Identification and characterization of cis-regulatory elements that target Polycomb in the mouse genome

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Philip Jermann
aus Würenlingen, AG

Basel, 2014

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Dirk Schübeler und Prof. Dr. Renato Paro.

Basel, den 20. Mai 2014

Prof. Dr. Jörg Schibler

TABLE OF CONTENTS

Abbreviations

bp	basepair
CBX	Chromobox protein
CGI	CpG island
ChIP	Chromatin immunoprecipitation
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DSP1	Dorsal switch protein 1
EED	Embryonic Ectoderm development
ESC	Embryonic stem cell
EZH	Enhancer of Zeste
GAF	GAGA factor
GTF	General transcription factor
HAT	Histone acetyltransferase
HMT	Histone methyltransferase
HP1	Heterochromatic protein 1
kb	kilobase
MBD	Methyl-binding domain
MLL	Mixed lineage leukemia
NMR	Non-methylated region
NP	Neural progenitor
nt	nucleotide
PC	Polycomb
PcG	Polycomb group
PCGF	Polycomb group ring finger
PCL	Polycomb-like
PH	Polyhomeotic
PHO	Pleiohomeotic
PIC	Pre-initiation complex
Pol II	RNA polymerase 2
PRC	Polycomb repressive complex
PRE	Polycomb responsive element

PSC	Posterior sex comb
qPCR	quantitative polymerase chain reaction
RING	Ring finger protein
RMCE	Recombinase-mediated cassette exchange
RNA	Ribonucleic acid
SUZ	Suppressor of Zeste
TF	Transcription factor
TFBS	Transcription factor binding site
YY1	Yin yang 1

Protein names are in capital letters irrespective of species.

Gene names are in italics.

SUMMARY

Multicellular organisms consist of numerous cell types, each serving a specific function. Remarkably, almost all cells within an organism contain the same genetic information. Nevertheless, each cell type interprets this information differently, resulting in cell type specific gene expression patterns. These expression patterns define cellular function and are acquired upon lineage commitment of a pluripotent cell. Once acquired, these patterns can be stably maintained throughout subsequent cell divisions. For example, upon differentiation of a stem cell pluripotency-associated genes need to be silenced, while lineage-specific genes have to be activated. The maintenance and propagation of these expression patterns is thought to be mediated at least in part via the posttranslational modification of chromatin components (Kouzarides 2007). These covalent modifications are deposited by specialized enzymes that modify specific histone residues (Meissner 2010). However, while many of the enzymes responsible for establishing these marks have been identified, how they are targeted to specific loci remains unclear.

Polycomb-group (PcG) proteins represent key regulators of gene expression, especially during early development where they play key roles in the stable repression of developmental regulators (Di Croce and Helin 2013). They form several complexes that mediate the modification of distinct histones. For example, the PRC2 complex mediates trimethylation of histone H3 at lysine 27 (H3K27me3), a chromatin mark essential for proper development of both flies and mammals (Papp and Muller 2006). However, despite the importance of this modification, it remains elusive how H3K27me3 is targeted to specific loci. In *Drosophila melanogaster*, it has been demonstrated that transcription factors (TFs) play a major role in guiding PcG complexes to specific DNA elements, termed Polycomb responsive elements (PREs) (Ringrose and Paro 2004). Efforts to identify similar DNA elements in mammals have proven less successful, with only a handful of PREs known today (Sing et al. 2009, Woo et al. 2010). Furthermore, it is unclear whether the correlation between TF binding and PcG recruitment observed in *D. melanogaster* is indeed reflecting a direct physical interaction or rather an indirect crosstalk involving other factors.

In this study, we aimed to investigate the mechanisms that facilitate PRC2 recruitment and deposition of its associated H3K27me3 mark in mammals. We hypothesized that recruitment of PcG complexes to specific loci is encoded within the target DNA sequence either in the form of TF binding sites or other sequence queues. To test this, we employed a reductionist approach and inserted a set of endogenous PRC2 targets in mouse embryonic stem (ES) cells into a defined ectopic locus. We then examined whether these ectopically inserted DNA sequences could recapitulate the H3K27me3 levels observed at endogenous loci. Indeed, all of the tested elements were able to reconstitute endogenous PRC2 and H3K27me3 patterns. Further dissection of these elements revealed that DNA sequences rich in CpG dinucleotides and as short as 220 bp are sufficient to establish an H3K27me3 domain. Furthermore, we found that cell-type specific recruitment is determined by the transcriptional state of the target locus. In particular, transcriptional activity regulated by TF binding to a proximal cis-regulatory element can efficiently block the acquisition of H3K27me3. Finally, by systematically mutating the identified recruiter elements we demonstrate that DNA methylation directly prevents the recruitment of H3K27me3 to the underlying DNA sequence.

Taken together, we propose a model whereby PRC2 recruitment and H3K27me3 deposition defines a default chromatin signature at transcriptionally inactive and unmethylated genomic regions. Furthermore, we show that TFs are involved in the recruitment of PRC2 by controlling the transcriptional activity of the target locus. This study therefore provides novel insights into the relationship between different gene regulatory mechanisms and broadens our understanding of the crosstalk between TFs and epigenetic modifications.

CHAPTER 1

1. INTRODUCTION

The blue prints for all life on earth are stored within a genetic code defined by the nucleotide sequence of deoxyribonucleic acids (DNA). DNA molecules are double stranded helices with each strand carrying the entire genetic information. Due to the large size of eukaryotic DNA, it is systematically organized within the nucleus to access stored information in an efficient manner. To achieve this, it is wrapped around a core histone protein complex forming a structure termed nucleosome. Nucleosomes subsequently make up the building blocks of chromosomes. Remarkably, almost every cell type in a multicellular organism contains an identical set of chromosomes, yet may serve drastically different functions. For example, the human body consists of several hundred different cell types, each carrying out specific functions. The fact that they all possess the same DNA implies regulatory mechanisms that allow for cell type-specific interpretation of the genetic information. This is reflected by distinct gene expression patterns acquired by cells during development. These patterns are set up by transcription factors (TFs) that can interact with specific DNA sequences to either promote or repress transcriptional activity. In addition to TFs, gene expression can be regulated by modification of DNA and histones that make up the core of the nucleosome. Such modifications are thought to alter DNA accessibility, leading to changes in gene expression.

Taken together, TF binding and chromatin modifications act in concert to dictate the specific gene expression patterns unique to each cell type. It is thus of critical interest to identify functional crosstalk between these processes in order to better understand gene regulation. Throughout the following chapters I will first discuss gene regulatory principles and then outline new insights into the crosstalk between TFs and epigenetic modifications.

1.1 Principles of gene regulation: from bacteria to mammals

Pioneering work in prokaryotes by Jacob and Monod in the 1960s revealed that a genetic locus contains three fundamental parts in addition to the coding sequence; the promoter sequence that is recognized by RNA polymerase, operator sequences that are bound by repressors in order to inhibit transcription, and activator elements that can be bound by factors that stimulate transcription (Jacob and Monod 1961). The authors proposed that in prokaryotes the ground state of transcription is non-restrictive, meaning that in the absence of repressors and activators transcriptional activity is determined by the quality of the promoter sequence alone. The complete silencing of a prokaryotic gene therefore requires the presence of a repressor. It was later demonstrated that repressors act by binding to specific DNA sequences and block the binding of RNA polymerase to the promoter (Ptashne 1967). Activators, on the other hand, are not required for basal transcriptional activity in prokaryotes and act only on generally weak promoters. They are able to stimulate transcriptional activity by directly interacting with RNA polymerase, resulting in either recruitment of the polymerase to the promoter, or stimulation of already bound polymerase (Hochschild and Dove 1998).

While bacterial RNA polymerase requires only one additional cofactor to initiate transcription, eukaryotic RNA polymerase requires many additional proteins that help position the polymerase to the gene promoter (Figure 1). One group of these factors consists of the general transcription factors (GTF), including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. These GTFs form the pre-initiation complex (PIC) with the help of another multi-subunit protein complex called the mediator complex (Figure 1) (Conaway and Conaway 2011). This assembly of protein complexes results in the correct positioning of RNA polymerase II (Pol II) at the transcription start site. Eukaryotic core promoters harbor specific DNA sequence elements in close proximity to the transcription start site, which can be recognized by the GTF TFIID in order to recruit additional subunits of the PIC (Maston et al. 2006). However, in contrast to prokaryotes, the ground state of transcription in eukaryotes is generally restrictive. This is mostly due to the presence of nucleosomes, which have been shown to act inhibitory to the binding of the transcriptional machinery (Felsenfeld 1992, Workman and Kingston 1998). Thus, transcription in eukaryotes

requires additional factors that help facilitate PIC assembly and subsequent transcription (Lorch et al. 1987, Morse 1989).

Such factors are known as activating transcription factors (TFs), which bind DNA in a sequence-specific manner. TFs can be classified into different groups based on the structure of their DNA binding domains (DBDs) that enable them to interact with DNA in a sequence specific manner. Usually, such TF binding sites (TFBS) are short and range between 6 and 12 basepairs (bp). Stimulation of PIC assembly and transcriptional activity is mediated either via direct interaction with specific subunits of the PIC, resulting in recruitment to the promoter, or by factors that modify chromatin to increase accessibility of the underlying DNA (Figure 1) (Kuras and Struhl 1999, Struhl 1999, Levine and Tjian 2003). This recruitment function of TFs is achieved via a second domain, the activating domain (AD). A prominent example is the yeast TF GAL4; when a GAL4 binding site is placed upstream of a reporter gene, GAL4 can bind to it and activate transcription of a gene driven by an otherwise silent promoter (Giniger et al. 1985). Furthermore, the function of the DBD and the AD can be separated from each other as they act independently, demonstrated in an elegant experiment where the GAL4 AD was fused to the DBD of the bacterial repressor LEX A (Brent and Ptashne 1985). This fusion protein is able to interact with and recruit the transcriptional machinery to a gene harboring LEX A binding sites and drive transcription of the gene via its GAL4 AD (Brent and Ptashne 1985). Remarkably, expression of GAL4 in other eukaryotes, including mammals, leads to activation of transgenes harboring GAL4 binding sites in proximity to the otherwise inactive promoter (Ptashne 2005). This indicates that the general principles of gene regulation by TFs are highly conserved in eukaryotic organisms. However, most mammalian TFs do not work alone to recruit the PIC. Instead, they need to act in concert to control tissue-specific gene expression, either via interactions between multiple copies of the same factor (Carey et al. 1990) or cooperation between different factors (Lin et al. 1990). This allows for much tighter regulation of expression levels, which is needed in multicellular organisms where cells need to communicate and interact with each other and serve distinct functions.

A third set of factors, the co-activators, can mediate interactions between TFs and the PIC (Figure 1). In contrast to TFs, co-activators do not directly interact with DNA but rather act via protein-protein interactions to modulate transcriptional activity (Lemon and Tjian 2000). Interaction between different TFs and co-activators is

required in order to orchestrate the crosstalk between various regulatory elements in higher eukaryotes. Metazoan regulatory modules contain many distal elements termed enhancers that can be up to 100 kilobases away from the gene (Ong and Corces 2011). These enhancer elements are brought into close proximity of promoters by looping the intervening DNA sequence (Figure 1) (Schoenfelder et al. 2010).

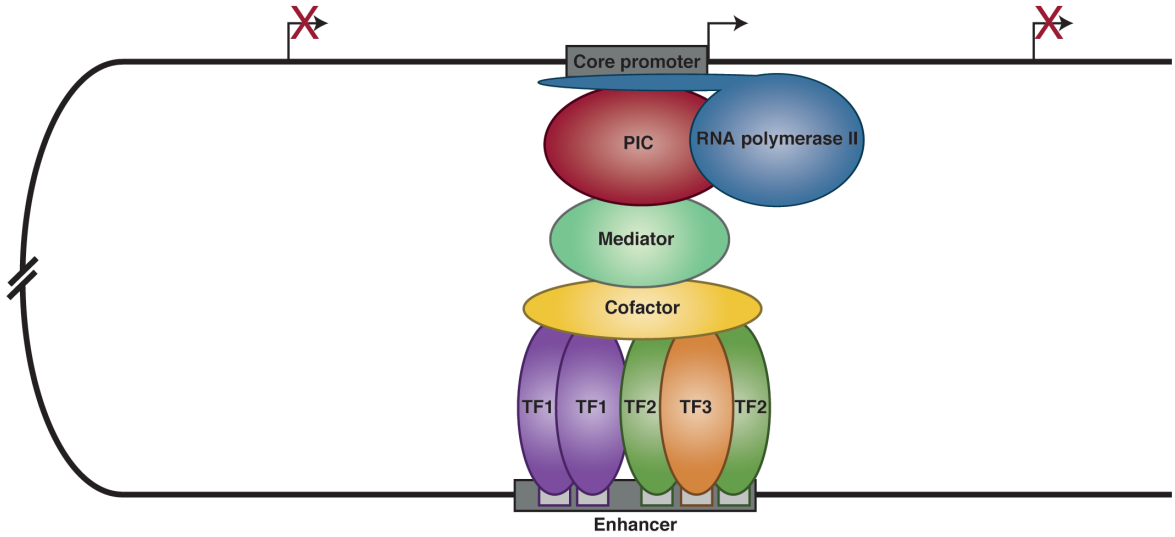

Figure 1 Basic mechanisms of metazoan gene regulation. The expression of genes in higher eukaryotes is regulated by transcription factors (TFs) that interact with various proximal and distal regulatory DNA sequences. Many of these factors act in concert to recruit cofactors that bring distal elements in proximity to the core promoter via the mediator complex. This results in the recruitment of the pre-initiation complex (PIC) and RNA polymerase II and subsequent transcription of the gene. Adapted from (Maston et al. 2006).

The human genome codes for around 1'700 – 2'000 TFs (Vaquerizas et al. 2009). This repertoire allows for a vast amount of combinatorial information, forming complex gene regulatory networks able to direct and maintain all the different gene expression patterns that define cellular identity.

As mentioned above, activation of eukaryotic transcription requires a permissive chromatin structure to make core promoters accessible for the transcriptional machinery. This is thought to be achieved via recruitment of specific chromatin modifying factors, whose enzymatic activity can either positively or negatively regulate transcriptional activity. In the following sections I will outline the mechanisms involved in these processes in order to provide a more detailed view on eukaryotic gene regulation.

1.2 Chromatin as a means of organizing DNA

The human genome consists of roughly $3 \cdot 10^9$ nucleotide pairs coding for 20'000- 30'000 genes and when fully extended measures around two meters. In order to organize such large molecules inside a nucleus measuring only roughly 6 µm in diameter (Alberts et al. 2008) DNA is wrapped around a protein complex and further folded into entities called chromosomes (Figure 2). This chromosomal structure can be observed in condensed mitotic and meiotic chromosomes during metaphase alignment. The proteins around which DNA is wrapped are called "histones" and were first described not long after the initial discovery of nucleic acids by Friedrich Miescher in 1871 [reviewed in (Dahm 2005)]. They were observed by Albrecht Kossel upon extraction of components of nucleated erythrocytes (Kossel 1884). The term chromatin was first used by Walther Flemming because he observed that the nucleus of a cell absorbed basophilic dyes (Flemming 1882). It took nearly a century before the single components of chromatin started to emerge. In the 1970s, it was demonstrated that chromatin digested with exogenous nucleases left roughly half of the DNA intact and protected. Remarkably, these protected regions were all between 100 – 200 bp long, leading to a model whereby chromatin is built up by a basic repeating structure (Clark and Felsenfeld 1971, Hewish and Burgoyne 1973). Electron microscopy and biochemical studies later defined the structure of chromatin to be a flexible chain of spherical particles that were termed nucleosomes (Oudet et al. 1975). It was demonstrated that a nucleosome consists of about 200 bp of DNA and four distinct core histones in a 1:1 ratio. Moreover, the authors showed that the identified building blocks of chromatin had the ability to self-assemble *in vitro* (Oudet et al. 1975).

Today, we know that four core histones, namely H2A, H2B, H3 and H4 form an octamer with two of each histone present. This structure represents the core particle of the nucleosome, around which 147 bp of DNA are wrapped 1.65 times (Figure 2) (Luger et al. 1997). Individual nucleosomes are separated from each other by 10 - 80 bp of linker DNA (Kornberg 1974, Luger et al. 1997). This "beads-on-a-string" structure (Figure 2) measures 11 nm in diameter and is mostly observed transcriptionally active regions of the genome. At inactive sites, it is thought to be further folded into a fiber roughly 30 nm in diameter via addition of the H1 linker

histone (Widom and Klug 1985). However, to date, such a 30 nm fiber has not been observed *in vivo*, despite the fact that artificial nucleosome arrays formed *in vitro* tend to acquire such a structure (Tremethick 2007).

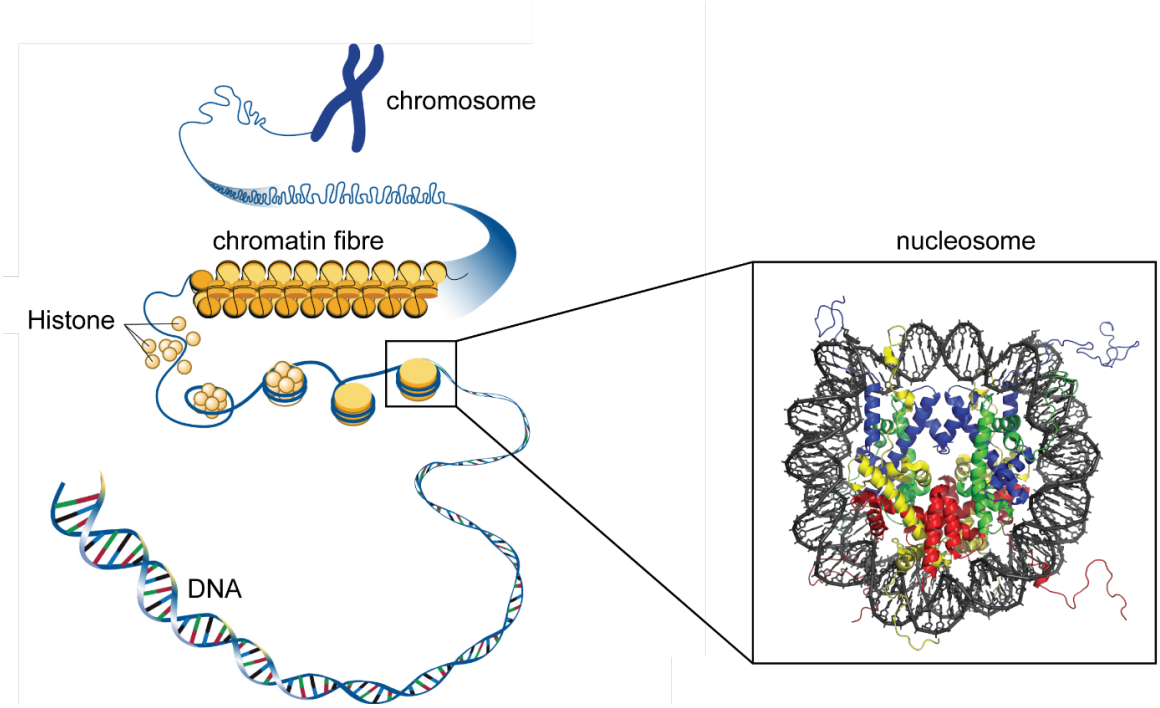

Figure 2 Chromatin compaction and nucleosome. DNA is located in the nucleus of each cell and undergoes several layers of compaction, forming chromatin. The basic subunit of chromatin is the nucleosome (right), consisting of an octamer of histones H2A, H2B, H3, and H4. An array of nucleosomes is then further organized into a chromatin fiber and eventually into a chromosome. Adapted from (Luger et al. 1997) and Darryl Leja, NHGRI.

1.2.1 Different states of chromatin

Chromatin can be divided into two distinct classes, heterochromatin and euchromatin. Initially, heterochromatin was defined as regions within nuclei that stained strongly with basic dyes (Figure 3). Strong staining at these domains indicated a state of high compaction. In contrast, euchromatin is more loosely structured. Today, the term heterochromatin is mainly applied to condensed and inactive regions of the genome, while euchromatin depicts sites of active transcription. However, within heterochromatin, a distinction is made between constitutive and facultative heterochromatin. Constitutive heterochromatin demarcates genomic loci that are always silent and condensed such as repetitive elements, gene-poor regions, and late replicating sequences (Trojer and Reinberg 2007). Facultative heterochromatin, on the other hand, is also transcriptionally silent, but can be decondensed and activated in response to specific stimuli (Trojer and

Reinberg 2007). To switch from an inactive heterochromatic state to a more accessible euchromatic state, chromatin structure has to be decondensed. To achieve this, specific protein complexes are able to interact with and modify nucleosomal entities. Thus, in addition to DNA sequence motifs that can be recognized by TFs, chromatin structure and organization plays a key role in the regulation of eukaryotic gene expression. In the next section I will expand on the processes involved in this type of gene regulation.

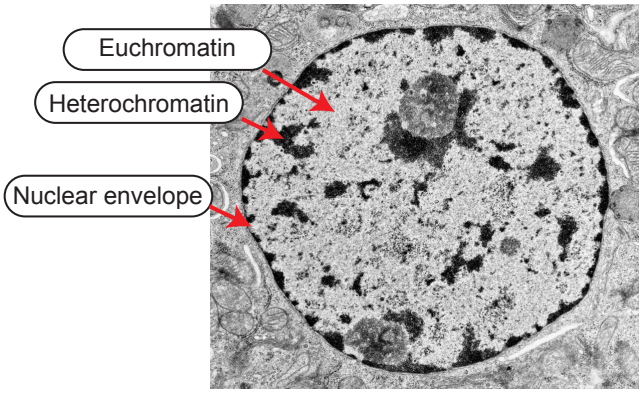

Figure 3 Nucleus staining showing hetero- and euchromatin. Histology slide depicting a cell nucleus stained for chromatin. The bright regions indicate active euchromatin and the black regions represent inactive condensed heterochromatin. Adapted from Histology Department of Yale University. Adapted from http://medcell.med.yale.edu/histology/cell_lab/images/euchromatin_and_heterochromatin.jpg

1.3 Epigenetic regulation of gene expression

Regulation of heritable gene expression patterns via modification of chromatin components is commonly described as “epigenetic”. This term was first shaped by Conrad Waddington to describe changes in gene expression that could not be explained by genetics (Waddington 1942). He therefore suggested epigenetics to be the “branch of biology that studies the causal relationship between genes and their products, which bring the phenotype into being” (Waddington 1942). Today, epigenetics is defined as processes that can heritably change the output of a genetic locus without altering the underlying DNA sequence. For example, most cells of a multicellular organism harbor the same genetic material (*i.e.* DNA sequence), yet they differ from one another significantly in regard to how they interpret this material. Moreover, gene expression patterns can be stably maintained and propagated to the daughter cell, even in the absence of the initial stimulus (*i.e.* TF binding). This maintenance is thought to be at least in part mediated by epigenetic modifications.

However, the mechanisms underlying epigenetic processes are still unclear. In the following sections, I will introduce the basic concepts of epigenetic gene regulation and outline the recent advances in our understanding thereof.

1.3.1 DNA methylation

The first epigenetic mark discovered was the covalent modification of DNA by addition of a methyl group at the 5th carbon of cytosine rings. At promoters, this modification can trigger transcriptional repression, mediated either directly by blocking TF-binding (Figure 4) (Iguchi-Ariga and Schaffner 1989) or via recruitment of proteins that specifically recognize methylated DNA (Nan et al. 1993). On the other hand, DNA methylation within mammalian gene bodies has been shown to correlate with transcriptional activity (Hellman and Chess 2007, Zilberman et al. 2007, Ball et al. 2009). DNA methylation has been implicated in several cellular processes such as genomic imprinting, X-chromosome inactivation in females, and suppression of repetitive elements [reviewed in (Weber and Schubeler 2007)]. In vertebrates, DNA methylation is exclusively deposited at cytosines that are followed by a guanine (CpG). In contrast, cytosine methylation in plants can occur in all sequence contexts (Henderson and Jacobsen 2007). Interestingly, the two invertebrates *Caenorhabditis elegans* and *Drosophila melanogaster* both lack DNA

methylation, while most other invertebrates show mosaic patterns of DNA methylation throughout the genome (Bird et al. 1979, Tweedie et al. 1997). Similar observations have been made in plants where DNA methylation is restricted to repetitive elements and gene bodies (Martienssen and Colot 2001, Zilberman et al. 2007). Vertebrates, on the other hand, display genome-wide DNA methylation outside of genic regions with around 80% of all CpGs methylated (Singer et al. 1979, Tweedie et al. 1997). Exceptions to this global methylation are active promoters, distal regulatory regions, and regions harboring a high density of CpG dinucleotides (Stadler et al. 2011).

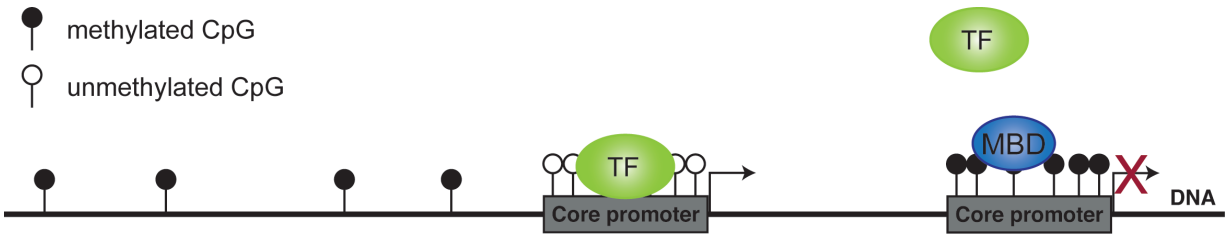

Figure 4 DNA methylation at promoters inhibits transcription. Schematic representation of the effects of DNA methylation on target gene expression. Unmethylated promoters (left) can be bound by activating TFs and therefore engage in transcription of the associated gene. On the other hand, methylation of a promoter (right) is generally associated with gene repression. This involves the binding of specific factors that contain an MBD domain that recognizes the methyl-group on the cytosine.

In the 1970s, CpG methylation was first suggested to be a mechanism of cellular memory when two independent groups showed that methylation patterns are copied to the daughter strand during DNA replication (Holliday and Pugh 1975, Riggs 1975). The authors proposed a mechanism whereby an enzyme would recognize existing patterns of DNA methylation and faithfully copy them during cell division. Such a mechanism would allow for the propagation of existing gene expression patterns even in the absence of the initial signal. Indeed, in 1983 Bestor and Ingram identified DNMT1, an enzyme with methyltransferase activity preferentially towards hemimethylated DNA (Bestor and Ingram 1983). It was later demonstrated that lack of DNMT1 leads to a global loss of DNA methylation in mouse embryonic stem cells (Li et al. 1992). Thus, DNMT1 was found to be the major enzyme responsible for maintaining DNA methylation patterns in mammals.

In 1982 Stewart et al. demonstrated that insertion of viral DNA into somatic cells resulted in expression of the viral genes. However, the viral genes were not expressed when inserted into pre-implantation embryos or mouse embryonic stem cells (ESCs), (Stewart et al. 1982). It was found that viral DNA became methylated

upon insertion in ESCs, which caused the silencing of viral gene expression. These experiments suggested the existence of enzymes capable of *de-novo* methylation, rather than just maintenance. In support of this, deletion of DNMT1 did not affect *de-novo* methylation of the inserted viral DNA in stem cells, supporting the idea of a separate set of methyltransferases (Lei et al. 1996). Sequence homology searches for the enzymatic domain of DNMT1 revealed several candidate proteins for *de-novo* methyltransferase activity. *In vitro* experiments confirmed that both DNMT3A and DNMT3B could *de-novo* methylate DNA (Okano et al. 1998). Loss of function experiments showed that cells lacking both of these proteins were no longer able to methylate and silence viral genes (Okano et al. 1999).

In summary, we now know that the enzymes DNMT3A and DNMT3B are able to *de-novo* methylate DNA during early development, while DNMT1 is responsible for maintaining the existing methylation patterns upon cell division (Figure 5) (Weber and Schubeler 2007).

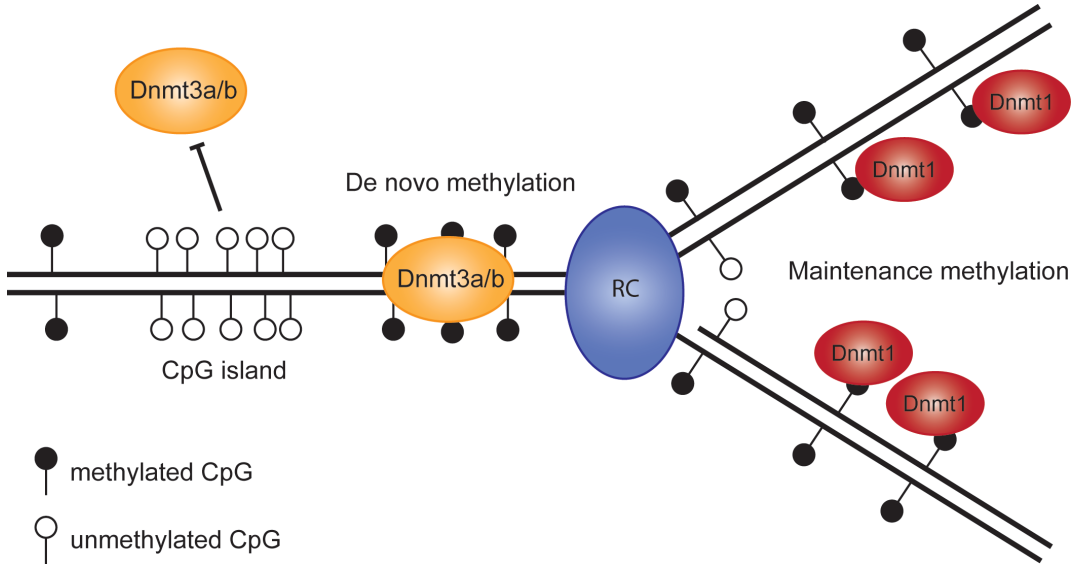

Figure 5 Mechanisms of de novo and maintenance methylation. Schematic depiction of the mechanisms involved in establishing and propagating DNA methylation states. Dnmt3a and Dnmt3b are cooperatively establishing new methylation marks. CpG dense regions are protected against such *de novo* methylation. During replication, preexisting DNA methylation patterns are copied onto the newly synthesized DNA strands by Dnmt1. RC: Replication complex.

1.3.1.1 CpG islands

Throughout the mammalian genome, CpG dinucleotides occur at only around one-fifth of the expected frequency (Russell et al. 1976). This is mainly attributed to spontaneous deamination of 5-methylcytosine, which converts methylated cytosine into a thymine and subsequent inefficient repair of G - T mismatches (Bird 1980).

This, together with the fact that around 80% of all CpGs in mammalian genomes are methylated has lead to a gradual global depletion of CpG dinucleotides over the course of evolution. In contrast, organisms that lack DNA methylation in the germline, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, harbor CpG dinucleotides at the expected frequency throughout the genome (Takai and Jones 2002). Despite genome-wide depletion of CpGs in mammals, there are regions in the genome where CpG dinucleotides cluster together and occur at the expected frequency. These CpG-rich regions are called CpG islands (CGI) and are unmethylated in most tissues (Bird 1986). Both the human and mouse genome contain roughly 25'000 CGIs with an average size of around 1 kb and roughly half of them lie within gene promoters (Illingworth et al. 2010). In fact, around 60% of all annotated gene promoters are associated with a CGI, making this the dominant type of promoter throughout the mammalian genome (Saxonov et al. 2006). Interestingly, however, this feature appears to be specific to warm-blooded vertebrates as only around 10% of promoters in cold-blooded vertebrates overlap with CGIs (Sharif et al. 2010). It has been proposed that CGI promoters are associated with genes that show similar activity in multiple tissues, often housekeeping genes, while non-CGI promoters represent tissue-specific genes that are activated upon distinct external stimuli (Sharif et al. 2010).

CGIs that do not lie within promoters are distributed throughout the genome mostly within gene-bodies or in intergenic regions. The function of these islands has been somewhat elusive, but several lines of evidence suggest that they are associated with transcriptional regulation of non-coding RNAs. For example, imprinting of the *Igf2r* gene is dependent on a non-coding transcript that initiates at a CGI within the *Igf2r* gene-body (Sleutels et al. 2002). Genome-wide analyses have revealed that around 40% of non-promoter CGIs are associated with transcriptional initiation and could therefore represent novel promoters (Illingworth et al. 2010, Maunakea et al. 2010). This number might increase in the future as more cell types are analyzed for transcriptional initiation.

The answer to the question how and why CGIs have emerged during evolution is still unclear. There are two possibilities that are currently being debated (Deaton and Bird 2011): (1) CGIs arose because they are enriched for cis-regulatory regions that are active in the germ-line and thus kept unmethylated. This would allow them to evade the gradual C to T erosion. (2) CGIs contain important regulatory

sequence motifs for TFs that have evolved specifically in organisms that harbor DNA methylation. Recent evidence points toward the second possibility, as factors have been identified that contain a specific structural domain (CXXC), which recognizes unmethylated CpG dinucleotides (Long et al. 2013).

Likewise, the mechanisms protecting CGIs from DNA methylation are still elusive, but are likely dependent on CpG density, presence of TFs, and modification of other chromatin components (Ooi et al. 2007, Weber et al. 2007, Lienert et al. 2011, Rose and Klose 2014). Despite the general absence of DNA methylation at CGIs, there are exceptions where CGIs get *de-novo* methylated during lineage commitment of pluripotent cells (Stein et al. 1982, Mohn et al. 2008, Payer and Lee 2008). However, it is thought that prior to DNA methylation, the target promoters become transcriptionally inactive via other mechanisms, including the modification of chromatin structure by post-translational modifications of histones (Mohn et al. 2008). This suggests a complex crosstalk between DNA methylation and other epigenetic modifications, some of which I will introduce in the following sections.

1.3.2 Chromatin remodeling and histone modifications

The key players that facilitate changes in chromatin structure are chromatin remodelers and chromatin modifiers. Chromatin remodelers consume energy in the form of ATP to physically expose the DNA masked within the nucleosome and thus make it accessible to DNA-binding proteins such as TFs (Clapier and Cairns 2009). This can be achieved by repositioning of existing nucleosomes, removal of a nucleosome, or just temporary unwrapping of DNA from the histone octamer. All chromatin remodelers have a conserved catalytic ATPase domain and share the ability to interact not only with DNA but also directly with histones. Chromatin remodelers can be classified into different families based on the presence of distinct domains that allow them to recognize specific post-translational modifications of histone proteins (Clapier and Cairns 2009).

In addition to chromatin remodelers, DNA accessibility can be regulated via post-translational modification of histones. The general structure of the four core histones is divided into two main parts, the globular domain and the N-terminal unstructured domain (Figure 6). Within the nucleosomal histone octamer, the N-terminal domains protrude from the central structure, making them accessible to

protein complexes that can interact with and chemically modify them (Figure 6) (Luger et al. 1997). It is thought that once established, these histone modification patterns can be propagated to the daughter cell upon division (Meissner 2010). Such a mechanism would allow for the stable maintenance of gene expression patterns during embryonic development even in the absence of the initial signal. In support of this model, loss of certain chromatin modifiers has been shown to impair ESC differentiation and cause embryonic lethality (Aloia et al. 2013).

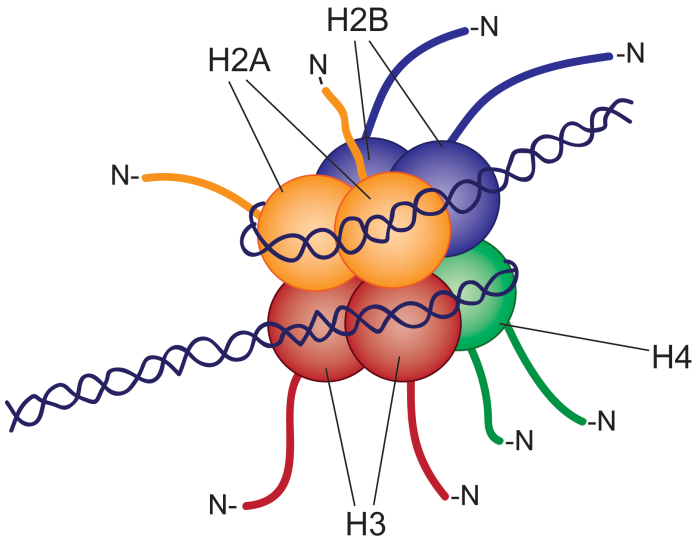

Figure 6 Histone octamer with protruding N-terminal tails. Two H2A-H2B and two H3-H4 dimers form an octamer, which represents the core of the histone. DNA is wrapped around this octamer 1.65 times to form the nucleosome. Each histone protrudes the nucleosome via its unstructured N-terminal domain. These histone tails can be accessed and modified by specific factors, leading to changes in the organization of chromatin structure.

Well over 50 histone modifications have been identified to date (Figure 7) (Kouzarides 2007). They include acetylation, methylation, ubiquitination, and phosphorylation, each having distinct effects on the surrounding chromatin environment (Koch et al. 2007, Kouzarides 2007). The first discovered histone modification was acetylation of lysine residues (Phillips 1963) and was suggested to be associated with active gene expression and proposed to increase DNA accessibility by neutralizing the positive charge of histone tails (Pogo et al. 1966, Hong et al. 1993, Megee et al. 1995). However, these conclusions were mostly driven by correlative data and it was not until the identification of the first histone acetyltransferase (HAT) in yeast that a direct relationship between histone acetylation and transcriptional activity was shown (Brownell et al. 1996). Concomitant with the discovery of the first HAT, Taunton et al. identified the first

histone deacetylase (HDAC), HD1, responsible for reversing the acetylation mark and promoting gene repression (Taunton et al. 1996).

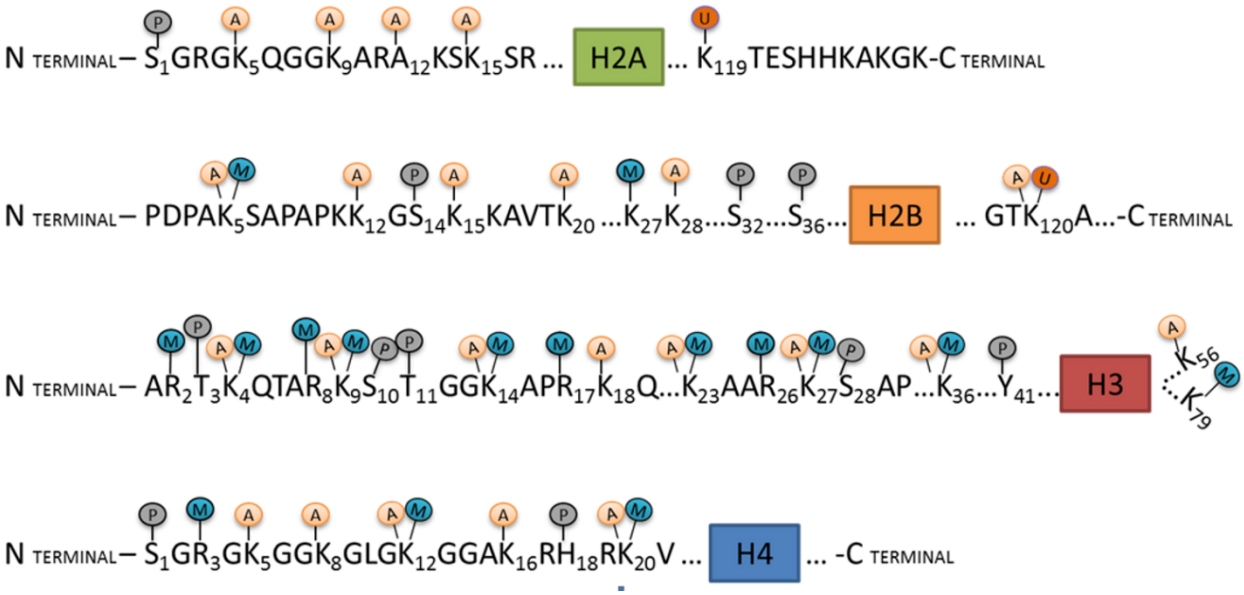

Figure 7 Selection of posttranslational modifications of N-terminal histone tails. Indicated are some of the known posttranslational modifications of specific amino acid residues within the N-terminal ends of the core histone proteins. A: acetylation, P: phosphorylation, U: ubiquitination, M: methylation. Adapted from (Xu 2013)

In addition to altering the DNA-histone contacts, histone modifications also generate docking sites for nuclear proteins. The first histone modification-reader was discovered in 1999 when Dhalluin et al. identified the bromodomain in the HAT P300/CBP-associated factor (Dhalluin et al. 1999). This particular domain within the protein forms a specific structure that can bind the acetyl group on the histone tail. Since this initial discovery, several additional histone modification readers have been identified, including readers of histone methylation and phosphorylation (Taverna et al. 2007, Musselman et al. 2012). Remarkably, many of these factors not only recognize a specific modification, but also the sequence surrounding them, therefore increasing their specificity.Once bound, these factors can recruit larger complexes which can then remodel nucleosomes or further modify histones (Musselman et al. 2012).

Due to the vast amount of possible combinations of histone marks on a single nucleosome, the existence of a “histone code” has been proposed (Strahl and Allis 2000). Recent advances in DNA sequencing technologies coupled with modification-specific antibodies enabled genome-wide analysis of chromatin modifications (Filion et al. 2010, Hawkins et al. 2010, Zhu et al. 2013). Chromatin maps in different cell

types have revealed association of specific histone marks with distinct gene expression patterns. However, efforts to dissect the histone code in *Drosophila melanogaster* have led to a rather simple classification of chromatin in just five major groups, despite the large number of theoretically possible combinations of chromatin marks (Filion et al. 2010). Studies in human cell lines have yielded similar results, yet a more variable number of chromatin states, ranging from 6 to 51 (Ernst and Kellis 2010, Ram et al. 2011, Hoffman et al. 2013).

Taken together, these studies define a set of common chromatin states including promoters marked by trimethylated histone H3 at lysine 4 (H3K4me3) and bound by polymerases, transcribed regions marked by trimethylated H3K36, enhancers characterized by monomethylated H3K4 and acetylated H3K27, Polycomb repressed regions marked by trimethylated H3K27, and heterochromatic regions characterized by the presence of trimethylated H3K9 (Ram et al. 2011, Hoffman et al. 2013).

In summary, histone modifications add an additional layer of information and are thought to aid in maintaining cell type-specific gene expression patterns. In the next section, I will describe the process of histone methylation in more detail and outline its relevance in gene regulation.

1.3.2.1 Histone methylation

Histone methylation is the process of covalently adding up to three methyl groups from the donor S-adenosylmethionine on the side-chains of lysine, arginine, and histidine. Methylation of lysine residues in histone proteins was first demonstrated in the 1960s (Allfrey and Mirsky 1964, Murray 1964). However, the first enzyme capable of transferring methyl groups onto histones, SUV39H1 was only recently discovered and has been demonstrated to be conserved from yeast to human (O'Carroll et al. 2000, Rea et al. 2000). Since then, additional histone methyltransferases (HMTs) have been identified based on homology searches (Black et al. 2012).

This discovery of many HMTs has lead to their classification into three main groups of enzymes that mediate histone methylation; the lysine-specific HMTs contain a 130 amino acid catalytic SET domain and mediate methylation of lysines 4, 9, 27, and 36 of histone H3 and lysine 20 of histone H4. The second group of HMTs

contains no SET domain and is involved in the methylation of lysine 79 of histone H3. The third set of HMTs is arginine-specific and methylates arginines 2, 17, and 26 of histone H3 and arginine 3 of histone H4. The HMTs in each group are highly specific for distinct amino acid residues within the histone tails and for the degree of methylation. For example, the lysine-specific HMTs SUV39H1/H2 specifically recognize H3K9 and di- or trimethylate it from a monomethylated state (Peters et al. 2001, Peters et al. 2003) while the HMT G9A preferentially mono- and dimethylates H3K9 (Tachibana et al. 2002). Similarly, MLL1 dimethylates H3K4, but when it associates with Ash2L and RbBP5 it is able to trimethylate the same lysine (Dou et al. 2006). Thus, intrinsic features of the HMTs as well as their interaction partners can regulate the amino acid specificity and the preferred degree of methylation.

In contrast to DNA methylation, histone lysine-methylation can be reversed by specific enzymes termed lysine demethylases (KDMs). The first discovered KDM was LSD1, which mainly reverses H3K4 methylation, but also shows some affinity towards methylated H3K9 when it interacts with the androgen receptor (Shi et al. 2004, Metzger et al. 2005). Subsequently, other KDMs were discovered and were classified into three distinct groups, the largest of which consists of the Jumonji C-domain containing KDMs (Klose et al. 2006).

Histone lysine methylation can have both activating and repressive effects on chromatin. As described above, H3K4 di-and trimethylation mediated by MLL1/2 and SET1 correlates well with gene activity throughout the genome and localizes mainly to active gene promoters and enhancer regions (Santos-Rosa et al. 2002, Schubeler et al. 2004, Barski et al. 2007, Mikkelsen et al. 2007). Functionally, it has been implicated in transcriptional elongation by interacting with Pol II in its initiated form (serine 5 phosphorylated). Additionally, TFIID can directly interact with H3K4me3 via its PHD domain, further emphasizing the crosstalk between histone modifications and TFs (Vermeulen et al. 2007). Due to the genome-wide anti-correlation with DNA methylation, especially at CGIs, H3K4me3 has been proposed to be a key factor in preventing DNA methylation at these sites (Ooi et al. 2007, Weber et al. 2007). H3K4me3 can also be found at inactive promoter regions in embryonic stem cells, where it co-localizes with the repressive H3K27me3 mark to form what has been termed bivalent domains (Bernstein et al. 2006). It has been proposed that it primes these promoters for later activation during lineage-commitment. In addition to H3K4 di- and trimethylation, monomethylation robustly demarcates distal cis-regulatory

regions. Together with p300-binding and H3K27 acetylation it predicts the genome-wide location of enhancer elements in the mammalian genome (Heintzman et al. 2009, Rada-Iglesias et al. 2011).

H3K36 trimethylation mediated by the HMT SETD2 also correlates with gene expression and is a good indicative marker for transcriptional activity (Tippmann et al. 2012). It localizes to the gene body, peaks near the 3' end of the gene, and associates with the elongating serine 2 phosphorylated form of Pol II (Bannister and Kouzarides 2011). H3K36me3 is believed to suppress inappropriate transcriptional initiation from cryptic start sites within the coding region of the gene, in part via the recruitment of DNA methylation and histone H3K4me3 demethylases to gene bodies (Carrozza et al. 2005, Joshi and Struhl 2005, Keogh et al. 2005, Fang et al. 2010). In contrast, H3K79 is less studied and believed to be involved in the activation of certain *Hox* genes and has a role in the DNA damage response and telomere silencing (Nguyen and Zhang 2011).

Histone lysine methylation is not only associated with gene activation, but can also serve repressive functions. In particular, methylation of H3K9, H3K27 and H4K20 has been demonstrated to play key roles in the formation of silent heterochromatin. H3K9 methylation is associated with the formation of constitutive heterochromatin, mainly at repetitive regions, including satellite sequences, ribosomal RNA clusters and pericentromeric chromatin (Mikkelsen et al. 2007, Filion et al. 2010, Ernst et al. 2011, Riddle et al. 2011). Di- and trimethylation of H3K9 is recognized by HP1, which forms a dimer and binds to the methyl mark via its chromodomain, resulting in the stabilization and spreading of heterochromatin (Hall et al. 2002).

H3K27 trimethylation is set by the HMT EZH2 in complex with EED and SUZ12 and marks inactive promoter regions of developmental regulators (Kuzmichev et al. 2002, Muller et al. 2002, Cao and Zhang 2004). While the H3K27me3 domains in ESCs are of short focal nature around promoters, these short domains can spread into larger domains upon differentiation (Hawkins et al. 2010, Zhu et al. 2013). The mechanisms involved in H3K27me3 mediated gene repression are discussed in more detail in the following section.

In summary, methylation of histone tails is a good indicator of transcriptional and regulatory activity. Together with DNA methylation, these modifications add an additional layer to the already complex regulatory network of transcription factors

(Figure 8). However, despite extensive mapping of these chromatin marks, the mechanisms whereby chromatin modifications influence transcriptional output are still poorly understood. It is important to investigate the relationship between all these factors and how they influence each other in order to better understand the regulatory cascade that leads to the phenotypic output of a genetic locus.

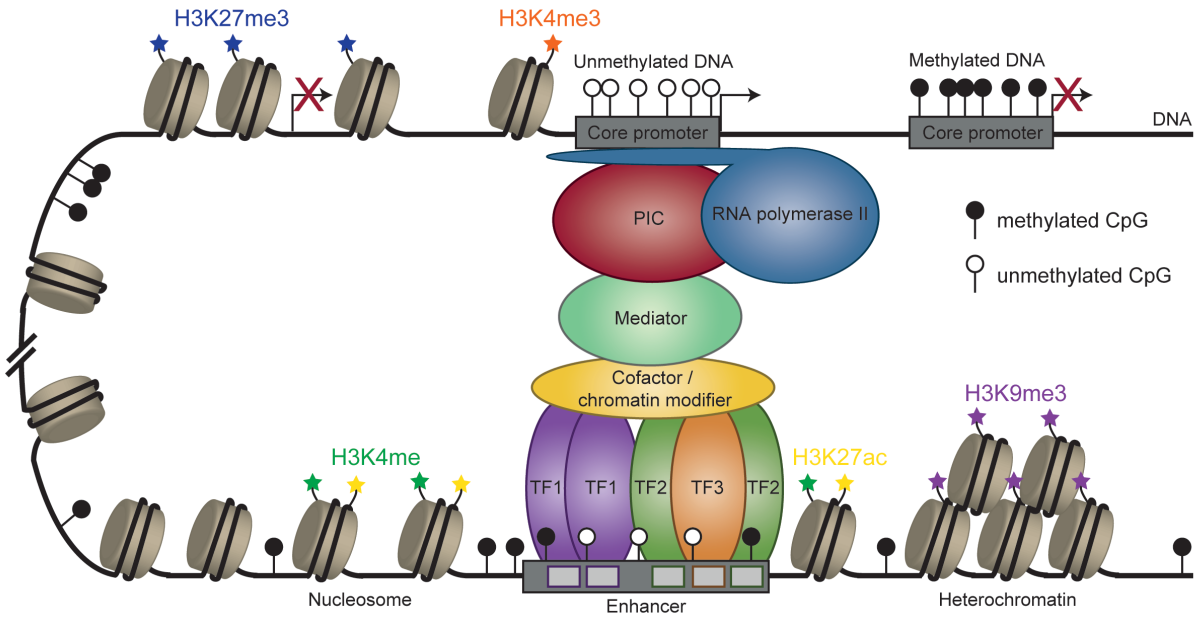

Figure 8 Complex interplay between TFs, DNA methylation and chromatin modifications leads to tightly regulated transcriptional output. Schematic representation of the interplay between epigenetic chromatin modifications and the transcriptional machinery. Inactive promoters are marked by H3K27me3, H3K9me3 or DNA methylation, while active promoters display an enrichment for H3K4me3. Distal regulatory elements such as enhancers are marked by H3K4 monomethylation and H3K27 acetylation and reduced levels of DNA methylation. Adapted from Anaïs Bardet and Maston et al. Annu. Rev. Genomics. Hum. Genet. 2012.

So far, I have discussed the importance of gene regulation in maintaining cell identity and the key processes involved in this regulation among which the modification of histones and DNA play a major role. But what regulates the deposition of these modifications? How does a cell decide when and where to modify chromatin?

1.4 Polycomb group proteins: developmental regulators

Probably the most studied group of chromatin modifiers in terms of genomic targeting is the polycomb group (PcG) of proteins. PcG proteins are key epigenetic regulators and play central roles in the regulation of genes involved in embryonic development and differentiation (Aloia et al. 2013). They were identified in the 1970s as regulators of homeotic (Hox) gene expression during embryonic development of

Drosophila melanogaster (Lewis 1978). *Hox* genes code for transcription factors that are required for shaping the body patterning and segmentation of fly embryos. Their expression is set up by early TFs that quickly decay after initiating the *Hox* gene expression patterns. These patterns, however, are maintained throughout development into adult stages (Moehrle and Paro 1994). Mutations in *Hox* genes cause characteristic phenotypes manifested in the aberrant development of body structures such as copies of structures that normally develop in different segments of the embryo. Accordingly, PcG mutants display defects in body-patterning where anterior segments are transformed toward more posterior segments and were therefore named after the observed phenotypes (Sato and Denell 1985). Because loss of PcG protein activity generally results in ectopic expression of *Hox* genes, they were classified as repressors. Furthermore, it was observed that PcG proteins do not act by themselves, but rather form complexes (Shao et al. 1999, Czermin et al. 2002, Kuzmichev et al. 2002). This was confirmed by studies that showed biochemical co-fractionation and cytological co-localization of PcG proteins (Shao et al. 1999, Cao et al. 2002). The first member of PcG proteins that was cloned and characterized was Polycomb (PC) (Paro and Hogness 1991). It was proposed to act on the level of chromatin because it harbors a chromodomain similar to HP1 and is thus able to interact with modified histones. Crosslinking PcG proteins to DNA by formaldehyde treatment further confirmed the direct interaction of PC with *Hox* gene loci (Orlando and Paro 1993).

Orthologs of *Drosophila* PcG proteins have since been identified in many multicellular organisms ranging from plants to humans (Schumacher and Magnuson 1997, Ross and Zarkower 2003, Hennig and Derkacheva 2009, Surface et al. 2010). Many of their functions appear to be conserved as outlined by their important roles during mouse embryonic development (Faust et al. 1995, O'Carroll et al. 2001, Suzuki et al. 2002, Voncken et al. 2003, Pasini et al. 2004, Boyer et al. 2006, Pasini et al. 2007). Furthermore, it has been demonstrated that the hallmark of PcG mediated silencing is the modification of histone tails, in particular methylation of lysine 27 on histone H3 (H3K27me3) and monoubiquitination of lysine 119 on histone H2A (H2AK119ub) (Cao et al. 2002, Muller et al. 2002, Wang et al. 2004, Papp and Muller 2006). The interdependence of these two chromatin marks evident as the H3K27me3 mark, set by the PRC2 complex, can be recognized by a second

complex, PRC1, which in turn deposits the H2AK119ub mark (Cao et al. 2002, Muller et al. 2002, Wang et al. 2004).

In addition to the regulation of developmental genes, PcG proteins have also been implicated in various cellular processes and diseases. First, PcG are key components of X-chromosome inactivation in mammals (Wang et al. 2001, Plath et al. 2003, de Napoles et al. 2004, Zhao et al. 2008). To achieve dosage compensation female cells, one X-chromosome is randomly chosen and inactivated in cells of the inner cell mass in early blastocysts. An initial step in this process is thought to be the recruitment of PcG proteins leading to the inactivation of the X-chromosome. This inactivation is irreversible during the lifetime of the cell and will be passed on to daughter cells during mitosis. Furthermore, PcG proteins have been implicated in the maintenance of pluripotency and cell-lineage specification (Boyer et al. 2006, Lee et al. 2006, Mohn et al. 2008). Nevertheless, ESCs lacking PcG proteins can be successfully generated, indicating that ESC self-renewal is not dependent on PcG proteins (Pasini et al. 2004, Chamberlain et al. 2008, Shen et al. 2008). However, *in vitro* differentiation of ESCs is severely affected by the loss of PcG proteins, emphasizing the key role of these proteins during lineage commitment (Chamberlain et al. 2008). In *D.melanogaster*, PcG proteins have been shown to target ncRNAs, which suggests a function in the regulation of microRNAs essential for development, apoptosis, and growth (Enderle et al. 2010). Additionally, aberrant expression and targeting of PcG proteins has been demonstrated to play key roles in the development and progression of a variety of different tumors by mediating the emergence and maintenance of cancer stem cells (Weikert et al. 2005, Bachmann et al. 2006, Collett et al. 2006, Suva et al. 2009, Mills 2010). In plants, H3K27me3 was shown to function in the regulation of vernalization by repressing *FLC,* a gene coding for a protein that represses flowering (Michaels and Amasino 1999, Sheldon et al. 1999, Schubert et al. 2006, Wood et al. 2006, De Lucia et al. 2008).

Despite the widespread recognition that PcG proteins are important for the repression of target genes, the exact mechanisms involved remain unclear. Transgenic experiments involving artificial tethering of PcG proteins proximal to a promoter driving a reporter gene suggest that recruitment of PcG complexes can induce repression of genes nearby (Sarma et al. 2008). However, whether this artificial tethering and subsequent recruitment reflects mechanisms that take place *in*

vivo remains elusive. In the following sections I will introduce in more detail the two main complexes formed by PcG proteins and their associated histone modifications.

1.4.1 Polycomb Repressive Complex 1 (PRC1)

The main components of the PRC1 complex in *D. melanogaster* are Polycomb (PC), Polyhomeotic (PH), Posterior sex combs (PSC), and Ring finger protein (RING), all of which are present in stoichiometric amounts (Di Croce and Helin 2013). In contrast, mammalian PRC1 is more diverse with several homologs for each component, resulting in different PRC1 variations (Figure 9) (Peterson et al. 2004, Li et al. 2010, Casanova et al. 2011, Gao et al. 2012, Hunkapiller et al. 2012). Canonical PRC1 consists of RING1A/B, homologs of RING, several chromodomain proteins (CBX) that are homologous to PC, one of six polycomb ring finger (PCGF) proteins that are similar to PSC, and three different polycomb-like (PCL) proteins, homologs of PH (Levine et al. 2002). The presence of CBX proteins enables the complex it to bind the H3K27me2/me3 chromatin marks via its chromodomain. It is thus believed that canonical PRC1 is recruited to target loci via prior deposition of the H3K27me3 mark by PRC2 (Cao et al. 2002, Min et al. 2003). Non-canonical PRC1 complexes lack CBX proteins and therefore cannot interact with the H3K27me3 mark and may instead bind genomic loci through mechanisms independent of the H3K27me3 mark (Farcas et al. 2012, Tavares et al. 2012, Wu et al. 2013). Shared among all PRC1 variants is the RING-domain containing protein, RING1B or RING1A, which acts as an ubiquitintransferase with H2AK119 as a substrate (Wang et al. 2004, Cao et al. 2005). It has been demonstrated that this histone mark is required for efficient silencing of target genes, but the precise mechanisms involved have yet to be determined (Wang et al. 2004, Cao et al. 2005). Furthermore, it has been proposed to inhibit RNA Pol II activity, possibly by blocking the phosphorylation of serine 2 at its C-terminal domain (Stock et al. 2007). Additionally, H2AK119ub prevents the eviction of the H2A-H2B dimers, a process that takes place during transcriptional elongation (Zhou et al. 2008). Another mechanism involves compaction of chromatin (Eskeland et al. 2010, Grau et al. 2011). This compaction reduces accessibility of the underlying chromatin to both transcription factors and chromatin remodelers such as SWI/SNF (Bantignies and Cavalli 2011). Furthermore, PRC1 has been demonstrated to directly interact with

the transcriptional machinery and inhibit transcriptional elongation (King et al. 2002, Zhou et al. 2008, Lehmann et al. 2012).

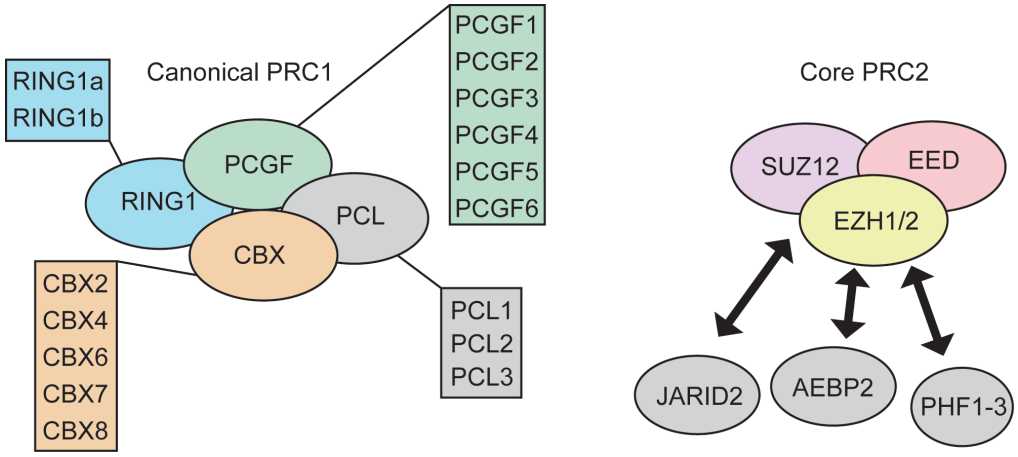

Figure 9 PRC1 and PRC2 complexes. Depicted are the two main complexes formed by mammalian PcG proteins. In the case of PRC1, the different subunits contain several optional factors that can be incorporated in order to form distinct subcomplexes. The RING1 subunit is present in all PRC1 subcomplexes and contains ubiquityltransferase activity towards histone H2AK119. The three core subunits of PRC2 have been shown to interact with several interaction partners that have been suggested to be involved in guiding the complex to its target loci. Adapted from (Di Croce and Helin 2013)

1.4.2 Polycomb Repressive Complex 2 (PRC2)

Mammalian PRC2 contains three core components: enhancer of zeste (EZH2), embryonic ectoderm development (EED), and suppressor of zeste 12 (SUZ12). The catalytic subunit of the complex, EZH2 bears a SET domain that catalyzes mono-, di-, and trimethylation of H3K27 (Cao et al. 2002, Muller et al. 2002, Schuettengruber et al. 2007). However, EZH2 on its own is inactive and must form a complex with EED and SUZ12 to gain catalytic activity (Cao and Zhang 2004, Pasini et al. 2004, Ketel et al. 2005). Accordingly, EED knock-out cells show complete loss of the H3K27me3 mark. Surprisingly, depletion of EZH2 results in dramatic loss of the mark, but not complete absence thereof. This is a result of its homolog, EZH1, which is partially redundant in function (Margueron et al. 2008). Another mechanism that regulates methyltransferase activity of EZH2 is the phosphorylation outside its active site (Chen et al. 2010, Kaneko et al. 2010, Wei et al. 2011). In addition to these three core components of the methyltransferase complex PRC2, there are several accessory proteins that regulate its methyltransferase activity as well as its recruitment to genomic sites (discussed below).

Despite the widespread recognition that H3K27me3 is a hallmark of gene repression, the mechanisms underlying this process are still elusive. As mentioned above, one function of the H3K27me3 mark could be that it acts as a docking site for PRC1. The chromodomain of the CBX component recognizes H3K27me3 and thus recruits the other components of PRC1 to sites marked by H3K27me3 in order to facilitate monoubiquitination of H2AK119. Additionally, the EED component of PRC2 also interacts with H3K27me3 via its aromatic cage structure and is thus thought to help propagate the H3K27me3 mark in order to maintain repressive chromatin domains and to transmit the histone mark from the mother to the daughter cells (Margueron et al. 2009, Xu et al. 2010). Furthermore, a recent study has emphasized the importance of the H3K27me3 mark in *D. melanogaster* by expressing a mutant form of histone H3 that cannot be methylated at lysine 27. Remarkably, replacement of endogenous histones with these mutant ones mimics the phenotypes observed in PcG mutants (Pengelly et al. 2013). Whether the observed phenotype is due to the absence of PRC1 as a result of the missing H3K27me3 mark remains to be determined.

In summary, PcG complexes comprise two main complexes that facilitate H3K27 trimethylation and H2AK119 monoubiquitination, respectively. These epigenetic marks play key roles in the maintenance of gene expression patterns and their absence leads to aberrant gene expression during development. However, despite their important role in gene regulation it is still controversial how PcG proteins find their way to their target sites. Interestingly, core PcG proteins do not have the ability to directly interact with DNA, suggesting the need for cofactors that recruit the complexes to their target loci. In the following sections I will introduce different models that have been proposed for PcG recruitment.

1.4.3 Genomic targeting of PcG proteins

1.4.3.1 Targeting PcG proteins in *Drosophila Melanogaster*

PcG proteins were initially discovered in *D. melanogaster* as repressors of *Hox* genes and as a result most of the work regarding PcG recruitment has emerged from studies of the *Hox* clusters in flies. Efforts to study cis-regulatory regions that control the segment specific *Hox* gene expression patterns during early development have revealed the existence of two groups of DNA regulatory elements; the initiator

elements and the maintenance elements. The initiator elements are DNA sequences bound by TFs that set the expression pattern of the target genes in the very early stages of development. As mentioned earlier, these TFs decay quickly while the expression patterns they established are maintained throughout development. This is largely dependent on the activity of the maintenance elements, termed Polycomb/Trithorax response elements (PREs) (Busturia et al. 1989, Simon et al. 1990, Simon et al. 1993, Chan et al. 1994, Chiang et al. 1995, Cavalli and Paro 1998). PREs are short DNA elements located several kb away from the transcription start site of the target gene [reviewed in (Ringrose and Paro 2004)]. In addition to being bound by PcG proteins, they can also be occupied by trithorax group (TrxG) proteins, which act antagonistically to the PcG proteins by maintaining an active state of the target gene via trimethylation of H3K4 (H3K4me3) (Schuettengruber et al. 2007). In that sense, PREs in *D. melanogaster* have dual potential for epigenetic maintenance of specific expression patterns. Furthermore, this maintenance is reversible in the sense that PREs can switch from an active to an inactive state and vice versa and therefore maintain the balance between gene activity and repression (Cavalli and Paro 1998, Cavalli and Paro 1999). This switch from a PcG repressed to a TrxG activated state can be induced via transcription through the PRE (Cavalli and Paro 1998, Schmitt et al. 2005).

The fact that PcG complexes bind PREs in a tissue-specific manner, but are ubiquitously expressed in all cells suggests that sequence specific DNA binding factors are involved in the recruitment of PcG proteins. However, since core PcG proteins do not harbor DNA binding capabilities themselves, there must be other factors that guide PcG complexes to their target loci. In order to identify specific sequence determinants that are common between PcG bound sites, several studies undertook efforts to identify novel PREs. Initially, cytological studies on polytene chromosomes that analyzed the co-localization of PcG proteins estimated the number of PREs in the fly genome to several hundred (Zink and Paro 1989, DeCamillis et al. 1992, Chinwalla et al. 1995). Biochemical and transgenic experiments revealed that all PREs share common characteristics, such as their ability to maintain the transcriptional state of a reporter gene when taken out of their endogenous context (Chan et al. 1994, Christen and Bienz 1994, Cavalli and Paro 1998, Sengupta et al. 2004). However, it was not until 2003 when Ringrose et al. developed a computational algorithm that could identify similarities between PREs,

based on the co-occurrence of transcription factor binding sites (Ringrose et al. 2003). In particular, the authors found binding sites for the sequence-specific DNA-binding factors GAG, ZESTE, PSQ, and PHO to be enriched in PREs. Interestingly, the binding sites do not occur on their own, but are present in clusters of pairs within the PREs. Of these proteins, PHO was already known to be involved in PcG-mediated silencing (Simon et al. 1992, Brown et al. 1998, Brown et al. 2003), but the mechanisms involved were still elusive. Similar observations have been made for the other two factors, ZESTE and GAF, which had previously been demonstrated to have activating and repressing functions (Hagstrom et al. 1997, Strutt et al. 1997, Decoville et al. 2001, Huang et al. 2002, Mulholland et al. 2003, Bejarano and Busturia 2004). In total, this algorithm predicted 167 candidate PREs, some of which were experimentally validated in transgenic assays testing their potential to repress a reporter in a PcG dependent manner (Ringrose et al. 2003). These studies lead to a model whereby PREs are defined by specific combinations of transcription factor binding sites.

A few years later, the emergence of microarray analysis and high-throughput sequencing enabled genome-wide occupancy studies with much higher resolution. Coupled with antibody-specific immunoprecipitation of formaldehyde crosslinked chromatin, analyses allowed for thorough evaluation of the computationally predicted model. In fact, three studies independently sought to map genome-wide binding profiles of PcG proteins (Negre et al. 2006, Schwartz et al. 2006, Tolhuis et al. 2006). Strikingly, there was only limited overlap between PRE prediction by Ringrose *et al.* and the *in vivo* binding data, with 73% - 94% of PcG binding sites lacking a corresponding predicted PRE. These observations suggest that there are PREs that are regulated by factors other than PHO, ZESTE and GAF. Furthermore, genome-wide occupancy data revealed that PHO associates not only with PcG targets but is also present at genes marked by active histone modifications such as H3K4me3 (Kwong et al. 2008, Oktaba et al. 2008, Schuettengruber et al. 2009). Moreover, Dejardin et al. constructed a synthetic PRE by inserting these binding sites into a bacterial backbone and showed that presence of binding sites for these three factors is not sufficient to create a PRE (Dejardin et al. 2005). The authors were, however, able to create a functional PRE by adding a motif for the homeotic gene regulator Dorsal switch protein 1 (DSP1). This construct successfully recruited PcG proteins and resulted in silencing of a downstream reporter (Dejardin et al. 2005).

In summary, PcG protein recruitment in *D. melanogaster* appears to be mediated by specific DNA sequences termed PREs and guided there by combinatorial binding of sequence specific TFs. However, many of these TFs are also associated with positive regulation of gene expression and it thus remains unclear whether their contribution to PcG recruitment is direct or indirect.

1.4.3.2 Targeting mammalian PcG proteins

After the discovery of PcG proteins in *D. melanogaster* and the identification of PREs as recruiter sequences, substantial efforts were made in searching for mammalian PREs. Despite these efforts, to date only two PREs have been discovered in vertebrates that fulfill the definition of a PRE (Sing et al. 2009, Woo et al. 2010). Instead, other mechanisms have been proposed to play a role in mammalian PcG recruitment, some of which will be discussed in more detail in the following sections.

Targeting non-canonical PRC1

Mammalian PcG proteins localize mainly to promoters overlapping with CGIs, leading to the hypothesis that CpG dinucleotides might be involved in recruiting of PcG complexes (Ku et al. 2008, Orlando et al. 2012). Indeed, two recent studies have linked the CxxC zinc-finger domain containing protein KDM2B to the recruitment of non-canonical PRC1 to CpG islands (Farcas et al. 2012, Wu et al. 2013). Wu et al. showed that KDM2B interacts with RING1B to form a non-canonical PRC1 that is required for H2AK119 ubiquitination of specific targets in mouse ESCs (Wu et al. 2013). Furthermore, loss of KDM2B results in reduced occupancy of PRC1 and deposition of H2AK119ub at CGIs (Wu et al. 2013). However, KDM2B occupies nearly all CGIs throughout the genome while PRC1 is present only at a subset of these, suggesting involvement of additional factors.

Some of these factors may include REST, RUNX1, and YY1 (Woo et al. 2010, Ren and Kerppola 2011, Dietrich et al. 2012, Yu et al. 2012, Woo et al. 2013). However, as depletion of these factors only affects a small set of target loci, they are unlikely to play a general role in recruitment. For example, YY1, the mammalian homologue of the *Drosophila* PHO, was suggested to regulate recruitment of PcG proteins to a specific locus within the HoxD cluster (Woo et al. 2010). Furthermore,

when placed upstream of a reporter gene, the element was able to repress its expression in a PcG dependent manner, a key characteristic of PREs in *D. melanogaster* (Woo et al. 2010). YY1 might therefore be a factor involved in targeting non-canonical PRC1 complexes that lack the CBX component. Nevertheless, the fact that YY1 does not correlate with PcG binding on a genome-wide level (Vella et al. 2012) makes it an unlikely candidate for a general recruitment factor.

Targeting mammalian PRC2

Similar to non-canonical PRC1, mammalian PRC2 has been suggested to be recruited by mechanisms involving factors that specifically recognize CGIs (Mendenhall et al. 2010). However, in contrast to PRC1, no CxxC-domain containing protein has been directly implicated in recruiting PRC2 to CGIs. Two proteins, JARID2 and AEBP2, have been co-purified with core PRC2 components and shown to be able to bind CG-rich DNA (Peng et al. 2009, Shen et al. 2009, Landeira et al. 2010, Li et al. 2010, Pasini et al. 2010). Furthermore, genome-wide occupancy studies have revealed a strong overlap between JARID2 and PRC2 core components, suggesting a role in regulating PRC2 targeting and / or activity (Peng et al. 2009, Landeira et al. 2010, Li et al. 2010, Pasini et al. 2010). JARID2 is a member of the Jumonji family of histone demethylases but lacks key catalytic residues required for demethylase activity (Li et al. 2010). It does, however, bear the DNA-binding domains, supporting its proposed role in targeting the PRC2 complex. Furthermore, knock-down of JARID2 in ESCs results in reduced levels of H3K27me3 at many target sites. However, the effect on H3K27me3 levels is relatively mild and most target genes remain silent. Thus, JARID2 might be involved in fine-tuning PRC2 methyltransferase activity, rather than recruiting the complex to DNA (Panning 2010). Whether this influence on catalytic activity of EZH2 is stimulatory or inhibitory is still controversial (Herz and Shilatifard 2010).

Another mechanism proposed to recruit PRC2 involves sequence-specific binding of TFs similar to what has been observed in *D. melanogaster*. This model is appealing as it would explain cell type-specific targeting of PRC2. As mentioned above, one such candidate factor is YY1, but poor correlation with PRC2 occupancy throughout the genome and mainly localized to active regions makes it unlikely to be a general recruitment factor (Vella et al. 2012). Two other candidates are SNAIL

(Herranz et al. 2008, Arnold et al. 2013) and REST (Dietrich et al. 2012, Arnold et al. 2013). REST has been demonstrated to play a role for in targeting and regulating H3K27 trimethylation (Dietrich et al. 2012, Arnold et al. 2013). In particular, Arnold et al. showed the importance of REST for efficient PRC2 targeting in neural progenitor cells. In contrast, depletion of REST in ESCs only mildly affected H3K27me3 levels. Similarly, SNAIL has been associated with only a small subset of targets and is therefore unlikely to play a general role in recruitment (Herranz et al. 2008, Arnold et al. 2013).

Another set of factors, which were shown to play a role in guiding PRC2 to its target sites are the Polycomb-like (PCL) factors. They were demonstrated to not only stimulate enzymatic activity, but also aid in recruiting PRC2 to specific sites in ESCs (Walker et al. 2010, Casanova et al. 2011, Hunkapiller et al. 2012). There are three main PCL proteins in mammals, each having distinct effects on PRC2 activity. For example, PCL1 and PCL3 have been shown to be important for genome-wide deposition of H3K27me3 and ESC self-renewal, while PCL2 was suggested to play a role in guiding PRC2 to its targets via its PHD finger (Sarma et al. 2008, Walker et al. 2010, Casanova et al. 2011). Furthermore, all three PCLs have recently been shown to interact with H3K36me3 via their Tudor domain (Ballare et al. 2012, Brien et al. 2012, Musselman et al. 2012). This interaction was suggested to be important to target PRC2 to previously active genes in order to repress them (Abed and Jones 2012).

Finally, interactions between noncoding RNAs (ncRNAs) and PcG components have been proposed to guide PcG complexes to their targets. In particular, several studies have suggested a role for these RNAs in the recruitment of PRC2. The basis for this model stems from studies on X-chromosome inactivation in females, which is initiated via transcription of a long non-coding RNA (lncRNA) called Xist (Brockdorff et al. 1992, Brown et al. 1992, Lee et al. 1999). Upon transcription of this ncRNA, it coats the entire X-chromosome in cis and is believed to initiate the formation of heterochromatin by recruiting PRC2 via a specific sequence within the transcript (Plath et al. 2003, Silva et al. 2003, Zhao et al. 2008). Another ncRNAs implicated in gene repression via PRC2 recruitment in humans is the lncRNA HOTAIR, which is transcribed from the *HOXD* cluster and has been proposed to act in trans to silence genes within the *HOXC* cluster (Rinn et al. 2007). Third, *Kcnqot1* seems to be involved in imprinting the *Kcnq1* cluster on mouse

chromosome 7 in a PRC2 dependent manner (Fitzpatrick et al. 2002, Pandey et al. 2008). In addition to lncRNAs, short ncRNAs have been proposed to be involved in PRC2 recruitment (Kanhere et al. 2010, Davidovich et al. 2013, Kaneko et al. 2013). However, despite the recent attention given to recruitment mechanisms involving ncRNAs, evidence for a direct interaction between ncRNAs and Polycomb proteins remains scarce and the observed interactions may be of indirect nature facilitated by other factors.

Taken together, several models for recruiting PcG complexes to their target loci have been proposed in recent years, all supported by experimental evidence (Figure 10). It thus remains a controversially debated question and it is likely that proper targeting is the result of a complex interplay between different mechanisms.

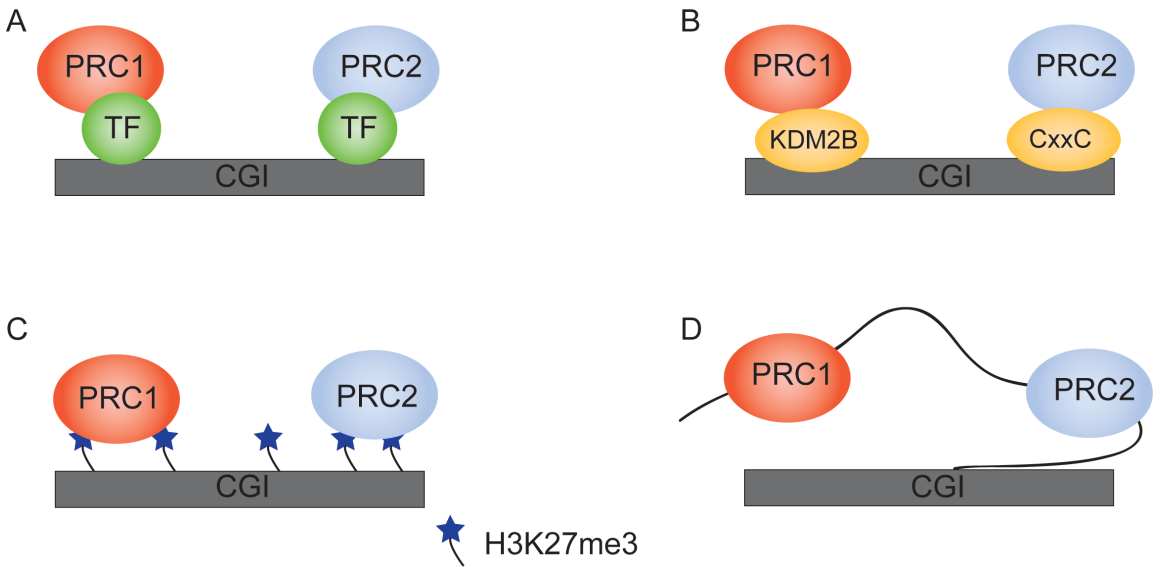

Figure 10 Proposed mechanisms that guide PcG complexes to their target loci. (A) Sequence specific TFs interact with PcG complexes and guide them to their targets. (B) The CxxC domain-containing protein KDM2B has been shown to be involved in the recruitment of non-canonical PRC1. An analogous factor has yet to be identified for PRC2. (C) Canonical PRC1 is recruited to H3K27me3-marked regions via its CBX component that contains a chromodomain. Similarly, the PRC2 component EED can bind to the H3K27me3 mark, a mechanism believed to be responsible for spreading the mark. (D) Non-coding RNAs have been shown to play a role in guiding PRC1 and PRC2 to their targets. Adapted from (Lanzuolo and Orlando 2012).

CHAPTER 2

2. SCOPE OF THIS THESIS

During development of an organism, pluripotent stem cells differentiate into specialized cell types carrying out specific functions. In order to achieve this, cell type-specific gene expression patterns are established upon lineage commitment and stably maintained during subsequent cell divisions. In particular, during differentiation of pluripotent stem cells, pluripotency genes get repressed and lineage-specific genes are activated. It is thought that histone modifications and DNA methylation play a key role in the maintenance of these patterns. Recent advances in the genome-wide analysis of these epigenetic marks have revealed distinct distributions throughout the genome, effectively distinguishing active and accessible euchromatin from inactive and closed heterochromatin. However, it remains unclear how specific modifications are targeted to distinct genomic loci in a temporally and spatially defined manner. It is essential to understand these mechanisms in order to

One epigenetic process that has been studied extensively in regard to genomic targeting is PcG-mediated gene repression. Stable repression by PcG proteins is thought to be mediated by deposition of the H3K27me3 mark. This mark is highly dynamic during the course of differentiation, emphasizing its relevance during development (Mohn et al. 2008). Work in both *D. melanogaster* and mammals suggests that TFs play a key role in the recruitment of PcG complexes to their target loci (Ringrose and Paro 2007, Herranz et al. 2008, Woo et al. 2010, Dietrich et al. 2012, Arnold et al. 2013). However, it is still unclear whether the observed functional interactions between TFs and PcG complexes are mediated via a direct physical interaction.

In addition to TFs, CpG-rich DNA sequences have been suggested to play a role in recruiting PcG complexes. This is made evident by the fact that the majority of H3K27me3 peaks are associated with CGIs (Ku et al. 2008, Orlando et al. 2012). Nevertheless, the direct role of CpG dinucleotides remains elusive.

Here, we aim to identify DNA sequences that can autonomously recruit the Polycomb machinery. Furthermore, we will dissect these elements in order to identify

specific sequence determinants that are required for efficient recruitment of PcG complexes and establishment of H3K27me3 domain. To this goal, we will employ a system that allows genomic targeting of murine embryonic stem cells at a specific locus, using CRE recombinase-mediated cassette exchange (Feng et al. 1999, Lienert et al. 2011). This reductionist approach will allow us to test the effect of minimal DNA sequence variations on the recruitment of PcG complexes and subsequent acquisition of H3K27me3.

CHAPTER 3

3. RESULTS

3.1 Summary

Multicellular organisms consist of various cell types, each serving a specialized function. Remarkably, all of these different cell types originate from one cell, the oocyte, which has the potential to give rise to every cell type within the organism. Therefore, almost all cells within an organism contain the same genetic information. Nevertheless, each cell type interprets this information differently, resulting in cell type-specific gene expression patterns. These expression patterns define a cell's function and are acquired upon lineage commitment of a pluripotent cell. Furthermore, once acquired, these patterns can be stably maintained throughout subsequent cell divisions via mechanisms involving the modification of chromatin structure by specialized enzymes.

A key group of such chromatin modifiers are the Polycomb-group (PcG) proteins. These proteins play key roles during early development by stably repressing developmental regulators. They function in part through methylation of histone H3 at lysine 27 (H3K27me3), an epigenetic mark associated with gene repression (Cao et al. 2002, Muller et al. 2002, Papp and Muller 2006). However, despite the importance of this epigenetic mark, it remains elusive how PcG complexes are guided to their target loci in a cell type-specific manner. Studies in *Drosophila melanogaster* and mammals suggest that TFs play a key role in the recruitment of PcG complexes (Ringrose and Paro 2007, Woo et al. 2010, Dietrich et al. 2012, Arnold et al. 2013). However, it is unclear whether this mechanism involves direct interactions between TFs and PcG proteins.

Here, we hypothesized that recruitment of PcG complexes to specific loci is encoded within the target DNA sequence. This could be in the form of TF binding sites or other sequence queues. To test this, we employed a reductionist approach to identify sequence determinants that guide PRC2 recruitment. In particular, we inserted a set of endogenous H3K27me3 targets in mouse ES cells into a defined ectopic locus. Then, we examined whether these ectopically inserted DNA

sequences could recapitulate endogenous H3K27me3 levels. Indeed, the tested elements were able to recruit PRC2 and acquire the H3K27me3 mark at levels comparable to the endogenous locus. Further dissection of these elements revealed that DNA sequences rich in CpG dinucleotides and as short as 220 bp are sufficient to establish an H3K27me3 domain. Furthermore, we found that cell type-specific recruitment is determined by the transcriptional state of the target locus. In particular, transcriptional activity regulated by TF binding to a proximal cis-regulatory element can efficiently block the acquisition of H3K27me3. Finally, using mutant DNA sequences we demonstrate that absence of DNA methylation at the target locus is necessary to efficiently recruit PRC2.

Taken together, we propose a model whereby PRC2 is recruited to transcriptionally inactive, unmethylated CpG-rich DNA. We observe that TFs play a key role in this process, albeit not via direct interaction with PcG components, but rather via regulation of transcriptional activity. We thus provide new insights into the role of epigenetic modifications in regulating gene expression patterns.

3.2 Submitted Manuscript

Proficient recruitment of H3K27 trimethylation by short sequences is counteracted by DNA methylation and enhancer activity

Philip Jermann[1,2], Leslie Hoerner[1], Lukas Burger[1,3] and Dirk Schübeler[1,2,4]

Working title: Sequence determinants of Polycomb recruitment

[1] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

[2] Faculty of Science, University of Basel, Petersplatz 1, CH 4053 Basel, Switzerland

[3] Swiss Institute of Bioinformatics, Biozentrum - University of Basel, 4056 Basel, Switzerland

[4] corresponding author. Contact Information**:** dirk@fmi.ch

Keywords: Chromatin, Polycomb, DNA methylation, stem cells

BIOLOGICAL SCIENCES: Genetics

3.2.1 Abstract

Trimethylation of histone H3 at lysine 27 (H3K27me3) is a chromatin mark associated with Polycomb mediated gene repression. Despite its critical role in development, it remains largely unclear how this mark is targeted to defined loci in mammalian cells. Here, we employ iterative genome editing in order to identify small DNA sequences capable of autonomously recruiting Polycomb. We inserted 28 DNA elements at a defined chromosomal position in mouse embryonic stem (ES) cells and assessed their ability to promote H3K27me3 deposition. Combined with deletion analysis we identified DNA elements as short as 220 nucleotides that correctly recapitulate endogenous H3K27me3 patterns. Functional Polycomb recruiter sequences are invariably CpG rich but require protection against DNA methylation. Alternatively, their activity can be blocked by placement of an active promoter-enhancer pair *in cis*. Taken together, these data support a model whereby PRC2 recruitment at specific targets in mammals is positively regulated by local CpG density yet obstructed by transcriptional activity or DNA methylation.

3.2.2 Significance statement

Polycomb repressive complex 2 functions in gene repression and acts by methylating histone H3 at lysine 27 (H3K27me3). Despite its relevance it remains elusive how this complex is recruited to its target sites in the genome. Here, we used repeated genomic targeting in embryonic stem cells to identify DNA sequence determinants that autonomously confer H3K27me3 recruitment. We show that surprisingly small CG-rich DNA sequences are sufficient to recruit H3K27me3, but only if they are devoid of DNA methylation and transcriptional activity. This study provides new insights into the mechanisms recruiting H3K27me3 and the crosstalk between diverse chromatin modifications.

3.2.3 Introduction

Polycomb Group (PcG) proteins are required for proper fine-tuning of gene expression and act in part through modifications of histones (Boyer et al. 2006, Pereira et al. 2010, Pengelly et al. 2013). Initially identified in *D. melanogaster* as regulators of homeobox gene expression and body segmentation, PcG proteins have since been described in mammals and implicated in many cellular processes, including maintenance of pluripotency and lineage-commitment (Lewis 1978, Di Croce and Helin 2013). PcG proteins form two main complexes, PRC1 and PRC2 (Levine et al. 2004). PRC1 monoubiquitylates histone H2A lysine 119 (H2AK119), while PRC2 harbors histone methyltransferase (HMT) activity towards histone H3 lysine 27 (H3K27) (Kuzmichev et al. 2002, de Napoles et al. 2004). Both chromatin marks are associated with gene repression and are essential for embryonic stem (ES) cell differentiation (Cao et al. 2002, Boyer et al. 2006). Mammalian PRC2 consists of 3 core proteins essential for its catalytic activity; Enhancer of zeste homologue 2 (Ezh2), Suppressor of zeste homologue (Suz12) and Embryonic ectoderm development (Eed). Genome-wide analysis of the H3K27me3 mark in different cell types and during *in vitro* differentiation of ES cells revealed the mark to be dynamic during lineage-commitment, suggesting that the complex is recruited to target sites in a cell-type specific manner (Bracken et al. 2006, Mohn et al. 2008).

The mechanisms by which PcG proteins are recruited to specific genomic sites are still elusive. In *D. melanogaster*, transgenic experiments showed that polycomb can be recruited to polycomb responsive elements (PREs) leading to repression of reporter genes (Chan et al. 1994, Cavalli and Paro 1998, Ringrose et al. 2003). Several TF binding motifs are enriched within fly PREs and might contribute to Polycomb recruitment (Ringrose et al. 2003, Muller and Kassis 2006). In mammals, PRC2 and the H3K27me3 mark localize mainly to transcriptionally inactive regions rich in CpG dinucleotides, referred to as CpG-islands (CGIs) (Tanay et al. 2007, Ku et al. 2008, Mendenhall et al. 2010, Lynch et al. 2012, Orlando et al. 2012). Since CGIs are regulatory regions they are under high selection for presence of regulatory motifs more complex than CG, which are occupied by transcription factors (TFs) (Cohen et al. 2011). It thus remains open whether high CpG dinucleotide frequency is indeed sufficient to direct PRC2 recruitment.

Here we aimed to identify DNA elements that can autonomously establish an H3K27me3 domain and dissect their function through iterative testing of sequence variants. To control for position effect and copy number, we employed recombination-mediated cassette exchange (RMCE) to insert DNA elements into the same genetic locus. These repeated insertions identify DNA fragments as short as 220 nucleotides (nt) capable of autonomously acquiring H3K27me3 in a heterologous genomic context. This ability is determined by the presence of unmethylated CpG dinucleotides, yet the cell-type specific acquisition of H3K27me3 during differentiation is dependent on the presence of a cell-type specific enhancer.

3.2.4 Results

Small DNA elements autonomously recruit H3K27me3 when placed at an ectopic locus in ES cells

Candidate recruiter sequences for PRC2 were chosen based on their enrichment for H3K27me3 and SUZ12 in genome-wide chromatin immunoprecipitation (ChIP) datasets from murine ES cells (Pasini et al. 2010, Tiwari et al. 2012). Since the majority of SUZ12 peaks lie within CGIs (Tanay et al. 2007, Ku et al. 2008) we asked if candidate CGIs could act autonomously to promote H3K27me3 when inserted in a heterologous genomic environment. We chose eight candidate CGIs ranging in size from 550nt to 1000nt and included elements that are putative promoters as well as elements located within exons and intergenic regions (Figure 11A-E, Figure 16C-E). In addition, we inserted a ninth CGI overlapping with the promoter of the exosome complex component Exosc9. This element does not harbor H3K27me3 at its endogenous site and served as a negative control (Figure 11F).

All nine elements were inserted separately into a previously described target site in the beta-globin locus using CRE-mediated recombination (Lienert et al. 2011) (Figure 16A). Importantly, the globin locus and several hundred kilobases of surrounding region contain no preexisting H3K27 methylation in stem cells (Figure 16B), making this site suitable to test for the autonomous ability of DNA elements to recruit this chromatin mark. Moreover the RMCE approach relies only on negative selection and the pre-existing marker gene is removed during the insertion, leaving behind only the sequence of interest flanked by loxP sites. Following verification of correct insertion, we tested for the presence of H3K27me3 and the core PRC2 component Suz12 at all elements by ChIP. For each of the eight putative recruiter elements, we observed the acquisition of H3K27me3 after insertion at the globin locus (Figure 11A-E). In contrast, no enrichment for either Suz12 or H3K27me3 was observed at the control sequence, excluding that the insertion of any heterologous genomic region leads to PRC2 recruitment and acquisition of H3K27me3 (Figure 11F). Importantly, the H3K27me3 levels detected at the eight functional elements are comparable to those observed at their endogenous loci, consistent with full sequence autonomy in recruitment of this chromatin mark. This is particularly intriguing in light of the relatively small sizes of tested sequences that range from

542 to 901 base pairs and demonstrates that elements of less than one kilobase can establish a local H3K27me3 domain. We also detected the PRC2 component Suz12 at all inserted candidate elements (Figure 11A-E, Figure 16C-E). In five of eight elements at levels similar to the endogenous site, while at three elements levels were reduced even though H3K27me3 levels were similar to the endogenous site.

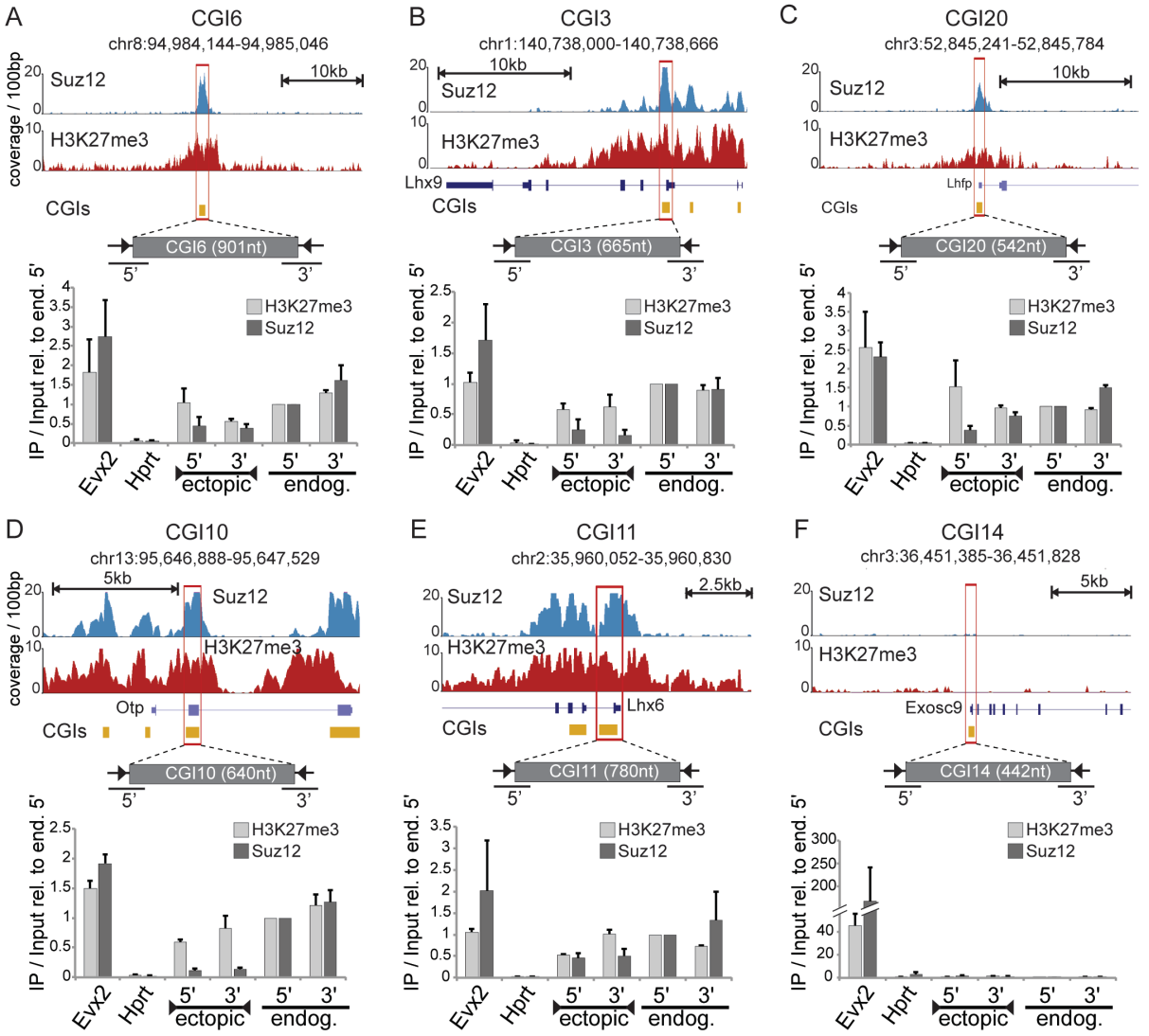

Figure 11 Short DNA sequences can establish H3K27me3 domains. (A)-(F) Top panels show Suz12 and H3K27me3 profiles around the candidate (A)-(E) and control (F) CGIs. Genomic coordinates indicate the location of the candidate CGI. Bottom panels show H3K27me3 and Suz12 ChIP-qPCR data at the ectopically inserted and endogenous loci.; All qPCR data are normalized to input DNA and shown relative to the 5' amplicon of the endogenous locus in order to better compare the levels between the ectopic and endogenous sites. Evx2 and Hprt were used as endogenous positive and negative controls, respectively. Error bars represent standard deviation between at least two biological replicates.

The fact that autonomous PRC2 recruitment and subsequent establishment of an H3K27me3 domain is observed for all cases tested suggests that such local sequence autonomy may be a general feature of polycomb recruitment in the mouse

genome. Next we asked if further deletions could identify even smaller sequence elements as functional recruiters.

Small functional polycomb recruiter elements cover less than two nucleosomes

To dissect length requirements and potentially identify regions essential for recruitment, we tested subfragments of the identified recruiter sequences ranging from 218 to 445 nt. The resulting eight fragments were separately inserted and Suz12 and H3K27me3 enrichments were measured (Figure 12, Figure 17). Enrichments varied between elements indicating that these smaller elements do not show a uniform ability to recruit a functional polycomb complex. For example, the 3' half of element 3 shows no recruitment of H3K27me3, while the 5' half shows strong recruitment, despite the fact that both are comparable in size (Figure 12C, Figure 17C). Importantly these insertion experiments show that elements as short as 218nt can be sufficient to create a local H3K27me3 domain (Figure 12D, Figure 17D). Thus DNA regions spanning less than two nucleosomes can be sufficient to recruit PRC2, leading to H3K27 methylation levels similar to their endogenous site.

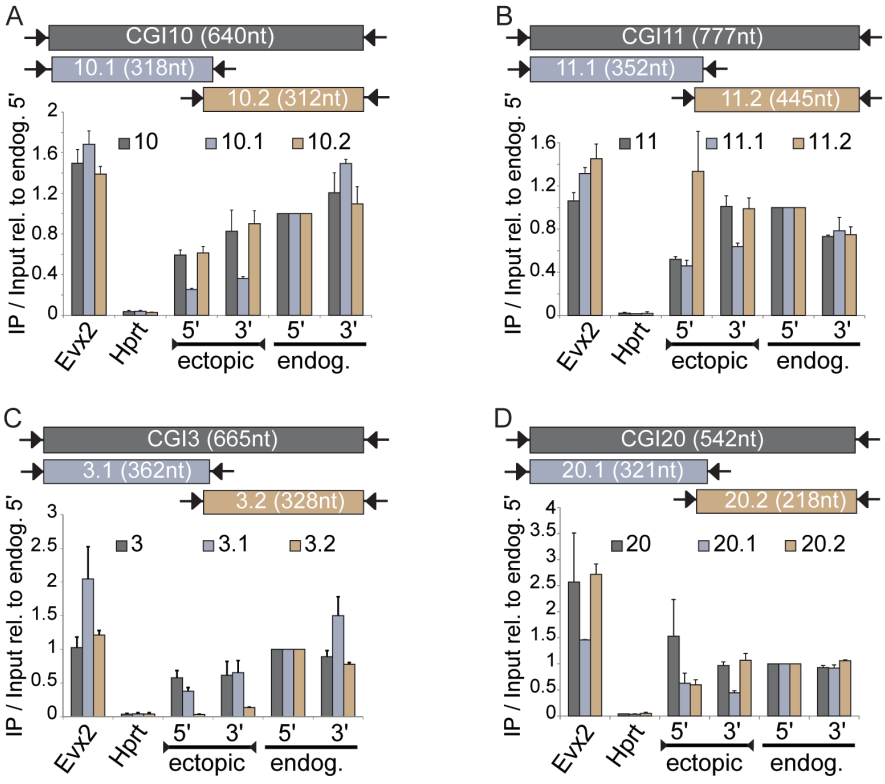

Figure 12 Dissection of H3K27me3 recruiters identifies 220 nt recruiter sequence. (A)-(D) Recruiter elements were dissected into smaller parts and analyzed for H3K27me3 enrichments by ChIP-qPCR. All qPCR

data are normalized to input DNA and shown relative to the 5' amplicon of the endogenous locus in order to better compare the levels between the ectopic and endogenous sites. Evx2 and Hprt were used as endogenous positive and negative controls, respectively. Error bars represent standard deviation between at least two biological replicates.

Cell-type specific reconstitution of H3K27me3 patterns requires endogenous context

The above experiments argue that CGIs that recruit H3K27me3 in stem cells function similarly when placed in an ectopic site. Since polycomb recruitment is variable between cell types we next asked whether sequences that do not harbor H3K27me3 in stem cells but gain it during differentiation recapitulate this behavior when placed at an ectopic site. As a test sequence we chose a 1.1kb CGI overlapping with the *Utf1* gene-body since it harbors no detectable H3K27me3 in stem cells but gains the mark during neuronal differentiation (Figure 13A). When inserted in the beta globin locus, however, this CGI recruits the H3K27me3 mark already in stem cells even though it is absent at the endogenous locus (Figure 13B). Thus, outside of the native genomic locus, this sequence acquires H3K27me3 in a cell-type independent manner. The *Utf1* locus codes for an established pluripotency factor (Kooistra et al. 2010) and is exclusively expressed in ESCs (Tippmann et al. 2012). Upon differentiation into neuronal progenitors, *Utf1* expression is shut off accompanied by the gain of H3K27me3 (Figure 13A). Expression of the *Utf1* gene is regulated by a proximal downstream enhancer positioned adjacent to the CGI (Chew et al. 2005). To determine whether the transcriptional context is required to regulate cell-type specific H3K27me3 recruitment, we inserted the entire *Utf1* locus including the promoter and enhancer. Addition of these *cis*-regulatory elements resulted in loss of H3K27me3 acquisition in stem cells, mimicking the endogenous locus (Figure 13C). Furthermore, upon *in vitro* differentiation of ESCs into NPs, we observed acquisition of H3K27me3 at the ectopic locus similar to the endogenous site, indicating that the dynamic pattern of this chromatin mark is now reconstituted at the ectopic site in both ESCs and NPs (Figure 13C). One possible explanation is that inclusion of the promoter and enhancer leads to transcriptional activation, which in turn counteracts H3K27me3 recruitment in stem cells. To test this, we sought to reduce the transcriptional activity of the promoter-enhancer pair with minimal changes in its sequence. The *Utf1* enhancer contains an OCT4/SOX2 binding site (Chew et al. 2005), both key transcription factors of the pluripotency network active

in stem cells. Genome-wide profiles of SOX2 in ES cells and NPs confirmed that the factor is bound to the Utf1 enhancer in ES cells, but not in NPs (Lodato et al. 2013). We thus inserted a construct with a 15 nt deletion within the enhancer, effectively removing the OCT4/SOX2 binding site while leaving the enhancer otherwise unchanged. Expression from this construct was analyzed by replacing the *Utf1* gene with a luciferase reporter as transcripts from the endogenous *Utf1* locus and ectopic site are otherwise indistinguishable (Figure 18A). After insertion we compared expression of both constructs revealing that deletion of the OCT4/SOX2 binding site abolishes transcription, leading to a 100-200 fold decrease in luciferase activity. To determine if this affects H3K27me3 acquisition we next inserted the enhancer mutation in combination with the *Utf1* gene. As expected, SOX2 binding was strongly reduced at the mutated enhancer when compared to the wild-type construct (Figure 18C). More importantly, however, and in contrast to the unaltered enhancer construct, this mutant now showed premature acquisition of H3K27me3 in stem cells (Figure 13D). Thus the presence of the *Utf1* promoter/enhancer pair is necessary to inhibit H3K27me3 of the overlapping CGI in stem cells and loss of enhancer activity is sufficient to cause premature recruitment of H3K27me3 to the *Utf1* CGI.

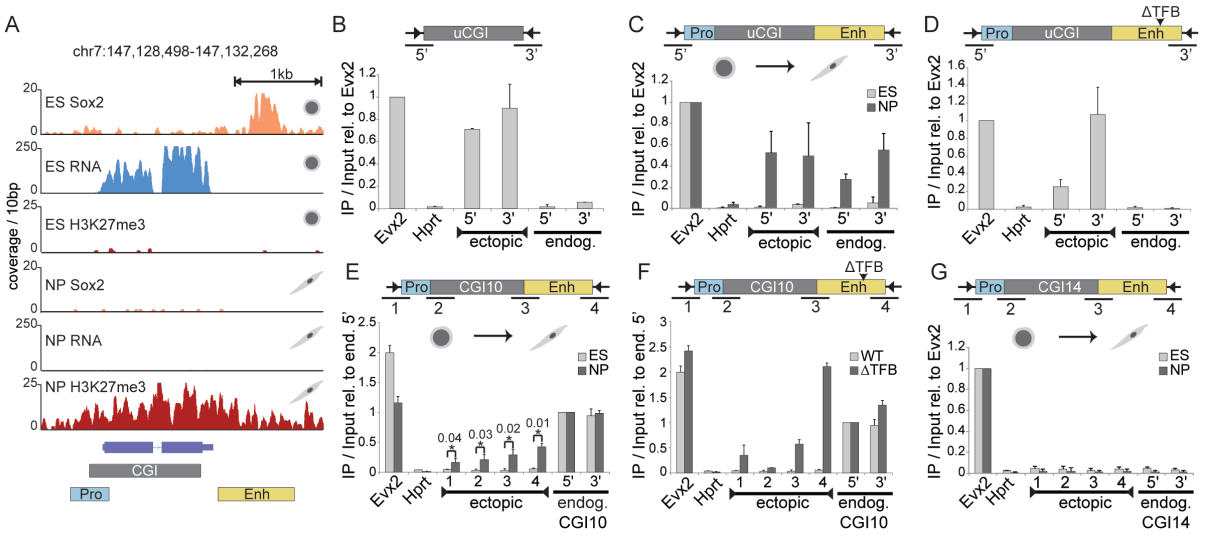

Figure 13 Transcriptional activity inhibits H3K27me3 acquisition. (A) Sox2, RNA-seq, and H3K27me3 profiles around the Utf1 locus in ESCs and NPs. (B) Ectopic insertion of the Utf1 CGI and corresponding H3K27me3 ChIP-qPCR enrichments in ESCs.; (C) Insertion of the whole Utf1 locus including its promoter and wild-type enhancer and corresponding H3K27me3 levels in ESCs and NPs.; (D) H3K27me3 enrichments at the Utf1 locus with the mutated enhancer in ESCs.; (E) Insertion of CGI10 flanked by the Utf1 promoter and enhancer and corresponding H3K27me3 enrichments in ES cells and NPs.; (F) Comparison of H3K27me3 levels between CGI10 in the context of the wild-type or mutant Utf1 enhancer in ESCs.; (G) As in E but with the negative control CGI14 placed between the Utf1 promoter and enhancer.; All qPCR data are normalized to input DNA and shown relative to the indicated amplicon (y-axis) of the endogenous locus in order to better compare the levels between the ectopic and endogenous sites. Evx2 and Hprt were used as endogenous positive and negative controls, respectively. Error bars represent standard deviation between at least two biological replicates. Numbers above asterisks indicate p-value.

Transcriptional context arbitrates recruitment abilities

To test whether the antagonism between recruitment and transcription was specific to the *Utf1* recruiter element, we replaced the *Utf1* gene-body CGI between the promoter-enhancer pair with the recruiter CGI10 (Figure 11D). Insertion of CGI10 in the context of the *Utf1* promoter-enhancer combination indeed inhibited H3K27me3 acquisition in ESCs (Figure 13E). Furthermore, upon differentiation into NPs, the CGI reacquired the H3K27me3 mark concomitant with loss of transcriptional activity (Figure 13E, Figure 18B). Using mutated constructs, we also observed that the OCT4/SOX2 binding site in the enhancer is essential for preventing H3K27me3 deposition in ESCs (Figure 13F). To further verify that H3K27me3 recruitment in NPs is indeed mediated by the CGI10 and not the *Utf1* promoter or enhancer, we repeated the experiment with the CGI10 replaced by the non-recruiting CGI14 (Figure 11F, Figure 13G). This construct no longer acquired H3K27me3 upon differentiation, confirming that the observed gain of this histone mark at the CGI10 construct was indeed due to the presence of the inserted CGI.

Thus cell type specific recruitment of the H3K27me3 mark can be counteracted by transcriptional activation mediated by surrounding regulatory regions.

DNA methylation and H3K27me3 recruitment exclude each other at CpG islands

All tested recruiter sequence elements are CpG dense as they originate from CGIs, suggesting that CpG dinucleotides might play an important role in the recruitment of the H3K27me3 mark to these target sites. In support of this model, we did not observe H3K27me3 at the CpG-free Luciferase reporter construct described above, even in the inactive state **(**Figure 18D), suggesting that CpG dinucleotides are required for H3K27me3 recruitment. Such a model has previously been proposed based on the strong presence of H3K27me3 at CGIs (Tanay et al. 2007, Ku et al. 2008) and the observation that two CpG-rich elements from the *e.coli* genome become H3K27 methylated when placed into stem cells as part of larger transgenes (Mendenhall et al. 2010, Lynch et al. 2012). The identification of small endogenous sequences that can autonomously recruit H3K27me3 enabled us to rigorously test this model. To determine if the density of CpGs and their individual

position is sufficient for recruitment, we synthetically modified three of the identified sequence elements (Figure 14). Specifically, we kept all CpG dinucleotides and their relative positions but replaced the rest of the sequence with random stretches of prokaryotic DNA from the *e.coli* genome. The purpose of this approach was to remove sequence cues (e.g. TF motifs) other than the CpG dinucleotides. Upon insertion, only one of three tested elements recruited Suz12 and showed H3K27me3 levels reminiscent of the wild-type fragment (Figure 14A, Figure 19A), while the other two displayed strongly reduced levels (Figure 14B,C and Figure 19B,C). Thus, CpG frequency and position cannot solely account for the ability to recruit H3K27me3. Notably, of the three modified recruiter elements above, the construct capable of recruitment also has the highest CpG density suggesting that very high CpG density could indeed be sufficient (Figure 14).

We have previously suggested that transcription factor binding within CGIs contributes to their low DNA methylation (Lienert et al. 2011). This could be relevant for the recruitment of the H3K27me3 mark since occupancy of DNA methylation and H3K27me3 in ES cells has been observed to be mutually exclusive at CGIs (Brinkman et al. 2012), leading to the hypothesis that DNA methylation could inhibit H3K27me3 acquisition. To explore whether increased DNA methylation could account for the absence of H3K27me3 at the two elements with modified sequence backbone we first determined their DNA methylation state. Bisulphite sequencing revealed that elements depleted in H3K27me3 upon alteration of the sequence backbone were indeed fully DNA methylated (Figure 14B,C and Figure 20A,B). Conversely, their wild-type counterparts showed little to no DNA methylation upon insertion (84% vs 19% and 94% vs 17%, respectively). The third element retaining H3K27me3 showed very low methylation even with the prokaryotic backbone (14% vs 0%, Figure 14A and Figure 20C). Thus, loss of Polycomb recruitment coincides with increased susceptibility to DNA methylation. This is compatible with the observation of increased H3K27 methylation upon global loss of DNA methylation (Brinkman et al. 2012, Lynch et al. 2012, Long et al. 2013, Reddington et al. 2013), which we also observe in our genome-wide datasets (Figure 21A). With this in mind, we also analyzed the DNA methylation of the smaller fragments described above (Figure 12C). Again the smaller fragment that lost the ability to recruit H3K27me3 became fully DNA methylated (Figure 20D). Taken together, this supports a model of

competition between DNA methylation and H3K27me3 recruitment at the tested CGIs.

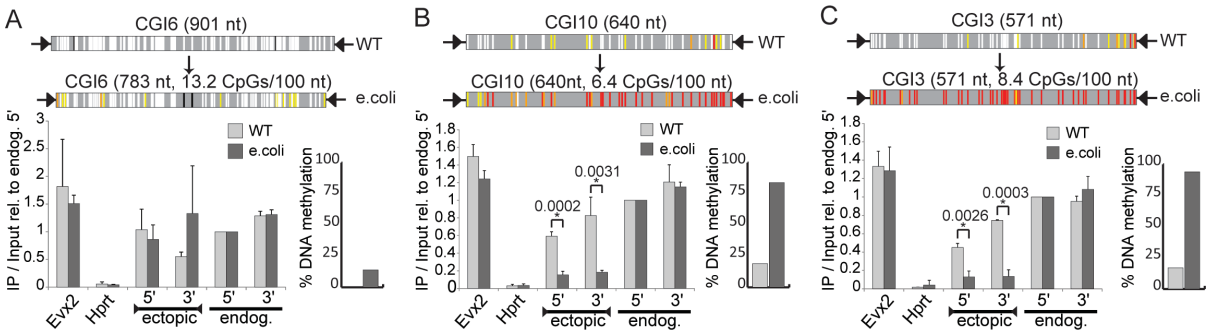

Figure 14 CpG dinucleotides cannot solely account for the establishment of H3K27me3 domains. (A)-(C) Replacement of all non-CpG nucleotides in CGI6 (A), CGI10 (B), and CGI3 (C) with prokaryotic DNA. Top panels represent CpG dinucleotide positions within the sequences and the color indicates the level of DNA methylation (white through red indicates 0% - 100%). Bar plots show H3K27me3 ChIP-qPCR levels at the tested sequences (left) and total DNA methylation percentages (right).; All qPCR data are normalized to input DNA and shown relative to the 5' amplicon of the endogenous locus in order to better compare the levels between the ectopic and endogenous sites. Evx2 and Hprt were used as endogenous positive and negative controls, respectively. Error bars represent standard deviation between at least two biological replicates. Numbers above asterisks indicate p-values.

Trimetylation of H3 at lysine 4 (H3K4me3) has been demonstrated to co-occupy PRC2 bound CpG islands in ES cells resulting in bivalent domains (Bernstein et al. 2006). When determining H3K4me3 levels at the three elements above, CGI3 and CGI10 both displayed low levels of H3K4me3 at the endogenous site and as wild-type inserts (Figure 19E,F). This signal was reduced upon sequence mutation in agreement with observations that H3K4me3 is absent at methylated CGIs (Weber et al. 2007). In case of CGI6 the wild-type insert showed high levels of H3K4me3 while the signal in the mutant was reduced even though it remains DNA unmethylated and maintains wild-type levels of H3K27me3 (Figure 19D, Figure 14A). One potential explanation is that mutation of the sequence backbone removed TF binding sites that positively affect H3K4 methylation at this sequence.

DNA methylation counteracts H3K27me3 recruitment

If DNA methylation at the two modified recruiter elements inhibits deposition of H3K27me3, removing DNA methylation should recover H3K27me3. To test this, we treated the respective cells with the DNMT inhibitor 5-Aza-2′-deoxycytidine (Juttermann et al. 1994), which resulted in loss of DNA methylation to levels closely reflecting those observed at wild-type CGIs (Figure 21B,C). Strikingly, reacquisition of H3K27me3 levels comparable to the corresponding wildtype sequences occurred

coincident with reduced DNA methylation (Figure 15A,B). We thus conclude that DNA methylation can directly block the ability of CpG rich DNA stretches to recruit the H3K27me3 mark.

The above results predict that a CGI that is not K27 methylated should be either DNA methylated or transcriptionally active. This is indeed what we observe at the level of the genome (Figure 21D) and at the CGI that was used as a negative control and which did not recruit PRC2 (CGI14, Figure 11F). This particular sequence is not DNA methylated after insertion but recruits efficiently RNA polymerase, similar to its endogenous site again in line with this model (Figure 20E and Figure 20F).

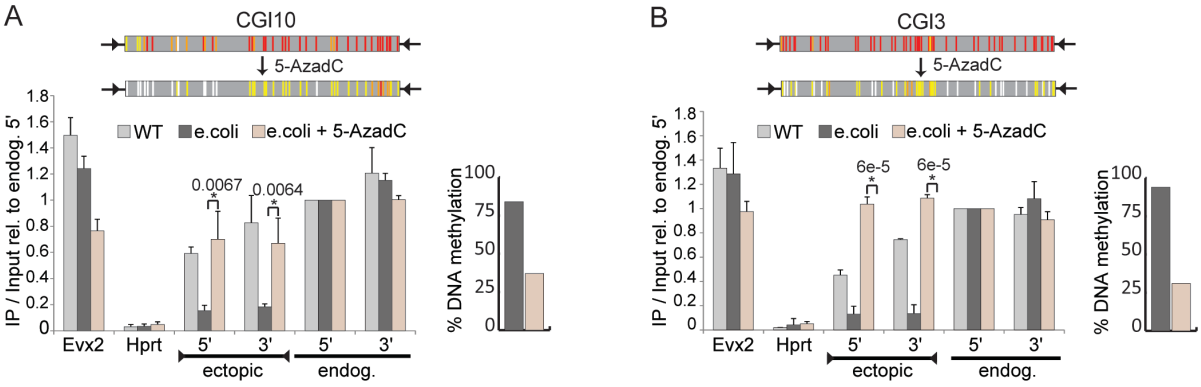

Figure 15 Removal of DNA methylation reestablishes acquisition of H3K27me3 at the modified fragments. (A)-(B) ESCs harboring the modified versions of fragments CGI10 (A) and CGI3 (B) were treated with 5-Aza-2′-deoxycytidine for four days. Top panels represent CpG dinucleotide positions within the sequences and the color indicates the level of DNA methylation (white through red indicates 0% - 100%). Bar plots show H3K27me3 ChIP-qPCR levels at the tested sequences (left) and total DNA methylation percentages (right). All qPCR data are normalized to input DNA and shown relative to the 5' amplicon of the endogenous locus in order to better compare the levels between the ectopic and endogenous sites. Evx2 and Hprt were used as endogenous positive and negative controls, respectively. Error bars represent standard deviation between at least two biological replicates. Numbers above asterisks indicate p-values.

3.2.5 Discussion

Using iterative genome editing, we define minimal DNA sequence elements capable of recruiting PRC2 and acquire H3K27me3 in stem cells. The genomic locus used in our studies is otherwise devoid of this mark, allowing us to identify DNA elements as short as 220 nt that autonomously create local H3K27me3 domains. Furthermore, we could recapitulate H3K27me3 dynamics upon differentiation into neural precursors, demonstrating that the cellular signal for recruiting PRC2 and H3K27me3 is encoded within the target DNA sequence. Of particular note, by genetically uncoupling the contribution of CpGs from the rest of the DNA sequence we show that CpG frequency and positioning is sufficient to recruit PRC2. Additionally, we demonstrate that DNA methylation directly inhibits H3K27me3 deposition at CpG rich sequences and that this inhibition can be directly reversed upon chemical removal of DNA methylation. These observations readily explain the occupancy of H3K27me3 and DNA methylation observed throughout the genome (Brinkman et al. 2012, Lynch et al. 2012, Long et al. 2013, Reddington et al. 2013). They extend previous studies that already suggested a role for CpG dinucleotides in PRC2 recruitment (Mendenhall et al. 2010, Lynch et al. 2012) but are also more comprehensive and provide novel mechanistic details.

Our data reveal that transcriptional context surrounding recruiter sequences is critical in H3K27me3 establishment. By coupling enhancer dependent activation to otherwise transcriptionally inert constructs, we were able to inhibit premature acquisition of the H3K27me3 mark. This suggests that active transcription prevents PRC2 binding, a finding confirmed by minimal deletions in motifs for activating transcription factors. Importantly, while the transcriptional state of a promoter is critical in H3K27me3 deposition, our data argue that it is codependent on CpG density. It is important to note that alternate transcriptional contexts have been implicated in recruiting the polycomb machinery to chromatin. For example, PRC2 has been suggested to interact with RNAs in a promiscuous manner leading to an elaborate model whereby nascent RNA accumulation results in a chromatin state refractory to H3K27me3 deposition (Davidovich et al. 2013). Our observations implicate transcriptional activity as an antagonistic factor for PRC2 activity and are thus compatible with these findings. Notably, our results argue that acquisition of

H3K27me3 is not the cause of target gene repression, but rather occurs in the absence of a transcriptional activator at CG-rich sequences.

TFs themselves have been implicated in modulating local levels of H3K27 methylation and in line with the above model many of these factors serve as transcriptional repressors (e.g. REST and SNAI1) (Herranz et al. 2008, Dietrich et al. 2012, Arnold et al. 2013). Notably, ES cells lacking REST display only subtle effects on H3K27me3 levels at target loci, while more pronounced implications are observed upon neuronal differentiation (Arnold et al. 2013). Additionally, loss of REST often accompanies transcriptional upregulation (Arnold et al. 2013), an observation compatible with our enhancer mutation experiments. A precedent for TF mediated H3K27me3 deposition was originally described in *Drosophila melanogaster* where, among others, the TF pleiohomeotic (Pho) seems to be involved in recruiting Polycomb (Brown et al. 1998, Brown et al. 2003, Ringrose et al. 2003, Klymenko et al. 2006). The mammalian homologue Yin-Yang1 (YY1), however displays mutually exclusive binding patterns with PRC2 in mouse stem cells (Vella et al. 2012).

Notably the fly genome lacks DNA methylation and CGIs and it is thus tempting to speculate that the adaptation to these genomic elements in mammals played a critical role in shaping H3K27me3 distribution. A critical question is thus how unmethylated CpGs are interpreted to become H3K27me3 domains in mammalian genomes. While several CXXC domain containing proteins can recognize the unmethylated CpG dinucleotide (Thomson et al. 2010, Long et al. 2013), it is unclear if they account for the observed genome-wide H3K27me3 distribution. Indeed, two recent studies have shown that the CXXC domain protein KDM2B interacts with PRC1 and binds to CGIs in stem cells and (Farcas et al. 2012, Wu et al. 2013). Importantly, however, KDM2B cannot solely account for PRC1 recruitment as the protein also binds active CGIs (Farcas et al. 2012, Wu et al. 2013). Furthermore, loss of KDM2B only results in a mild reduction of PRC1 recruitment at relatively few targets (Farcas et al. 2012, Wu et al. 2013). Another factor implicated in PRC2 recruitment is the lysine methyltransferase cofactor Jarid2 (Peng et al. 2009, Shen et al. 2009, Landeira et al. 2010, Li et al. 2010, Pasini et al. 2010). While Jarid2 may be involved in this process, it actually binds CG-rich DNA with relatively low affinity (Kim et al. 2003, Li et al. 2010) and its loss results only in mild H3K27me3 changes (Shen et al. 2009, Li et al. 2010) indicating that Jarid2 may

fine-tune rather than directly recruit PRC2. Our findings are in agreement with the concept that PRC1 and PRC2 are recruited to transcriptionally inactive, unmethylated CpG-rich DNA sequences (Klose et al. 2013). We envision a model whereby unmethylated CGIs are PRC2 occupied by default in the absence of active transcription. Presence of the H3K27me3 mark may serve to suppress sporadic expression of target genes, while concurrently allowing for future transcriptional activation. The dispensability of PRC2 in ES cells lends support to this model, as well as the fact that only few genes are upregulated in PRC2 null stem cells (Boyer et al. 2006, Leeb et al. 2010).

Through analysis of several sequence variants in a controlled system we were able to delineate a set of guidelines for H3K27me3 recruitment in mouse stem cells. Critical parameters for this establishment include CpG density, transcriptional competency and DNA methylation. The robustness of these criteria is exemplified by our experiments recapitulating Polycomb recruitment on an element spanning less than two nucleosomes.

3.2.6 Methods

Cell lines and cell culture. TC-1 ES cells (Lienert et al. 2011) and Dnmt^{TKO} cells (Tsumura et al. 2006) were cultured on 0.2% gelatin coated dishes. Growth medium consisted of DMEM (Invitrogen) supplemented with 15% fetal calf serum (Invitrogen), 1x non-essential amino acids (Invitrogen), 1 mM L-glutamine, LIF, and 0.001% β-mercaptoethanol. Differentiation was performed as previously described (Bibel et al. 2007). Treatment with 5-Aza-2′-deoxycytidine (Sigma) was performed over a course of four days at a concentration of 0.1 μM.

Recombination mediated cassette exchange. For targeted insertion, DNA fragments were cloned into a plasmid containing a multiple cloning site flanked by two inverted L1 *LoxP* sites (L1-poly-1L). Promoter regions were amplified from TC-1 ES cell genomic DNA (primers in **Table S1**) or synthesized (modified CGIs, **Table S2**) and TF binding site deletions were introduced by Quick-Change PCR. RMCE was performed as described (Lienert et al. 2011) with slight modifications: TC-1 ES cells were selected under hygromycin (250 μg ml-1, Roche) for 10 days. Next, 4 × 106 cells were electroporated (Amaxa nucleofection, Amaxa) with 25 μg of L1-poly-1L plasmid and 15 μg of pIC-Cre. Selection with 3 μM Ganciclovir (Roche) was started 2 days after transfection and continued for 7–10 days. The surviving population was then diluted and selected for another 7-10 days. Single clones were tested for successful insertion events by PCR using site-specific primers.

Chromatin IP. ChIP was carried out as previously described (Weber et al. 2007): Cells were cross-linked in medium containing 1% formaldehyde for 10 min at room temperature. The reaction was quenched with 150 mM Glycine on ice. Crosslinked cells were washed twice with ice-cold PBS and incubated on ice 5 min in Buffer1 containing 10 mM Tris pH 8.0, 10 mM EDTA pH 8.0, 0.5 mM EGTA, and 0.25% Triton X-100. The cells were centrifuged 5 min at 600g followed by 5 min incubation on ice in Buffer2 consisting of 10mM Tris pH 8.0, 1 mM EDTA, 0.5 mM EGTA and 200 mM NaCl. Finally, cells were centrifuged again and resuspended in lysis buffer consisting of 50 mM HEPES/KOH pH 7.5, 500mM NaCl, 1mM EDTA, 1% Triton X-100, and 1x Protease Inhibitor cocktail (Roche). The crosslinked chromatin lysate was sonicated for 15 (stem cells) or 20 cycles (neuronal progenitors) of 30 sec each

using a next-gen Diagenode Bioruptor on high setting. Sonicated chromatin was divided into 70 µg aliquots and either stored at -80°C or used immediately for ChIP. In the latter case aliquots were pre-cleared with BSA and tRNA blocked Protein A or G sepharose beads for 1h, rotating at 4°C. 5% of pre-cleared lysate was kept as an input sample and the rest was then incubated overnight with the corresponding antibodies, rotating at 4°C and subsequently with 30 µl Protein A or G beads for 3h at 4°C. Chromatin-bound beads were then washed twice with lysis buffer containing protease inhibitors at room temperature for 5 min followed by a single 5 min wash with DOC buffer consisting of 10mM Tris Ph 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% DOC, and 1 mM EDTA. Beads were then transferred to a fresh tube and bound chromatin was eluted in elution buffer containing 1% SDS and 100 mM $NaHCO_3$ in two rounds of 20 min, rotating at room temperature. Eluted chromatin and input sample were treated with 50 µg ml^{-1} RNase A for 30min at 37°C and subsequently reverse-crosslinked by addition of 200 µg ml^{-1} Proteinase K and 3h incubation at 55°C followed by overnight incubation at 65°C. DNA was then purified by phenol / chlorophorm extraction, precipitated in ice-cold ethanol, and resuspended in 40 µl (IP) or 50 µl (Input DNA) TE pH 8.0. Quantitative PCR (qPCR) was performed using SYBR Green chemistry (Applied Biosystems) and 1/80th of the ChIP reaction or 20 ng of input chromatin per PCR reaction, respectively. Standard curves and primer efficiencies for each primer pair used in qPCR were determined by measuring signal intensity of a dilution series of input DNA ranging from 100 ng to 0.1 ng. All IP measurements were normalized to the corresponding input values. Samples were excluded from analysis when the experiment clearly failed due to technical issues (no enrichments for positive controls). Error bars in all figures represent standard error from at least two biological replicates. Significances were calculated using a one-tailed, unpaired two-sample T-test. The following antibodies were used in this study: H3K27me3 (Millipore, #07-449 and Abcam, #ab6002), Suz12 (Cell Signaling, #3737), Pol II (Santa Cruz, #SC899), H3K4me3 (Abcam, ab8580), and Sox2 (R&D Systems, AF2018). Primers used for qPCR are listed in **Table S3**. H3K27me3 ChIP in TKO-133 cells was subjected to high-throughput sequencing on an Illumina GAII sequencer using standard Illumina library preparation kits and protocols (GEO accession GSE56110). For details regarding sequencing data analysis see supplementary methods.

High-throughput sequencing data analysis. The July 2007 Mus musculus genome assembly (NCBI37/mm9) provided by the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/genome/guide/mouse/) and the Mouse Genome Sequencing Consortium (http://www.sanger.ac.uk/Projects/M_musculus/) was used as a basis for all analyses. A nonredundant, nonoverlapping set of promoters (n=19,534) was obtained by using the annotation of known RefSeq transcripts from the University of California, Santa Cruz genome browser (http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/refGene.txt.gz, downloaded on April 24, 2013), generating 2,000-bp windows centered around RefSeq transcription start sites and removing all overlapping windows. CpG island annotation ("cpgIslandExt") was downloaded from the UCSC genome browser (http://genome.ucsc.edu/) using the R package rtracklayer (Lawrence et al. 2009) and all islands were resized to a length of 1 kb centered around their annotated midpoint. Fully methylated, low-methylated and unmethylated regions were identified applying the R package MethylSeekR (Burger et al. 2013) to mouse embryonic stem cell whole-genome bisulfite data (Stadler et al. 2011). The following published datasets were used for analysis: H3K27me3 in ES cells (2 replicates, GEO accessions GSM632032-GSM632034), H3K27me3 in DNMT triple knockout ES cells (GEO accession 56110), Input chromatin (GEO accession: GSM671103), Pol II (2 replicates, GEO accessions GSM747547-GSM747548). Alignment and quantification of ChIP-seq samples was performed using the qAlign and qCount function of the R package QuasR, version 1.2.2 (http://www.bioconductor.org/packages/release/bioc/html/QuasR.html). For ChIP-seq quantification (single-end), read counts were assigned to the 5' coordinate of the mapped read and shifted by 60 nts towards the 3' end (ie assuming a fragment length of 120 nts, shift=60 in the qCount/qProfile functions). To account for differences in read depths in the various samples, the total counts for all regions (promoters or CpG islands) were normalized to the minimal number of counts observed in any of the samples. Region levels were then determined as log2(x+8), where x is the number of normalized read counts and 8 is a pseudocount (to reduce the sampling noise). For K27me3 and PolIII wt, reads counts from each replicate were pooled. To generate the heatmap in Figure 21D, counts were aggregated in consecutive bins of 50nts for each region (ie 20 bins for

1kb) and enrichments (over input) were first calculated separately in each bin and then averaged over the two neighbouring bins. Methylation levels in each bin were determined as the ratio of summed-up methylated counts divided by the summed-up total counts for all CpGs overlapping the bin, only retaining bins with at least 5 total counts. For bins, which were not covered or did not contain any CpGs, methylation levels were inferred by linear interpolation from neigbhbouring bins. To avoid overplotting, 1000 randomly chosen CpG islands were used for visualization. These regions were clustered via kmeans-clustering with 3 clusters using Z-scores of Pol II and K27me3 (vs input) enrichments as well as Z-scores of methylation. For locus specific ChIP-seq profiles shown throughout figures 1, S1, 3, and S4 we generated wiggle tracks depicting genome-wide coverage. Published data was used for H3K4me3 (GSM207618), Sox2 (GSE35496), Suz12 (GSM480680), and RNA (GSE34473).

Methylation analysis. Cells were lysed in cell lysis buffer containing 20 mM Tris pH 8.0, 4 mM EDTA, 20 mM NaCl, 1 % SDS, and 0.7 mg/ml Proteinase K, followed by incubation at 55°C for 5 hours. Genomic DNA was extracted by phenol-chloroform extraction and precipitated in EtOH. 2ug of genomic DNA was bisulfite converted with the EpiTec Bisulfite Kit (QIAGEN). Regions of interest were amplified by PCR and cloned by TOPOTA cloning (Invitrogen). Bisulfite PRC primers were designed to be complementary to the converted DNA and did not contain CpG dinucleotides to avoid biases (see **Table S4**). Sequences were analyzed using BiQ Analyzer (Bock et al. 2005).

RNA expression. RNA was isolated from cells using Trizol / chloroform extraction. Briefly, cell pellets were resuspended in 1 mL of Trizol Reagent and incubated for 5 min at room temperature, followed by the addition of 0.2 mL chlorophorm and another incubation for 3 min at room temperature. The Trizol / chlorophorm solution was centrifuged for 15 min at 4°C and 12'000g. The aqueous phase was transferred into a fresh 1.5 mL eppendorf tube followed by the addition of 0.5 mL isopropanol. This mixture was incubated for 10 min at room temperature, followed by 10 min centrifugation at 4°C and 12'000g. The pellet was washed in 75% EtOH and then air-dried. The RNA pellet was treated with DNaseI for 50 min at 37°C in a volume of 40

µL to remove residual DNA contaminants. cDNA synthesis was performed using the PrimeScript RT reagent Kit (Takara). Corresponding cDNA levels were measured by quantitative PCR and normalized to endogenous LaminB levels.

Luciferase activity measurement. Lysate preparation and measurement of luciferase activity was done according to the Luciferase Assay Kit (Promega). Luminescence was normalized to total protein content in the lysate measured with a BCA Protein Assay Kit (Pierce).

3.2.7 Acknowledgements

The authors thank C. Wirbelauer for technical support and M. Lorincz and members of the Schübeler laboratory, in particular P. Ginno, for critical reading of the manuscript. Research in the laboratory of D.S. is supported by the Novartis Research Foundation, the European Union (NoE ‘‘EpiGeneSys’’ FP7- HEALTH-2010-257082 and the ‘‘Blueprint’’ consortium FP7-282510), the European Research Council (EpiGePlas), the SNF Sinergia program, and the Swiss initiative in Systems Biology (RTD Cell Plasticity). The authors declare no competing financial interests.

3.2.8 Supplemental Material

Supplemental Figures

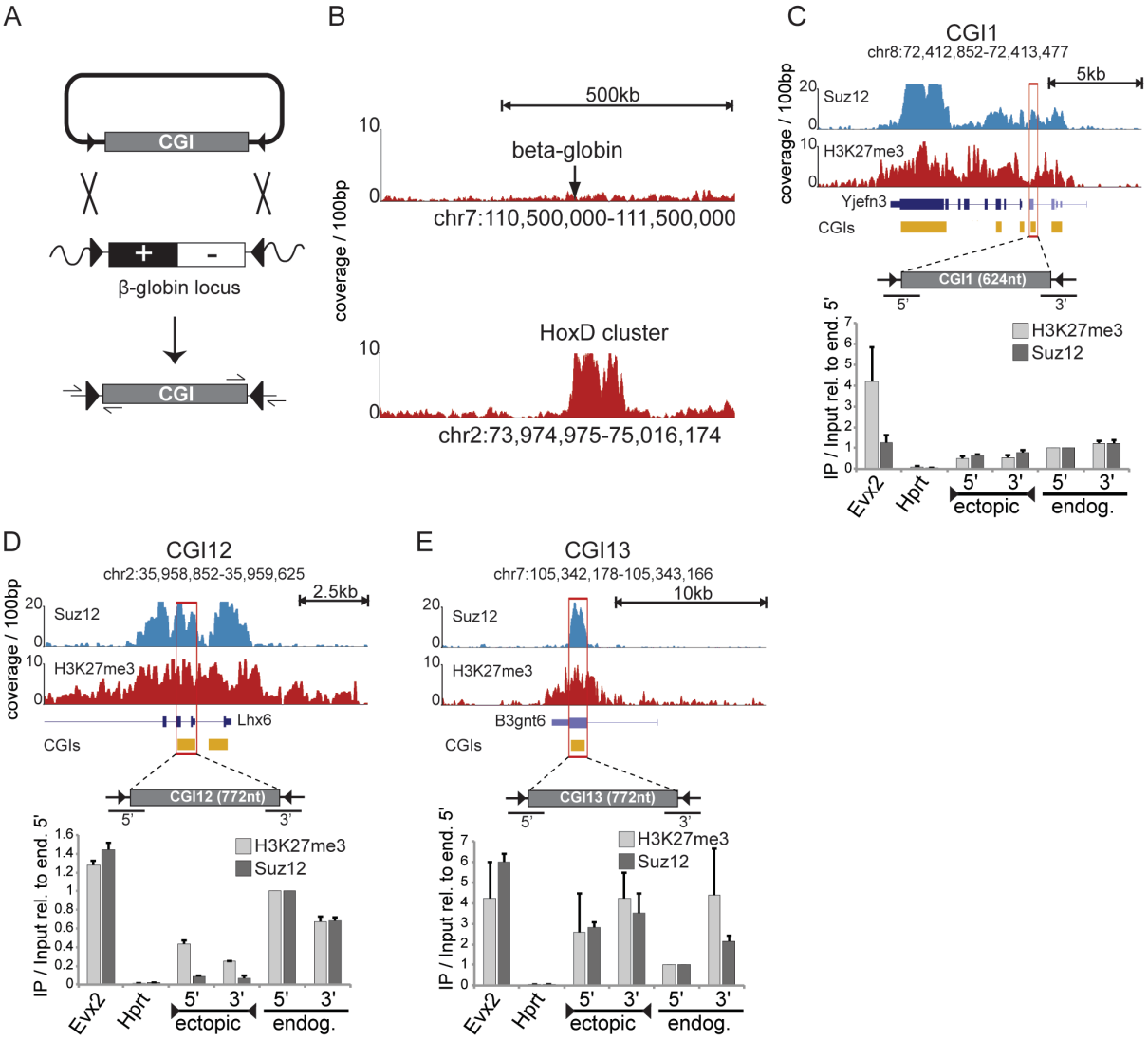

Figure 16 Short DNA sequences reconstitute endogenous H3K27me3 patterns when inserted ectopically. (A) Schematic representation of ectopic insertion of DNA sequences by recombination-mediated cassette exchange (RMCE, see text for details).; (B) H3K27me3 profiles within 500 kb around the Hbb-y locus (top) and the HoxD gene cluster (bottom).; (C)-(E) Top panels show Suz12 and H3K27me3 profiles around the candidate CGIs. Genomic coordinates indicate location of the candidate CGI. Bottom panel shows H3K27me3 and Suz12 ChIP-qPCR data at the ectopically inserted and endogenous loci.; ChIP-qPCR data in all panels are shown relative to the 5' amplicon of the corresponding endogenous locus. Error bars indicate standard deviation between at least two biological replicates.

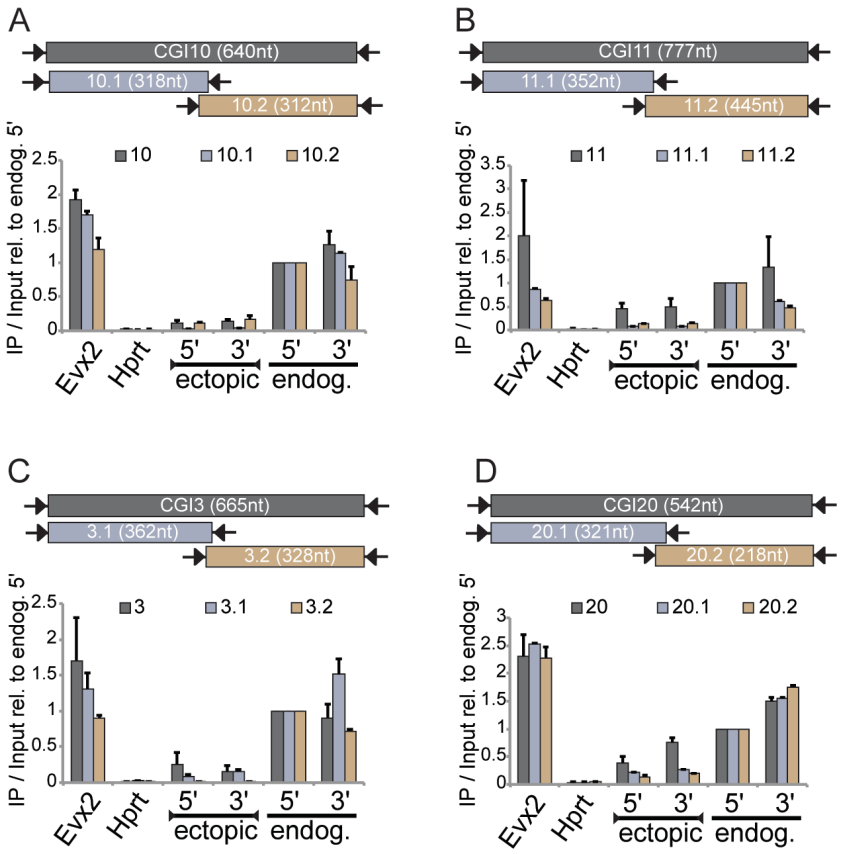

Figure 17 Minimal recruiter elements are bound by Suz12. A)-(D) Recruiter elements were dissected into smaller parts and analyzed for Suz12 enrichments by ChIP-qPCR. Error bars represent standard deviation between at least two biological replicates.

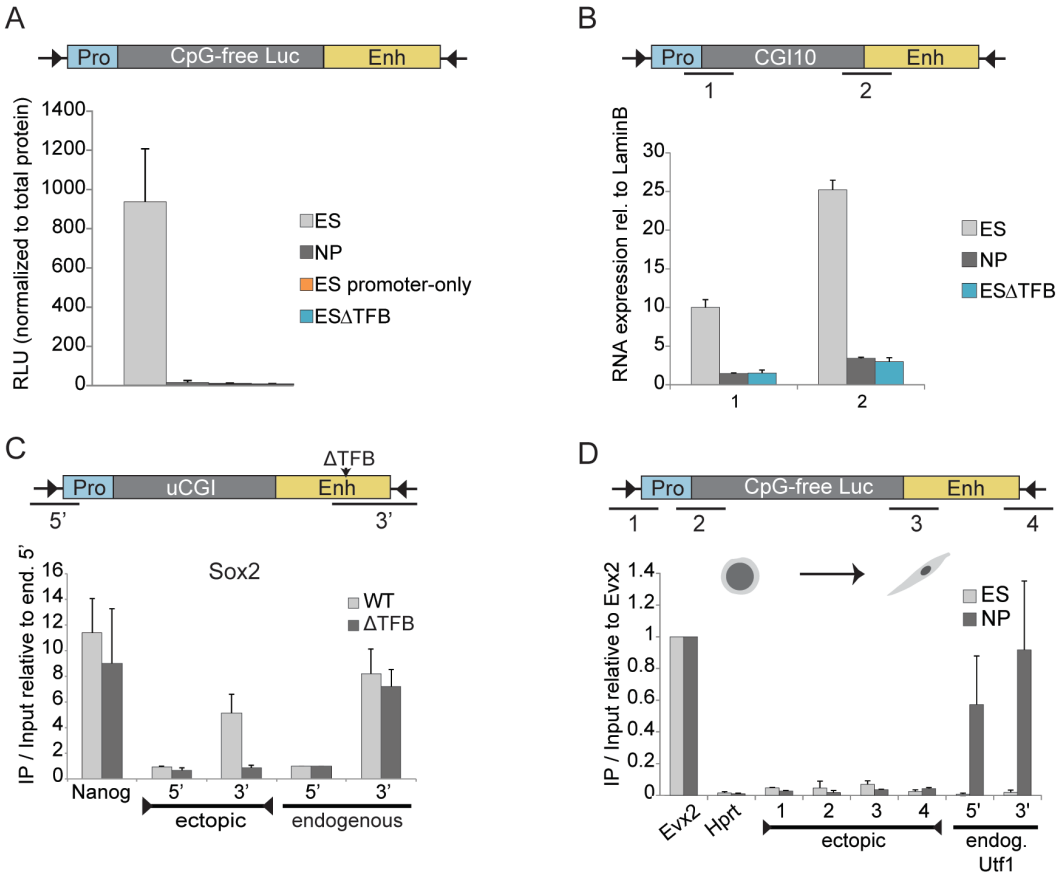

Figure 18 Ectopic *Utf1* promoter-enhancer activity reflects endogenous pattern. (A) A CpG-free Luciferase gene driven by the *Utf1* promoter with or without the enhancer was stably inserted in the beta-globin locus and activity was measured in ESCs and / or NPs. ΔTFB indicates OCT4/SOX2 binding site deletion.; (B) CGI10 was inserted between the wild-type or mutated Utf1 promoter-enhancer pair and RNA levels were measured in ESCs and NPs. (C) Comparison of Sox2 ChIP enrichments between the *Utf1* locus with either the wild-type or mutant enhancer in ES cells; (D) H3K27me3 enrichments at the Luciferase construct in ES cells and NPs.; Error bars in all panels represent standard deviation between at least two biological replicates.

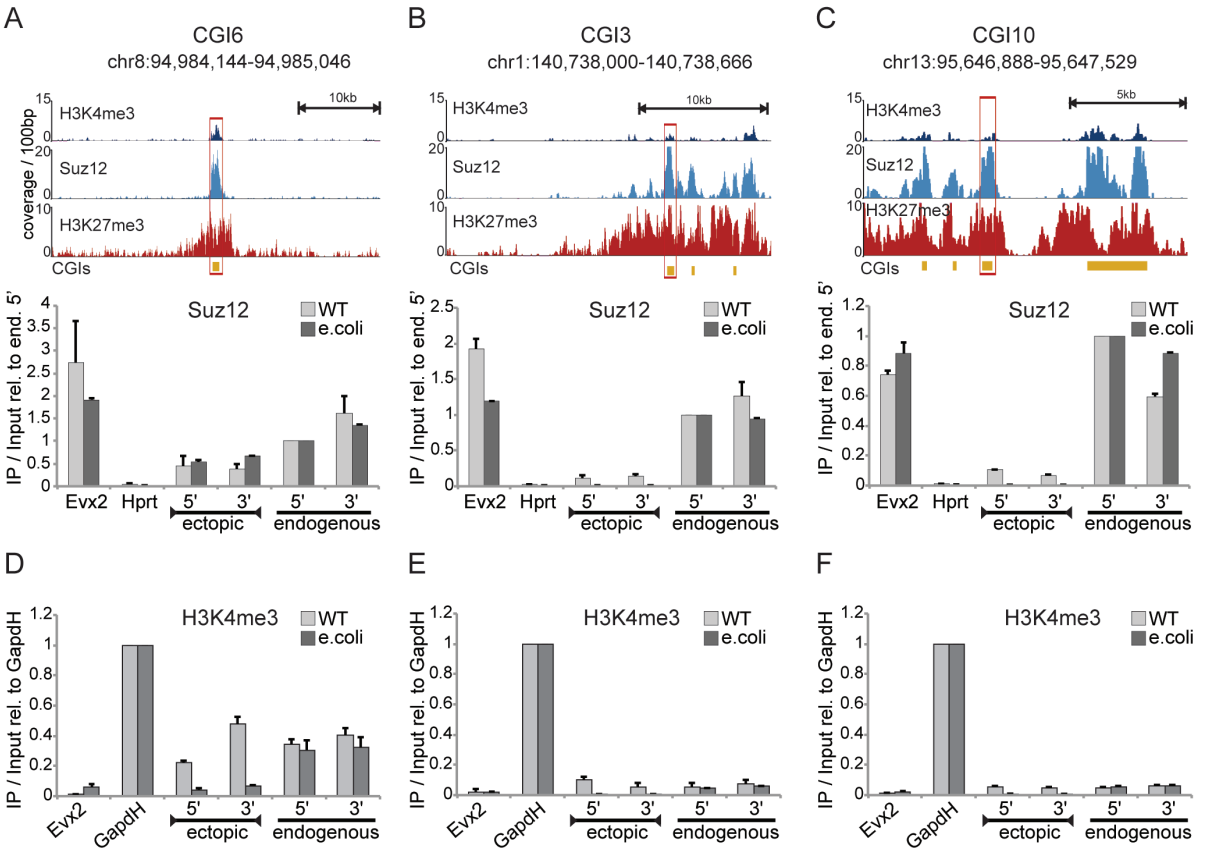

Figure 19 Suz12 and H3K4me3 occupancy at the wild-type and mutated CGIs. (A)-(C). Top panels show H3K4me3, Suz12, and H3K27me3 profiles around the candidate CGIs. Genomic coordinates indicate location of the tested elements. Bottom panels show Suz12 ChIP-qPCR data at the wild-type and mutated elements.; (D)-(F) H3K4me3 ChIP enrichments at the wild-type and mutated elements. Error bars in all panels represent standard deviation between at least two biological replicates.

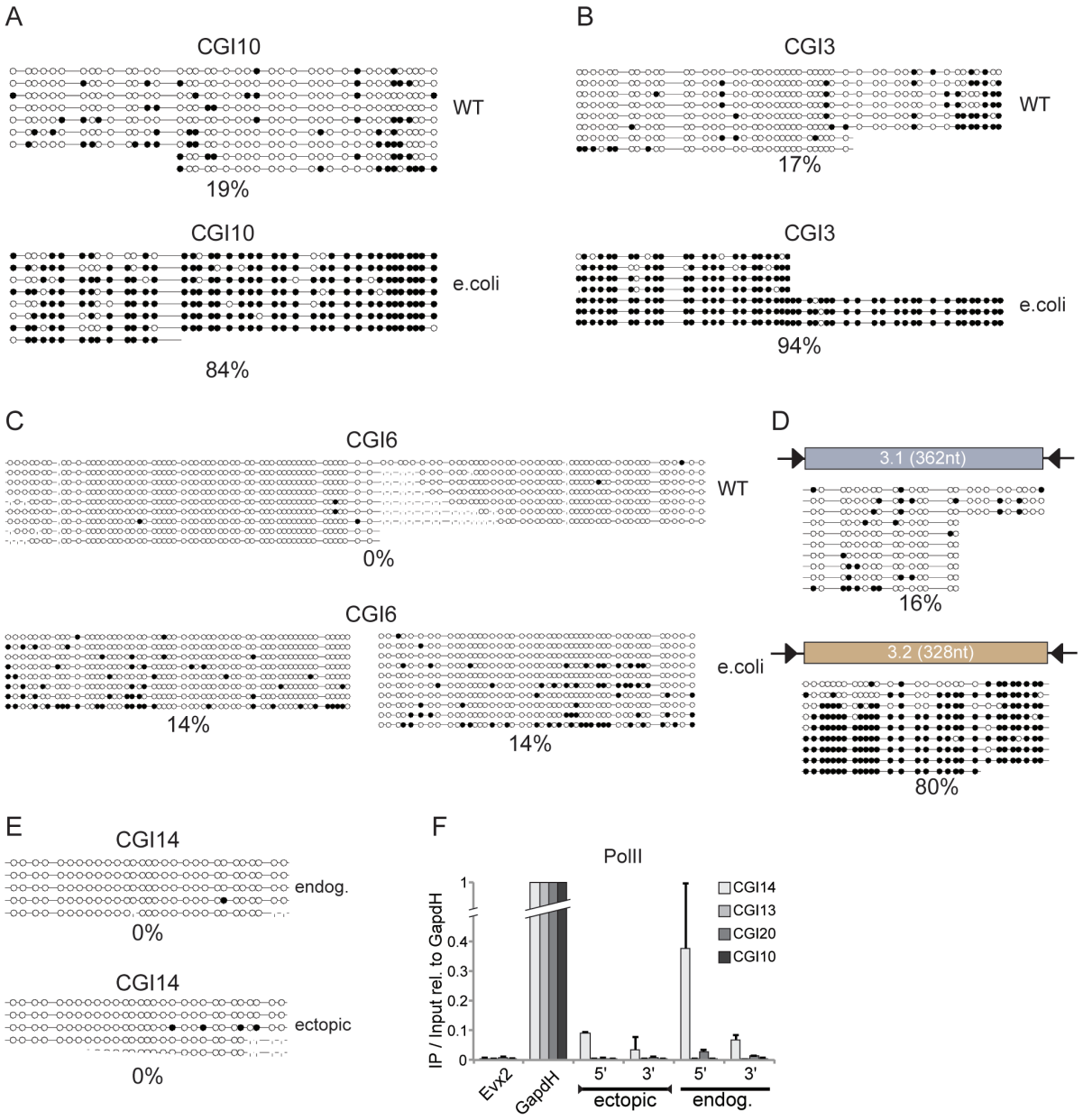

Figure 20 DNA methylation levels at the modified recruiter sequence elements (A)-(C) Bisulphite sequencing data for CGI10 (A), CGI3 (B), and CGI6 (C). Top panels show DNA methylation levels at the wild-type sequences, bottom panels at the mutant elements.; (D) Bisulphite sequencing data for ectopic insertion of CGI 3.1 and 3.2 in ES cells. (E) Bisulphite sequencing data for CGI14 at the endogenous and ectopic locus.; (F) Pol II ChIP-qPCR data for CGIs 14, 13, 20, and 10.

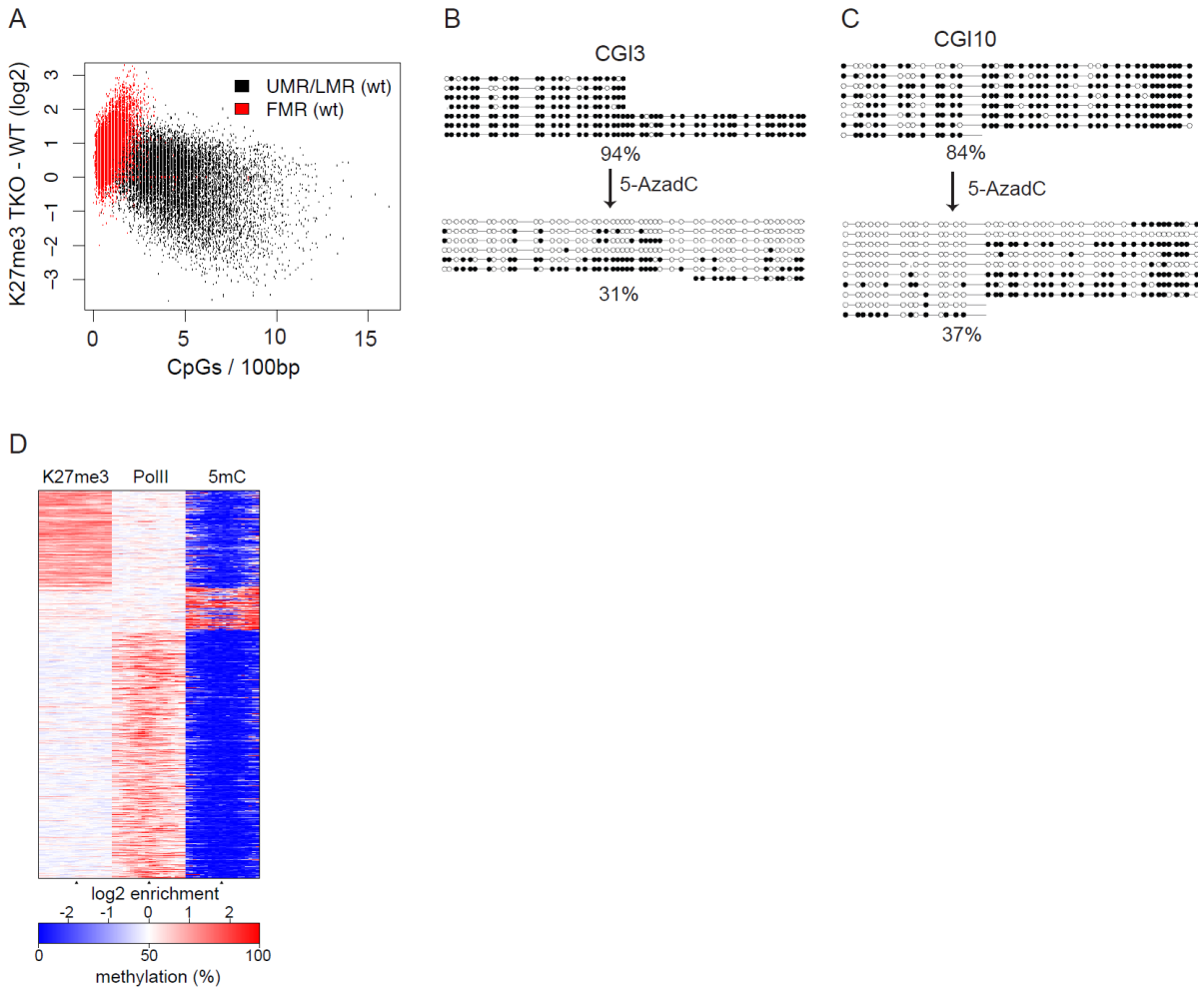

Figure 21 DNA methylation inhibits H3K27me3 acquisition. (A) Scatterplot showing H3K27me3 in DNMT triple knockout ES cells at all gene promoters in relation to the DNA methylation state in wild-type ES cells.; (B)-(C) Bisulphite DNA methylation profiles at the modified fragments CGI3 (B) and CGI10 (C) before and after treatment with 5-Aza-2'-deoxycytidine. (D) Heatmap comparing K27me3, Pol II and DNA methylation at all CGIs.

Primers used in this study

Table S1 Primers used to clone DNA fragments from genomic DNA

Fragment Name	Sense Primer	Anti-sense Primer
CGI1	AAAAAGGATCCGCTAGGGTAACCAGGCTGTC	TTTTTAAGCTTCGCCCTGCTTTGCTATGT
CGI3 665nt	AAAAAGATATCGTCTTCTATCCCAGGGCAAGG	TTTTTAAGCTTCTTGTCCCTGTTCCCCAAC
CGI3 571nt	AAAAAGGATCCCGGGCGGATCCGGCCTGGCG	AAAAAATCGATCGCGTCTCTACGCTTCTGCA
CGI6	AAAAAGGATCCATGCCTAGGTCAGCTATTCAA CA	TTTTTATCGATCAAAATGCCTTCTTTTCCTTG
CGI10	AAAAAGGATCCTATTTCAACCTAGCTGTCCAAT TC	TTTTTAAGCTTTCCTCCCCTTCCCTCCTT
CGI11	AAAAAGATATCGGTGGTGGTGGTTGAGAAAA	TTTTTAAGCTTCTCCAGCTCTCCCAGCAG
CGI12	AAAAAGGATCCCTACCCTGACTCACCTTGAGC	TTTTTAAGCTTGCTTTAAATGCCCACTACAAA
CGI13	AAAAAGATATCACATGAGCAGCATCTCATAAG G	TTTTTAAGCTTAGGAACACAGGCAGGAGAAA
CGI14	AAAAAGGATCCACTGCAGATGCTAATCTCCAC TG	AAAAAAAGCTTATACTGCCCTGAGGCTTGG
CGI20	AAAAAGGATCCCAGGGTTTGACCCAGAAAAG	AAAAAAAGCTTAGGACCTCAGCTGAACTATGC
3.2	AAAAAGATATCCCGGGAGACTCGCTGTTT	TTTTTAAGCTTCTTGTCCCTGTTCCCCAAC
3.1	AAAAAGATATCGTCTTCTATCCCAGGGCAAGG	TTTTTAAGCTTCCGGGAAACAGCGAGTCT
10.2	AAAAAGGATCCATCACCACCGTGGGCTCTA	TTTTTAAGCTTTCCTCCCCTTCCCTCCTT
10.1	AAAAAGGATCCTATTTCAACCTAGCTGTCCAAT TC	TTTTTAAGCTTGCAGCAGAGTGGCTCCTTC
11.2	AAAAAGATATCCTCCTCTTGATTCGCCTATG	TTTTTAAGCTTCTCCAGCTCTCCCAGCAG
11.1	AAAAAGATATCGGTGGTGGTGGTTGAGAAAA	AAAAAAAGCTTCATAGGCGAATCAAGAGGAG
20.2	AAAAAGGATCCCAGACTCGCCTCGATCTCC	AAAAAAAGCTTAGGACCTCAGCTGAACTATGC
20.1	AAAAAGGATCCCAGGGTTTGACCCAGAAAAG	TTTTTAAGCTTGGAGATCGAGGCGAGTCTG
Utf1 Locus	AAAAAATCGATGAGTACATCGCTGTCTTTTGAC AC	AAAAAGCCGGCCAAGCCCTGGGACCATCT
utf1 CGI	AAAAAGGATCCATGCTGCTTCGTCCCCGGA	AAAAAATCGATTTCTGGAGAAGAGGACTGATAACAA A
Oct4/Sox2 deletion	GAAGCTGCTGCCGGCACTTCACGGCTCATCCT GAGGCTC	GAGCCTCAGGATGAGCCGTGAAGTGCCGGCAGCA GCTTC

Table S2 Prokaryotic DNA sequences used to replace non-CpG dinucleotides

Name	prokaryotic sequence
CGI3	AAATATTGAATACACATGGATTGAAAACTATGACACAGCCCATTTAACATCAAAGAAAGTTAATAAGTTTGAGAT TATGGATCAAGTTGATGATGATAGACTTGCACAACTTATTCCTGATTTTTATGTCTTTCCAGAAAAAAGTGTAAG CTATAATATACTAAAGCAAGGTAAGCATGCTTTTATTTTGAGCATTGGTAACAGAGCAATAATGCATTGTGCAA GGTAAACTAAAATAGATAACAAAAAGATGGAACCTCAGCAAAGTCAGGCATTTATACTCTTTTGAATACATAGA AAATTGATATCAATATAATGAAAATATCAAATATTTGCATATAAATATAATCTTAAAGTTCAGTCTATTTAATGTT CAATGAAATATTTCTGCCTGTATAATCTTTAAAGATGTTGAACATATATTCACATTAAATATGATTATGTACTTGT TACAAGGATAAGGTTATATATGAATAAAGTTACAAAAACAGCTATTGTCCCCAGAGCCTCTTCAGCCATCTATT TGGGAGCAAACAATTTCATTCCAACTCATAACCCCAGCATATAAATC
CGI6	AGCAATCAAAAAAGACAAGCAACTTTCAGCAGAAGGCTCTATCCTTGAAGCTGGATGGTAGCTGTAGATAAAC ATTTCTTTTTGAAGGCAAAAAAATAAAATATTCTGTAAGTCCTCCACAGTTCTGCCAAGACAGCTTAGATCTGG

	AATGACAAGGTCATTGTCACTTCAGCTTTATACAGGCACTCTCTATCAGAATGTTGTTTATTACTGCCCAGAGT TTTGTTGGTATTCATCTACCATTTTTTTGACAAGGCAAAACATTACAGAAATTACAATGCTTAGAAAAAATCCCA GAGCCAGGCAAATGGCAGCAGGGTATTTATTTTTTATTTCTGGTGAGCTTATAGTTACACAGCCATCACATGC AAAAACCAATAAGGAAACCTGTGATTTTCAGCTCTACATCACCCTGCAAATCTCTGTCACTTCTAATATAAAAAT AGGGAGAAATGATGGAGCTTATATTCATTGGTGTAAAAAATATGCTTAATAGCACCATTTCTATGAGTTACCCT GATGTTGTAATTGCATGTATAGAACATAAGGTGTCTCTGGAAGCATTCAGGGCAATTGAGGCAGGTAACAAAT TTCCATATTAACTATTATTAATAGAACTCATTAATTGTTTTATTAATTAGTACCCCTCCAGTGTTCTGGAGGGGA TATTCATATTTTTTAAGAGTGACTATTTATGAAAACTTGCATAACAAAGGGAATTGTGACATTGATTCACCAGAG ATATTTCTGCTGGTTTGCTCTCTCATTAGAATTTAACACTAA
CGI10	AGATGTACTTGATCTCAATAATTTGTAACCACAAAATATTTGTTATGGTGCAAAAATAACACATTTAATTTATTGA TTATAAAGGGCTTTAATTTTTGGCCCTTTTATTTTTGGTGTTATGTTTTTAAATTGTCTATAAGTGCCAAAAATTA CATGTTTTGTCTTCTGTTTTTGTTGTTTTAATGTAAATTTTGACCATTTGGTCCACTTTTTTCTGCTGTCATTGTC ACTTCAGCTTTATACAGGCACTCTCTATCAGAATGTTGTTTATTACTGCCCAGAGTTTTGTTGGTATTCATCTAC CATTTTTTTGACAAGGCAAAACATTACAGAAATTACAATGCTTAGAAAAAATCCCAGAGCCAGGCAAATGGCAG CAGGGTATTTATTTTTTATTTCTGGTGAGCTTATAGTTACACAGCCATCACTTGAATGGATATTATCCATATAGT GAATTTGTTGATGATGAATTCATCTGTGCTAAAAATGTTAGTTTAATAAAATATTGAAAGTGACCTGTAATAACA GTTGTTGTTGATTGAGAACAAATAAGTTTATGTGAAAAATATATAAATACATTAGCTGGTCTTGTGTGTCATTTT ATTTTTTTTTGTTGCTAACACAGGGATATGAACAATAA

Table S3 qPCR Primers

Name	Used for	sequence
FL30	RMCE upstream s	TCTTGGAAGAGAAACTCTTAGGG
FL40	RMCE downstr. as	TGTATACAGATCTACCAACATTACGA
FM83	Evx2 s	CGCAGCCCATCATTAAGAC
FM84	Evx2 as	CGGACAAACTGGAGAACCTC
FM81	HoxA9 s	AAGAAGGAAAAGGGGAATGG
FM82	HoxA9 as	TCACCTCGCCTAGTTTCTGG
Hprt s1	Hprt s	CCAAGACGACCGCATGAGAG
Hprt as1	Hprt as	CAACGGAGTGATTGCGCATT
FM85	Gapdh s	CTCTGCTCCTCCCTGTTCC
FM86	Gapdh as	TCCCTAGACCCGTACAGTGC
m1_RT5_as_short	CGI1 5' as	GACAGCCTGGTTACCCTAGC
m1_RT3_s_short	CGI1 3' s	GGGAAGGCCTAAGACATAGC
ML1_5_s	CGI1 5' endog. s	GCTAGGGTAACCAGGCTGTC
ML1_3_as	CGI1 3' endog. as	GCCCTGCTTTGCTATGTCTT
m3_RT5_as_short	CGI3 5' as	CCTTGCCCTGGGATAGAAGAC
m3_RT3_s_short	CGI3 3' s	GAAGCGTAGAGACGCGTTG
ML3_RT5_f	CGI3 5' endog. s	CCCCCTATTAACTGCACCAA
ML3_RT3_r	CGI3 3' endog. as	GCGTCTCTACGCTTCTGCAT
ML3_RT5_r	CGI3 5' endog. as	TCCTGAGCTCTCCAGTCTTTG
ML3_RT3_f	CGI3 3' endog. s	GAGATTCCGTGGAAGAGCAT
ML11_RT5_s	CGI11 5' endog. s	GTGGTGGTGGTGGTAGTGGT
ML11_RT5_as	CGI11 5' endog. as	TCAATATCCCCGCTTCAATT
ML11_RT3_s	CGI11 3' endog.s	GAAGAGAGGCGCAGAGTG
ML11_RT3_as	CGI11 3' endog. as	AGCCTCGGTTCTCCAGCTC
m11_RT3_s_short	CGI11 3' s	AAGAGAGGCGCAGAGTGG
m11_RT5_as_short	CGI11 5' as	GATGCTCCTAGCGCTCTG
ML13_RT5_s	CGI13 5' endog. s	AGGGGACATCAGTGTGCATC

ML13_RT5_as	CGI13 5‘	GTGCAGCTGCCCAATGTG
ML13_RT3_s	CGI13 3‘ endog. s	AAGCCAGTCAGGAGGTTCG
ML13_RT3_as	CGI13 3‘ endog. as	AGCCGCAGGTTCAAGAGTC
m13_RT3_s_short	CGI13 3‘ s	GCCAAAGACCGGCGAGT
ML14_RT5_s	CGI14 5‘ endog. s	TGCAGATGCTAATCTCCACTG
ML14_RT5_as	CGI14 5‘ endog. as	GAAAACCGACGTCATCCAG
ML14_RT3_s	CGI14 3‘ endog. s	CGCAATTGAGGAGAAGAAGG
ML14_RT3_as	CGI14 3‘ endog. as	AATGGGGAGAGTCCACACTG
m14_RT_3_s_short	CGI14 3‘ s	AGTGTGGACTCTCCCCATTC
m14_RT_5_as_short	CGI14 5‘ as	GCATCGCAGTGGAGATTAGC
ML20_RT5_s	CGI20 5‘ endog. s	AGGCTACTTCCCATCCTGGT
ML20_RT5_as	CGI20 5‘ endog. as	GCGGCAGACAGACTGAGAAT
ML20_RT3_s	CGI20 3‘ endog. s	GCCCGACCGGGTAAGTAG
ML20_RT3_as	CGI20 3‘ endog. as	CACGGGGAAGTTCTTGCAG
m20_RT5_as_short	CGI20 5‘ as	CTTTTCTGGGTCAAACCCTG
ML10_RT3_r	CGI10 3‘ endog. as	CTTCCCTCCTTGCCTTCCT
ML10_RT5_f	CGI10 5‘ endog. s	TCAACCTAGCTGTCCAATTCC
ML10_RT5_r	CGI10 5‘ endog. as	GTACTGTCCTGGCGACGTG
ML10_RT3_f	CGI10 3‘ endog. s	CACAGCTCAACGAATTGGAG
m10_RT3_s_short	CGI10 3‘ s	AGGAAGGCAAGGAGGGAAG
m10_RT5_as_short	CGI10 5‘ as	GGTTTCCCAGGAATTGGA
ML6_RT3_fwd	CGI6 3‘ endog. s	GTCCCGCGACCACAAAGT
ML6_RT3_rev	CGI6 3‘ endog. as	AAAATGCCTTCTTTTCCTTGC
ML6_RT5_fwd	CGI6 5‘ endog. s	TGCCTAGGTCAGCTATTCAACA
ML6_RT5_rev	CGI6 5‘ endog. r	GGGCTGGGAGCAATTACAG
m6_RT_3_s_short	CGI6 3‘ s	CTCTGGACTACCAGCAAGGA
m6_RT5_as_short	CGI6 5‘ as	CCAAGGTGGTGAAGTGTAAGAG
ML12_RT3_r	CGI12 3‘ endog. as	ATTGTTCAGCGCCGGTTTAT
ML12_RT5_f	CGI12 5‘ endog. s	CTACCCTGACTCACCTTGAGC
ML12_RT5_r	CGI12 5‘ endog. as	CGCTCTGGACAAGGACGA
ML12_RT3_f	CGI12 3‘ endog. s	GAGGGGGAGGGAATGCAG
m12_RT3_s_short	CGI12 3‘ s	GCGCTGAACAATGAGTCCTAAC
m12_RT5_as_short	CGI12 5‘ as	CTGGACCGGTATCTGCTCA
Utf1_CGI_RT3_s	Utf1 CGI 3‘ s	GGACTTGCGCCAATAAAGC
Utf1noP_CGI_RT5_as	Utf1 CGI 5‘ as	GGACGAAGCAGCATGGATC
Utf1_enh_RT5_as	Utf1 CGI end. 3‘ as	AGTCCGAATTCATTCACAGGA
Utf1e_end_RT3_s	Utf1 endog. 3‘ s	TGACTTGTGGTGTGGACCTC
Utf1e_end_flank_as	Utf1 endog. 3‘ as	CATTTGCAGAGTGGGCTCAT
Utf1e_RT3_s_new	Utf1 enhancer 3‘ s	CAGATGGTCCCAAGGTCACA
Utf1e_RT5_as	Utf1 enhancer 5‘ as	AGGCCCCAGCCTCTCTCTAT
Utf1p_RT3_s	Utf1 promoter 3‘ s	AGGAGCCCTCCTCTCTGGT
Utf1p_RT5_as	Utf1 promoter 5‘ as	TGCCCTGCTACACAAAACAG
Utf1p_upstr_RT	Utf1 endog. 5‘ s	CAGATGGTTGTGAAACACCAG
6coli_RT3_s	cCGI6 3‘ s	TGCTGGTCGGCTCTCTCATC
6coli_RT5_as	cCGI6 5‘ as	GTGCTCGTCTTCGTTGATTGC
3coli_RT3_s	cCGI3 3‘ s	GATTTCATTCCCGCGCATAAC
3coli_RT5_as	cCGI3 5‘ as	GCTGTGTCCGCGTTTTCAAC
10coli_RT3_s	cCGI10 3‘ s	ATCGCGTGGTCTTGTGTGT

10coli_RT5_as	cCGI10 5‘ as	TTACGGCCCTTTATCGTCAAC
10_left_RT3_s	10.1 3‘ s	CACTCTGCTGCAAGCTTATC
10_right_RT5_as	10.2 5‘ as	TAGAGCCCACGGTGGTGATG
11_left_RT3_s	11.1 3‘ s	GCCTCGCCTCCTCTTGATTC
11_right_RT5_as	11.2 5‘ as	CATAGGCGAATCAAGAGGAG
20_left_RT3_s	20.1 3‘ s	CCAGACTCGCCTCGATCTCC
20_right_RT5_as	20.2 5‘ as	GGAGATCGAGGCGAGTCTG
m3_left_RT3_s	3.1 3‘ s	CCGGGAGACTCGCTGTTT
m3_right_RT5_as	3.1 5‘ as	GAAACAGCGAGTCTCCCG
Nanog s	Nanog promoter	CCCAGGGAGGTTGAGAGTTC
Nanog as	Nanog promoter	AGCCGCCAAGTTCACAAAG
Utf1_TFB_s	Utf1 enhancer TFB	GGAGGGCTTAGGTGCAGGTAG
Utf1_TFB_as	Utf1 enhancer TFB	TCCTCAGGACTTCCCTTAGCC

Table S4 Bisulfite Primers

Name	**Used for**	**Sequence**
MamBis1	RMCE upstream	ATTAAATAAAATGAAAGTTTTGGAAGAG
FL277	RMCE downstr.	ATATAAAATAATAACAATATATACAAATCTACCAAC
10_WT_bis_as_new	CGI10 5‘	CTAAATAACCCCCAAAATCAAAAC
10_WT_bis_s_new	CGI10 3‘	TATTTAGGGGATTTGGTATTTAATT
3left_bis_as	3.1 5‘	ACTTACTAAACAAAATTCCCTTTCC
3left_bis_s	3.1 3‘	TAATTGTATTAAGGTTTGTTTTGGG
3right_bis_as	3.2 5‘	AATAAACTAAAAACCCCAAAC
3right_bis_s	3.2 3‘	TTTTTTTTATGATATTTTGGATGTGG
6coli_inv_bis_as	cCGI6 3‘	TCTATAAATTACCCTAATATTATAC
6coli_inv_bis_s	cCGI6 5‘	AAAAATAAATATTTTGTT
6WT_bis_s	CGI6 5‘	GGAGAGGTAGTAAGAAAGGTTATTAA
6WT_bis_i_as	CGI6 3‘	AACCTTTCTTACTACCTCTCCAAAC
m3_bis_WT_as	CGI3 5‘	AAAACCACATCCAAAATATCATAAAA
m3_bis_WT_s	CGI3 3‘	GGAAAGGGAATTTTGTTTAGTAAGTA
3_coli_bis_s_new	cCGI3 3‘	GATGTTTAATGAAATATTTTTGTTTGTAT
m3_bis_coli_as	cCGI3 5‘	CAAACAAAAATATTTCATTAAACATC
10_WT_bis_as_new	CGI10 5‘	CTAAATAACCCCCAAAATCAAAAC
10_WT_bis_s_new	CGI10 3‘	TATTTAGGGGATTTGGTATTTAATT
10coli_bis_as4	cCGI10 5‘	AAACTCTAAACAATAATAAACAACATTCTAATAAA
m10_bis_coli_s_2	cCGI10 3‘	TTTATTAGAATGTTGTTTATTATTGTTTAGAGTTT
14_bis_as2	CGI14	AAAATCCACACTAAACAACCCC
14_bis_s_end	CGI14	TAGAGTGAGATTTTGTTTTAATAAAAA

3.3 Additional results

3.3.1 Utf1 promoter region is required to fully inhibit PRC2 recruitment

Since we observed that loss of OCT4 / SOX2 binding at the *Utf1* locus could prevent acquisition of H3K27me3, we wanted to investigate whether this inhibitory effect required the presence of the *Utf1* promoter. To this end, we inserted the *Utf1* locus without its promoter and tested if this construct would still be able to prevent H3K27me3 acquisition (Figure 22). Indeed, the observed levels of H3K27me3 at this construct were significantly reduced in comparison to the *Utf1* CGI inserted without the enhancer (Figure 22). Nevertheless, the H3K27me3 levels were still significantly above background, indicating that the OCT4 / SOX2 binding to the enhancer is not sufficient to completely block H3K27me3 acquisition, but instead requires the presence of the promoter and possibly the transcriptional machinery. However, it has to be noted that the qPCR amplicons at the 3' end of the inserted constructs are not the same due to presence or absence of the enhancer. It is therefore possible that the observed difference in H3K27me3 levels at the 3' end is partially caused by the fact that different amplicons were analyzed. To exclude this possibility one could use a construct that contains the enhancer with the OCT4 / SOX2 binding site deleted instead of the complete absence of the enhancer.

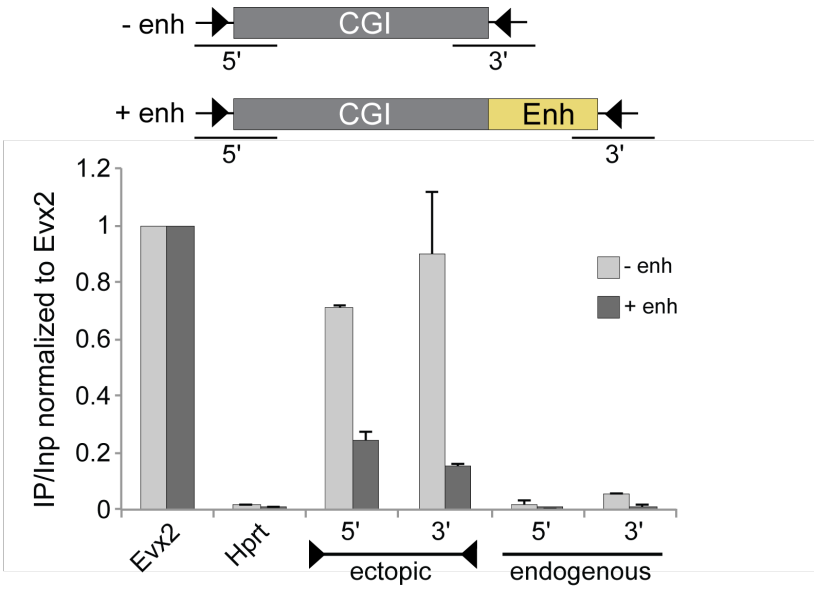

Figure 22 Utf1 enhancer is not sufficient to fully prevent H3K27me3 recruitment. Comparison between H3K27me3 levels at the Utf1 CGI with and without the enhancer. TF binding at the enhancer has a significant effect on H3K27me3 recruitment, even in the absence of the promoter. All data normalized to input and shown relative to Evx2 levels.

3.3.2 Inhibition of H3K27me3 recruitment is independent of promoter directionality

In order to test whether transcription through the CGI is required for inhibition of H3K27me3 recruitment, we inversed the promoter sequence relative to the *Utf1* CGI and enhancer (Figure 23). Insertion of this construct and subsequent measurement of H3K27me3 levels revealed that similar to the wild-type locus H3K27me3 deposition was still inhibited. These data suggest that the directionality of the promoter sequence does not play a major role in preventing the deposition of H3K27me3 in ES cells. We therefore conclude that transcription through the CGI is not required to prevent deposition of H3K27me3. Notably, however, it has previously been shown that active promoters in ES cells often produce short anti-sense transcripts peaking at around 250 nt upstream of the TSS (Seila et al. 2008). It is therefore possible that the presence such anti-sense transcripts originating from the *Utf1* promoter might play a role in preventing H3K27me3.

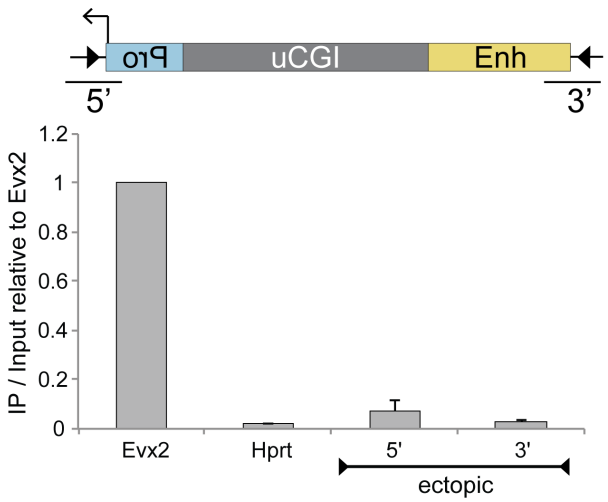

Figure 23 Directionality of Utf1f promoter sequence does not influence H3K27me3 recruitment in ES cells. The promoter sequence of the Utf1 locus was inverted to prevent transcription through the CGI. H3K27me3 levels at the ectopic were subsequently measured by ChIP. Similar to the wild-type construct, H3K27me3 is not recruited, indicating that promoter directionality does not play a significant role in preventing H3K27me3 deposition.

3.3.3 DNA methylation and H3K27me3 mark different sets of promoters

The observation that DNA methylation and H3K27me3 are mutually exclusive throughout the genome in stem cells prompted us to further investigate the sequence composition of the underlying DNA sequences. To this end, we analyzed the levels of H3K27me3 and DNA methylation at all non-overlapping promoters (n=19'517) in the mouse genome and compared them to RNA Pol II levels in order to define active and inactive promoters (Figure 24). While active promoters are virtually devoid of DNA methylation and the H3K27me3 mark, inactive promoters are either marked by H3K27me3 or DNA methylation (Figure 24A). Next, we wanted to know whether there is any significant difference in the sequence composition between the H3K27me3 modified and DNA methylated promoters. Remarkably, when plotted against CpG density, the DNA-methylated promoters are exclusively CpG-poor, while the H3K27me3 levels scale with increasing CpG density (Figure 24B). This demonstrates that high CpG density seems to have an inhibitory effect on DNA methylation as observed previously (Lienert et al. 2011), while at the same time favoring Polycomb binding.

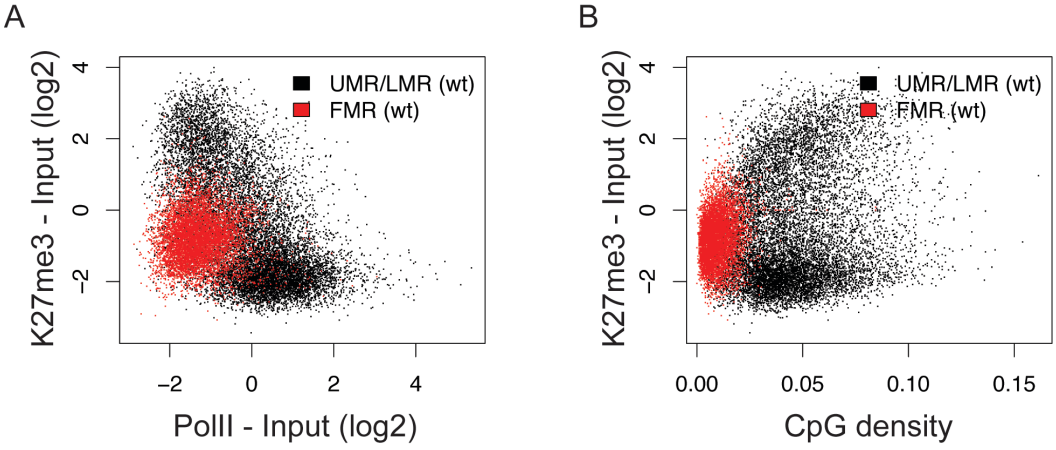

Figure 24 DNA methylation and H3K27me3 occupy separate sets of promoters in the mouse ES cells. (A) Genome-wide comparison of Pol II, H3K27me3 and DNA methylation levels at all non-overlapping promoters (2kb window around transcription start site) in mouse ESCs. Pol II negative promoters are marked by either H3K27me3 or DNA methylation. (B) Same as (A), but instead of Pol II, chromatin marks are shown against CpG density. This shows that H3K27me3 marks promoters of high CpG density, while DNA methylation is found at low CpG density regions. UMR: unmethylated region; LMR: Low methylated regions; FMR: Fully methylated region.

Taken together, these genome-wide data confirm our observations at the single insertions, suggesting that the antagonistic relationship between DNA methylation and H3K27me3 are conserved throughout the genome. Furthermore,

CpG density appears to have an important role in defining the epigenetic state of a promoter.

CHAPTER 4

4. CONCLUSIONS AND OUTLOOK

Recent advances in DNA sequencing technologies have enabled the genome-wide analysis of epigenetic marks and have revealed striking correlations between chromatin modifications and gene expression patterns (Suzuki and Bird 2008, Hawkins et al. 2010, Meissner 2010, Zhou et al. 2011, Zhu et al. 2013). Therefore, epigenetic modifications have been proposed to directly influence transcriptional output and maintain expression patterns established by TFs (Kouzarides 2007). However, despite their important role during the development of an organism, it remains unclear how the enzymes that facilitate epigenetic modifications are targeted to specific loci in a tissue-specific manner.

Here, we employed a reductionist approach involving repeated genomic insertions coupled with iterative dissection and manipulation of short DNA sequences to address this fundamental question. We focused on the recruitment of PcG complexes, as this repressive system is highly relevant during development. Using mouse embryonic stem cells as a model system, we demonstrate that DNA sequences as short as 220 bp can be sufficient to recruit the PRC2 complex and establish an H3K27me3 domain. Furthermore, by systematically mutating these elements we show that unmethylated CpG-dense sequences are sufficient to facilitate this recruitment. Finally, we demonstrate that tissue-specific establishment of a H3K27me3 domain is abrogated by the activity of active adjacent cis-regulatory regions. Furthermore, H3K27me3 acquisition can be reestablished when activating TFBS are mutated.

Taken together, we propose a model whereby recruitment of PRC2 and subsequent establishment of an H3K27me3 domain represents a default state of transcriptionally inactive and unmethylated CpG-dense genomic regions (Figure 25).

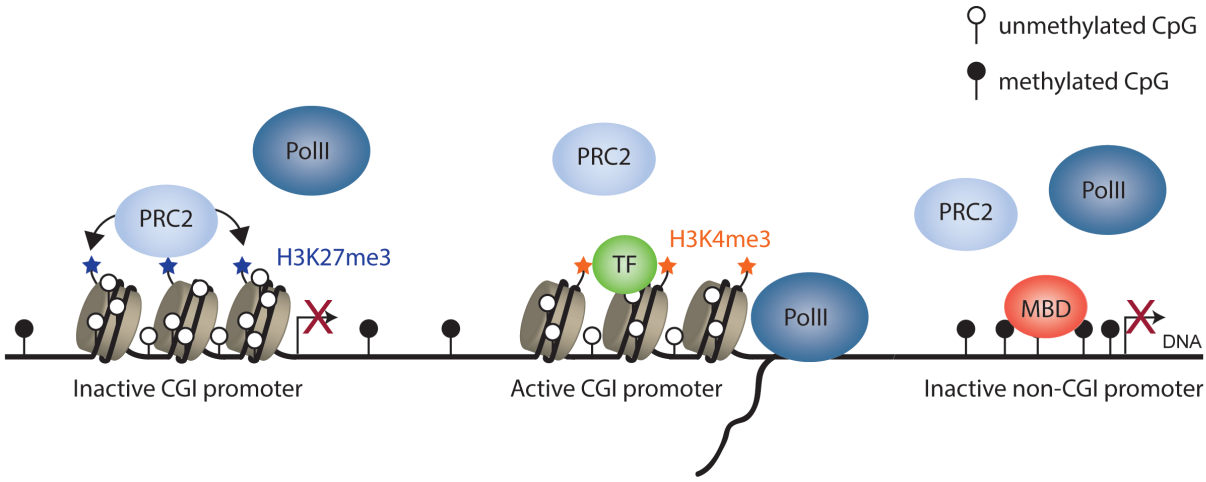

Figure 25 Schematic representations of the PRC2 recruitment mechanisms suggested here. The data presented in this study argue for a model whereby H3K27me3 marks transcriptionally inactive promoter regions if they are unmethylated (left). On the other hand, regions that are DNA methylated block the binding of PRC2 and therefore prevent deposition of the H3K27me3 mark (right). Active promoters are enriched in the H3K4me3 mark, which together with TF binding and RNA production inhibits the binding and activity of PRC2.

4.1 Crosstalk between H3K27me3 and DNA methylation

In this study we demonstrate an important interplay between two repressive epigenetic marks, namely H3K27me3 and DNA methylation. Similar observations have been made between DNA methylation and chromatin states, suggesting that DNA methylation plays an important role in the regulation of epigenetic modifications (Rose and Klose 2014). For example, a recent study investigating the conservation of non-methylated regions (NMRs) in seven vertebrate species showed that absence of DNA methylation is a highly conserved feature of vertebrate promoters (Long et al. 2013). Interestingly, while roughly half of all promoters in humans and mice are associated with CGIs, other vertebrates such as the frog or zebrafish display a much smaller overlap of only around 10%. Accordingly, the correlation between H3K27me3 and CGIs observed in humans and mice is not seen in the frog and zebrafish (van Heeringen et al. 2013). Instead, H3K27me3 domains in these species overlap largely with NMRs. This suggests that the observed correlation between CpG density and H3K27me3 in mammals could be simply due to the strong overlap with unmethylated regions.

To test this experimentally and expand on the role of CpG dinucleotides in recruiting PRC2 in mammals, further insertion experiments using fragments with various CpG densities would be necessary. Additionally, the relationship between DNA methylation and H3K27me3 could be more extensively studied by inserting

these fragments into *Dnmt1/3a/3b* triple-knockout cells. Such a setup would bypass the need for chemical treatment with methylation inhibiting factors and thus allow for direct comparison of H3K27me3 recruitment in the absence of DNA methylation in a more efficient manner.

4.2 The role of transcription factors in PcG recruitment

In light of the different models that have been suggested for Polycomb recruitment, the seemingly simple mechanism proposed here appears controversially different from other models. For example, in *D. melanogaster*, where most of the early work on Polycomb recruitment was performed, established models suggest that recruitment of PcG complexes to PREs is facilitated by a direct interaction with sequence specific TFs [reviewed in (Muller and Kassis 2006)]. Furthermore, the *D. melanogaster* genome does not contain CGIs and lacks global DNA methylation. Hence, at first glance, our proposed model for PcG recruitment in mammals seems to contradict those established in the fly. However, these seemingly different modes of recruitment are not mutually exclusive and have many factors in common. First, similar to what has been reported in *D. melanogaster,* we observe a direct correlation between TF occupancy and H3K27me3 deposition. However, we propose that the link between the two is not based on direct physical interaction. Instead, our data suggest that H3K27me3 recruitment is secondary to TF activity and thus responsive to the transcriptional state of a locus (Figure 13). Considering the limited evidence for direct interaction of PcG proteins with TFs in *D. melanogaster*, a similar mechanism might be at work in flies. This is supported by the fact that transcriptional activity opposes PcG occupancy in *D.melanogaster* (Cavalli and Paro 1998, Schmitt et al. 2005). In particular, it was demonstrated that transcription through a *Drosophila* PRE could induce a switch from a PcG mediated “off” state to an active state maintained by TrxG proteins, indicating that PcG occupancy is sensitive to the transcriptional state (Cavalli and Paro 1998, Schmitt et al. 2005). Similar observations have also been made in plants deposition of the H3K27me3 mark at the *FLC* locus during vernalization is dependent on the transcriptional state of the locus (Buzas et al. 2011).

Thus, we propose that instead of being directly recruited via sequence specific TFs, PcG complexes scan along the genome and sense the chromatin state

as well as transcriptional activity of a locus. One mechanism through which this scanning for inactive genes could be facilitated might involve the interaction of PRC2 with short non-coding RNAs. Recent studies investigating these interactions revealed promiscuous binding of RNAs to PRC2 (Zhao et al. 2010, Davidovich et al. 2013, Kaneko et al. 2013). Interestingly, Kaneko et al. observed widespread interactions between PRC2 and short ncRNAs originating from active gene promoters and propose that these interactions are a means for PRC2 to detect transcriptional activity (Kaneko et al. 2013). Furthermore, the authors propose that interaction with these RNAs could negatively affect enzymatic activity and thus lead to decreased H3K27me3 deposition at active loci. Such a mechanism would be compatible with the model proposed here whereby transcriptional activity inhibits H3K27me3 recruitment.

4.3 H3K27me3 as a means to reinforce repressive chromatin

The model for Polycomb recruitment that we propose here raises several questions in terms of functional relevance of this chromatin mark. If it is not necessary to induce gene repression, then what role does it play in gene regulation? Depletion of PcG proteins in stem cells has minimal consequences regarding transcriptional activity (Chamberlain et al. 2008, Leeb et al. 2010). However, it is also clear that PcG proteins are key regulators of embryonic development and differentiation of pluripotent stem cells. This is made evident in the absence of canonical PcG complexes, which results in early embryonic lethality caused by defects in cellular differentiation (Faust et al. 1995, O'Carroll et al. 2001, Suzuki et al. 2002, Voncken et al. 2003, Pasini et al. 2004, Boyer et al. 2006, Pasini et al. 2007).

Based on our observation that H3K27me3 deposition may represent a default state of inactive CpG-rich chromatin, we propose that PcG proteins serve to maintain a repressed state and protect the target genes from aberrant transcription. This could be especially important during differentiation when tight regulation of developmental regulators is essential. It is possible that basal transcriptional output even in the absence of specific activators could produce enough protein to disturb the overall balance and result in aberrant functionality of the cell. Thus, targeting of promoters below a certain basal level of transcriptional activity could be a mechanism to reinforce the repressed state and ensure dampen transcriptional

noise. Such a mechanism would allow recruitment of basal transcriptional components while concomitantly dampening aberrant transcriptional output. Conversely, to overcome the repressive state maintained by PcG, activating TFs would cause a switch from a H3K27me3 repressed state into an H3K4me3 marked active state. A similar mechanism involving H3K4me3 might then be used to prevent H3K27me3 from aberrantly repressing active genes (Schmitges et al. 2011, Voigt et al. 2012).

Loss of PRC2 activity in stem cells has only very mild effects on gene expression, suggesting that this cell type might not be ideal for studying the mechanisms underlying H3K27me3 mediated repression. Instead, it may be more informative to study these aspects in differentiated cells, where loss of H3K27me3 could have more drastic consequences on gene expression. To achieve this, one could generate a conditional *Eed*-knockout stem cell line to study loss of PRC2 in a differentiated type, such as neural progenitors or terminal neurons.

In summary, this study provides novel insights into the targeting mechanisms of the PRC2 complex and H3K27me3 deposition. Moreover, it emphasizes the crosstalk between epigenetic marks and broadens our understanding of how epigenetic modifications influence gene expression.

CHAPTER 5

5. ACKNOWLEDGEMENTS

First, I would like to thank Dirk for letting me conduct my PhD thesis in his group. I would also like to thank him for his continuous support throughout my time in the group, especially his motivational encouragement during the times when things didn't go as planned.

Furthermore, I would like to thank all the current and past members of the lab who have contributed to a great atmosphere both in and outside of the lab. I especially thank Paul for all the lively discussions and helpful input for my project.

I would also like to thank Kristina and my family for their continuous support.

Finally, I would like to thank the members of my thesis committee, Renato Paro and Marc Bühler, for taking their time and providing helpful feedback during committee meetings and for evaluating my thesis .

CHAPTER 6

6. BIBLIOGRAPHY

Abed, J. A. and R. S. Jones (2012). "H3K36me3 key to Polycomb-mediated gene silencing in lineage specification." Nat Struct Mol Biol **19**(12): 1214-1215.

Alberts, B., J. H. Wilson and T. Hunt (2008). Molecular biology of the cell. New York, Garland Science.

Allfrey, V. G. and A. E. Mirsky (1964). "Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis." Science **144**(3618): 559.

Aloia, L., B. Di Stefano and L. Di Croce (2013). "Polycomb complexes in stem cells and embryonic development." Development **140**(12): 2525-2534.

Arnold, P., A. Scholer, M. Pachkov, P. J. Balwierz, H. Jorgensen, M. B. Stadler, E. van Nimwegen and D. Schubeler (2013). "Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting." Genome research **23**(1): 60-73.

Bachmann, I. M., O. J. Halvorsen, K. Collett, I. M. Stefansson, O. Straume, S. A. Haukaas, H. B. Salvesen, A. P. Otte and L. A. Akslen (2006). "EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast." J Clin Oncol **24**(2): 268-273.

Ball, M. P., J. B. Li, Y. Gao, J. H. Lee, E. M. LeProust, I. H. Park, B. Xie, G. Q. Daley and G. M. Church (2009). "Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells." Nat Biotechnol **27**(4): 361-368.

Ballare, C., M. Lange, A. Lapinaite, G. M. Martin, L. Morey, G. Pascual, R. Liefke, B. Simon, Y. Shi, O. Gozani, T. Carlomagno, S. A. Benitah and L. Di Croce (2012). "Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity." Nat Struct Mol Biol **19**(12): 1257-1265.

Bannister, A. J. and T. Kouzarides (2011). "Regulation of chromatin by histone modifications." Cell Res **21**(3): 381-395.

Bantignies, F. and G. Cavalli (2011). "Polycomb group proteins: repression in 3D." Trends Genet **27**(11): 454-464.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-837.

Bejarano, F. and A. Busturia (2004). "Function of the Trithorax-like gene during Drosophila development." Dev Biol **268**(2): 327-341.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber and E. S. Lander (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-326.

Bestor, T. H. and V. M. Ingram (1983). "Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA." Proc Natl Acad Sci U S A **80**(18): 5559-5563.

Bibel, M., J. Richter, E. Lacroix and Y. A. Barde (2007). "Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells." Nat Protoc **2**(5): 1034-1043.

Bird, A. P. (1980). "DNA methylation and the frequency of CpG in animal DNA." Nucleic Acids Res **8**(7): 1499-1504.

Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." Nature **321**(6067): 209-213.

Bird, A. P., M. H. Taggart and B. A. Smith (1979). "Methylated and unmethylated DNA compartments in the sea urchin genome." Cell **17**(4): 889-901.

Black, J. C., C. Van Rechem and J. R. Whetstine (2012). "Histone lysine methylation dynamics: establishment, regulation, and biological impact." Mol Cell **48**(4): 491-507.

Bock, C., S. Reither, T. Mikeska, M. Paulsen, J. Walter and T. Lengauer (2005). "BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing." Bioinformatics **21**(21): 4067-4068.

Boyer, L. A., K. Plath, J. Zeitlinger, T. Brambrink, L. A. Medeiros, T. I. Lee, S. S. Levine, M. Wernig, A. Tajonar, M. K. Ray, G. W. Bell, A. P. Otte, M. Vidal, D. K. Gifford, R. A. Young and R. Jaenisch (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-353.

Bracken, A. P., N. Dietrich, D. Pasini, K. H. Hansen and K. Helin (2006). "Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions." Genes Dev **20**(9): 1123-1136.

Brent, R. and M. Ptashne (1985). "A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor." Cell **43**(3 Pt 2): 729-736.

Brien, G. L., G. Gambero, D. J. O'Connell, E. Jerman, S. A. Turner, C. M. Egan, E. J. Dunne, M. C. Jurgens, K. Wynne, L. Piao, A. J. Lohan, N. Ferguson, X. Shi, K. M. Sinha, B. J. Loftus, G. Cagney and A. P. Bracken (2012). "Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation." Nat Struct Mol Biol **19**(12): 1273-1281.

Brinkman, A. B., H. Gu, S. J. Bartels, Y. Zhang, F. Matarese, F. Simmer, H. Marks, C. Bock, A. Gnirke, A. Meissner and H. G. Stunnenberg (2012). "Sequential ChIP-

bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk." Genome research **22**(6): 1128-1138.

Brockdorff, N., A. Ashworth, G. F. Kay, V. M. McCabe, D. P. Norris, P. J. Cooper, S. Swift and S. Rastan (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**(3): 515-526.

Brown, C. J., B. D. Hendrich, J. L. Rupert, R. G. Lafreniere, Y. Xing, J. Lawrence and H. F. Willard (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**(3): 527-542.

Brown, J. L., C. Fritsch, J. Mueller and J. A. Kassis (2003). "The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing." Development **130**(2): 285-294.

Brown, J. L., D. Mucci, M. Whiteley, M. L. Dirksen and J. A. Kassis (1998). "The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1." Mol Cell **1**(7): 1057-1064.

Brownell, J. E., J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth and C. D. Allis (1996). "Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation." Cell **84**(6): 843-851.

Burger, L., D. Gaidatzis, D. Schubeler and M. B. Stadler (2013). "Identification of active regulatory regions from DNA methylation data." Nucleic Acids Res **41**(16): e155.

Busturia, A., J. Casanova, E. Sanchez-Herrero and G. Morata (1989). "Structure and function of the bithorax complex genes of Drosophila." Ciba Found Symp **144**: 227-238; discussion 239-242, 290-225.

Buzas, D. M., M. Robertson, E. J. Finnegan and C. A. Helliwell (2011). "Transcription-dependence of histone H3 lysine 27 trimethylation at the Arabidopsis polycomb target gene FLC." Plant J **65**(6): 872-881.

Cao, R., Y. Tsukada and Y. Zhang (2005). "Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing." Mol Cell **20**(6): 845-854.

Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones and Y. Zhang (2002). "Role of histone H3 lysine 27 methylation in Polycomb-group silencing." Science **298**(5595): 1039-1043.

Cao, R. and Y. Zhang (2004). "The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3." Curr Opin Genet Dev **14**(2): 155-164.

Cao, R. and Y. Zhang (2004). "SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex." Mol Cell **15**(1): 57-67.

Carey, M., Y. S. Lin, M. R. Green and M. Ptashne (1990). "A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives." Nature **345**(6273): 361-364.

Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn and J. L. Workman (2005). "Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription." Cell **123**(4): 581-592.

Casanova, M., T. Preissner, A. Cerase, R. Poot, D. Yamada, X. Li, R. Appanah, K. Bezstarosti, J. Demmers, H. Koseki and N. Brockdorff (2011). "Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells." Development **138**(8): 1471-1482.

Cavalli, G. and R. Paro (1998). "The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis." Cell **93**(4): 505-518.

Cavalli, G. and R. Paro (1999). "Epigenetic inheritance of active chromatin after removal of the main transactivator." Science **286**(5441): 955-958.

Chamberlain, S. J., D. Yee and T. Magnuson (2008). "Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency." Stem cells **26**(6): 1496-1505.

Chan, C. S., L. Rastelli and V. Pirrotta (1994). "A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression." The EMBO journal **13**(11): 2553-2564.

Chen, S., L. R. Bohrer, A. N. Rai, Y. Pan, L. Gan, X. Zhou, A. Bagchi, J. A. Simon and H. Huang (2010). "Cyclin-dependent kinases regulate epigenetic gene silencing through phosphorylation of EZH2." Nat Cell Biol **12**(11): 1108-1114.

Chew, J. L., Y. H. Loh, W. Zhang, X. Chen, W. L. Tam, L. S. Yeap, P. Li, Y. S. Ang, B. Lim, P. Robson and H. H. Ng (2005). "Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells." Molecular and cellular biology **25**(14): 6031-6046.

Chiang, A., M. B. O'Connor, R. Paro, J. Simon and W. Bender (1995). "Discrete Polycomb-binding sites in each parasegmental domain of the bithorax complex." Development **121**(6): 1681-1689.

Chinwalla, V., E. P. Jane and P. J. Harte (1995). "The Drosophila trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites." EMBO J **14**(9): 2056-2065.

Christen, B. and M. Bienz (1994). "Imaginal disc silencers from Ultrabithorax: evidence for Polycomb response elements." Mech Dev **48**(3): 255-266.

Clapier, C. R. and B. R. Cairns (2009). "The biology of chromatin remodeling complexes." Annu Rev Biochem **78**: 273-304.

Clark, R. J. and G. Felsenfeld (1971). "Structure of chromatin." Nat New Biol **229**(4): 101-106.

Cohen, N. M., E. Kenigsberg and A. Tanay (2011). "Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection." Cell **145**(5): 773-786.

Collett, K., G. E. Eide, J. Arnes, I. M. Stefansson, J. Eide, A. Braaten, T. Aas, A. P. Otte and L. A. Akslen (2006). "Expression of enhancer of zeste homologue 2 is significantly associated with increased tumor cell proliferation and is a marker of aggressive breast cancer." Clin Cancer Res **12**(4): 1168-1174.

Conaway, R. C. and J. W. Conaway (2011). "Function and regulation of the Mediator complex." Curr Opin Genet Dev **21**(2): 225-230.

Czermin, B., R. Melfi, D. McCabe, V. Seitz, A. Imhof and V. Pirrotta (2002). "Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites." Cell **111**(2): 185-196.

Dahm, R. (2005). "Friedrich Miescher and the discovery of DNA." Dev Biol **278**(2): 274-288.

Davidovich, C., L. Zheng, K. J. Goodrich and T. R. Cech (2013). "Promiscuous RNA binding by Polycomb repressive complex 2." Nature structural & molecular biology **20**(11): 1250-1257.

De Lucia, F., P. Crevillen, A. M. Jones, T. Greb and C. Dean (2008). "A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization." Proc Natl Acad Sci U S A **105**(44): 16831-16836.

de Napoles, M., J. E. Mermoud, R. Wakao, Y. A. Tang, M. Endoh, R. Appanah, T. B. Nesterova, J. Silva, A. P. Otte, M. Vidal, H. Koseki and N. Brockdorff (2004). "Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation." Dev Cell **7**(5): 663-676.

Deaton, A. M. and A. Bird (2011). "CpG islands and the regulation of transcription." Genes Dev **25**(10): 1010-1022.

DeCamillis, M., N. S. Cheng, D. Pierre and H. W. Brock (1992). "The polyhomeotic gene of Drosophila encodes a chromatin protein that shares polytene chromosome-binding sites with Polycomb." Genes Dev **6**(2): 223-232.

Decoville, M., E. Giacomello, M. Leng and D. Locker (2001). "DSP1, an HMG-like protein, is involved in the regulation of homeotic genes." Genetics **157**(1): 237-244.

Dejardin, J., A. Rappailles, O. Cuvier, C. Grimaud, M. Decoville, D. Locker and G. Cavalli (2005). "Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1." Nature **434**(7032): 533-538.

Dhalluin, C., J. E. Carlson, L. Zeng, C. He, A. K. Aggarwal and M. M. Zhou (1999). "Structure and ligand of a histone acetyltransferase bromodomain." Nature **399**(6735): 491-496.

Di Croce, L. and K. Helin (2013). "Transcriptional regulation by Polycomb group proteins." Nature structural & molecular biology **20**(10): 1147-1155.

Dietrich, N., M. Lerdrup, E. Landt, S. Agrawal-Singh, M. Bak, N. Tommerup, J. Rappsilber, E. Sodersten and K. Hansen (2012). "REST-Mediated Recruitment of Polycomb Repressor Complexes in Mammalian Cells." PLoS Genet **8**(3): e1002494.

Dou, Y., T. A. Milne, A. J. Ruthenburg, S. Lee, J. W. Lee, G. L. Verdine, C. D. Allis and R. G. Roeder (2006). "Regulation of MLL1 H3K4 methyltransferase activity by its core components." Nat Struct Mol Biol **13**(8): 713-719.

Enderle, D., C. Beisel, M. B. Stadler, M. Gerstung, P. Athri and R. Paro (2010). "Polycomb preferentially targets stalled promoters of coding and noncoding transcripts." Genome Res.

Ernst, J. and M. Kellis (2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome." Nat Biotechnol **28**(8): 817-825.

Ernst, J., P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis and B. E. Bernstein (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types." Nature **473**(7345): 43-49.

Eskeland, R., M. Leeb, G. R. Grimes, C. Kress, S. Boyle, D. Sproul, N. Gilbert, Y. Fan, A. I. Skoultchi, A. Wutz and W. A. Bickmore (2010). "Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination." Mol Cell **38**(3): 452-464.

Fang, R., A. J. Barbera, Y. Xu, M. Rutenberg, T. Leonor, Q. Bi, F. Lan, P. Mei, G. C. Yuan, C. Lian, J. Peng, D. Cheng, G. Sui, U. B. Kaiser, Y. Shi and Y. G. Shi (2010). "Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation." Mol Cell **39**(2): 222-233.

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

Faust, C., A. Schumacher, B. Holdener and T. Magnuson (1995). "The eed mutation disrupts anterior mesoderm production in mice." Development **121**(2): 273-285.

Felsenfeld, G. (1992). "Chromatin as an essential part of the transcriptional mechanism." Nature **355**(6357): 219-224.

Feng, Y. Q., J. Seibler, R. Alami, A. Eisen, K. A. Westerman, P. Leboulch, S. Fiering and E. E. Bouhassira (1999). "Site-specific chromosomal integration in mammalian

cells: highly efficient CRE recombinase-mediated cassette exchange." J Mol Biol **292**(4): 779-785.

Filion, G. J., J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker and B. van Steensel (2010). "Systematic protein location mapping reveals five principal chromatin types in Drosophila cells." Cell **143**(2): 212-224.

Fitzpatrick, G. V., P. D. Soloway and M. J. Higgins (2002). "Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1." Nat Genet **32**(3): 426-431.

Flemming, W. (1882). "Zellsubstanz, Kern und Zelltheilung."

Gao, Z., J. Zhang, R. Bonasio, F. Strino, A. Sawai, F. Parisi, Y. Kluger and D. Reinberg (2012). "PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes." Mol Cell **45**(3): 344-356.

Giniger, E., S. M. Varnum and M. Ptashne (1985). "Specific DNA binding of GAL4, a positive regulatory protein of yeast." Cell **40**(4): 767-774.

Grau, D. J., B. A. Chapman, J. D. Garlick, M. Borowsky, N. J. Francis and R. E. Kingston (2011). "Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge." Genes Dev **25**(20): 2210-2221.

Hagstrom, K., M. Muller and P. Schedl (1997). "A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex." Genetics **146**(4): 1365-1380.

Hall, I. M., G. D. Shankaranarayana, K. Noma, N. Ayoub, A. Cohen and S. I. Grewal (2002). "Establishment and maintenance of a heterochromatin domain." Science **297**(5590): 2232-2237.

Hawkins, R. D., G. C. Hon, L. K. Lee, Q. Ngo, R. Lister, M. Pelizzola, L. E. Edsall, S. Kuan, Y. Luu, S. Klugman, J. Antosiewicz-Bourget, Z. Ye, C. Espinoza, S. Agarwahl, L. Shen, V. Ruotti, W. Wang, R. Stewart, J. A. Thomson, J. R. Ecker and B. Ren (2010). "Distinct epigenomic landscapes of pluripotent and lineage-committed human cells." Cell Stem Cell **6**(5): 479-491.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis and B. Ren (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.

Hellman, A. and A. Chess (2007). "Gene body-specific methylation on the active X chromosome." Science **315**(5815): 1141-1143.

Henderson, I. R. and S. E. Jacobsen (2007). "Epigenetic inheritance in plants." Nature **447**(7143): 418-424.

Hennig, L. and M. Derkacheva (2009). "Diversity of Polycomb group complexes in plants: same rules, different players?" Trends Genet **25**(9): 414-423.

Herranz, N., D. Pasini, V. M. Diaz, C. Franci, A. Gutierrez, N. Dave, M. Escriva, I. Hernandez-Munoz, L. Di Croce, K. Helin, A. Garcia de Herreros and S. Peiro (2008). "Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor." Mol Cell Biol **28**(15): 4772-4781.

Herz, H. M. and A. Shilatifard (2010). "The JARID2-PRC2 duality." Genes Dev **24**(9): 857-861.

Hewish, D. R. and L. A. Burgoyne (1973). "Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease." Biochem Biophys Res Commun **52**(2): 504-510.

Hochschild, A. and S. L. Dove (1998). "Protein-protein contacts that activate and repress prokaryotic transcription." Cell **92**(5): 597-600.

Hoffman, M. M., J. Ernst, S. P. Wilder, A. Kundaje, R. S. Harris, M. Libbrecht, B. Giardine, P. M. Ellenbogen, J. A. Bilmes, E. Birney, R. C. Hardison, I. Dunham, M. Kellis and W. S. Noble (2013). "Integrative annotation of chromatin elements from ENCODE data." Nucleic Acids Res **41**(2): 827-841.

Holliday, R. and J. E. Pugh (1975). "DNA modification mechanisms and gene activity during development." Science **187**(4173): 226-232.

Hong, L., G. P. Schroth, H. R. Matthews, P. Yau and E. M. Bradbury (1993). "Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA." J Biol Chem **268**(1): 305-314.

Huang, D. H., Y. L. Chang, C. C. Yang, I. C. Pan and B. King (2002). "pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex." Mol Cell Biol **22**(17): 6261-6271.

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

Iguchi-Ariga, S. M. and W. Schaffner (1989). "CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation." Genes Dev **3**(5): 612-619.

Illingworth, R. S., U. Gruenewald-Schneider, S. Webb, A. R. Kerr, K. D. James, D. J. Turner, C. Smith, D. J. Harrison, R. Andrews and A. P. Bird (2010). "Orphan CpG islands identify numerous conserved promoters in the mammalian genome." PLoS Genet **6**(9): e1001134.

Jacob, F. and J. Monod (1961). "Genetic regulatory mechanisms in the synthesis of proteins." J Mol Biol **3**: 318-356.

Joshi, A. A. and K. Struhl (2005). "Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation." Mol Cell **20**(6): 971-978.

Juttermann, R., E. Li and R. Jaenisch (1994). "Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation." Proceedings of the National Academy of Sciences of the United States of America **91**(25): 11797-11801.

Kaneko, S., G. Li, J. Son, C. F. Xu, R. Margueron, T. A. Neubert and D. Reinberg (2010). "Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA." Genes Dev **24**(23): 2615-2620.

Kaneko, S., J. Son, S. S. Shen, D. Reinberg and R. Bonasio (2013). "PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells." Nature structural & molecular biology **20**(11): 1258-1264.

Kanhere, A., K. Viiri, C. C. Araujo, J. Rasaiyaah, R. D. Bouwman, W. A. Whyte, C. F. Pereira, E. Brookes, K. Walker, G. W. Bell, A. Pombo, A. G. Fisher, R. A. Young and R. G. Jenner (2010). "Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2." Mol Cell **38**(5): 675-688.

Keogh, M. C., S. K. Kurdistani, S. A. Morris, S. H. Ahn, V. Podolny, S. R. Collins, M. Schuldiner, K. Chin, T. Punna, N. J. Thompson, C. Boone, A. Emili, J. S. Weissman, T. R. Hughes, B. D. Strahl, M. Grunstein, J. F. Greenblatt, S. Buratowski and N. J. Krogan (2005). "Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex." Cell **123**(4): 593-605.

Ketel, C. S., E. F. Andersen, M. L. Vargas, J. Suh, S. Strome and J. A. Simon (2005). "Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes." Mol Cell Biol **25**(16): 6857-6868.

Kim, T. G., J. C. Kraus, J. Chen and Y. Lee (2003). "JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor." J Biol Chem **278**(43): 42247-42255.

King, I. F., N. J. Francis and R. E. Kingston (2002). "Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro." Mol Cell Biol **22**(22): 7919-7928.

Klose, R. J., S. Cooper, A. M. Farcas, N. P. Blackledge and N. Brockdorff (2013). "Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins." PLoS Genet **9**(8): e1003717.

Klose, R. J., E. M. Kallin and Y. Zhang (2006). "JmjC-domain-containing proteins and histone demethylation." Nat Rev Genet **7**(9): 715-727.

Klymenko, T., B. Papp, W. Fischle, T. Kocher, M. Schelder, C. Fritsch, B. Wild, M. Wilm and J. Muller (2006). "A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities." Genes Dev **20**(9): 1110-1122.

Koch, C. M., R. M. Andrews, P. Flicek, S. C. Dillon, U. Karaoz, G. K. Clelland, S. Wilcox, D. M. Beare, J. C. Fowler, P. Couttet, K. D. James, G. C. Lefebvre, A. W. Bruce, O. M. Dovey, P. D. Ellis, P. Dhami, C. F. Langford, Z. Weng, E. Birney, N. P. Carter, D. Vetrie and I. Dunham (2007). "The landscape of histone modifications across 1% of the human genome in five human cell lines." Genome Res **17**(6): 691-707.

Kooistra, S. M., V. van den Boom, R. P. Thummer, F. Johannes, R. Wardenaar, B. M. Tesson, L. M. Veenhoff, F. Fusetti, L. P. O'Neill, B. M. Turner, G. de Haan and B. J. Eggen (2010). "Undifferentiated embryonic cell transcription factor 1 regulates ESC chromatin organization and gene expression." Stem cells **28**(10): 1703-1714.

Kornberg, R. D. (1974). "Chromatin structure: a repeating unit of histones and DNA." Science **184**(4139): 868-871.

Kossel, A. (1884). "Ueber einen peptoartigen bestandteil des zellkerns." Physiol Chem **8**: 511-515.

Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.

Ku, M., R. P. Koche, E. Rheinbay, E. M. Mendenhall, M. Endoh, T. S. Mikkelsen, A. Presser, C. Nusbaum, X. Xie, A. S. Chi, M. Adli, S. Kasif, L. M. Ptaszek, C. A. Cowan, E. S. Lander, H. Koseki and B. E. Bernstein (2008). "Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains." PLoS Genet **4**(10): e1000242.

Kuras, L. and K. Struhl (1999). "Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme." Nature **399**(6736): 609-613.

Kuzmichev, A., K. Nishioka, H. Erdjument-Bromage, P. Tempst and D. Reinberg (2002). "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein." Genes Dev **16**(22): 2893-2905.

Kwong, C., B. Adryan, I. Bell, L. Meadows, S. Russell, J. R. Manak and R. White (2008). "Stability and dynamics of polycomb target sites in Drosophila development." PLoS Genet **4**(9): e1000178.

Landeira, D., S. Sauer, R. Poot, M. Dvorkina, L. Mazzarella, H. F. Jorgensen, C. F. Pereira, M. Leleu, F. M. Piccolo, M. Spivakov, E. Brookes, A. Pombo, C. Fisher, W. C. Skarnes, T. Snoek, K. Bezstarosti, J. Demmers, R. J. Klose, M. Casanova, L. Tavares, N. Brockdorff, M. Merkenschlager and A. G. Fisher (2010). "Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators." Nat Cell Biol **12**(6): 618-624.

Lanzuolo, C. and V. Orlando (2012). "Memories from the polycomb group proteins." Annu Rev Genet **46**: 561-589.

Lawrence, M., R. Gentleman and V. Carey (2009). "rtracklayer: an R package for interfacing with genome browsers." Bioinformatics **25**(14): 1841-1842.

Lee, J. T., L. S. Davidow and D. Warshawsky (1999). "Tsix, a gene antisense to Xist at the X-inactivation centre." Nat Genet **21**(4): 400-404.

Lee, T. I., R. G. Jenner, L. A. Boyer, M. G. Guenther, S. S. Levine, R. M. Kumar, B. Chevalier, S. E. Johnstone, M. F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H. L. Murray, J. P. Zucker, B. Yuan, G. W. Bell, E. Herbolsheimer, N. M. Hannett, K. Sun, D. T. Odom, A. P. Otte, T. L. Volkert, D. P. Bartel, D. A. Melton, D. K. Gifford, R. Jaenisch and R. A. Young (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-313.

Leeb, M., D. Pasini, M. Novatchkova, M. Jaritz, K. Helin and A. Wutz (2010). "Polycomb complexes act redundantly to repress genomic repeats and genes." Genes Dev **24**(3): 265-276.

Lehmann, L., R. Ferrari, A. A. Vashisht, J. A. Wohlschlegel, S. K. Kurdistani and M. Carey (2012). "Polycomb repressive complex 1 (PRC1) disassembles RNA polymerase II preinitiation complexes." J Biol Chem **287**(43): 35784-35794.

Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch and E. Li (1996). "De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells." Development **122**(10): 3195-3205.

Lemon, B. and R. Tjian (2000). "Orchestrated response: a symphony of transcription factors for gene control." Genes Dev **14**(20): 2551-2569.

Levine, M. and R. Tjian (2003). "Transcription regulation and animal diversity." Nature **424**(6945): 147-151.

Levine, S. S., I. F. King and R. E. Kingston (2004). "Division of labor in polycomb group repression." Trends in biochemical sciences **29**(9): 478-485.

Levine, S. S., A. Weiss, H. Erdjument-Bromage, Z. Shao, P. Tempst and R. E. Kingston (2002). "The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans." Mol Cell Biol **22**(17): 6070-6078.

Lewis, E. B. (1978). "A gene complex controlling segmentation in Drosophila." Nature **276**(5688): 565-570.

Li, E., T. H. Bestor and R. Jaenisch (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell **69**(6): 915-926.

Li, G., R. Margueron, M. Ku, P. Chambon, B. E. Bernstein and D. Reinberg (2010). "Jarid2 and PRC2, partners in regulating gene expression." Genes Dev **24**(4): 368-380.

Lienert, F., C. Wirbelauer, I. Som, A. Dean, F. Mohn and D. Schubeler (2011). "Identification of genetic elements that autonomously determine DNA methylation states." Nat Genet.

Lin, Y. S., M. Carey, M. Ptashne and M. R. Green (1990). "How different eukaryotic transcriptional activators can cooperate promiscuously." Nature **345**(6273): 359-361.

Lodato, M. A., C. W. Ng, J. A. Wamstad, A. W. Cheng, K. K. Thai, E. Fraenkel, R. Jaenisch and L. A. Boyer (2013). "SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state." PLoS Genet **9**(2): e1003288.

Long, H. K., N. P. Blackledge and R. J. Klose (2013). "ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection." Biochemical Society transactions **41**(3): 727-740.

Long, H. K., D. Sims, A. Heger, N. P. Blackledge, C. Kutter, M. L. Wright, F. Grutzner, D. T. Odom, R. Patient, C. P. Ponting and R. J. Klose (2013). "Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates." eLife **2**: e00348.

Lorch, Y., J. W. LaPointe and R. D. Kornberg (1987). "Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones." Cell **49**(2): 203-210.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.

Lynch, M. D., A. J. Smith, M. De Gobbi, M. Flenley, J. R. Hughes, D. Vernimmen, H. Ayyub, J. A. Sharpe, J. A. Sloane-Stanley, L. Sutherland, S. Meek, T. Burdon, R. J. Gibbons, D. Garrick and D. R. Higgs (2012). "An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment." The EMBO journal **31**(2): 317-329.

Margueron, R., N. Justin, K. Ohno, M. L. Sharpe, J. Son, W. J. Drury, 3rd, P. Voigt, S. R. Martin, W. R. Taylor, V. De Marco, V. Pirrotta, D. Reinberg and S. J. Gamblin (2009). "Role of the polycomb protein EED in the propagation of repressive histone marks." Nature **461**(7265): 762-767.

Margueron, R., G. Li, K. Sarma, A. Blais, J. Zavadil, C. L. Woodcock, B. D. Dynlacht and D. Reinberg (2008). "Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms." Molecular cell **32**(4): 503-518.

Martienssen, R. A. and V. Colot (2001). "DNA methylation and epigenetic inheritance in plants and filamentous fungi." Science **293**(5532): 1070-1074.

Maston, G. A., S. K. Evans and M. R. Green (2006). "Transcriptional regulatory elements in the human genome." Annu Rev Genomics Hum Genet **7**: 29-59.

Maunakea, A. K., R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang and J. F. Costello (2010). "Conserved role of intragenic DNA methylation in regulating alternative promoters." Nature **466**(7303): 253-257.

Megee, P. C., B. A. Morgan and M. M. Smith (1995). "Histone H4 and the maintenance of genome integrity." Genes Dev **9**(14): 1716-1727.

Meissner, A. (2010). "Epigenetic modifications in pluripotent and differentiated cells." Nat Biotechnol **28**(10): 1079-1088.

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

Metzger, E., M. Wissmann, N. Yin, J. M. Muller, R. Schneider, A. H. Peters, T. Gunther, R. Buettner and R. Schule (2005). "LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription." Nature **437**(7057): 436-439.

Michaels, S. D. and R. M. Amasino (1999). "The gibberellic acid biosynthesis mutant ga1-3 of Arabidopsis thaliana is responsive to vernalization." Dev Genet **25**(3): 194-198.

Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander and B. E. Bernstein (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature **448**(7153): 553-560.

Mills, A. A. (2010). "Throwing the cancer switch: reciprocal roles of polycomb and trithorax proteins." Nat Rev Cancer **10**(10): 669-682.

Min, J., Y. Zhang and R. M. Xu (2003). "Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27." Genes Dev **17**(15): 1823-1828.

Moehrle, A. and R. Paro (1994). "Spreading the silence: epigenetic transcriptional regulation during Drosophila development." Dev Genet **15**(6): 478-484.

Mohn, F., M. Weber, M. Rebhan, T. C. Roloff, J. Richter, M. B. Stadler, M. Bibel and D. Schubeler (2008). "Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors." Mol Cell **30**(6): 755-766.

Morse, R. H. (1989). "Nucleosomes inhibit both transcriptional initiation and elongation by RNA polymerase III in vitro." The EMBO journal **8**(8): 2343-2351.

Mulholland, N. M., I. F. King and R. E. Kingston (2003). "Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA." Genes Dev **17**(22): 2741-2746.

Muller, J., C. M. Hart, N. J. Francis, M. L. Vargas, A. Sengupta, B. Wild, E. L. Miller, M. B. O'Connor, R. E. Kingston and J. A. Simon (2002). "Histone methyltransferase activity of a Drosophila Polycomb group repressor complex." Cell **111**(2): 197-208.

Muller, J. and J. A. Kassis (2006). "Polycomb response elements and targeting of Polycomb group proteins in Drosophila." Curr Opin Genet Dev **16**(5): 476-484.

Murray, K. (1964). "The Occurrence of Epsilon-N-Methyl Lysine in Histones." Biochemistry **3**: 10-15.

Musselman, C. A., M. E. Lalonde, J. Cote and T. G. Kutateladze (2012). "Perceiving the epigenetic landscape through histone readers." Nat Struct Mol Biol **19**(12): 1218-1227.

Nan, X., R. R. Meehan and A. Bird (1993). "Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2." Nucleic Acids Res **21**(21): 4886-4892.

Negre, N., J. Hennetin, L. V. Sun, S. Lavrov, M. Bellis, K. P. White and G. Cavalli (2006). "Chromosomal distribution of PcG proteins during Drosophila development." PLoS Biol **4**(6): e170.

Nguyen, A. T. and Y. Zhang (2011). "The diverse functions of Dot1 and H3K79 methylation." Genes Dev **25**(13): 1345-1358.

O'Carroll, D., S. Erhardt, M. Pagani, S. C. Barton, M. A. Surani and T. Jenuwein (2001). "The polycomb-group gene Ezh2 is required for early mouse development." Mol Cell Biol **21**(13): 4330-4336.

O'Carroll, D., H. Scherthan, A. H. Peters, S. Opravil, A. R. Haynes, G. Laible, S. Rea, M. Schmid, A. Lebersorger, M. Jerratsch, L. Sattler, M. G. Mattei, P. Denny, S. D. Brown, D. Schweizer and T. Jenuwein (2000). "Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression." Mol Cell Biol **20**(24): 9423-9433.

Okano, M., D. W. Bell, D. A. Haber and E. Li (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-257.

Okano, M., S. Xie and E. Li (1998). "Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases." Nat Genet **19**(3): 219-220.

Oktaba, K., L. Gutierrez, J. Gagneur, C. Girardot, A. K. Sengupta, E. E. Furlong and J. Muller (2008). "Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila." Dev Cell **15**(6): 877-889.

Ong, C. T. and V. G. Corces (2011). "Enhancer function: new insights into the regulation of tissue-specific gene expression." Nat Rev Genet **12**(4): 283-293.

Ooi, S. K., C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S. P. Lin, C. D. Allis, X. Cheng and T. H. Bestor (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." Nature **448**(7154): 714-717.

Orlando, D. A., M. G. Guenther, G. M. Frampton and R. A. Young (2012). "CpG island structure and trithorax/polycomb chromatin domains in human cells." Genomics **100**(5): 320-326.

Orlando, V. and R. Paro (1993). "Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin." Cell **75**(6): 1187-1198.

Oudet, P., M. Gross-Bellard and P. Chambon (1975). "Electron microscopic and biochemical evidence that chromatin structure is a repeating unit." Cell **4**(4): 281-300.

Pandey, R. R., T. Mondal, F. Mohammad, S. Enroth, L. Redrup, J. Komorowski, T. Nagano, D. Mancini-Dinardo and C. Kanduri (2008). "Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation." Mol Cell **32**(2): 232-246.

Panning, B. (2010). "Fine-tuning silencing." Cell Stem Cell **6**(1): 3-4.

Papp, B. and J. Muller (2006). "Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins." Genes Dev **20**(15): 2041-2054.

Paro, R. and D. S. Hogness (1991). "The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila." Proc Natl Acad Sci U S A **88**(1): 263-267.

Pasini, D., A. P. Bracken, J. B. Hansen, M. Capillo and K. Helin (2007). "The polycomb group protein Suz12 is required for embryonic stem cell differentiation." Molecular and cellular biology **27**(10): 3769-3779.

Pasini, D., A. P. Bracken, M. R. Jensen, E. Lazzerini Denchi and K. Helin (2004). "Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity." EMBO J **23**(20): 4061-4071.

Pasini, D., P. A. Cloos, J. Walfridsson, L. Olsson, J. P. Bukowski, J. V. Johansen, M. Bak, N. Tommerup, J. Rappsilber and K. Helin (2010). "JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells." Nature **464**(7286): 306-310.

Payer, B. and J. T. Lee (2008). "X chromosome dosage compensation: how mammals keep the balance." Annu Rev Genet **42**: 733-772.

Peng, J. C., A. Valouev, T. Swigut, J. Zhang, Y. Zhao, A. Sidow and J. Wysocka (2009). "Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells." Cell **139**(7): 1290-1302.

Pengelly, A. R., O. Copur, H. Jackle, A. Herzig and J. Muller (2013). "A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb." Science **339**(6120): 698-699.

Pereira, C. F., F. M. Piccolo, T. Tsubouchi, S. Sauer, N. K. Ryan, L. Bruno, D. Landeira, J. Santos, A. Banito, J. Gil, H. Koseki, M. Merkenschlager and A. G. Fisher (2010). "ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency." Cell Stem Cell **6**(6): 547-556.

Peters, A. H., S. Kubicek, K. Mechtler, R. J. O'Sullivan, A. A. Derijck, L. Perez-Burgos, A. Kohlmaier, S. Opravil, M. Tachibana, Y. Shinkai, J. H. Martens and T. Jenuwein (2003). "Partitioning and plasticity of repressive histone methylation states in mammalian chromatin." Mol Cell **12**(6): 1577-1589.

Peters, A. H., D. O'Carroll, H. Scherthan, K. Mechtler, S. Sauer, C. Schofer, K. Weipoltshammer, M. Pagani, M. Lachner, A. Kohlmaier, S. Opravil, M. Doyle, M. Sibilia and T. Jenuwein (2001). "Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability." Cell **107**(3): 323-337.

Peterson, A. J., D. R. Mallin, N. J. Francis, C. S. Ketel, J. Stamm, R. K. Voeller, R. E. Kingston and J. A. Simon (2004). "Requirement for sex comb on midleg protein interactions in Drosophila polycomb group repression." Genetics **167**(3): 1225-1239.

Phillips, D. M. (1963). "The presence of acetyl groups of histones." Biochem J **87**: 258-263.

Plath, K., J. Fang, S. K. Mlynarczyk-Evans, R. Cao, K. A. Worringer, H. Wang, C. C. de la Cruz, A. P. Otte, B. Panning and Y. Zhang (2003). "Role of histone H3 lysine 27 methylation in X inactivation." Science **300**(5616): 131-135.

Pogo, B. G., V. G. Allfrey and A. E. Mirsky (1966). "RNA synthesis and histone acetylation during the course of gene activation in lymphocytes." Proc Natl Acad Sci U S A **55**(4): 805-812.

Ptashne, M. (1967). "Specific binding of the lambda phage repressor to lambda DNA." Nature **214**(5085): 232-234.

Ptashne, M. (2005). "Regulation of transcription: from lambda to eukaryotes." Trends Biochem Sci **30**(6): 275-279.

Rada-Iglesias, A., R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn and J. Wysocka (2011). "A unique chromatin signature uncovers early developmental enhancers in humans." Nature **470**(7333): 279-283.

Ram, O., A. Goren, I. Amit, N. Shoresh, N. Yosef, J. Ernst, M. Kellis, M. Gymrek, R. Issner, M. Coyne, T. Durham, X. Zhang, J. Donaghey, C. B. Epstein, A. Regev and B. E. Bernstein (2011). "Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells." Cell **147**(7): 1628-1639.

Rea, S., F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis and T. Jenuwein (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases." Nature **406**(6796): 593-599.

Reddington, J. P., S. M. Perricone, C. E. Nestor, J. Reichmann, N. A. Youngson, M. Suzuki, D. Reinhardt, D. S. Dunican, J. G. Prendergast, H. Mjoseng, B. H. Ramsahoye, E. Whitelaw, J. M. Greally, I. R. Adams, W. A. Bickmore and R. R. Meehan (2013). "Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes." Genome biology **14**(3): R25.

Ren, X. and T. K. Kerppola (2011). "REST interacts with Cbx proteins and regulates polycomb repressive complex 1 occupancy at RE1 elements." Mol Cell Biol **31**(10): 2100-2110.

Riddle, N. C., A. Minoda, P. V. Kharchenko, A. A. Alekseyenko, Y. B. Schwartz, M. Y. Tolstorukov, A. A. Gorchakov, J. D. Jaffe, C. Kennedy, D. Linder-Basso, S. E. Peach, G. Shanower, H. Zheng, M. I. Kuroda, V. Pirrotta, P. J. Park, S. C. Elgin and G. H. Karpen (2011). "Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin." Genome Res **21**(2): 147-163.

Riggs, A. D. (1975). "X inactivation, differentiation, and DNA methylation." Cytogenet Cell Genet **14**(1): 9-25.

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

Ringrose, L. and R. Paro (2007). "Polycomb/Trithorax response elements and epigenetic memory of cell identity." Development **134**(2): 223-232.

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

Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal and H. Y. Chang (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell **129**(7): 1311-1323.

Rose, N. R. and R. J. Klose (2014). "Understanding the relationship between DNA methylation and histone lysine methylation." Biochim Biophys Acta.

Ross, J. M. and D. Zarkower (2003). "Polycomb group regulation of Hox gene expression in C. elegans." Dev Cell **4**(6): 891-901.

Russell, G. J., P. M. Walker, R. A. Elton and J. H. Subak-Sharpe (1976). "Doublet frequency analysis of fractionated vertebrate nuclear DNA." J Mol Biol **108**(1): 1-23.

Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor and T. Kouzarides (2002). "Active genes are tri-methylated at K4 of histone H3." Nature **419**(6905): 407-411.

Sarma, K., R. Margueron, A. Ivanov, V. Pirrotta and D. Reinberg (2008). "Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo." Molecular and cellular biology **28**(8): 2718-2731.

Sato, T. and R. E. Denell (1985). "Homoeosis in Drosophila: anterior and posterior transformations of Polycomb lethal embryos." Dev Biol **110**(1): 53-64.

Saxonov, S., P. Berg and D. L. Brutlag (2006). "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." Proc Natl Acad Sci U S A **103**(5): 1412-1417.

Schmitges, F. W., A. B. Prusty, M. Faty, A. Stutzer, G. M. Lingaraju, J. Aiwazian, R. Sack, D. Hess, L. Li, S. Zhou, R. D. Bunker, U. Wirth, T. Bouwmeester, A. Bauer, N. Ly-Hartig, K. Zhao, H. Chan, J. Gu, H. Gut, W. Fischle, J. Muller and N. H. Thoma (2011). "Histone methylation by PRC2 is inhibited by active chromatin marks." Mol Cell **42**(3): 330-341.

Schmitt, S., M. Prestel and R. Paro (2005). "Intergenic transcription through a polycomb group response element counteracts silencing." Genes Dev **19**(6): 697-708.

Schoenfelder, S., I. Clay and P. Fraser (2010). "The transcriptional interactome: gene expression in 3D." Curr Opin Genet Dev **20**(2): 127-133.

Schubeler, D., D. M. MacAlpine, D. Scalzo, C. Wirbelauer, C. Kooperberg, F. van Leeuwen, D. E. Gottschling, L. P. O'Neill, B. M. Turner, J. Delrow, S. P. Bell and M. Groudine (2004). "The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote." Genes Dev **18**(11): 1263-1271.

Schubert, D., L. Primavesi, A. Bishopp, G. Roberts, J. Doonan, T. Jenuwein and J. Goodrich (2006). "Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27." EMBO J **25**(19): 4638-4649.

Schuettengruber, B., D. Chourrout, M. Vervoort, B. Leblanc and G. Cavalli (2007). "Genome regulation by polycomb and trithorax proteins." Cell **128**(4): 735-745.

Schuettengruber, B., M. Ganapathi, B. Leblanc, M. Portoso, R. Jaschek, B. Tolhuis, M. van Lohuizen, A. Tanay and G. Cavalli (2009). "Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos." PLoS biology **7**(1): e13.

Schumacher, A. and T. Magnuson (1997). "Murine Polycomb- and trithorax-group genes regulate homeotic pathways and beyond." Trends Genet **13**(5): 167-170.

Schwartz, Y. B., T. G. Kahn, D. A. Nix, X. Y. Li, R. Bourgon, M. Biggin and V. Pirrotta (2006). "Genome-wide analysis of Polycomb targets in Drosophila melanogaster." Nat Genet **38**(6): 700-705.

Seila, A. C., J. M. Calabrese, S. S. Levine, G. W. Yeo, P. B. Rahl, R. A. Flynn, R. A. Young and P. A. Sharp (2008). "Divergent transcription from active promoters." Science **322**(5909): 1849-1851.

Sengupta, A. K., A. Kuhrs and J. Muller (2004). "General transcriptional silencing by a Polycomb response element in Drosophila." Development **131**(9): 1959-1965.

Shao, Z., F. Raible, R. Mollaaghababa, J. R. Guyon, C. T. Wu, W. Bender and R. E. Kingston (1999). "Stabilization of chromatin structure by PRC1, a Polycomb complex." Cell **98**(1): 37-46.

Sharif, J., T. A. Endo, T. Toyoda and H. Koseki (2010). "Divergence of CpG island promoters: a consequence or cause of evolution?" Dev Growth Differ **52**(6): 545-554.

Sheldon, C. C., J. E. Burn, P. P. Perez, J. Metzger, J. A. Edwards, W. J. Peacock and E. S. Dennis (1999). "The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation." Plant Cell **11**(3): 445-458.

Shen, X., W. Kim, Y. Fujiwara, M. D. Simon, Y. Liu, M. R. Mysliwiec, G. C. Yuan, Y. Lee and S. H. Orkin (2009). "Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells." Cell **139**(7): 1303-1314.

Shen, X., Y. Liu, Y. J. Hsu, Y. Fujiwara, J. Kim, X. Mao, G. C. Yuan and S. H. Orkin (2008). "EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency." Mol Cell **32**(4): 491-502.

Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero and Y. Shi (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." Cell **119**(7): 941-953.

Silva, J., W. Mak, I. Zvetkova, R. Appanah, T. B. Nesterova, Z. Webster, A. H. Peters, T. Jenuwein, A. P. Otte and N. Brockdorff (2003). "Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes." Dev Cell **4**(4): 481-495.

Simon, J., A. Chiang and W. Bender (1992). "Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products." Development **114**(2): 493-505.

Simon, J., A. Chiang, W. Bender, M. J. Shimell and M. O'Connor (1993). "Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products." Developmental biology **158**(1): 131-144.

Simon, J., M. Peifer, W. Bender and M. O'Connor (1990). "Regulatory elements of the bithorax complex that control expression along the anterior-posterior axis." EMBO J **9**(12): 3945-3956.

Sing, A., D. Pannell, A. Karaiskakis, K. Sturgeon, M. Djabali, J. Ellis, H. D. Lipshitz and S. P. Cordes (2009). "A vertebrate Polycomb response element governs segmentation of the posterior hindbrain." Cell **138**(5): 885-897.

Singer, J., J. Roberts-Ems, F. W. Luthardt and A. D. Riggs (1979). "Methylation of DNA in mouse early embryos, teratocarcinoma cells and adult tissues of mouse and rabbit." Nucleic Acids Res **7**(8): 2369-2385.

Sleutels, F., R. Zwart and D. P. Barlow (2002). "The non-coding Air RNA is required for silencing autosomal imprinted genes." Nature **415**(6873): 810-813.

Stadler, M. B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Scholer, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari and D. Schubeler (2011). "DNA-binding factors shape the mouse methylome at distal regulatory regions." Nature **480**(7378): 490-495.

Stein, R., Y. Gruenbaum, Y. Pollack, A. Razin and H. Cedar (1982). "Clonal inheritance of the pattern of DNA methylation in mouse cells." Proc Natl Acad Sci U S A **79**(1): 61-65.

Stewart, C. L., H. Stuhlmann, D. Jahner and R. Jaenisch (1982). "De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells." Proc Natl Acad Sci U S A **79**(13): 4098-4102.

Stock, J. K., S. Giadrossi, M. Casanova, E. Brookes, M. Vidal, H. Koseki, N. Brockdorff, A. G. Fisher and A. Pombo (2007). "Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells." Nat Cell Biol **9**(12): 1428-1435.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-45.

Struhl, K. (1999). "Fundamentally different logic of gene regulation in eukaryotes and prokaryotes." Cell **98**(1): 1-4.

Strutt, H., G. Cavalli and R. Paro (1997). "Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression." EMBO J **16**(12): 3621-3632.

Surface, L. E., S. R. Thornton and L. A. Boyer (2010). "Polycomb group proteins set the stage for early lineage commitment." Cell Stem Cell **7**(3): 288-298.

Suva, M. L., N. Riggi, M. Janiszewska, I. Radovanovic, P. Provero, J. C. Stehle, K. Baumer, M. A. Le Bitoux, D. Marino, L. Cironi, V. E. Marquez, V. Clement and I. Stamenkovic (2009). "EZH2 is essential for glioblastoma cancer stem cell maintenance." Cancer Res **69**(24): 9211-9218.

Suzuki, M., Y. Mizutani-Koseki, Y. Fujimura, H. Miyagishima, T. Kaneko, Y. Takada, T. Akasaka, H. Tanzawa, Y. Takihara, M. Nakano, H. Masumoto, M. Vidal, K. Isono and H. Koseki (2002). "Involvement of the Polycomb-group gene Ring1B in the specification of the anterior-posterior axis in mice." Development **129**(18): 4171-4183.

Suzuki, M. M. and A. Bird (2008). "DNA methylation landscapes: provocative insights from epigenomics." Nat Rev Genet **9**(6): 465-476.

Tachibana, M., K. Sugimoto, M. Nozaki, J. Ueda, T. Ohta, M. Ohki, M. Fukuda, N. Takeda, H. Niida, H. Kato and Y. Shinkai (2002). "G9a histone methyltransferase

plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis." Genes Dev **16**(14): 1779-1791.

Takai, D. and P. A. Jones (2002). "Comprehensive analysis of CpG islands in human chromosomes 21 and 22." Proc Natl Acad Sci U S A **99**(6): 3740-3745.

Tanay, A., A. H. O'Donnell, M. Damelin and T. H. Bestor (2007). "Hyperconserved CpG domains underlie Polycomb-binding sites." Proceedings of the National Academy of Sciences of the United States of America **104**(13): 5521-5526.

Taunton, J., C. A. Hassig and S. L. Schreiber (1996). "A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p." Science **272**(5260): 408-411.

Tavares, L., E. Dimitrova, D. Oxley, J. Webster, R. Poot, J. Demmers, K. Bezstarosti, S. Taylor, H. Ura, H. Koide, A. Wutz, M. Vidal, S. Elderkin and N. Brockdorff (2012). "RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3." Cell **148**(4): 664-678.

Taverna, S. D., H. Li, A. J. Ruthenburg, C. D. Allis and D. J. Patel (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nat Struct Mol Biol **14**(11): 1025-1040.

Thomson, J. P., P. J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A. R. Kerr, A. Deaton, R. Andrews, K. D. James, D. J. Turner, R. Illingworth and A. Bird (2010). "CpG islands influence chromatin structure via the CpG-binding protein Cfp1." Nature **464**(7291): 1082-1086.

Tippmann, S. C., R. Ivanek, D. Gaidatzis, A. Scholer, L. Hoerner, E. van Nimwegen, P. F. Stadler, M. B. Stadler and D. Schubeler (2012). "Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels." Molecular systems biology **8**: 593.

Tiwari, V. K., M. B. Stadler, C. Wirbelauer, R. Paro, D. Schubeler and C. Beisel (2012). "A chromatin-modifying function of JNK during stem cell differentiation." Nature genetics **44**(1): 94-100.

Tolhuis, B., E. de Wit, I. Muijrers, H. Teunissen, W. Talhout, B. van Steensel and M. van Lohuizen (2006). "Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster." Nat Genet **38**(6): 694-699.

Tremethick, D. J. (2007). "Higher-order structures of chromatin: the elusive 30 nm fiber." Cell **128**(4): 651-654.

Trojer, P. and D. Reinberg (2007). "Facultative heterochromatin: is there a distinctive molecular signature?" Mol Cell **28**(1): 1-13.

Tsumura, A., T. Hayakawa, Y. Kumaki, S. Takebayashi, M. Sakaue, C. Matsuoka, K. Shimotohno, F. Ishikawa, E. Li, H. R. Ueda, J. Nakayama and M. Okano (2006). "Maintenance of self-renewal ability of mouse embryonic stem cells in the absence

of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b." Genes to cells : devoted to molecular & cellular mechanisms **11**(7): 805-814.

Tweedie, S., J. Charlton, V. Clark and A. Bird (1997). "Methylation of genomes and genes at the invertebrate-vertebrate boundary." Mol Cell Biol **17**(3): 1469-1475.

van Heeringen, S. J., R. C. Akkers, I. van Kruijsbergen, M. A. Arif, L. L. Hanssen, N. Sharifi and G. J. Veenstra (2013). "Principles of nucleation of H3K27 methylation during embryonic development." Genome research.

Vaquerizas, J. M., S. K. Kummerfeld, S. A. Teichmann and N. M. Luscombe (2009). "A census of human transcription factors: function, expression and evolution." Nat Rev Genet **10**(4): 252-263.

Vella, P., I. Barozzi, A. Cuomo, T. Bonaldi and D. Pasini (2012). "Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells." Nucleic acids research **40**(8): 3403-3418.

Vermeulen, M., K. W. Mulder, S. Denissov, W. W. Pijnappel, F. M. van Schaik, R. A. Varier, M. P. Baltissen, H. G. Stunnenberg, M. Mann and H. T. Timmers (2007). "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4." Cell **131**(1): 58-69.

Voigt, P., G. LeRoy, W. J. Drury, 3rd, B. M. Zee, J. Son, D. B. Beck, N. L. Young, B. A. Garcia and D. Reinberg (2012). "Asymmetrically modified nucleosomes." Cell **151**(1): 181-193.

Voncken, J. W., B. A. Roelen, M. Roefs, S. de Vries, E. Verhoeven, S. Marino, J. Deschamps and M. van Lohuizen (2003). "Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition." Proc Natl Acad Sci U S A **100**(5): 2468-2473.

Waddington, C. H. (1942). "Canalization of development and the inheritance of acquired characters." Nature genetics **150**: 563-565.

Walker, E., W. Y. Chang, J. Hunkapiller, G. Cagney, K. Garcha, J. Torchia, N. J. Krogan, J. F. Reiter and W. L. Stanford (2010). "Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation." Cell Stem Cell **6**(2): 153-166.

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

Wang, J., J. Mager, Y. Chen, E. Schneider, J. C. Cross, A. Nagy and T. Magnuson (2001). "Imprinted X inactivation maintained by a mouse Polycomb group gene." Nat Genet **28**(4): 371-375.

Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan and D. Schubeler (2007). "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome." Nat Genet **39**(4): 457-466.

Weber, M. and D. Schubeler (2007). "Genomic patterns of DNA methylation: targets and function of an epigenetic mark." Curr Opin Cell Biol **19**(3): 273-280.

Wei, Y., Y. H. Chen, L. Y. Li, J. Lang, S. P. Yeh, B. Shi, C. C. Yang, J. Y. Yang, C. Y. Lin, C. C. Lai and M. C. Hung (2011). "CDK1-dependent phosphorylation of EZH2 suppresses methylation of H3K27 and promotes osteogenic differentiation of human mesenchymal stem cells." Nat Cell Biol **13**(1): 87-94.

Weikert, S., F. Christoph, J. Kollermann, M. Muller, M. Schrader, K. Miller and H. Krause (2005). "Expression levels of the EZH2 polycomb transcriptional repressor correlate with aggressiveness and invasive potential of bladder carcinomas." Int J Mol Med **16**(2): 349-353.

Widom, J. and A. Klug (1985). "Structure of the 300A chromatin filament: X-ray diffraction from oriented samples." Cell **43**(1): 207-213.

Woo, C. J., P. V. Kharchenko, L. Daheron, P. J. Park and R. E. Kingston (2010). "A region of the human HOXD cluster that confers polycomb-group responsiveness." Cell **140**(1): 99-110.

Woo, C. J., P. V. Kharchenko, L. Daheron, P. J. Park and R. E. Kingston (2013). "Variable requirements for DNA-binding proteins at polycomb-dependent repressive regions in human HOX clusters." Molecular and cellular biology **33**(16): 3274-3285.

Wood, C. C., M. Robertson, G. Tanner, W. J. Peacock, E. S. Dennis and C. A. Helliwell (2006). "The Arabidopsis thaliana vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3." Proc Natl Acad Sci U S A **103**(39): 14631-14636.

Workman, J. L. and R. E. Kingston (1998). "Alteration of nucleosome structure as a mechanism of transcriptional regulation." Annu Rev Biochem **67**: 545-579.

Wu, X., J. V. Johansen and K. Helin (2013). "Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation." Mol Cell **49**(6): 1134-1146.

Xu, C., C. Bian, W. Yang, M. Galka, H. Ouyang, C. Chen, W. Qiu, H. Liu, A. E. Jones, F. MacKenzie, P. Pan, S. S. Li, H. Wang and J. Min (2010). "Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2)." Proc Natl Acad Sci U S A **107**(45): 19266-19271.

Xu, Y. Z. K., C. (2013). Chromatin Remodelling During Host-Bacterial Pathogen Interaction. Chromatin Remodelling. D. D. Radzioch.

Yu, M., T. Mazor, H. Huang, H. T. Huang, K. L. Kathrein, A. J. Woo, C. R. Chouinard, A. Labadorf, T. E. Akie, T. B. Moran, H. Xie, S. Zacharek, I. Taniuchi, R. G. Roeder, C. F. Kim, L. I. Zon, E. Fraenkel and A. B. Cantor (2012). "Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors." Mol Cell **45**(3): 330-343.

Zhao, J., T. K. Ohsumi, J. T. Kung, Y. Ogawa, D. J. Grau, K. Sarma, J. J. Song, R. E. Kingston, M. Borowsky and J. T. Lee (2010). "Genome-wide identification of polycomb-associated RNAs by RIP-seq." Mol Cell **40**(6): 939-953.

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

Zhou, V. W., A. Goren and B. E. Bernstein (2011). "Charting histone modifications and the functional organization of mammalian genomes." Nat Rev Genet **12**(1): 7-18.

Zhou, W., P. Zhu, J. Wang, G. Pascual, K. A. Ohgi, J. Lozach, C. K. Glass and M. G. Rosenfeld (2008). "Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation." Mol Cell **29**(1): 69-80.

Zhu, J., M. Adli, J. Y. Zou, G. Verstappen, M. Coyne, X. Zhang, T. Durham, M. Miri, V. Deshpande, P. L. De Jager, D. A. Bennett, J. A. Houmard, D. M. Muoio, T. T. Onder, R. Camahort, C. A. Cowan, A. Meissner, C. B. Epstein, N. Shoresh and B. E. Bernstein (2013). "Genome-wide chromatin state transitions associated with developmental and environmental cues." Cell **152**(3): 642-654.

Zilberman, D., M. Gehring, R. K. Tran, T. Ballinger and S. Henikoff (2007). "Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription." Nat Genet **39**(1): 61-69.

Zink, B. and R. Paro (1989). "In vivo binding pattern of a trans-regulator of homoeotic genes in Drosophila melanogaster." Nature **337**(6206): 468-471.

Philip Jermann

Bläsiring 43
4057 Basel, Switzerland
+41798728915
jermann.philip@gmail.com

Personal Data

30. September 1983, Baden AG, Switzerland; single

Education

2009 – 2014 **Ph.D. in Genetics, Friedrich Miescher Institute, Basel, Switzerland**
Ph.D. thesis: „Identification and characterization of cis-regulatory elements that target Polycomb in the mouse genome"

2007 – 2009 **Master in Biochemistry, ETH Zürich, Zürich, Switzerland**
Master thesis: „Activating cullin-RING ubiquitin ligases by neddylation"

2004 – 2007 **ETH Zürich, Zürich, Switzerland**
Bachelor degree in Biology

2000 – 2004 **Gymnasium, Baden, Switzerland**
Matura

1999 – 2000 **Silver Lake Regional High School, Kingston, MA, USA**
Stay abroad

Research and Work Experience

Friedrich Miescher Institute, Basel, Switzerland

2009 - now **Ph.D. student,** *Department for Epigenetics,* group of Dirk Schübeler

- Design, execution, and analysis of research projects studying the effects of epigenetic modifications on gene expression patterns
  - Analysis of protein-DNA interactions by chromatin-immunoprecipitation coupled with high-throughput sequencing and qPCR
  - Analysis of DNA methylation state via sequencing of bisulfite-converted DNA
  - Bioinformatic analysis of NextGen-sequencing data in *R*

ETH Zürich, Zürich, Switzerland

2008 - 2009 **Master thesis,** *Institute for Biochemistry,* group of Matthias Peter

- Biochemical analysis of protein-protein interactions involved in DNA-repair
  - Protein purification and analysis via Western-Blot

2008 **Semester thesis,** *Institute for Molecular Systems Biology,* group of Lukas Pelkmans

- siRNA-screen to identify factors involved in virus uptake

Additional Courses

July 2012 **Computational Statistics for Genome Biology,** Brixen, Austria

- Bioinformatic analysis of High-throughput DNA sequencing data

June 2012 **Epigenetics meets Systems Biology, Institute of Science,** Rehovot, Israel

- Workshop for the communication between different fields of research

SKILLS

General
- Excellent **presentation and communication skills**
- Hands-on experience with **high-throughput DNA sequencing**
- Extensive experience using **quantitative PCR**
- Broad knowledge in the fields of **epigenetics and genetics**
- **Stem cell culture** and *in vitro* differentiation
- **Project management** and **collaboration** with researchers in an **interdisciplinary and international** environment
- Hands-on experience with various **molecular biology** techniques
- Writing and critical reading of ***peer-reviewed* research articles**

IT skills
- **Bioinformatic analysis** of NextGen-sequencing data in *R*
- Experience in working with **Unix/Linux** systems
- Excellent knowledge of **MS Office**
- Extensive experience with **Adobe Photoshop and Illustrator**

Languages
- **German** (mother tongue)
- **English** (proficient written and spoken)
- **French** (good written and spoken)
- **Swedish** (good written and spoken)

INTERNATIONAL PRESENTATIONS AND POSTERS (SELECTION)

October 2013 **TriRhena Chromatin Club,** Freiburg, Germany – „Identification and characterization of short DNA elements that target Polycomb in the mouse genome" – *presentation*

May 2013 **EMBO Conference Series: Chromatin and Epigenetics,** Heidelberg, Germany – „Deciphering the mechanisms of Polycomb targeting in the mouse genome" – *poster*

Nov. 2012 **Sinergia Meeting 2012,** Staufen, Germany – „Functional characterization of cis-regulatory elements that target Polycomb in the mouse genome" – *presentation*

June 2012 **Epigenetics meets Systems Biology,** Rehovot, Israel – „Targeting principles of the Polycomb group proteins" – *poster*

May 2011 **Joint PhD retreat, CRG,** Barcelona, Spanien – *poster*

HOBBIES AND INTERESTS

Snowboarding, Computers, Hiking, Photography

PUBLICATIONS

P Jermann „Identification and characterization of cis-regulatory elements that target Polycomb in the mouse genome", PhD thesis

P Jermann, L Hoerner, L Burger, D Schübeler, „Proficient recruitment of H3K27 trimethylation by short sequences is counteracted by DNA methylation and enhancer activity", in revision at *Proceedings of the National Academy of Sciences (PNAS).*