score equal to or greater than 40 in the C2 immunoprecipitation. (B) Immunofluorescence signal using the AEBP2 antibody is, in addition to being nuclear, also cytoskeleton-like. Scale bar is 20 µm.

This list includes a number of cytoskeleton-associating proteins as annotated by the UniProt database, which might be consistent with a cytoskeleton-like staining I observed in differentiated cells using the AEBP2 antibody (Figure 6.17B). The predicted molecular weight of some proteins would also match the observed molecular weight as judged by migration in the gel. However, membrane and cytoskeleton-like proteins are often highly expressed and may be contamination in the immunoprecipitation. I can therefore not definitively conclude what the contaminating protein is.

6.10 Progress towards making an *Aebp2* gene-trapped mouse

Cells with a targeted and inverted cassette in a single allele were analysed for ploidy by metaphase spread. It was discovered that the F2 line, and its parent A6, had 41 chromosomes, most likely due to gain of chromosome 8, a very common form of trisomy in mouse ES cells (Kim et al., 2013). Therefore, karyotypes of targeted ES cell lines were screened and it was found that line A7 had 40 chromosomes (Figure 6.18A). The cassette in these cells was inverted by nucleofection of the *FlpE* expressing plasmid, and correct inversion of the B6 descendant was verified by Southern Blot (data not shown). Presence of the correct number of chromosomes was then verified (Figure 6.18B) and both this line and its parent A7 were used for injection into C57 black 6 (C57BL/6) blastocysts. Offspring with high levels of chimerism were identified (Figure 6.18C) and mated with C57BL/6 mice. Heterozygote *Aebp2* gene-trapped mice were identified by PCR (Figure 6.18D). These heterozygotes are currently being mated with 129/1 mice to generate more heterozygotes. Subsequently heterozygote crosses will be set up to generate homozygous *Aebp2* gene-trapped mice (from A7) and homozygous *Aebp2* conditional gene-trapped mice (from B6), where the gene-trap can then be activated by crossing these mice to Cre-expressing mouse lines.

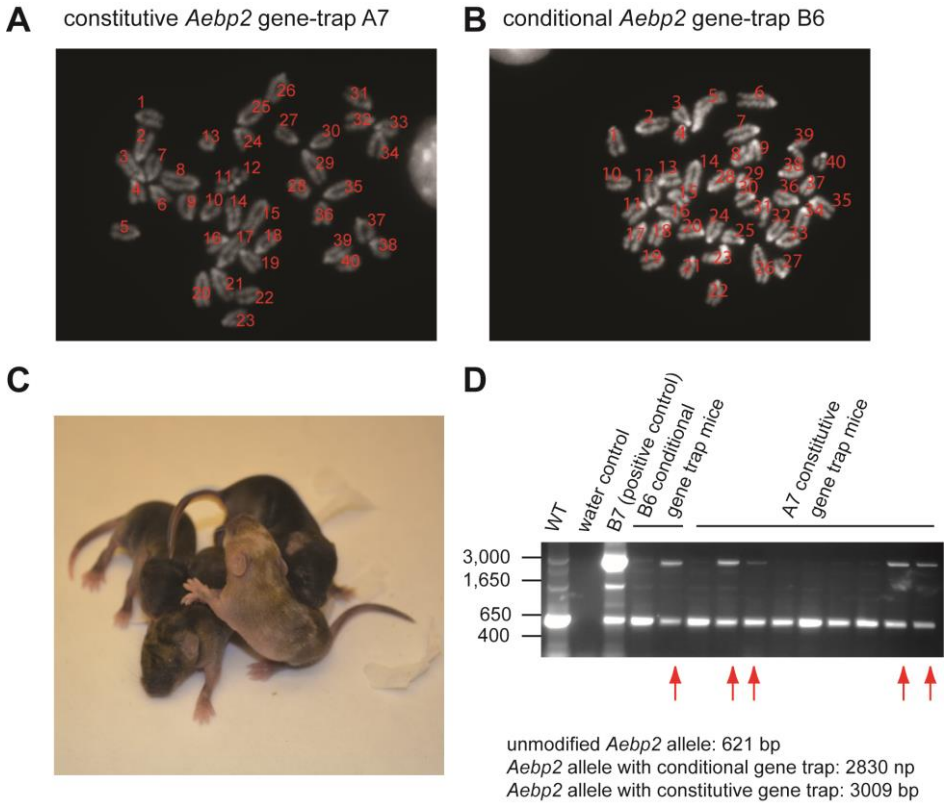


Figure 6.18: Progress towards making *Aebp2* gene-trapped mice. Metaphase spreads show the correct number of chromosomes for the constitutive *Aebp2* gene-trapped line A7 (A) and the conditional *Aebp2* gene-trapped line B6 (B). (C) Mice with a high level of chimerism were obtained after blastocyst injection. (D) After mating chimeric mice with WT mice, heterozygotes were obtained as observed by coat colour and confirmed by PCR. Heterozygote PCR products are indicated by a red arrow.

6.11 Discussion

I have successfully inserted a gene-trap immediately upstream of exon 2 of the *Aebp2* gene, in both alleles, in mouse ES cells. Southern Blotting has confirmed the correct genetic configuration of the gene-trap in control and trapped cells and *Aebp2* transcript levels in the gene-trapped cells are severely depleted. It was therefore a surprise to find that protein levels were unaffected. Upon further analysis, it was shown that a band migrating slightly faster than the band previously presumed to be AEBP2 is absent in the homozygous *Aebp2* gene-trapped ES cells. Together with the analysis of immunoprecipitations by the AEBP2 antibody in WT and homozygous *Aebp2* gene-

trapped ES cells, this shows that the full length AEBP2 protein is absent or severely depleted in the gene-trapped ES cells. That insertion into this site would cause AEBP2 depletion is consistent with a report that a constitutive gene-trap inserted in a similar location leads to embryonic lethality, presumably due to absence of the AEBP2 protein (Kim et al., 2011).

The immunoprecipitation and immunofluorescence experiments demonstrated that the AEBP2 antibody certainly recognizes the AEBP2 protein, but may also recognize a cytoskeletal associated protein. Considering that the AEBP2 antibody may recognize another protein requires a re-interpretation of previously obtained results. For example, it makes it more difficult to estimate how much FS2-AEBP2 is overexpressed relative to endogenous AEBP2 (Figure 4.3). It also means results from ChIP and from immunofluorescence need to be interpreted with some caution. In the case of ChIP, the signal was strongly reduced in the absence of EED (Figure 5.7) making it likely that ChIP signal is from AEBP2, as a cytoskeletal protein would not be expected to be dependent on EED, or indeed bind DNA at polycomb sites. Furthermore, a ChIP experiment for AEBP2 in the homozygous *Aebp2* gene-trapped cells displayed strongly reduced signal (Figure 7.3A in the following chapter).

Another issue that has arisen is that the *Aebp2* targeted ES cells are aneuploid. Aneuploidy is very common in ES cell lines (Kim et al., 2013) and I have the appropriate control wild type cells (the cells without *Cre* nucleofected) to eliminate any observation arising from this fact. However, when embryos with the correct genetic configuration are obtained, ES cell lines with correct ploidy can be derived from these blastocysts. Another

concern could be the relatively high passage number (P55) of the cell lines due to the high number of steps in generating the homozygous *Aebp2* gene-trapped line. Deriving ES cell lines from blastocysts will therefore also help me correct for any non-viable mutations that may have accumulated during the *in vitro* passaging of the ES cells in order to create the homozygous *Aebp2* gene-trapped line.

I have successfully generated ES cells with a gene-trap inserted into both *Aebp2* alleles, successfully trapping *Aebp2* mRNA transcripts and reducing AEBP2 protein. These cell lines are being used to generate *Aebp2* gene-trapped mice to analyse the function of AEBP2 *in vivo*.

7. Preliminary analysis of *Aebp2* gene-trapped cells

7.1 Introduction

Having generated homozygous *Aebp2* gene-trapped cells, I subsequently analysed the effects of AEBP2 depletion in mouse ES cells. Research into the effects of depletion of PRC2 core components and PRC2 associating factors, as well as the *Aebp2* homologue in *Drosophila melanogaster*, has been reported and a summary of the findings is presented here.

7.1.1 Effects of depletion of PRC2 core components

Several groups have studied the effects of genetic removal of the core components of the PRC2 complex. The main defect seen in embryos depleted for EZH2, EED or SUZ12 is a failure to complete gastrulation. Gastrulation initiates at E6.5 and finishes around E7.5-8.0. At the start of gastrulation, cells on the posterior side of the epiblast form a primitive node, from which a region known as the primitive streak forms. The primitive streak elongates towards the ventral region of the embryo and cells migrate across it to eventually differentiate into the three germ layers: ectoderm, mesoderm and endoderm. Interestingly, first expression of any *Hox* gene is detected at E7.2, which is after the appearance of defects observed in PRC2 mutants, leading to the suggestion that death may not be due to misregulation of the *Hox* genes (Wang et al., 2002; Deschamps and van Nes, 2005).

Mutations in the *Eed* locus were found to be embryonic lethal (Niswander et al., 1988) and later the *Eed* lethal mutation was characterised and has been used in many studies (including in this study, described in 5.1.2) (Schumacher et al., 1996). *Eed*$^{-/-}$ embryos fail to complete gastrulation, a phenotype which is most obviously seen at E8.5. WT embryos at E8.5 show extensive axial development including subdivision of the mesoderm into somites, node and notochord, whilst *Eed*$^{-/-}$ embryos appear to lack embryonic endoderm and have much less, seemingly undifferentiated, endoderm, and generally resemble E7.5 embryos (Faust et al., 1995). In contrast to the lack of embryonic endoderm, extraembryonic endoderm is overproduced and forms a bulge. Earlier in development, at E7.5, differences can also be seen between WT and *Eed*$^{-/-}$ embryos: mutant embryos are smaller and often resemble E6.0-6.5 embryos. It was shown that a hypomorphic allele of *Eed* resulted in a posterior transformation along the axial skeleton (Faust et al., 1995; Wang et al., 2002).

Similar phenotypes were described for homozygous *Suz12*$^{-/-}$ and *Ezh2*$^{-/-}$ knockout mice. E8.5 *Suz12*$^{-/-}$ embryos resemble E7.5 WT embryos (Pasini et al., 2004). There is no clear distinction between the three embryonic layers and there is an overproduction of extraembryonic endoderm. Like *Eed*$^{-/-}$ embryos, the neurectoderm is poorly developed and there is no indication that organogenesis has initiated. *Ezh2*$^{-/-}$ embryos also do not complete gastrulation (O'Carroll et al., 2001). In summary, studies of murine development demonstrate that mutant embryos have gastrulation defects in the absence of PRC2 core components.

The effects of genetic depletion of PRC2 core components have also been studied in the context of ES cells. It was found that ES cells lacking EZH2, SUZ12 or EED could be maintained in an undifferentiated state (Morin-Kensicki et al., 2001; Montgomery et al., 2005; Pasini et al., 2007; Chamberlain et al., 2008; Margueron et al., 2008; Shen et al., 2008). *Suz12*$^{-/-}$ and *Ezh2*$^{-/-}$ ES cells did not have decreased levels of pluripotency markers such as *Nanog* and *Oct4*, whilst some reduction was observed in *Eed*$^{-/-}$ ES cells; however, in *Eed*$^{-/-}$ cells no change in expression for NANOG and OCT4 target genes was observed. *Suz12*$^{-/-}$, *Ezh2*$^{-/-}$ and *Eed*$^{-/-}$ ES cells can give rise to teratomas, indicating that they could still differentiate into cell types of different lineages. Interestingly, whilst *Suz12*$^{-/-}$ and *Ezh2*$^{-/-}$ ES cells were morphologically indistinguishable from WT cells, *Eed*$^{-/-}$ cells were found to grow more slowly and to have a tendency to differentiate, perhaps consistent with the decreased levels of pluripotency markers (Boyer et al., 2006; Chamberlain et al., 2008; Shen et al., 2008).

In all PRC2 core protein mutants global levels of H3K27me3 and H3K27me2 were depleted whilst H3K27 monomethylation appeared unaffected. Importantly, in *Ezh2*$^{-/-}$ ES cells, some PcG target genes remain marked by H3K27me3 because of the compensatory action of EZH1, and when EZH1 was depleted in addition no H3K27 methylation was detected, including monomethylation (Shen et al., 2008). It is possible that monomethylation detected in the *Eed*$^{-/-}$ and *Suz12*$^{-/-}$ cells may be due to cross-reactivity of the antibody with H3K9me1, as Lysine-9 is embedded in a very similar amino acid sequence as Lysine-27 on the H3 tail (Perez-Burgos et al., 2004; Sarma et al., 2004). Presumably due to a loss of H3K27me3, some PcG target genes were derepressed in *Suz12*$^{-/-}$ and *Eed*$^{-/-}$ cells. PcG target genes were derepressed in *Ezh2*$^{-/-}$ cells only after RNAi-mediated knock-down of *Ezh1*.

In the absence of PRC2, not all PcG target genes become derepressed. For example, it was found that 27% of genes that are marked by H3K27me3 in WT cells become at least 2-fold derepressed in *Eed*$^{-/-}$ cells and that there is a bias towards derepression of bivalent genes versus genes marked only by H3K27me3, suggesting that the presence of activating marks makes a gene more likely to become derepressed (Chamberlain et al., 2008). Therefore, maybe not all H3K27me3-genes become derepressed due to the absence of activating marks; alternatively, their repression is maintained by other repressors also present. In some cases, this is PRC1, as in *Ring1b Eed* double knockout cells more PcG target genes are upregulated (Leeb et al., 2010). However, a large proportion remained repressed. These may be repressed by other repressors still. It is also possible that the upregulation of certain PcG repressed genes would be incompatible with the maintenance of cells in culture, leading to a selective loss of these cells.

Additionally, it was found that PRC2 core component mutant cells displayed defects in differentiation. Mutant ES cells were found to be unable to fully induce lineage specific genes, whilst failing to fully extinguish pluripotency markers. It was reported that *Suz12*$^{-/-}$ ES cells were unable to differentiate into neurons after treatment with all-trans retinoic acid (ATRA), whereas neuron formation was seen during differentiation of *Eed*$^{-/-}$ ES cells (Pasini et al., 2007). Morphologically, *Suz12*$^{-/-}$ ES cell derived embryoid bodies were often smaller and displayed a disorganised structure. Taken together, absence of PRC2 core components leads to an upregulation of PcG targets in undifferentiated ES cells, yet leads to a failure to fully induce these genes during differentiation.

7.1.2 Effects of depletion of PRC2 associating factors

Depletion of JARID2 and members of the polycomblike family has effects on embryonic development and polycomb function. *Jarid2* was identified as a regulator of neural development (Takeuchi et al., 1995) and later was described to be involved in heart development and haematopoiesis (Kitajima et al., 1999; Ohno et al., 2004). The exact time point at which embryos die varies in different studies, and most likely depends on genetic background. Typically it occurs between E10.5 and E15.5. The effect of *Jarid2* depletion in ES cells has shown that JARID2 is required for full enrichment of PRC2 and PRC1 at PcG target sites (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010). However, it has also been shown that JARID2 enrichment is dependent on PRC2, indicating that PRC2 and JARID2 recruitment are inter-dependent. However, the effect on H3K27me3 enrichment by *Jarid2* depletion is unclear. Globally, no effect on H3K27me3 was observed upon JARID2 depletion in ES cells (Shen et al., 2009; Li et al., 2010), whilst Peng et al. (2009) report a modest increase in *Xenopus laevis*. Some studies report a decrease of H3K27me3 at PcG target genes (Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010) whilst others report no change or an increase (Peng et al., 2009; Shen et al., 2009). In *Drosophila melanogaster* it was reported that *jarid2* mutants displayed increased H3K27me3 (Herz et al., 2012a). Importantly, *Jarid2* depleted cells do not display derepression of PcG target genes as was shown for depletion of the PRC2 core components. In fact, PcG target gene expression appeared to be slightly lower in *Jarid2* depleted cells as compared to the wild type controls (Shen et al., 2009; Landeira et al., 2010). Nonetheless, like PRC2 core component mutants, *Jarid2* depleted cells showed defects in differentiation.

The effect of depletion of PRC2 associating protein *polycomblike* (*pcl*) and its mammalian homologues on PRC2 function has also been studied. *Pcl* was first identified in *Drosophila melanogaster* as an interactor of E(z) (O'Connell et al., 2001; Tie et al., 2003). It was shown that homozygous mutants die at the end of embryogenesis (Nekrasov et al., 2007). Subsequently homozygous mutant clones in imaginal discs were analysed, and it was found that both H3K27me3 and H3K27me1 were reduced as assessed by immunofluorescence. By ChIP, it was also found that H3K27me1, H3K27me2 and H3K27me3 levels were reduced at PcG target genes. Some target genes became derepressed, whilst others did not, even though H3K27me3 levels were reduced there too. PRC1 enrichment was not affected by *pcl* depletion.

Mammalian genomes contain 3 homologues of *pcl*: *PCL1/PHF1, PCL2/MTF2, PCL3/PHF19* and it appears that depletion of each of these has slightly different effects. It was found that knockdown of PCL1 and PCL3 led to a decrease in H3K27me3 (Cao et al., 2008; Sarma et al., 2008; Brien et al., 2012; Hunkapiller et al., 2012; Cai et al., 2013), whereas knockdown of PCL2 did not seem to affect H3K27me3 levels (Walker et al., 2010; Casanova et al., 2011; Li et al., 2011b). Similarly, effects on PRC2-target gene expression have been demonstrated for PCL1 and PCL3 (Cao et al., 2008; Sarma et al., 2008; Hunkapiller et al., 2012; Cai et al., 2013) yet no effects or varying effects were reported for PCL2 (Walker et al., 2010; Casanova et al., 2011). On the other hand, recruitment of PRC2 was affected by knockdown of PCL2 and PCL3 (Casanova et al., 2011; Li et al., 2011b; Brien et al., 2012; Hunkapiller et al., 2012; Cai et al., 2013) but was reported as unaltered for PCL1. The observation that differentiation defects have also been reported for PCL2 and PCL3 (Walker et al., 2010; Brien et al., 2012) but not for PCL1 is perhaps consistent with the defects observed in PRC2 recruitment.

Summarising the findings for JARID2 and PCL proteins, it seems that mutations in PRC2-associating factors typically do not lead to major changes in H3K27me3 levels or PcG target expression, but do lead to differentiation defects.

7.1.3 Effects of depletion of *Aebp2* and *Jing*, the *Drosophila* homologue

So far, one study has addressed the effects of *Aebp2* depletion through use of a gene-trap. As described in Chapter 6, Kim et al. (2011) made use of a commercially available ES cell line in which one *Aebp2* allele contained a gene-trap inserted just upstream of the second exon. No homozygotes were found at E10.5 or E14.5, and the authors proposed that the *Aebp2* gene-trap is embryonic lethal before E10.5. Homozygote phenotypes were not further investigated. The consequences of heterozygous depletion of *Aebp2* was analysed and it was found that 50-70% of heterozygous mice, which were viable and fertile, showed defects similar to the human disease Waardenburg syndrome, including hypopigmentation and an enlarged colon. These defects are caused by developmental defects in neural-crest derived ganglia and melanocytes. *Aebp2* expression was detected in neural crest cells of E9.5 embryos and in tissues derived from neural crest cells in E15.5 embryos. Comparing published datasets for *Aebp2* and *Ezh2* expression, it appears that they are co-expressed in most tissues at E9.5, but that *Aebp2*'s expression pattern is more restricted than that of *Ezh2* at E15.5, where there is a strong enrichment for *Aebp2* at the footplate, handplate, thymus, midbrain and reproductive system (Kim et al, 2011; Caretti et al, 2004; Diez-Roux et al, 2011; Ningxia et al, 2011). Based on these observations, Kim and colleagues proposed a critical role for *Aebp2* in the migration and development of neural crest neurons. However, the role of *Aebp2* during early development and in ES cells remains unclear.

As discussed in the Introduction, *Drosophila melanogaster* contains an *Aebp2* homologue called *jing*. *jing* function has typically been studied by analysing genetic interactions with other developmentally important genes. Roles have been described for *jing* in border cell migration (Liu and Montell, 2001), the Central Nervous System (CNS) (Sedaghat et al., 2002; Sedaghat and Sonnenfeld, 2002), developing tracheae (Sedaghat et al., 2002; Sedaghat and Sonnenfeld, 2002; Morozova et al., 2010), the developing leg (Culi et al., 2006), and leg disc regeneration (McClure and Schubiger, 2008). It is currently unclear to which extent these phenotypes rely on a possible interaction with PRC2 (Culi et al., 2006; Herz et al., 2012a) and might be independent of PRC2, possibly mediated by the large N-terminal part of *jing* that bears no homology to *Aebp2*. Taken together, it is clear *jing* is important for the development of many organs.

7.1.4 Investigating *Aebp2* gene-trapped phenotypes

I wanted to study phenotypes of homozygous *Aebp2* gene-trapped ES cells. Based on reports on depletion of PRC2 core components and PRC2 associating factors, I decided to test PRC2 stability, levels of H3K27me3, recruitment of PRC2 and PRC1 and PcG target gene expression in the absence of AEBP2.

7.2 PRC2 stability is unaffected in the absence of AEBP2

The structure presented by Ciferri et al. (2012) indicated a bridging role for AEBP2 in the PRC2 complex. They also found that EM structures could not be obtained with the EED-SUZ12-EZH2-RBBP4 complex, yet could after addition of AEBP2, suggesting that the stability of the PRC2 complex could be affected by the absence of AEBP2.

Therefore, I tested PRC2 stability by gel filtration. I found that SUZ12, EED and EZH2 retained an elution profile identical to WT cells, indicating that PRC2 stability is not affected in homozygous *Aebp2* gene-trapped cells.

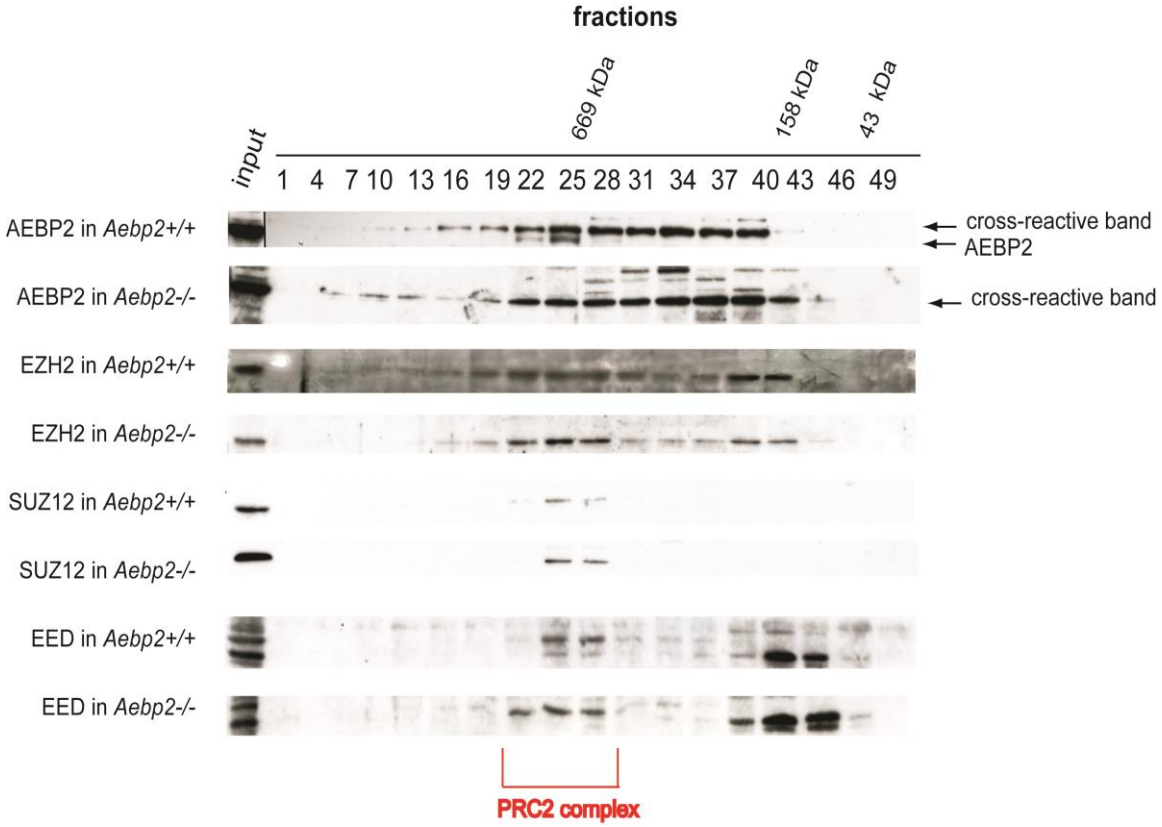


Figure 7.1: PRC2 stability is not affected in *Aebp2* gene-trapped cells. Elution profiles from the Superose 6 gel filtration column show that PRC2 core components elute in the same fractions in *Aebp2* gene-trapped cells and WT cells.

7.3 H3K27me3 levels and PcG recruitment is unaffected in *Aebp2* gene-trapped cells

I then tested whether global levels of H3K27me3 were altered due to the absence of AEBP2. It has been shown that the presence of AEBP2 increased PRC2-mediated methylation levels *in vitro* (Cao and Zhang, 2004). However, the *Aebp2* gene-trapped cells did not show a depletion of global levels of H3K27me3 as might be expected, but

instead showed a slight increase, though the extent of this increase was variable across biological repeats (Figure 7.2).

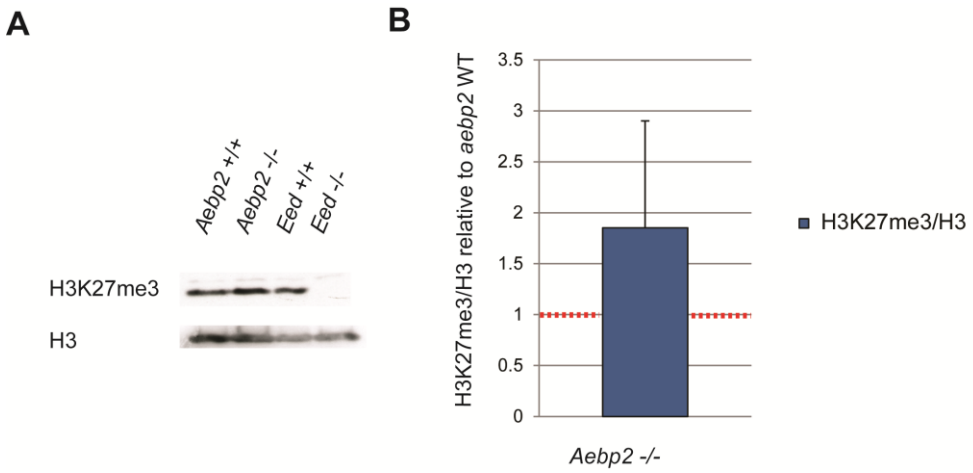
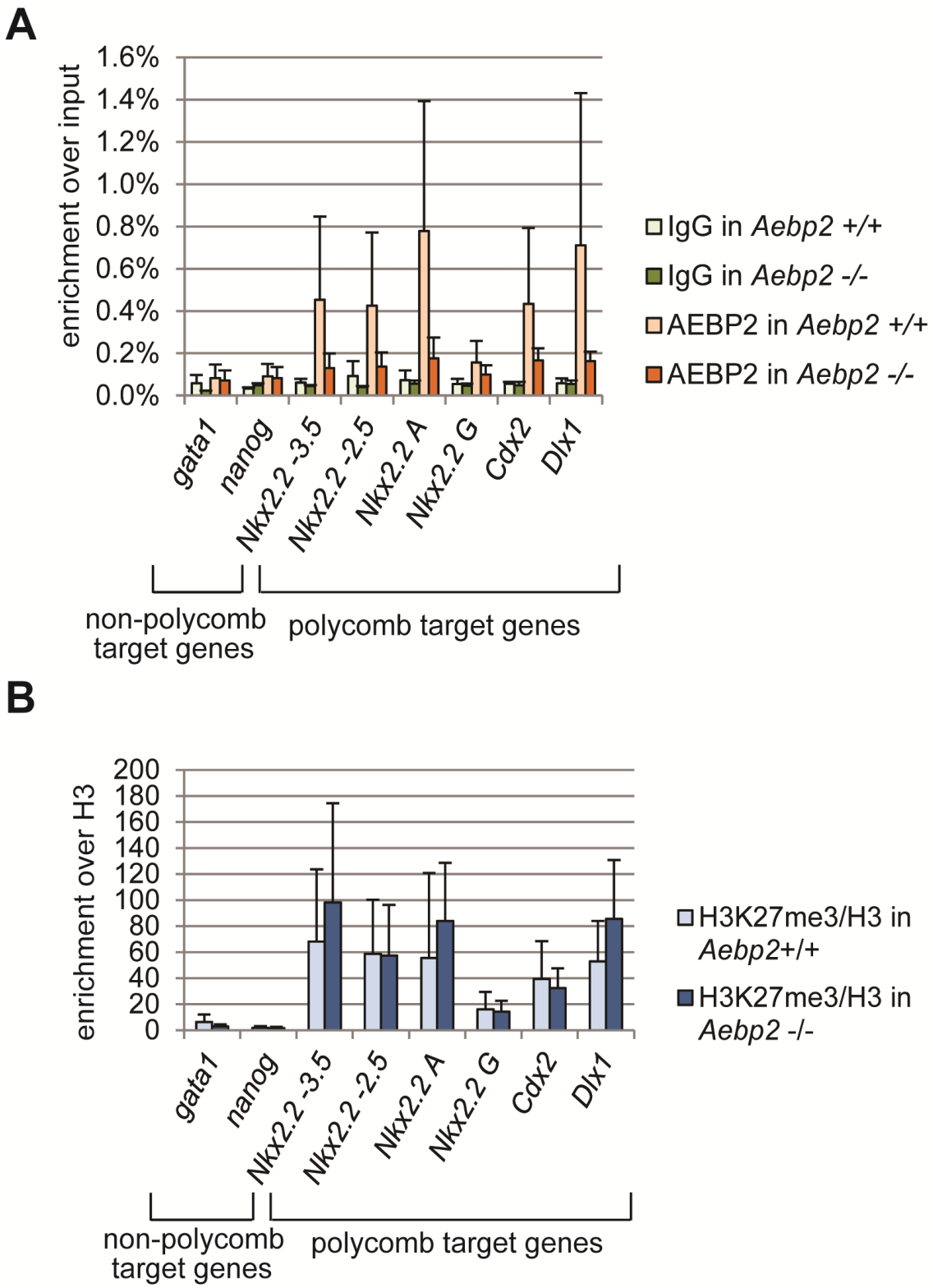


Figure 7.2: H3K27me3 levels are not reduced in the absence of AEBP2. (A) H3K27me3 and H3 levels by immunoblotting. *Eed*$^{+/+}$ and *Eed*$^{-/-}$ cells were used as a positive and negative control. (B) H3K27me3 levels in *Aebp2* gene-trapped cells relative to WT cells, normalised to H3 levels. A variable increase, rather than decrease, in H3K27me3 levels was observed for the *Aebp2* gene-trapped cells relative to WT. Error bar represents standard deviation of three biological repeats.

I then tested H3K27me3 levels by ChIP. It was found that, as expected, AEBP2 enrichment was reduced to background levels in the *Aebp2* gene-trapped cells (Figure 7.3). H3K27me3 enrichment levels, however, were not different between WT and mutant cells, nor were SUZ12 levels or levels of the PRC1-component RING1B (Figure 7.3B, C). Together, these results suggest that AEBP2 is not important for overall H3K27me3 levels or local levels at the loci tested, or for the recruitment of PRC2 or PRC1.


A
enrichment over input
1.6%
1.4%
1.2%
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
gata1
nanog
Nkx2.2 -3.5
Nkx2.2 -2.5
Nkx2.2 A
Nkx2.2 G
Cdx2
Dlx1
IgG in Aebp2 +/+
IgG in Aebp2 -/-
AEBP2 in Aebp2 +/+
AEBP2 in Aebp2 -/-
non-polycomb target genes
polycomb target genes
B
enrichment over H3
200
180
160
140
120
100
80
60
40
20
0
gata1
nanog
Nkx2.2 -3.5
Nkx2.2 -2.5
Nkx2.2 A
Nkx2.2 G
Cdx2
Dlx1
H3K27me3/H3 in Aebp2+/+
H3K27me3/H3 in Aebp2 -/-
non-polycomb target genes
polycomb target genes

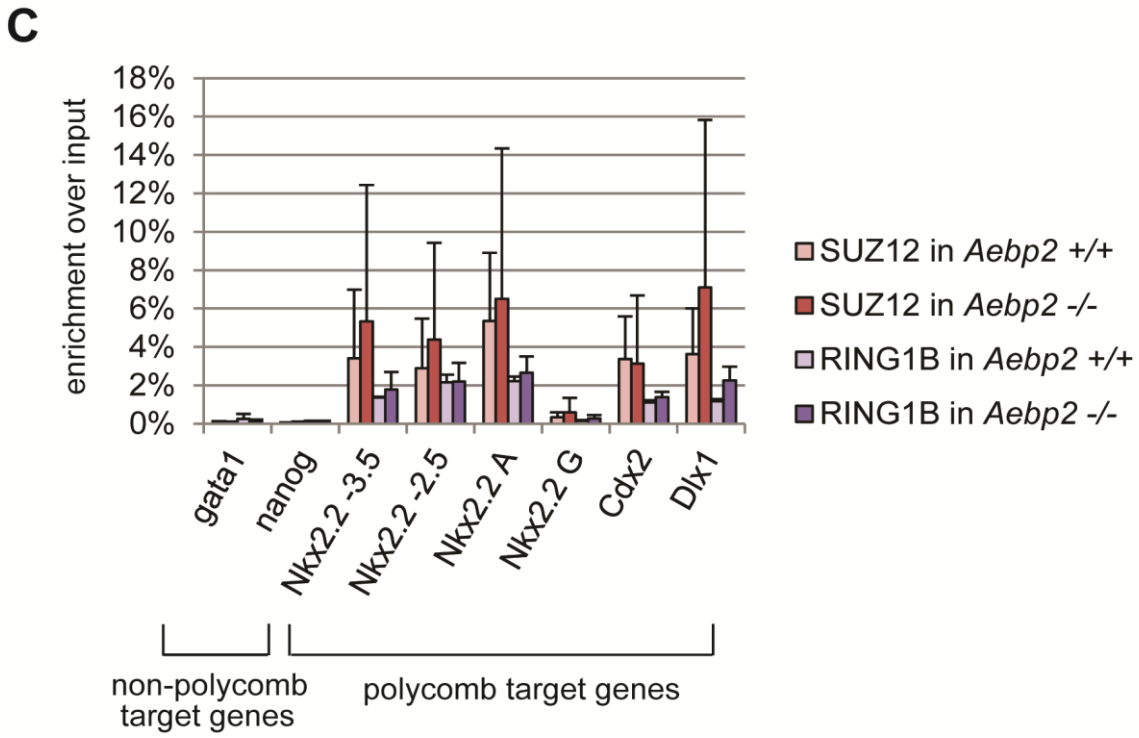


Figure 7.3: H3K27me3 and PcG recruitment is not affected in *Aebp2* gene-trapped cells. IgG and AEBP2 enrichment (A), H3K27me3 enrichment (B), and SUZ12 and RING1B enrichment (C) at PcG target and non-target genes. Error bars represent standard deviation of 3 biological repeats.

7.4 PcG target genes are not derepressed in homozygous *Aebp2* gene-trapped cells

In order to test whether AEBP2 is important for PRC2-mediated repression, I tested the expression of several PcG target genes, as well as the level of pluripotency marker *Rex1* and PRC2 core component *Eed* (Figure 7.4). It was shown that *Rex1* levels are unchanged indicating that there is likely no change in level of pluripotency of *Aebp2* gene-trapped cells, similar to PRC2 mutants. Also, levels of *Eed* mRNA are unchanged, consistent with an unchanged level of EED protein (Figure 7.1). Levels of PcG target gene expression seem unchanged (*Sox9, HoxA4, Mrg1*) or lower (*Ebf1, Pax3, Irx3*), although typically the amount of biological variation was large, so that *Aebp2* was the only gene that was significantly downregulated.

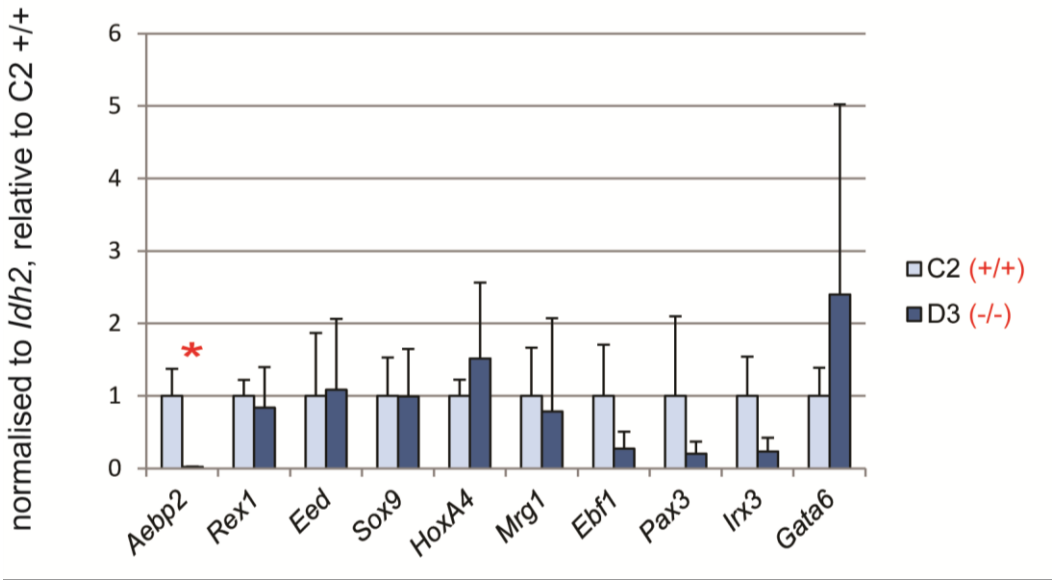


Figure 7.4: PcG target genes are not derepressed in *Aebp2* gene-trapped cells. Asterisk indicates statistical significance in a Student's T-test ($p \leq 0.05$). Error bars show standard deviation of three biological repeats.

7.5 Discussion

From my preliminary analysis of the homozygous *Aebp2* gene-trapped cells, it appears that AEBP2 depletion has no effect on PRC2 stability, global H3K27me3 levels and H3K27me3 levels at specific PcG targets, PRC2 or PRC1 recruitment to PcG target genes or PcG target gene expression.

Although AEBP2 appears to play a pivotal structural role in an EM structure generated of the AEBP2-PRC2 complex (Ciferri et al., 2012), AEBP2 is a substoichiometric component (Smits et al., 2013) and therefore it is not surprising that the PRC2 remains stable in the absence of AEBP2. The observation that the elution profile from the gel filtration column remains the same in the *Aebp2* gene-trapped cells suggests that inclusion of AEBP2 does not lead to a much bigger radius of the PRC2 complex.

H3K27me3 levels appeared to be slightly elevated in the *Aebp2* gene-trapped cells; however, further biological repeats are required to establish this observation. Nonetheless, both from the analysis of global H3K27me3 levels by immunoblot, as well as analysis of H3K27me3 levels at specific targets by ChIP-qPCR, it seems clear that H3K27me3 levels are not reduced in the absence of AEBP2. This is surprising given that it was shown that AEBP2 stimulates PRC2-mediated H3K27me3 deposition *in vitro* (Cao and Zhang, 2004). However, it is possible that AEBP2 association with the PRC2 complex could prohibit binding of other PRC2-associating factors (see Chapter 5) which may also stimulate H3K27me3 deposition (e.g. PCL1) and therefore, in the absence of AEBP2, these factors could compensate for AEBP2 loss. Alternatively, there may be other proteins *in vivo* that negatively regulate the ability of AEBP2 to stimulate H3K27me3 deposition.

Perhaps the most surprising finding is that the absence of AEBP2 does not appear to affect PRC2 or PRC1 recruitment. Where other PRC2 associating factors such as JARID2 and PCL2 also did not alter H3K27me3 or PcG target gene expression, they did have an effect on PRC2 recruitment. This observation, in conjunction with data from DNA binding experiments (see Chapter 3) suggests that AEBP2 is not a PRC2-recruiting factor as has previously been suggested (Kim et al., 2009a). Because only a limited set of loci was tested, the generality of this observation needs to be investigated by analysing more loci or performing ChIP-seq. A further possibility is that defects may not be present in the maintenance of PcG silencing, but there could be a defect in establishing silencing

de novo. In this regard, it will be interesting to see if there are any defects of PcG recruitment during XCI or during differentiation.

PcG target expression was not derepressed in *Aebp2* gene-trapped cells. This is not surprising because depletion of other PRC2 associating factors also do not display this phenotype. RNA-seq is needed to confirm whether any or no PcG target genes or any other genes are affected by the absence of AEBP2.

All mutations in PRC2 core components and PRC2-associating factors lead to defects in differentiation. Therefore, it is of great importance to test the ability of AEBP2 depleted cells to differentiate. Observations that embryos with a constitutively activated gene-trap in the *Aebp2* loci die before E10.5 (Kim et al., 2011) point to a defect in gastrulation as was observed for mutants in the PRC2 core components (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004), or in early organogenesis, which may indicate an *in vivo* defect in differentiation. Designing an embryoid body formation protocol that leads to a very reproducible induction of lineage-specific genes in control wild type cells has proven difficult so far, as is also evidenced by many fold difference in induction of genes in published reports (e.g., Landeira et al., 2010; Pasini et al., 2010). Therefore in future studies it will be important to optimise the protocol or use a directed differentiation protocol to assess the role of AEBP2 in differentiation.

To understand the role of AEBP2 it will also be possible to analyse the homozygous *Aebp2* gene-trapped embryos. This will first confirm previous reports about the lethality

of *Aebp2* gene-trapped embryos and then further refine analysis of defects at different developmental time points. It will also be possible to cross the conditionally gene-trapped lines to mouse lines expressing *Cre* under the control of various promoters. For example, the importance of extraembryonic *Aebp2* could be assessed by crossing the conditional AEBP2 mouse line to a mouse line expressing *Cre* under the control of the Sox2 promoter, which results in ubiquitous knockout of AEBP2 in the embryonic tissues (but not extraembryonic tissue) at E6.5 (Hayashi et al., 2002). If such embryos are not embryonic lethal, it will be particularly interesting to analyse phenotypes associated with Polycomb proteins such as homeotic transformations, changes in hematopoietic cell number and skeletal defects. The conditionally gene-trapped line could also be crossed to mouse lines that express *Cre* at later stages to assess the role of AEBP2 in later development. Additionally, it will be interesting to assess whether there are any genetic interactions between AEBP2 and other Polycomb components, which can be achieved by crossing the *Aebp2* gene-trapped line to mouse lines mutated in Polycomb components, such as a mouse homozygous for a hypomorphic *Ring1b* allele (Suzuki et al., 2002). A further possible crossing is a crossing with the RosaCreERT2 mouse line, which expresses *Cre* ubiquitously but is restricted to the cytoplasm in the absence of tamoxifen, but imported into the nucleus when exposed to tamoxifen, which would lead to an inducible gene-trap set up (Hameyer et al., 2007). It would also be useful to derive ES cells from RosaCreERT2 *Aebp2* gene-trap blastocysts so that any phenotypes seen in *in vitro* experiments using the homozygous *Aebp2* gene-trapped ES cells will not be obscured by any adaptation events that may occur in these cells when kept in culture over many passages.

In summary, preliminary research shows that the absence of AEBP2 does not lead to a reduction of H3K27me3 or changes in PRC2 and PRC1 recruitment. Further studies are required to analyse the role of AEBP2 during development and differentiation.

8. General discussion

8.1 AEBP2 in the PRC2 complex

AEBP2 has been identified as an associating factor of the PRC2 complex in a series of *in vitro* and *in vivo* studies. I have also shown that AEBP2 associates with PRC2 as determined by mass spectrometry (Chapter 5). Therefore AEBP2 might be expected to colocalise with PRC2 at least at some location. Using both a model of X chromosome inactivation and endogenous X chromosome inactivation occurring during the differentiation of ES cells, I have shown that AEBP2, like PRC2 and associating factor PCL2, becomes enriched on the inactive X during XCI. Additionally, I have shown that endogenous AEBP2 is enriched at similar locations as H3K27me3, and epitope-tagged AEBP2 is enriched at PRC2 targets genomewide. My analysis suggests that AEBP2 does not localise to any places other than PRC2 targets although this point requires further investigation. Whether AEBP2 localises to all, or a subset of, PRC2 targets is difficult to determine, because high ChIP-seq background levels prevent assigning AEBP2 peaks unequivocally to some loci.

It would be interesting to do further analysis on this dataset as this may reveal a differential enrichment for different classes of PRC2 targets (such as developmental regulators involved in different lineages) which could provide insights into AEBP2's role within the PRC2 complex. It would also be interesting to analyse the distribution of AEBP2 in differentiated tissues, especially in those in which AEBP2 has been proposed to play a role during development (Kim et al., 2011). It would also be important to verify

the ChIP-seq experiment by testing the distribution of endogenous AEBP2. As the antibody developed by myself does immunoprecipitate AEBP2 (as shown in Figure 6.16) and I have obtained promising results in ChIP-qPCR, I am hopeful that this will be possible.

8.2 PRC2 targeting

It has been suggested that AEBP2 could be required for targeting the PRC2 complex in a sequence-specific manner (Kim et al., 2009a). However, my results do not support this hypothesis.

Firstly, *in vitro* binding experiments have failed to show a physiological interaction between AEBP2 and DNA for both the full length protein and the zinc finger domain expressed by itself (Chapter 3). As it is hard to prove a non-interaction, the possibility that AEBP2 binds DNA (be this in a sequence-specific manner or not) cannot be excluded. However, my results illustrated the importance of using controls to show that any interaction detected is definitely due to the protein of interest and not due to a contamination. As both studies that reported an interaction between AEBP2 and DNA did not include these controls, I believe their results should be interpreted with caution. A study showing robust interaction between AEBP2 and DNA should at the least include an EMSA with supershift or show loss of binding after mutation of binding interphases. Ideally further methods such as NMR would also be used to demonstrate protein-DNA interaction.

A second finding arguing against AEBP2 being a sequence-specific targeting factor is the observation that, in the absence of EED, AEBP2 is largely depleted from PRC2 targets as observed by ChIP-qPCR. This demonstrates that AEBP2 localisation is dependent on the PRC2 complex. If AEBP2 was a targeting factor it would be expected that it could localise independently of EED. An example of such behaviour is shown by JARID2, which is a PRC2 associating factor and is important for PRC2 targeting (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010). JARID2 can still be recruited to PRC2 targets in the absence of EED and in the absence of EZH2, although enrichment levels are lower (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li et al., 2010). In contrast to the lack of recruitment observed by ChIP-qPCR, I have shown that AEBP2 can still be recruited to a model of the inactive X in the absence of EED. However, AEBP2 recruitment could either be due to AEBP2 itself or due to its association with SUZ12, which also still localises there.

Thirdly, in homozygous *Aebp2* gene-trapped cells, neither PRC2 nor PRC1 enrichment levels appear to be affected by ChIP-qPCR. This suggests that AEBP2 cannot be essential for PRC2 recruitment. It is possible that AEBP2 contributes to targeting redundantly with JARID2 or another PRC2 associating factor. To analyse this it would be interesting to generate a double knockout cell line and observe the effects on PRC2 targeting.

Taken together, it seems unlikely that AEBP2 plays an essential role in PRC2 targeting. It is possible it could still contribute to targeting during the establishment of new PRC2

domains and the resolution of old ones. To investigate this, targeting of PRC2 during differentiation or XCI needs to be investigated.

8.3 Composition of PRC2 complexes

During my study of the association of AEBP2 with the PRC2 complex, I made two interesting observations.

Firstly, I found that in *Eed*$^{-/-}$ cells, SUZ12 and AEBP2 proteins could still be detected and localised to an Xist-silenced chromosome, albeit at somewhat lower levels than in wild type cells. This is surprising because it is a widely held view that in the absence of EED, the other core components of PRC2 become degraded and PRC2 functionality is removed. However, upon close inspection of the literature it is clear that SUZ12, and to a lesser extent EZH2, are still present in *Eed*$^{-/-}$ cells and possibly interact (Shen et al., 2008; Shen et al., 2009; Dietrich et al., 2012). The fact that SUZ12 and AEBP2 are seen to localise to an inactivated chromosome demonstrates that a PRC2 subcomplex may retain some of its function. It would be interesting to further investigate this question by performing ChIP-seq to identify whether SUZ12 may still be somewhat enriched at target sites in the absence of EED. However, I did confirm that H3K27me3 levels are undetectable as previously published (Montgomery et al., 2005) suggesting that any retained function is unlikely to be related to histone modification. One way to analyse this would be to generate double knockout cells of e.g. *Eed* and *Suz12* and observe whether any phenotypes are exacerbated. Another approach would be to see if chemical inhibitors which block the HMT activity of EZH1/2 would create phenotypes that are

identical to or less severe than *Ezh1/2* deletions (Knutson et al., 2012; McCabe et al., 2012).

Secondly, I found that epitope-tagged AEBP2 immunoprecipitated JARID2, yet did not immunoprecipitate PRC2 associating factors of the polycomblike family, or the poorly characterised PRC2 associating factor esPRC2p48 (Chapter 5). A survey of published immunoprecipitation experiments showed that AEBP2 and JARID2 were generally mutually exclusive with PCL members and esPRC2p48. Together, my results and published reports suggest that there are distinct PRC2 complexes, likely with specialised roles. What these roles may be requires further investigation. JARID2 and PCL proteins have been implicated in targeting and so each may target one PRC2 version to particular regions of the genome. However, ChIP-seq experiments typically show that PRC2 associating factors are enriched at all PRC2 targets. Another possibility is that the associating factors can in turn recruit other proteins to aid PRC2 function, such as has been suggested for PCL3 (Brien et al., 2012). One way to characterise different roles of the associating factors may be to perform co-immunoprecipitations in less stringent conditions which could uncover lower-affinity binding partners. Further structural information about how associating factors interact with the PRC2 complex may provide a rationale for exclusivity, whilst creating double knockout mutants may demonstrate whether some of these factors are acting redundantly.

8.4 The role of AEBP2

It has long been established that PRC2 represses developmentally important genes, and it has been shown that it does so through tri-methylating H3K27. That the catalytic subunit

EZH2 cannot trimethylate H3K27 on its own has highlighted the importance of the other two core components of PRC2: SUZ12 and EED (Cao and Zhang, 2004; Pasini et al., 2004; Ketel et al., 2005; Nekrasov et al., 2005). Exactly why these two factors are required for enzymatic activity is unclear but likely do so through allosteric activation of the SET domain (Ciferri et al., 2012).

Many associating factors have been described in the last decade and their roles are less clear. Recently, many interactions between chromatin marks and PRC2 core and associating factors have been reported (Tie et al., 2007; Margueron et al., 2009; Schmitges et al., 2011; Yuan et al., 2011; Ballaré et al., 2012; Brien et al., 2012; Musselman et al., 2012; Cai et al., 2013). Together, a picture is painted of a PRC2 complex that uses its (sub)stoichiometric subunits to "read" the state of chromatin, adjusting its activity accordingly. Experiments show that these interactions are most important during differentiation, when the distribution of H3K27me3 is altered (Pasini et al., 2007; Chamberlain et al., 2008; Shen et al., 2008).

The recently presented EM structure of PRC2 in complex with AEBP2 indicates that AEBP2 is present at the core of the PRC2 complex. Modelling of the PRC2 complex onto a dinucleosome indicated that AEBP2 might be far away from histone tails and may therefore not be involved in reading histone marks. However, AEBP2 could contribute to chromatin reading by binding to linker DNA, as in the model the zinc finger of AEBP2 was positioned close to linker DNA. This interaction could be dependent on the methylation status of the DNA and thus aid recognition of CpG islands. Importantly, the EM structure did not include the large N-terminal acidic and neutral domains that make

up almost half of full length AEBP2. This region is predicted to be largely unstructured. It would be interesting to know whether inclusion of this domain may alter possible interactions with DNA or the other PRC2 components.

Interestingly, AEBP2 was found to be present in only 20% of PRC2 complexes in HeLa cells (Smits et al., 2013). Although the percentage of AEBP2-containing PRC2 complexes in ES cells is not known, it seems unlikely it would be stoichiometric, especially given the mutual exclusion with the PCL family proteins. Therefore, despite its position at the heart of the structure as presented by Ciferri et al. (2012), it is unlikely that AEBP2 plays an essential role in PRC2 function, but rather enhances it.

AEBP2 inclusion could alter enzymatic activity of the complex, which has been suggested based on *in vitro* experiments (Cao and Zhang, 2004). However, the fact that H3K27me3 levels are not altered by *Aebp2* depletion suggests this is not the case. Therefore, taking the data presented in this study together with published data, it seems that AEBP2 does not target PRC2, nor is it important for the level of H3K27me3, nor does it appear that it binds histone marks, though it may bind DNA. So what is the role of AEBP2? The PRC2 complex seems to be an integrator of many signals and given the position that AEBP2 occupies within the complex, I hypothesize that AEBP2 may control the way chromatin signature is interpreted and the way this is relayed allosterically to other PRC2 members.

AEBP2 may achieve this in different ways: Firstly, inclusion of AEBP2 could predispose PRC2 to be in a particular conformation (Ciferri et al., 2012). This conformation could lead to increased interaction with particular chromatin marks, whilst decreasing interaction with others. Secondly, AEBP2 could regulate the presence or absence of other associating factors as has been discussed in 8.3, which could influence PRC2 functionality. AEBP2 activity itself may be regulated by the expression of different isoforms or post-translational modifications such as phosphorylation.

Together this highlights the requirement for more structural information: PRC2 in complex with other associating factors might provide a rationale for the inclusion and exclusion of associating factors. A PRC2 complex in association with DNA and nucleosomes will allow us to understand which parts of the PRC2 complex are likely to engage in chromatin reading whilst other parts may interact with other proteins or are available for posttranslational modifications which could modify PRC2 activity.

Meanwhile, *in vivo* and *in vitro* work will establish the requirement for AEBP2 during differentiation and early development. It will be interesting to study the effect of AEBP2 depletion on PRC2 biology during a time window in which many genes undergo changes in PRC2 regulation, and also the time window during which PRC2 activity has proven to be the most important. Furthermore, the conditional knockout that I created will be useful for establishing the role of AEBP2 in later development, where it has been suggested to be important for neural crest cells (Kim et al., 2011).

9. Permissions for the reproduction of figures

Figure 1.1: Obtained and adapted from Wikimedia Commons under CC BY-SA 3.0 license

Figure 1.2: Sparmann and van Lohuizen (2006): obtained through RightsLink License number 3226001277306

Figure 1.3: Mak (2007): obtained through RightsLink License number 3223611243799

Figure 1.4: Parrish et al. (2007): CC BY-NC 3.0 license

Figure 1.6: Gao et al. (2012): obtained through RightsLink License number 3226001509703

Figure 1.8: Obtained using Ensembl 72 (Flicek et al., 2013)

Figure 1.9: Ciferri et al. (2012): obtained under CC BY 3.0 license

Figure 3.1: Brayer et al. (2008) obtained through RightsLink License number 3226020069708

Figure 3.7B: Lawson et al. (2004): obtained through RightsLink License number 3226011258370

Figure 3.7C: Umezawa et al. (2008): obtained through RightsLink; no license number given

Figure 4.1: Klose et al. (2013): obtained under CC BY 3.0 license

Figure 4.9: Stock et al. (2007): obtained through RightsLink License number 3226021219228

Figure 6.2A, Figure 6.3A, Figure 6.4A have been adapted from unpublished diagrams received from I. Stancheva.

10.Bibliography

Abe, K., Niwa, H., Iwase, K., Takiguchi, M., Mori, M., Abe, S.I., and Yamamura, K.I. (1996). Endoderm-specific gene expression in embryonic stem cells differentiated to embryoid bodies. Experimental cell research *229*, 27-34.

Abed, J.A., and Jones, R.S. (2012). H3K36me3 key to Polycomb-mediated gene silencing in lineage specification. Nature structural & molecular biology *19*, 1214-1215.

Adhvaryu, K.K., Morris, S.A., Strahl, B.D., and Selker, E.U. (2005). Methylation of histone H3 lysine 36 is required for normal development in Neurospora crassa. Eukaryot Cell *4*, 1455-1464.

Agger, K., Cloos, P.A., 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.

Akam, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. Development *101*, 1-22.

Akasaka, T., Kanno, M., Balling, R., Mieza, M.A., Taniguchi, M., and Koseki, H. (1996). A role for mel-18, a Polycomb group-related vertebrate gene, during the anteroposterior specification of the axial skeleton. Development *122*, 1513-1522.

Akkers, R.C., van Heeringen, S.J., Jacobi, U.G., Janssen-Megens, E.M., Francoijs, K.J., Stunnenberg, H.G., and Veenstra, G.J.C. (2009). A Hierarchy of H3K4me3 and H3K27me3 Acquisition in Spatial Gene Regulation in Xenopus Embryos. Developmental cell *17*, 425-434.

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.

Allahverdi, A., Yang, R.L., Korolev, N., Fan, Y.P., Davey, C.A., Liu, C.F., and Nordenskiold, L. (2011). The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. Nucleic acids research *39*, 1680-1691.

An, S., Yeo, K.J., Jeon, Y.H., and Song, J.J. (2011). Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. The Journal of biological chemistry *286*, 8369-8374.

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., and Walter, J. (2012). In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS genetics *8*, e1002750.

Armstrong, J.A., Papoulas, O., Daubresse, G., Sperling, A.S., Lis, J.T., Scott, M.P., and Tamkun, J.W. (2002). The Drosophila BRM complex facilitates global transcription by RNA polymerase II. The EMBO journal *21*, 5245-5254.

Arnold, P., Scholer, A., Pachkov, M., Balwierz, P.J., Jorgensen, H., Stadler, M.B., van Nimwegen, E., and Schubeler, D. (2013). Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. Genome research *23*, 60-73.

Atanassov, B.S., Koutelou, E., and Dent, S.Y. (2011). The role of deubiquitinating enzymes in chromatin regulation. FEBS letters *585*, 2016-2023.

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.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., and Fisher, A.G. (2006). Chromatin signatures of pluripotent cell lines. Nature cell biology *8*, 532-538.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., Carlomagno, T., Benitah, S.A., and Di Croce, L. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nature structural & molecular biology *19*, 1257-1265.

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.

Bartova, E., Galiova, G., Krejci, J., Harnicarova, A., Strasak, L., and Kozubek, S. (2008). Epigenome and chromatin structure in human embryonic stem cells undergoing differentiation. Developmental dynamics : an official publication of the American Association of Anatomists *237*, 3690-3702.

Ben-Saadon, R., Zaaroor, D., Ziv, T., and Ciechanover, A. (2006). The polycomb protein Ring1B generates self atypical mixed ubiquitin chains required for its in vitro histone H2A ligase activity. Molecular cell *24*, 701-711.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., Schreiber, S.L., and Lander, E.S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

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

Bhattacharya, D., Talwar, S., Mazumder, A., and Shivashankar, G.V. (2009). Spatio-temporal plasticity in chromatin organization in mouse cell differentiation and during Drosophila embryogenesis. Biophysical journal *96*, 3832-3839.

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. Cell *152*, 1270-1284.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396-398.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic acids research *8*, 1499-1504.

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. Molecular cell *38*, 179-190.

Blackledge, N.P., Long, H.K., Zhou, J.C., Kriaucionis, S., Patient, R., and Klose, R.J. (2012). Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA. Nucleic acids research *40*.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 19*, Unit 19 10 11-21.

Boa, S., Coert, C., and Patterton, H.G. (2003). Saccharomyces cerevisiae Set1p is a methyltransferase specific for lysine 4 of histone H3 and is required for efficient gene expression. Yeast *20*, 827-835.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., Gnirke, A., Eggan, K., and Meissner, A. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439-452.

Borsani, G., Tonlorenzi, R., Simmler, M.C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Willard, H.F., Avner, P., and

Ballabio, A. (1991). Characterization of a Murine Gene Expressed from the Inactive X-Chromosome. Nature *351*, 325-329.

Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). The SANT domain: a unique histone-tail-binding module? Nature reviews Molecular cell biology *5*, 158-163.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., and Young, R.A. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., Bell, G.W., Otte, A.P., Vidal, M., Gifford, D.K., Young, R.A., and Jaenisch, R. (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 & development *20*, 1123-1136.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255-256.

Brayer, K.J., Kulshreshtha, S., and Segal, D.J. (2008). The protein-binding potential of C2H2 zinc finger domains. Cell Biochem Biophys *51*, 9-19.

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.

Brennan, R.G., and Matthews, B.W. (1989). The Helix-Turn-Helix DNA-Binding Motif. Journal of Biological Chemistry *264*, 1903-1906.

Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., Lohan, A.J., Ferguson, N., Shi, X., Sinha, K.M., Loftus, B.J., Cagney, G., and Bracken, A.P. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nature structural & molecular biology *19*, 1273-1281.

Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., Mccabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of Position and Exclusive Expression of Mouse Xist from the Inactive X-Chromosome. Nature *351*, 329-331.

Brockdorff, N. (2011). Chromosome silencing mechanisms in X-chromosome inactivation: unknown unknowns. Development *138*, 5057-5065.

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA *19*, 429-442.

Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., and Vallier, L. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., Kimura, H., Ragoussis, J., Teichmann, S.A., and Pombo, A. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157-170.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A Gene from the Region of the Human X-Inactivation Center Is Expressed Exclusively from the Inactive X-Chromosome. Nature *349*, 38-44.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The Human Xist Gene - Analysis of a 17 Kb Inactive X-Specific Rna That Contains Conserved Repeats and Is Highly Localized within the Nucleus. Cell *71*, 527-542.

Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.L., and Kassis, J.A. (1998). The Drosophila polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. Molecular cell *1*, 1057-1064.

Brown, J.L., Fritsch, C., Mueller, J., and Kassis, J.A. (2003). The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. Development *130*, 285-294.

Brown, R.S., Sander, C., and Argos, P. (1985). The Primary Structure of Transcription Factor Tfiiia Has 12 Consecutive Repeats. FEBS letters *186*, 271-274.

Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., and Sixma, T.K. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. The EMBO journal *25*, 2465-2474.

Burdach, J., O'Connell, M.R., Mackay, J.P., and Crossley, M. (2012). Two-timing zinc finger transcription factors liaising with RNA. Trends in biochemical sciences *37*, 199-205.

Burgold, T., Spreafico, F., De Santa, F., Totaro, M.G., Prosperini, E., Natoli, G., and Testa, G. (2008). The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. Plos One *3*, e3034.

Cahan, P., and Daley, G.Q. (2013). Origins and implications of pluripotent stem cell variability and heterogeneity. Nature reviews Molecular cell biology *14*, 357-368.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D., Strahl, B.D., Song, J., and Wang, G.G. (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Molecular cell *49*, 571-582.

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. Molecular cell *15*, 57-67.

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

Cao, R., Wang, H., He, J., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2008). Role of hPHF1 in H3K27 methylation and Hox gene silencing. Molecular and cellular biology *28*, 1862-1872.

Caretti, G., Di Padova, M., Micales, B., Lyons, G.E., and Sartorelli, V. (2004). The Polycomb EA2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes & development *18*, 2627-2638.

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., and Workman, J.L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Casanova, M., Preissner, T., Cerase, A., Poot, R., Yamada, D., Li, X., Appanah, R., Bezstarosti, K., Demmers, J., Koseki, H., and Brockdorff, N. (2011). Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. Development *138*, 1471-1482.

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., 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.

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes & development *20*, 2223-2237.

Choi, J.K., and Howe, L.J. (2009). Histone acetylation: truth of consequences? Biochemistry and cell biology = Biochimie et biologie cellulaire *87*, 139-150.

Choi, J.K. (2010). Contrasting chromatin organization of CpG islands and exons in the human genome. Genome Biol *11*, R70.

Chopra, V.S., Hong, J.W., and Levine, M. (2009). Regulation of Hox gene activity by transcriptional elongation in Drosophila. Current biology : CB *19*, 688-693.

Chopra, V.S., Hendrix, D.A., Core, L.J., Tsui, C., Lis, J.T., and Levine, M. (2011). The polycomb group mutant esc leads to augmented levels of paused Pol II in the Drosophila embryo. Molecular cell *42*, 837-844.

Christy, R.J., Yang, V.W., Ntambi, J.M., Geiman, D.E., Landschulz, W.H., Friedman, A.D., Nakabeppu, Y., Kelly, T.J., and Lane, M.D. (1989). Differentiation-Induced Gene-Expression in 3t3-L1 Preadipocytes - Ccaat Enhancer Binding-Protein Interacts with and Activates the Promoters of 2 Adipocyte-Specific Genes. Genes & development *3*, 1323-1335.

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. eLife *1*, e00005.

Clemson, C.M., Hall, L.L., Byron, M., McNeil, J., and Lawrence, J.B. (2006). The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. Proceedings of the National Academy of Sciences of the United States of America *103*, 7688-7693.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes & development *26*, 1714-1728.

Copur, O., and Muller, J. (2013). The histone H3-K27 demethylase Utx regulates HOX gene expression in Drosophila in a temporally restricted manner. Development *140*, 3478-3485.

Coré, N., Bel, S., Gaunt, S.J., Aurrand-Lions, M., Pearce, J., Fisher, A., and Djabali, M. (1997). Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. Development *124*, 721-729.

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

Cross, B., Chen, L.H., Cheng, Q., Li, B.Z., Yuan, Z.M., and Chen, J.D. (2011). Inhibition of p53 DNA Binding Function by the MDM2 Protein Acidic Domain. Journal of Biological Chemistry *286*.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nature genetics *22*, 323-324.

Cuddapah, S., Roh, T.Y., Cui, K., Jose, C.C., Fuller, M.T., Zhao, K., and Chen, X. (2012). A novel human polycomb binding site acts as a functional polycomb response element in Drosophila. Plos One *7*, e36365.

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.

Dahl, J.A., Reiner, A.H., Klungland, A., Wakayama, T., and Collas, P. (2010). Histone H3 Lysine 27 Methylation Asymmetry on Developmentally-Regulated Promoters Distinguish the First Two Lineages in Mouse Preimplantation Embryos. Plos One *5*.

De Cegli, R., Iacobacci, S., Flore, G., Gambardella, G., Mao, L., Cutillo, L., Lauria, M., Klose, J., Illingworth, E., Banfi, S., and di Bernardo, D. (2013). Reverse engineering a mouse embryonic stem cell-specific transcriptional network reveals a new modulator of neuronal differentiation. Nucleic acids research *41*, 711-726.

de la Cruz, C.C., Fang, J., Plath, K., Worringer, K.A., Nusinow, D.A., Zhang, Y., and Panning, B. (2005). Developmental regulation of Suz 12 localization. Chromosoma *114*, 183-192.

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., Koseki, H., and Brockdorff, N. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Developmental cell *7*, 663-676.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science *328*, 1161-1164.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes & development *25*, 1010-1022.

del Mar Lorente, M., Marcos-Gutierrez, C., Perez, C., Schoorlemmer, J., Ramirez, A., Magin, T., and Vidal, M. (2000). Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice. Development *127*, 5093-5100.

Denis, H., Ndlovu, M.N., and Fuks, F. (2011). Regulation of mammalian DNA methyltransferases: a route to new mechanisms. EMBO reports *12*, 647-656.

Deschamps, J., and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. Development *132*, 2931-2942.

Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Sodersten, E., and Hansen, K. (2012). REST-Mediated Recruitment of Polycomb Repressor Complexes in Mammalian Cells. PLoS genetics *8*.

Dodd, I.B., Micheelsen, M.A., Sneppen, K., and Thon, G. (2007). Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification. Cell *129*, 813-822.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576-578.

Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W., and Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. Journal of embryology and experimental morphology *87*, 27-45.

Dura, J.M., Randsholt, N.B., Deatrick, J., Erk, I., Santamaria, P., Freeman, J.D., Freeman, S.J., Weddell, D., and Brock, H.W. (1987). A complex genetic locus, polyhomeotic, is required for segmental specification and epidermal development in D. melanogaster. Cell *51*, 829-839.

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., Gingeras, T.R., Misteli, T., and Meshorer, E. (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 ubiquitin ligase to chromatin. Molecular cell *28*, 107-120.

Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M., and Koseki, H. (2008). Polycomb group proteins

Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development *135*, 1513-1524.

Endoh, M., Endo, T.A., Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W.A., Vidal, M., Bernstein, B.E., and Koseki, H. (2012). Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS genetics *8*, e1002774.

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

Evans, M.J. (1972). The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. Journal of embryology and experimental morphology *28*, 163-176.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Fan, Y., Nikitina, T., Morin-Kensicki, E.M., Zhao, J., Magnuson, T.R., Woodcock, C.L., and Skoultchi, A.I. (2003). H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. Molecular and cellular biology *23*, 4559-4572.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell *123*, 1199-1212.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., Koseki, H., Brockdorff, N., Ponting, C.P., Kessler, B.M., and Klose, R.J. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. eLife *1*, e00205.

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

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 & development *17*, 1870-1881.

Fisher, C.L., and Fisher, A.G. (2011). Chromatin states in pluripotent, differentiated, and reprogrammed cells. Current opinion in genetics & development *21*, 140-146.

Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., Garcia-Giron, C., Gordon, L., Hourlier, T., Hunt, S., Juettemann, T., Kahari, A.K., Keenan, S., Komorowska, M., Kulesha, E., Longden, I., Maurel, T., McLaren, W.M., Muffato, M., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H.S., Ritchie, G.R., Ruffier, M., Schuster, M., Sheppard, D., Sobral, D., Taylor, K., Thormann, A., Trevanion, S., White, S., Wilder, S.P., Aken, B.L., Birney, E., Cunningham, F., Dunham, I., Harrow, J., Herrero, J., Hubbard, T.J., Johnson, N., Kinsella, R., Parker, A., Spudich, G., Yates, A., Zadissa, A., and Searle, S.M. (2013). Ensembl 2013. Nucleic acids research *41*, D48-55.

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., 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.

Friberg, A., Oddone, A., Klymenko, T., Muller, J., and Sattler, M. (2010). Structure of an atypical Tudor domain in the Drosophila Polycomblike protein. Protein science : a publication of the Protein Society *19*, 1906-1916.

Friedman, J.M., Jones, P.A., and Liang, G. (2009). The tumor suppressor microRNA-101 becomes an epigenetic player by targeting the polycomb group protein EZH2 in cancer. Cell Cycle *8*, 2313-2314.

Fritsch, C., Brown, J.L., Kassis, J.A., and Muller, J. (1999). The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. Development *126*, 3905-3913.

Gan, Q.A., Schones, D.E., Eun, S.H., Wei, G., Cui, K.R., Zhao, K.J., and Chen, X. (2010). Monovalent and unpoised status of most genes in undifferentiated cell-enriched Drosophila testis. Genome Biol *11*.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Molecular cell *45*, 344-356.

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

Gaytan de Ayala Alonso, A., Gutierrez, L., Fritsch, C., Papp, B., Beuchle, D., and Muller, J. (2007). A genetic screen identifies novel polycomb group genes in Drosophila. Genetics *176*, 2099-2108.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., and Nekrutenko, A. (2005). Galaxy: a platform for interactive large-scale genome analysis. Genome research *15*, 1451-1455.

Gilbert, N., Boyle, S., Sutherland, H., de Las Heras, J., Allan, J., Jenuwein, T., and Bickmore, W.A. (2003). Formation of facultative heterochromatin in the absence of HP1. Embo Journal *22*, 5540-5550.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol *11*, R86.

Goo, Y.H., Sohn, Y.C., Kim, D.H., Kim, S.W., Kang, M.J., Jung, D.J., Kwak, E., Barlev, N.A., Berger, S.L., Chow, V.T., Roeder, R.G., Azorsa, D.O., Meltzer, P.S., Suh, P.G., Song, E.J., Lee, K.J., Lee, Y.C., and Lee, J.W. (2003). Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Molecular and cellular biology *23*, 140-149.

Gu, H., Marth, J.D., Orban, P.C., Mossmann, H., and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. Science *265*, 103-106.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell *7*, 249-257.

Gutierrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N., and Muller, J. (2012). The role of the histone H2A ubiquitinase Sce in Polycomb repression. Development *139*, 117-127.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., Yang, X., Amit, I., Meissner, A., Regev, A., Rinn, J.L., Root, D.E., and Lander, E.S. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295-300.

Hall, T.M.T. (2005). Multiple modes of RNA recognition by zinc finger proteins. Curr Opin Struc Biol *15*, 367-373.

Hameyer, D., Loonstra, A., Eshkind, L., Schmitt, S., Antunes, C., Groen, A., Bindels, E., Jonkers, J., Krimpenfort, P., Meuwissen, R., Rijswijk, L., Bex, A., Berns, A., and Bockamp, E. (2007). Toxicity of ligand-dependent Cre recombinases and generation of a conditional Cre deleter mouse allowing mosaic recombination in peripheral tissues. Physiol Genomics *31*, 32-41.

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. Nature cell biology *10*, 1291-U1289.

Harvey, A.C., and Downs, J.A. (2004). What functions do linker histones provide? Molecular microbiology *53*, 771-775.

Hatayama, M., and Aruga, J. (2010). Characterization of the tandem CWCH2 sequence motif: a hallmark of inter-zinc finger interactions. Bmc Evol Biol *10*.

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., and Crabtree, G.R. (2012). Dynamics and memory of heterochromatin in living cells. Cell *149*, 1447-1460.

Hayashi, S., Lewis, P., Pevny, L., and McMahon, A.P. (2002). Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Mechanisms of development *119*, S97-S101.

He, A., Shen, X., Ma, Q., Cao, J., von Gise, A., Zhou, P., Wang, G., Marquez, V.E., Orkin, S.H., and Pu, W.T. (2012). PRC2 directly methylates GATA4 and represses its transcriptional activity. Genes & development *26*, 37-42.

He, G.P., Muise, A., Li, A.W., and Ro, H.S. (1995). A eukaryotic transcriptional repressor with carboxypeptidase activity. Nature *378*, 92-96.

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. The Journal of biological chemistry *274*, 14678-14684.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nature cell biology *15*, 373-384.

Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C.D., and Spector, D.L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell *107*, 727-738.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford,

G.E., and Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics *39*, 311-318.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.

Herrera, R., Ro, H.S., Robinson, G.S., Xanthopoulos, K.G., and Spiegelman, B.M. (1989). A Direct Role for C/Ebp and the Ap-I-Binding Site in Gene-Expression Linked to Adipocyte Differentiation. Molecular and cellular biology *9*, 5331-5339.

Herz, H.M., Mohan, M., Garrett, A.S., Miller, C., Casto, D., Zhang, Y., Seidel, C., Haug, J.S., Florens, L., Washburn, M.P., Yamaguchi, M., Shiekhattar, R., and Shilatifard, A. (2012a). Polycomb repressive complex 2-dependent and -independent functions of Jarid2 in transcriptional regulation in Drosophila. Molecular and cellular biology *32*, 1683-1693.

Herz, H.M., Mohan, M., Garruss, A.S., Liang, K., Takahashi, Y.H., Mickey, K., Voets, O., Verrijzer, C.P., and Shilatifard, A. (2012b). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes & development *26*, 2604-2620.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.W., Lyou, Y., Townes, T.M., Schubeler, D., and Gilbert, D.M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. PLoS biology *6*, e245.

Hoffmann, A., Barz, T., and Spengler, D. (2006). Multitasking C2H2 zinc fingers link zac DNA binding to coordinated regulation of p300-histone acetyltransferase activity. Molecular and cellular biology *26*, 5544-5557.

Hon, G.C., Hawkins, R.D., and Ren, B. (2009). Predictive chromatin signatures in the mammalian genome. Human molecular genetics *18*, R195-201.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research *37*, 1-13.

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

Hublitz, P., Albert, M., and Peters, A.H. (2009). Mechanisms of transcriptional repression by histone lysine methylation. The International journal of developmental biology *53*, 335-354.

Hunkapiller, J., Shen, Y., Diaz, A., Cagney, G., McCleary, D., Ramalho-Santos, M., Krogan, N., Ren, B., Song, J.S., and Reiter, J.F. (2012). Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. PLoS genetics *8*, e1002576.

Ingham, P.W. (1985). A clonal analysis of the requirement for the trithorax gene in the diversification of segments in Drosophila. Journal of embryology and experimental morphology *89*, 349-365.

Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science *334*, 194.

Inoue, H., Nojima, H., and Okayama, H. (1990). High efficiency transformation of Escherichia coli with plasmids. Gene *96*, 23-28.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Molecular & cellular proteomics : MCP *4*, 1265-1272.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nature cell biology *10*, 353-360.

Jorgensen, H.F., Giadrossi, S., Casanova, M., Endoh, M., Koseki, H., Brockdorff, N., and Fisher, A.G. (2006). Stem cells primed for action: polycomb repressive complexes restrain the expression of lineage-specific regulators in embryonic stem cells. Cell Cycle *5*, 1411-1414.

Jung, J., Kim, T.G., Lyons, G.E., Kim, H.R., and Lee, Y. (2005). Jumonji regulates cardiomyocyte proliferation via interaction with retinoblastoma protein. The Journal of biological chemistry *280*, 30916-30923.

Kalantry, S., and Magnuson, T. (2006). The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. PLoS genetics *2*, e66.

Kalantry, S., Mills, K.C., Yee, D., Otte, A.P., Panning, B., and Magnuson, T. (2006). The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. Nature cell biology *8*, 195-202.

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 genetics *5*, e1000506.

Kaneko, S., Li, G., Son, J., Xu, C.F., Margueron, R., Neubert, T.A., and Reinberg, D. (2010). Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. Genes & development *24*, 2615-2620.

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., Pombo, A., Fisher, A.G., Young, R.A., and Jenner, R.G. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Molecular cell *38*, 675-688.

Karakuzu, O., Wang, D.P., and Cameron, S. (2009). MIG-32 and SPAT-3A are PRC1 homologs that control neuronal migration in Caenorhabditis elegans. Development *136*, 943-953.

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.

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., Boone, C., Emili, A., Weissman, J.S., Hughes, T.R., Strahl, B.D., Grunstein, M., Greenblatt, J.F., Buratowski, S., and Krogan, N.J. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

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

Kim, H., Kang, K., Ekram, M.B., Roh, T.Y., and Kim, J. (2011). Aebp2 as an epigenetic regulator for neural crest cells. Plos One *6*, e25174.

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., Guermah, M., McGinty, R.K., Lee, J.S., Tang, Z., Milne, T.A., Shilatifard, A., Muir, T.W., and Roeder, R.G. (2009b). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell *137*, 459-471.

Kim, T.G., Chen, J., Sadoshima, J., and Lee, Y. (2004). Jumonji represses atrial natriuretic factor gene expression by inhibiting transcriptional activities of cardiac transcription factors. Molecular and cellular biology *24*, 10151-10160.

Kim, T.G., Jung, J., Mysliwiec, M.R., Kang, S., and Lee, Y. (2005). Jumonji represses alpha-cardiac myosin heavy chain expression via inhibiting MEF2 activity. Biochemical and biophysical research communications *329*, 544-553.

Kim, Y.M., Lee, J.Y., Xia, L., Mulvihill, J.J., and Li, S. (2013). Trisomy 8: a common finding in mouse embryonic stem (ES) cell lines. Molecular cytogenetics *6*, 3.

Kitajima, K., Kojima, M., Nakajima, K., Kondo, S., Hara, T., Miyajima, A., and Takeuchi, T. (1999). Definitive but not primitive hematopoiesis is impaired in jumonji mutant mice. Blood *93*, 87-95.

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

Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling-an emerging perspective on targeting polycomb repressor proteins. PLoS genetics *9*, e1003717.

Klug, A. (2010). The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. Annual Review of Biochemistry, Vol 79 *79*, 213-231.

Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Muller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes & development *20*, 1110-1122.

Knutson, S.K., Wigle, T.J., Warholic, N.M., Sneeringer, C.J., Allain, C.J., Klaus, C.R., Sacks, J.D., Raimondi, A., Majer, C.R., Song, J., Scott, M.P., Jin, L., Smith, J.J., Olhava, E.J., Chesworth, R., Moyer, M.P., Richon, V.M., Copeland, R.A., Keilhack, H., Pollock, R.M., and Kuntz, K.W. (2012). A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. Nat Chem Biol *8*, 890-896.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS biology *2*, 991-1003.

Kohoutek, J., Li, Q., Blazek, D., Luo, Z., Jiang, H., and Peterlin, B.M. (2009). Cyclin T2 is essential for mouse embryogenesis. Molecular and cellular biology *29*, 3280-3285.

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

Krejci, J., Uhlirova, R., Galiova, G., Kozubek, S., Smigova, J., and Bartova, E. (2009). Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation. Journal of cellular physiology *219*, 677-687.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., Adli, M., Kasif, S., Ptaszek, L.M., Cowan, C.A., Lander, E.S., Koseki, H., and Bernstein, B.E. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS genetics *4*, e1000242.

Kuehn, M.R., Bradley, A., Robertson, E.J., and Evans, M.J. (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature *326*, 295-298.

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 & development *16*, 2893-2905.

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

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.

Laible, G., Wolf, A., Dorn, R., Reuter, G., Nislow, C., Lebersorger, A., Popkin, D., Pillus, L., and Jenuwein, T. (1997). Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at S. cerevisiae telomeres. The EMBO journal *16*, 3219-3232.

Laity, J.H., Dyson, H.J., and Wright, P.E. (2000). DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers. Journal of molecular biology *295*, 719-727.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., Brookes, E., Pombo, A., Fisher, C., Skarnes, W.C., Snoek, T., Bezstarosti, K., Demmers, J., Klose, R.J., Casanova, M., Tavares, L., Brockdorff, N., Merkenschlager, M., and Fisher, A.G. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nature cell biology *12*, 618-624.

Landeira, D., and Fisher, A.G. (2011). Inactive yet indispensable: the tale of Jarid2. Trends in cell biology *21*, 74-80.

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 Biol *10*, R25.

Lawson, C.L., Swigon, D., Murakami, K.S., Darst, S.A., Berman, H.M., and Ebright, R.H. (2004). Catabolite activator protein: DNA binding and transcription activation. Curr Opin Struc Biol *14*, 10-20.

Leahy, A., Xiong, J.W., Kuhnert, F., and Stuhlmann, H. (1999). Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation. The Journal of experimental zoology *284*, 67-81.

Lee, J.T., Strauss, W.M., Dausman, J.A., and Jaenisch, R. (1996). A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell *86*, 83-94.

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

Lee, S., Lee, D.K., Dou, Y., Lee, J., Lee, B., Kwak, E., Kong, Y.Y., Lee, S.K., Roeder, R.G., and Lee, J.W. (2006a). Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. Proceedings of the National Academy of Sciences of the United States of America *103*, 15392-15397.

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., Koseki, H., Fuchikami, T., Abe, K., Murray, H.L., Zucker, J.P., Yuan, B., Bell, G.W., Herbolsheimer, E., Hannett, N.M., Sun, K., Odom, D.T., Otte, A.P., Volkert, T.L., Bartel, D.P., Melton, D.A., Gifford, D.K., Jaenisch, R., and Young, R.A. (2006b). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

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. The Journal of cell biology *178*, 219-229.

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 & development *24*, 265-276.

Lehmann, L., Ferrari, R., Vashisht, A.A., Wohlschlegel, J.A., Kurdistani, S.K., and Carey, M. (2012). Polycomb repressive complex 1 (PRC1) disassembles RNA

polymerase II preinitiation complexes. The Journal of biological chemistry *287*, 35784-35794.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current biology : CB *13*, 1192-1200.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565-570.

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

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Current opinion in genetics & development *21*, 175-186.

Li, H.B., Muller, M., Bahechar, I.A., Kyrchanova, O., Ohno, K., Georgiev, P., and Pirrotta, V. (2011a). Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in Drosophila melanogaster. Molecular and cellular biology *31*, 616-625.

Li, H.B., Ohno, K., Gui, H., and Pirrotta, V. (2013). Insulators target active genes to transcription factories and polycomb-repressed genes to polycomb bodies. PLoS genetics *9*, e1003436.

Li, X.Z., Isono, K., Yamada, D., Endo, T.A., Endoh, M., Shinga, J., Mizutani-Koseki, Y., Otte, A.P., Casanova, M., Kitamura, H., Kamijo, T., Sharif, J., Ohara, O., Toyada, T., Bernstein, B.E., Brockdorff, N., and Koseki, H. (2011b). Mammalian Polycomb-Like Pcl2/Mtf2 Is a Novel Regulatory Component of PRC2 That Can Differentially Modulate Polycomb Activity both at the Hox Gene Cluster and at Cdkn2a Genes. Molecular and cellular biology *31*, 351-364.

Lindeman, L.C., Andersen, I.S., Reiner, A.H., Li, N., Aanes, H., Ostrup, O., Winata, C., Mathavan, S., Muller, F., Alestrom, P., and Collas, P. (2011). Prepatterning of Developmental Gene Expression by Modified Histones before Zygotic Genome Activation. Developmental cell *21*, 993-1004.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y., Pape, U.J., Poidinger, M., Chen, Y., Yeung, K., Brown, M., Turpaz, Y., and Liu, X.S. (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*, R83.

Liu, Y., and Montell, D.J. (2001). Jing: a downstream target of slbo required for developmental control of border cell migration. Development *128*, 321-330.

Llano, E., Gomez, R., Gutierrez-Caballero, C., Herran, Y., Sanchez-Martin, M., Vazquez-Quinones, L., Hernandez, T., de Alava, E., Cuadrado, A., Barbero, J.L., Suja, J.A., and Pendas, A.M. (2008). Shugoshin-2 is essential for the completion of meiosis but not for mitotic cell division in mice. Genes & development *22*, 2400-2413.

Lo, S.M., Ahuja, N.K., and Francis, N.J. (2009). Polycomb group protein Suppressor 2 of zeste is a functional homolog of Posterior Sex Combs. Molecular and cellular biology *29*, 515-525.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K.Y., Sung, K.W., Lee, C.W., Zhao, X.D., Chiu, K.P., Lipovich, L., Kuznetsov, V.A., Robson, P., Stanton, L.W., Wei, C.L., Ruan, Y., Lim, B., and Ng, H.H. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature genetics *38*, 431-440.

Long, H.K., Sims, D., Heger, A., Blackledge, N.P., Kutter, C., Wright, M.L., Grutzner, F., Odom, D.T., Patient, R., Ponting, C.P., and Klose, R.J. (2013). Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. eLife *2*, e00348.

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 angstrom resolution. Nature *389*, 251-260.

Luis, N.M., Morey, L., Di Croce, L., and Benitah, S.A. (2012). Polycomb in stem cells: PRC1 branches out. Cell Stem Cell *11*, 16-21.

Lukas, J., Lukas, C., and Bartek, J. (2011). More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. Nature cell biology *13*, 1161-1169.

Lynch, M.D., Smith, A.J., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., Meek, S., Burdon, T., Gibbons, R.J., Garrick, D., and Higgs, D.R. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. The EMBO journal *31*, 317-329.

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372-373.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., Plath, K., and Hochedlinger, K. (2007). Directly

reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55-70.

Mak, T.W. (2007). Gene targeting in embryonic stem cells scores a knockout in Stockholm. Cell *131*, 1027-1031.

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. Current biology : CB *12*, 1016-1020.

Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nature structural biology *10*, 882-891.

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. Molecular cell *32*, 503-518.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., Pirrotta, V., Reinberg, D., and Gamblin, S.J. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343-349.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590-604.

Martin, G.R., Epstein, C.J., Travis, B., Tucker, G., Yatziv, S., Martin, D.W., Clift, S., and Cohen, S. (1978). X-Chromosome Inactivation during Differentiation of Female Teratocarcinoma Stem-Cells Invitro. Nature *271*, 329-333.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 7634-7638.

McCabe, M.T., Ott, H.M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G.S., Liu, Y., Graves, A.P., Della Pietra, A., Diaz, E., LaFrance, L.V., Mellinger, M., Duquenne, C., Tian, X.R., Kruger, R.G., McHugh, C.F., Brandt, M., Miller, W.H., Dhanak, D., Verma, S.K., Tummino, P.J., and Creasy, C.L. (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature *492*, 108-+.

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

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

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. Journal of Cell Biology *147*, 1399-1408.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Developmental cell *10*, 105-116.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S., and Bernstein, B.E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Miller, J., Mclachlan, A.D., and Klug, A. (1985). Repetitive Zinc-Binding Domains in the Protein Transcription Factor Iiia from Xenopus Oocytes. Embo Journal *4*, 1609-1614.

Milne, T.A., Sinclair, D.A., and Brock, H.W. (1999). The Additional sex combs gene of Drosophila is required for activation and repression of homeotic loci, and interacts specifically with Polycomb and super sex combs. Molecular & general genetics : MGG *261*, 753-761.

Milne, T.A., Kim, J., Wang, G.G., Stadler, S.C., Basrur, V., Whitcomb, S.J., Wang, Z., Ruthenburg, A.J., Elenitoba-Johnson, K.S., Roeder, R.G., and Allis, C.D. (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. Molecular cell *38*, 853-863.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes & development *25*, 742-754.

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 & development *17*, 1823-1828.

Mohd-Sarip, A., van der Knaap, J.A., Wyman, C., Kanaar, R., Schedl, P., and Verrijzer, C.P. (2006). Architecture of a polycomb nucleoprotein complex. Molecular cell *24*, 91-100.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Molecular cell *30*, 755-766.

Molitor, A., and Shen, W.H. (2013). The polycomb complex PRC1: composition and function in plants. Journal of genetics and genomics = Yi chuan xue bao *40*, 231-238.

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. Current biology : CB *15*, 942-947.

Montgomery, N.D., Yee, D., Montgomery, S.A., and Magnuson, T. (2007). Molecular and functional mapping of EED motifs required for PRC2-dependent histone methylation. Journal of molecular biology *374*, 1145-1157.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di Croce, L. (2012). Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell *10*, 47-62.

Morey, L., Aloia, L., Cozzuto, L., Benitah, S.A., and Di Croce, L. (2013). RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. Cell reports *3*, 60-69.

Morgan, H.D., Sutherland, H.G., Martin, D.I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nature genetics *23*, 314-318.

Morin-Kensicki, E.M., Faust, C., LaMantia, C., and Magnuson, T. (2001). Cell and tissue requirements for the gene eed during mouse gastrulation and organogenesis. Genesis *31*, 142-146.

Morozova, T., Hackett, J., Sedaghat, Y., and Sonnenfeld, M. (2010). The Drosophila jing gene is a downstream target in the Trachealess/Tango tracheal pathway. Dev Genes Evol *220*, 191-206.

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.

Muntean, A.G., Tan, J., Sitwala, K., Huang, Y., Bronstein, J., Connelly, J.A., Basrur, V., Elenitoba-Johnson, K.S., and Hess, J.L. (2010). The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. Cancer cell *17*, 609-621.

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

Musselman, C.A., Avvakumov, N., Watanabe, R., Abraham, C.G., Lalonde, M.E., Hong, Z., Allen, C., Roy, S., Nunez, J.K., Nickoloff, J., Kulesza, C.A., Yasui, A., Cote, J., and Kutateladze, T.G. (2012). Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nature structural & molecular biology *19*, 1266-1272.

Mysliwiec, M.R., Kim, T.G., and Lee, Y. (2007). Characterization of zinc finger protein 496 that interacts with Jumonji/Jarid2. FEBS letters *581*, 2633-2640.

Nakano, T., Kodama, H., and Honjo, T. (1994). Generation of lymphohematopoietic cells from embryonic stem cells in culture. Science *265*, 1098-1101.

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.

Naughton, C., Sproul, D., Hamilton, C., and Gilbert, N. (2010). Analysis of Active and Inactive X Chromosome Architecture Reveals the Independent Organization of 30 nm and Large-Scale Chromatin Structures. Molecular cell *40*, 397-409.

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

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

Neri, F., Zippo, A., Krepelova, A., Cherubini, A., Rocchigiani, M., and Oliviero, S. (2012). Myc regulates the transcription of the PRC2 gene to control the expression of developmental genes in embryonic stem cells. Molecular and cellular biology *32*, 840-851.

Ng, S.Y., Johnson, R., and Stanton, L.W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. The EMBO journal *31*, 522-533.

Nicol, J.W., Helt, G.A., Blanchard, S.G., Jr., Raja, A., and Loraine, A.E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics *25*, 2730-2731.

Nishikawa, S.I., Nishikawa, S., Hirashima, M., Matsuyoshi, N., and Kodama, H. (1998). Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. Development *125*, 1747-1757.

Niswander, L., Yee, D., Rinchik, E.M., Russell, L.B., and Magnuson, T. (1988). The albino deletion complex and early postimplantation survival in the mouse. Development *102*, 45-53.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes & development *12*, 2048-2060.

Norris, D.P., Patel, D., Kay, G.F., Penny, G.D., Brockdorff, N., Sheardown, S.A., and Rastan, S. (1994). Evidence that random and imprinted Xist expression is controlled by preemptive methylation. Cell *77*, 41-51.

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. Molecular and cellular biology *21*, 4330-4336.

O'Connell, S., Wang, L., Robert, S., Jones, C.A., Saint, R., and Jones, R.S. (2001). Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein. The Journal of biological chemistry *276*, 43065-43073.

O'Loghlen, A., Munoz-Cabello, A.M., Gaspar-Maia, A., Wu, H.A., Banito, A., Kunowska, N., Racek, T., Pemberton, H.N., Beolchi, P., Lavial, F., Masui, O., Vermeulen, M., Carroll, T., Graumann, J., Heard, E., Dillon, N., Azuara, V., Snijders, A.P., Peters, G., Bernstein, E., and Gil, J. (2012). MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. Cell Stem Cell *10*, 33-46.

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

Ohno, T., Nakajima, K., Kojima, M., Toyoda, M., and Takeuchi, T. (2004). Modifiers of the jumonji mutation downregulate cyclin D1 expression and cardiac cell proliferation. Biochemical and biophysical research communications *317*, 925-929.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*, 644-649.

Oktaba, K., Gutierrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E.M., and Muller, J. (2008). Dynamic Regulation by Polycomb Group Protein Complexes Controls Pattern Formation and the Cell Cycle in Drosophila. Developmental cell *15*, 877-889.

Orkin, S.H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. Cell *145*, 835-850.

Orlando, D.A., Guenther, M.G., Frampton, G.M., and Young, R.A. (2012). CpG island structure and trithorax/polycomb chromatin domains in human cells. Genomics *100*, 320-326.

Palmer, D.K., O'Day, K., and Margolis, R.L. (1990). The centromere specific histone CENP-A is selectively retained in discrete foci in mammalian sperm nuclei. Chromosoma *100*, 32-36.

Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., and Margolis, R.L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proceedings of the National Academy of Sciences of the United States of America *88*, 3734-3738.

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.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes & development *10*, 1991-2002.

Papp, B., and Muller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes & development *20*, 2041-2054.

Parrish, J.Z., Emoto, K., Jan, L.Y., and Jan, Y.N. (2007). Polycomb genes interact with the tumor suppressor genes hippo and warts in the maintenance of Drosophila sensory neuron dendrites. Genes & development *21*, 956-972.

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

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

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

Pastor, W.A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nature reviews Molecular cell biology *14*, 341-356.

Pavletich, N.P., and Pabo, C.O. (1993). Crystal-Structure of a 5-Finger Gli-DNA Complex - New Perspectives on Zinc Fingers. Science *261*, 1701-1707.

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.

Pengelly, A.R., Copur, O., Jackle, H., Herzig, A., and Muller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. Science *339*, 698-699.

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.

Pereira, C.F., Piccolo, F.M., Tsubouchi, T., Sauer, S., Ryan, N.K., Bruno, L., Landeira, D., Santos, J., Banito, A., Gil, J., Koseki, H., Merkenschlager, M., and Fisher, A.G. (2010). ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. Cell Stem Cell *6*, 547-556.

Perez-Burgos, L., Peters, A.H., Opravil, S., Kauer, M., Mechtler, K., and Jenuwein, T. (2004). Generation and characterization of methyl-lysine histone antibodies. Methods in enzymology *376*, 234-254.

Petruk, S., Sedkov, Y., Johnston, Danika M., Hodgson, Jacob W., Black, Kathryn L., Kovermann, Sina K., Beck, S., Canaani, E., Brock, Hugh W., and Mazo, A. (2012). TrxG and PcG Proteins but Not Methylated Histones Remain Associated with DNA through Replication. Cell *150*, 922-933.

Philpott, A., Krude, T., and Laskey, R.A. (2000). Nuclear chaperones. Seminars in cell & developmental biology *11*, 7-14.

Pierce, G.B. (1967). Teratocarcinoma: model for a developmental concept of cancer. Current topics in developmental biology *2*, 223-246.

Pirity, M.K., Locker, J., and Schreiber-Agus, N. (2005). Rybp/DEDAF is required for early postimplantation and for central nervous system development. Molecular and cellular biology *25*, 7193-7202.

Pirrotta, V., and Li, H.B. (2012). A view of nuclear Polycomb bodies. Current opinion in genetics & development *22*, 101-109.

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.

Pullirsch, D., Hartel, R., Kishimoto, H., Leeb, M., Steiner, G., and Wutz, A. (2010). The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. Development *137*, 935-943.

Qin, J., Whyte, W.A., Anderssen, E., Apostolou, E., Chen, H.H., Akbarian, S., Bronson, R.T., Hochedlinger, K., Ramaswamy, S., Young, R.A., and Hock, H. (2012). The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. Cell Stem Cell *11*, 319-332.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

Ramsden, C.M., Powner, M.B., Carr, A.J., Smart, M.J., da Cruz, L., and Coffey, P.J. (2013). Stem cells in retinal regeneration: past, present and future. Development *140*, 2576-2585.

Rastan, S., and Robertson, E.J. (1985). X-Chromosome Deletions in Embryo-Derived (Ek) Cell-Lines Associated with Lack of X-Chromosome Inactivation. Journal of embryology and experimental morphology *90*, 379-388.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Remeseiro, S., Cuadrado, A., Carretero, M., Martinez, P., Drosopoulos, W.C., Canamero, M., Schildkraut, C.L., Blasco, M.A., and Losada, A. (2012). Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. The EMBO journal *31*, 2076-2089.

Ren, X.J., and Kerppola, T.K. (2011). REST Interacts with Cbx Proteins and Regulates Polycomb Repressive Complex 1 Occupancy at RE1 Elements. Molecular and cellular biology *31*, 2100-2110.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2012). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. The EMBO journal *31*, 593-605.

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., and Chang, H.Y. (2007). Functional

demarcation of active and silent chromatin domains in human HOX loci by Noncoding RNAs. Cell *129*, 1311-1323.

Robertson, E., Bradley, A., Kuehn, M., and Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature *323*, 445-448.

Robinson, P.J.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D. (2008). 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. Journal of molecular biology *381*, 816-825.

Robinton, D.A., and Daley, G.Q. (2012). The promise of induced pluripotent stem cells in research and therapy. Nature *481*, 295-305.

Rosenfeld, J.A., Wang, Z., Schones, D.E., Zhao, K., DeSalle, R., and Zhang, M.Q. (2009). Determination of enriched histone modifications in non-genic portions of the human genome. BMC genomics *10*, 143.

Rossetto, D., Avvakumov, N., and Cote, J. (2012). Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics : official journal of the DNA Methylation Society *7*, 1098-1108.

Rowbotham, S.P., Barki, L., Neves-Costa, A., Santos, F., Dean, W., Hawkes, N., Choudhary, P., Will, W.R., Webster, J., Oxley, D., Green, C.M., Varga-Weisz, P., and Mermoud, J.E. (2011). Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. Molecular cell *42*, 285-296.

Rugg-Gunn, P.J., Cox, B.J., Ralston, A., and Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proceedings of the National Academy of Sciences of the United States of America *107*, 10783-10790.

Russo, V.E.A., Martienssen, R.A., and Riggs, A.D. (1996). Epigenetic mechanisms of gene regulation (Plainview, N.Y.: Cold Spring Harbor Laboratory Press).

Ryan, R.F., and Darby, M.K. (1998). The role of zinc finger linkers in p43 and TFIIIA binding to 5S rRNA and DNA. Nucleic acids research *26*, 703-709.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome research *20*, 761-770.

Saramaki, A., Diermeier, S., Kellner, R., Laitinen, H., Vaisanen, S., and Carlberg, C. (2009). Cyclical chromatin looping and transcription factor association on the regulatory

regions of the p21 (CDKN1A) gene in response to 1alpha,25-dihydroxyvitamin D3. The Journal of biological chemistry *284*, 8073-8082.

Sarma, K., Nishioka, K., and Reinberg, D. (2004). Tips in analyzing antibodies directed against specific histone tail modifications. Methods in enzymology *376*, 255-269.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., and Reinberg, D. (2008). Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Molecular and cellular biology *28*, 2718-2731.

Savla, U., Benes, J., Zhang, J., and Jones, R.S. (2008). Recruitment of Drosophila Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. Development *135*, 813-817.

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.

Scheuermann, J.C., Gutierrez, L., and Muller, J. (2012). Histone H2A monoubiquitination and Polycomb repression: the missing pieces of the puzzle. Fly *6*, 162-168.

Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods *48*, 240-248.

Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., Bunker, R.D., Wirth, U., Bouwmeester, T., Bauer, A., Ly-Hartig, N., Zhao, K., Chan, H., Gu, J., Gut, H., Fischle, W., Muller, J., and Thoma, N.H. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Molecular cell *42*, 330-341.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. Genes & development *19*, 697-708.

Schnutgen, F., Doerflinger, N., Calleja, C., Wendling, O., Chambon, P., and Ghyselinck, N.B. (2003). A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. Nature biotechnology *21*, 562-565.

Schnutgen, F., De-Zolt, S., Van Sloun, P., Hollatz, M., Floss, T., Hansen, J., Altschmied, J., Seisenberger, C., Ghyselinck, N.B., Ruiz, P., Chambon, P., Wurst, W., and von Melchner, H. (2005). Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. Proceedings of the National Academy of Sciences of the United States of America *102*, 7221-7226.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. The EMBO journal *25*, 3110-3122.

Schorderet, P., and Duboule, D. (2011). Structural and Functional Differences in the Long Non-Coding RNA <italic>Hotair</italic> in Mouse and Human. PLoS genetics *7*, e1002071.

Schuettengruber, B., and Cavalli, G. (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development *136*, 3531-3542.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional Anatomy of Polycomb and Trithorax Chromatin Landscapes in Drosophila Embryos. PLoS biology *7*, 146-163.

Schuettengruber, B., Martinez, A.M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nature reviews Molecular cell biology *12*, 799-814.

Schumacher, A., Faust, C., and Magnuson, T. (1996). Positional cloning of a global regulator of anterior-posterior patterning in mice. Nature *384*, 648.

Schwaiger, M., Stadler, M.B., Bell, O., Kohler, H., Oakeley, E.J., and Schubeler, D. (2009). Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome. Genes & development *23*, 589-601.

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. Nature genetics *38*, 700-705.

Sedaghat, Y., Miranda, W.F., and Sonnenfeld, M.J. (2002). The jing Zn-finger transcription factor is a mediator of cellular differentiation in the Drosophila CNS midline and trachea. Development *129*, 2591-2606.

Sedaghat, Y., and Sonnenfeld, M. (2002). The rjing gene is required for embryonic brain development in Drosophila. Development Genes and Evolution *212*, 277-287.

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. Molecular and cellular biology *18*, 3586-3595.

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.

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. Molecular cell *32*, 491-502.

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.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annual review of biochemistry *81*, 65-95.

Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: cis-regulatory element annotation system. Bioinformatics *25*, 2605-2606.

Shirato, H., Ogawa, S., Nakajima, K., Inagawa, M., Kojima, M., Tachibana, M., Shinkai, Y., and Takeuchi, T. (2009). A jumonji (Jarid2) protein complex represses cyclin D1 expression by methylation of histone H3-K9. The Journal of biological chemistry *284*, 733-739.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., 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. Developmental cell *4*, 481-495.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of Polycomb gene silencing: knowns and unknowns. Nature Reviews Molecular Cell Biology *10*, 697-708.

Simon, J.A., and Kingston, R.E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Molecular cell *49*, 808-824.

Sims, R.J., 3rd, and Reinberg, D. (2006). Histone H3 Lys 4 methylation: caught in a bind? Genes & development *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.

Skene, P.J., and Henikoff, S. (2013). Histone variants in pluripotency and disease. Development *140*, 2513-2524.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688-690.

Smith, A.G., Nichols, J., Robertson, M., and Rathjen, P.D. (1992). Differentiation inhibiting activity (DIA/LIF) and mouse development. Developmental biology *151*, 339-351.

Smith, K.N., Singh, A.M., and Dalton, S. (2010). Myc represses primitive endoderm differentiation in pluripotent stem cells. Cell Stem Cell *7*, 343-354.

Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A., and Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic acids research *41*, e28.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. Nature reviews Cancer *6*, 846-856.

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 research *16*, 890-900.

Srinivasan, S., Dorighi, K.M., and Tamkun, J.W. (2008). Drosophila Kismet regulates histone H3 lysine 27 methylation and early elongation by RNA polymerase II. PLoS genetics *4*, e1000217.

Steffen, P.A., Fonseca, J.P., Ganger, C., Dworschak, E., Kockmann, T., Beisel, C., and Ringrose, L. (2013). Quantitative in vivo analysis of chromatin binding of Polycomb and Trithorax group proteins reveals retention of ASH1 on mitotic chromatin. Nucleic acids research *41*, 5235-5250.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiology and molecular biology reviews : MMBR *64*, 435-459.

Stevens, L.C. (1967). Origin of testicular teratomas from primordial germ cells in mice. Journal of the National Cancer Institute *38*, 549-552.

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. Nature cell biology *9*, 1428-1435.

Suzuki, M., Mizutani-Koseki, Y., Fujimura, Y., Miyagishima, H., Kaneko, T., Takada, Y., Akasaka, T., Tanzawa, H., Takihara, Y., Nakano, M., Masumoto, H., Vidal, M., Isono, K., and Koseki, H. (2002). Involvement of the Polycomb-group gene Ring1B in the specification of the anterior-posterior axis in mice. Development *129*, 4171-4183.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Takagi, N., and Martin, G.R. (1984). Studies of the Temporal Relationship between the Cytogenetic and Biochemical Manifestations of X-Chromosome Inactivation during the Differentiation of Lt-1 Teratocarcinoma Stem-Cells. Developmental biology *103*, 425-433.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., and Higashinakagawa, T. (1995). Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. Genes & development *9*, 1211-1222.

Takeuchi, T., Watanabe, Y., Takano-Shimizu, T., and Kondo, S. (2006). Roles of jumonji and jumonji family genes in chromatin regulation and development. Developmental dynamics : an official publication of the American Association of Anatomists *235*, 2449-2459.

Takihara, Y., Tomotsune, D., Shirai, M., Katoh-Fukui, Y., Nishii, K., Motaleb, M.A., Nomura, M., Tsuchiya, R., Fujita, Y., Shibata, Y., Higashinakagawa, T., and Shimada, K. (1997). Targeted disruption of the mouse homologue of the Drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. Development *124*, 3673-3682.

Tan-Wong, S.M., French, J.D., Proudfoot, N.J., and Brown, M.A. (2008). Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. Proceedings of the National Academy of Sciences of the United States of America *105*, 5160-5165.

Tanaka, Y., Katagiri, Z., Kawahashi, K., Kioussis, D., and Kitajima, S. (2007). Trithorax-group protein ASH1 methylates histone H3 lysine 36. Gene *397*, 161-168.

Tanaka, Y., Kawahashi, K., Katagiri, Z., Nakayama, Y., Mahajan, M., and Kioussis, D. (2011). Dual function of histone H3 lysine 36 methyltransferase ASH1 in regulation of Hox gene expression. Plos One *6*, e28171.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., Wutz, A., Vidal, M., Elderkin, S., and Brockdorff, N. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664-678.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

Thomas, J.O. (1999). Histone H1: location and role. Current opinion in cell biology *11*, 312-317.

Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.

Thompson, J.R., and Gudas, L.J. (2002). Retinoic acid induces parietal endoderm but not primitive endoderm and visceral endoderm differentiation in F9 teratocarcinoma stem cells with a targeted deletion of the Rex-1 (Zfp-42) gene. Molecular and cellular endocrinology *195*, 119-133.

Tie, F., Prasad-Sinha, J., Birve, A., Rasmuson-Lestander, A., and Harte, P.J. (2003). A 1-megadalton ESC/E(Z) complex from Drosophila that contains polycomblike and RPD3. Molecular and cellular biology *23*, 3352-3362.

Tie, F., Stratton, C.A., Kurzhals, R.L., and Harte, P.J. (2007). The N terminus of Drosophila ESC binds directly to histone H3 and is required for E(Z)-dependent trimethylation of H3 lysine 27. Molecular and cellular biology *27*, 2014-2026.

Tolhuis, B., Blom, M., Kerkhoven, R.M., Pagie, L., Teunissen, H., Nieuwland, M., Simonis, M., de Laat, W., van Lohuizen, M., and van Steensel, B. (2011). Interactions among Polycomb domains are guided by chromosome architecture. PLoS genetics *7*, e1001343.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909-918.

Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Buhler, M., Margueron, R., and Schneider, R. (2013). Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. Cell *152*, 859-872.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. The EMBO journal *13*, 3822-3831.

Tsukada, Y., 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.

Umezawa, Y., Shimada, T., Kori, A., Yamada, K., and Ishihama, A. (2008). The uncharacterized transcription factor YdhM is the regulator of the nemA gene, encoding N-Ethylmaleimide reductase. J Bacteriol *190*, 5890-5897.

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., van Lohuizen, M., and et al. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. Genes & development *8*, 757-769.

van der Stoop, P., Boutsma, E.A., Hulsman, D., Noback, S., Heimerikx, M., Kerkhoven, R.M., Voncken, J.W., Wessels, L.F., and van Lohuizen, M. (2008). Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells. Plos One *3*, e2235.

Van Holde, K.E. (1989). Chromatin (New York ; London: Springer-Verlag).

Varambally, S., Cao, Q., Mani, R.S., Shankar, S., Wang, X., Ateeq, B., Laxman, B., Cao, X., Jing, X., Ramnarayanan, K., Brenner, J.C., Yu, J., Kim, J.H., Han, B., Tan, P., Kumar-Sinha, C., Lonigro, R.J., Palanisamy, N., Maher, C.A., and Chinnaiyan, A.M. (2008). Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science *322*, 1695-1699.

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922-U143.

Vernimmen, D., Lynch, M.D., De Gobbi, M., Garrick, D., Sharpe, J.A., Sloane-Stanley, J.A., Smith, A.J.H., and Higgs, D.R. (2011). Polycomb eviction as a new distant enhancer function. Genes & development *25*, 1583-1588.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181-193.

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

Voncken, J.W., Roelen, B.A., 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. Proceedings of the National Academy of Sciences of the United States of America *100*, 2468-2473.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Molecular and cellular biology *20*, 2108-2121.

Waddington, C.H. (1957). The strategy of the genes : a discussion of some aspects of theoretical biology (London: Allen & Unwin).

Waddington, C.H. (2012). The epigenotype. 1942. International journal of epidemiology *41*, 10-13.

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, J., Mager, J., Chen, Y., Schneider, E., Cross, J.C., Nagy, A., and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse Polycomb group gene. Nature genetics *28*, 371-375.

Wang, J., Mager, J., Schnedier, E., and Magnuson, T. (2002). The mouse PcG gene eed is required for Hox gene repression and extraembryonic development. Mammalian genome : official journal of the International Mammalian Genome Society *13*, 493-503.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004b). Hierarchical recruitment of polycomb group silencing complexes. Molecular cell *14*, 637-646.

Wang, L., Zeng, X.Z., Chen, S., Ding, L.Y., Zhong, J., Zhao, J.C., Wang, L.G., Sarver, A., Koller, A., Zhi, J.Z., Ma, Y.P., Yu, J.D., Chen, J.J., and Huang, H.J. (2013). BRCA1 is a negative modulator of the PRC2 complex. Embo Journal *32*, 1584-1597.

Watanabe, S., Radman-Livaja, M., Rando, O.J., and Peterson, C.L. (2013). A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. Science *340*, 195-199.

Weake, V.M., and Workman, J.L. (2008). Histone ubiquitination: triggering gene activity. Molecular cell *29*, 653-663.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nature genetics *41*, 246-250.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid-Leukemia Inhibitory Factor Maintains the Developmental Potential of Embryonic Stem-Cells. Nature *336*, 684-687.

Wolfe, S.A., Nekludova, L., and Pabo, C.O. (2000). DNA recognition by Cys(2)His(2) zinc finger proteins. Annu Rev Bioph Biom *29*, 183-212.

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.

Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J., and Kingston, R.E. (2013). Variable Requirements for DNA-Binding Proteins at Polycomb-Dependent Repressive Regions in Human HOX Clusters. Molecular and cellular biology *33*, 3274-3285.

Workman, J.L., and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annual review of biochemistry *67*, 545-579.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-393.

Wu, M., Wang, P.F., Lee, J.S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Molecular and cellular biology *28*, 7337-7344.

Wu, X., Johansen, J.V., and Helin, K. (2013). Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. Molecular cell *49*, 1134-1146.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Molecular cell *5*, 695-705.

Wutz, A. (2011). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nature reviews Genetics *12*, 542-553.

Wutz, A. (2013). Epigenetic regulation of stem cells : the role of chromatin in cell differentiation. Advances in experimental medicine and biology *786*, 307-328.

Yamamoto, K., Sonoda, M., Inokuchi, J., Shirasawa, S., and Sasazuki, T. (2004). Polycomb group suppressor of zeste 12 links heterochromatin protein 1alpha and enhancer of zeste 2. The Journal of biological chemistry *279*, 401-406.

Yan, J., Zierath, J.R., and Barres, R. (2011). Evidence for non-CpG methylation in mammals. Experimental cell research *317*, 2555-2561.

Ye, T., Krebs, A.R., Choukrallah, M.A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic acids research *39*, e35.

Young, N.L., DiMaggio, P.A., Plazas-Mayorca, M.D., Baliban, R.C., Floudas, C.A., and Garcia, B.A. (2009). High throughput characterization of combinatorial histone codes. Molecular & cellular proteomics : MCP *8*, 2266-2284.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940-954.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., Slukvin, II, and Thomson, J.A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Yu, T.X., Chen, X., Zhang, W., Colon, D., Shi, J.D., Napier, D., Rychahou, P., Lu, W.G., Lee, E.Y., Weiss, H.L., Evers, B.M., and Liu, C.M. (2012). Regulation of the Potential Marker for Intestinal Cells, Bmi1, by beta-Catenin and the Zinc Finger Protein KLF4 IMPLICATIONS FOR COLON CANCER. Journal of Biological Chemistry *287*, 3760-3768.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. The Journal of biological chemistry *286*, 7983-7989.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & development *25*, 2227-2241.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhang, Z., Jones, A., Sun, C.W., Li, C., Chang, C.W., Joo, H.Y., Dai, Q., Mysliwiec, M.R., Wu, L.C., Guo, Y., Yang, W., Liu, K., Pawlik, K.M., Erdjument-Bromage, H., Tempst, P., Lee, Y., Min, J., Townes, T.M., and Wang, H. (2011). PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. Stem Cells *29*, 229-240.

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

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. Molecular cell *40*, 939-953.

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.K., Shahab, A., Kuznetsov, V.A., Bourque, G., Oh, S., Ruan, Y., Ng, H.H., and Wei, C.L. (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell *1*, 286-298.

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. Molecular cell *29*, 69-80.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., Bennett, D.A., Houmard, J.A., Muoio, D.M., Onder, T.T., Camahort, R., Cowan, C.A., Meissner, A., Epstein, C.B., Shoresh, N., and Bernstein, B.E. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642-654.