Involvement of short RNAs in polycomb-mediated gene repression

Carla Cristina
Gonçalves de Araújo

Submitted to University College London for the degree of Master of Philosophy

September 2014

UCL Cancer Institute
Division of Infection and Immunity
Paul O'Gorman Building
University College London

I, Carla Cristina Gonçalves de Araújo, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Carla Cristina Gonçalves de Araújo
29/9/2014

Abstract

Polycomb group proteins maintain cell identity by repressing developmental regulator genes specific for other cell types. There are two main complexes: Polycomb repressive complex 1 (PRC1) and 2 (PRC2). PRC2 methylates histone H3 lysine 27 (H3K27me3), creating a binding site for PRC1 that ubiquitinates H2AK119. Polycomb target genes are associated with stalled RNA polymerase II (RNAPII), and the initiation marker H3K4me3, known as bivalent chromatin. Our laboratory has demonstrated that short RNAs are transcribed from the promoter region of these genes in human T-cells, while the work carried out as part of the present thesis demonstrates that short RNAs are also transcribed in murine embryonic stem cells (ESCs). This indicates that they are conserved across different species and cell types. Northern blotting for RNAs ≤200 nucleotides extracted from murine ES cell deficient for PRC2 and PRC1 revealed that short RNA production is independent of Polycomb activity. When cells differentiate and Polycomb-target genes become activated, short RNAs are depleted. Given that PRC2 interacts with RNA, this loss of short RNAs might allow gene activation. Additionally, polycomb response elements (PRE) have been detected in *Drosophila*. These elements are necessary and sufficient for polycomb recruitment. A recently identified PRE, HOXD11.12, recruits PRC2 in human mesenchymal stem cells (MSC). It is hypothesized that PRE activity is due to the transcription of short RNAs. Blotting for RNA extracted from MSC identified short RNAs transcribed from D11.12. Moreover, these short RNAs can form the same secondary structure as the previously-identified short RNAs and are also located at a CpG island. Furthermore, RASL12 and YBX2 behave as PREs while D11.12 from active HOXD11 enhances gene expression, potentially also acting as a Trithorax response element (TRE).

Acknowledgements

Above all, I would like to express my sincere gratitude to my supervisor Dr. Richard Jenner for his scientific teachings, advice and guidance throughout this research project and for his knowledge and immense patience, all of which have been priceless. I am also very grateful to my tutor Professor Benny Chain for listening to me and guiding me in difficult times. I also thank my secondary supervisor Professor Greg Towers for his opinions. And I would like to acknowledge the Medical Research Council (MRC) for the funding.

My appreciation also goes to the other members of this laboratory, in particular to Aditi S. Kanhere and Russell D. Bowman for their definite support at different levels and great technical help. I also thank the Amanda G. Fisher's and Richard A. Young's laboratories for their indispensible collaboration. As well as Juan Funes for providing me with tumorigenic mesenchymal stem cells.

Finally, I want to thank my family, my friend Filipe Oliveira as well as the professional support of the Day Hospital (Dr Luísa, Dr Carla, Dr Silvie, and nurses) for being with me through hard times and giving me encouragement to finalize my studies.

Contents

Figures

Tables

Abbreviations

ANT-C	Antennapedia Complex
4C	Chromosome Conformation Capture on Chip
A	Adenine
AEBP2	Adipocyte Enhancer Binding Protein 2
ANRIL	Antisense Non-coding RNA in the INK4 Locus
Arg or R	Arginine
ARID	AT-Rich Interactive Domain
ASH1	Absent, small, or homeotic discs 1
Ash1	Absent, small, or homeotic discs 1
ASXL1	Additional sex comb-like 1
BAP1	BRCA1-associated protein-1
BMI1	B-cell-specific Moloney murine leukemia virus integration site 1
bp	Base Pairs
BRG1	Brahma-related gene 1
BRM	Brahma
BX-C	Bithorax Complex
bxd	Bithoraxoid
C	Cytosine
CBP	CREB-binding protein
CBX	Polycomb Chromobox Protein
ChIP	Chromatin Immunoprecipitation
cKrox/Th-POK	Krüppel-related zinc-finger protein/T-helper-inducing POZ/Krüppel-like factor
COMPASS	Complex proteins associated with Set1
CTCF	CCCTC-binding factor
CTD	Carboxy-Terminal Domain
DNMTs	DNA methyltransferases
dNTP	Deoxyribonucleic triphosphate
dRAF	Drosophila RING-associated factors
dSfmbt	SCM-related gene containing four MBT domains
Dsp1	Dorsal switch protein 1
dsRNA	Double stranded RNA
DZNep	S-adenosylhomocysteine hydrolase inhibitor 3-Deazaneplanocin A
EED	Embryonic Ectoderm Development
EZH2	Enhancer of Zeste Homologue 2
G	Guanine
GAF	GAGA Factor
GH	Grainyhead
GRO-seq	Global ru-on-sequencing
GTF	General transcription factor
H	Histones
HCNEs	Highly Conserved Noncoding Elements
HDAC	Histone Deacetylases
HMT	Histone Methyltransferase

HOTAIR	HOX antisense intergenic RNA
HOX	Homeobox
HSC	Hematopoietic Stem Cell
Hsp	Heat Shock Promoter
ICM	Inner cell mass
iPSC	Induced pluripotente stem cell
ISWI	Imitation switch
JARID2	Juminji AT-Rich Interactive Domain 2
jmj	Juminji
kb	Kilobase
Kcnq1	Potassium voltagegated channel, subfamily Q, member 1
Kcnq1ot1	Kcnq1 overlapping transcript 1
kr	Kreisler
lincRNA	Long intergenic ncRNA
Lys or K	Lysines
Mbd3	methyl CpG-binding domain protein 3
MBLR	Mel18 and Bmi1-like ring finger
MLL	Mixed-lineage Leukemia
mRNA	Messenger RNA
MSC	Mesenchymal Stem Cell
NCP	Nucleosome core particle
ncRNA	Non-coding RNA
NPC	Neural Progenitor Cells
NSPC1	Nervous System Pc1
nt	Nucleotide
Pc	Polycomb
PcG	Polycomb group Proteins
PCGFs	PcG ring fingers
Ph	Polyhomeotic
PHC	Polyhomeotic-like protein
Pho	Pleiohomeotic
Phol	Pleiohomeotic-like
PhoRC	Pho-Repressive Complex
PIC	Preinitiation complex
pp	Primer-primer complementary sequence
PR-DUB	Polycomb Repressive Deubiquitinase
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PRE	Polycomb responsive elements
pre-miRNA	Precursor micro RNA
pre-rRNA	Precursor ribosomal RNA
pri-miRNA	Primary micro RNA
Psc	Posterior sex comb
Psq	Pipsqueak
RbAp46/48	Retinoblastoma-Associated Proteins 46 and 48
RepA	A-repeat region
RIP	RNA immuneprecipitation
RNAPI	RNA polymerase I

RNAPII	RNA Polymerase II
RNAPIII	RNA polymerase III
rRNA	Ribosomal RNA
RYBP	RING1 and YY1 Binding Protein
S2P	Ser2 Phosphorylated
S5P	Ser5 Phosphorylated
seq	Sequencing
Ser or S	Serine
SETDB1	SET domain, bifurcated 1
shRNA	short hairpin RNA
SMARCB1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1
snRNPs	Small nuclear ribonucleoproteins.
SP1	Specificity protein 1
SS	Splice site
ssRNA	Single stranded RNA
SUZ12	Suppressor of Zeste 12
SWI/SNF	SWItch/Sucrose NonFermentable
T	Thymine
TFs	Transcription Factors
TRE	Trithorax response element
tRNA	Transfer RNA
TRRAP	transformation/transcription domain-associated protein
TrxG	Trithorax group
TSS	Transcription-Starting Site
Ubx	Ultrabothorax
WCE	Whole cell extract
WDR5	WD repeat-containing protein 5
Xi	Inactive X chromosome
XIST	X-inactive-specific transcript
YAF2	YY1 Associated Factor 2
YY1	Yin Yang-1

Chapter 1 – Introduction

During early development of multicellular organisms pluripotent cells differentiate into multiple, distinctly(Koontz et al., 2001) specialized cells. As every cell contains exactly the same genetic information but with different sets of genes being expressed/silenced, the process of cell differentiation requires meticulous regulation of gene expression, particularly those that encode for transcription factors and determine the anterior-posterior axis and segment identity of metazoan organisms. These genes encode proteins with a homeobox domain and are named HOX genes. It is also fundamental that HOX and other developmental genes are maintained in the correct spatial and temporal expression pattern over multiple rounds of mitotic cell division and cell differentiation, which is assured by a set of proteins of the polycomb group. This ensures preservation of cell identity and proper body pattern formation throughout the development.

The study of deregulation in gene orchestration is important, as it is a critical pathway leading to diseases such as developmental disorders and cancer. Understanding of mechanisms leading to such gene silencing or activation is a prerequisite for identifying possible ways of counteracting them therapeutically.

1.1 Embryonic development

Embryonic development starts with cellular division and subsequent and gradual specialization of totipotent and pluripotent cells (cells that can become any or nearly any cell type, respectively) into cells that perform specific functions (Evans and Hunter, 2002). During embryogenesis, differentiating cells go through several stages of compartmentalization for commitment and to perform specific functions. Cellular differentiation involves meticulous regulation and orchestration of expression of

developmental genes, accomplished through communication between neighboring cells and their environment.

Drosophila embryogenesis is followed by three larval stages and pupal stage where metamorphosis happens to form an adult fly. Mammalian embryonic development, on the other hand, progresses in two major stages: embryogenesis, where critical stages for the formation of the embryo happens; and fetal development, where the organs of the embryo grow and mature.

1.1.1 Gene regulation in *Drosophila* embryogenesis

The first set of transcription factors-encoding genes activated on the onset of zygotic transcription is the gap genes, such as the *hunchback* and *Kruppel* genes. These are segmentation genes and are regulated by maternal genes. Together they are responsible for the periodic expression of other segmentation genes, pair-rule genes, such as *paired* gene, and the localized expression of several homoeotic genes (Gutjahr et al., 1993; Stanojevic et al., 1989). During early embryogenesis, segmentation genes establish cellular patterns of gene repression including the Hox genes, which control metamerization (Duncan, 1986; Hodgson et al., 1997; Mulholland et al., 2003; Strutt and Paro, 1997; Tie et al., 2001). Besides metamerization, segmentation genes are also expressed during the development of nervous system (Doe and Scott, 1988; Doe et al., 1988; Patel et al., 1989). Hox genes are homeotic genes that encode for transcription factors (TFs) involved in the development of body structures according to the anterior-posterior specified body plan containing a homeobox DNA sequence. This box encodes for a DNA- or RNA- binding homeodomain. However, not all homeobox-containing genes are homeotic, and some segmentation genes also contain a homeobox sequence (Gutjahr et al., 1993; Slack, 2012). Hox genes start being expressed at the phylotypic

stage and their initial pattern of expression is regulated by the gap and pair-rule genes (Bienz and Muller, 1995). In *Drosophila melanogaster* there are two clusters of Hox genes: the Bithorax Complex (BX-C) and the Antennapedia Complex (ANT-C). These genes are crucial for the anterior-posterior development in embryogenesis (Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2004).

1.1.2 Mammalian embryogenesis

The embryogenesis in mammals has 5 major stages (listed in order of occurrence): fertilization, cleavage, blastulation, gastrulation and organogenesis (Figure 1). The process starts with the fertilization of the egg cell (ovum) by a sperm cell, (spermatozoon) forming the zygote. Cleavage follows this step with multiple divisions of the zygote without growth in volume; thus, each mother cell is divided into daughter cells with half of the size. This stage is under maternal effect as protein expression is provided by maternal messenger RNA. After four cellular divisions a Morula is formed, and at this point, there is the onset of zygotic transcription where the zygote starts expressing its own genes. The process continues with the blastulation, where a layer of cells surrounding a fluid is formed called the bastula, and it is followed by the formation of an inner agglomeration of cells designated ‘inner cell mass’ (ICM) giving rise to a blastocyst (the outer layer of cells is trophoblast), and ends with the implantation of the embryo. In gastrulation there is cell movement from the inner cell mass of the blastocyst to form three germ layers: endoderm, mesoderm, and ectoderm. This process is accompanied by region specification where cells become committed to the formation of an anteriorposterior body plan – phylotypic stage. This gives rise to an embryonic disk with a primitive streak, a thickening of the embryonic disk along the median line from the rostro to half of the embryo. The embryo is then ready for organogenesis. This stage starts with neurulation, and which occurs simultaneously with transversal segmentation

of paraxial mesoderm. The segmentation gives origin to different body parts from which structures derive. There is primitive formation of heart, and then other structures follow. (Gutjahr et al., 1993; Slack, 2012).

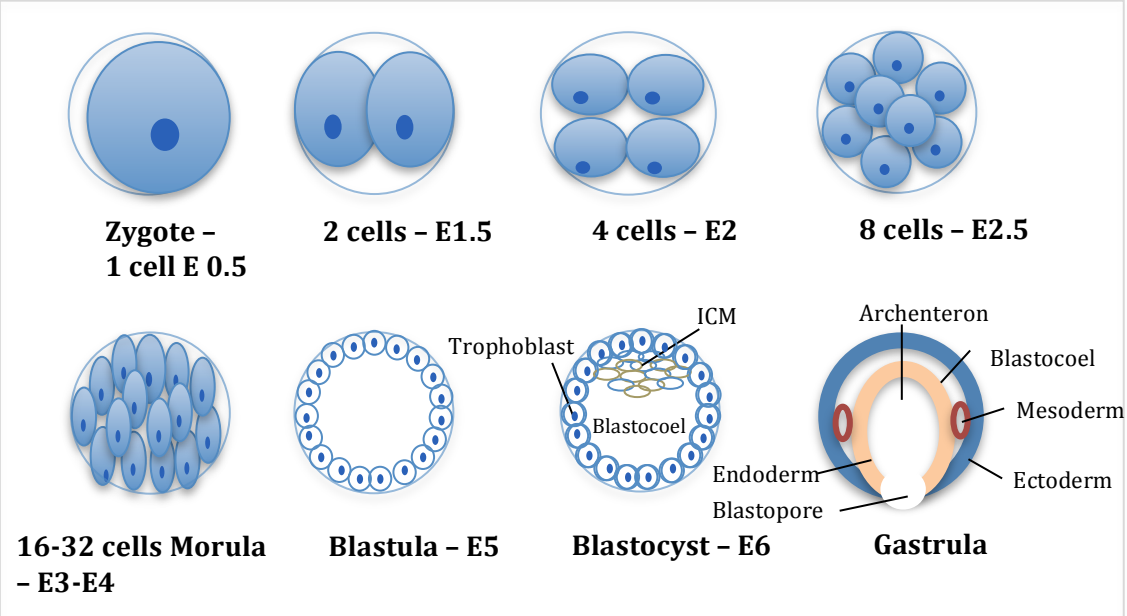


Figure 1: Steps in human embryogenesis. After fertilization there are successive cleavages producing a 32-cells Morula, which arrange themselves to produce a cavity called Blastocoel. With the formation of the Blastocyst cell specialization begins. The Inner Cell Mass (ICM) gives rise to ESCs that will form the three germ layers of the gastrula.

1.2 Cellular differentiation and chromatin

ESCs are derived from the inner cell mass of the blastocyst. They can differentiate into a broad spectrum of cells (pluripotency) and they can propagate continuously (self-renewal). OCT4, SOX2, and NANOG are key to the pluripotency of ESCs (Avilion et al., 2003; Chambers et al., 2003; Mitsui et al., 2003; Nichols et al., 1998). OCT4 promotes segregation of ICM and their commitment to the embryonic lineage in the pre-implantation embryo, while CDX2 determines trophoblast lineage (Nichols et al., 1998; Niwa et al., 2005; Strumpf et al., 2005). Other proteins that are chromatin modifiers (better described from the next chapter onwards) also play a role in cell fate determination during the pre- and post-implantation embryogenesis. Examples

in the pre-implantation embryo are SETDB1 (SET domain, bifurcated 1; also known as ESET and KMT1E), a histone modifier that is crucial for implantation by recruiting OCT4 to silence developmental regulators in ICM, which would otherwise become trophoblasts, including CDX2 (Bilodeau et al., 2009; Lohmann et al., 2010; Yeap et al., 2009; Yuan et al., 2009); and *Mbd3 (*methyl CpG-binding domain protein 3 gene), a chromatin remodeller that also upholds ICM development by preventing expression of trophoblast-specific genes (Kaji et al., 2007). Other examples include: MOF, which directly binds to *Nanog*, *Oct4* and *Sox2*; and other chromatin modifiers such as TIP60, TRRAP (transformation/transcription domain-associated protein), BRG1, and SMARCB1 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1) (Bultman et al., 2000; Bultman et al., 2006; Ho et al., 2009; Kidder et al., 2009; Li et al., 2012; Lohmann et al., 2010; Yeap et al., 2009; Yuan et al., 2009).

Chromatin remodeling has an effect on gene expression and therefore on cellular differentiation. In undifferentiated cells, like ESC, chromatin is mostly open, meaning that it is accessible to transcription factors and specific modifications that allow genomic transcription of both coding and non-coding elements (Efroni et al., 2008). Thus, undifferentiated cells are low in heterochromatin, and their chromatin is less condensed when compared to differentiated/adult cells. As cells differentiate, global chromatin remodelling occurs and exchanges of chromatin proteins are reduced (Ho and Crabtree, 2010; Park et al., 2004a). Expectantly, multipotent adult stem cells such as hematopoietic stem cells and neural progenitor cells, which occur at low amounts in tissues, have an intermediate chromatin state with only a sub-set of ‘stemness’ genes active (Schuettengruber et al., 2009).

1.3 Chromatin structure and epigenetics

Chromatin is composed of a long DNA molecule wrapped around histones and other bound proteins. There are five types of histone designated H1, H2A, H2B, H3 and H4, which are highly conserved among eukaryotes (Szenker et al., 2011). The simplest form of chromatin is a 'beads-on-a-string' fiber, each bead being the smallest unit of chromatin and designated the nucleosome.

Chromatin structure is largely influenced by epigenetic modification, modification caused by environmental factors that alter gene expression without changes to the DNA sequence and which are heritable, mostly by a processes not fully understood.

1.3.1 Chromatin structure

Chromatin structure affects transcription (activation or repression) and other processes involving DNA, including replication, DNA repair, and recombination (Li et al., 2007). Changes to chromatin compaction by chromatin remodelers, epigenetic modifications and histone variants makes DNA accessibility easier or more difficult.

1.3.1.1 The Nucleosome

The nucleosome is composed of four different histones, a H3/H4 tetramer and two H2A/H2B dimers, forming an octamer wrapped by 145-147 base pairs (bp) of DNA in a 1.8 helical turn (Figure 2). Chromatinized DNA in the nucleus also contains H1 bound to the region of exit and entry points of the DNA to the nucleosome.

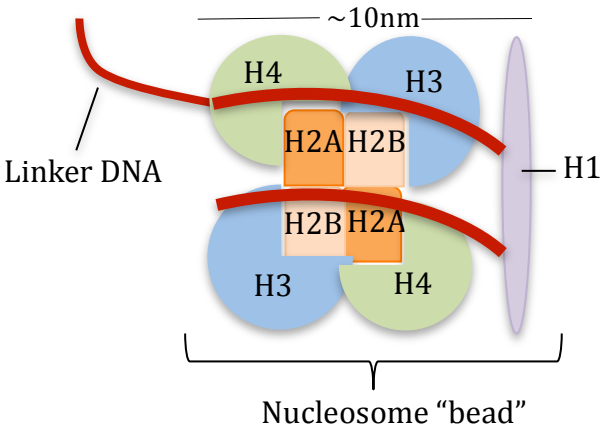


Figure 2: A nucleosome. Eight histone proteins (H4, H3, H2B, H2A) are wrapped around by 145-147 bp of DNA. H1 keeps DNA wrapped around the octomer.

In the genome there are nucleosome-free regions, and other regions with defined nucleosomes positions (Jiang and Pugh, 2009; Lee et al., 2007; Yuan et al., 2005). Nucleosome assembly on DNA is dependent on DNA sequence and other surrounding proteins (Segal and Widom, 2009). H1 is more abundant in repressed loci than in activated genes (Cockerill, 2011). Nucleosome plus H1 and other proteins form a 10nm fibre in vitro. In the presence of 0.5 mM $MgCl_2$ or 60 mM NaCl, this fibre coils for further compaction into a 30 nm 'solenoid' fibre, but it is unclear if such a structure exists in cells. Further compaction into loops and rosettes is thought to form chromosomes and this compaction can reach 10,000-fold.

Nucleosomes interfere with transcription regulation through modifications on histone tails, DNA methylation, variation in histone composition through incorporation of histone variants and through histone rearrangement/displacement by chromatin remodeling factors, such as histone chaperones and ATP-dependent chromatin remodelers (Campos and Reinberg, 2009; Kouzarides, 2007; Li et al., 2007). Differences in histone variants include changes in their primary amino-acids sequence that can range from a few changes to changes in large domains and they confer different

properties to the nucleosome (Szenker et al., 2011; Talbert and Henikoff, 2010). Regarding H3, there are seven variants: H3.1 (mammalian-specific), and H3.2 are both canonical thus expressed during S phase where DNA is actively replicating; three H3.3 variants are associated with actively transcribing chromatin (G1, and G2 phases); and there are also the primate-specific H3.Y and H3.X (Allshire and Karpen, 2008; Wiedemann et al., 2010; Witt et al., 1996). Correlating with H3.3 presence at active genes, nucleosomes containing this variant are less stable and unable to recruit H1 that is involved in chromatin compaction (Braunschweig et al., 2009; Jin and Felsenfeld, 2007). Histone H2A comprises the canonical H2A (most abundant), the variant H2A.Z (vertebrate variants H2A.Z-1 and H2A.Z-2) that confers stability to the nucleosome thus avoiding eviction of the H2A/H2B dimer, H2A.X involved in DNA repair, H2A.Bbd (H2A.B) associated with active genes during spermatogenesis, and macroH2A (mH2A1 and mH2A2) ((Bonisch and Hake, 2012; Hoch et al., 2007; Park et al., 2004b; Soboleva et al., 2012; Talbert and Henikoff, 2010). Interestingly, when H2A.Z is combined with H3.3 the nucleosome becomes more unstable and with higher turnover rates than other combination of these histones (Jin and Felsenfeld, 2007). RNA Polymerase II (RNAPII) recruits facilitator of active transcription (FACT), which evicts the H2A/H2B dimer while RNAPII passes through the nucleosome during transcription; the dimer is then replaced after the polymerase has passed (Reinberg and Sims, 2006).

1.3.1.2 Histone modifications

The N-terminus tail of histones is flexible, protruding outwards, and can be subjected to posttranslational covalent modifications such as methylation, acetylation, ubiquitination, and phosphorylation (Bernstein et al., 2007; Bonasio et al., 2010; Zentner and Henikoff, 2013)**.** Most of these modifications occur on lysines (Lys or K) residues in the tails. Methylation can also occur on arginine (Arg or R) whereas

phosphorylation occurs on serines and threonines (Li et al., 2007). Histone posttranslational modifications can lead to either a permissive or repressive environment for transcription by causing changes in chromatin structure or by recruiting other factors. Histone modifications that promote transcription are acetylation (ac) on Lys and Arg residues of histones H3 and H4 (for example H3K27ac) (Margueron and Reinberg, 2011), di- or tri-methylation (me) of Lys4 residue of H3 (H3K4me2/3), H3K36me, H3K79me, and monoubiquitination (ub1) of H2BK123, while modifications that are correlated with transcriptional repression are methylation on H3K9, H3K27 and H4K20 and ubiquitination (ub) of Lys119 (H2BK119ub) (Kouzarides, 2002; Martin and Zhang, 2005; Nguyen and Zhang, 2011; Peterson and Laniel, 2004). Histone acetylation directly alters chromatin structure by neutralizing positively charged lysine residues on the histone tails, loosening their interaction with negatively charged DNA. Other histone modifications form binding sites for other chromatin regulatory proteins. During gene activation, methylation of H3K4 forms a binding site for activator enzymes such as acetylases and nucleosome remodelers (Pray-Grant et al., 2005; Santos-Rosa et al., 2003; Sims et al., 2005; Wysocka et al., 2005). SET domain-containing Trithorax group of proteins (TrxG) (described in section 1.5), such as the MLL group of proteins, catalyze di- and tri-methylation of H3K4 at transcription start sites (Schuettengruber et al., 2011). H3K36me is an elongation mark that locates along the coding regions of active genes contrary to H3K27me3 that localizes at promoter regions of silenced genes (Mikkelsen et al., 2007). Histone modifications have specific patterns of distribution on the genome and many of them are known to be dynamic (Bernstein et al., 2007; Lanzuolo and Orlando, 2007; Li et al., 2007).

Histone modifications can be epigenetic. As such, these marks are transmitted from mother to daughter cell during mitosis, thus keeping the gene in the same transcriptional state (Margueron and Reinberg, 2010; Probst et al., 2009). Epigenetic

modifications are therefore a memory of past stimuli that can be changed in response to opposite signals. Whilst epigenetic modifications like H3K9me and DNA methylation (a cis modification that occurs on CpG dinucleotides) are well studied and understood**,** the process by which inheritance of H3K4me3, H3K27me3 and other modifications happens is not so well understood and require more research on the subject.

1.3.2 Epigenetic modifications

Two processes are considered to be the basis for epigenetic changes: histone modifications and DNA methylation. The links between these two processes is not well understood, however it is known that a specific histone methyltransferases can recruit DNA methyltransferases to target genes (Vire et al., 2006). Although some epigenetic effects can be reversible, X-chromosome inactivation and imprinting are permanent.

1.2.2.1 Methylation of Lysine-27 on H3

H3K27me2/3 is a very important and broad repressive histone mark. It is a dynamic modification catalyzed by polycomb repressive complex 2 (PRC2, which is better described in chapter 1.4.4) and removed by the demethylases of the Jumonji C family: UTX (also named KDM6A), and JMJD3 (also known as KDM6B) (Agger et al., 2007; De Santa et al., 2007; Hong et al., 2007). UTX, unlike JMJD3, contains a tetratricopeptide motif for which predictions indicate functions in protein-protein interactions, and it is mainly enriched at promoter compared to protein-coding regions (Agger et al., 2007). Both demethylases regulate HOX gene clusters.

H3K27me3 is present in 15% of the H3 proteins in mammalian ESCs (Peters et al., 2003). In the fruit fly and mammals, the CREB-binding protein (CBP)-mediated histone mark H3K27ac is believed to antagonize H3K27 methylation and is enriched in

its absence (Tie et al., 2009). A few studies (Azuara et al., 2006; Francis et al., 2009; Hansen et al., 2008; Margueron et al., 2009) have given glimpses of how H3K27me3 might be preserved through DNA replication. Once H3K27me3 is established, PRC2 binds to it at the G1 phase of cell division, and this is sufficient for both the recruitment of the three core subunits of this complex and to direct H3K27me3 on the new daughter strand during DNA replication. PcG can physically be maintained on the chromatin during replication *in vitro* (Francis et al., 2009).

1.2.2.2 DNA methylation

DNA methylation is a known epigenetic modification that happens on position 5 of cytosines (5mC) of the dinucleotide CpG palindrome. It causes gene silencing when located on a transcription start site (TSS) but stimulates transcription and may even influence splicing when located in the gene body (Jones, 2012). It occurs in most animals; exceptions being, for instance, the nematode worm *Caenorhabditis elegans* that has no methylation and the fruit fly *Drosophila melanogaster* that rather methylates CpT dinucleotides at a low percentage. On the other hand, vertebrates account for the group of animals with highest amount of methylated CpG (Bird, 2002), they have over 85% of CpG sites methylated. Another phenomenon is the occurrence of CpG islands, DNA regions with 200 base pair or more that are CpG-rich (ten times higher than average in the genome) that associate with 60% of promoters, including all of the housekeeping genes and half of tissue specific genes (Antequera and Bird, 1993; Larsen et al., 1992). Only less than 10% of these regions are permanently methylated (Bergman and Cedar, 2013; Bird, 2002). An example of this is X-chromosome inactivation and genomic imprinting. 5mC is also important for chromosome stability by repressing repeated regions such as centromeres allowing proper chromosome segregation during mitosis, and by repressing transposable elements (Moarefi and Chedin, 2011). CpG

island origin has been suggested to result from the deoxyribonucleic triphosphate pool present at a determined time-frame window of S phase (Cross et al., 1991). However this factor might contribute, if this were the only factor, C:G incorporation would occur along the entire body of the gene rather than enriched at the 5'end and promoter. Another suggestion comes from the fact that replication initiation sites co-localize with CpG islands and have different properties of transcription elongation, being possible that their components could render difficult the access of enzymes that catalyze DNA methylation (Antequera, 2003; Antequera and Bird, 1999). Maintainance of DNA methylation is accomplished by the three DNA methyltransferases (DNMT3a, DNMT3b and DNMT1). *De novo* DNA methylation is mediated by DNMT3a, DNMT3b whereas DNMT1 uses as substrate hemimethylated DNA duplex to include a methylation diagonally on the new strand, a process important for example in imprinting (Wood and Oakey, 2006). *De novo* methylation by DNMT3a at intergenic or non-promoter-proximal regions also has the capability of inhibiting silencing by PcG in neurogenic genes being fundamental for neurogenesis (Wu et al., 2010).

1.4 Polycomb

1.4.1 The discovery of polycomb group proteins in *Drosophila*

Polycomb (Pc) is a gene first identified in the fruit fly *Drosophila melanogaster* in 1947 by P. Lewis for its essential role in maintaining Hox genes in a repressed state as concluded from random mutagenesis screens and observation of their phenotypic effects (Breen and Duncan, 1986; Duncan, 1982; Gaytan de Ayala Alonso et al., 2007; Jurgens et al., 1971; Lewis, 1978; Paro and Hogness, 1991; Struhl and Akam, 1985). This way, like *Pc*, a number of other genes were identified and therefore designated polycomb group proteins (PcG) (Breen and Duncan, 1986; Duncan, 1982; Grimaud et al., 2006b; Jurgens et al., 1971; Lewis, 1978; Sato and Denell, 1985; Struhl and Akam,

1985). PcG proteins are involved in regulation of gene expression by keeping genes repressed that were initially regulated by segmentation genes. The expression of segmentation genes fades, and Hox genes are then kept repressed by polycomb group (PcG) proteins. PcG proteins do not only silence Hox genes. The first indication of this phenomenon came in 1989 where it was observed that the Pc protein also binds to PcG genes (Zink and Paro, 1989). Later genomic distribution including genome-wide studies of PcG proteins distribution revealed they occupy genes encoding for transcription regulators involved in development and responsible for cell differentiation (e.g. neurogenesis), morphogenesis, and signalling pathway both in flies and mammals (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006; Negre et al., 2006; Schwartz et al., 2006; Squazzo et al., 2006; Tolhuis et al., 2006). The expression of some polycomb proteins becomes barely detectable at the end of embryogenesis (Gutjahr et al., 1995). Polycomb-mediated silencing is conserved from plants to humans, thus reflecting the importance of this mechanism (Birve et al., 2001; Cao et al., 2002; Czermin et al., 2002; Franke et al., 1992; Kuzmichev et al., 2002; Levine et al., 2002; Ross and Zarkower, 2003; Schumacher and Magnuson, 1997; Shao et al., 1999; Simon and Kingston, 2013; Tie et al., 2001).

PcG proteins are found in the form of multi-subunit complexes. There are five identified polycomb complexes in *Drosophila*: Polycomb repressive complex 1 (PRC1), PRC2, Pho-Repressive Complex (PhoRC), dRING-associated factors (dRAF) and Polycomb Repressive Deubiquitinase (PR-DUB) (Francis et al., 2004; Lanzuolo and Orlando, 2012). With the exception of dRAF all their components have orthologues in mammals (Klymenko et al., 2006; Lanzuolo and Orlando, 2012; Scheuermann et al., 2010; Schuettengruber and Cavalli, 2010; Simon and Kingston, 2009). The first indication that PcG proteins form complexes was in *Drosophila* from the observation made by co-immunoprecipitation showing that the proteins polycomb (Pc), polyhomeotic (Ph) and posterior sex comb (Psc), which all belong to PRC1, co-

precipitate (DeCamillis et al., 1992; Franke et al., 1992). Similar observations were made for Esc and E(z) (belonging to PRC2) and further biochemical purifications confirmed the existence of two separate complexes (Jones et al., 1998; Shao et al., 1999; Tie et al., 1998).

1.4.2 Polycomb group proteins in mammals

Genome-wide chromatin immunoprecipitation (ChIP-Chip) analysis in mouse and human ESCs (Boyer et al., 2006; Lee et al., 2006) showed that PRC proteins bind to homeodomain-encoding genes of transcriptional regulators - Hox genes – that are involved in anterior-posterior development, as well as other developmental genes. Mammals have four clusters of HOX genes: HOX A, B, C, and D; each localised on different chromosomes. Studies show that PRC proteins also bind to cell cycle and proliferation genes (Negre et al., 2006; Oktaba et al., 2008; Schwartz et al., 2006; Squazzo et al., 2006), and genes encoding components of signal transduction pathways such as Wingless and Notch (Janody et al., 2004; Tolhuis et al., 2006). Other processes where PcG have been involved include X-inactivation (Silva et al., 2003; Wutz, 2011; Zhao et al., 2008), genomic imprinting (Martinez and Cavalli, 2006; Terranova et al., 2008; Wolff et al., 2011), senescence (Aguilo et al., 2011; Bracken et al., 2007; Jacobs et al., 1999), cancer (Simon and Lange, 2008), cell reprogramming (Pereira et al., 2010), and immunity through maintenance of thymic epithelium (Liu et al., 2013), lymphopoiesis in B-cell development (Su et al., 2003), and in controlling CD4+ T-cells differentiation into T helper 1 (Th1) and Th2 (Tumes et al., 2013). The many implications of this group of proteins show their importance for a variety of treatments that can be reflected on cancer, tissue regeneration and the induction of pluripotent stem cells.

1.4.3 The PRC1 Complex

PcG proteins are conserved from flies to humans (Levine et al., 2002). In mammals, there are two main complexes of polycomb proteins - PRC1 and PRC2 – with a higher number of homologues and isoforms (from 16 different genes in *Drosophila* to 37 in mice and humans) that can result in molecular redundancy (Whitcomb et al., 2007).

Table 1: Polycomb group complexes and its respective subunits. Based on Lanzuolo and Orlando (2012)

Drosophila	**Mammals**
Polycomb repressive complex 1 (PRC1)	
Ph	HPH1-3
Psc	PCGF1-6
Pc	CBX2-8
dRybp	RYBP, YAF2
Sce	Ring1A-B
Polycomb repressive complex 2 (PRC2)	
E(z)	EZH1, EZH2
Su(z)12	SUZ12
Esc	EED
Nurf55	RbAp46/48
Jarid2	JARID2
Jing	AEBP2
Pcl	PCL1-3
Pho repressive complex (PhoRC)	
Pho	YY1
dSfmbt	SFMBT1
Polycomb repressive deubiquitinase (PR-DUB)	
Calypso	BAP1
ASX	ASXL1

PRC1 is composed of six groups of subunits (Table 1). The components of this complex and respective *Drosophila*/mammalian orthologues are: Ph/HPH1-3 (polycomb-like protein); Pcs/ PCGF1-6; Pc/CBX family (Polycomb Chromobox protein); Sce (Sex comb extra)/RING1A/B and dRybp/RYBP (Ring1/YY1-binding factor) and YY1-associated factor 2 (YAF2) homologues (Francis et al., 2001; Lanzuolo and Orlando, 2012; Levine et al., 2002; Simon and Kingston, 2009). PCGFs (PcG ring fingers) group is constituted by the paralogues NSPC1 (nervous system Pc1)/PCGF1, MEL18/PCGF2, PCGF3, BMI1/PCGF4, PCGF5, MBLR (Mel18 and Bmi1-like ring finger)/PCGF6. Each paralogue has been identified to be part of a distinct PRC1, forming six different complexes, PRC1.1-PRC1.6. In particular, PRC1.2 and PRC1.4 can form two sub-complexes of unique subunit composition, one that contains CBX and HPH, and another that rather contains RYBP/YAF2. These two sets of subunits are mutually exclusive as they compete for binding to RING1B. This non-CBX-containing PRC1 or RYBP-PRC1 is designated non-canonical or variant PRC1 (Gao et al., 2012). RYBP-PRC1 contains other associated proteins such as L3MBTL2 or KDM2b (Farcas et al., 2012; He et al., 2013; Qin et al., 2012; Trojer et al., 2011; Wu et al., 2013). RYBP and YAF2 are homologues and mutually exclusive too.

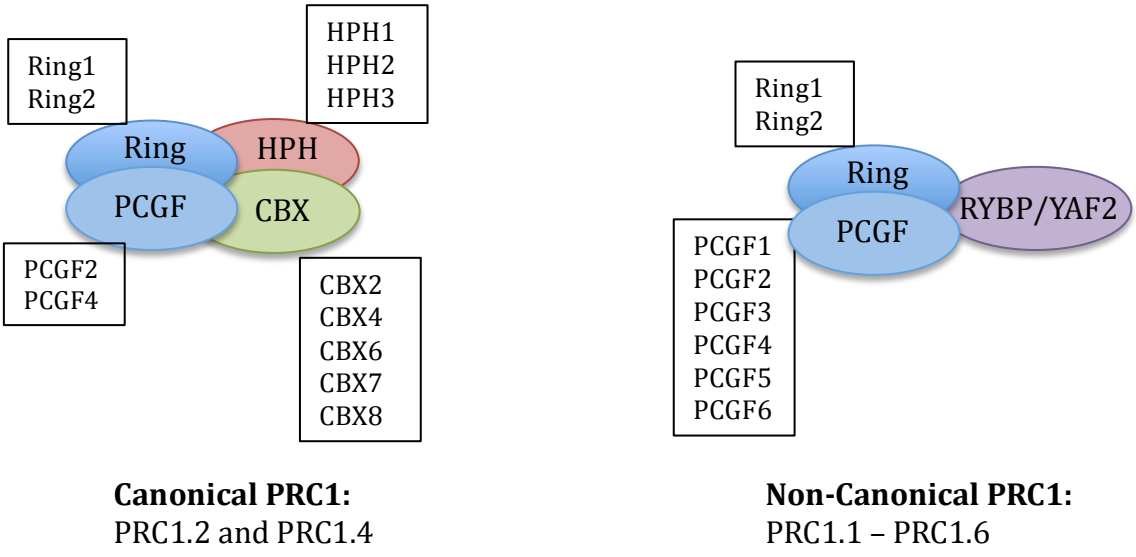


Figure 3: PRC1 complex. Representation of canonical (PRC1.1-PRC1.6) and non-canonical PRC1 complex (PRC1.2 and PRC1.4). Canonical and non-canonical PRC1.2 and PRC1.4 are mutually exclusive.

1.3.3.1 PRC1-mediated gene repression

RING1A/B, an ubiquitin E3 ligase, promotes the mono-ubiquitination of histone H2A at lysine 119 (H2AK119ub1) (Bentley et al., 2011; Brown et al., 1998; Cao et al., 2005; Kallin et al., 2009; Wang et al., 2004a). Reports show that H2A mono-ubiquitination by PRC1 contributes to gene silencing by blocking RNAPII elongation in ESCs (Brookes et al., 2012; Stock et al., 2007; Zhou et al., 2008), concordant with studies showing that PRC1 and RNAPII co-localize at target genes (Breiling et al., 2001; Dellino et al., 2004; Min et al., 2011). If H2AK119ub1 is lost, even though H3K27me3 is present, repression is also lost (Bernstein et al., 2006a; de Napoles et al., 2004; Wang et al., 2004a). It has also been shown experimentally that this gene silencing mark is heritable (de Napoles et al., 2004). H2AK119ub1 co-localises with H3K4me3 at a higher percentage than H3K27me3, 97% to 79% (Brookes et al., 2012).

Blocking RNAPII transcriptional elongation is not the only way of PRC1 repressiveness. Indeed, PRC1 silences target genes through various mechanisms: it inhibits ATP-dependent chromatin remodelling of the hSWI/SNF complex, and induces chromatin compaction in a manner that requires nucleosomes but not histone modifications such as H2AK119ub1 since it can be accomplished in tail-less histones (Eskeland et al., 2010; Francis et al., 2004; King et al., 2005; Margueron et al., 2008; Shao et al., 1999). Psc (Polycomb sex comb), a *Drosophila* ortholog of human PCGF1-6, is the main subunit part of the PRC1 core components responsible for this type of repression, since mutation in this subunit but not in dRING1 cause gene activation. However, loss of ubiquitination through mutation of Ring1 leads to complete loss of repressive activity at some other genes (Gutierrez et al., 2012). PRC1 also recruits histone deacetylases (HDACs) (Tie et al., 2001; van der Vlag and Otte, 1999). The diverse PRC1 complexes occupy distinct genomic loci but their target genes show

similar enrichments of gene ontology (GO) terms (Gao et al., 2012). PRC1 contains a single unit of each subgroup of the complex, and they are interdependent of one another for gene repression (Maertens et al., 2009). Nonetheless, orthologs can co-localize through the binding of multiple PRC1 complexes simultaneously (Maertens et al., 2009; Pemberton et al., 2014).

1.4.4 The PRC2 complex

PRC2 is composed of four core subunits, where the catalytic component is the SET domain of the subunit enhancer of zeste homologue 2 (EZH2) (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002), which is conserved from flies to humans. The other three core subunits are suppressor of Zeste 12 (SUZ12), embryonic ectoderm development (EED), and Retinoblastoma-associated proteins 46 and 48 (RbAp46/48; also known as RBBP7/4; and Nurf55 in *Drosophila*) (Cao and Zhang, 2004a; Margueron and Reinberg, 2011; Nowak et al., 2011) and each is relevant for the functionality of the complex. Non-stoichiometric subunits of PRC2 are adipocyte enhancer-binding protein 2 AEBP2 (Peng et al., 2009), JARID2 (Chen et al., 2011; Landeira et al., 2010; Pasini et al., 2010; Peng et al., 2009), and PHF1 (PCL1), PHF19 (PCL3), MTF2 (PCL2) which are the mammalian orthologs for the fruit fly polycomb-like (Pcl) protein (Margueron and Reinberg, 2011; Simon, 2010). HDACs, although not part of PRC2, also interact with this complex. JARID2 (jumonji (jmj) AT-rich interactive domain (ARID) 2) belongs to jmj family of proteins, whilst AEBP2 (adipocyte enhancer binding protein 2) is a zinc finger protein (He et al., 1999; Kim et al., 2009), evolutionarily well conserved from flies to mammals (Kim et al., 2009) that is present in murine brain tissue, which has been shown to co-purify with the mammalian PRC2 (Cao et al., 2002; Cao and Zhang, 2004a, b) and shares the same target loci (Kim et al., 2009).

1.4.4.1 PRC2-mediated gene repression

PRC2 is firstly recruited to its genomic site to catalyse methylation of H3K27 (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002). The SET domain of EZH2 is required for recruitment to target genes (Margueron et al., 2008) while EED and SUZ12 are necessary for the enzymatic activity of EZH2 (Cao and Zhang, 2004b; Montgomery et al., 2005; Pasini et al., 2004) . Esc (EED ortologue) binds directly to histone H3 and this binding is necessary for the catalytic activity, which indicates that PRC2 interacts with chromatin through histones rather than binding directly to DNA (Tie et al., 2007). Similarly, mammalian EED binds H3K27me3, suggesting a mechanism by which this histone mark may be propagated (Margueron et al., 2009). EED may also function in PRC2 repressive activity through its interaction with HDACs (van der Vlag and Otte, 1999).

The PRC2 component EZH2 has a variant paralogue, EZH1, which might confer different properties to the complex and these two paralogs are mutually exclusive in the complex (Margueron et al., 2008). EZH1 can replace EZH2 in PRC2 and is particularly predominant in adult non-dividing cells, and is capable of polynucleosome compaction. EZH2 is highly expressed during embryogenesis and in proliferating cells. Complexes containing EZH2 have higher HMT activity than EZH1-containing complexes (Bracken et al., 2003; Margueron et al., 2008; Shen et al., 2008). Although EZH1 methylates H3K27 weakly, it mediates gene repression mainly through chromatin compaction accomplished through binding to a few nucleosomes and bringing them together (Margueron et al., 2008). Still, EZH1 seems to complement EZH2 in maintaining stem cell identity (Shen et al., 2008). Another function related with transcriptional activation as also been reported for EZH1 (Mousavi et al., 2012; Stojic et al., 2011). The H3K27me3 mark enables condensation

of chromatin and inhibition of chromatin remodelling, facilitated probably by providing a binding site for PRC1 (Cao et al., 2002; Czermin et al., 2002; Francis et al., 2001; Kirmizis et al., 2004; Kuzmichev et al., 2002; Muller et al., 2002; Shao et al., 1999).

There are also four distinct isoforms of EED due to alternative translational starting sites. PRC2 has been categorized as the form of the complex containing the largest isoform, EED-1, and can assist methylation of both H3K27 and H1K26, whereas complexes containing the shortest isoforms, EED-3 and EED-4, named PRC3, can only methylate H3K27 (Kuzmichev et al., 2004). The EED-2 isoform is expressed only in cancer and undifferentiated ES cells, and is part of another complex, PRC4, which preferentially methylates H1 (Kuzmichev et al., 2005).

JARID2 has essential roles in tissue development and has been identified as a transcriptional repressor with demethylase activity (Takeuchi et al., 2006). JARID2, like other PcG proteins, is also required for the differentiation of mouse ESCs (Pasini et al., 2010). The jmjC domain is a characteristic feature of the jmj family of H3K27-specific demethylases but due to amino-acids substitutions it is inactive in JARID2 (Cloos et al., 2008; Swigut and Wysocka, 2007). The ARID domain, present in JARID2, binds to DNA specially enriched for CG and GA dinucleotides, but is not present in every jmj protein. JARID2 binds to more than 90% of previously mapped PcG target genes and is sufficient to recruit PcG proteins (Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009). This recruitment and tethering of PRC2 is dependent on the ARID domain DNA-binding ability (Kim et al., 2003; Li et al., 2010; Pasini et al., 2010; Peng et al., 2009). There are reports saying that JARID2 inhibits HMT activity of PRC2 (Peng et al., 2009; Zhang et al., 2011), and controversially, other reports show that it rather stimulates HMTase activity (Li et al., 2010; Mejetta et al., 2011). It is rather possible that JARID2 fine-tunes the H3K27me3 level in vivo as

suggested by Shen *et al* (2009). JARID2 can bind to DNA potentially through its ARID domain, thus facilitating anchoring of PRC2; ChIP-Seq shows a significant binding overlap between the two (Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009). JARID2 inhibition leads to loss but not complete abolition of PcG binding and a subsequent reduction of H3K27me3 and H2AK119ub levels on target genes (Landeira et al., 2010; Li et al., 2010; Shen et al., 2009). Besides the ARID domain, JARID2 contains a C5HC2 zinc finger and both may synergize to bind to DNA (Kim et al., 2004; Kim et al., 2003; Li et al., 2010). Landeira et al. (2010) further shows that the presence of Ser 5 phosphorylated RNAPII correlates with the presence of JARID2, implying that JARID2 is required for the poised state of PRC engaged genes and subsequent reactivation.

There are 3 PCL proteins, which bind to subsets of PRC2 binding sites and are also involved in PcG recruitment (Aloia et al., 2013; Brien et al., 2012; Hunkapiller et al., 2012; Walker et al., 2010). PHF19 (PCL3) binds to H3K36me3 through its Tudor domain also recruiting the H3K36me3 demethylase NO66 (Brien et al., 2012). Knockdown of PHF19 results in reduction of both H3K27me3 and SUZ12 binding (Hunkapiller et al., 2012). In *Drosophila*, it is required for high levels of H3K27me3 but not H3K27me1 and H3K27me2 (Nekrasov et al., 2007). RbAp46/48 (PRC2) are histone-binding proteins that are also present in HDAC and co-repressor complexes (Cao et al., 2002; Kuzmichev et al., 2002).

1.4.5 Canonical and noncanonical PcG gene repression

In canonical PcG repression, PRC1 and PRC2 often co-occupy target sites (Boyer et al., 2006; Ku et al., 2008; Schwartz et al., 2006). PRC2 is firstly recruited to the target genes exerting its catalytic activity; PRC1 subsequently follows by binding to

H3K27me3 through the CBX chromodomain, which recognizes this modification. Therefore, H3K27me3 functions as a docking site for PRC1 recruitment and targeting (Figure 3) (Cao et al., 2002; Fischle et al., 2003; Min et al., 2003). In murine ESCs, CBX7 is probably the main recruiter of PRC1 as it is more predominant in these cells. However, as cells differentiate, CBX7 fades and only a subset of canonical PRC1 (CBX2- and CBX4-containing PRC1) prevails, as they are fundamental for proper differentiation (Aloia et al., 2013; Gao et al., 2012).

Noncanonical PRC1 is a variant that does not contain CBX or PHC and is composed of one PCGF paralog, RYBP and RING1A/B. RYBP and CBX are mutually exclusive (Gao et al., 2012; Kalenik et al., 1997). This means that a different recruitment mechanism of PRC1 must exist that is independent of PRC2. RYBP binds to RING1B through its C-terminal domain and this is important for its functionality by enhancing H2AK119 ubiquitination independently of PRC2 (Gao et al., 2012; Tavares et al., 2012) Recent work in mouse ESCs also indicates that RYBP-PRC1 can be recruited independently of PRC2 to unmethylated CpG islands via KDM2b (Farcas et al., 2012; He et al., 2013; Wu et al., 2013). RYBP-PRC1 and CBX7-PRC1 target overlapping genomic sites, being the genes bound by both the most repressed (Morey et al., 2013). Additionally, RYBP has been demonstrated to possess repressive activity, by binding to YY1 and possibly establishing a link between YY1 and PcG (Garcia et al., 1999).

Concordant with different mechanisms of PRC1 silencing, PRC1 and PRC2 can bind genomic sites independently of each other's action (Ku et al., 2008; Richly et al., 2010; Sing et al., 2009; Trojer et al., 2011). Due to this independent binding of PRC1 and PRC2, questions regarding the role of H3K27me3 have been raised. Supporting recruitment of PRC1 independently of H3K27me3, there is a low abundance of H3K27me3 (Nekrasov et al., 2007) at the core of PRE regions whereas PC is highly

enriched, and that H3K27me3 occupies broad domains whereas PC rather occupies narrow regions (Papp and Muller, 2006; Schwartz et al., 2006). Furthermore, after the absence of the lysine tri-methylation, the loss of PRC1 is gradual (Ohno et al., 2008). Not only recruitment of PRC1 can be independent of H3K27me3 as PRC1-dependent H2AK119ub1 leads to PRC2 recruitment and subsequent H3K27 trimethylation (Blackledge et al., 2014). Additionally, and as mentioned above in section 1.2.2.1, H3K27me3 can recruit PRC2 (Hansen et al., 2008). Histone variants also play a role in polycomb targeting. Histone H3.3 is required for higher rates of histone turnover and has been demonstrated to be involved in the establishment of H3K27me3 via promoting interaction of its histone chaperone with PRC2 (Banaszynski et al., 2013).

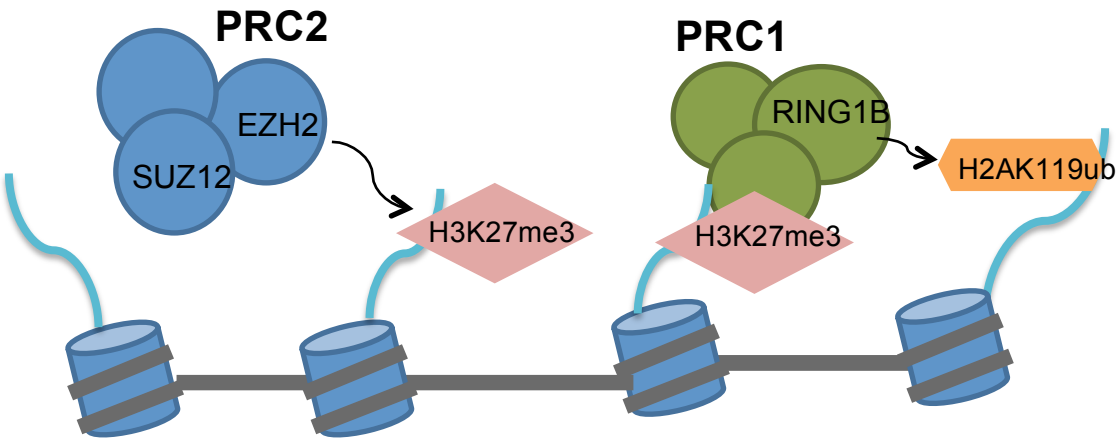


Figure 4: Representation of the canonical mechanism of polycomb mediated gene repression. PRC2 is recruited to the genomic site, trimethylates H3K27, which servesas a binding site for PRC1. RING1B, the catalytic subunit of PRC1 the monoubiquitanates H2AK119, which cause RNAPII to stall.

1.4.6 Other PcG complexes

Biochemical purification of pleiohomeotic (Pho) revealed that it forms a third PcG complex in *Drosophila*, known as Pho repressive complex (PhoRC) (Klymenko et al., 2006). This complex is composed of Pho, pleiohomeotic-like (Phol) and SCM-related gene containing four MBT domains (dSfmbt). The transcription factor yin yang-

1 (YY1) has been described as the mammalian homolog of Pho (Brown et al., 1998; Klymenko et al., 2006; Wang et al., 2004a); it is conserved between vertebrates (Pisaneschi et al., 1994; Shi et al., 1991); and is ubiquitous. While it is known that Pho and Pho-like proteins are important for polycomb repressiveness in *Drosophila*, the role of YY1 in mammals is not clear (Atchison et al., 2003; Srinivasan and Atchison, 2004).

1.5 Trithorax group proteins

Trithorax group proteins (TrxG) are counterparts of the PcG proteins that also maintain gene expression patterns through epigenetic modifications (Maeda and Karch, 2006; Ringrose and Paro, 2004), but act antagonistically to them (Moehrle and Paro, 1994). They have long-term effects in *Drosophila* Hox genes but there is no certainty of such effect in vertebrates (Maeda and Karch, 2006; Ringrose and Paro, 2004). Expression of the genes regulated by the PcG-TrxG system are fine tuned and dynamic along the development (Ringrose, 2007), this being essential for embryonic development and cell differentiation (Christophersen and Helin, 2010; Schumacher and Magnuson, 1997) and stem cells and differentiated cells identity by binding to essential developmental regulators, cell growth and proliferation factors (Mendenhall and Bernstein, 2008; Ringrose and Paro, 2004; Ringrose et al., 2003). Like PcG, they are conserved across species (Ringrose and Paro, 2004; Schuettengruber et al., 2007).

Proteins of the Trithorax group act in multimeric complexes that can be divided in two groups: 1) SET domain-containing histone methyltransferases; 2) ATP-dependent chromatin-remodelling factors. Some other TrxG proteins can bind to DNA and act as chromatin remodelers and histone modifiers (Schuettengruber et al., 2011). Complexes of the first group include COMPASS that catalyses the trimethylation of H3K4 (Wu et al., 2008); COMPASS-like (MLL1-2 in one complex and MLL2-4 in

another complex) that has tumor suppressor activity (Yang and Hua, 2007) and catalyses of the acetylation of H4K16 (Gu et al., 1992). The demethylase UTX also belongs to the COMPASS-like complex, which removes H3K27me3 mediated by PRC2 (Agger et al., 2007); and ASH1 that has methyltransferase acivity for H3K36. To the second group belong SWI/SNF that binds to acetylated histones via its bromodomain (part of Brahma; BRM and BRG1) (Chatterjee et al., 2011); ISWI that recognizes H3K4me3 via its PHD finger (Wysocka et al., 2006); and CHD1-8 that bind to H3K4me3 and have histone deacetylase activity (Gaspar-Maia et al., 2009; Schuettengruber et al., 2011).

Studying of the mechanisms underlying Trx function can open new doors to cell regeneration, longevity and environmental stresses as well as possibly other venues (Greer et al., 2011; Klebes et al., 2005; Siebold et al., 2010).

1.6 Genomic profiling of polycomb function

PcG proteins, known gene repressors, have some targeted genes that have low to high levels of activity (Brookes et al., 2012; Nishiyama et al., 2009; Tolhuis et al., 2006; Young et al., 2011). However, this might be due to allelic differences in expression where one allele is expressed and the other one is silent (Brookes et al., 2012); this is possibly to be the case of imprinted genes. 20% of silent genes do not show ChIP-seq signal for H3K27me3, PRC1 or 2, nor RNAPII. From the same study (Brookes et al., 2012), PRC-target silent genes may show the presence of H3K27me3 only (no RNAPII present), which mainly happens in differentiated cells (with signalling and stimuli genes) where developmental regulators specific to other cell types have to be maintained in a silent state to maintain cell identity; or presence of both PRCs

deposited marks plus RNAPII, which occur more on developmental and metabolism-related genes important for embryogenesis.

1.6.1 Bivalency

Numerous promoters of mammalian developmental genes are associated with bivalent chromatin, a term coined by Bernstein *et al*. (2006a) because it contains two antagonistic histone marks: H3K27me3, a marker for gene repression, and tri-methylation at H3K4, a marker of transcription initiation present in active genes (Azuara et al., 2006; Bernstein et al., 2006a). Bivalent genes are silent but some show a low level of expression (Brookes et al., 2012; Voigt et al., 2013). Bivalent domains are dominant at the pre-implantation stage and are a common feature of ESC (Alder et al., 2010) but not of the *Drosophila* epigenome (Schuettengruber et al., 2009). Bivalency occurs in different cell types in mammalian species, including humans (Pan et al., 2007; Voigt et al., 2013; Zhao et al., 2007), and other vertebrates too (Vastenhouw et al., 2010). Bivalency is a particular signature of ESCs where mouse and human bivalency largely overlap (Voigt et al., 2013). Bivalent marks co-exist in the same nucleosome (Voigt et al., 2012), although not in the same histone (Voigt et al., 2012; Young et al., 2009), meaning that they are not a result of cell heterogeneity. To further support this conclusion, a study (Marks et al., 2012) took advantage of 2i medium (that functions by inhibiting both Erk signaling and glycogen kinase 3 (Ying et al., 2008), thus preserving the naive state of ESC by keeping the expression of lineage-specific genes at a lower level and thus preserving cellular homogeneity (Nichols and Smith, 2009)), to culture ESC and still found bivalent domains, although fewer in number. Bernstein *et al*. (2006a) studied bivalent domains in highly conserved noncoding elements (HCNEs), which are mostly present at genes that encode for TFs such as Hox and Sox, Fox and

Pax gene families (Bernstein et al., 2006a). Especially the regulatory regions of vertebrate Hox clusters contain HCNEs in and around the genes existing around 200 of these loci (Bejerano et al., 2004; Lindblad-Toh et al., 2005; Woolfe et al., 2005). 75% of H3K27me3 domains (5 kb) in HCNEs are also methylated on K4 where 50% of these are present at TFs genes (Bernstein et al., 2006a; Lee et al., 2006). In a broader genomic study, Mikkelsen *et al*. (2007) identified three different classes of promoter according to CpG content. 99% of CpG-high promoters – promoters of housekeeping genes or developmental genes – were enriched for H3K4me3, 22% of which were bivalent. There is therefore a correlation between CpG islands, bivalency and conserved regions.

Upon cell differentiation, most bivalent domains resolve into monovalent H3K27me3 (some of which become DNA methylated) (Mohn et al., 2008), H3K4me3 or neither mark (Mikkelsen et al., 2007). Nonetheless, many new bivalent sites are formed in progenitor stem cells that also have a ESC-like prevalence of bivalent domains, as it is the case in neural progenitor cells (NPC), mesenchymal stem cells (MSC), hematopoietic stem cell (HSC), showing this is a highly dynamic process (Cui et al., 2009; Mikkelsen et al., 2007; Mohn et al., 2008; Paige et al., 2012). Bivalency occurs in cultured ESCs as in the inner cells of the blastocyst of the developing embryo (Rugg-Gunn et al., 2010). In the trophoblast and extra-embryonic endoderm stem cells, H3K9me3 replaces H3K27me3 mark and function as bivalent marks (Rugg-Gunn et al., 2010).

It is also worthy to note that bivalency can also happen with the repressive mark H2Aub1, which blocks RNAPII transcription, where 97% of the marks coincide with H3K4me3 whereas 79% of H3K27me3 overlaps H3K4me3. The three marks can co-exist (Brookes et al., 2012).

1.6.2 Elongation by RNAPII and transcriptional pausing

Production of mRNA starts with the assembly of a pre-initiation complex (PIC) together with general transcription factors (GTFs) and a mediator complex, being these recruited to gene promoter. It was initially thought that blocking such recruitment was the only way of regulating gene expression. However, it was found that regulation could happen at the level of elongation too (Figure 4). Promoter-proximal pausing, first identified in *Drosophila heat shock* promoters, for example *HSP70* (Gilmour and Lis, 1986; Rougvie and Lis, 1988), is associated with transcripts of around 20-60 nucleotides (nt) (Rasmussen and Lis, 1993; Rougvie and Lis, 1988). Other ground-breaking examples of such studied genes are HIV TAR, from which transcripts of 59-nt are produced from the HIV LTR in unstimulated cells (Kao et al., 1987); and *MYC* (Bentley and Groudine, 1986). In these three cases, pausing is due to the action of two factors: DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF). The repressive activity is antagonized by the recruitment of positive transcription elongation factor (P-TEFb) (Marshall and Price, 1992). Subsequent genome-wide studies demonstrated that this type of transcriptional regulation is a common feature in *Drosophila* and mammalian genomes too (Guenther et al., 2007; Nechaev and Adelman, 2008; Price, 2008; Zeitlinger et al., 2007), probably to allow rapid and efficient expression upon physiological changes.

The ends of the integrated HIV provirus comprise a long terminal repeat (LTR) segmented into U3, R, and U5 (Knipe, 2006). In naïve resting CD4 T-cells the provirus is in a latent state but transcription is still initiated, producing a short RNA named TAR (trans-activation response element). TAR is a *cis* RNA required for the activation of the provirus through binding of the viral Tat (trans-activator of transcription) protein, which occurs when T-cells encounter an antigen and are

activated. A significant feature of the TAR RNA is its stem-loop secondary structure that is required for Tat binding (Feng and Holland, 1988). Initiation of transcription takes place at the U3/R border of the 5'LTR but in the absence of Tat transcription is aborted about +55 to +59 downstream of TSS and RNAPII stalls (Kao et al., 1987). At HIV, as well as in cellular genes, RNAPII phosphorylated at Ser5 but not Ser2 stalls downstream of the TSS (Yamaguchi et al., 2013). Activation of transcriptional elongation occurs after recruitment of the elongation factor P-TEFb, in the HIV case by Tat. The CDK9 subunit of P-TEFb phosphorylates RNA polymerase II Ser2 (Kim et al., 2002) as well as NELF and the DSIF subunit Spt5 causing them to dislodge (Fujinaga et al., 2004; Ivanov et al., 2000). Additionally, EZH2 has been linked to the HIV LTR (Friedman et al., 2011), which was found tri-methylated at H3K27 and ubiquitinated on H2A (Kim et al., 2011). Knockdown of EZH2 leads to reactivation of the provirus (Friedman et al., 2011).

Pausing can also happen further into productive elongation (Adelman and Lis, 2012; Brookes et al., 2012). Such genes, for example Lhx5, Pitx1, and Zfp503, show higher expression levels upon de-repression than genes where RNAPII is paused near the TSS, such as Fgf5, Kcnc4, and Lrat (Brookes et al., 2012). Consistent with the presence of a transcription initiation marker (H3K4me3) at polycomb-target genes, RNAPII is recruited to the promoter, starts transcription, but stalls and ceases productive elongation (Chopra et al., 2009; Dellino et al., 2004; Stock et al., 2007). RNAPII is present at a substantial percentage of H3K27me3-associated silent genes from *Drosophila* to humans (Guenther et al., 2007; Muse et al., 2007; Zeitlinger et al., 2007). At perichromatic regions, border regions of condensed chromatin, transcription is mainly paused and this region is also rich in polycomb proteins (Cmarko et al., 1999; Ruthenburg et al., 2007; Shahbazian and Grunstein, 2007; Trentani et al., 2003).

The largest subunit of RNAPII, RPB1, contains a Carboxy-terminal domain (CTD) that comprises the heptapeptide consensus sequence N-Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7-C repeated 26 times in yeast, 42 in *Drosophila* and 52 times in mammals. The residues in this sequence are subjected to post-translational modification, particularly the phosphorylation of Serine (Ser) residues (Patturajan et al., 1998). RNAPII can be unphosphorylated, Ser5 phosphorylated (S5P), and hyperphosphorylated (S5P + S2P). Phosphorylation of Ser 5 is essential for transcription initiation, capping and recruitment of Lys4 methyltransferases; while escape from the pause position, transcription elongation, polyadenylation, splicing and recruitment of Lys36 HMT require phosphorylation at Ser 2 (Brookes et al., 2012; Cadena and Dahmus, 1987; Sims and Reinberg, 2004). Therefore, paused RNAPII is phosphorylated only at Ser5 at bivalent genes that are silent, and have no H3K36me3; however S5P can also be found on coding regions at low levels (Brookes et al., 2012). PRC genes are mainly S5P, and hyperphosphorylated RNAPII is not detected above background levels (Brookes et al., 2012; Stock et al., 2007). Such polycomb active genes encode for transcription factors involved in ESC identity such as Hmga2, Klf4 and Tbx3. Levels of Nanog, a ESC mark of pluripotency, influences whether PcG target genes are expressed or not. RNAPII stalling at PcG target genes has been attributed to the H2AK119ub1 modification by PRC1 (Brookes et al., 2012; Stock et al., 2007).

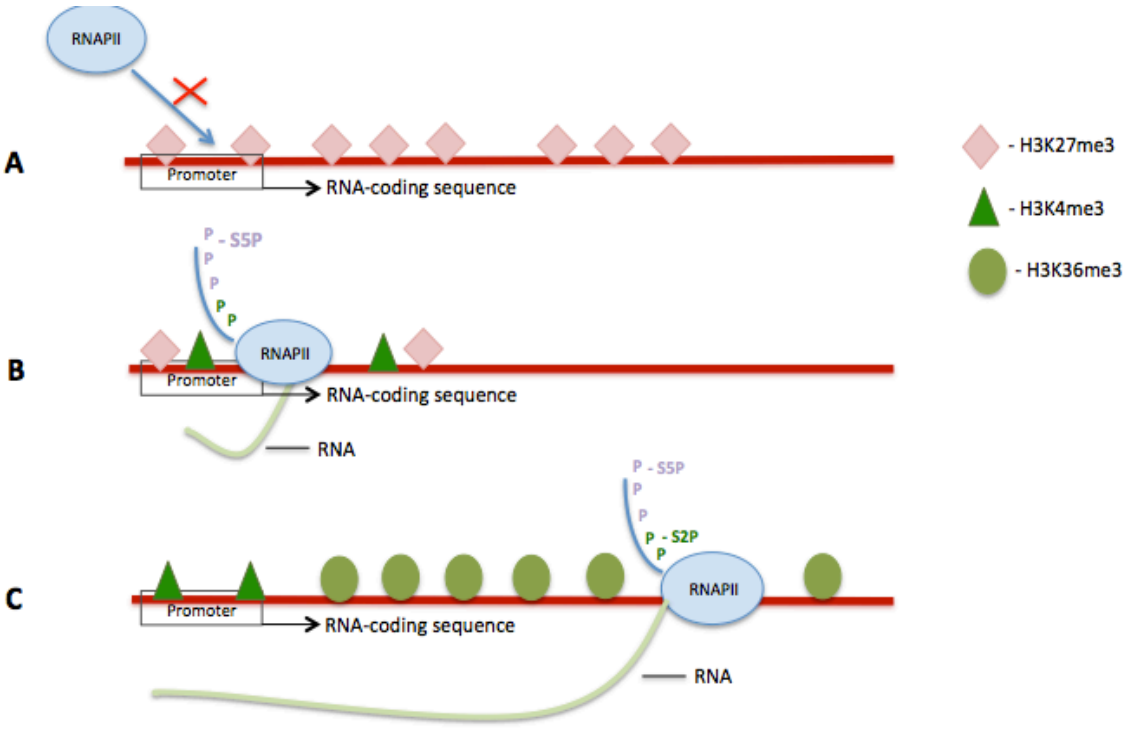


Figure 5: Gene transcription; A – Silent genes: transcription is inhibited at the level of initiation where the recruitment of RNAPII to the promoter is inhibited; Gene site is highly methylated on H3K27. **B** – Silent gene: transcription is inhibited at the level of elongation. RNAPII is phosphorylated at Serine-5 only and the promoter is methylated both at H3K27 and H3K4. **C** – Gene is expressed: RNAPII is phophorylated at both Serine-5 and -2; promoter is methylated at H3K4 only and the body of the gene is trimethylated at H3K36.

1.7 Polycomb recruitment to chromatin

1.7.1 Polycomb Responsive Elements in flies

Polycomb responsive elements (PREs) are cis-regulatory sequences initially identified in transgenic *Drosophila* embryos and larvae that confer polycomb repressive activity (Muller and Bienz, 1991; Simon et al., 1993), thus recruiting PRC1 and PRC2 complexes (Oktaba et al., 2008). Identification of PREs has proven difficult given the lack of identifiable sequence homology. Using algorithms based on several aspects that seem relevant to *Drosophila* PREs, quite accurate predictions have been attained for this animal model. Unfortunately, it has been harder to find it in mammals where only three PRE sequences have been identified so far.

PREs of the *Drosophila* Hox clusters, like the bithoraxoid PRE, can be located thousands of base pairs away, 20-30 kb, from the transcription start site (Chan et al., 1994; Sengupta et al., 2004), whereas PREs like *engrailed* (Kassis, 1994) and many other *Drosophila* PREs are located close to the transcription start sites (Bloyer et al., 2003; Schwartz et al., 2006).

In *Drosophila*, PcG proteins interact with chromatin by binding to Pho and Pho-like proteins (Brown et al., 1998; Wang et al., 2004a). Genome-wide binding profile of PhoRC, PRC1, and PRC2 shows that the three complexes co-localise on a large set of genomic sites, identified as PREs, and that Pho not only directly interacts with Pc (PRC1) and Esc/E(z) (PRC2) but it is also required for the recruitment of both complexes (Oktaba et al., 2008; Schwartz et al., 2006; Wang et al., 2004a). But, little is known regarding the regulation of Pho binding to PREs. It is likely that the carboxy-terminal binding protein (CtBP) interacts with Pho in its recruitment as analysis of the PREs *scr* and *engrailed* in wild-type and mutated *Drosophila* embryos revealed that loss of this protein results in an increased expression of intergenic transcripts and in a reduced binding of Pho (Basu and Atchison, 2010). A binding site of 17bp has been identified for Pho (Brown et al., 1998), with a consensus sequence of only four base pairs, CCAT, considered the core binding site of Pho (Ringrose et al., 2003).

Pho/Phol alone are not sufficient to recruit PcG complexes in vitro (Brown et al., 2003; Dejardin et al., 2005), and Pho/Phol double mutants don't generally interfere with PcG binding in polytene chromosomes (Brown et al., 2003; Wang et al., 2004b). Other DNA-binding proteins must be important for PRE functionality and many have been identified in Drosophila. These proteins include GAGA factor (GAF: also known as TRL), pipsqueak (Psq), zeste, dorsal switch protein 1 (Dsp1), grainyhead (GH) and specificity protein 1 (SP1) or Luna (also known as KLF) (Simon and Kingston, 2009). In an attempt to predict PREs on a large scale, Ringrose and colleagues designed an

alignment-independent algorithm that searches for one or more possible motifs for each of the proteins GAF, Psq, zeste and Pho/Phol. They found that clustered pairs rather than single cluster motif were required to distinguish between *Drosophila* PREs and non-PREs (Ringrose and Paro, 2007; Ringrose et al., 2003). Another study suggested that although Gaf, Zeste and Pho sites are necessary, they are not sufficient to make a PRE (Dejardin et al., 2005). Taking this into consideration, Rehmsmeier and co-workers constructed another prediction algorithm by incorporating the Dsp1, GH, and Sp1/KLF sites, and including comparative genomic data from four *Drosophila* genomes it improved the overlapping of predicted PREs with published polycomb ChIP data from 20% to 34% (Hauenschild et al., 2008). Recently, it has been shown that Dsp1 is important for PcG recruitment given that abolishment of Dsp1 binding to certain PREs causes loss of PcG proteins binding, thus showing relevance in their recruitment (Dejardin et al., 2005; Wang et al., 2010). None of these proteins, however, are considered part of the polycomb group since mutations do not lead to PcG phenotype (Schuettengruber et al., 2009).

Genetic evidence demonstrates that Drosophila PREs also function as TREs (Trithorax response element), as these sequences are bound by both TrxG and PcG proteins (Chan et al., 1994; Orlando et al., 1998; Papp and Muller, 2006; Tillib et al., 1999) and supported by the existence of bivalency at PcG-target sites (Schuettengruber et al., 2009). Trithorax group (TrxG) proteins are required for active gene transcription (Ringrose and Paro, 2004). They include the proteins Trithorax (Trx) and Ash1 (absent, small, or homeotic discs 1), two SET domain-containing proteins that methylate H3K4 (Klymenko and Muller, 2004). An example is the bithorax *Fab-7* PRE/TRE, which is capable of recruiting PcG proteins as well as acting as a TRE that can maintain the memory of active chromatin states during embryogenesis (Cavalli and Paro, 1999; Klymenko and Muller, 2004). The relative levels of PcG, and Trx at target genes

determine the chromatin state, thus creating a delicate and dynamic balance (Schwartz et al., 2010). Interestingly, the PRE binding proteins Zeste, Gaf and Pipsqueak also appear to have an activatory function (Decoville et al., 2001; Hagstrom et al., 1997; Huang et al., 2002).

1.7.2 Polycomb Responsive Elements in mammals

Although much has been investigated about PRE/TREs in *Drosophila*, in mammals, PREs are not well defined. PRE-*kr (Kreisler)*, is the first mammalian PRE to be identified (Sing et al., 2009). It is 3kb long and regulates expression of the mouse *MafB/Kreisler* gene. PRE-*kr* was capable of causing repression of a reporter gene in a PcG-dependent manner and both PRC1 and 2 complexes can bind the PRE sequence, albeit PRC2 binding is weak, which suggests different binding requirements. Although prediction algorithms did not identify the PRE-*kr*, it contains a 450-bp highly conserved region between human, mouse and chicken, and the consensus Pho/YY1-binding sites and GAGAG motifs (Sing et al., 2009).

Another mammalianPRE, identified in the human Hox cluster between the genes HOXD11 and HOXD12, is designated HOXD11.12 (Woo et al., 2010). It is 1.8 kb long, is enriched for both H3K4me3 and H3K27me3, has a highly conserved sequence of 237-bp, and contains GC-rich sequences. Characteristics consistent with *Drosophila* PREs include YY1 binding sites, nucleosome depletion and repression by PcG protein maintained throughout differentiation. Woo *et al*. (2010) showed that deletion of the YY1 binding motif and especially the highly conserved region, as well as interaction with the protein RYBP, all interfere with PRE repressive ability and especially PRC1 binding.

Common characteristics of these two PREs are: regions of high sequence

conservation (Lee et al., 2006), and GC-rich sequences (Ku et al., 2008). GC-rich regions are preferentially bound by PcG (Lynch et al., 2012), and seem to be sufficient to recruit PRC2 (Mendenhall et al., 2010). Other characteristics of PREs include nuclease-hypersensitivity, revealing nucleosome depletion at these sites (common with Drosophila PREs) (Mohd-Sarip et al., 2006; Muller and Kassis, 2006; Papp and Muller, 2006), and chromatin flanking a PRE sequence that contains nucleosomes marked with H3K27me3 (Pan et al., 2007; Schwartz et al., 2006).

1.6.3 The involvement of YY1

YY1 is a multifunctional protein (Bushmeyer et al., 1995) that performs roles in DNA repair (Wu et al., 2007) and transcription regulation, capable of acting as both an activator (Cai et al., 2007; He et al., 2010; Seto et al., 1991) and a repressor (Atchison et al., 2003) of genes, due to the possession of distinct domains (Bushmeyer et al., 1995). Its effect on the silencing of transcription can be accomplished through multiple mechanisms (Galvin and Shi, 1997; Guo et al., 1995), for example, by interacting with transcription factor LSF to inhibit the expression of HIV-1 provirus – a mechanism not well understood (Romerio et al., 1997).

Pho, YY1, and Phol are all sequence-specific DNA binding proteins that contain related zinc finger DNA-binding domains (Brown et al., 2003; Brown et al., 1998). YY1 interacts with the protein RYBP, which in turn interacts with the PRC1 components RING1A, RING1B, and M33/CBX2. YY1 has also been reported to interact with the PRC2 and PRC1 subunits EED and BMI1 respectively (Caretti et al., 2004; Satijn et al., 2001). Conversely, another report indicates that YY1 does not interact with PRC2 (Li et al., 2010). The discrepancy of these results might be due to different cell types or antibody specificity.

YY1 is the mammalian homolog of Pho and HMGB2 is the mammalian homolog of Dsp1 (Gabellini et al., 2002). Krüppel-related zinc-finger protein/T-helper-inducing POZ/Krüppel-like factor (cKrox/Th-POK) is the vertebrate orthologue to the *Drosophila* Gaf, and are linked to PcG recruitment (Matharu et al., 2010). While it is known that Pho and Phol are important for Polycomb repressiveness in *Drosophila*, the role of YY1 in mammals is not fully clear. The observation that the consensus binding sequences of Pho and YY1 have a perfect match suggests that YY1 is a PcG protein. (Brown et al., 1998). Studies in Drosophila show that YY1 recruits (Caretti et al., 2004) and interacts with PcG proteins, like EED (Atchison et al., 2003; Palacios et al., 2010; Satijn et al., 2001; Srinivasan and Atchison, 2004), and Pho also interacts and co-localises with both PRC1 and PRC2 (Klymenko et al., 2006; Mohd-Sarip et al., 2002; Oktaba et al., 2008; Wang et al., 2011). Nonetheless, a study in mouse ESC has evidenced that YY1 does not interact (Vella et al., 2012) nor co-localise with PRC2 (Squazzo et al., 2006). Among other observations that YY1 might constitute a PcG protein that its binding sites are present and required for the four known mammalian polycomb responsive elements: PRE-*kr*, HOXD11.12, HOXC11.12 and HOXB4.5 (Woo et al., 2010, 2013). It functionally compensates for Pho in Drosophila (Atchison et al., 2003; Srinivasan et al., 2005), YY1 is a developmental gene regulator essential for embryogenesis and tissue formation (Donohoe et al., 1999; He et al., 2007). However, YY1 distribution does not correlate with PcG genome-wide (Squazzo et al., 2006), but it is possible that YY1 is implicated in PRC recruitment for a specific set of genes only, what would reflect the variety of mechanisms involved (Simon and Kingston, 2013).

1.7.4 Other potential mammalian recruiters

Proteins such as JARID2, AEBP2, and PCL may act to tether PcG to chromatin (Casanova et al., 2011; da Rocha et al., 2014; Kim et al., 2009; Landeira et al., 2010; Nekrasov et al., 2007). Tet1 was found to demethylate DNA at CpG islands in mouse (Tan and Shi, 2012) and is also important for PRC2 binding to chromatin and both PRC2 and Tet1 co-localise at more than 95% of target sites (Wu et al., 2011). SNAIL is another protein that interacts with EZH2 and SUZ12, at the E-cadherin gene of mouse ESCs (Herranz et al., 2008). Runx1/CBFb recruits PRC1 in a subset of PcG targets in a PRC2-dependent manner (Yu et al., 2012). REST is required for the recruitment of both complexes in a sub-fraction of Pc-regulated neural genes (Dietrich et al., 2012). Other proposed proteins are NurD (Reynolds et al., 2012), and PLZF (Barna et al., 2002). Despite PRE binding proteins being mostly absent in vertebrates, prominent features of PcG proteins are their association with promoter regions, bivalent domains, CpG-rich promoters, high sequence conservation, and also with ncRNAs.

A new consensus in the field is that CpG islands are involved in polycomb recruitment. Introduction of CpG island or GC-rich DNA into the genome of mESCs is sufficient for PRC2 binding, suggesting that PRC2 is recruited to these sites (Ku et al., 2008; Mendenhall et al., 2010). However, CpG islands that are associated with activator factors or that are DNA methylated do not recruit PRC2 (Lynch et al., 2012; Mendenhall et al., 2010).

Recent studies (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014) have shown that PRC2 and H3K27me3 can also form and be recruited by the variant PRC1 through H2AK119ub1. PRC2 subunits that bind to this modification are JARID2 and AEBP2, which subsequently stimulate H3K27me3 (Kalb et al., 2014). Loss of H2AK119ub1 leads to subsequent loss of PRC2 and H3K27me3 occupancy genome-wide. Variant PRC1 is probably recruited to CpG islands through recognition of non-

methylated CpG by the CXXC-zinc finger domain of KDM2B (Blackledge et al., 2014). It was also demonstrated that hypomethylated CpG islands are in fact sufficient to recruit PRC1 and PRC2 and that PcG recruitment is default to these regions; counteraction of polycomb recruitment at these sites is accomplished by histone H3 tail modifications (Cooper et al., 2014).

1.8 The role of polycomb in ESC pluripotency, differentiation and development

ESCs are derived from the inner cell mass of the blastocyst. They can differentiate into a broad spectrum of cells (pluripotency) and they can propagate continuously (self-renewal). OCT4, SOX2, and NANOG are key to the pluripotency of ESCs (Avilion et al., 2003; Chambers et al., 2003; Mitsui et al., 2003; Nichols et al., 1998) and they co-localise with PRC2 at developmental genes (Boyer et al., 2005; Lee et al., 2006). PcG function is particularly relevant in neural, hematopoietic and epidermal stem cell proliferation, self-renewal and differentiation (Aloia et al., 2013; Mohn et al., 2008; Molofsky et al., 2003).

It is fundamental that Hox and other developmental genes are coordinated in a correct spatial and temporal expression pattern over multiple rounds of mitotic cell division and differentiation, thus ensuring a correct formation of the body structure. Genome-wide analyses have revealed that PcG proteins are highly enriched at the promoter region of hundreds of developmental regulators in ESC that would otherwise induce cell differentiation (Azuara et al., 2006; Boyer et al., 2006; Lee et al., 2006; Mikkelsen et al., 2007; Negre et al., 2006). Hence, genes necessary for a specific cell type are expressed while unnecessary genes are repressed. ESCs have been highly used

for the study of polycomb group proteins given the relevance of these proteins in pluripotency and differentiation.

During embryogenesis, PcG proteins regulate anterior-posterior development (Simon et al., 1992). Mutations lead to ectopic expression of developmental genes and therefore to segment defects. Mutations in Eed, Ezh2, Suz12, Rybp and Ring1b (not Ring1A), even cause lethality by gastrulation arrest (Alkema et al., 1995). Knockout of Ring1B leads to premature differentiation of neural stem cells (Roman-Trufero et al., 2009), and knockouts of Jarid2 and Kdm2b cause improper neural development (Fukuda et al., 2011; Takeuchi et al., 1995). Mel18 or Bmi1 deficiency causes anterior-posterior defects at the axial skeleton plus immune deficiency, while simultaneous mutation of both components cause lethality (Akasaka et al., 1996; Alkema et al., 1995; van der Lugt et al., 1994).

1.9 Polycomb disregulation in cancer

Mutations as well as alterations in epigenetic modifications can result in repression of tumor suppressors or genomic stabilizer genes and lead to the formation of carcinogenic cells (Benetatos et al., 2013; Bracken and Helin, 2009; Richly et al., 2011).

Deregulation of PcG protein expression, particularly EZH2 overexpression or gain of function is correlated with progression of several types of cancer including prostate, lymphoma, breast, melanoma, bladder, gastric, and renal cancers and also correlated with poor prognostic (Simon and Lange, 2008; Velichutina et al., 2010). EZH2 is particularly important as it silences the expression of over 200 tumor suppressors (Simon and Lange, 2008). EZH2 harbor several heterozygous mutations at the tyrosine 641 of the C-terminal SET domain, which do not correlate with loss of function but rather an enhancement of function of H3K27me2 on its substrate

(Sneeringer et al., 2010; Yap et al., 2011) and are found in a low but significant percentage on a two different types of lymphomas (Morin et al., 2010). Another mutation found at a lower percentage in lymphoma cell lines is Alanine to Glycine mutation on position 677 of the EZH2 protein that has an enhancement of function as the previous described mutation but acts on H3K27, H3K27me1, and H3K27me2.

Glioblastoma is an aggressive form of cancer, with a high mortality rate, that arises from astrocytes and which accounts for 20% of all brain tumors in children (Saran, 2002). A frequent mutation in the glioblastoma tissue occurs in histone H3.3, which is incorporated during brain development, and consists of a replacement of lysine 27 by methionine (K27M). H3 Lysine 27 amino-acid is a common site for histone modifications such as the referred methylation and acetylation. Therefore, mutation at this position can lead to alteration in gene expression, in a way by inhibiting the enzymatic activity of PRC2 (Khuong-Quang et al., 2012; Lewis et al., 2013; Schwartzentruber et al., 2012).

In all types of endometrial stromal tumors (EST) the zinc fingers of SUZ12 and JAZF1 are fused. Such fusion occurs in stromal nodules too indicating that stromal proliferation from an initially benign tumor can develop into EST (Koontz et al., 2001). Also JAZF1-PCL1 fusions in ESTs. EZH2 also cooperates with other histone modifiers such as DNMTs (Vire et al., 2006), and HDAC (Tie et al., 2001; van der Vlag and Otte, 1999), which are found and linked to various types of cancer. Sites that are marked by H3K27me3 may become CpG methylated; DNA methylation sites are, to some extent, also marked by H3K27me3 and EZH2 (Fahrner et al., 2002; McGarvey et al., 2006; McGarvey et al., 2008). PRC1 has also been linked to cancer given that medulloblastoma cells overexpress BMI1 (Leung et al., 2004).

Targeting PcG proteins can be a strategy to reduce metastasis and increase effectiveness of chemotherapeutic drugs (Bracken and Helin, 2009; Crea et al., 2011; Croonquist and Van Ness, 2005; Kemp et al., 2012; Varambally et al., 2002). An example of a drug that depletes cellular levels of PRC2 core components and associated H3K27me3 is S-adenosylhomocysteine hydrolase inhibitor 3-Deazaneplanocin A (DZNep). This drug has low *in vivo* cytotoxicity (Bray et al., 2000), and induces apoptosis of cancer cells leaving normal cells alive (Tan et al., 2007). Another potential group of therapeutic drugs constitutes S-adenosyl-L-methionine (SAM)- competitive inhibitors as the SET domain contains a pocket for a SAM methyl donor (Copeland et al., 2009). It will be a clinical challenge to develop combined drugs to target methyltransferases (for histone and DNA) as well as HDACs.

1.10 Non-coding RNAs

Only around 2% of the mammalian genome is transcribed into coding messenger RNA (mRNA), and at least 90% of transcribed RNA is non-coding (ncRNA) (Bertone et al., 2004; Birney et al., 2007; Cheng et al., 2005; Djebali et al., 2012; Kapranov et al., 2002; Rinn et al., 2003), which means that it is not translated into proteins, giving rise to thousands of transcripts, most of which are uncharacterized. This phenomenon is known as 'pervasive transcription' and has been described in most eukaryotic organisms (Berretta and Morillon, 2009) (Mercer et al., 2009), with the transcriptome of human ESCs comprising around 10^5-$5x10^4$ transcripts (Brockdorff, 2013).

Comparing the amount of ncRNA with the amount of genes (~20,000-25,000) between organisms suggests that rather ncRNA may be critically important for determining the complexity of eukaryotes (Costa, 2010). Concordant with their wide variety, they have a myriad of functions many of which are regulatory at different levels

(transcription, post-transcriptional processing, and regulators of protein activity) (Britten and Davidson, 1969; Geisler and Coller, 2013; Jacob and Monod, 1961). ncRNA can be divided into small ncRNA (sRNA), smaller than 200 nt, and long ncRNA (lncRNA), longer than 200 nt. The diverse types of ncRNA include ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), small nucleolar ncRNA (snoRNA), micro RNA (miRNA), small interfering RNA (siRNA), piwi-interacting RNA (piRNA), and long non-coding RNA such as long intergenic RNA (lincRNA) and others. The former three types of RNA mentioned above can be described as classical non-coding RNAs. In eukaryotes about 80% of all RNA is rRNA; tRNA makes up 15% whereas mRNA accounts for only 5% (Warner, 1999). Lin-4 miRNA (Lee et al., 1993; Wightman et al., 1993), and XIST (Borsani et al., 1991; Brockdorff et al., 1992; Brown et al., 1992) were the first non-classical ncRNAs with assigned functions.

1.10.1 'Housekeeping' ncRNAs – rRNA, tRNA, snRNA

The central dogma of molecular biology views the existence of three types of RNA involved in protein synthesis: the mRNA, which is a copy of the DNA blueprint for the amino-acid sequence; the rRNA which forms a ribonucleoprotein complex of two subunits, called the ribosome, that bind together during mRNA translation; and the tRNA that transports the amino-acids to the ribosomes. Sequences of rRNA genes are highly conserved. There are four different types of rRNA in eukaryotes designated 28S, 18S, and 5.8S that result from the cleavage of a single long precursor (pre-rRNA), and 5S that is transcribed from a separate gene and is not processed extensively. In the nucleolus, pre-rRNA assembles with ribosomal proteins that are imported from pre-ribosomal particles. With the exception of 5S precursor rRNA, rRNA is cleaved twice, firstly to form 18S and 28S + 5.8S, then secondly to form the three different rRNAs.

Cleavage of pre-rRNA is accomplished by a complex of proteins with snoRNA (a subgroup of snRNA) that form small nucleolar riboprotein complex (snoRNPs). RNAPI transcribes this pre-rRNA from a single promoter in the nucleolus whilst 5S rRNA, together with tRNA and snRNA, is transcribed by RNAPIII in the nucleus. tRNA, like rRNA, is also processed by cleavage of pre-tRNA at the 3' and 5' ends by two different ribozymes. CCA sequence is then added to the 3' end and some bases are modified. snRNAs range in size from 50-200 nt. snRNAs, particularly the U-type snRNAs (U1, U2, U5, and U4/U6), are important regulatory molecules involved in pre-mRNA splicing. Splicing takes place in large complexes, called spliceosomes, composed of these snRNAs and proteins to form small nuclear ribonucleoproteins (snRNPs). For the splicing process, firstly U1 binds the 5'splice site (SS) consensus sequence of the pre-mRNA, followed by U2 binding to the branch consensus sequence point. The branch corresponds to a point where the 5'SS becomes ligated to the intron. Then a complex formed by U4/U6 and U5 joins the spliceosome as U1 is released and a lariat-like structure is formed with U5 binding the 3' SS and causing excision of the intron (Cooper, 2000).

1.10.2 MicroRNA and siRNA

MicroRNAs (miRNAs) belong to a class of non-coding RNAs that have a post-transcriptional gene expression regulatory function. miRNAs are single stranded RNAs (ssRNA) approximately 19-25 nucleotides in length and were first identified in the nematode *C. elegans* (Kim, 2005; Lee et al., 1993; Wightman et al., 1993), the founding members being *lin-4* and *let-7*. MiRNAs also occur in flies, mice and humans and hundreds of these RNAs exist (He and Hannon, 2004) that act in combination with RNA-binding proteins to bind to the mRNA 3'untranslated region (UTR), through base pairing, and consequently gene expression is silenced by inhibiting translation. With the

exception of miRNAs located on Alu-repetitive regions, which are transcribed by RNAPIII, primary miRNAs are transcribed by RNAPII to form a primary (pri)-miRNA molecule of hundreds to thousands of nucleotides long that is cleaved by the RNase III enzyme Drosha to form a double-stranded (ds) hairpin-shaped precursor miRNA (pre-miRNA; ~70-nt). pre-miRNA is transported to the cytoplasm of the cell and is cleaved again by another RNase III enzyme, Dicer (Bernstein et al., 2001; Bushati and Cohen, 2007; Ketting et al., 2001; Tijsterman and Plasterk, 2004). Subsequently, Argonaute2 (Ago2) is recruited and forms a trimeric ribonucleoprotein complex designated RNA-induced silencing complex (RISC) (Gregory et al., 2005; Liu et al., 2004; Maniataki and Mourelatos, 2005). RISC and its target mRNA accumulate in processing-bodies (P-bodies) where mRNA degradation happens (Liu et al., 2005; Pillai et al., 2005; Sen and Blau, 2005). There are suggested biological function of miRNA targeting in embryogenesis including germ layer formation, morphogenesis and organogenesis, control of developmental timing, neuronal patterning, haematopoietic cell differentiation, and cell proliferation and death (Brennecke et al., 2003; Chang et al., 2004; Chen et al., 2004; Johnston and Hobert, 2003; Lee et al., 1993; Pauli et al., 2011; Reinhart et al., 2000; Wightman et al., 1993; Yekta et al., 2004). Changes in the expression profile of miRNAs have been detected in tumors; indeed miRNAs can act as tumor suppressors such as miR-15a and miR-16-1or act as oncogenes such as miR-155 miR-17-92 (Calin and Croce, 2006). Another type of small RNA, siRNA, differs from miRNA in that it can have several different sources and it cleaves target mRNAs that are either endogenous or viral RNA (Forstemann et al., 2007; Tomari et al., 2007).

1.10.3 New classes of ncRNAs

1.10.3.1 Promoter-proximal short RNAs

Short RNAs around TSS of active protein-coding promoters, termed TSS-associated RNAs (TSSa-RNAs) or promoter-associated short RNAs have been identified, which are transcribed within 1.0-1.5 kb from the TSS and are smaller than 200 nt, some of them <22nt (Core et al., 2008; Kapranov et al., 2007; Seila et al., 2008; Taft et al., 2009). Such RNAs are predominantly associated with active and CpG islands promoters. Core *et al* study used a global run-on-sequencing (GRO-seq) assay, where a ribonucleotide analog is added to BrU-tag, to confirm that RNAPII is actively engaged in transcription (Core et al., 2008). Both studies found that short promoter-associated RNAs are transcribed in the sense and unexpectedly in antisense direction too (Core et al., 2008; He et al., 2008; Seila et al., 2008). TSSa-RNAs sense locate downstream of the TSS and peak at around +0 and +50 nt whereas antisense ones are detected upstream of TSS around -100 and -300 (Core et al., 2008; Seila et al., 2008). A relevant feature of these short RNAs is their highly significant association with CpG islands (Flynn et al., 2011; Seila et al., 2008). Another type of small RNA, identified up to 5 kb upstream of promoter and designated "PROMPTs" have been found to be very unstable (Preker et al., 2008). These transcripts are actively transcribed by RNAPII (Core et al., 2008) and are not produced by Dicer, as they remain present in dicer-/- ESCs (Seila et al., 2008).

Given that short RNAs have been discovered around transcription start sites (Core et al., 2008; Kapranov et al., 2007; Seila et al., 2008; Taft et al., 2009) and that polycomb target genes are often associated with H3K4me3 (Bernstein et al., 2006a) and RNA Pol II (Brookes et al., 2012; Stock et al., 2007), our laboratory hypothesized that short RNAs are also produced from repressed polycomb target genes. Using

microarrays, our laboratory identified short RNAs transcribed from the 5' region of protein-coding genes in human CD4+ T-cells. The previous studies (Core et al., 2008; Kapranov et al., 2007; Seila et al., 2008; Taft et al., 2009) showed that short RNAs are transcribed mainly from active genes, but the short RNAs identified in our laboratory are transcribed from both mRNA-producing and non-mRNA-producing genes, indicating that they are also transcribed from silent and polycomb-associated genes. This set of silent genes contained RNAPII that was positionally associated with the short RNAs.

1.9.3.2 Long non-coding RNAs (lncRNA)

The advent of microarray and next-generation sequencing technologies have allowed genome-wide identification of thousands of new transcripts (Carninci, 2009; Jacquier, 2009; Mattick et al., 2010). LncRNAs can be sense or antisense, and they can also act in *cis* or in *trans,* as already exemplified above. As far as known, most of lncRNAs are capped and polyadenylated and even though are not exported from the nucleus to the cytoplasm like mRNA but have no or little open reading frame (ORF) (Carninci et al., 2005). LncRNA show low sequence conservation except for a subset of these RNAs at exons (Brockdorff, 2013; Brosius, 2005; Struhl, 2007). This fact indicates that structure might be mainly responsible for ncRNAs activity rather than sequence. LncRNA control processes like imprinting, such as *Airn* (antisense to Igf2r RNA non-coding) and *Kcnq1ot1* (Kcnq1-overlapping transcript 1); dosage compensation as for *Xist* (X chromosome inhibition) (Nagano and Fraser, 2011; Wilusz et al., 2009); gene regulation, which can be in *cis* as it is the case of the lncRNA at the dihydrofolate reductase (*DHFR*) locus that binds to the *DHFR* promoter to inhibit its expression, or in *trans* it is the case of *HOTAIR* (HOX antisense intergenic RNA) (Pauli et al., 2011). They have also been linked to cell cycle regulation and pluripotency

(Nagano and Fraser, 2011; Wilusz et al., 2009) and can be developmentally regulated (Mercer et al., 2009). lncRNAs have also been found important in other cell types that not ESCs, for example the long non-coding RNA named Braveheart that is required for the cardiovascular lineage commintment (Klattenhoff et al., 2013).

A class of lncRNA is contained within intergenic regions and therefore designated large intergenic RNA (lincRNA). LincRNAs can have a myriad of functions (Geisler and Coller, 2013; Guttman et al., 2009; Guttman et al., 2011; Huarte et al., 2010; Hung et al., 2011; Orom et al., 2010; Wang et al., 2011; Wilusz et al., 2008). They exhibit a chromatin profile identical to active genes, thus being transcribed from H3K4 and H3K36 methylated regions (Guttman and Rinn, 2012). Some of them have enhancer-like function (Orom et al., 2010; Wang et al., 2011), and at least 30% of lincRNAs interact with one of multiple chromatin-regulatory complexes (Guttman et al., 2011). These RNAs have suggested functions in ESC pluripotency and differentiation (Guttman et al., 2011; Pauli et al., 2011), for instance the lincRNA-RoR is a regulator of reprograming of cells into induced pluripotent cells (iPSCs) (Loewer et al., 2010). lncRNAs have also been found important in other cell types than ESCs, for example the long non-coding RNA named Braveheart that is required for the cardiovascular lineage commitment (Klattenhoff et al., 2013).

1.10.4 Polycomb-RNA interactions

An important characteristic of RNA is that not only can it interact with protein complexes but it can also bind to chromatin bringing the possibility of ncRNA functioning to target protein complexes to specific genomic sites (Guttman and Rinn, 2012). Although there is no evidence that ncRNAs are involved in the targeting of PcG

proteins in Drosophila, RNA may form a linkage between PcG and chromatin in mammals (Beisel and Paro, 2011). Recruitment of polycomb complexes to mammalian target genes is still not well understood. However, mammalian PRC2 has been reported to interact with a wide range of lncRNAs (Rinn and Chang, 2012; Wang and Chang, 2011) derived from intergenic and intragenic regions, and which have been suggested to be PcG recruiters (Pandey et al., 2008; Rinn et al., 2007; Zhao et al., 2008).

Central to X-inactivation in female mammals (a process indispensable for chromosome dosage compensation) is the *cis* acting ~17 kb X-inactive-specific transcript (XIST) that accumulates on the inactive X-chromosome (Xi) (Brockdorff et al., 1992; Brown et al., 1992; Plath et al., 2003). The first indication of a correlation between PRC2 and the Xi was the observation of an enrichment of EED/EZH2 and H3K27me2/3 on the Xi territory (Duthie et al., 1999; Mak et al., 2002; Plath et al., 2003; Silva et al., 2003). Further experiments confirmed a correlation between XIST expression and PRC2 recruitment (Kohlmaier et al., 2004; Mak et al., 2004; Plath et al., 2003). XIST contains a structure called RepA (A-repeat region) embedded within XIST intron 1. Rep A consists of A-rich spacer sequence followed by a 26-nt GC-rich core sequence forming a double stem-loop that repeats itself 7.5x to form a structure of 1.6 kb (Brockdorff et al., 1992; Brown et al., 1992; Zhao et al., 2008). PRC2, particularly EZH2, as shown by EMSA, binds to RepA through a conserved double stem-loop (Zhao et al., 2008). However, the interaction is not very specific given that Ezh2 can also interact with antisense RNA. Furthermore, RepA alone is not the key recruiter of PRC2 as RepA-lacking XIST can also recruit PRC2 (Kohlmaier et al., 2004). The localization domain of XIST, RepC, binds to YY1. YY1 also tethers to this non-coding RNA, bringing it to X nucleation centre (Jeon and Lee, 2011). Inactive X chromosome (Xi) is largely occupied by H3K27me3, and monoubiquitylation of H2A has also been linked to Xi (Bernstein et al., 2006b; de Napoles et al., 2004; Smith et al., 2004). Members of PRC1 have been shown to associate with Xi including most of the CBX

members (not CBX4) (Bernstein et al., 2006b; de Napoles et al., 2004; Hernandez-Munoz et al., 2005; Plath et al., 2004). CBX7, is particularly enriched in the inactive X chromosome (Bernstein et al., 2006b). Recently, the chromatin remodeller ATRX has been demonstrated to directly interact with RepA/Xist, which is required for PRC2 loading on XIST (Sarma et al., 2014).

Another PRC2-interacting ncRNA is the 2.2-kb HOTAIR (Hox antisense intergenic RNA) located on human chromosome 12, which acts in *trans* (Gupta et al., 2010; Rinn et al., 2007). HOTAIR originates from the *HOXC* locus but appears to function to repress the *HOXD* cluster region on chromosome 2 by interacting with PRC2, through its 5'-end, and adding H3K27me3 marks across a 40kb region (Rinn et al., 2007). How HOTAIR is directed to its targeting sites is unknown. *HOTAIR* is over expressed in some primary tumors; this overexpression induces genome-wide retargeting of PRC2 and can be indicative of metastasis and poor prognosis (Gupta et al., 2010; Kogo et al., 2011). This ability of HOTAIR to confer cellular invasiveness is specifically dependent on PRC2 (Gupta et al., 2010). The HOTAIR 3'-end has been found to interact with LSD1, a component of CoREST/REST that has demethylase activity against H3K4 (Tsai et al., 2010). It seems that lncRNAs may contain multiple binding sites for different proteins, thus directing and combining different histone modifications. Similar to HOTAIR, thousands of other lncRNAs have been identified in different organisms that associate with chromatin remodeling complexes, including PRC2, and this interaction is required for their function (Khalil et al., 2009).

Another PRC2-interacting lncRNAs is the 91kb transcript from the imprinted potassium voltagegated channel, subfamily Q, member 1 (*Kcnq1*) cluster, termed Kcnq1 overlapping transcript 1 (Kcnq1ot1), located on mouse chromosome 7. In this case, the cluster of paternal alleles is repressed by the expression of Kcnq1ot1, and deletion of the Kcnq1ot1 promoter results in loss of imprinting of most genes of that cluster (Fitzpatrick et al., 2002). Kcnq1ot1 requires direct interaction of PRC2 with Kcnq1ot1

ncRNA for spreading of H3K27me3 throughout the locus (Pandey et al., 2008).

BMI1 and CBX7 have been reported to repress the *INK4b/ARF/INK4a* locus, a master regulator of cellular senescence in case of stress (Bernard et al., 2005; Gil et al., 2004). This repression is mediated by ncRNA ANRIL (antisense non-coding RNA in the INK4 locus), which binds to the CBX7 subunit of PRC1 through its chromo domain, and causes repression in *cis* (Yap et al., 2010). ANRIL is an antisense noncoding RNA that spans along 30–40 kb of the *INK4b/ARF/INK4a* locus (Pasmant et al., 2007). Human MOV10, a putative RNA helicase previously implicated in post-transcriptional gene silencing, associates with CBX7 and is required for repression of *Ink4a/Arf* and the presence of H3K27me3 (El Messaoudi-Aubert et al., 2010).

In this introduction I have reviewed the involvement of the polycomb group of proteins in animal development, their function and what is known about their mechanism of action. It is still poorly understood how these proteins interact with chromatin despite the advances in recognizing ncRNA as an intermediate. It is not known how these RNAs are regulated, and how specifically they interact with polycomb proteins. Although the functions of a subset of lncRNAs have been identified and correlated with changes in gene expression, the majority of ncRNAs have no assigned function, opening the possibility of new functions potentially linked with protein binding. It is also being considered how environmental factors affect such RNA functionality and regulation.

Aims of this project

PREs are well defined in Drosophila allowing for their prediction in this organism but not so well characterized in mammals due to lack of consistency found in this sequence. However, it has been observed that they contain PRC2-interacting CpG island and that PRC2 also interacts with RNA, leading to the hypothesis that these might constitute features of mammalian PREs.

Some genes are regulated at the level of elongation and poised for activation. As RNAPII and transcription positive marks are associated with polycomb-targeted genes, I hypothesised that short RNAs are transcribed from DNA regions around the TSS of many genes regulated by polycomb, and that these same regions might behave a PREs like in *Drosophila*.

This project has two principal aims. Firstly, to investigate whether short RNAs are transcribed from repressed polycomb target genes and whether this is a result of polycomb activity. Secondly, to investigate whether the DNA regions that produce short RNAs behave as PREs.

Chapter 2 – Methods

2.1 RNA purification and fractionation

Total RNA was purified using TRizol was as follows: $1x10^7$ pelleted cells per millilitre of TRizol were dissolved by pipetting and vortexing, while lysate was incubated for 5 min at room temperature (RT). 0.2 ml of chloroform was added, the lysate shaken by hand for 15 seconds and incubated for 5min at RT. The organic and aqueous layers were then separated by centrifugation at 12,000g for 15 min at 4ºC. 0.5 ml of the top aqueous layer was pipetted into a fresh tube and 0.5 ml of isopropanol was added. This was vortexed and RNA allowed to precipitate by incubating at RT for 10 min, then pelleted by centrifugation at 12,000g for 15 min at 4ºC. The supernatant was discarded and the pellet washed by adding 1ml of 75% Ethanol, followed by vortexing and centrifuging at 7500g for 10min at 4ºC. Finally, pellet was dried for 5 min at RT and dissolved in RNase-free water (Ambion). When required, total RNA was fractionated into long and short (≤200 nt) using a mirVana miRNA kit (Ambion) by following the manufacturer's instructions. Purified RNA was quantified using a nanodrop.

For short RNA purification using RNAzol, $1x10^7$ cells were dissolved in 1ml of RNAzol, vortexed and 0.4 ml of double distilled water added into the lysate. The solution was mixed by manual stirring for 15 seconds and incubated for 10 min at RT. Cell debris was pelleted by centrifuging at 12,000g for 15min at 4ºC, and 1 ml of the aqueous phase, containing total RNA, was transferred to a new tube. 0.4 ml of 75% ethanol was added, solution was vortexed, incubated for 10 min, and centrifuged at 12,000 g for 8 min at 4ºC to precipitate long RNA. 1ml of supernatant containing short RNA was pipette into a new tube and stored at -20ºC. The long RNA fraction was washed by adding 0.4 ml of 75% ethanol, vortexed and

centrifuged 8,000 g for 3 min. Supernatant was discarded and the pellet washed a second time. Residual supernatant was removed with a micropipette and RNA resuspended in RNase-free water without drying. For shrt RNA purification, 0.8 ml of isopropanol added to the samples, vortexed and precipitated at -20ºC for 30 min, then centrifuged at 12,000 g for 20 min. The short RNA pellet was washed twice in 0.4 ml of 70% isopropanol, centrifuged at 8,000 g for 3 min, and re-suspended in water the same way as long RNA.

RNA was treated with DNase-turbo (Ambion) at 37ºC for 30 minutes. For total purified RNA, DNase was stopped by addition of $1/10^{th}$ volume of terminator mix (0.1M EDTA pH 8.0, 1 mg/ml glycogen). Total and short DNased RNAs were purified by ethanol precipitation at -20ºC for 30 min, then centrifuged and washed as described above. Total, long, and short RNAs were examined with an Agilent Bioanalyzer to confirm RNA quality and fractionation size. Short RNA was purified from PBMCs, neurons and ES cells with TRizol and from MSC and drug-treated CEM cells with RNAzol.

In a test experiment to analyse differences between the two methods, RNA was purified in parallel from CEM cells with TRizol and RNAzol. Equivalent amounts of each fraction purified with RNAzol were also pulled together and re-fractionated using the mirVana miRNA kit.

2.2 Northern blotting

5 μg of DNase-treated short RNA was mixed with loading buffer (Ambion) and heated to 95ºC for 5min, and then loaded into 15% acrylamide-7M Urea TBE Novex gels (Invitrogen). Denatured RNA was resolved in 1X TBE in parallel with radio-labelled Century and Decade markers (Ambion) for 1hr at 200V. The RNA

was electroblotted to a Nytran Supercharge membrane (Whatman) for 1.5hr with a starting current of 200mA. 0.5X TBE was used as the running buffer. Fixation to the membrane was performed by UV-crosslinking using 1200mJ, followed by baking it for 1hr at 80ºC. The 49-nt northern probes used to detect the short RNAs in human cells (PBMCs, SH-SY5Y) were designed based on the 60-nt microarray probes (Table 3). The mouse probes were chosen based on homology to the human microarray probes. Probes to detect novel short RNA in the HOXD11.12 PRE were designed across the entire element (Table 4) based on the DNA sequence extracted by Woo *et al*. (2010). The probes were produced by Integrated DNA Technologies, are 49 nucleotides long, and contain a 3' StarFire extension system for labelling. Radioactive labelling was carried out according to manufacturer's instructions, and MicroSpin G-25 (GE HealthCare) columns were used to purify labelled probes. Before probing, membranes were pre-hybridised in UltraHyb buffer (Ambion) at 35ºC for 30 min. Labelled and purified probes were added into the buffer, and left to hybridize at 35ºC for 16hr. The washes were performed with a solution composed of 2X SSC and 0.5% SDS, and consisted of three washes at 35ºC for 5min each with a fourth wash at 42ºC for 5-10min. The membranes were then exposed to a phosphor screen for at least 48hr. The phosphor screen was scanned using a STORM phosphoimager (Molecular Dynamics). Blotting for the murine knockout cell lines was performed with the help of lab colleague Aditi Kanhere.

2.3 Cell culture

Mesenchymal stem cells

Immortalised Mesenchymal Stem Cells (MSC) (Funes et al., 2007) were cultured at 37°C and 5% CO2 in MesenCult® MSC Basal Medium (Human) with 10% MesenCult® Mesenchymal Stem Cell Stimulatory Supplement (Human) until ~60% confluency. Human primary MSC have limited proliferative potential *in vitro*.

Neuronal cells

SH-SY5Y cells were cultured in DMEM with 10% FCS until 40-50% confluent and then terminally differentiated in DMEM with 5% FCS and 10μM retinoic acid (Sigma) for 7 days.

PBMC purification

Peripheral blood mononuclear cells were isolated from standard buffy coat by gradient centrifugation using lymphoprep. Blood was diluted 1 in 4 HBSS. Lymphoprep and HBSS were warmed to 37ºC. 15 ml of Lymphoprep were added to six 50 ml Falcon tubes and 35 ml of diluted buffy coat was carefully layered on top and centrifugated at 1600 rpm, brake zero, for 30 minutes. Interphase layer was removed into a new tube. Then spun at 1600rpm for 10 minutes, brake nine; supernatant discarded and pellet gently resuspended in 50 ml HBSS.

CEM cells

CEM were grown in RPMI-1640 medium supplemented with 10% Fetal calf serum (FCS) and 1% streptomycin/ Penicillin.

Murine ES cells

The growth of Ezh2 knockout cell lines (Ezh2-1) (Landeira et al., 2010; Su et al., 2003) was performed by Filipe Pereira and Cynthia Fisher, while the growth of Ring1B knockout cell lines (ES-ERT2) (Endoh et al., 2008; Stock et al., 2007) was carried out by Emily Brooks. Murine ES differentiation into motor neurons (Wichterle et al., 2002; Wichterle and Peljto, 2008) was accomplished by Warren Whyte.

2.4 Quantitative reverse-transcription PCR

Total RNA was purified from the mouse ES cell differentiation stages, Ring1B and Ezh2 knockout cell lines, and drug treated CEM cells, and treated with DNase turbo. Samples were then reverse transcribed with SuperScriptII (Invitrogen) primed with oligo-dT or random primers for RNA extracted from CEM cells for the RNAPII inhibition study described below. Controls with no reverse transcriptase were performed to ensure that qPCR products are cDNA derivates. Samples were subjected to SYBR green quantitative PCR (Qiagen) containing 5-15 ng cDNA and a primer concentration of 300nM. Change in the expression of Ybx2, Msx1, Hes5 and Pcdh8, in the murine cell lines, was calculated relative to day 0 and normalized to Actin using the formula 'gene of interest'/Actin. The amount of nascent Actin RNA in the RNAPII inhibition experiment was measured by comparing the Ct to that of 5S rRNA using the formula $2^{(Ct(5S\ rRNA) - Ct(\beta\text{-Actin}))}$.

2.5 Chromatin immunoprecipitation

The Chromatin immunoprecipitation (ChIP) protocol followed that of Lee and colleges (Lee et al., 2006). MSC were grown as described above and crosslinked by addition of 1% formaldehyde to the culture medium for 10 minutes at room temperature. Formaldehyde was quenched with 0.125M of glycine, the cells washed twice with cold PBS, scraped from the plate surface, washed twice with cold PBS and then flash frozen and stored at -80ºC. Cells were lysed by re-suspension in lysis buffer 1 (50mM Hepes KOH pH7.5, 40mM NaCl, 1mM EDTA, 10% glycerol, 0.5% IGEPAL CA-630, 0.25% Triton X-100) rocked for 10 min at 4ºC, centrifuged and the entire process repeated with lysis buffer 2 (10mM Tris pH 8, 200mM NaCl, 1mM EDTA, 0.5mM EGTA). Extracted nuclei were re-suspended in buffer 3 (10mM Tris pH 8, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauryl sarcosine) and sonicated on ice at 24W for 5 minutes total (pulses of 30s separated for gaps of 1 minute) with a Misonix Sonicator 3000. Complete protease inhibitor (Roche) was added to each buffer. 100 µl Dynal Protein G magnetic beads were coated with 10 µg of respective antibody by 8 hr incubation at 4ºC, and washed. Antibodies used were: H3K27me3 (Abcam ab6002), H3K4me3 (Abcam ab8580), and H3 (Abcam ab1791). Sonicated cell lysate (whole cell extract) from 5x10^7 MSCs was immunoprecipitated overnight with the antibody-coated beads at 4ºC. 100 µl of whole cell extract was preserved to be used as a control. Beads were washed 5 times with RIPA buffer (25mM Tris-HCL, 150mM NaCl, 1%NP-40, 1% sodium deoxycholate, 0.1% SDS) and once with TE containing 50 mM NaCl. To elute the immunocomplex from the beads these were incubated at 65ºC for 1hr with vortexing every 10 minutes, and then pelleted. The supernatant was then left incubating for an additional 6 hr in order to reverse the

protein-DNA crosslinks. Control whole cell extract was also reverse crosslinked. Samples were then treated with 8 µl RNAseA (10 µg/µl), and 4 µl proteinase K (20 µg/µl). Immunoprecipitated and whole cell extract DNA were purified by phenol:chloroform:isoamyl alcohol phase separation and ethanol precipitation and then quantified using a nanodrop. Quantitative PCR (qPCR) was performed using the SYBR green PCR (Qiagen) system. Samples contained 5-15 ng cDNA and a primer concentration of 900 nM for HOXD11.12, 600 nM for HOXC8, and 300 nM for Actin primers. To quantify ChIP vs input DNA, serial 1/10 dilutions of whole cell extract (WCE) were prepared to construct a standard curve.

2.6 RNA secondary structure

The stem-loop-stem structure motif in RepA interacts with PRC2 (Zhao et al., 2008). Using the RepA RNA structure as a model, prediction of RNA structures was performed with the help of Aditi Kanhere. Structures were identified within 200nt of sequence surrounding probes that detected short RNAs using RNAmotif (Macke et al., 2001). Free-energy structures were predicted using RNAfold (Hofacker and Stadler, 2006).

2.7 RNA polymerase II inhibition

CEM cells were grown to confluency ($1x10^6$ cells/ml), and $1x10^8$ cells were re-suspended in 100 ml of medium per treatment. Cells were treated with 10 µM Flavopiridol (4.56mM stock concentration; Invitrogen), or 5µg/ml ActinomycinD (10mg/ml stock concentration; Sigma) for 10 hours. A control sample received no drug treatment. One hour after adding the drugs, 2ml of cell suspension were aliquoted to use as a control for drug inhibition of RNAPII by stimulating the cells to

differentiate, which was accomplished with 50 ng/μL of PMA and 500 ng/ μL of Ionomycin added into one ml of each sample, the other one ml was used as a control. Therefore, after the 10 hours treatment, control samples for drug inhibition were stained with anti-CD69-PE antibody (Miltenyi). First cells were washed in PBS containing 4% FCS, 5 μL antibody were added and incubated for 15 min at 4ºC in the dark. Cells were crosslinked with 2% formaldehyde and washed in PBS before being analysed by flow cytometry.This will allow the drug to inhibit for the transcription of mRNA including CD69. After drug treatment, test cells were centrifuged at 12,000rpm for 5 min and resuspended in RNAzol for a final concentration $2x10^7$ cells/ml.

2.8 Cloning

The HOXD11.12 PRE was identified from Woo *et al* (2010), and potentially new PRE sequences were identified from regions with short RNAs flanked by H3K27me3 peaks. Primers underlying the H3K27me3 peaks were designed to amplify these regions by PCR (Table 2). The sequences between each of the two peaks were amplified from genomic DNA from human PBMCs. Pfx50 or Pfu DNA polymerases (Invitrogen) were used because they produce fragments with blunt ends. The fragment sizes were confirmed by gel electrophoresis, then purified and ligated with T4 ligase (Promega) into the pCR-Blunt Vector (Invitrogen). A ligation control was performed with water. The ligation products where transformed into TOP10 *E. coli* competent cells (Invitrogen) by heat shock. Cells were grown in Luria-Bertani broth (LB) at 37ºC overnight, and the plasmids were extracted with Miniprep Kit (QIAgen). The presence of insert and orientation was checked by sequencing. The subcloned inserts were separated from the pCR-Blunt vector by digestion with XbaI for MSX1, SpeI and XbaI for YBX2, and SpeI

and NheI for RASL12, all have compatible overhang ends. To linearise the luciferase vector, YY1pLuc was digested with NheI (Promega). Following 2hr digestion, the vector was treated with CIAP (Promega) to avoid self- ligation. Presence and orientation of insert were checked by sequencing. Ligation, transformation and plasmid purification were performed as above.

2.9 Deletions in D11.12 PRE

The northern blot probes that detected short RNAs corresponded to a DNA sequence of 150 base pairs. To delete this short RNA sequence, the region of HOXD11.12 upstream of the short RNA area (5'D11.12) from HOXD11.12 YY1pLuc was amplified to produce a fragment with recognition sites for SacI at one end of PCR product and NheI at the other end. The 5'HOXD11.12 fragment and empty YY1pLuc vector were digested with these enzymes, ligated together, and the resultant clone linearized with AgeI and NheI. The HOXD11.12 region downstream of short RNA (3'HOXD11.12) was amplified with primers enclosing recognition sites for AgeI and NheI, and ligated into the vector containing the 5'HOXD11.12. HOXD11.12 with deleted short RNA sequence (ΔshortRNAD11.12) was confirmed by sequencing. HOXD11.12 PRE with deleted conserved region was obtained from Woo *et al* (2010). Disruption of short RNA secondary structure was performed using site directed mutagenesis, following the QuikChange protocol (Stratagene) (Liu and Naismith, 2008). According to this method, forward and reverse primers align only for 17 nt (primer-primer complementary sequence; pp) and the non-overlapping (no) comprises a longer sequence that requires a higher melting temperature. A mutation was created for each first and second short RNAs of HOXD11.12 consisting of a 10nt long change in DNA sequence located on the pp region of each primer, designed using RNAFold to search for a mutation that would disrupt the secondary structure. 30 ng of template DNA (D11.12pYY1Luc) were used

in a 50 μL reaction with Pfx Accuprime polymerase. PCR cycles used for the first and second mutation were as follows: 15 times (95ºC, 1min; Tm no -5ºC, 1min; 68ºC, 8min) plus Tm pp -5ºC, 1min and 68ºC, 30 min. 10 μL of PCR reaction plus loading buffer were ran by gel electrophoresis (0.8% agarose) using template plasmid as a control to confirm amplification. After the first cycle and before the second one the plasmid was cloned into TOP10 *E. coli* and isolated. Samples were sequenced to check for any unwanted mutation, and subsequently miniprepped.

2.10 Luciferase assay

The following procedure is adapted from the work of Woo *et al.* (2010). The parental pTranslucent (pLuc) (Panomics) expresses a firefly luciferase reporter gene from a Herpes simplex virus thymidine kinase promoter, but has low expression levels (Woo et al., 2010). In order to promote luciferase expression, another vector was used with an YY1 enhancer sequence introduced upstream of the promoter (YY1pLuc). New PREs and D11.12 PRE were inserted immediately upstream of the YY1 enhancer. Transfected constructs were as follows: pLuc, YY1pLuc, D11.12 YY1pLuc, ΔshortD11.12 YY1pLuc, and control, pIRESneo3 LTR C20orf112mut Luc, a control previously used in our lab $3x10^3$ cells were seeded into each well of a 96-well plate. Cells were transfected with Fugene HD (Roche) 12 hrs later using a 6:2 Fugene:DNA ratio and adding to each well 75% of standard volume. The Renilla luciferase plasmid (pRL-TK) (Promega) was used as the assay control. Both luciferases were measured with the Dual-Luciferase Reporter Assay System (Promega) 48 hr post-transfection. Results were first normalized to pRL-TK and then further normalized so that pLuc RLU represented 0% activity and YY1pLuc 100%.

Chapter 3 – Regulation of short RNA transcription at polycomb target genes

3.1 Introduction

The small RNAs identified by Seila *et al*. (2008) and Core *et al*. (2008) were detected in ESCs and cell lines. Our laboratory also detected thousands of short RNAs from promoter, introns, and exons of protein-coding genes of human CD4+ T-cells indicating that production of small RNAs is a common feature of the transcriptome of somatic cells (Kanhere et al., 2010). Short RNAs detected in our laboratory are concentrated at 700 bp either side of the TSS, are 50-200 bp long, are from sense strand, and some of them were located on the same genes as TSSa-RNAs but on different positions. Furthermore, they could be detected from both active and silent genes equally, whereas only small RNAs have been previously detected only at active genes.

Concordant with production of short RNA, RNAPII was located at the site of these short RNAs. H3K4me3 was also found present at the same location as RNAPII on these silent genes, with H3K27me3 peaks flanked by RNAPII and H3K4me3. These features correlate with described transcriptional pausing, a common mechanism of regulating polycomb-target genes. These results show that these short RNA-producing silent genes are targets of polycomb. In *Drosophila,* Polycomb often targets TSSs with a stalled RNAPII and is linked to the production of short RNAs (Enderle et al., 2011; Kharchenko et al., 2011).

Using RepA – the Xist RNA element responsible for PRC2 binding – as a positive control, our laboratory utilized electrophoretic mobility shift assay (EMSA) and RNA immunoprecipitation (RNA IP) to successfully demonstrate that short RNAs also interact with PRC2. Furthermore, our laboratory has identified SUZ12 as the binding subunit, and that binding is dependent on the secondary structure of the short RNAs. SUZ12 was not observed to bind to the RepA sequence encoded by ssDNA,

dsDNA or a RNA-DNA duplex. This interaction was also true in living cells and when incorporating the short RNA in the HIV LTR and submitting it to luciferase activity it lead to gene repression in *cis*.

Given the production of short RNAs at paused polycomb targeted genes, and that PRC1 inhibits RNAPII elongation through H2A ubiquitination, we hypothesized that short RNAs are a by-product of polycomb activity and therefore, dependent on H3K27me3 and H2AK119ub. Furthermore, polycomb-tarfeted neuronal genes constitute a subset of polycomb targets, which means that as ESCs differentiate to primary motor neurons (PMN) and the genes becomes activated short RNAs should resolve.

3.2 Results

3.2.1 Purifying different lengths of short RNA

The RNAs identified in our laboratory are short in length. RNA purification begins with total RNA extraction, followed by fractionation which can be accomplished by two methods: TRIzol or RNAzol. In order to assess whether TRIzol and RNAzol methods of RNA purification are equivalent, RNA was purified from a CEM cell line using the two methods in parallel. TRIzol purifies total RNA, which then requires fractionation using the mirVana system, while RNAzol purification separates short (≤ 200 nt) and long RNA fractions in one step. To better address any possible difference between the two methods, RNA was purified with RNAzol alone, with TRIzol an mirVana and thirdly, by pooling equal proportions of the short and long RNA fractions purified by RNAzol back together and re-fractionated using the mirVana system. Northern blotting for short RNA from *HOXC6* has been shown to produce short RNA in T-cells and MSC. This blotting revealed that the short RNA fraction purified directly with RNAzol lacks the ~200nt short RNA that are present in RNA purified with TRIzol and mirVana. This RNA can be observed when

long and short RNAzol fractions are pooled together with long RNA and re-fractionated through total RNA purification with the mirVana system (Figure 6). Therefore, ~200 nt RNAs fractionate into the long RNA fraction upon purification via RNAzol. It was consequently determined that TRIzol and mirVana were most suitable for subsequent experiments as it was possible to maintain a greater ranger of short RNA species.

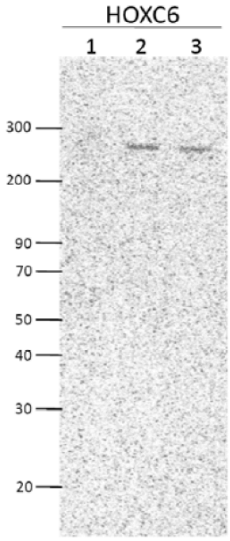


Figure 6: Test for the fractionation of small RNA by different methods. Northern blotting in CEM cell line to compare two different methods of RNA purification and short RNA fractionation: A two step method involving first total RNA purification (with TRIzol) and then fractionation (with mirVana); and the other method that purifies and fractionates in a single step (with RNAzol). Lane 1 – short RNA purified with RNAzol, Lane 2 – short RNA purified with RNAzol, the two fractions were pooled together and re-purified with the mirVana system, Lane 3 – short RNA purification with TRIzol and the mirVana system.

3.2.2 Short RNAs transcribed from silent genes are conserved between human and mouse

Our laboratory has previously detected short RNAs in primary resting $CD4^+$ T-cells. It was first verified whether short RNAs could indeed be produced from the promoter region of silent genes in mouse ESCs. Moreover, conservation is a sign of functional significance. To analyze this, H3K27me3 and PRC2 ChIP-Chip data from Boyer *et al*. (2006) were used to identify genes targeted by Polycomb in murine ES cells. These set of genes was compared to the set of genes producing short RNAs

identified in T-cells, to select for common genes silenced in both cells types and to identify short RNA sequences conserved between human and mouse (performed by laboratory colleague Aditi Kanhere). Probes for northern blotting were designed to detect murine *Pcdh8, Ybx2, Hes5* and *Msx1* short RNAs. The blots showed short RNA bands for a11 these genes, therefore I have demonstrated that they are also transcribed in murine ES cells (Figure 7A and B). This supports strong evidence that production of short RNAs at Polycomb targets is conserved between these mammalian species and between ES cells and T cells.

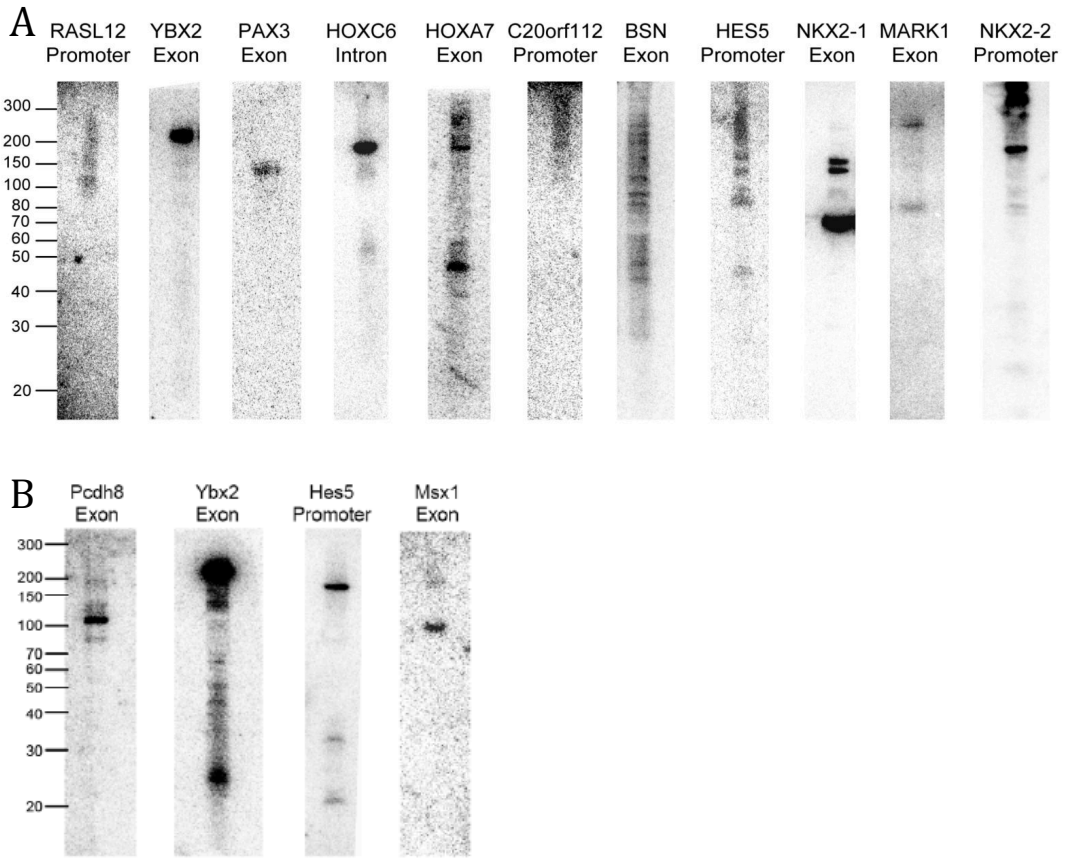


Figure 7: Short RNA is conserved between human and mouse. (A) Northern blotting for detection of short RNAs extracted from human CD4+ T-cells. *RASL12, YBX2, PAX3, HOXC6, HOXA7, C20orf112, BSN, HES5, NKX2-1, MARK1,* and *NKX2-2* are repressed in this cell type. The size of markers is shown on the left side in nucleotides. Short RNAs are sized around 200–50 nt on the probed genes. (B) Northern blotting for detection of short RNAs extracted from wild-type murine ES cells. *Pcdh8*, *Ybx2, Hes5* and *Msx1* genes are repressed in murine ES cells. The size of markers are shown on the left side in nucleotides. RNAs are sized around 100 nt on exons of Pcdh8 and Msx1; and around 200 nt for Ybx2 exon and Hes5 promoter. Short RNA production is therefore shown to be conserved from mouse to human.

3.2.3 Short RNA transcription is independent on polycomb activity

Given that Polycomb-repressed genes contain stalled RNA polymerase, due to ubiquitination of H2AK119, short RNA might be a by-product of PRC2 repressive activity. To evaluate the validity of this assumption, further work was carried out in collaboration with Amanda Fisher's group. They have generated a murine ES cell line, Ezh2-1.3, that uses a tamoxifen-inducible cre-lox system to create a null mutation to delete the catalytic SET domain of Ezh2, which was subsequently cultured in the presence of tamoxifen for five days by Filipe Pereira.

To confirm that the inducible knockout is effective, H3K27me3 and Ezh2 were analysed for five days after tamoxifen treatment by western blotting (performed by Keijo Viiri). As indicated in the western blot, the Ezh2 SET domain is abolished one day after addition of tamoxifen and a truncated form of Ezh2 appears, which is accompanied by loss of H3K27me3 (Figure 8A). In contrast, blotting for short RNAs showed that short RNAs do not show a change in abundance four to five days after adding tamoxifen (Figure 8B). 5S rRNA was used as loading control. In order to confirm that short RNA transcription are also unaffected by loss of PRC2 function, JARID2$^{-/-}$ murine knockout from Fisher's laboratory was utilized (Figure 8C). Northern blotting revealed that short RNAs are also transcribed from Msx1 and Hes5 in these cell lines, consistent with the results from the Ezh2 deletion experiment. These results demonstrate that production of short RNA is indeed independent and upstream of H3K27 methylation.

H2A ubiquitination has been suggested to be involved in polymerase stalling (Stock et al., 2007), implying that this might be the polycomb component responsible for the short RNA production. H2A ubiquitination is catalysed by the PRC1 subunit Ring1B. A murine ES cell line, Es-ERT2, was used to test the dependency of short RNAs on H2AK119ub. Es-ERT2 contains a tamoxifen-inducible conditional knockout of the Ring1B and is also homozygous null for the functional homologue Ring1A. Following addition of tamoxifen, cells are progressively depleted of H2Aub but preserve the overall levels of PRC2 and H3K27me3 (Endoh et al., 2008; Stock et al., 2007). These cells were grown by Emily Brooks (MRC CSC). Blotting for short RNAs at *Hes5, Msx1, and Ybx2*, showed that loss of Ring1b had no effect on short RNA transcription (Figure 8D). Es-ERT2 and Ezh2-1.3, cells could have been treated with tamoxifen for more than 3 or 5 days respectively; however, longer treatments induce cell differentiation because of the activation of differentiation-associated genes. This indicates that short RNAs are transcribed independently of Ring1B activity and independently of the H2Aub-mediated block to elongation and are not degraded within the time-frame of tamoxifin induction.

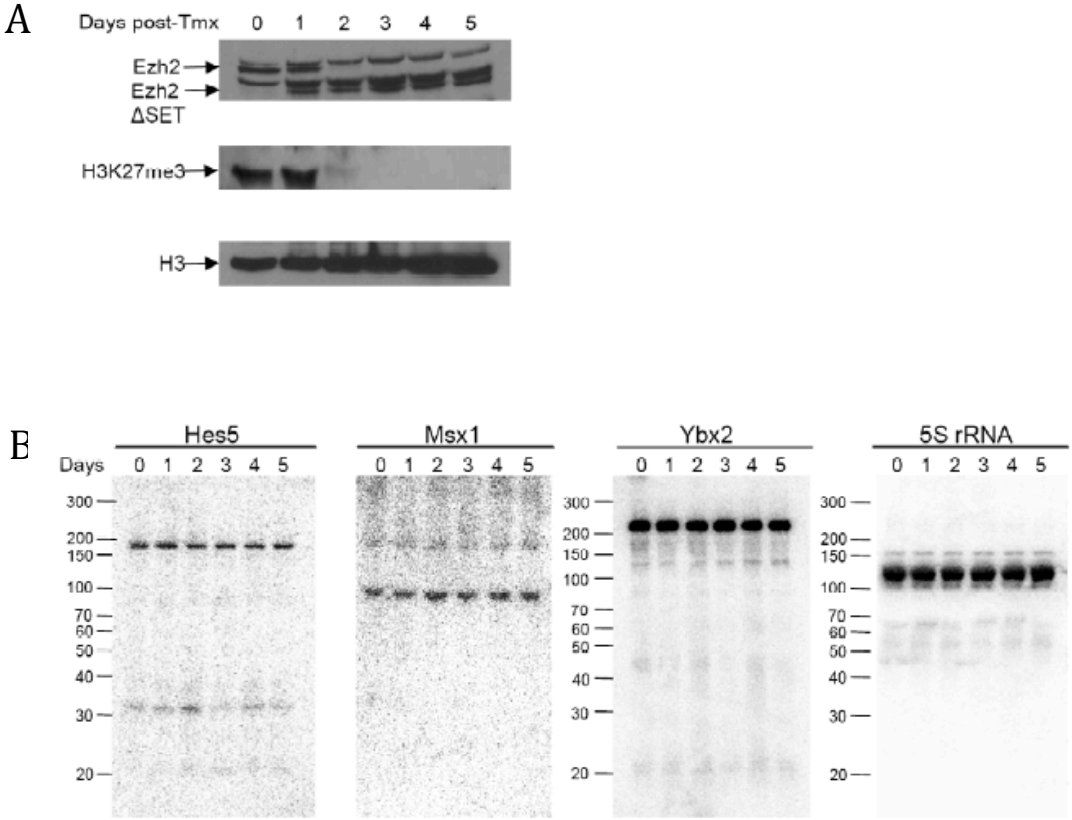


Figure 8: Transcription of short RNAs from cells lacking Polycomb subunits. (A) Western blotting for H3K27me3, and Ezh2, in Ezh2-1.3 cell line (Ezh2 knockout) (by Keijo Viiri) to measure the efficacy of deletion of the SET domain, which contains histone methyltransferase activity. Measurements were made before the addition of tamoxifen (day zero) and over a time course of five days after addition of tamoxifen. Two days after the addition of tamoxifen, Ezh2 is mutated and H3K27me3 disappears. Antibody against H3 was used as a loading control. **(B)** Northern blotting of short RNA in murine Ezh2-1.3 cell line at the genes *Ybx2*, *Msx1*, and *Hes5*. Short RNAs were measured in a 5-day treatment course after addition of tamoxifen. Day zero was used as a control before addition of tamoxifen. Production of short RNA is maintained constant along the time course treatment. Short RNAs have a size of around 200 nt in Hes5 and Ybx2, and around 90 nt for Msx1.

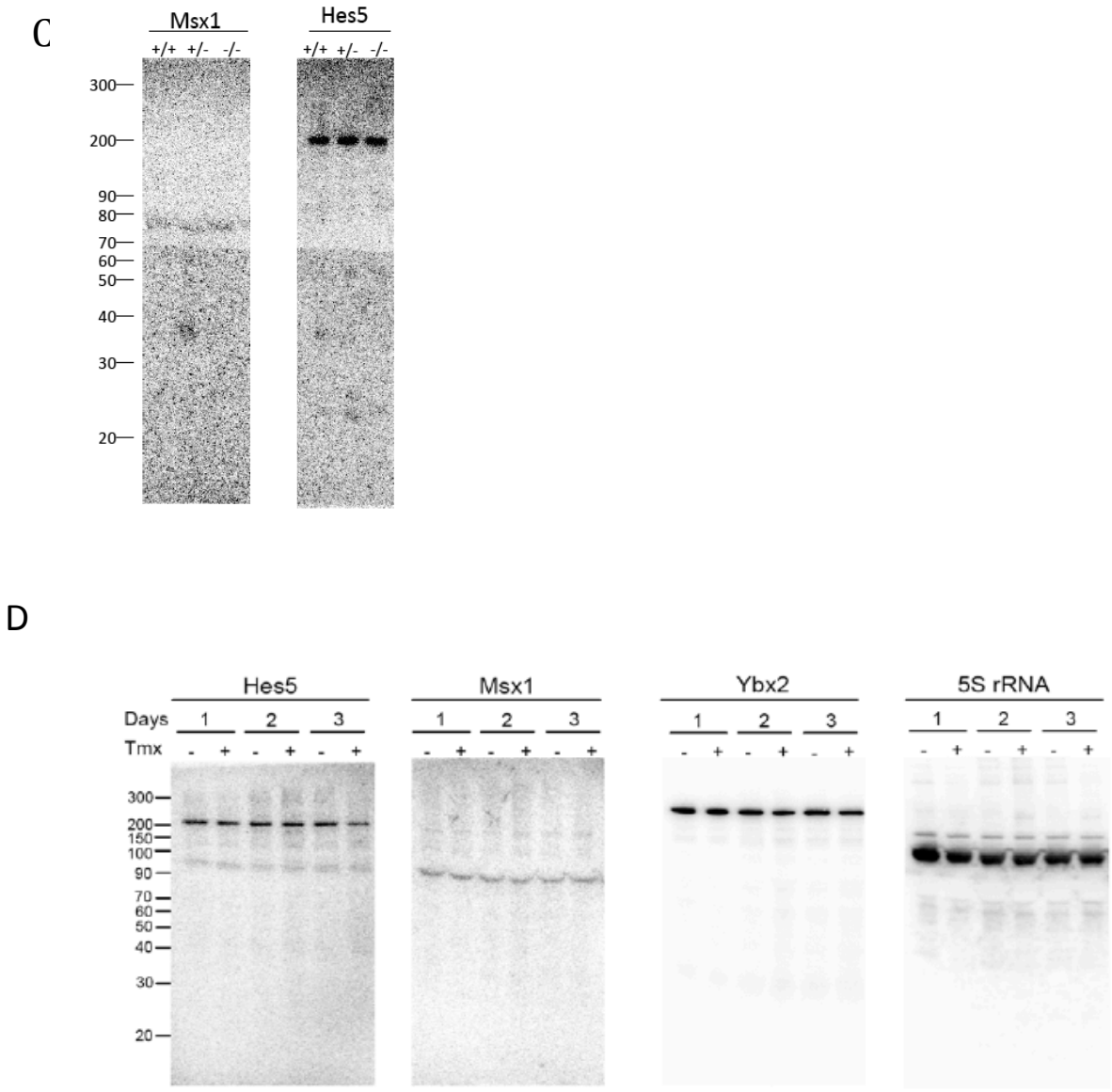


Figure 8: Transcription of short RNAs from cells lacking Polycomb subunits. (C) Northern blotting for short RNA transcribed from *Msx1* and *Hes5* in Jarid2 knockout cells, wild type (+/+), JARID2 heterozygous (+/-), and JARID2 homozygous (-/-) from JM8 mouse ESCs. Short RNA was purified using a single step method with RNAzol. **(D)** Northern blotting in ES-ERT2 (Ring1B knockout) cell line. Short RNAs were measured over a three day time course with and without (control) addition of tamoxifen. Short RNAs have a size of around 200 nt in Hes5 and Ybx2, and around 90 nt for Msx1. Therefore, production of short RNAs is independent of polycomb repressive activity. 5S rRNA was used as a loading control.

To measure whether polycomb loss resulted in derepression of these short RNA-producing genes, mRNA expression was measured at these genes by qPCR (Figure 9A and B). RNA was reverse-transcribed using oligo-dT primers and qPCR performed with primers that spanned an intron. It was found that none of the genes were upregulated upon loss of EZH2 and only *Msx1* increased expression upon loss of Ring1B. Therefore, short RNA production remains constant even upon gene derepression caused by loss of polycomb binding. This means that the short RNAs act upstream of Polycomb activity and are not a by-product of the repressive state it maintains.

To verify that short RNA expression remains constant even if loss of Polycomb activity causes mRNA induction, expression data (Shen et al., 2008) was sought to identify genes that are derepressed after Ezh2 deletion. *Pax3* increases in expression in Ezh2-1.3. Northern blotting shows that short RNA remains constant in Ring1B knockouts (Figure 9C). Therefore, even at genes that are upregulated upon polycomb loss, short RNA production remains constant. It can be concluded from these experiments that PRC1 and PRC2 does not function to regulate short RNA production at polycomb target genes. Also, microarray data could have been compared with ChIP-Seq data to identify polycomb target genes that are up-regulated upon Ezh2 and Ring1 deletion before conducting the experiments to select only for genes that polycomb targets and which are derepressed upon polycomb removal.

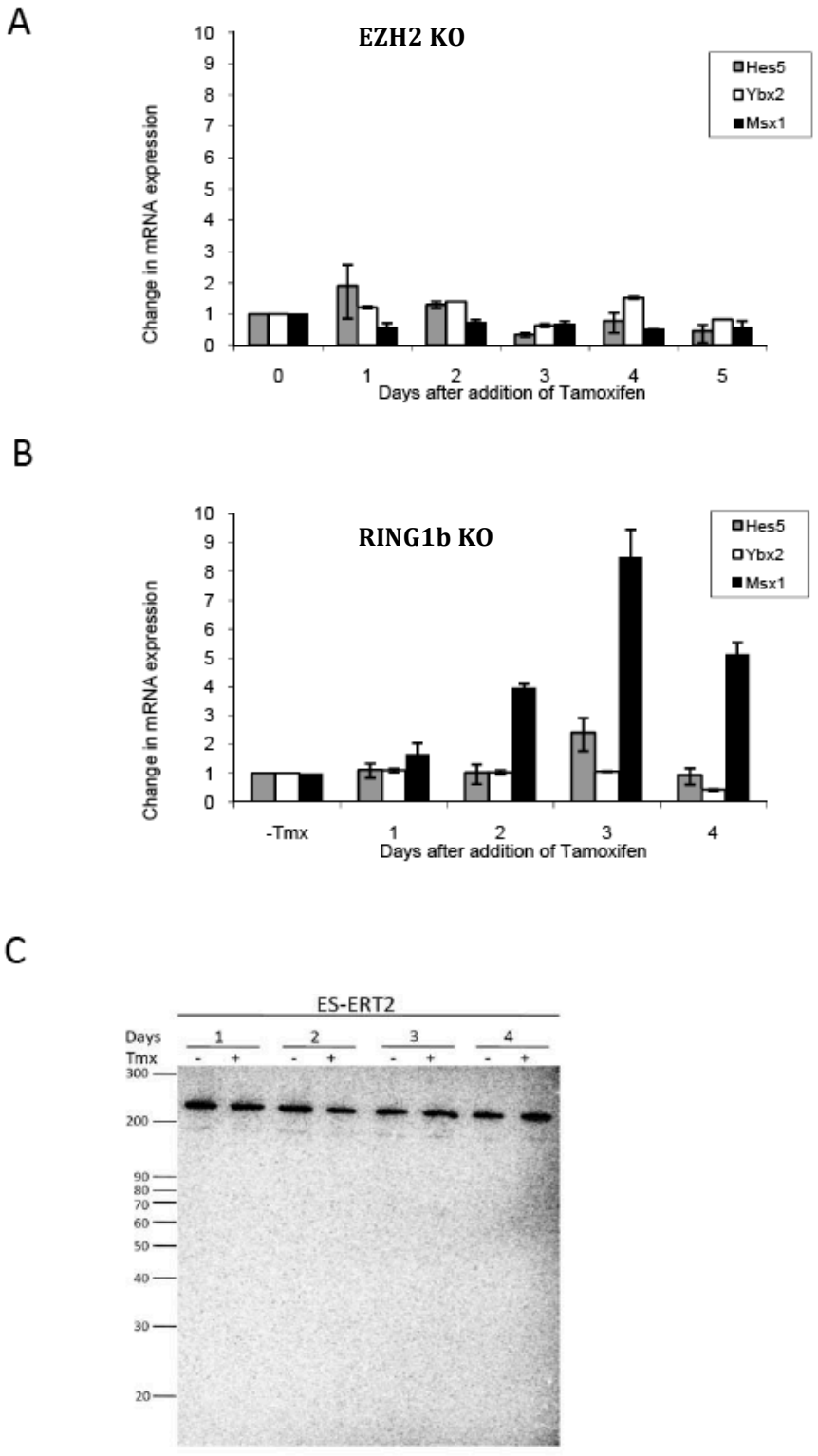


Figure 9: mRNA expression at genes producing short RNAs. (A) Quantitative PCR to measure changes in mRNA levels in the genes *Hes5* (grey), *Ybx2* (white), and *Msx1* (black) in Ezh2-1.3 cells upon knockout of the polycomb Ezh2 subunit of PRC2. Measurements were performed during a 5-day treatment with tamoxifen. Messenger RNA levels were normalised to Actin and to the time zero control. (B) Quantitative PCR for measurement of mRNA in ES-ERT2 cells before and after addition of tamoxifen to induce deletion of Ring1b, showing that only Msx1 is upregulated. (C) Northern blotting for short RNA transcribed from *Pax3*, which is derepressed in Ring1B knockout cell lines.

3.2.4 Short RNAs are lost from polycomb target genes active in other cell types

Evidence from our laboratory indicates that short RNAs might bind to the PRC2 complex. If short RNAs are necessary for Polycomb repression one might expect that loss of Polycomb gene activation is accompanied by loss of short RNAs. Gene Ontology describes the function of genes and the relation between them. To assess a possible correlation between transcriptional states and short RNA, we made use of the fact that many PRC-targeted silent genes in T-cells have a neuronal function using Gene Ontology analysis (performed by Aditi Kanhere). Therefore, predictions were made for genes repressed in T-cells but active in neurons using gene expression data in T-cells (in house data) and looking at the function of the gene.

RNA from the neuroblastoma cell line SH-SY5Y was used in northern blots to look for the absence of short RNAs at genes silent in T-cells and active in neurons. As expected, the neuronal genes *FOXN4, HEY1, MARK1, NKX2-2, BSN* and *HES5* show short RNAs in PBMCs but in the SH-SY5Y neuronal cell line these are reduced (Figure 10A). On the other hand, short RNAs are expressed at equal levels in both PBMCs and neuroblastoma cells for the genes *YBX2* and *NKX2-1*, which are silent in both cell types. These results suggest that short RNAs are indeed a characteristic of Polycomb-silenced genes.

To observe the loss of short RNA during gene activation in a dynamic system, murine ES cells were differentiated into motor neurons (MN) (by Warren Whyte in Rick Young's lab) by addition of retinoic acid (RA) into embryonic bodies (EB) (Wichterle et al., 2002). *Hes5* and *Pcdh8* are repressed in ESCs but activated in MN as confirmed by the increase in mRNA levels analysed by quantitative reverse-PCR (qPCR) (Figure 10B). Using northern blotting, it was found that short RNAs gradually decrease as the genes are activated (Figure 10C). For *Hes5*, the progressive loss of the ~190 nt RNA is accompanied by the appearance of smaller RNAs, suggesting that there is degradation of the short RNA. Supporting these results, our laboratory has shown that short RNAs interact with PRC2, hence depletion of short RNAs could destabilise the association of PRC2 with chromatin allowing gene activation. It is interesting to note that gene activation during differentiation was coupled to a loss of short RNAs (Figure 10C) but gene activation (of Msx1 and Pax3) due to polycomb deletion was not (Figure 8B and D). Therefore, perhaps short RNAs may only disappear upon natural physiological cell differentiation processes.

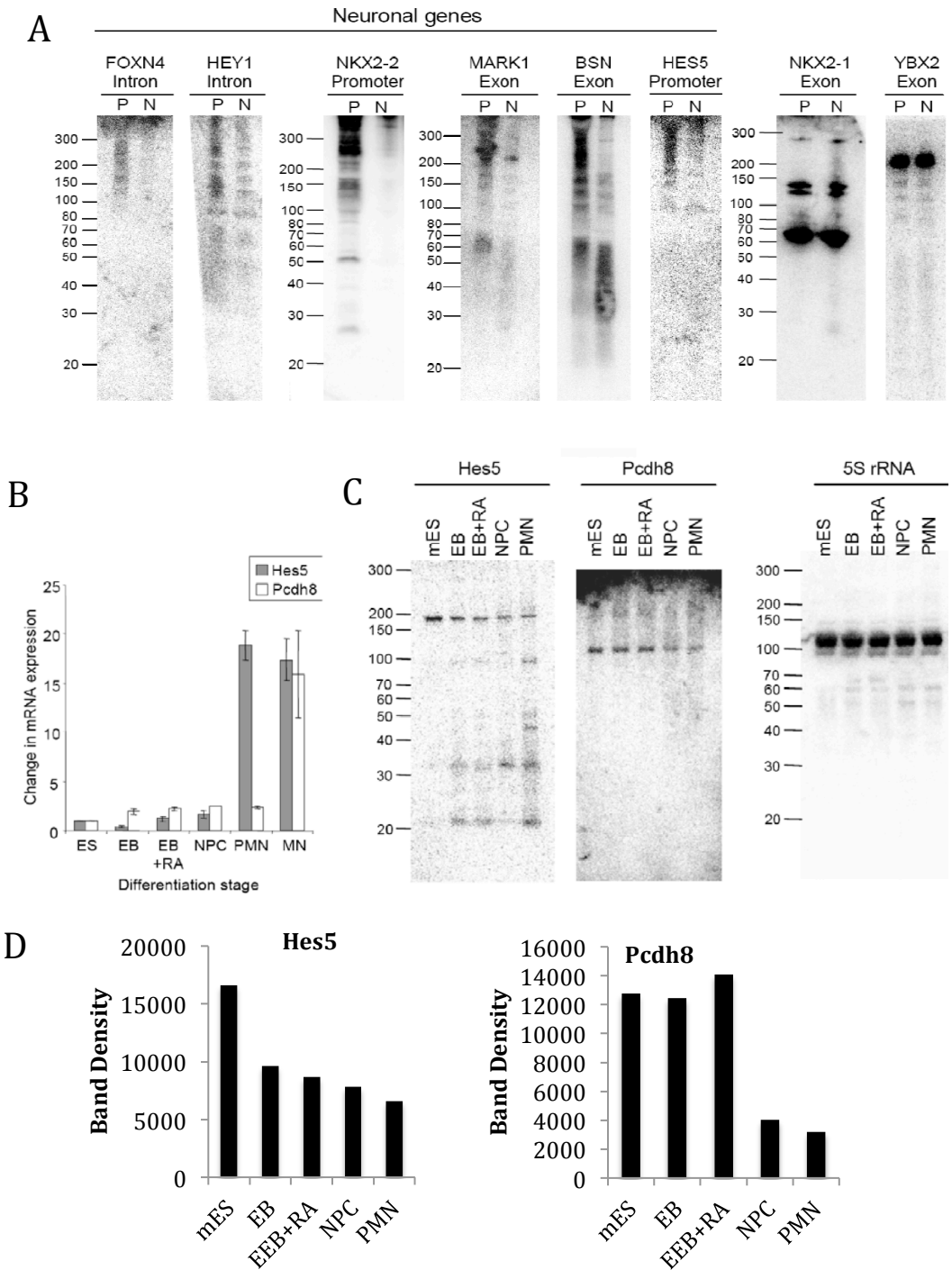


Figure 10: Loss of short RNAs upon gene activation. (A) Northern blotting for short RNAs at the neuronal genes *FOX4, HEY1, NKX2-1, MARK1, BSN*, and *HES5*. These genes are active in SH-SY5Y neuronal cell line, and repressed in PBMCs. *YBX2* and *NKX2-1* are repressed in both cell types, and were used as controls. P – Peripheral blood mononuclear cells (PBMCs); N – SH-SY5Y neuronal cell line. (B) Quantitative PCR showing increased levels of transcription of *Hes5* (grey) and *Pcdh8* (white) during ES cell differentiation to motor neurons. Primers were designed at the 3' end of each gene. mRNA levels were normalised to Actin. EB, embryoid bodies (day 2); EB+RA, EB treated with retinoic acid (day 2, 8 hours later); NPC, neuronal precursor cells (day 3), PMN, motor neuron precursors (day 4); MN, motor neuron (day 7). (C) Northern blotting for short RNA from *Hes5* and *Pcdh8*, which become activated during the step-wise 4-day differentiation to PMN.**(D) Density of the major band on the northern blots of figure 10C. Data was analysed using the software TotalLab Quant.**

3.2.5 Short RNAs are stable transcripts

A previous report has shown that the RNAPII present at polycomb-target genes is phosphorylated at Ser-5 but not at Ser-2 (Stock et al., 2007). In order to demonstrate that this is also true of the RNAPII transcribing the short RNAs, CEM cells were treated with Flavopiridol (FLAV), an inhibitor of P-TEFb phosphorylation of Ser-2, and Actinomycin D (ACTD), which inhibits RNAPII initiation. It was first confirmed that these compounds were specific for RNAPII by performing qPCR on RNAPII-transcribed β-Actin 5'end mRNA using primers for the first exon and first intron, and RNAPIII-transcribed 5S rRNA, as well as via analysis of RNAPI-transcribed 18s and 28S rRNA integrity using a Bioanalyser, in which peaks at the corresponding sizes were observed (Figure 11A and B). To confirm that the drug treatment was efficient, expression of the activation marker CD69 was stimulated by PMA and Ionomycin. If the drugs are inhibiting the transcription of mRNA, then the cell should not be capable of upregulating CD69. FACS analysis for stained CD69 confirms that drug treated cells show a drastic reduction compared to non-treated cells (Figure 11C). Short RNAs were then purified from control and drug-treated cells and blotted for *NKX2-1* and *YBX2*. These two genes were chosen since previous experiments from the laboratory of Richard Jenner showed that they have a good short RNA signal in this cell line. Hypothesizing that short RNAs are products of initiated but not fully-elongating RNAPII, no changes were expected in short RNA expression in Flavopiridol treated cells, and reduced expression in Actinomycin D treated cells. However, short RNAs were still found to be expressed (Figure 11D) after treating the cells with Actinomycin D. These results indicate that short RNAs are stable transcripts and that such types of analyses may not be suitably informative for RNAPII phosphorylation state. It is postulated that RNAPII is the polymerase most likely responsible for short RNA transcription, considering its presence at polycomb target genes (Kanhere et al., 2010; Stock et al., 2007). However, these results also open the possibility that short RNAs are rather transcribed by RNAPIII or RNAPI.

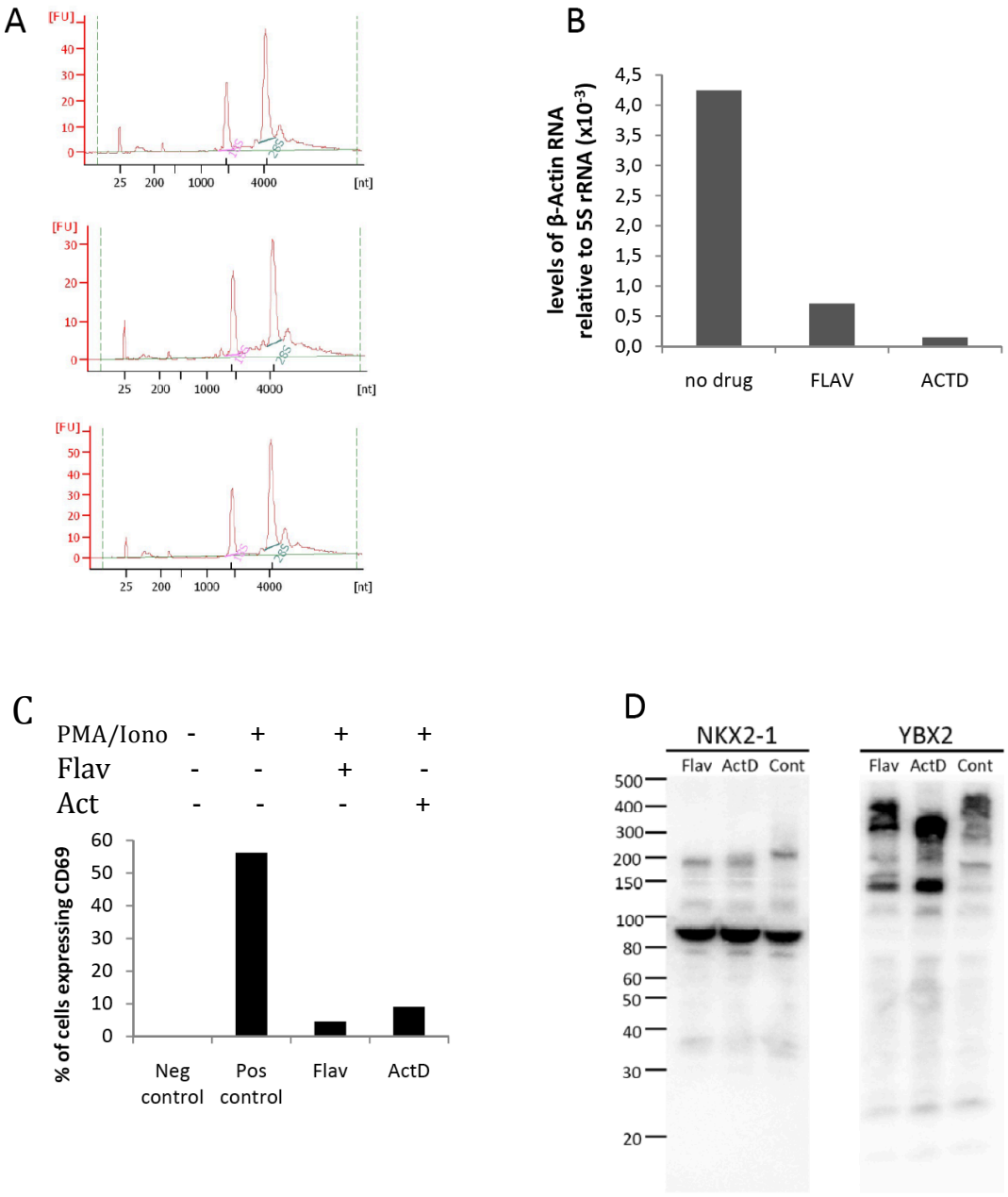


Figure 11: Effect of inhibition of RNAPII transcription on short RNA transcription. (A) Bioanalyser traces showing integrity of RNA extracted from CEM cell treated with Flavoporidol (top), Actinomycin D (middle), and no drug control (bottom). (B) Quantitative PCR showing reduced levels of β-Actin nascent RNA relative to 5S rRNA in drug-treated cells. Flav – Flavopiridol; ActD – ActinomycinD. (C) FACS analysis of percentage of control and PMA/Ionomycin stimulated CEM cells expressing the T-cell receptor CD69. Upon inhibition of RNApII with Flavopiridol (Flav) and Actinomycin D (ActD) the cells cannot efficiently produce CD69. (D) Northern blotting for short RNA in CEM cells treated with Flavopiridol (Flav), ActinomycinD (ActD), and control with no drug. NKX2-1 and YBX2 genes are repressed in the CEM cell line.

3.3 Discussion

3.3.1 Short RNAs are transcribed in the absence of Polycomb

When levels of H3K27me3 or H2Aub are lost by knocking-out their respective catalytic proteins, short RNAs transcribed from these repressed Polycomb-target genes remain unchanged. This occurs even when PcG knockdown results in de-repression of some PRC targeted genes. Since *Hes5* and *Ybx2* do not show an increase in mRNA expression upon Ring1B knockout, and *Hes5, Ybx2* and *Msx1* are not upregulated upon Ezh2 knockout, ChIP-qPCR could have been used to confirm that these genes are polycomb targets. The reason the three genes did not behave in the same way upon polycomb knockout could be due to different physiologic characteristics, such as lack of activators for *Hes5* and *Ybx2*, or RNA might be transcribed from these genes upon polycomb knockout, but not be fully processed. Regarding the use of Pax3 to confirm short RNA expression upon mRNA up-regulation upon polycomb knockout, ChIP-qPCR could have been performed to confirm Ezh2 binding at *Pax3*.

JARID2 is probably involved in PRC2 recruitment and according to Landeira *et al.* (2010) it is also involved in the recruitment of S5P RNAPII. However, JARID2 knock down does not completely abolish S5P RNAPII recruitment, hence the presence of short RNAs in JARID2 knock down cells is expected. In constrast, genes which are naturally activated during ES cell differentiation into motor neurons exhibit depletion and even degradation of short RNAs. This indicates that their invariance upon Ezh2 and Ring1 loss does not merely reflect their stability. Furthermore, this suggests that

cell differentiation is associated with specific signals or transcription factors that lead to reduction in short RNA; in contrast, merely removing polycomb components does not. Another curiosity relies on the mechanisms that lead to short RNA degradation, which due to its apparent stability could be due o the action of exosomes. There is no certainty that these short RNAs are produced in the absence of longer RNA forms and they might instead result from breaking down of a long nascent RNA molecule. Accordingly, Brookes *et al*. (2012) showed that PRC targets, which are highly interlinked with S5P RNAPII, produce transcripts that do not mature into mRNA.

Considering that PRC2 interacts with RNA, short RNAs might function upstream of Polycomb, stabilizing the association of PRC2 with chromatin. Short RNAs could act like the ncRNAs *HOTAIR*, *Xist*, (Pandey et al., 2008; Rinn et al., 2007; Zhao et al., 2008), and other large intergenic non-coding (linc) RNAs, a group a RNAs containing more than 3300 members of which 20% associate with PRC2 and are required for its repression (Khalil et al., 2009). Moreover, the polycomb protein Cbx7 also interacts with chromatin in an RNA-dependent manner (Bernstein et al., 2006a). ANRIL is an antisense RNA from the *Ink4a*/*Arf* locus that binds to the Cbx7 protein (Yap et al., 2010). Indeed, the interaction of the newly identified short RNAs with PRC2 has been confirmed in our laboratory. Xist is a ncRNA involved in X chromosome inactivation through PRC2. It possesses a repetitive element named RepA that has a double-stem loop structure which has been found to directly interact with PRC2 (Zhao et al., 2008). Structural analysis of the short RNA in our

laboratory indicated that similar to the RepA element, these short RNAs also form a double stem-loop structure which is required for the PRC2 binding in vitro. The exact mechanism by which these short RNAs affect Polycomb-chromatin interaction is currently unclear, but the RNAs could possibly function in the interaction of PRC2 with its target loci.

3.3.2 Short RNAs are stable transcripts

Repressed polycomb-target genes contain bivalent chromatin, have RNAPII associated with their promoter region and produce short RNAs, indicative of expression regulated at the level of transcription elongation. Stalled RNAPII is phosphorylated at Ser-5 only, while Ser-2 phophorylation of the RNAPII CTD (hyperphosphorylated RNAPII) is required for a productive elongation. Genes targeted by polycomb exhibit S5P RNAPII only (Stock et al., 2007). In an attempt to demonstrate that short RNAs are transcribed by RNAPII phosphorylated at Ser-5 but not at Ser-2, cells were treated with Flavopiridol, which inhibits Ser-2 phosphorlation only, or Actinomycin D, which inhibits RNAPII initiation. I found that upon total inhibition of RNAPII, short RNAs did not disappear. It is therefore likely that these short RNAs are stable transcripts. The fact that the short RNAs are detected even though they are not associated with actively transcribing polymerase is reflective of their stability. It would be also interesting to analyze the presence of the short RNAs upon knock-down of PcG members or look at the drug treated short RNAs during

cellular differentiation. ChIP could also have been performed to detect the different states of RNAPII. It is currently unclear as to how these short RNAs are generated. They might be products of RNAPII stalling or derived from broken unspliced long RNAs, which is a possibility supported by the fact that Kapranov and colleagues have been able to identify overlapping long and short RNAs (Kapranov et al., 2007).

3.4 Summary of Chapter 3

1) Short RNAs are transcribed from polycomb target genes in murine ES cells in addition to human T-cells. Consequently, they are likely to be a general feature of the mammalian transcriptome.

2) Production of short RNAs is independent of Polycomb activity, indicating the possibility that they may function upstream of PRC2.

3) Short RNAs are depleted from genes in cells where those genes are active, and this can be observed as a dynamic process when genes become activated through cell differentiation. This suggests that they might function in PcG mediated repression.

4) Short RNA levels are not affected by inhibition of RNAPII by Actinomycin D. This either implies that they have a long half-life, or are transcribed by RNAPIII or RNAPI.

Chapter 4 - The role of short RNAs in polycomb response element function

4.1 Introduction

The mammalian genome produces thousands of transcripts most of which are non-coding RNA. Beside the extensively studied interactions of polycomb elements with lncRNAs, small RNAs have also been linked to PcG-mediated repression. It has been demonstrated that, in *Drosophila*, polycomb requires RNA interference (RNAi) to function (Grimaud et al., 2006a). Other non-micro-RNAs were identified at polycomb-target genes that we designated short RNA and which seem to be a common feature among higher order organisms. Given that short RNAs are transcribed upstream of polycomb recruitment, they bind PRC2 (Kanhere et al., 2010) and that they get degraded as polycomb repressed genes become activated, short RNAs might have a function related with polycomb recruitment. Although Polycomb response elements (PREs) have been identified and characterized in *Drosophila*, evidence for their presence in mammals has been lacking. This is due to the fact that PcG-binding is not a critical factor to determine a PRE sequence. The PcG protein Pho is consistently associated with Drosophila PREs, yet the presence of its consensus sequence is not enough to define a PRE. The same is true for its mammalian homolog YY1, which is a multi-functional protein that interacts with several regulators including EED and BMI1. In *Drosophila* some PREs were correctly identified based on DNA-binding motifs, but such were absent in mammalian sequences. Therefore, new features must characterise PREs in mammals. A newly identified human PRE, designated HOXD11.12, revealed

some striking similarities with the chromatin regions around the short RNAs our laboratory identified. Such similarities include flanking H3K27me3 peaks (Kanhere et al., 2010) conserved regions within the PRE, a location at CpG-islands, and nuclease-hypersensitivity (Woo et al., 2010). Nucleosome depletion is characteristic of *Drosophila* PREs too, and is closely associated with the histone variant H3.3, which is more easily disassembled. These shared features suggest that short RNA transcription sites might function as PREs, and that this may be due to the interaction of short RNAs with PRC2. The microarray that Jenner used to identify short RNAs was limited to regions around protein-coding gene start sites. HOXD11.12 is located in between the HOXD11 and HOXD12 genes and is not represented on the array. Therefore, it was unknown whether short RNAs are transcribed from HOXD11.12.

It is well established in *Drosophila* that PRE can act as a TRE (Trithorax response element), thus the sequence having a dual function (Chang et al., 1995; Tillib et al., 1999). Furthermore, they have the potential to switch states between them (Beuchle et al., 2001; Cavalli and Paro, 1998, 1999; Klymenko and Muller, 2004). Some of the factors used to construct the algorithm that predicts PRE/TRE participate in both activation and repression such as Zeste, Gaf and Pipsqueak (Decoville et al., 2001; Dejardin and Cavalli, 2004; Hagstrom et al., 1997; Huang et al., 2002). Another relevant point is the involvement of a ncRNA enconded by the PRE/TRE *bithoraxoid (bxd*) in the recruitment of the methyltransferase ASH1 to activate *Ultrabothorax (Ubx*) gene expression (Sanchez-Elsner et al., 2006). A lncRNA transcribed from the 5' end of the HOXA locus called HOTTIP, also interacts with WDR5-MLL complex to induce H3K4 trimethylation and HOXA activation (Wang et al., 2011).

4.2 Results

4.2.1 Short RNAs are transcribed from the HOXD11.12 PRE

Comparing the data from our laboratory to that published by Woo *et al* (2009), such similarities included an identical pattern of H3K27me3 distribution, high CpG content, and low nucleosome occupancy (analysed by Aditi Kanhere). This evidence has led to the hypothesis that short RNAs might be transcribed from PRE elements and that these are responsible for PRE activity. A closer look at the HOXD11.12 region on the UCSC Human Genome Browser re-enforced this hypothesis by revealing the presence of a TSS located between the two H3K27me3 peaks (Figure 12).

Woo *et al* (2009) identified the Polycomb recruitment activity of the HOXD11.12 PRE in ES-cell derived mesenchymal stem cells (MSC), where the HOXD11 and HOXD12 genes are repressed. The similarities between our short RNA transcription sites and the HOXD11.12 PRE lead to the question of whether short RNAs are also transcribed in HOXD11.12. To this end, an immortalised MSC line derived from bone marrow was used (Funes et al., 2007). This cell line has a different origin to those used by Woo *et al.* (2010), which are derived from H1 and H9 human ESCs. Chromatin immunoprecipitation with antibodies against H3K4me3, and H3K27me3 was performed followed by qPCR analysis to verify that HOXD11.12 is

also targeted by polycomb in these cells (Figure 13A). However, given the low enrichment of H3K27me3, it would have been useful to perform ChIP-qPCR for PRC2 to ensure that this gene is being targeted by polycomb.

Given the resemblance of the HOXD11.12 PRE to the short RNA loci we had identified, we hypothesized that the mammalian PRE might also produce short RNAs. To test this, fifteen probes, ≤ 100 nucleotides apart, were designed across the entire HOXD11.12 region. Northern blotting with these probes revealed that short RNAs are transcribed from HOXD11.12 around the alternative HOXD11 TSS (Figure 13B). These RNAs are about 20 to 30 nucleotides long and occupy a region 150 nucleotides long. Short RNA bands can also be observed at the control gene *HOXC6* in MSCs, shown by our laboratory to be repressed and to produce short RNA in T-cells, and enriched for H3K27me3 in MSC (Woo et al., 2010). These results reveal that short RNAs are transcribed from the HOXD11.12 PRE in MSCs.

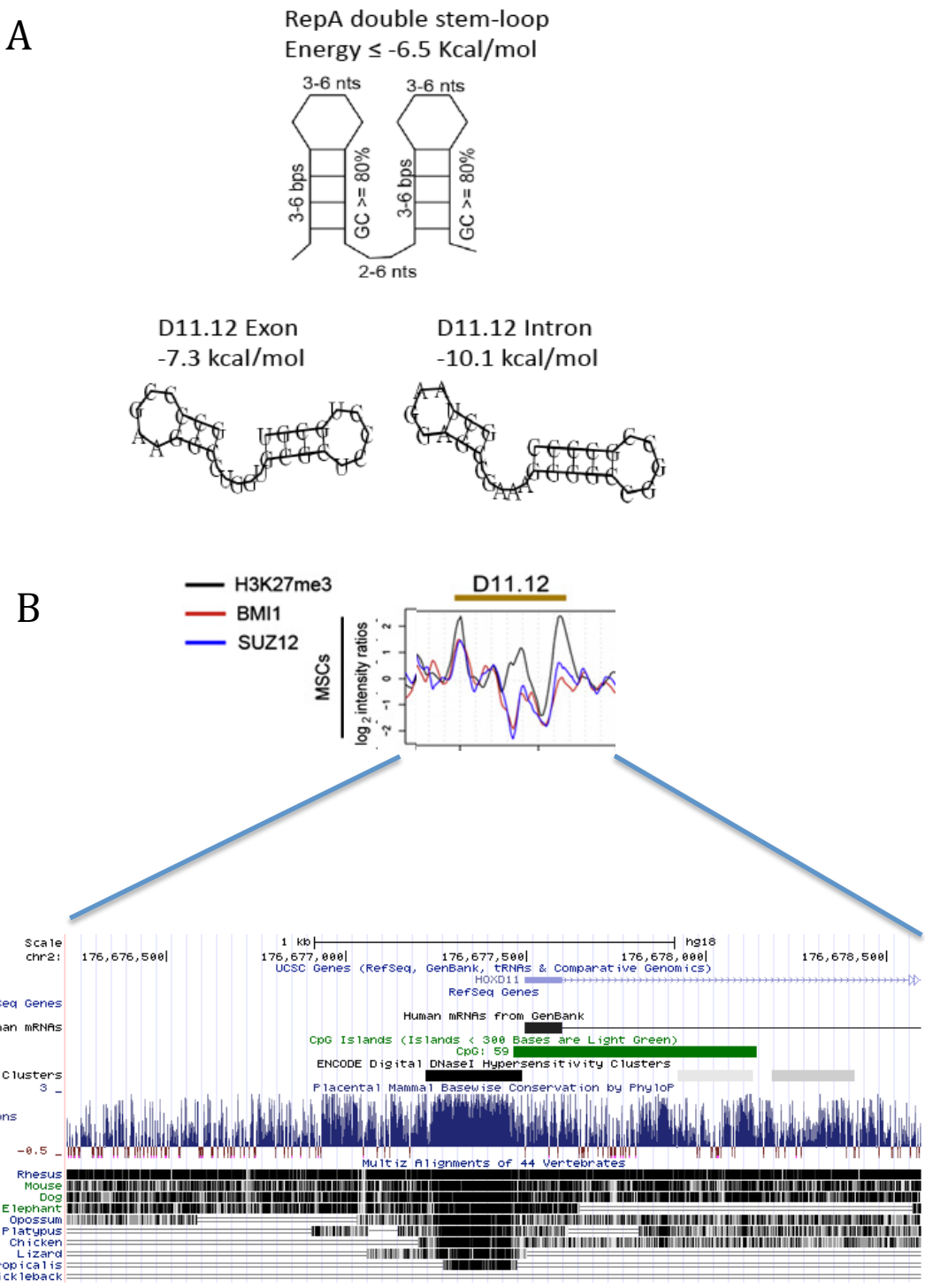
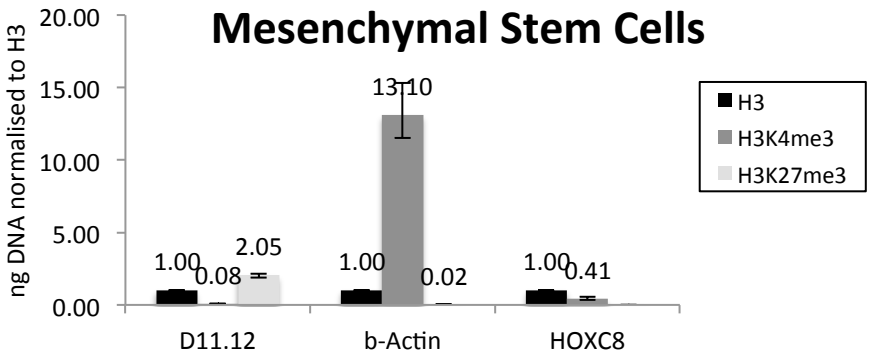
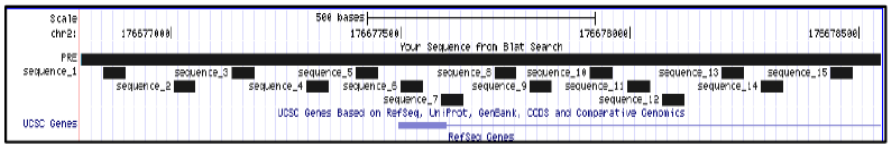


Figure 12: Analysis of short RNAs transcribed from HOXD11.12. (A) Secondary structure derived from the RepA ncRNA (top) used to identify similar structures within short RNAs.Predicted secondary structures found in short RNAs transcribed from the D11.12 PRE (bottom). (B) Localisation of the HOXD11.12 PRE and the sequences with predicted double stem-loop structures for short RNAs 1 and 2 at the UCSC genome browser. HOXD11.12 is localised between H3K27me3 peaks (top panel; from Woo, 2010) and contains an alternative TSS for HOXD11 (bottom panel). The short RNAs are situated on a CpG island.

A



B



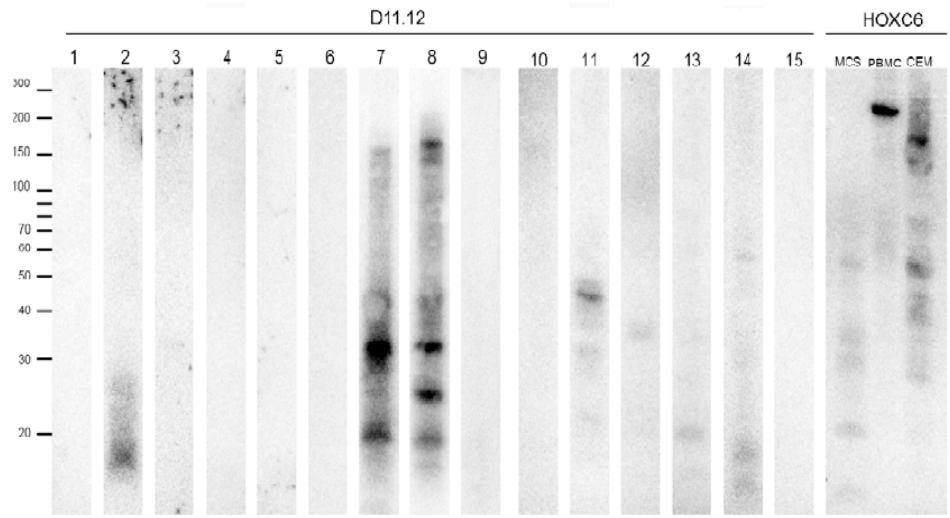


Figure 13: Detection of short RNAs transcribed from HOXD11.12. (A) Quantitative PCR for detection of H3 (black) H3 K4me3 (grey) and H3K27me3 (white). β-Actin was used as control. HOXC8 is repressed in wild type Mesenchymal Stem Cells but not in carcinogenic transformed (5hit) mesenchymal stem cells (Funes JM *et al.* 2007). DNA levels were normalised to H3. (B) Northern blotting detecting short RNAs at the HOXD11.12 PRE. 15 probes were designed spanning the entire HOXD11.12 sequence (top panel). *HOXC6* is repressed in T-cells (PBMCs and CEM) and in MSCs and was used as a positive control.

4.2.2 HOXD11.12 short RNAs contain stem-loop structures and are located at a CpG island

The PRC2 binding site within mouse Xist RepA is a double stem-loop structure that is repeated 7 times (Figure 12A) (Zhao et al., 2008). Moreover, the novel short RNAs identified in our laboratory also have the potential to form this structure. For this reason, the short RNAs identified in HOXD11.12 were examined to find whether they could also form the same structure. A general structural motif was derived based on the RepA sequence, and RNAmotif was used to detect sequences that would produce this structure in the immediate surroundings of probes number 7 and 8, and could fold into a double stem-loop structure, and to confirm the free-energy values for these structures were under -6.5kcal/mol. Two ~30 nt long sequences with predicted double stem-loop were identified (Figure 12A).

Furthermore, it has been shown that PcG proteins localise to CpG islands (Ku et al., 2008), and that these islands might be important for Polycomb recruitment. Sites of short RNA transcription were also commonly associated with CpG islands. Consistent with this, the HOXD11.12 short RNAs are also transcribed from a CpG island (Figure 12B). These results indicate that short RNAs transcribed from HOXD11.12 PRE have the potential to interact with PRC2.

4.2.3 Role of short RNAs in HOXD11.12 PRE activity

Given the presence of short RNAs in HOXD11.12, I sought to determine whether these short RNAs are required for HOXD11.12 PRE activity. To this end, HOXD11.12 was cloned into a luciferase reporter vector that contains an YY1 region and a promoter from the TATA box of Herpes simplex virus thymidine Kinase (pYY1-Luc) as performed by Woo *et al.* (2009). In addition, mutations in HOXD11.12 were

generated; deleted conserved region (marked by the green box in Figure 14A), deleted short RNA region (marked by the red box in Figure 14A) and disrupted RNA secondary structure with point mutations (Figure 14B), in order to allow the significance of these DNA sequences in PRE activity to be assessed. The UCSC genome browser was used to design primers that would amplify the sequences upstream and downstream of the region to be deleted and the halves then ligated together. Site directed mutagenesis was used to disrupt the short RNA structure, by designing primers that align at the mutated region and have no-overlapping much longer than the pp region of the primers. This allows for amplification of the mutated plasmid in a single step. The vectors were then transfected into MSCs produced by Funes et al. (2007). Measurement of luciferase activity revealed that the HOXD11.12 insert in the pYY1-Luc vector did not have repressive activity in MSCs, despite the fact that the endogenous region is trimethylated at H3K27 (Figure 15A). However, H3K27 trimethylation was only 2-fold enriched and this might be a reason why no repressive activity was observed. The conditions of MSC cell culture were altered in an attempt to induce cellular differentiation into chondrocytes and observe possible changes in gene expression. To this end, MSCs were transfected in high and low confluency and for each of these conditions transfection from cells that grew with serum and another set without added serum was performed. Chondrocytes are formed when MSCs are left at high confluency with no added serum (Augello and De Bari, 2010), therefore results could be compared between induced chondrocytes and MSCs. In none of these conditions did HOXD11.12 exhibit repressive activity (Figure 15B). In an attempt to identify a cell type in which the HOXD11.12 sequence has a repressive activity, I experimented transfecting it into other cell types – Hela and SH-SY5Y. Again, in both cell lines exogenous HOXD11.12 failed to repress the reporter gene. In fact, all these cell types HOXD11.12 rather has an activatory effect, acting instead like a Thritorax response element (TRE).

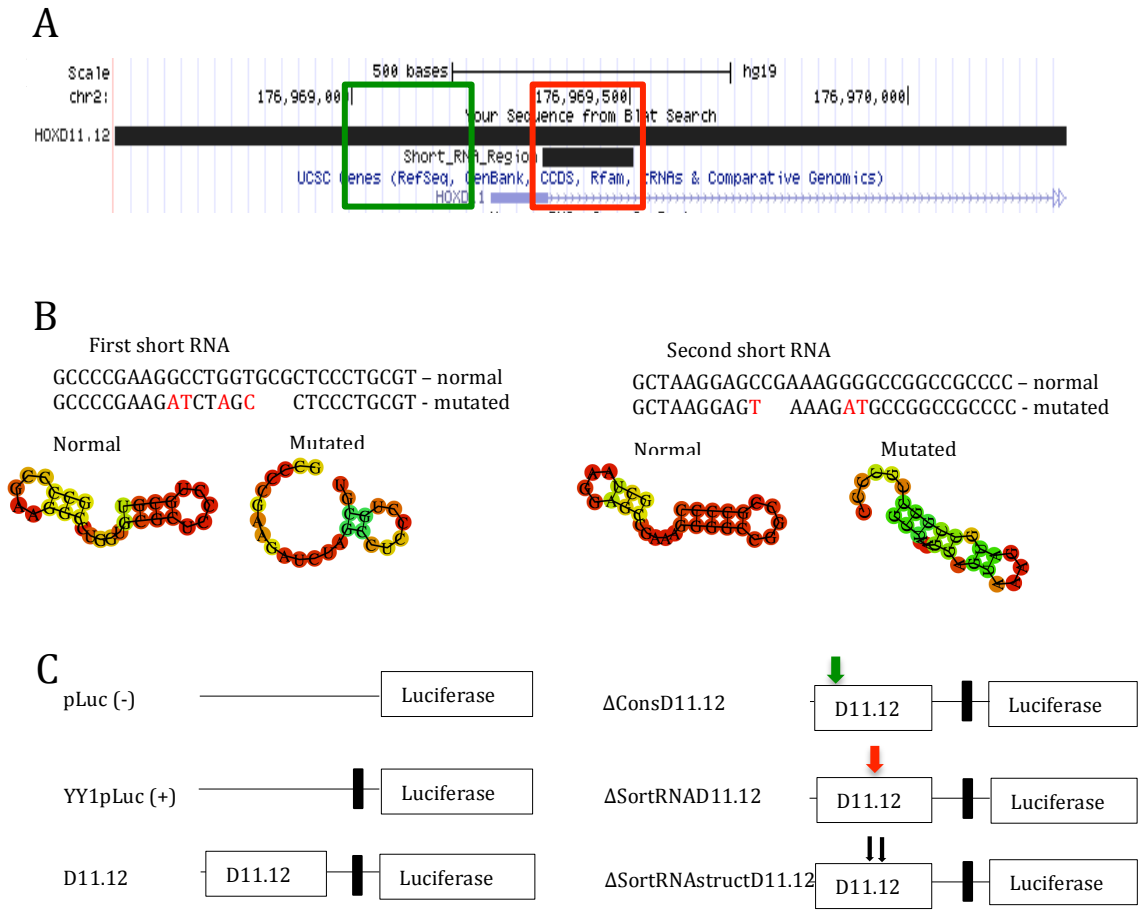


Figure 14: Constructs for assessment of features in HOXD11.12. (A) – UCSC genome browser representation of HOXD11.12, located at the alternative TSS of HOXD11. The green box indicates the part of the conserved region that was deleted; the red box indicates a second deleted region from which short RNAs were detected (black bar). (B) – DNA sequences in HOXD11.12 near probe 7 and 8 (first and second short RNAs) with normal sequence and respective secondary structure, and with mutated sequences and respective disrupted double-stem loop structure; the stem-loop structures were predicted by RNAfold. (C) – Schematic representation of the vectors used for the luciferase assay. The black box represents the YY1 enhancer element, the green arrow indicates the site of deletion of the conserved region, the red arrow indicates the region of deletion of the short RNA sequence, the two black arrows indicate the point mutation for disruption of the secondary structure of the short RNAs.

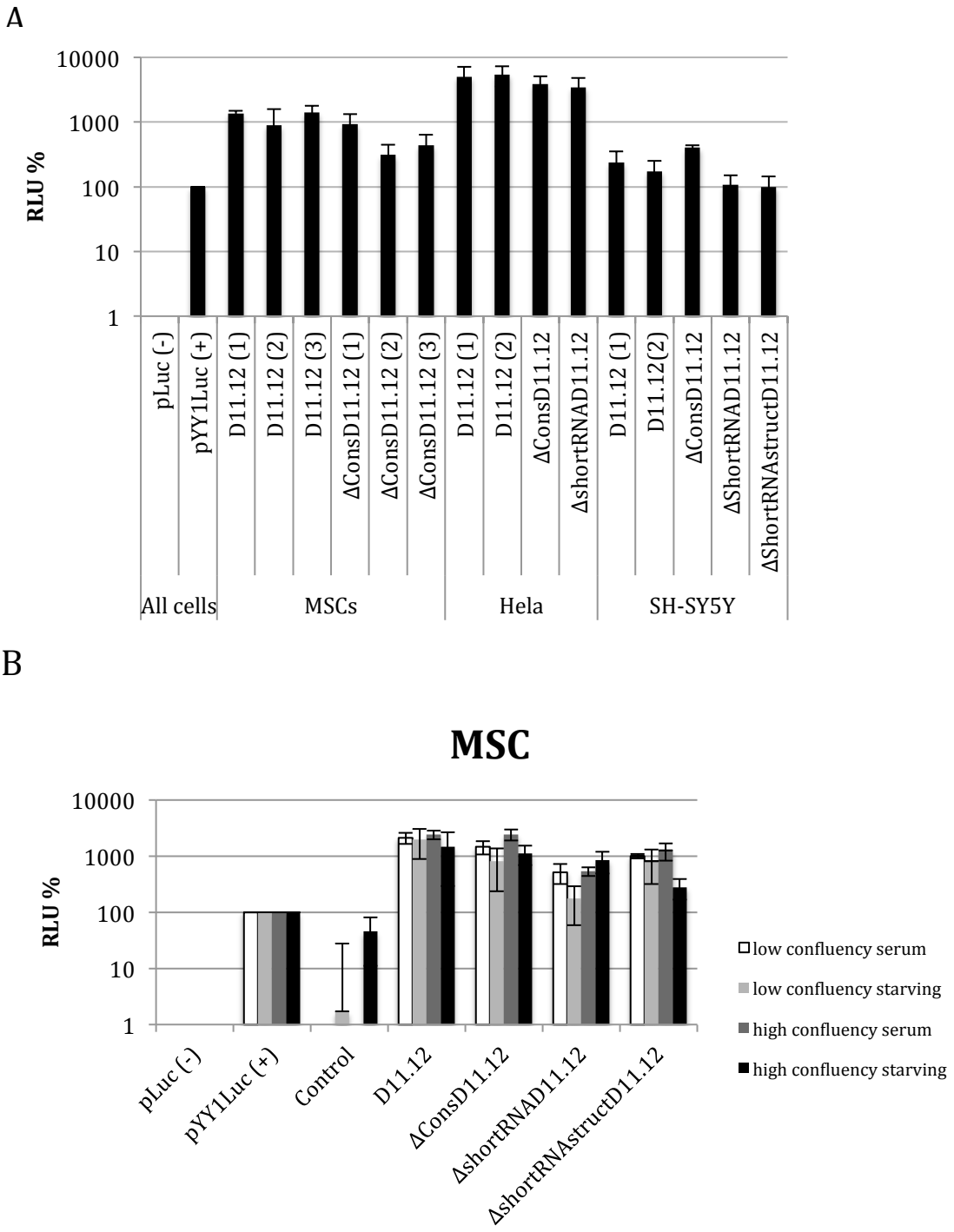


Figure 15: Activity of HOXD11.12 PRE/TRE. (A) Luciferase expression in percentage of Relative Light Units (RLU %) from a vector containing one of the potential PRE D11.12 upstream of the promoter, measured in transformed mesenchymal stem cells (MSCs, 5 hit; Funes JM *et al.* 2007), Hela, and the neuronal cell line SH-SY5Y. ΔConsD11.12, ΔshortRNAD11.12, and ΔshortRNAstructD11.12 respectively correspond to D11.12 sequence with deleted conserved region, D11.12 sequence with deleted shortRNA region, and with mutation of the shortRNA region that disturbs the double stem loop structure. PLuc (-), and YY1pLuc (+) represent the negative and the positive controls, respectively. RLU is normalized so that pLuc(-) is 1% and pYY1Luc(+) is 100%, according to the method of Woo et al. (1), (2), and (3) represent independent clones. (B) Luciferase expression in percentage of Relative Light Units (RLU %) from a vector containing the HOXD11.12 sequence upstream of the promoter, measured in transformed MSCs. Clone details as in (A). Cells were placed in several condition: low confluency with serum (white), low confluency without serum (starving; light grey), high confluency with serum (dark grey), and high confluency starving (black).

4.2.4 Other short RNA loci can potentially behave as PREs

HOXD11.12 shares similarities with the short RNAs regions identified in our laboratory, as described on section 4.2.1. These regions include the two flanked H3K27me3 peaks with short RNAs in the centre, and which contains a TSS. The regions analysed here correspond to HOXD11.12, YBX2, RASL12, and MSX1. To confirm that these DNA sequences similarly to HOXD11.12 have repressive activity, they were cloned into the pYY1Luc as before and individually transfected them into the MSCs, Hela, and SH-SY5Y cells. Not every sequence was repressive in every cell type (Figure 16A). MSX1 was not repressive in any of the cells used and instead had an enhanced activity in MSC and Hela. RASL12 was repressive only in SH-SY5Y, and YBX2 was repressive in Hela and SH-SY5Y. Again, different conditions of MSCs cell culture were tested to see if it would have any effect on luciferase activity by stimulating chondrocytes formation. MSX1 failed to gain repressive activity in serum-starved, high confluence cells and there was also little effect on the two other potential PREs, RASL12 and YBX2 (Figure 16B). In an attempt to determine whether the MSX1 short RNA loci could be repressive, cell types were analyzed for MSX1 that has high levels of H3K27me3 using the UCSC genome browser. MSX1 was positive for H3K27me3 in H1-ES cells and NHEK (Figure 17A), meaning that this cell line could be a good cell line to try this transfection.

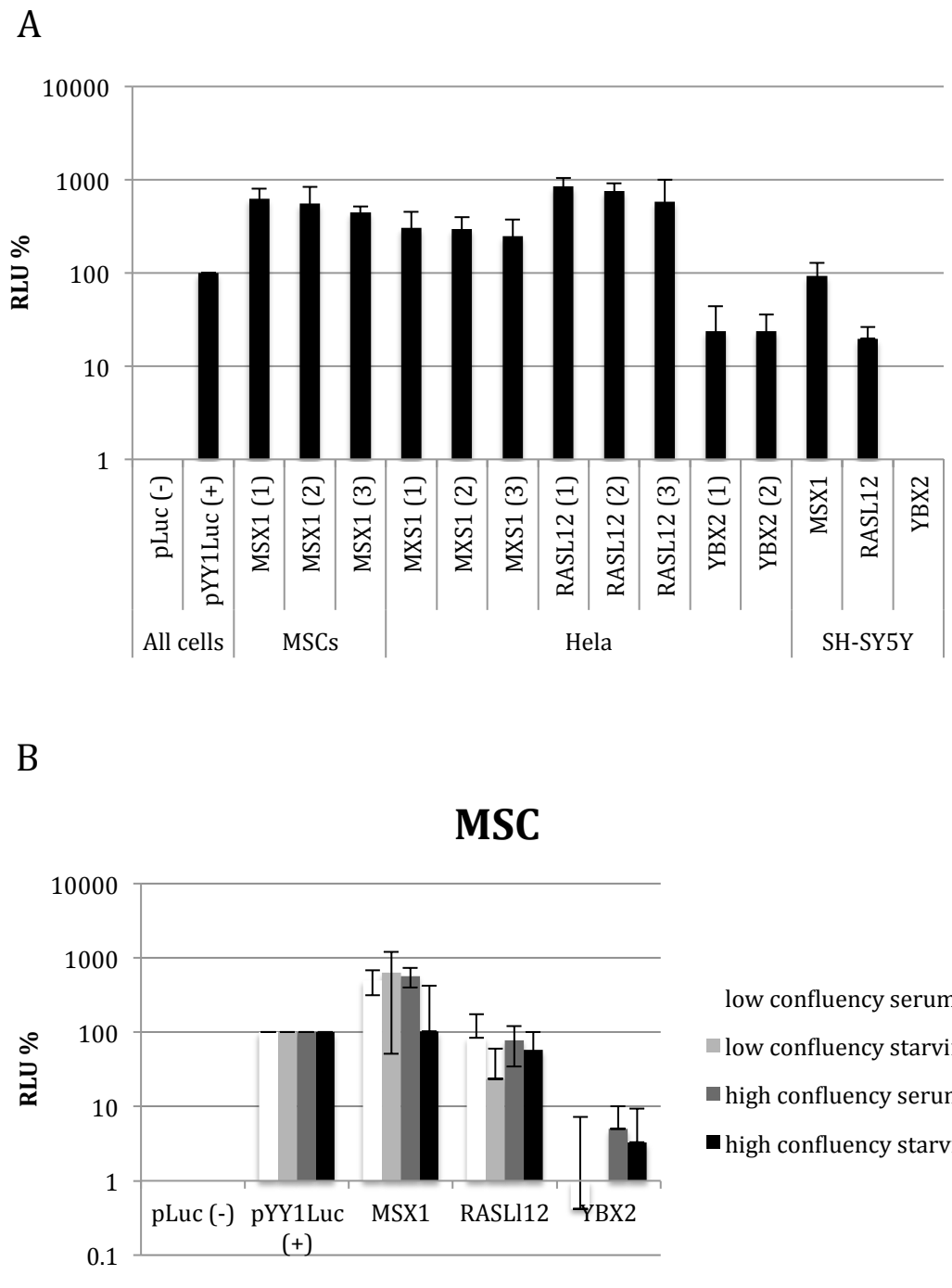


Figure 16: Potential PRE activity of other short RNA loci. (A) Luciferase expression in percentage of Relative Light Units (RLU %) from a vector containing one of the potential PRE MSX1, RASL12, and YBX2 upstream of the promoter, measured in transformed MSCs (5 hit; Funes JM *et al.* 2007), Hela, and neuronal cell line SH-SY5Y. PLuc (-), and YY1pLuc (+) represent the negative and the positive controls, respectively. (1), (2), and (3) represent potential PRE sequence obtained from different clones. (B) Luciferase expression in percentage of Relative Light Units (RLU %) from a vector containing one of the potential PRE MSX1, RASL12, and YBX2, measured in transformed MSCs,. PLuc (-), and YY1pLuc (+) represent the negative and the positive controls, respectively. Cells were placed in several condition: low confluency containing serum (white), low confluency with no serum (starving; light grey), high confluency with serum (dark grey), and high confluency starving (black).

Figure 17: Trimethylation levels in different cell types. (A) - H1ESC (green) and NHEK (pink) cells have high levels of H3K27me3 on the *MSX1* gene. (B) – H1ESC (green), HUVEC (orange), and NHEK (pink) cells have high levels H3K27me3 on HOXD11.12. Both figures were generated and data extracted from the UCSC genome browser. The y-axis represents number of reads.

4.3 Discussion

4.3.1 Short RNAs are transcribed from the HOXD11.12 polycomb response element

PRE elements are regulatory sequences are defined as being necessary and sufficient for the recruitment of PcG proteins. Identification of PREs in *Drosophila* had proven difficult due to the lack of sequence homology, apart from the presence of the consensus Pho/YY1 binding sites. Eventually, the identification of several PRE-binding DNA-binding proteins, such as Pho allowed the construction of PRE prediction algorithms in *Drosophila*. Unfortunately, apart from the mammalian non-functional homologue of Pho, YY1, there are no known mammalian homologues for these factors. However, short RNAs transcribed from polycomb-targeted genes in mammals bind to PRC2 and thus may mediate PRC2 recruitment.

Recently, two mammalian PREs have been identified, one in mice and the other in humans. Surprisingly, there is no conservation of the YY1 binding site in the mouse PRE-*kr*. There are though a few similarities between the *Drosophila* PREs and the newly found human HOXD11.12 PRE. These include nucleosome depletion, and high sequence conservation across species. The loci surrounding short RNAs produced from mammalian Polycomb targets share all these common features and are also located on CG rich regions, implicated in PRC2 recruitment in mammals. CpG islands and nucleosome-free areas have a relative overlap, which might mean that CpG di nucleotide frequency contributes to the absence of nucleosomes, perhaps because these sites are frequently sites of transcription initiation. Moreover, nucleosome depletion is conserved from Drosophila to humans indicating its importance, probably for the recruitment of proteins.

Similar between HOXD11.12 and the short RNA loci identified in our laboratory, was the H3K27me3 distribution pattern around the TSS, which was identical between the HOXD11.12 sequence and DNA sequences from where the short RNAs are transcribed. Furthermore, analysis on the UCSC browser revealed that just as short RNAs are transcribed from the 5' region of target genes, HOXD11.12 also contains an alternative TSS for *HOXD11*, which lies in between the two H3K27me3 peaks. Although some *Drosophila* PREs are located far from the promoter region in the HOX cluster, most are located at TSS (Oktaba et al., 2008; Schwartz et al., 2006).

Given the similarities between HOXD11.12 and our short RNA regions, it is possible that HOXD11.12 also produces short RNAs. Using northern blotting, found that this was indeed the case. Some RNAs such as Xist, HOTAIR and Kcnq1ot1 bind to PcG proteins. Short RNAs identified in our laboratory have double stem-loops that

also interact with PRC2. RNAs from the HOXD11.12 PRE have a structure similar to that of the RepA subunit of Xist, which also binds to PRC2, thus indicating a potential similar function for the HOXD11.12 short RNAs.

4.3.2 Activity of potential PREs and TREs

As mentioned above, HOXD11.12 shares a few similarities with the short RNA loci identified in our laboratory including the fact that HOXD11.12 also produces short RNAs. This lead to the question of whether these short RNAs are required for HOXD11.12 PRE activity. The transfection of the HOXD11.12pYY1Luc vector from Woo *et al*. (2010) into immortalised MSCs gave contradictory results compared to those obtained by that article. This is possibly due to the use of a different type of MSCs to that used by Woo; I used bone-marrow derived cells whereas Woo *et al*. (2010) used MSCs differentiated from H1 and H9 hESC. According to Young *et al*. (2011) there are three profiles of H3K27me3: 1) broad domains and these genes are silenced, 2) bivalent genes, that are also silenced, 3) narrow peaks at the promoter of active genes. It is possible that the HOXD11.12 behaves as the third profile of H3K27me3. It is also possible that endogenous and exogenous HOXD11.12 do not behave in the same way, hence, ChIP-QPCR for detection of H3K27me3 at the vector would be useful. It is also possible that given that the HOX clusters are involved in development and differentiation, which are highly regulate by PcG, and cancer cells show changes in expression of PcG proteins (Alharbi et al., 2013; Cillo et al., 1995; Jin and Sukumar, 2010), it is probable that transformed MSCs have changed expression pattern of HOX genes making it possible that the expression of alternative TSS of HOXD11 is altered. When attempting to use chondrocytes to analyze HOXD11.12 there are a couple of possible reasons for the negative results. One is that chondrocytes have active HOXD11, the second is that MSCs did not properly differentiate into chondrocytes. To check for this last possibility, potential chondrocytes could have been stained for Alcien

Blue, a dye that stains sulfated proteoglycans deposits thus indicating the presence of functional chondrocytes. Given that the vector might not be chromatinized, an alternative to the method used would be integrating the vector into the genome, which could be accomplished by lentivirus infection or homologous recombination into gene deserts (Lienert et al., 2011). Although endogenous HOXD11.12 in MSCs seems to be repressed, its level of H3K27me3 is not very high. Therefore, it was reasonable to look in the UCSC genome browser for cell types where HOXD11.12 is marked by high levels of H3K27me3. I found that HOXD11.12 was highly positive for H3K27me3 in H1-ES, HUVEC and NHEK (Figure 17B) for which transfection of HOXD11.12 in these cells could be possible.

Other cell types (Hela and SH-SY5Y) were transfected with the vectors in an attempt to identify other cellular environments in which the sequences behaved as PREs and repressed luciferase expression. YBX2 in Hela, and SH-SY5Y cells and RASL12 in SH-SY5Y cells have repressive effects on luciferase expression, meaning that they might indeed be mammalian PREs. Inversion of the PRE-kr leads to loss of its function. It would be interesting to check whether inversion of these potential PREs would have the same effect. Other relevant experiments would be, likely what was done in HOXD11.12, to delete the short RNA site in both YBX2 and RASL12 to analyze the relevance they have on PcG recruitment and following positive results to mutate these short RNA sequences in order no disrupt double stem-loop structure to assign its significance in PcG interaction.

Unexpectedly, HOXD11.12 has an activatory effect in bone-marrow derived MSCs, Hela, and neuronal SH-SY5Y cells, potentially acting like a TRE, perhaps because the endogenous *HOXD11.12* region is active in these cells. Indeed, a report shows that HOXD11 is expressed in motor neurons (Misra et al., 2009). In fact, the PRE sequence can reversibly switch between silencing and activation, also acting as a TRE (Beuchle et al., 2001; Cavalli and Paro, 1998, 1999; Klymenko and Muller, 2004). A recent study suggests that the switch occurs at the transition of direction of

transcription of ncRNAs, making it possible that, besides the sense short RNAs we identified, there could also be antisense short RNAs (Herzog et al., 2014). Furthermore, it is not surprising that PRE can act as TRE as polycomb proteins generally associate with promoters. It is possible that like Drosophila PREs, mammalian PREs can alternatively act as TREs in the absence of PRC binding. To confirm this possibility I would have to analyze the recruitment of proteins belonging to the Trithorax complex such as MLL and SET1.

4.4 Summary of chapter 4

1) Short RNAs are transcribed from the human HOXD11.12 PRE, suggesting that PRE function may be mediated by PcG-RNA interactions.

2) D11.12 short RNAs present the same characteristics as previously identified short RNAs. They contain a potential double stem-loop structure, and are located within a CpG island.

3) The HOXD11.12 PRE also possesses activatory function, depending on the cellular context, potentially acting as a TRE.

4) The RASL12 and YBX2 short RNA loci are repressive in SH-SY5Y cells but not MSCs and Hela cells, whereas YBX2 is repressive in MSCs and Hela , but not SH-SY5Y cells. This demonstrates that the repressive effects of such loci is cell-type specific.

Chapter 5 - Conclusions

As part of the work carried out for this thesis, investigations have been carried out regarding aspects of non-coding RNA transcription and the recruiting mechanism of Polycomb group proteins to their target regions. In order to best analyze purified and fractionated RNA, two different methods were compared: single step RNAzol, and two step TRIzol/MiRvana methods. It is concluded that TRIzol/MiRvana method gave better quality RNA because use of RNAzol led to loss of ~ 200nt RNA species.

Previously, our laboratory observed that RNA Pol II is present at Polycomb target genes and that short RNA transcripts are present. This work has been extended to show that the production of these RNAs is conserved across species as they are produced in human primary resting $CD4^+$ T-cells and in mouse ESCs. By extension, this indicates there might be functional significance for these short RNAs. It has also been successfully demonstrated that such RNAs are very stable by observing their presence in cells treated with RNAPII inhibitors.

Polycomb proteins have no known DNA-binding sites and it is not clear how they target mammalian genes. A question asked was whether short RNAs produced at repressed polycomb targeted genes would also express these short RNAs when the gene was active. Therefore, in our laboratory, the production of short RNA from neuronal genes was compared between differentiated SH-SY5Y neuronal cells, where these genes were active and PBMC cells were these genes were silenced by polycomb. It was observed that short RNAs were present only at genes in the cell type in which they were repressed by polycomb. Furthermore, differentiation of ES cells to motor neurons demonstrated a gradual loss of short RNAs as genes became activated. Hence there is a link between short RNA production and gene activity. On these lines, it became

important to test whether the presence of short RNA was a mere by-product of polycomb silencing activity at these genes or whether they were produced upstream of polycomb activity. Ezh2 and Ring1B knockout ESC lines were utilised to test and subsequently verify that production of short RNAs at this repressed genes is independent of PRC activity. It is questionable whether the short RNAs are actually a product of longer RNA degradation. Not every gene begins mRNA expression upon polycomb knockout, a phenomenon not currently well-understood and possibly indicates the existence of different mechanisms of polycomb action, or some genes might not reactivate when deleted polycomb because the transcription factors necessary are nor present.

Attempts were made to test whether short RNAs could be functional rather than just merely being waste products of RNAPII pausing. It was later further demonstrated in the laboratory that these short RNAs bind to the SUZ12 subunit, and have a double stem loop structure similar to the RepA motif in the Xist RNA, a structure that seems to be important for interaction with PRC2. It is important to make future in vitro (EMSA) and in vivo (RNA IP or Chip) experiments with mutations that disrupt the double stem loop structure to draw conclusions regarding the importance of short RNA and its specific secondary structure.

Using computational algorithms to identify specific sequences enriched at polycomb target sites, it was possible to predict Drosophila PREs with some accuracy (Ringrose and Paro, 2004; Ringrose et al., 2003). Unfortunately, the same algorithms could not predict PREs in mammals due to lack of homologies between sequences. Therefore, different features must rule the determination of mammalian PREs. It was noticed that there are similarities between the identified short RNA regions and a newly identified mammalian PRE, HOXD11.12. Such similarities include a nucleosome-

depleted region, a CpG island, high sequence conservation and flanking H3K27me3 peaks. Given these reasons and similarities between the short RNA regions and the HOXD11.12 PRE, tests were carried out to determine whether this PRE also produces short RNAs, for which positive results were obtained. Overall, it is indicative that these five features might constitute typical characteristics of a mammalian PRE. The HOXD11.12 RNAs also have the potential to form a double-stem loop structure and hence have the potential to interact with PRC2. Further analysis of the activity of HOXD11.12 showed that HOXD11.12 acts as an enhancer of expression in cells where HOXD11 is expressed. Therefore, this PRE could potentially also function as a TRE. Other DNA sequences (YBX2 and RASL12) with a series of characteristics identical to HOXD11.12, for which respective genes are repressed in those cells types, produce short RNAs and can potentially behave as a PRE. It is important to test whether PRC structures are present at these sites by ChIP-qPCR. The importance of these features is based on the facts that conserved region of HOXD11.12 is essential for abundance of polycomb proteins at the site; PRC2 is associated with CpG island, although it is not very well understood why, being possible that the reason is due to KDM2b binds to unmethylated DNA at CpG; furthermore, nucleosome depletion allows for accessibility of proteinic complexes. These features are associated with polycomb protein binding allowing for its repressive activity.

Chapter 6 - Future Research

Although short RNAs are transcribed upstream of polycomb activity and that they bind to SUZ12, it has not been demonstrated that they are crucial for silencing of respective genes, nor that the DNA sequences comprising these short RNAs might be PREs. It is also unknown whether these short RNAs are transcribed by RNAPII phosphorylated only at Ser5. It is also of importance to analyze whether short RNAs are a product of polymerase stalling or derived from long RNA processing. A list of future studies to be carried out in order to answer relevant remaining questions are presented as follows:

Experiments involving transcriptional analysis:

1. A well-known ncRNA, *Kcnq1ot1,* is transcribed from the antisense strand of the *Kcnq1* cluster (Pandey et al., 2008). ANRIL is another antisense ncRNA identified (Yap et al., 2010). Additionally, thousands of novel short RNAs mapped genome-wide are antisense to known genes (Kapranov et al., 2010). Although short RNAs identified in our laboratory are transcribed from the sense strand, this evidence indicates there is a probability of antisense RNA being transcribed from polycomb target genes. Future research could therefore include designing northern probes to detect antisense long and short RNA at polycomb target genes.

2. Although microarray analysis in $CD4^+$ T-cells from our laboratory shows that short RNAs are transcribed from silent genes (no detected mRNA), this does not rule out the possibility of longer non-poly-adenylated RNA being transcribed from the same locus. Specifically, it is important to

understand why transcription of these long RNAs is not associated with H3K79me2 or H3K36me3. A plausible reason can be found in the explanation that they are unspliced RNAs transcribed by Ser-5 phosphorylated Pol II (Hargreaves et al., 2009). It is possible that short RNAs are actually products from such unprocessed long RNAs. Supporting this idea is the overlap between novel long and short RNAs transcriptional fragments (Kapranov et al., 2007). Northern blotting for detection of long RNA could be performed to test whether there long RNA at this short RNA producing genes.

3. To prove that short RNAs are transcribed by RNAPII phosphorylated at Ser-5 only, CEM cells can be treated with drugs that inhibit P-TEFb from phosphorylating the Ser-2, for example, the drugs Flavopiridol and DRB. In parallel, total inhibition of RNAPII initiation can be accomplished with Actinomycin D or Alpha-Amanatin. However, previous experiments revealed that these short RNA are very stable as northern bands persist from cells treated with Actinomycin D, and Amanatin. A new method is required which only detects new transcripts. To address this issue, new RNA transcripts could be evaluated by pulse-chase experiments using radiolabelled nucleotides. Therefore, only RNA newly transcribed after the addition of the drugs, would be detected by northern blotting.

4. To test whether other RNA polymerases might transcribe the short RNAs, different RNA polymerases inhibitors should be tested with pulse-chase and compared. Thus, Flavopiridol and DRB do not inhibit RNAPII transcription initiation only, while Triptolide inhibits both RNAPII and RNAPI transcription initiation. On the other hand, Alpha-Amanitin inhibits transcription of RNAPII and RNAPIII but not RNAPI (Bensaude, 2011).

5. Short RNA is degraded/depleted upon gene activation during neuronal differentiation but not when polycomb is depleted. RNA degradation can be achieved by ribonucleases, namely: endonucleases that cut RNA internally, 5′ exonucleases that hydrolyze RNA from the 5′ end, and 3′ exonucleases that degrade RNA from the 3′ end. A systematic knockdown of components of these pathways during cell differentiation would be highly beneficial to observe and identify the exact pathway, which lowers H3K27me3 levels.

Experiments involving analysis of PREs:

6. For a sequence to function as a PRE, it would have to recruit Polycomb proteins. To this end, ChIP could be used for several polycomb components, with subsequent analysis of enrichment by qPCR.

7. To confirm that the HOXD11.12 PRE in immortalised MSCs also confers repressiveness, this PRE was cloned into a luciferase vector, as performed by Woo *et al.* (2010). However, it was not possible to recapitulate the repressive effect of the HOXD11.12 sequence on luciferase expression in bone marrow derived MSCs in the present work. Therefore, it may be that HOXD11.12 behaves differently in these cells compared to the ESC-derived MSCs used by Woo et al. and therefore using these cells may be required. Other cell types containing high levels of H3K27 methylation on the HOXD11.12 region have been identified, and transfection of these cells may reveal that the HOXD11.12 sequence is repressive in these cell types. Another promising strategy that could be employed is the

performance of nucleofection instead of tranfection with electroporation, given that it allows better cell viability and transfection rates of up to 95% (Zeitelhofer et al., 2007).

8. In order to better examine the potential PRE sequences in a chromatinized state, the vector could be integrated into the genome by lentivirus infection or by homologous recombination into a gene desert or through the CRISPR technique described in Wilkinson and Wiedenheft (2014).

9. Our laboratory also showed that PRC2 binds to short RNA and that the secondary structure is fundamental for this binding in vitro. EMSA experiments using wild-type and mutated short RNAs in HOXD11.12 PRE would show whether this is also the case for the short RNA at this potential PRE. Cell extracts from *Escherichia coli* expressing recombinant PRC2 would be used for EMSA.

10. It would be also important to test if short RNAs at HOXD11.12 are important for PRE function in cells by showing that it binds to PRC2. To this end, native RNA immunoprecipitation for SUZ12 could be carried out and checked for short RNA enrichment by qPCR. By constructing DNA mutations that lead to short RNA disruption, it would be possible to confirm if they are indeed important for PRC recruitment. Compensatory mutations that restore the RNA structure could then be introduced and checked for PRC2 binding and luciferase expression.

11. A characteristic of *Drosophila* PcG system is that they maintain repressive states as the cells divide. Woo *et al.* (2010) reported that HOXD11.12 repressive ability could be heritably transmitted through cell differentiation (Woo *et al.*, 2010). This feature could also be tested in other short RNA loci by making use of the mESC's differentiation to motor neurons. Genes repressed in both murine stem cells and neurons, such as *NKX2-1* and *YBX2*, could be used for such analyses. *Pcdh8* becomes activated upon differentiation, and could be used as a positive control. This experiment would also show whether short RNA transcription from PRE elements is a common feature between mice and humans.

12. PREs can also act as TREs. In order to analyse the TRE functionality, ChIP for Trithorax proteins (e.g. MLL and SET1) could be performed in transfected cells showing high luciferase activity. By RNA IP in cell extract or by EMSA in vitro, it would be beneficial to check for TrxG binding to short RNA, as it has been shown that the lncRNA HOTTIP can can target proteins of the trithorax complex to the HOXA locus (Wang et al., 2011). Our laboratory has tested for sense-strand short RNA, and it is possible that the RNA that binds to TrxG is rather anti-sense strand (Herzog et al., 2014).

Search for other possible short RNA function and interactions:

13. Binding of PRC1 to polycomb target genes is not directly dependent on H3K27 methylation by PRC2 (Blackledge et al., 2014). Moreover, the CBX7 subunit of PRC1 has an RNA binding domain (Bernstein et al., 2006a) and has been shown to bind the ncRNA ANRIL (Yap et al., 2010). EMSAs and RNA IP could reveal whether this domain also binds short RNA at polycomb target genes.

14. shRNA-medicated knockdown of short RNAs was not found to be efficient (Russell Bouwman and Richard Jenner, personal communication). Thus other methods would be needed to assess the functionality of the short RNAs. It would be possible to transfect into MSCs a YY1pLuc vector with a single mutation that disrupt the short RNA stem-loop and then create a compensatory mutation somewhere else that restores the RNA structure and where the DNA is still mutated.

Annexes

Table 2: List of primers used for amplification in different experiments.

Experiment	Target	Forward	Reverse
qPCR for mRNA in ES cells defficient for Ezh2 and Ring1B or during differentiation to neurons	Hes5 mRNA	TTTGTATGGGTGGGTGCATGT	AAGCCTTCAGAACAGCCTGTGT
	Pcdh8 mRNA	ATTGGGATTTTATCTTTCACCAGAA	CCACAGACTCAAGATCTACAAGTTGTT
	Msx1 mRNA	CCAGCCCTATAGAAAGCAAGGA	CCCCTCAGAGCAATGCTTTG
	Ybx2 mRNA	CCCACCACCATACTGGAGTGA	AGAGTCAGTTGGAAACTGGCAGAT
	Actin mRNA	TTGTCCCCCCAACTTGATGT	CCCTGGCTGCCTCAACAC
qPCR for gene expression upon inhibition of RNAPII in CEM cells	Actin 5' mRNA	ACCATGGATGATATCGCC	TCGGCTGGCCGGCTTAC
	5S rRNA	GATCTCGGAAGCTAAGCAGG	AAGCCTACAGCACCCGGTAT
Amplification of potential PREs from human genome	MSX1 DNA	GGACAGAAGTTCAGAGCGAG	CAGCTCCTACTGCGAGAAAG
	RASL12 DNA	GCAGGAGCTAGCACTAGATCC	CTGGCTTGAAACGAACACAT
	YBX2 DNA	ACAGAGGGAGACCTTGTGTC	GCTTCAAGGTATTCCTACCCT
Deletion of the D11.12 sequence that contains short RNAs	5'D11.12 DNA	CAGTGCAAGTGCAGGTGCCAGAAC	CGCGCTAGCACCGGTGACTCTGTGGCCTTGTCC
	3'D11.12 DNA	CGCGACCGGTCAAGGGTCCGCGCTAAGGA	CGCGCTAGCCAGTGTGATGGATATCTG
Disruption of Short RNAs secondary structure	Short RNA 1 DNA	CCGAAGATCTAGCCTCCCTGCGTTTCGGGTGGGGC	GGAGGCTAGATCTTCGGGGCGCAGCCAGCCGCTC
	Short RNA 2 DNA	GGAGTAAAGATGCCGGCCGCCCCCTTCCCCTAT	CCGGCATCTTTACTCCTTAGCGCGACCCTTGCC

Table 3: Northern blotting probes used for detection of short RNAs in murine ES cells, and human CEM, PBMCs and SH-SY5Y cell lines.

Gene	Species	Sequence
5S rRNA	Human & mouse	TTAGCTTCCGAGATCAGACGAGATCGGGCGCGTTCAGGGTGGTATGGCC
BSN	Human	GCGGTGCTCACACTCTCGGCGCCGCCGCTGCCGCCGCCATCTCCCAGCT
FOXN4	Human	CAATGCCCGGCATTGCCCGGGAGGAGGGAGCAAAGCCGACCCTGCAAGG
HES5	Human	GATGCCGGGAGCCCCGCGCCTCAATATGCTGCCTTTTCCCAGGCCGCCA
Hes5		GATGCTGAGAGCCCCGCGCCTCAATATGCTGCCTTTTCCCAGGCCGCGG
HEY1	Human	CGCGGCAGGCCTGCGCTCGCCTCCCGCTCTGGCTCGGCTCCGCTCCGCC
HOXC6	Human	GCCATTAGCACCAATTATTAGAGAGATCCCGAGTGCCCAGGACCCTCCC
MARK1	Human	GGCGCGAATGTCTCGGCTCGGTCCGCGCGGTCACAGCCACCGCCGCCGC
Msx1	Mouse	GCTTCCTGTGATCGGCCATGAGGGCCTCCACGCTGAAGGGCAGGAGTGA
NKX2-1	Human	ATGAGCGAGCGAGTCTGGGGACGAACCCTGGGGCCGCACTGTTGGTCTA
NKX2-2	Human	GGAGGAGGGAAAAAATCCTCTTTAACATTCACCGGTTCCTACCTCCCCG
Pax3	Mouse	GGTGAGGGAGGGTGGTGACGAGGCAGGAACACGTCCCAAGTCTCCTCCC
Pcdh8	Mouse	ATCCTCTTCGAACGTGCTGTATCGGACTGTCTTGCTCTGGGCCACTGAG
YBX2	Human	CTCATCCCGCCGGGTCCAGTACCGGCCACAGCCGCCACCGCCCCGGCCC

Table 4: Northern blotting probes spanning the HOXD11.12 PRE.

Order of probes on D11.12 from 5' to 3'	Sequence
1	CAACTGATTGAATGCCAGATGAAGCTGTTTTTTTCTCGTTTTAAAACTG
2	GGGTGAAGGGGCATCCCCCCACATGAATATTTACAACGGTTCCAGATTT
4	AGGGATTCTGAGTCATTTCATTTATTGCCACTACAGTTCTGCAAGAAAG
5	CACCCAAGATAAGACACTAACTTGACCTTAACTTTGTCAGGGCGCCCCT
6	AAACGCAGGGAGCGCACCAGGCCTTCGGGGCGCAGCCAGCCGCTCCGCG
7	CAGCCTCCGGACCCGCGCGTCCCGCTCGAGAACCTACCGTGAAGACCCC
8	CCCCCCGACCCACTCCACAACCCAGAGCGCTTTCATTCCAGTCCCAGCC
9	AGCCATCTGCAGTGGCGCGCCGGTGCATAGGGGAAGGGGGCGGCCGGCC
10	CAGGCAGCCCGGCGCCCCCAGTATGGAGCCTGCTGACCGCCGAAGAGGG
11	GAGGACTTCAACCCGGGTCTTTATGTGTCTGGGGATTCGCAAAATTCTC
12	GGATCTCTCTCCTAAAATGGCTCCCTGGACGCTGCACACTGGCCCTGCG
13	AGCTGGCACCCCCGCCCAGATTCCAGCTGACATTCTTCTCCTCAACCCA
14	AACACTGCGCAAATAGGTCTGTGGCGACTGGAGGGCATCGAGCCTCCCC
15	CCCACCGTGCGGACAGACAGAGAAAAGGCGGGTGGTCTGAGCTAATATC

References

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature reviews Genetics *13*, 720-731.

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.

Aguilo, F., Zhou, M.M., and Walsh, M.J. (2011). Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. Cancer research *71*, 5365-5369.

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 theanteroposterior specification of the axial skeleton. Development (Cambridge, England) *122*, 1513-1522.

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 (Cambridge, England) *137*, 2483-2492.

Alharbi, R.A., Pettengell, R., Pandha, H.S., and Morgan, R. (2013). The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia *27*, 1000-1008.

Alkema, M.J., van der Lugt, N.M., Bobeldijk, R.C., Berns, A., and van Lohuizen, M. (1995). Transformation of axial skeleton due to overexpression of bmi-1 in transgenic mice. Nature *374*, 724-727.

Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nature reviews Genetics *9*, 923-937.

Aloia, L., Di Stefano, B., and Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. Development (Cambridge, England) *140*, 2525-2534.

Antequera, F. (2003). Structure, function and evolution of CpG island promoters. Cellular and molecular life sciences : CMLS *60*, 1647-1658.

Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Proceedings of the National Academy of Sciences of the United States of America *90*, 11995-11999.

Antequera, F., and Bird, A. (1999). CpG islands as genomic footprints of promoters that are associated with replication origins. Current biology : CB *9*, R661-667.

Atchison, L., Ghias, A., Wilkinson, F., Bonini, N., and Atchison, M.L. (2003). Transcription factor YY1 functions as a PcG protein in vivo. The EMBO journal *22*, 1347-1358.

Augello, A., and De Bari, C. (2010). The regulation of differentiation in mesenchymal stem cells. Human gene therapy *21*, 1226-1238.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development *17*, 126-140.

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

Banaszynski, L.A., Wen, D., Dewell, S., Whitcomb, S.J., Lin, M., Diaz, N., Elsasser, S.J., Chapgier, A., Goldberg, A.D., Canaani, E.*, et al.* (2013). Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell *155*, 107-120.

Barna, M., Merghoub, T., Costoya, J.A., Ruggero, D., Branford, M., Bergia, A., Samori, B., and Pandolfi, P.P. (2002). Plzf mediates transcriptional repression of HoxD gene expression through chromatin remodeling. Developmental cell *3*, 499-510.

Basu, A., and Atchison, M.L. (2010). CtBP levels control intergenic transcripts, PHO/YY1 DNA binding, and PcG recruitment to DNA. Journal of cellular biochemistry *110*, 62-69.

Beisel, C., and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. Nature reviews Genetics *12*, 123-135.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. Science (New York, NY) *304*, 1321-1325.

Benetatos, L., Voulgaris, E., Vartholomatos, G., and Hatzimichael, E. (2013). Non-coding RNAs and EZH2 interactions in cancer: long and short tales from the transcriptome. International journal of cancer Journal international du cancer *133*, 267-274.

Bensaude, O. (2011). Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? Transcription *2*, 103-108.

Bentley, D.L., and Groudine, M. (1986). A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature *321*, 702-706.

Bentley, M.L., Corn, J.E., Dong, K.C., Phung, Q., Cheung, T.K., and Cochran, A.G. (2011). Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. The EMBO journal *30*, 3285-3297.

Bergman, Y., and Cedar, H. (2013). DNA methylation dynamics in health and disease. Nature structural & molecular biology *20*, 274-281.

Bernard, D., Martinez-Leal, J.F., Rizzo, S., Martinez, D., Hudson, D., Visakorpi, T., Peters, G., Carnero, A., Beach, D., and Gil, J. (2005). CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the Ink4a/Arf locus. Oncogene *24*, 5543-5551.

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

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

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

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

Berretta, J., and Morillon, A. (2009). Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO reports *10*, 973-982.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S.*, et al.* (2004). Global identification of human transcribed sequences with genome tiling arrays. Science (New York, NY) *306*, 2242-2246.

Beuchle, D., Struhl, G., and Muller, J. (2001). Polycomb group proteins and heritable silencing of Drosophila Hox genes. Development (Cambridge, England) *128*, 993-1004.

Bienz, M., and Muller, J. (1995). Transcriptional silencing of homeotic genes in Drosophila. BioEssays : news and reviews in molecular, cellular and developmental biology *17*, 775-784.

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

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development *16*, 6-21.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E.*, et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Birve, A., Sengupta, A.K., Beuchle, D., Larsson, J., Kennison, J.A., Rasmuson-Lestander, A., and Muller, J. (2001). Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. Development (Cambridge, England) *128*, 3371-3379.

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

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

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. Science (New York, NY) *330*, 612-616.

Bonisch, C., and Hake, S.B. (2012). Histone H2A variants in nucleosomes and chromatin: more or less stable? Nucleic acids research *40*, 10719-10741.

Borsani, G., Tonlorenzi, R., Simmler, M.C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C.*, et al.* (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *351*, 325-329.

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.*, et al.* (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.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

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

Bracken, A.P., and Helin, K. (2009). Polycomb group proteins: navigators of lineage pathways led astray in cancer. Nature reviews Cancer *9*, 773-784.

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

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

Braunschweig, U., Hogan, G.J., Pagie, L., and van Steensel, B. (2009). Histone H1 binding is inhibited by histone variant H3.3. The EMBO journal *28*, 3635-3645.

Bray, M., Driscoll, J., and Huggins, J.W. (2000). Treatment of lethal Ebola virus infection in mice with a single dose of an S-adenosyl-L-homocysteine hydrolase inhibitor. Antiviral research *45*, 135-147.

Breen, T.R., and Duncan, I.M. (1986). Maternal expression of genes that regulate the bithorax complex of Drosophila melanogaster. Developmental biology *118*, 442-456.

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.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell *113*, 25-36.

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

Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. Science (New York, NY) *165*, 349-357.

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

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515-526.

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

Brosius, J. (2005). Waste not, want not--transcript excess in multicellular eukaryotes. Trends in genetics : TIG *21*, 287-288.

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., 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 (Cambridge, England) *130*, 285-294.

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.

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G.*, et al.* (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Molecular cell *6*, 1287-1295.

Bultman, S.J., Gebuhr, T.C., Pan, H., Svoboda, P., Schultz, R.M., and Magnuson, T. (2006). Maternal BRG1 regulates zygotic genome activation in the mouse. Genes & development *20*, 1744-1754.

Bushati, N., and Cohen, S.M. (2007). microRNA functions. Annual review of cell and developmental biology *23*, 175-205.

Bushmeyer, S., Park, K., and Atchison, M.L. (1995). Characterization of functional domains within the multifunctional transcription factor, YY1. The Journal of biological chemistry *270*, 30213-30220.

Cadena, D.L., and Dahmus, M.E. (1987). Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. The Journal of biological chemistry *262*, 12468-12474.

Cai, Y., Jin, J., Yao, T., Gottschalk, A.J., Swanson, S.K., Wu, S., Shi, Y., Washburn, M.P., Florens, L., Conaway, R.C.*, et al.* (2007). YY1 functions with INO80 to activate transcription. Nature structural & molecular biology *14*, 872-874.

Calin, G.A., and Croce, C.M. (2006). MicroRNA signatures in human cancers. Nature reviews Cancer *6*, 857-866.

Campos, E.I., and Reinberg, D. (2009). Histones: annotating chromatin. Annual review of genetics *43*, 559-599.

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, 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 (New York, NY) *298*, 1039-1043.

Cao, R., and Zhang, Y. (2004a). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Current opinion in genetics & development *14*, 155-164.

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

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

Carninci, P. (2009). Molecular biology: The long and short of RNAs. Nature *457*, 974-975.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C.*, et al.* (2005). The transcriptional landscape of the mammalian genome. Science (New York, NY) *309*, 1559-1563.

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

Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell *93*, 505-518.

Cavalli, G., and Paro, R. (1999). Epigenetic inheritance of active chromatin after removal of the main transactivator. Science (New York, NY) *286*, 955-958.

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

Chan, C.S., Rastelli, L., and Pirrotta, V. (1994). A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. The EMBO journal *13*, 2553-2564.

Chang, S., Johnston, R.J., Jr., Frokjaer-Jensen, C., Lockery, S., and Hobert, O. (2004). MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. Nature *430*, 785-789.

Chang, Y.L., King, B.O., O'Connor, M., Mazo, A., and Huang, D.H. (1995). Functional reconstruction of trans regulation of the Ultrabithorax promoter by the products of two antagonistic genes, trithorax and Polycomb. Molecular and cellular biology *15*, 6601-6612.

Chatterjee, N., Sinha, D., Lemma-Dechassa, M., Tan, S., Shogren-Knaak, M.A., and Bartholomew, B. (2011). Histone H3 tail acetylation modulates ATP-dependent remodeling through multiple mechanisms. Nucleic acids research *39*, 8378-8391.

Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science (New York, NY) *303*, 83-86.

Chen, F., Li, Y., Wang, L., and Hu, L. (2011). Knockdown of BMI-1 causes cell-cycle arrest and derepresses p16INK4a, HOXA9 and HOXC13 mRNA expression in HeLa cells. Medical oncology (Northwood, London, England) *28*, 1201-1209.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G.*, et al.* (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science (New York, NY) *308*, 1149-1154.

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.

Christophersen, N.S., and Helin, K. (2010). Epigenetic control of embryonic stem cell fate. The Journal of experimental medicine *207*, 2287-2295.

Cillo, C., Wilmore, H., Barba, P., and Brown, K. (1995). Expression and structure of hox genes in wilms-tumor. International journal of oncology *7*, 1145-1150.

Cloos, P.A., Christensen, J., Agger, K., and Helin, K. (2008). Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes & development *22*, 1115-1140.

Cmarko, D., Verschure, P.J., Martin, T.E., Dahmus, M.E., Krause, S., Fu, X.D., van Driel, R., and Fakan, S. (1999). Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. Molecular biology of the cell *10*, 211-223.

Cockerill, P.N. (2011). Structure and function of active chromatin and DNase I hypersensitive sites. The FEBS journal *278*, 2182-2210.

Cooper, G.M. (2000). The Cell: A Molecular Approach (2000).

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R.*, et al.* (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell reports *7*, 1456-1470.

Copeland, R.A., Solomon, M.E., and Richon, V.M. (2009). Protein methyltransferases as a target class for drug discovery. Nature reviews Drug discovery *8*, 724-732.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (New York, NY) *322*, 1845-1848.

Costa, F.F. (2010). Non-coding RNAs: Meet thy masters. BioEssays : news and reviews in molecular, cellular and developmental biology *32*, 599-608.

Crea, F., Duhagon Serrat, M.A., Hurt, E.M., Thomas, S.B., Danesi, R., and Farrar, W.L. (2011). BMI1 silencing enhances docetaxel activity and impairs antioxidant response in prostate cancer. International journal of cancer Journal international du cancer *128*, 1946-1954.

Croonquist, P.A., and Van Ness, B. (2005). The polycomb group protein enhancer of zeste homolog 2 (EZH 2) is an oncogene that influences myeloma cell growth and the mutant ras phenotype. Oncogene *24*, 6269-6280.

Cross, S., Kovarik, P., Schmidtke, J., and Bird, A. (1991). Non-methylated islands in fish genomes are GC-poor. Nucleic acids research *19*, 1469-1474.

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

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.

da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C*., et al.* (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. Molecular cell *53*, 301-316.

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

De Santa, F., Totaro, M.G., Prosperini, E., Notarbartolo, S., Testa, G., and Natoli, G. (2007). The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. Cell *130*, 1083-1094.

DeCamillis, M., Cheng, N.S., Pierre, D., and Brock, H.W. (1992). The polyhomeotic gene of Drosophila encodes a chromatin protein that shares polytene chromosome-binding sites with Polycomb. Genes & development *6*, 223-232.

Decoville, M., Giacomello, E., Leng, M., and Locker, D. (2001). DSP1, an HMG-like protein, is involved in the regulation of homeotic genes. Genetics *157*, 237-244.

Dejardin, J., and Cavalli, G. (2004). Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. The EMBO journal *23*, 857-868.

Dejardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., and Cavalli, G. (2005). Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. Nature *434*, 533-538.

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

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*, e1002494.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F*., et al.* (2012). Landscape of transcription in human cells. Nature *489*, 101-108.

Doe, C.Q., and Scott, M.P. (1988). Segmentation and homeotic gene function in the developing nervous system of Drosophila. Trends in neurosciences *11*, 101-106.

Doe, C.Q., Smouse, D., and Goodman, C.S. (1988). Control of neuronal fate by the Drosophila segmentation gene even-skipped. Nature *333*, 376-378.

Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E., and Shi, Y. (1999). Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. Molecular and cellular biology *19*, 7237-7244.

Duncan, I. (1986). Control of bithorax complex functions by the segmentation gene fushi tarazu of D. melanogaster. Cell *47*, 297-309.

Duncan, I.M. (1982). Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. Genetics *102*, 49-70.

Duthie, S.M., Nesterova, T.B., Formstone, E.J., Keohane, A.M., Turner, B.M., Zakian, S.M., and Brockdorff, N. (1999). Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. Human molecular genetics *8*, 195-204.

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

El Messaoudi-Aubert, S., Nicholls, J., Maertens, G.N., Brookes, S., Bernstein, E., and Peters, G. (2010). Role for the MOV10 RNA helicase in polycomb-mediated repression of the INK4a tumor suppressor. Nature structural & molecular biology *17*, 862-868.

Enderle, D., Beisel, C., Stadler, M.B., Gerstung, M., Athri, P., and Paro, R. (2011). Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome research *21*, 216-226.

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

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

Evans, M., and Hunter, S. (2002). Source and nature of embryonic stem cells. Comptes rendus biologies *325*, 1003-1007.

Fahrner, J.A., Eguchi, S., Herman, J.G., and Baylin, S.B. (2002). Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. Cancer research *62*, 7213-7218.

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

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

Feng, S., and Holland, E.C. (1988). HIV-1 tat trans-activation requires the loop sequence within tar. Nature *334*, 165-167.

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.

Fitzpatrick, G.V., Soloway, P.D., and Higgins, M.J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nature genetics *32*, 426-431.

Flynn, R.A., Almada, A.E., Zamudio, J.R., and Sharp, P.A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proceedings of the National Academy of Sciences of the United States of America *108*, 10460-10465.

Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287-297.

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

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

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

Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H.W., and Paro, R. (1992). Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. The EMBO journal *11*, 2941-2950.

Friedman, J., Cho, W.K., Chu, C.K., Keedy, K.S., Archin, N.M., Margolis, D.M., and Karn, J. (2011). Epigenetic silencing of HIV-1 by the histone H3 lysine 27 methyltransferase enhancer of Zeste 2. Journal of virology *85*, 9078-9089.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Molecular and cellular biology *24*, 787-795.

Fukuda, T., Tokunaga, A., Sakamoto, R., and Yoshida, N. (2011). Fbxl10/Kdm2b deficiency accelerates neural progenitor cell death and leads to exencephaly. Molecular and cellular neurosciences *46*, 614-624.

Funes, J.M., Quintero, M., Henderson, S., Martinez, D., Qureshi, U., Westwood, C., Clements, M.O., Bourboulia, D., Pedley, R.B., Moncada, S.*, et al.* (2007). Transformation of human mesenchymal stem cells increases their dependency on oxidative phosphorylation for energy production. Proceedings of the National Academy of Sciences of the United States of America *104*, 6223-6228.

Gabellini, D., Green, M.R., and Tupler, R. (2002). Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell *110*, 339-348.

Galvin, K.M., and Shi, Y. (1997). Multiple mechanisms of transcriptional repression by YY1. Molecular and cellular biology *17*, 3723-3732.

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.

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

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

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.

Geisler, S., and Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nature reviews Molecular cell biology *14*, 699-712.

Gil, J., Bernard, D., Martinez, D., and Beach, D. (2004). Polycomb CBX7 has a unifying role in cellular lifespan. Nature cell biology *6*, 67-72.

Gilmour, D.S., and Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Molecular and cellular biology *6*, 3984-3989.

Greer, E.L., Maures, T.J., Ucar, D., Hauswirth, A.G., Mancini, E., Lim, J.P., Benayoun, B.A., Shi, Y., and Brunet, A. (2011). Transgenerational epigenetic inheritance of longevity in Caenorhabditis elegans. Nature *479*, 365-371.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell *123*, 631-640.

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

Grimaud, C., Negre, N., and Cavalli, G. (2006b). From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *14*, 363-375.

Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C.M., and Canaani, E. (1992). The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene. Cell *71*, 701-708.

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.

Guo, B., Odgren, P.R., van Wijnen, A.J., Last, T.J., Nickerson, J., Penman, S., Lian, J.B., Stein, J.L., and Stein, G.S. (1995). The nuclear matrix protein NMP-1 is the transcription factor YY1. Proceedings of the National Academy of Sciences of the United States of America *92*, 10526-10530.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L*., et al.* (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071-1076.

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 (Cambridge, England) *139*, 117-127.

Gutjahr, T., Frei, E., and Noll, M. (1993). Complex regulation of early paired expression: initial activation by gap genes and pattern modulation by pair-rule genes. Development (Cambridge, England) *117*, 609-623.

Gutjahr, T., Frei, E., Spicer, C., Baumgartner, S., White, R.A., and Noll, M. (1995). The Polycomb-group gene, extra sex combs, encodes a nuclear member of the WD-40 repeat family. The EMBO journal *14*, 4296-4306.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P.*, et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L.*, et al.* (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295-300.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339-346.

Hagstrom, K., Muller, M., and Schedl, P. (1997). A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex. Genetics *146*, 1365-1380.

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*, 1291-1300.

Hargreaves, D.C., Horng, T., and Medzhitov, R. (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. Cell *138*, 129-145.

Hauenschild, A., Ringrose, L., Altmutter, C., Paro, R., and Rehmsmeier, M. (2008). Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species. PLoS biology *6*, e261.

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.

He, L., and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nature reviews Genetics *5*, 522-531.

He, Y., Dupree, J., Wang, J., Sandoval, J., Li, J., Liu, H., Shi, Y., Nave, K.A., and Casaccia-Bonnefil, P. (2007). The transcription factor Yin Yang 1 is essential for oligodendrocyte progenitor differentiation. Neuron *55*, 217-230.

He, Y., Kim, J.Y., Dupree, J., Tewari, A., Melendez-Vasquez, C., Svaren, J., and Casaccia, P. (2010). Yy1 as a molecular link between neuregulin and transcriptional modulation of peripheral myelination. Nature neuroscience *13*, 1472-1480.

He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. Science (New York, NY) *322*, 1855-1857.

Hernandez-Munoz, I., Lund, A.H., van der Stoop, P., Boutsma, E., Muijrers, I., Verhoeven, E., Nusinow, D.A., Panning, B., Marahrens, Y., and van Lohuizen, M. (2005). Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. Proceedings of the National Academy of Sciences of the United States of America *102*, 7635-7640.

Herranz, N., Pasini, D., Diaz, V.M., Franci, C., Gutierrez, A., Dave, N., Escriva, M., Hernandez-Munoz, I., Di Croce, L., Helin, K.*, et al.* (2008). Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. Molecular and cellular biology *28*, 4772-4781.

Herzog, V.A., Lempradl, A., Trupke, J., Okulski, H., Altmutter, C., Ruge, F., Boidol, B., Kubicek, S., Schmauss, G., Aumayr, K.*, et al.* (2014). A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. Nature genetics *46*, 973-981.

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

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proceedings of the National Academy of Sciences of the United States of America *106*, 5181-5186.

Hoch, D.A., Stratton, J.J., and Gloss, L.M. (2007). Protein-protein Forster resonance energy transfer analysis of nucleosome core particles containing H2A and H2A.Z. Journal of molecular biology *371*, 971-988.

Hodgson, J.W., Cheng, N.N., Sinclair, D.A., Kyba, M., Randsholt, N.B., and Brock, H.W. (1997). The polyhomeotic locus of Drosophila melanogaster is transcriptionally and post-transcriptionally regulated during embryogenesis. Mechanisms of development *66*, 69-81.

Hofacker, I.L., and Stadler, P.F. (2006). Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics (Oxford, England) *22*, 1172-1176.

Hong, S., Cho, Y.W., Yu, L.R., Yu, H., Veenstra, T.D., and Ge, K. (2007). Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. Proceedings of the National Academy of Sciences of the United States of America *104*, 18439-18444.

Huang, D.H., Chang, Y.L., Yang, C.C., Pan, I.C., and King, B. (2002). pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. Molecular and cellular biology *22*, 6261-6271.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M.*, et al.* (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell *142*, 409-419.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P.*, et al.* (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nature genetics *43*, 621-629.

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.

Ivanov, D., Kwak, Y.T., Guo, J., and Gaynor, R.B. (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. Molecular and cellular biology *20*, 2970-2983.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of molecular biology *3*, 318-356.

Jacobs, J.J., Kieboom, K., Marino, S., DePinho, R.A., and van Lohuizen, M. (1999). The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. Nature *397*, 164-168.

Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nature reviews Genetics *10*, 833-844.

Janody, F., Lee, J.D., Jahren, N., Hazelett, D.J., Benlali, A., Miura, G.I., Draskovic, I., and Treisman, J.E. (2004). A mosaic genetic screen reveals distinct roles for trithorax and polycomb group genes in Drosophila eye development. Genetics *166*, 187-200.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119-133.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nature reviews Genetics *10*, 161-172.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & development *21*, 1519-1529.

Jin, K., and Sukumar, S. (2010). BRCA1: linking HOX to breast cancer suppression. Breast cancer research : BCR *12*, 306.

Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. Nature *426*, 845-849.

Jones, C.A., Ng, J., Peterson, A.J., Morgan, K., Simon, J., and Jones, R.S. (1998). The Drosophila esc and E(z) proteins are direct partners in polycomb group-mediated repression. Molecular and cellular biology *18*, 2825-2834.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews Genetics *13*, 484-492.

Jurgens, M.H., Blunn, C.T., and Peo, E.R., Jr. (1971). Vitamin D2 and cholesterolemia in the growing rat. The Journal of nutrition *101*, 153-159.

Kaji, K., Nichols, J., and Hendrich, B. (2007). Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. Development (Cambridge, England) *134*, 1123-1132.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W., Muller, C.W., Vermeulen, M., and Muller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nature structural & molecular biology *21*, 569-571.

Kalenik, J.L., Chen, D., Bradley, M.E., Chen, S.J., and Lee, T.C. (1997). Yeast two-hybrid cloning of a novel zinc finger protein that interacts with the multifunctional transcription factor YY1. Nucleic acids research *25*, 843-849.

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.

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

Kao, S.Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature *330*, 489-493.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science (New York, NY) *296*, 916-919.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L.*, et al.* (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science (New York, NY) *316*, 1484-1488.

Kapranov, P., Ozsolak, F., Kim, S.W., Foissac, S., Lipson, D., Hart, C., Roels, S., Borel, C., Antonarakis, S.E., Monaghan, A.P.*, et al.* (2010). New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism. Nature *466*, 642-646.

Kassis, J.A. (1994). Unusual properties of regulatory DNA from the Drosophila engrailed gene: three "pairing-sensitive" sites within a 1.6-kb region. Genetics *136*, 1025-1038.

Kemp, C.D., Rao, M., Xi, S., Inchauste, S., Mani, H., Fetsch, P., Filie, A., Zhang, M., Hong, J.A., Walker, R.L.*, et al.* (2012). Polycomb repressor complex-2 is a novel target for mesothelioma therapy. Clinical cancer research : an official journal of the American Association for Cancer Research *18*, 77-90.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes & development *15*, 2654-2659.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A.*, et al.* (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proceedings of the National Academy of Sciences of the United States of America *106*, 11667-11672.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T.*, et al.* (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480-485.

Khuong-Quang, D.A., Buczkowicz, P., Rakopoulos, P., Liu, X.Y., Fontebasso, A.M., Bouffet, E., Bartels, U., Albrecht, S., Schwartzentruber, J., Letourneau, L.*, et al.* (2012). K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. Acta neuropathologica *124*, 439-447.

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 (Dayton, Ohio) *27*, 317-328.

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

Kim, H.G., Kim, K.C., Roh, T.Y., Park, J., Jung, K.M., Lee, J.S., Choi, S.Y., Kim, S.S., and Choi, B.S. (2011). Gene silencing in HIV-1 latency by polycomb repressive group. Virology journal *8*, 179.

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., Kraus, J.C., Chen, J., and Lee, Y. (2003). JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. The Journal of biological chemistry *278*, 42247-42255.

Kim, V.N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. Nature reviews Molecular cell biology *6*, 376-385.

Kim, Y.K., Bourgeois, C.F., Isel, C., Churcher, M.J., and Karn, J. (2002). Phosphorylation of the RNA polymerase II carboxyl-terminal domain by CDK9 is directly responsible for human immunodeficiency virus type 1 Tat-activated transcriptional elongation. Molecular and cellular biology *22*, 4622-4637.

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

Kirmizis, A., Bartley, S.M., Kuzmichev, A., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2004). Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes & development *18*, 1592-1605.

Klattenhoff, C.A., Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P.A., Steinhauser, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S*., et al.* (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell *152*, 570-583.

Klebes, A., Sustar, A., Kechris, K., Li, H., Schubiger, G., and Kornberg, T.B. (2005). Regulation of cellular plasticity in Drosophila imaginal disc cells by the Polycomb group, trithorax group and lama genes. Development (Cambridge, England) *132*, 3753-3765.

Klymenko, T., and Muller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO reports *5*, 373-377.

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.

Knipe, D.M.H., Peter M. (2006). Fields Virology, 5th edn (Lippincott Williams & Wilkins).

Kogo, R., Shimamura, T., Mimori, K., Kawahara, K., Imoto, S., Sudo, T., Tanaka, F., Shibata, K., Suzuki, A., Komune, S*., et al.* (2011). Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer research *71*, 6320-6326.

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*, E171.

Koontz, J.I., Soreng, A.L., Nucci, M., Kuo, F.C., Pauwels, P., van Den Berghe, H., Dal Cin, P., Fletcher, J.A., and Sklar, J. (2001). Frequent fusion of the JAZF1 and JJAZ1 genes in endometrial stromal tumors. Proceedings of the National Academy of Sciences of the United States of America *98*, 6348-6353.

Kouzarides, T. (2002). Histone methylation in transcriptional control. Current opinion in genetics & development *12*, 198-209.

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

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

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.

Kuzmichev, A., Margueron, R., Vaquero, A., Preissner, T.S., Scher, M., Kirmizis, A., Ouyang, X., Brockdorff, N., Abate-Shen, C., Farnham, P.*, et al.* (2005). Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. Proceedings of the National Academy of Sciences of the United States of America *102*, 1859-1864.

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.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M.*, et al.* (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nature cell biology *12*, 618-624.

Lanzuolo, C., and Orlando, V. (2007). The function of the epigenome in cell reprogramming. Cellular and molecular life sciences : CMLS *64*, 1043-1062.

Lanzuolo, C., and Orlando, V. (2012). Memories from the polycomb group proteins. Annual review of genetics *46*, 561-589.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095-1107.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

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

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nature genetics *39*, 1235-1244.

Leung, C., Lingbeek, M., Shakhova, O., Liu, J., Tanger, E., Saremaslani, P., Van Lohuizen, M., and Marino, S. (2004). Bmi1 is essential for cerebellar development and is overexpressed in human medulloblastomas. Nature *428*, 337-341.

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

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565-570.

Lewis, P.W., Muller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science (New York, NY) *340*, 857-861.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.

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, X., Li, L., Pandey, R., Byun, J.S., Gardner, K., Qin, Z., and Dou, Y. (2012). The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network. Cell stem cell *11*, 163-178.

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schubeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. Nature genetics *43*, 1091-1097.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C*., et al.* (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature *438*, 803-819.

Liu, B., Liu, Y.F., Du, Y.R., Mardaryev, A.N., Yang, W., Chen, H., Xu, Z.M., Xu, C.Q., Zhang, X.R., Botchkarev, V.A*., et al.* (2013). Cbx4 regulates the proliferation of thymic epithelial cells and thymus function. Development (Cambridge, England) *140*, 780-788.

Liu, H., and Naismith, J.H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC biotechnology *8*, 91.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science (New York, NY) *305*, 1437-1441.

Liu, J., Rivas, F.V., Wohlschlegel, J., Yates, J.R., 3rd, Parker, R., and Hannon, G.J. (2005). A role for the P-body component GW182 in microRNA function. Nature cell biology *7*, 1261-1266.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S*., et al.* (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nature genetics *42*, 1113-1117.

Lohmann, F., Loureiro, J., Su, H., Fang, Q., Lei, H., Lewis, T., Yang, Y., Labow, M., Li, E., Chen, T*., et al.* (2010). KMT1E mediated H3K9 methylation is required for the maintenance of embryonic stem cells by repressing trophectoderm differentiation. Stem cells (Dayton, Ohio) *28*, 201-212.

Lynch, M.D., Smith, A.J., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L*., et al.* (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. The EMBO journal *31*, 317-329.

Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A., and Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic acids research *29*, 4724-4735.

Maeda, R.K., and Karch, F. (2006). The ABC of the BX-C: the bithorax complex explained. Development (Cambridge, England) *133*, 1413-1422.

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

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.

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., and Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. Science (New York, NY) *303*, 666-669.

Maniataki, E., and Mourelatos, Z. (2005). A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. Genes & development *19*, 2979-2990.

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.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

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

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nature reviews Genetics *11*, 285-296.

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.*, et al.* (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590-604.

Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. Molecular and cellular biology *12*, 2078-2090.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nature reviews Molecular cell biology *6*, 838-849.

Martinez, A.M., and Cavalli, G. (2006). The role of polycomb group proteins in cell cycle regulation during development. Cell cycle (Georgetown, Tex) *5*, 1189-1197.

Matharu, N.K., Hussain, T., Sankaranarayanan, R., and Mishra, R.K. (2010). Vertebrate homologue of Drosophila GAGA factor. Journal of molecular biology *400*, 434-447.

Mattick, J.S., Taft, R.J., and Faulkner, G.J. (2010). A global view of genomic information--moving beyond the gene and the master regulator. Trends in genetics : TIG *26*, 21-28.

McGarvey, K.M., Fahrner, J.A., Greene, E., Martens, J., Jenuwein, T., and Baylin, S.B. (2006). Silenced tumor suppressor genes reactivated by DNA demethylation do not return to a fully euchromatic chromatin state. Cancer research *66*, 3541-3549.

McGarvey, K.M., Van Neste, L., Cope, L., Ohm, J.E., Herman, J.G., Van Criekinge, W., Schuebel, K.E., and Baylin, S.B. (2008). Defining a chromatin pattern that characterizes DNA-hypermethylated genes in colon cancer cells. Cancer research *68*, 5753-5759.

Mejetta, S., Morey, L., Pascual, G., Kuebler, B., Mysliwiec, M.R., Lee, Y., Shiekhattar, R., Di Croce, L., and Benitah, S.A. (2011). Jarid2 regulates mouse epidermal stem cell activation and differentiation. The EMBO journal *30*, 3635-3646.

Mendenhall, E.M., and Bernstein, B.E. (2008). Chromatin state maps: new technologies, new insights. Current opinion in genetics & development *18*, 109-115.

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.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nature reviews Genetics *10*, 155-159.

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

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.

Misra, M., Shah, V., Carpenter, E., McCaffery, P., and Lance-Jones, C. (2009). Restricted patterns of Hoxd10 and Hoxd11 set segmental differences in motoneuron subtype complement in the lumbosacral spinal cord. Developmental biology *330*, 54-72.

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

Moarefi, A.H., and Chedin, F. (2011). ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. Journal of molecular biology *409*, 758-772.

Moehrle, A., and Paro, R. (1994). Spreading the silence: epigenetic transcriptional regulation during Drosophila development. Developmental genetics *15*, 478-484.

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.

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

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.

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

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

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.

Morin, R.D., Johnson, N.A., Severson, T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., Kuchenbauer, F.*, et al.* (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nature genetics *42*, 181-185.

Mousavi, K., Zare, H., Wang, A.H., and Sartorelli, V. (2012). Polycomb protein Ezh1 promotes RNA polymerase II elongation. Molecular cell *45*, 255-262.

Mulholland, N.M., King, I.F., and Kingston, R.E. (2003). Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. Genes & development *17*, 2741-2746.

Muller, J., and Bienz, M. (1991). Long range repression conferring boundaries of Ultrabithorax expression in the Drosophila embryo. The EMBO journal *10*, 3147-3155.

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.

Muller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. Current opinion in genetics & development *16*, 476-484.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nature genetics *39*, 1507-1511.

Nagano, T., and Fraser, P. (2011). No-nonsense functions for long noncoding RNAs. Cell *145*, 178-181.

Nechaev, S., and Adelman, K. (2008). Promoter-proximal Pol II: when stalling speeds things up. Cell cycle (Georgetown, Tex) *7*, 1539-1544.

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

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.

Nguyen, A.T., and Zhang, Y. (2011). The diverse functions of Dot1 and H3K79 methylation. Genes & development *25*, 1345-1358.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell stem cell *4*, 487-492.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.

Nishiyama, A., Xin, L., Sharov, A.A., Thomas, M., Mowrer, G., Meyers, E., Piao, Y., Mehta, S., Yee, S., Nakatake, Y.*, et al.* (2009). Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. Cell stem cell *5*, 420-433.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917-929.

Nowak, A.J., Alfieri, C., Stirnimann, C.U., Rybin, V., Baudin, F., Ly-Hartig, N., Lindner, D., and Muller, C.W. (2011). Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding sites. The Journal of biological chemistry *286*, 23388-23396.

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.

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.

Oktaba, K., Gutierrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E., 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.

Orlando, V., Jane, E.P., Chinwalla, V., Harte, P.J., and Paro, R. (1998). Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. The EMBO journal *17*, 5141-5150.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.

Paige, S.L., Thomas, S., Stoick-Cooper, C.L., Wang, H., Maves, L., Sandstrom, R., Pabon, L., Reinecke, H., Pratt, G., Keller, G.*, et al.* (2012). A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell *151*, 221-232.

Palacios, D., Mozzetta, C., Consalvi, S., Caretti, G., Saccone, V., Proserpio, V., Marquez, V.E., Valente, S., Mai, A., Forcales, S.V.*, et al.* (2010). TNF/p38alpha/polycomb signaling to Pax7 locus in satellite cells links inflammation to the epigenetic control of muscle regeneration. Cell stem cell *7*, 455-469.

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

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

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.

Park, S.H., Park, S.H., Kook, M.C., Kim, E.Y., Park, S., and Lim, J.H. (2004a). Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells. Ultrastructural pathology *28*, 229-238.

Park, Y.J., Dyer, P.N., Tremethick, D.J., and Luger, K. (2004b). A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. The Journal of biological chemistry *279*, 24274-24282.

Paro, R., and Hogness, D.S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proceedings of the National Academy of Sciences of the United States of America *88*, 263-267.

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

Pasini, D., Cloos, P.A., 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.

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

Patel, N.H., Martin-Blanco, E., Coleman, K.G., Poole, S.J., Ellis, M.C., Kornberg, T.B., and Goodman, C.S. (1989). Expression of engrailed proteins in arthropods, annelids, and chordates. Cell *58*, 955-968.

Patturajan, M., Schulte, R.J., Sefton, B.M., Berezney, R., Vincent, M., Bensaude, O., Warren, S.L., and Corden, J.L. (1998). Growth-related changes in phosphorylation of yeast RNA polymerase II. The Journal of biological chemistry *273*, 4689-4694.

Pauli, A., Rinn, J.L., and Schier, A.F. (2011). Non-coding RNAs as regulators of embryogenesis. Nature reviews Genetics *12*, 136-149.

Pemberton, H., Anderton, E., Patel, H., Brookes, S., Chandler, H., Palermo, R., Stock, J., Rodriguez-Niedenfuhr, M., Racek, T., de Breed, L*., et al.* (2014). Genome-wide co-localization of Polycomb orthologs and their effects on gene expression in human fibroblasts. Genome biology *15*, R23.

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.

Pereira, C.F., Piccolo, F.M., Tsubouchi, T., Sauer, S., Ryan, N.K., Bruno, L., Landeira, D., Santos, J., Banito, A., Gil, J.*, et al.* (2010). ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. Cell stem cell *6*, 547-556.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y.*, et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Molecular cell *12*, 1577-1589.

Peterson, C.L., and Laniel, M.A. (2004). Histones and histone modifications. Current biology : CB *14*, R546-551.

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. Science (New York, NY) *309*, 1573-1576.

Pisaneschi, G., Ceccotti, S., Falchetti, M.L., Fiumicino, S., Carnevali, F., and Beccari, E. (1994). Characterization of FIII/YY1, a Xenopus laevis conserved zinc-finger protein binding to the first exon of L1 and L14 ribosomal protein genes. Biochemical and biophysical research communications *205*, 1236-1242.

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 (New York, NY) *300*, 131-135.

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

Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R., 3rd, and Grant, P.A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature *433*, 434-438.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science (New York, NY) *322*, 1851-1854.

Price, D.H. (2008). Poised polymerases: on your mark...get set...go! Molecular cell *30*, 7-10.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nature reviews Molecular cell biology *10*, 192-206.

Qin, J., Whyte, W.A., Anderssen, E., Apostolou, E., Chen, H.H., Akbarian, S., Bronson, R.T., Hochedlinger, K., Ramaswamy, S., Young, R.A.*, et al.* (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.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proceedings of the National Academy of Sciences of the United States of America *90*, 7923-7927.

Reinberg, D., and Sims, R.J., 3rd (2006). de FACTo nucleosome dynamics. The Journal of biological chemistry *281*, 23297-23301.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.

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.

Richly, H., Aloia, L., and Di Croce, L. (2011). Roles of the Polycomb group proteins in stem cells and cancer. Cell death & disease *2*, e204.

Richly, H., Rocha-Viegas, L., Ribeiro, J.D., Demajo, S., Gundem, G., Lopez-Bigas, N., Nakagawa, T., Rospert, S., Ito, T., and Di Croce, L. (2010). Transcriptional activation of polycomb-repressed genes by ZRF1. Nature *468*, 1124-1128.

Ringrose, L. (2007). Polycomb comes of age: genome-wide profiling of target sites. Current opinion in cell biology *19*, 290-297.

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

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

Ringrose, L., Rehmsmeier, M., Dura, J.M., and Paro, R. (2003). Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. Developmental cell *5*, 759-771.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annual review of biochemistry *81*, 145-166.

Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., Hartman, S., Harrison, P.M., Nelson, F.K., Miller, P., Gerstein, M.*, et al.* (2003). The transcriptional activity of human Chromosome 22. Genes & development *17*, 529-540.

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

Roman-Trufero, M., Mendez-Gomez, H.R., Perez, C., Hijikata, A., Fujimura, Y., Endo, T., Koseki, H., Vicario-Abejon, C., and Vidal, M. (2009). Maintenance of undifferentiated state and self-renewal of embryonic neural stem cells by Polycomb protein Ring1B. Stem cells (Dayton, Ohio) *27*, 1559-1570.

Romerio, F., Gabriel, M.N., and Margolis, D.M. (1997). Repression of human immunodeficiency virus type 1 through the novel cooperation of human factors YY1 and LSF. Journal of virology *71*, 9375-9382.

Ross, J.M., and Zarkower, D. (2003). Polycomb group regulation of Hox gene expression in C. elegans. Developmental cell *4*, 891-901.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795-804.

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.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Molecular cell *25*, 15-30.

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

Santos-Rosa, H., Schneider, R., Bernstein, B.E., Karabetsou, N., Morillon, A., Weise, C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2003). Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. Molecular cell *12*, 1325-1332.

Saran, F. (2002). Recent advances in paediatric neuro-oncology. Current opinion in neurology *15*, 671-677.

Sarma, K., Cifuentes-Rojas, C., Ergun, A., Del Rosario, A., Jeon, Y., White, F., Sadreyev, R., and Lee, J.T. (2014). ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. Cell *159*, 869-883.

Satijn, D.P., Hamer, K.M., den Blaauwen, J., and Otte, A.P. (2001). The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in Xenopus embryos. Molecular and cellular biology *21*, 1360-1369.

Sato, T., and Denell, R.E. (1985). Homoeosis in Drosophila: anterior and posterior transformations of Polycomb lethal embryos. Developmental biology *110*, 53-64.

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.

Schuettengruber, B., and Cavalli, G. (2010). The DUBle life of polycomb complexes. Developmental cell *18*, 878-880.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

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*, e13.

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., and Magnuson, T. (1997). Murine Polycomb- and trithorax-group genes regulate homeotic pathways and beyond. Trends in genetics : TIG *13*, 167-170.

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.

Schwartz, Y.B., Kahn, T.G., Stenberg, P., Ohno, K., Bourgon, R., and Pirrotta, V. (2010). Alternative epigenetic chromatin states of polycomb target genes. PLoS genetics *6*, e1000805.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tonjes, M.*, et al.* (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends in genetics : TIG *25*, 335-343.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science (New York, NY) *322*, 1849-1851.

Sen, G.L., and Blau, H.M. (2005). Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. Nature cell biology *7*, 633-636.

Sengupta, A.K., Kuhrs, A., and Muller, J. (2004). General transcriptional silencing by a Polycomb response element in Drosophila. Development (Cambridge, England) *131*, 1959-1965.

Seto, E., Shi, Y., and Shenk, T. (1991). YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. Nature *354*, 241-245.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. Annual review of biochemistry *76*, 75-100.

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., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303-1314.

Shen, X., Liu, Y., Hsu, Y.J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Molecular cell *32*, 491-502.

Shi, Y., Seto, E., Chang, L.S., and Shenk, T. (1991). Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. Cell *67*, 377-388.

Siebold, A.P., Banerjee, R., Tie, F., Kiss, D.L., Moskowitz, J., and Harte, P.J. (2010). Polycomb Repressive Complex 2 and Trithorax modulate Drosophila longevity and stress resistance. Proceedings of the National Academy of Sciences of the United States of America *107*, 169-174.

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., Chiang, A., and Bender, W. (1992). Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. Development (Cambridge, England) *114*, 493-505.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. Developmental biology *158*, 131-144.

Simon, J.A. (2010). Chromatin compaction at Hox loci: a polycomb tale beyond histone tails. Molecular cell *38*, 321-322.

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.

Simon, J.A., and Lange, C.A. (2008). Roles of the EZH2 histone methyltransferase in cancer epigenetics. Mutation research *647*, 21-29.

Sims, R.J., 3rd, Chen, C.F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. The Journal of biological chemistry *280*, 41789-41792.

Sims, R.J., 3rd, and Reinberg, D. (2004). From chromatin to cancer: a new histone lysine methyltransferase enters the mix. Nature cell biology *6*, 685-687.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D., and Cordes, S.P. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell *138*, 885-897.

Slack, J.M.W. (2012). Essential Developmental Biology (Wiley-Blackwell).

Smith, K.P., Byron, M., Clemson, C.M., and Lawrence, J.B. (2004). Ubiquitinated proteins including uH2A on the human and mouse inactive X chromosome: enrichment in gene rich bands. Chromosoma *113*, 324-335.

Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proceedings of the National Academy of Sciences of the United States of America *107*, 20980-20985.

Soboleva, T.A., Nekrasov, M., Pahwa, A., Williams, R., Huttley, G.A., and Tremethick, D.J. (2012). A unique H2A histone variant occupies the transcriptional start site of active genes. Nature structural & molecular biology *19*, 25-30.

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, L., and Atchison, M.L. (2004). YY1 DNA binding and PcG recruitment requires CtBP. Genes & development *18*, 2596-2601.

Srinivasan, L., Pan, X., and Atchison, M.L. (2005). Transient requirements of YY1 expression for PcG transcriptional repression and phenotypic rescue. Journal of cellular biochemistry *96*, 689-699.

Stanojevic, D., Hoey, T., and Levine, M. (1989). Sequence-specific DNA-binding activities of the gap proteins encoded by hunchback and Kruppel in Drosophila. Nature *341*, 331-335.

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.

Stojic, L., Jasencakova, Z., Prezioso, C., Stutzer, A., Bodega, B., Pasini, D., Klingberg, R., Mozzetta, C., Margueron, R., Puri, P.L*., et al.* (2011). Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells. Epigenetics & chromatin *4*, 16.

Struhl, G., and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. The EMBO journal *4*, 3259-3264.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nature structural & molecular biology *14*, 103-105.

Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development (Cambridge, England) *132*, 2093-2102.

Strutt, H., and Paro, R. (1997). The polycomb group protein complex of Drosophila melanogaster has different compositions at different target genes. Molecular and cellular biology *17*, 6773-6783.

Su, I.H., Basavaraj, A., Krutchinsky, A.N., Hobert, O., Ullrich, A., Chait, B.T., and Tarakhovsky, A. (2003). Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. Nature immunology *4*, 124-131.

Swigut, T., and Wysocka, J. (2007). H3K27 demethylases, at long last. Cell *131*, 29-32.

Szenker, E., Ray-Gallet, D., and Almouzni, G. (2011). The double face of the histone variant H3.3. Cell research *21*, 421-434.

Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K*., et al.* (2009). Tiny RNAs associated with transcription start sites in animals. Nature genetics *41*, 572-578.

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.

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.

Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nature reviews Molecular cell biology *11*, 264-275.

Tan, J., Yang, X., Zhuang, L., Jiang, X., Chen, W., Lee, P.L., Karuturi, R.K., Tan, P.B., Liu, E.T., and Yu, Q. (2007). Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. Genes & development *21*, 1050-1063.

Tan, L., and Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development (Cambridge, England) *139*, 1895-1902.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H*., et al.* (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664-678.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Developmental cell *15*, 668-679.

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

Tie, F., Furuyama, T., and Harte, P.J. (1998). The Drosophila Polycomb Group proteins ESC and E(Z) bind directly to each other and co-localize at multiple chromosomal sites. Development (Cambridge, England) *125*, 3483-3496.

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., and Harte, P.J. (2001). The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. Development (Cambridge, England) *128*, 275-286.

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.

Tijsterman, M., and Plasterk, R.H. (2004). Dicers at RISC; the mechanism of RNAi. Cell *117*, 1-3.

Tillib, S., Petruk, S., Sedkov, Y., Kuzin, A., Fujioka, M., Goto, T., and Mazo, A. (1999). Trithorax- and Polycomb-group response elements within an Ultrabithorax transcription maintenance unit consist of closely situated but separable sequences. Molecular and cellular biology *19*, 5189-5202.

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

Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell *130*, 299-308.

Trentani, A., Testillano, P.S., Risueno, M.C., and Biggiogera, M. (2003). Visualization of transcription sites at the electron microscope. European journal of histochemistry : EJH *47*, 195-200.

Trojer, P., Cao, A.R., Gao, Z., Li, Y., Zhang, J., Xu, X., Li, G., Losson, R., Erdjument-Bromage, H., Tempst, P*., et al.* (2011). L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. Molecular cell *42*, 438-450.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science (New York, NY) *329*, 689-693.

Tumes, D.J., Onodera, A., Suzuki, A., Shinoda, K., Endo, Y., Iwamura, C., Hosokawa, H., Koseki, H., Tokoyoda, K., Suzuki, Y.*, et al.* (2013). The polycomb protein Ezh2 regulates differentiation and plasticity of CD4(+) T helper type 1 and type 2 cells. Immunity *39*, 819-832.

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.*, 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 Vlag, J., and Otte, A.P. (1999). Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. Nature genetics *23*, 474-478.

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P.*, et al.* (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature *419*, 624-629.

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*, 922-926.

Velichutina, I., Shaknovich, R., Geng, H., Johnson, N.A., Gascoyne, R.D., Melnick, A.M., and Elemento, O. (2010). EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. Blood *116*, 5247-5255.

Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T., and Pasini, D. (2012). Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. Nucleic acids research *40*, 3403-3418.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M.*, et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

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.

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, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Molecular cell *43*, 904-914.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A.*, et al.* (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120-124.

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, R., Taylor, A.B., Leal, B.Z., Chadwell, L.V., Ilangovan, U., Robinson, A.K., Schirf, V., Hart, P.J., Lafer, E.M., Demeler, B.*, et al.* (2010). Polycomb group targeting through different binding partners of RING1B C-terminal domain. Structure (London, England : 1993) *18*, 966-975.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends in biochemical sciences *24*, 437-440.

Whitcomb, S.J., Basu, A., Allis, C.D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. Trends in genetics : TIG *23*, 494-502.

Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110*, 385-397.

Wichterle, H., and Peljto, M. (2008). Differentiation of mouse embryonic stem cells to spinal motor neurons. Current protocols in stem cell biology *Chapter 1*, Unit 1H 1 1-1H 1 9.

Wiedemann, S.M., Mildner, S.N., Bonisch, C., Israel, L., Maiser, A., Matheisl, S., Straub, T., Merkl, R., Leonhardt, H., Kremmer, E.*, et al.* (2010). Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. The Journal of cell biology *190*, 777-791.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Wilkinson, R., and Wiedenheft, B. (2014). A CRISPR method for genome engineering. F1000prime reports *6*, 3.

Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008). 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell *135*, 919-932.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes & development *23*, 1494-1504.

Witt, O., Albig, W., and Doenecke, D. (1996). Testis-specific expression of a novel human H3 histone gene. Experimental cell research *229*, 301-306.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Kohler, C. (2011). High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm. PLoS genetics *7*, e1002126.

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.

Wood, A.J., and Oakey, R.J. (2006). Genomic imprinting in mammals: emerging themes and established theories. PLoS genetics *2*, e147.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K.*, et al.* (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS biology *3*, e7.

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science (New York, NY) *329*, 444-448.

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, S., Shi, Y., Mulligan, P., Gay, F., Landry, J., Liu, H., Lu, J., Qi, H.H., Wang, W., Nickoloff, J.A*., et al.* (2007). A YY1-INO80 complex regulates genomic stability through homologous recombination-based repair. Nature structural & molecular biology *14*, 1165-1172.

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. (2011). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nature reviews Genetics *12*, 542-553.

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell *121*, 859-872.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P*., et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.

Yamaguchi, Y., Shibata, H., and Handa, H. (2013). Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. Biochimica et biophysica acta *1829*, 98-104.

Yang, Y., and Hua, X. (2007). In search of tumor suppressing functions of menin. Molecular and cellular endocrinology *265-266*, 34-41.

Yap, D.B., Chu, J., Berg, T., Schapira, M., Cheng, S.W., Moradian, A., Morin, R.D., Mungall, A.J., Meissner, B., Boyle, M*., et al.* (2011). Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. Blood *117*, 2451-2459.

Yap, K.L., Li, S., Munoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Molecular cell *38*, 662-674.

Yeap, L.S., Hayashi, K., and Surani, M.A. (2009). ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. Epigenetics & chromatin *2*, 12.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science (New York, NY) *304*, 594-596.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.

Young, M.D., Willson, T.A., Wakefield, M.J., Trounson, E., Hilton, D.J., Blewitt, M.E., Oshlack, A., and Majewski, I.J. (2011). ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. Nucleic acids research *39*, 7415-7427.

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.

Yu, M., Mazor, T., Huang, H., Huang, H.T., Kathrein, K.L., Woo, A.J., Chouinard, C.R., Labadorf, A., Akie, T.E., Moran, T.B.*, et al.* (2012). Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. Molecular cell *45*, 330-343.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science (New York, NY) *309*, 626-630.

Yuan, P., Han, J., Guo, G., Orlov, Y.L., Huss, M., Loh, Y.H., Yaw, L.P., Robson, P., Lim, B., and Ng, H.H. (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes & development *23*, 2507-2520.

Zeitelhofer, M., Vessey, J.P., Xie, Y., Tubing, F., Thomas, S., Kiebler, M., and Dahm, R. (2007). High-efficiency transfection of mammalian neurons via nucleofection. Nature protocols *2*, 1692-1704.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nature genetics *39*, 1512-1516.

Zentner, G.E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. Nature structural & molecular biology *20*, 259-266.

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.*, et al.* (2011). PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. Stem cells (Dayton, Ohio) *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 (New York, NY) *322*, 750-756.

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.*, et al.* (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.

Zink, B., and Paro, R. (1989). In vivo binding pattern of a trans-regulator of homoeotic genes in Drosophila melanogaster. Nature *337*, 468-471.