universität wien

MASTERARBEIT

Titel der Masterarbeit

Characterisation of transcribed intergenic Polycomb target sites in mammalian embryonic stem cells

Verfasserin

Melita Ticevic BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 066877
Studienrichtung lt. Studienblatt:	Masterstudium Genetik und Entwicklungsbiologie
Betreuerin / Betreuer:	Dr. Leonie Ringrose

ABSTRACT

The highly conserved Polycomb (PcG) and Trithorax (TrxG) proteins work antagonistically to regulate and maintain gene expression pattern in flies and mammals. They constitute a regulatory system that is crucial for the dynamic change and retention of gene expression states. PcG and TrxG proteins contribute to cellular memory during the development of multicellular organisms and are essential regulators of developmentally important genes.

It is known from studies performed in *Drosophila* that PcG and TrxG proteins act through cis-regulatory DNA elements also called Polycomb response elements (PREs). In *Drosophila*, PcG and TrxG proteins bind via sequence-specific DNA binding proteins to PREs and thus regulate gene expression patterns.
In mammals, little is known about the involvement of PcG and TrxG proteins in gene regulation. A recent study from our group based on genome-wide transcription and binding profiles identified transcribed intergenic Polycomb sites (TIP sites), and postulated that these sites could be the equivalent to the *Drosophila* PRE in mammals [1]. The authors validated these findings with transient transfections of TIP site containing vectors and showed that these constructs repressed the reporter gene in murine ES cells. It had previously been shown that many *Drosophila* PREs are transcribed and it has been proposed that their non-coding transcripts are involved in the recruitment of PcG or TrxG proteins to their target sites.

This study aimed to determine whether TIP sites are indeed mammalian PREs, and whether the non-coding transcripts they produce are important for their regulatory properties. To address this question I have generated transgenic murine ES cell lines with integrated TIP sites flanked by a Luciferase reporter gene. All constructs were integrated at an identical genomic location, to exclude genomic position effects and to allow comparisons. I used the transgenic ES cell lines to analyze the functional characteristics of TIP sites in a genomic context. To determine the regulatory effects of TIP sites on gene expression reporter assays were performed. I show that the Nkx2-9 TIP site can reduce the expression of the Luciferase reporter gene to one third compared to the control without the TIP site. Furthermore, to gain better understanding of the involvement of PcG proteins in the repression of the reporter gene, a knockdown of PcG proteins was performed. However the results of this analysis were inconclusive. Finally, ncRNA analysis was performed to determine the role of the non-coding transcripts in reporter gene expression. This analysis showed that every analyzed TIP

site is transcribed and moreover that this transcription may influence that of their corresponding genes. This study shows that TIP sites integrated into chromatin show strikingly different properties to those observed in transient assays [1], and illustrates the importance of studying gene regulatory elements in the correct chromatin context.

ZUSAMMENFASSUNG

Die hochkonservierten PcG- und TrxG-Proteine regulieren und erhalten bestimmte Gen-Expressionsmuster sowohl in *Drosophila* als auch in Wirbeltieren aufrecht. In ihren Funktionen ergänzen sie sich als Antagonisten und steuern einen wichtigen Beitrag zum epigenetischen Gedächtnis von Zellidentitäten während der Entwicklung höherer Organismen bei. Sie sind unentbehrlich für die dynamische Regulation von Genen, die in humanen und murinen ES Zellen für Entwicklungsprozesse wichtig sind.

Aus Studien, die an *Drosophila* durchgeführt wurden ist bekannt, dass PcG- wie auch TrxG-Proteine über sogenannte PREs (Polycomb Response Elements) agieren. PREs sind DNA Sequenzen, welche ein bestimmtes Muster in der Basenfolge enthalten, das von PcG- und TrxG-Proteinen erkannt werden kann.
In Vertebraten sind PREs bislang schwer zu identifizieren gewesen. In einer kürzlich veröffentlichten Studie aus unserer Arbeitsgruppe, welche auf genomweiten Transkriptions- und Proteinbindungsprofilen basiert, wurden sogenannte TIPs (transkribierte, intergenische Polycomb-Sequenzen) vorgestellt [1]. Die Autoren vermuten, dass die TIP sites das Äquivalent zum *Drosophila* PRE in Vertebraten sind. Diese Vermutung wird durch transiente Transfektionen mit Vektoren, welche TIP sites tragen, validiert. Die Ergebnise haben gezeigt, dass TIP sites die Expression eines Reportergens in murinen ES Zellen reprimieren können.
Weiters wurde gezeigt, dass viele Drosophila PREs transkribiert werden und es wurde angenommen, dass diese nicht-kodierenden Transkripte in die Rekrutierung von PcG- und TrxG-Proteinen zu ihren Zielorten involviert sein könnten.

In dieser Arbeit habe ich mir zum Ziel gesetzt herauszufinden, ob TIP sites tatsächlich Vertebraten PREs sind und ob die nicht-kodierenden Transkripte, welche aus ihnen hervorgehen, wichtig für ihre regulatorischen Funktionen sind. Ich habe transgene murine ES Zellen generiert, indem ich TIP sites in das Genom von ES Zellen stabil integriert habe. Alle Konstrukte wurde an einen indenten genetischen Locus integriert, um einen Vergleich zu ermöglichen und Einflüsse durch genomische Positionseffekte zu verhindern. Ich habe die transgenen Zellen benutzt, um die Funktionen von den TIPs, die nun im Chromatin der genomischen DNA eingebunden sind, zu analysieren. Durch den Einsatz eines in den transgenen Zellen enthaltenen Luciferase-Reportergens konnte ich eine dreifache Reduzierung dieses Reporters feststellen. Da Zellen, die kein TIP enthalten, keine Reduktion in der Expression des Reportergens

aufweisen, wissen wir, dass die Reduktion dieses Reportergens abhängig vom Nkx2-9 TIP ist. Deshalb nehmen wir an, dass der Nkx2-9 TIP Polycomb-Proteine rekrutieren und ein Vertebraten PRE sein könnte. Zum Nachweis, dass die Repression des Reportergens von Polycomb-Proteinen abhängig ist, haben wir im Gegenversuch die Produktion von Polycomb-Proteinen durch siRNAs unterbrochen. Jedoch waren die Ergebnisse dieser Experimente nicht beweiskräftig.
Desweiteren haben zahlreiche Studien, darunter die oben erwähnte der Kollegen [1], nachgewiesen, dass bei der Rekrutierung von Polycomb-Proteinen und anschließenden Genrepression auch nicht-kodierende RNS beteiligt sind. Aus diesem Grund wurde eine Analyse der nicht-kodierenden Transkripte in diese Arbeit mit einbezogen. Die Ergebnisse zeigen, dass jede TIP site transkribiert wird und, dass die Transkription dieser die Transkription der zugehörigen Gene beeinflussen könnte.

Diese Studie macht deutlich, dass die Resultate von TIP sites, welche in den genomischen Kontext von murinen ES Zellen integriert wurden, auffällige Differenzen aufweisen verglichen mit den Ergebnissen aus transienten Analysen. Dadurch unterstreicht diese Studie die Wichtigkeit, regulatorische Elemente in ihren korrekten genetischen Hintergrund zu analysieren.

CONTENTS

1 INTRODUCTION

1.1 POLYCOMB GROUP PROTEINS

What is the difference between a liver cell, an endothelial cell and an embryonic stem cell? Clearly, these cells have different functions and morphologies but what is the underlying cause of these differences? It is the gene expression pattern, a combination of genes that are in a transcriptional active or on-state and genes in a repressed, silenced or off-state, which determines cellular identity, function and morphology. The active state is associated with an open chromatin structure and active gene regions are usually transcribed, allowing several DNA-binding proteins and enzyme such as the RNA Polimerase II to access the DNA, which is wrapped around histones. These proteins and enzymes can transcribe the DNA or repair it in the case of single or double strand breaks. The off-state is not associated with active transcription. In contrast, these genes and DNA sections have a compact chromatin structure, which does not allow proteins to access and bind DNA. This leads to transcriptional shutdown of regions whose expression is not necessary for the cell in a specific developmental period or in a specific differentiated cell type.

An important aspect in the development of vertebrates is the ability of ES cells to self-renew while at the same time bearing the ability to differentiate into any given cell type. In ES cells, genes crucial for self-renewal and “stemness” are turned on and actively transcribed whereas genes pivotal for differentiation are turned off and not transcribed [2, 3]. However, to date it is not yet fully understood how gene expression is regulated.

The Polycomb group (PcG) of proteins is crucial for the determination of gene expression and normal development of multicellular organisms. They are highly conserved in flies and mammals and play a major role in silencing genes important for development [4, 5]. In *Drosophila*, they are able to maintain repressed transcriptional states after the initiating determining transcription factors have disappeared [6, 7].
PcG genes were first identified in *Drosophila melanogaster* [8, 9]. They were named after Polycomb (PC) and extra sex comb (ESC) mutations, which cause an additional sex comb phenotype on the second and third leg of *D. melanogaster.* It is now known that this phenotype is caused by misexpressions of homeotic (*Hox*) genes in the antennapedia (ANT-C) and bithorax (BX-C) complexes of *D. melanogaster*. PcG

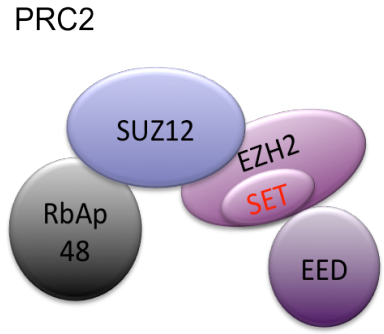
proteins and their antagonistic players the TrxG proteins are crucial for the maintenance of the correct expression of body-patterning genes including the *Hox* genes during the development of *Drosophila* [5]. Their importance in mammalian development is marked by gastrulation arrest and lethality in PcG gene null mouse mutants [10, 11]. How PcG proteins work as transcriptional regulators is not yet fully understood, but it is known that they act through several main protein complexes, which are highly conserved throughout evolution. In the case of Polycomb, the two main multi-protein complexes in both flies and mammals are Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2).

1.2 Polycomb Repressive Complexes (PRC1 and PRC2)

1.2.1 PRC2

In vertebrates, the PcG proteins EED, EZH2, SUZ12 and the histone-binding protein RbAp46/48 are the core components of the PRC2 protein complex [12, 13].

PRC2 catalyzes H3K27me2 and H3K27me3 through the subunit EZH2, which has been shown to exhibit methyltransferase activity via its SET domain [14-16]. The outcome of this histone modification is believed to be transcriptional repression of target genes [17]. Moreover, studies have shown that EZH2 KO mouse embryos die upon gastrulation, demonstrating an essential role of EZH2 in mouse development [18]. Several authors have reported that binding of SUZ12 and EED is required for EZH2 activity [12, 13].



In vivo and *in vitro* experiments have suggested an important role of SUZ12 in regulation of PRC2 [13]. It has been reported that SUZ12 can stimulate the activity of PRC2 by recruiting the histone-binding protein RbAp48 to the complex, which stabilizes the protein complex [13].

Figure 1 Vertebrate PRC2 core components.

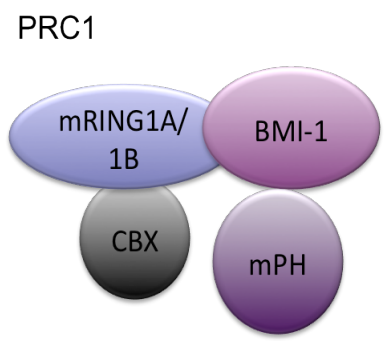
Furthermore, it has been shown that H3K27me2/me3 is lost in mouse embryos lacking the SUZ12 protein. This confirms that SUZ12 is necessary for EZH2 HMT activity [19, 20]. During mouse embryogenesis SUZ12 KO mutants have proliferative defects and are not viable [13].

The core component EED has, similar to SUZ12 and EZH2, been shown to be important for mouse development, since EED KO mice exhibit growth defects and die upon gastrulation [18, 21]. Moreover, EED functions as a scaffold protein since it can link EZH2 to histone H3 substrates and thereby stabilize the complex at target sites [22].
In addition, it has been reported that the histone-binding protein RbAp48/46 can bind to PRC2 and by this optimize the enzymatic activity of PRC2 [19].

In summary, PRC2 is important for the establishment of H3K27me3, a repressive mark leading to transcriptional shutdown of target genes. Recognition of H3K27me3 by the PRC2 complex is essential for proper function. In concordance, H3K27me3 marks stimulate PRC2 activity. Recent studies have demonstrated that both EZH2 and EED can bind specifically H3K27me3 nucleosomes and that this read-out is responsible for the propagation of the repressive chromatin mark [23, 24].

1.2.2 PRC1

The core components of mammalian PRC1 are RING1A/1B, mPH, BMI1 and CBX [25-31]. PRC1 can recognize and bind H3K27me3, which is thought to be a repressive histone modification.



Through its chromo-domain the CBX protein can mediate recognition and binding of H3K27me3 [32]. Due to this it is thought that PRC1 works downstream of PRC2. Nonetheless, PRC1 has also been reported to target genes independently of PRC2 [33].

Figure 2 Vertebrate PRC1 core components.

In addition, the RING1A/1B protein contains a ring-domain and by this it can monoubiquitylate Lysine 119 of Histone 2A (H2AK119ub1) [34]. This posttranslational

modification of H2A has been shown to influence the RNA Pol II activity. *In vitro* studies have demonstrated that H2AK119ub1 can block RNA Pol II activity leading to transcriptional arrest [35, 36].

Furthermore, Francis and colleagues have shown that the PRC1 core complex can lead to chromatin compaction *in vitro* [37]. Moreover, PRC1 is believed to stabilize nucleosomal arrays at its target genes by blocking chromatin remodeling enzymes, which would normally allow access to condensed DNA sections and repressed genes [29, 37, 38]. These data endorse the relevance of the PRC1 complex in the maintenance of gene repression.

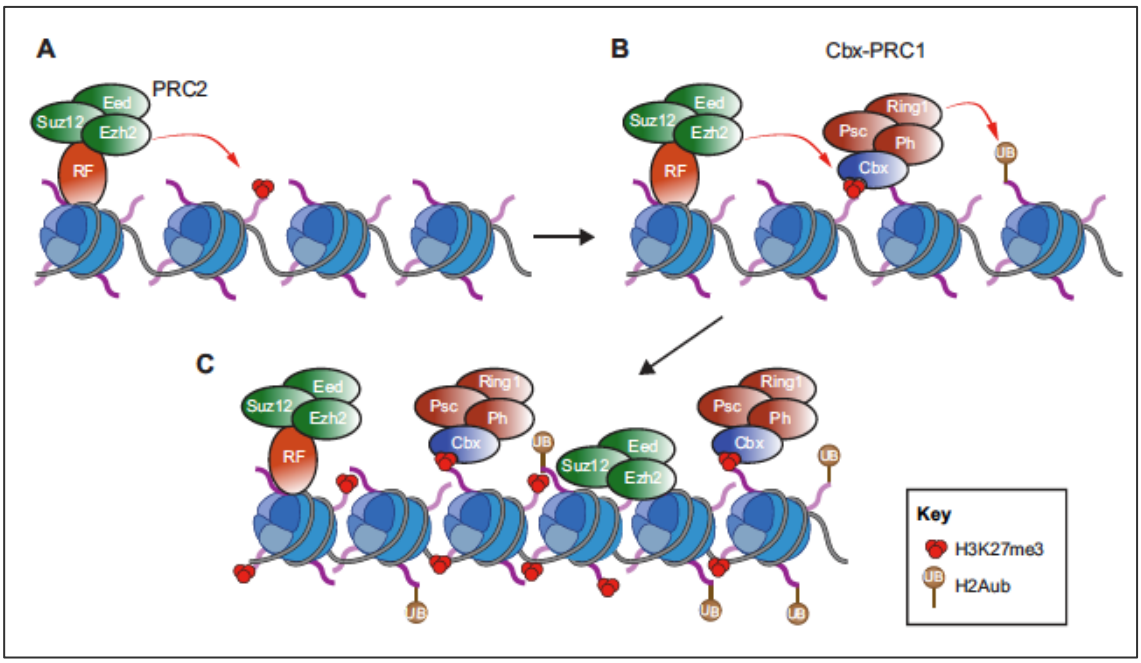


Figure 3 Mechanisms of PRC1 and PRC2.

(A) Deposition of H3K27me3 marks by PRC2. (B) Recognition and binding of H3K27me3 by the PRC1 protein CBX and (C) ubiquitination of H2AK119. RF = Recruiting factors such as Jarid2 target PRC2 to specific loci. Figure adapted from [39].

1.3 HISTONE MODIFICATIONS

In this section I especially want to present the H3K27me3 histone mark, which is important in PcG protein mediated gene silencing. For better understanding I will also include acetylation of histone residues and H3K36me3 as these marks have functions, which are antagonistic to those of most histone methylation marks. I will not cover any other histone modifications, since these were not the subjects of this thesis.

1.3.1 Histone acetylation and H3K27me3

Post-translational histone modifications can change the chromatin structure in such a way that the interaction between DNA and histones becomes loose, allowing other proteins to access the DNA and transcribe or repair it. This is the case for acetylated histone marks, such as acetylation of H3K27 (H3K27ac). Acetylation of histone residues is equivalent to the addition of negatively charged residues to histones, which overall destabilize the interaction between the negatively charged DNA and the positively charged histones [40]. Actively transcribed genes are associated with histones, which are hyper-acetylated. This is also the reason why acetylation of histones is known to be an active mark. In some cases, the chromatin structure might be changed resulting from the replacement of one histone by another variant (e.g. H3 becomes replaced by H3.3 [41]) or the recruitment of other regulatory proteins.

The PRC2 core protein EZH2 catalyzes di- and tri-methylation of H3K27 as mentioned above. H3K27me3 is an important histone modification, which correlates highly with repressed genes and therefore is thought to be a repressive mark [17]. PRC2 can bind to this histone mark via EED, one of its core subunits, which in turn stimulates the activity of the complex [42].
As a result of post-translational modification, H3K27 can be mono- di or tri-methylated. Methylation of H3K27 occurs gradually, where di-methylation results from H3K27me and tri-methylation from H3K27me2 [14]. It has been proposed that H3K27me2 is an important intermediary product because it might prevent acetylation of H3K27 and thus chromatin will rather stay compact and ready to be methylated further resulting in H3K27me3 [14, 17].
Both H3K27me2 and H3K27me3 are associated with facultative heterochromatin (fHC), which is regulated in a developmentally specific manner [43], whereas mono-methylated H3K27 is associated with constitutive heterochromatin (cHC) [44]. In

general, heterochromatin is associated with condensed chromatin architecture and low levels of transcription. For example, repetitive genomic sequences, centromeres and telomeres are considered as cHC. These regions are heterochromatinized stably although moderate transcription at these loci is still possible [45, 46]. In contrast to cHC, fHC has a more diverse chromatin structure and is sometimes found to be as condensed as it is in cHC, e.g. in the case of the inactive X-chromosome. Still, at specific loci such as promoters of inactive genes it has also been reported to be indistinguishable from euchromatic regions, which have a loose chromatin architecture and are associated with actively transcribed gene regions [43].

Zee *et al*. have shown that tri-methylation of H3K27 is a stable mark with a low turn-over compared to other methylated histone residues [47]. It is believed that the recruitment of regulatory factors, such as the PRC1 subunit CBX or the PRC2 subunits EZH2 and EED is crucial for the establishment of H3K27me3. The chromo-domain of the PRC1 subunit CBX binds H3K27me3 specifically and its affinity to this mark increases with the augmented abundance of H3K27me3 [48]. Moreover, it has been demonstrated that H3K27me3 and H3K36me3, an active mark associated with transcriptional elongation, are located at different loci [49]. This again suggests that H3K27me3 is associated with repressed and transcriptionally silent genetic regions whereas the active H3K36me3 mark is rather associated with transcriptionally active regions, such as transcription start sites.

1.4 POLYCOMB RESPONSE ELEMENTS (PREs)

PcG proteins act through cis-regulatory DNA elements called Polycomb response elements (PREs), to maintain gene expression patterns. In *Drosophila*, PREs contain multiple short DNA motifs. Several sequence-specific DNA binding proteins recognize these motifs and recruit PcG proteins to these sites. Thus, in *Drosophila* a PRE carries the information to recruit PcG or TrxG proteins. These DNA motifs have been used to predict computationally and identify further PREs [50].

In the past years, transgenic studies in flies have revealed that PREs can maintain the gene expression pattern that is established by an adjacent enhancer through many cell divisions even though the initial activating or repressing factors have gone [6, 7]. Due to this, they are often described as epigenetic memory elements. Furthermore, Cavalli and Paro have been able to show that PREs are switchable elements, which can

switch between an active and a silent state upon an experimentally induced change of the promoter status [51]. To date, PREs are best characterized in the *Drosophila* Hox-complexes. At this locus some PREs are located kilobases away from their promoters and form higher-order chromatin interactions as they interact through a looping-form with repressed target genes [31]. In transgenic *Drosophila* assays three functional hallmarks of a PRE have been characterized: (1) Silencing of a reporter gene at an ectopic locus, (2) loss of silencing of the reporter gene upon genetic PcG depletion and (3) recruitment of PcG proteins to the PRE at the ectopic locus [5, 52].

The problem of identifying mammalian PREs is compounded by the fact that the molecular mechanisms underlying mammalian PcG recruitment are not fully understood. The recruitment of PcG to PREs is more complex in mammals compared to flies. Mammalian PcG protein are involved in many different tissue-specific processes throughout development [52]. Another obstacle in identifying vertebrate PREs is the fact that PcG protein mediated H3K27me3 tends to spread over large domains, including the PRE itself, the promoter and the gene body [53]. This makes it difficult to identify PREs solely from PcG binding profiles.
Recently several vertebrate PREs, which share functional characteristics reminiscent of *Drosophila* PREs have been identified [54-56] and (Heinen *et al.*, manuscript in preparation). Although the Polycomb repressive complexes are conserved between *Drosophila* and vertebrates, there is only one DNA binding protein that seems to have a conserved function in PcG recruitment in both flies and mammals, namely the YY1 (*Drosophila* homologue PHO; GCCAT) protein [57]. It has been shown that YY1 rescues gene silencing when introduced into PHO-deficient *Drosophila* embryos [58]. In a recent study Woo and colleagues analyzed an intergenic region of the human HOXD cluster for PcG protein association [55]. The identified region is called D11.12 and was shown to repress a reporter gene and to associate with PRC1 and PRC2 proteins in human mesenchymal stem cells (MSC). The 1.8 kb region contains a cluster of YY1 binding sites and a highly conserved region of 237 bp. Mutations of the YY1 binding sites led to a partial loss in reporter gene repression and impede binding of BMI1 at this locus, but not SUZ12 binding. In contrast, a deletion of the highly conserved sequence resulted in a loss of repression and PRC1 and PRC2 binding. This led to the conclusion that YY1 binding sites might be involved in PRC1 recruitment whereas the highly conserved region is important for both PRC1 and PRC2 recruitment. Another study performed by Mendenhall *et al*. demonstrated that YY1 binding sites are not necessary for PRC2 recruitment. Moreover, the authors point out that there is no overlap between YY1 binding sites and PRC1 or PRC2 at a global level.

Instead, these sites co-localize with genomic sites which are marked by H3K4me3 only [59]. YY1 might not be involved in PRC2 recruitment directly though it might interact with other regulatory proteins and with PRC1 and impact PcG mediated gene silencing this way.

Other authors try to find similarities in the recruiting mechanism of PcG proteins to the PREs by introducing presumptive vertebrate PREs into the Drosophila genome. Sing and colleagues have identified a putative vertebrate PRE called the PRE-*kr* [54]. It regulates the expression of the *Maf-B/Kreissler* gene in mice rhomdomeres. The endogenous PRE-*kr* is associated with PRC1 and PRC2 components and reporter constructs have been shown to recruit PcG proteins as well. The analysis of the PRE-*kr* has shown, that the element contains YY1 and GAF (GAGAG) binding sites but their involvement in PRC1 and PRC2 recruitment has not been investigated. Instead, the PRE-*kr* was introduced into *Drosophila* where it has been shown to mediate reporter (mini-white) silencing. Moreover, the putative regulatory elements recruited PcG proteins to polytene chromosomes. Thus, the authors show that the PRE-*kr* acts as a PRE in *Drosophila* as well [54].
In *Drosophila*, somatic chromosomes are paired. This pairing is important for the increase in silencing in flies. This phenomenon is called pairing sensitive silencing (PSS) and it is a good indication of PRE activity. Cuddapah *et al.* have identified a potential human PRE in resting T-cells [60]. This 3 kb long PRE (SLCA17-PRE) contains YY1 and GAF binding sites and was shown to act as a pairing-sensitive silencer in *Drosophila* as well. These data indicate that PcG proteins might be recruited to vertebrate PREs in a similar way as in *Drosophila*. However, the fact that extrinsically introduced vertebrate elements are silenced in the *Drosophila* genome might be a protection and defense mechanism to shut down unknown DNA.

The intense search for vertebrate PREs in the past few years has brought interesting details on PcG proteins and their associated regulatory proteins und nucleic acids to light. However, these studies have also revealed a high level of complexity and many potentially redundant mechanisms.

1.5 NON-CONDING RNAS IN PCG PROTEIN RECRUITMENT

In mammals, the identification of PREs involves more than the characterization of respective DNA binding proteins. Recent studies have revealed that PcG proteins might be attracted to their sites of action not only through DNA binding proteins, but also through non-coding (nc) RNAs [14, 31], reviewed in [61]. Several authors have shown that the PRC2 protein SUZ12 co-immunoprecipitates and interacts with ncRNAs [62, 63]. In addition, Zhao *et al.* have demonstrated that EZH2 interacts with ncRNAs *in vivo* [62, 64].

The shutdown of the female X-chromosome involves the expression of the ncRNA *Xist*, which coats the Xi and leads to trimethylation of H3K27 and thereby to heterochromatization [42]. It has been shown that PRC2 can interact with the 1.6 kb long ncRNA RepA transcribed from *Xist* [64]. Additionally, it has been reported that the large intergenic non-coding (linc) RNA HOTAIR, which is transcribed from the mammalian HOXC locus, can bind and target the PRC2 complex to the HOXD locus [65]. Furthermore, it has been demonstrated that HOTAIR can bridge the PRC2 complex to LSD1, a demethylase that catalyzes demethylation of H3K4me2, an active histone mark, and that is required for gene repression in *Drosophila* [66]. Moreover, the authors observed that a knockdown of the lincRNA HOTAIR led to decreased binding of SUZ12 and LSD1 on the HOXD locus and also to a loss of H3K27me3 but to an increase in H3K4me2. Strikingly, overexpression of HOTAIR led to the conclusion that a CG-rich motif, which plays an important role in PRC2 binding [53, 67], is involved in HOTAIR targeting [66]. In this context it seems likely that ncRNAs, which are crucial in imprinted gene regulation and X-inanctivation are also associated with silencing of protein-coding genes through the PRC2 complex [68].

Genes that are repressed by PcG proteins are very often associated with H3K4me3 and RNA Pol II at transcriptional start sites and promoters, which indicates a potential to initiate transcription at these loci [2, 35]. Due to this, Kanhere and colleagues wanted to examine whether short transcripts resulting from these regions were associated with PcG proteins. Indeed they found that regions generating short ncRNAs but no mRNAs are enriched for H3K27me3 and interact with PRC2 through a stem-loop structure, reminiscent of the interaction between *Xist* and PRC2 [63, 64].

1.6 TRANSCRIBED INTERGENIC POLYCOMB (TIP) TARGET SITES

Transcribed intergenic Polycomb target sites, also called TIP sites or TIPs, represent a novel class of vertebrate Polycomb target sites and were described for the first time by Hekimoglu-Balkan and colleagues [1]. These sites are, as their name suggests, transcribed during differentiation and they recruit PcG proteins to the sites of transcription, also in a developmentally regulated manner. It has been reported previously that PREs can be transcribed into ncRNAs in both flies and mammals in a developmentally regulated manner [61, 69]. Transcription from these sites can regulate PcG function and recruitment [70]. Another characteristic of TIP sites is that they are more conserved than intronic sequences, though less conserved than protein coding genes. Furthermore, it was shown that only a small percentage of TIP sites contain predicted CpG islands. This indicates that CpG islands are not required for PcG recruitment to these sites.

In their study, Hekimoglu-Balkan *et al.* performed ChIP analyses at different stages of neural differentiation and found that TIP sites recruit PcG proteins to their site of transcription. In contrast to coding regions, H3K27me3 and SUZ12 enrichment profiles overlapped precisely with regions of TIP site transcription (Figure 1). These results build up a basis on which motif based computational PRE prediction can be performed, as there is no PcG spreading at TIP sites.

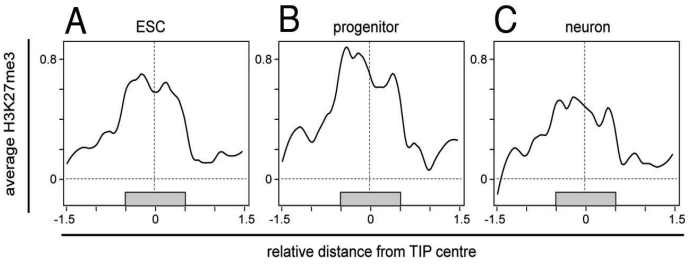


Figure 4 Average H3K27me3 enrichment profiles in ES cells during differentiation overlap precisely with sites of TIP site transcription [1].

In addition to PcG protein enrichment profiles, reporter assays were performed in transient transfection assays on selected sites. These studies showed that TIP sites repressed transcription of a reporter gene independent of the direction of transcription from the TIP site itself. Moreover, the data indicate that the transcript from one TIP site (TIP site 3, also called Nkx2-9 TIP site) is required for full repression of the reporter gene [1].

1.6.1 Utf1 TIP site

The Utf1 TIP site is located on the murine chromosome 7 [1, 71]. In ES cells, transcription from the Utf1 TIP site was detected but no enrichment in H3K27me3 (Figure 5). In neurons, no transcription from the TIP site was detected and H3K27me3 levels increased, whereas H3K4me3 levels decreased [1].
The corresponding Utf1 gene, also located on chromosome 7, has one transcript (Ensembl) and is expressed in ESCs.

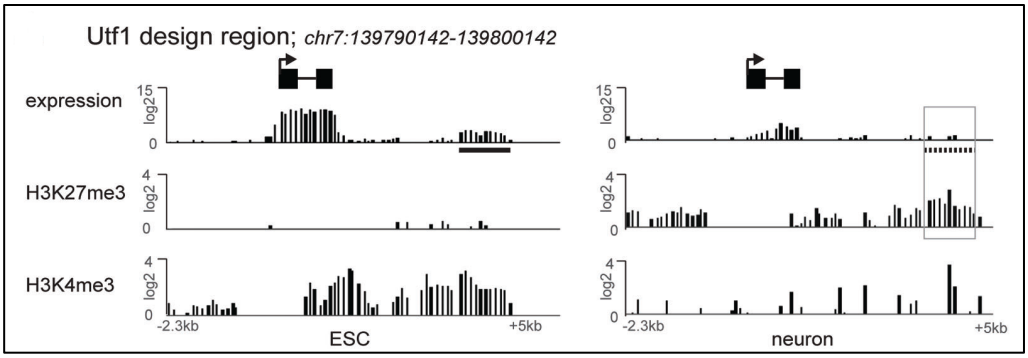


Figure 5 Expression and ChIP tracks for the Utf1 design region [1].
The exon-intron structure of the corresponding Utf1 gene is shown on top. The intergenic transcript is indicated as expressed, when the bar below the TIP site is solid. The intergenic transcript is not expressed if the bar below the TIP site is dashed. The grey box indicates an overlap between TIP site transcription and H3K27me3 enrichment.

1.6.2 Nkx2-9 TIP site

The Nkx2-9 TIP site is located on the murine chromosome 12 [1, 71]. In ES cells, both transcription from the TIP site and H3K27me3 enrichment is detected (Figure 6). Upon differentiation into neurons, transcription from the Nkx2-9 TIP site is not detected and histone methylation modifications change in the same way as for the Utf1 TIP site; H3K27me3 is more enriched whereas H3K4me3 is reduced [1].
The corresponding gene has one transcript (Ensembl) and is expressed in ESCs. The expression level of the gene decreases upon differentiation into neurons.

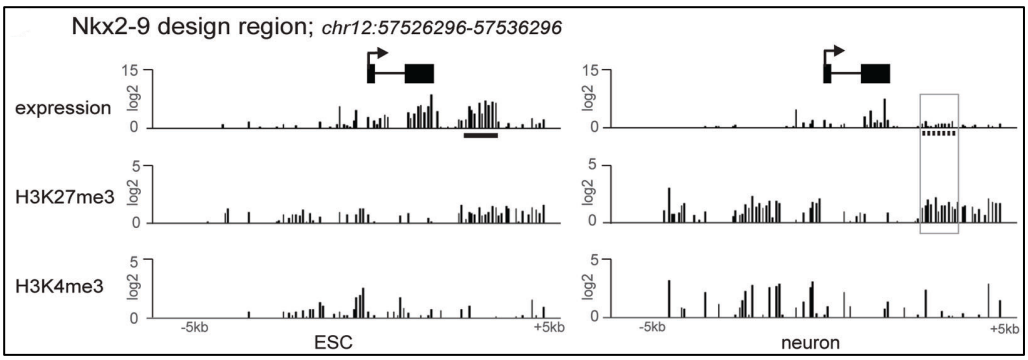


Figure 6 Expression and ChIP tracks for the Nkx2-9 design region [1].

The exon-intron structure of the corresponding Nkx2-9 gene is shown on top. The intergenic transcript is indicated as expressed, when the bar below the TIP site is solid. The intergenic transcript is not expressed if the bar below the TIP site is dashed. The grey box indicates an overlap between TIP site transcription and H3K27me3 enrichment.

1.6.3 Meis1 TIP site

The Meis1 TIP site is located at the murine chromosome 11 [1, 71]. No transcription from this TIP site is detected in ES cells and H3K27me3 levels are elevated compared to those in neurons [1]. Upon differentiation, H3K4me4 marks are enriched. In addition, the TIP site is transcribed and expression from the corresponding gene is detected as well (Figure 7). The Meis1 gene has 8 transcripts (Ensembl).

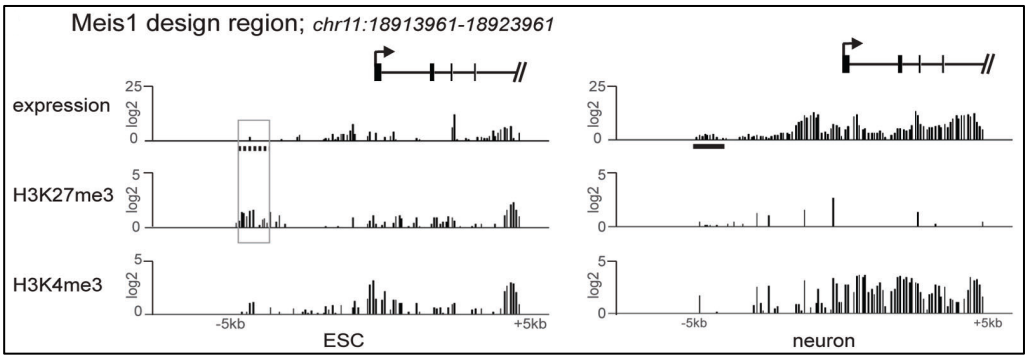


Figure 7 Expression and ChIP tracks for the Meis1 design region [1].

The exon-intron structure of the corresponding Meis1 gene is shown on top. The intergenic transcript is indicated as expressed, when the bar below the TIP site is solid. The intergenic transcript is not expressed if the bar below the TIP site is dashed. The grey box indicates an overlap between TIP site transcription and H3K27me3 enrichment.

Overall, the data demonstrate that TIP sites can target PcG proteins to their site of transcription. Presumably due to PcG recruitment, repression of transcription from a reporter gene was achieved. TIP sites have a very low or no CpG island content and

they are transcribed. For one TIP site (TIP site 3 or Nkx2-9 TIP site) it was shown that its transcription is necessary for full reporter gene repression. Nonetheless, a TIP site can only be regarded as a vertebrate PRE, if these characteristics can be maintained and the data reproduced once the TIP sites have been integrated into the genomic DNA at an ectopic locus.

2 AIM OF THIS THESIS

The aim of this thesis is to contribute to the understanding of PcG protein regulated gene expression in a mammalian system using murine ES cells. This work is especially based on studies performed by Hekimoglu-Balkan and colleagues, which for the first time described PcG protein recruitment to transcribed intergenic sequences (TIP sites) in mammals [1]. Since this study used transient assays in murine ES cells, we have made it our goal to understand the functional properties of TIP sites upon stable integration into chromatin of murine ES cells. Central questions were whether the TIP sites are able to repress transcription of a reporter gene upon stable integration into the genomic DNA. And if so, whether this repression was dependent on PcG proteins. Furthermore we have investigated the role of ncRNA transcripts, which arise from the TIP sites, in PcG protein recruitment and PcG protein mediated gene repression.

Overall, the aim of this thesis is the analysis and characterization of possible vertebrate PREs.

3 MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 Chemicals

- 6x Loading Dye (Thermo Scientific)
- Agarose (Sigma-Aldrich)
- Ampicillin (Amresco)
- BSA, 10mg/ml (New England Bio Labs)
- DMSO (Sigma-Aldrich)
- Ethanol (Merck)
- GeneRuler DNA Ladder mix, ready to use (Thermo Scientific)
- Hygromycin B (Invitrogen)
- Iso-propanol (Sigma-Aldrich)
- Kanamycin (Sigma-Aldrich)
- Lipofectamine 2000 (Life Technologies)
- Opti-MEM (Life Technologies)
- Perfecthyb Plus (Sigma-Aldrich)
- Phenol:Chloroform:Isoamylalcohol 25:24:1 (Fluka)
- Radioactivity: (α-^{32}P)-dCTP, 3000 Ci/mmol, 250 µCi (Perkin Elmer)
- Sodium hydroxide (NaOH) (Fluka)
- SYBR Safe DNA gel stain (Invitrogen)

3.1.2 Kits

- Dual Luciferase Reporter Assay System (Promega)
- High Pure RNA Isolation Kit (Roche)
- illustra ProbeQuant G-50 Micro Columns Radiolabeled Probe purification Kit (GE Healthcare)
- Immobilon-Ny+ Membrane, Charged Nylon, 0.45 µm, 30 cm x 3.3 m Roll (Millipore)
- Kapa fast ready mix with dye (Peqlab)

- Kapa2G Robust PCR Kit (Peqlab)
- Luciferase Assay System (Promega)
- peqGOLD Hot Start Mix Y (Peqlab)
- Prime-It RmT Random Primer Labeling Kit (Agilent Technologies)
- PureYield Plasmid Midiprep System (Promega)
- QIAprep Spin Miniprep Kit (Qiagen)
- QIAquick Gel Extraction Kit (Qiagen)
- Quick Ligation Kit (New England Biolabs)
- SYBR Green Jump Start Taq Ready Mix for Quantitative PCR (Sigma)
- Transcriptor Reverse Transcriptase (Roche)
- TURBO DNA-Free (Ambion)
- Whatman Chromatography Paper 46 cm x 57 cm (Schleicher and Schuell)

3.1.3 Standard Buffers and Solutions

1x PBS:

8g NaCl (137mM), 200mg KCl (2.7mM), 1.44g Na_2HPO_4 (10mM), 240mg KH_2PO_4O (2mM), add 1L dH_2O

Prepared by the IMBA service department

1x TE:

10mM Tris, 1 mM EDTA add 1L dH_2O, adjust pH with acetic acid

Prepared by the IMBA service department

10x TAE:

48.4g Tris (0.4M), 3.7g EDTA (0.01M), add 1L dH_2O, adjust pH with acetic acid

Prepared by the IMBA service department

5M NaCl:

292.2g NaCl (5M), add 1L dH_2O

Prepared by the IMBA service department

3M NaAc (pH 5.2):

408.1g 3M $NaAc3H_2O$, add 1 L ddH_2O, adjust pH with acetic acid

Prepared by the IMBA service department

LB media (pH 7.5):

5g NaCL, 10g Tryptone, 5g Yeast extract, add 1 L dH_2O, adjust pH with NaOH

Prepared by the IMBA service department

LB Amp^R agar plates (pH 7.5):

16g Agar, 5g NaCl, 2g Tryptone, 5g Yeast extract, adjust pH with NaOH

Prepared by the IMBA service department

SOB media:

20 g Tryptone, 5g Yeast Extract, 2ml 5M NaCl, 2.5 ml 1M $MgCl_2$, 10ml 1M MgSO4, add 1L dH_2O, sterilized by autoclaving

Prepared by the IMBA service department

SOC media:

0.9 ml 20% glucose, 0.5 ml 1M $MgCl_2$, add 50 ml SOB media

20x SSC:

175.3 g NaCl, 88.2 g Na3C6H5O7•2H2O (= sodium citrate), add 1L dH2O, adjust pH with HCl

Prepared by the IMBA service department

Low stringency buffer:

2x SSC, 0.1 %SDS, add 1L dH2O

Ultra high stringency buffer:

0.5x SSC, 0.1 % SDS, add 1L dH2O

Depurination solution:

0.25 M HCL: 25 ml of 37 % HCL add 1 L dH_2O

Denaturation solution:

0.5 N NaOH (20.0 g NaOH), 1.5 M NaCl (87.7 g NaCl), add 1L dH2O to final volume

Neutralization solution:

500 ml 1M Tris pH 7.5, 300 ml 5M NaCl

3.1.4 Buffers, solutions and media used for the cell culture

+/+ media:

10% FCS (Gibco)
1x L-glutamine (Sigma-Aldrich, 100x, 200mM))
1x Penicillin/Streptomycin (Sigma-Aldrich, 100x)
1x Non-Essential Amino Acids (Sigma-Aldrich, 100x)
1x Sodium Pyruvate (Sigma-Aldrich, 100x, 100mM)
0.1mM β-mercaptoethanol (Merck)
1000U LIF (Esgro)
add 500 ml DMEM (prepared by the IMBA service department)
stored at 4°C up to one week

+/- media:

10% FCS (Gibco)
1x L-glutamine (Sigma-Aldrich, 100x, 200mM))
1x Penicillin/Streptomycin (Sigma-Aldrich, 100x)
1x Non-Essential Amino Acids (Sigma-Aldrich, 100x)
1x Sodium Pyruvate (Sigma-Aldrich, 100x, 100mM)
0.1mM β-mercaptoethanol (Merck)
ad 500 ml DMEM (prepared by the IMBA service department)
stored at 4°C

Freeze media:

50% FCS (Gibco), 40% +/+ media, 10% DMSO (Sigma-Aldrich)

1x Trypsin:

Dilution prepared using 10x Trypsin (Gibco)

0.1% Gelatin solution:

Prepared using 2% Gelatin (Sigma), add 500 ml 1x PBS

Cell lysis buffer:

0.1 M EDTA (pH 8.0), 10 mM NaCl, 0.5% N-Lauroylsarcosine, 10 mM Tris-Cl (pH 8.0)

3.1.5 Competent E. coli strain

DH5α, F⁻_80*lac*ZΔM15 Δ*pho*A gyrA96
Provided by the IMBA service department

3.1.6 Murine Embryonic stem cells

Murine 46c ($Sox1^{GFP}$) embryonic stem cells were used for all cell culture experiments in the course of this study [1], [72].

3.1.7 Plasmids

pGL4.26 TK hygrofloxed containing a hygromycin coding region flanked by loxP sites, a TK promoter, a luciferase reporter gene (luc2) and a *beta-lactamase* (Amp^R) coding region. The basic plasmid was obtained from Promega and modified by Robert Heinen and Tanja Drexel.

Rosa26 contains a beta-lactamase (Amp^R) coding region and homologous arms targeting the murine gene trap Rosa26 locus on chromosome 6 [73].

pRL-TK is a wildtype *Renilla* luciferase (Rluc) control reporter vector. It contains a TK promoter and a *beta-lactamase* (Amp^R) coding region. The vector was obtained from Promega.

pEGFP-N3 is a GFP reporter gene containing a Kanamycin resistance. The Plasmid was obtained from BD Biosciences Clontech.

CRE contains a Cre recombinase and a Kanamycin resistance gene. It was obtained from Vector Biolabs.

3.1.8 Enzymes

- AscI (Thermo Scientific, 5 U/µl)
- AsiSI (New England Biolabs, 10 000 U/ml)
- BamHI (Thermo Scientific, 10 U/µl)

- HindIII (Thermo Scientific, 10 U/µl)
- FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific, 1 U/µl)
- PacI (Thermo Scientific, 10 U/µl)
- Psp1406I (AclI) (Thermo Scientific, 10 U/µl)
- SacI (Thermo Scientific, 10 U/µl)
- XhoI (Thermo Scientific, 10 U/µl)
- XmaI (New England Biolabs, 10 000 U/ml)
- Proteinase K (Sigma-Aldrich)

3.1.9 Oligonucleotides (5' – 3')

Oligo Name	Sequence (5' - 3')	Purpose
pGl4.26mcsfwd	AGTGCAGGTGCCAGAACATT	Sequencing, Colony PCR
pGl4.26mcsrev	AACAGTACCGGATTGCCAAG	Sequencing
ROSA26mcs-fwd	GAGTTCTCTGCTGCCTCCTG	Sequencing, Colony PCR
ROSA26mcs-rev	TCTTCTGGGCAGGCTTAAAG	Sequencing, Colony PCR
Luc2-seq-fwd	GCTACTTGATCTGCGGCTTT	Sequencing
Luc2-seq-rev	AAAGCCGCAGATCAAGTAGC	Sequencing
FlpTK-eff-LCb	CAAACCCTAACCACCGCTTA	Sequencing, Colony PCR
pGL4.26mcsdo_test a	GGTCATGGCCGCAATAAAAT	Sequencing
Rosa diagA	TCTACCAGAGCCTCGTGGAC	Diagnostic PCR
Rosa dcloningB	GCTAGTCGACCACCGCCCCACACT TATT	Diagnostic PCR
Ro-diag-luc-2a	GGACTTGGACACCGGTAAGA	Amplification of internal probe for southern blot
Ro-diag-luc-2b	GTCCACGAACACAACACCAC	Amplification of internal probe for southern blot
Ro-diag-mm-s3a	TTTGCCTGGGTATTGCCTAC	Amplification of external probe for southern blot
Ro-diag_mm-S3b	CCTAGGATCTTGGCTTGCAC	Amplification of external probe for southern blot
Ro_dTg2a	ACGTTTCCGACTTGAGTTGC	PCR to confirm excision of hyg^R coding region
pGl4.26 Lucpa	AGATCCGCGAGATTCTCATT	PCR to confirm excision of hyg^R coding region

Table 1 Oligonucleotides (5' – 3')

3.1.10 Primer sequences for qPCR of transcripts from TIP sites and adjacent genes

Name of targeted transcript	Forward primer (5' - 3')	Reverse primer (5' - 3')
Gapdh	TGCGACTTCAACAGCAACTC	CTTGCTCAGTGTCCTTGCTG
Suz12	GACAGAAGCCAGAGACGACCT	TTCCTGCATAGGAGCCATCAT
Ezh2	GTGACAGAGAAGCAGGGACTG	CACCACTCCACTCCACATTCT
Utf1 TIP site	AGTTTCCCTCTTCTCCACAGC	GGGTTGTCATGGGTCTTCTCT
Nkx2-9 TIP site	TGCCTAAAGCCACCCTAGACT	AAAATCTTGCTCTGGGACTGG
Meis1 TIP site	GTTCTCTGCCCTCCAAATACC	GCCAGCATTCCTCCATTATCT
Utf1 gene	CCTCCTCTCTGGTGAGGCCA	GCATCAGACGTAGTCACCGGG
Nkx2-9 gene	TTCCCAGCCTACCAGCACTTA	CTTTGGTGTGGCTGTAGCTCT
Meis1 gene	TCTGCCACCGGTATATTAGCTG	TGAAAGTCGCATACCTGGTCA

Table 2 Primer sequences for qPCR

3.1.11 RNA Interference

Name of RNAi	Sequence	Company
Non-targeting siRNA #1	UGGUUUACAUGUCGACUAA	Dharmacon, ON-TARGET plus (D-001810-01-05)
Non-targeting siRNA #3	UGGUUUACAUGUUUUCUGA	Dharmacon, ON-TARGET plus (D-001810-03-05)
Non-targeting siRNA #4	UGGUUUACAUGUUUUCCUA	Dharmacon, ON-TARGET plus (D-001810-04-05)
Ezh2 siRNA #5	GCACAAGUCAUCCCGUUAA	Dharmacon, ON-TARGET plus Set of 4 (J-040882-05)
Ezh2 siRNA #6	CAGAGAAUGUGGAUUUAUA	Dharmacon, ON-TARGET plus Set of 4 (J-040882-06)
Ezh2 siRNA #7	GGGAUGAAGUUCUGGAUCA	Dharmacon, ON-TARGET plus Set of 4 (J-040882-07)
Ezh2 siRNA #8	GGUAAAUGCUCUUGGUCAA	Dharmacon, ON-TARGET plus Set of 4 (J-040882-08)
Suz12 #5	UUACAUGUCUCAUCGAAAU	Dharmacon, ON-TARGET plus Set of 4 (J-040180-05)
Suz12 #6	GGAUAGAUGUUUCAAUCAA	Dharmacon, ON-TARGET plus Set of 4 (J-040180-06)
Suz12 #7	GUAAAGAGAACACCUAUCA	Dharmacon, ON-TARGET plus Set of 4 (J-040180-07)
Suz12 #8	CAACUUACAUUUACCGGUU	Dharmacon, ON-TARGET plus Set of 4 (J-040180-08)

Table 3 RNA interference

3.1.12 Locked Nucleic Acids

Name of LNA	Sequence (5' - 3')	Company
Scrambled	GACGGTAACYAGGCGA	EXIQON
Utf1 antisense2	GGTGCTGGGAAATGTGAG	EXIQON
Nkx2-9 antisense1	CCTCACCCTTGTTAAATC	EXIQON

Table 4 Locked nucleic acids

3.1.13 Software

- CLC Main Workbench (CLCbio, version 6.6.2)
- Gene Construction Kit (Textco, version 3.5.3)
- NCBI Primer designing tool (http://www.ncbi.nlm.nih.gov)
- Webcutter 2.0 (http://rna.lundberg.gu.se/cutter2/)

3.2 METHODS

3.2.1 Cloning TIP sites and the En2 PRE into the pGL4.26 TK hygrofloxed vector

TIP sites were cloned into the pGL4.26 vector containing the TK promoter and a hyg^R gene. Additionally, as a control for the expression level of the luc2 reporter gene, the En2 PRE was cloned into the pGL4.26 TK hygrofloxed vector. Previous TIP and PRE constructs contained a minimal promoter in front of the TIP site or the PRE and were provided by Betül Hekimoglu-Balkan and Robert Heinen.

TIP site	Enzyme 5'	Enzyme 3'	Fragment size (bp)
Utf1	SacI	SacI	3142
Nkx2-9	SacI	SacI	2758
Meis1	XhoI	XhoI	3510
En2	SacI	HindIII	1600

Table 5 Enzymes used for cloning of TIP sites and the En2 PRE into the pGL4.26 TK hygrofloxed vector.

In each reaction set-up, 2 µg of DNA were digested for 1 hour at 37°C with the appropriate amount of enzyme following the manufacturer's protocol. To prevent self-annealing of the digested vector backbone, FAST AP was added during the last 5 minutes of the restriction reaction.

Agarose gel electrophoresis was performed to separate the digested DNA (1% agarose gel, 100V, 30 min). According to their expected size, the bands were cut out and purified using the QIAquick Gel Extraction Kit. DNA concentration was determined using the UV/VIS Spectrophotometer Nanodrop ND-1000 (Peqlab) at 260 nm.
During the ligation process, a five-fold molar excess of the insert was combined with 50ng of the digested pGL4.26 TK hygrofloxed vector using the Quick Ligation Kit (New England Biolabs). 2 µl of the ligation mixture were transformed in 80 µl of competent DH5α E.coli cells and incubated on ice for 10 minutes. Heat shock was applied at 42°C for 50 seconds. Subsequently the cell suspension was placed on ice for another 5 minutes. 750 µl of SOC media were added and the cell suspension was incubated for 1 hour at 37°C, shaking vigorously. The cell suspension was plated on LB Amp^R agar plates and incubated over night at 37°C.

Colony PCR was performed to indentify clones positive for the constructs of interest using the peqGOLD Hot Start Mix Y (Peqlab) and 50 µM of each primer with the following program: 4min at 94°C, 40 cycles of 30sec at 94°C – 30sec at 61°C – 1min/kb at 72°C, 10min at 72°C.
The identified positive colonies were inoculated overnight in 5 ml LB media with added ampicillin at 37°C. Plasmid DNA was purified using the QIAprep Spin Miniprep Kit (Qiagen). The constructs of interest were validated by sequencing (IMBA service department).

3.2.2 Cloning TIP sites and the En2 PRE into the Rosa26 Vector

For the purpose of homologous recombination, TIP sites and the En2 PRE had to be cloned into the Rosa26 vector.

TIP site	Enzyme 5'	Enzyme 3'	Fragment size (bp)
Utf1	AscI	AsiSI	7890
Nkx2-9	AscI	AsiSI	7492
Meis1	AscI	AsiSI	8257
En2	AscI	AsiSI	6304

Table 6 Enzymes used for cloning of TIP sites and the En2 PRE into the Rosa26 targeting vector.

The restriction digest, ligation, transformation, colony PCR and sequencing were performed as described above.
Constructs of interest were inoculated overnight in 50 ml of LB media with ampicillin and plasmid DNA was purified using the PureYield Plasmid Midiprep System (Promega). As an additional control a restriction analysis was performed.

Plasmid DNA	Total size (bp)	Enzyme	Restriction sites	Position of sites (cuts)	Fragment size (bp)
Rosa26 TK hyg^R Utf1	17833	XmaI	3	5246, 6066, 8973	14106, 2907, 820
Rosa26 TK hyg^R Nkx2-9	17435	XmaI	2	5246, 8575	14106, 3329
Rosa26 TK hyg^R Meis1	18201	XhoI	4	5565, 9080, 15237, 15257	8509, 6157, 3515, 20
Rosa26 TK hyg^R En2	16247	XmaI	3	5246, 5714, 7387	14106, 1673, 468

Table 7 Enzymes used for the restriction analysis of the Rosa26 constructs.

3.2.3 Transfections

For transient transfections 1.75 x 10^5 46c murine ES cells were co-transfected with the pGL4.26 TK hygrofloxed plasmid DNA containing the TK promoter, hyg^R gene and the TIP sites or the En2 PRE, 600ng of EGFP vector (BD Biosciences Clontech) and 25ng of pRL-TK vector (Promega) using the Lipofectamine 2000 reagent (Invitrogen).

For siRNA transfections, 1.5 x 10^5 46c murine ES cells were co-transfected with 30 nM of double-stranded siRNA, 25ng of pRL-TK vector (Promega) and EGFP vector (BD Biosciences Clontech) to obtain a total DNA amount of 800ng using the Lipofectamine 2000 reagent (Invitrogen). The siRNAs targeted either mouse Suz12 mRNA or mouse Ezh2 mRNA.

For the LNA transfections, 1.5 x 10^5 46c murine ES cells were co-transfected with 10 nM of LNA, and EGFP vector (BD Biosciences Clontech) to obtain a total DNA amount of 600ng using the Lipofectamine 2000 reagent (Invitrogen). The LNAs targeted either mouse Nkx2-9 TIP site ncRNA or mouse Utf1 TIP site ncRNA.

For the excision of the hygromycin cassette in transgenic cell lines, 1.5 x 10^5 46c murine ES cells were co-transfected with 400ng of Cre and EGFP vector (BD Biosciences Clontech) to obtain a total DNA amount of 800ng using the Lipofectamine 2000 reagent (Invitrogen). Double-selection was performed with media +/+ with added hygromycin (200µg/ml).

Expression of the luc2 reporter gene was determined using the Dual Luciferase Reporter Assay System (Promega) or the Luciferase Reporter Assay System (Promega) following the manufacturer's protocol.

3.2.4 Stable Integration of TIP sites and the En2 PRE constructs into the murine Rosa26 locus

In order to obtain better targeting efficiency of the murine Rosa26 locus in 46c ES cells, the Rosa26 plasmid DNA containing the TK promoter, hyg^R gene and the TIP sites or the En2 PRE were linearized with restriction enzymes. Rosa26 plasmid DNA containing the Meis1 TIP site was linearized using the restriction enzyme Psp1406I, whereas Rosa26 plasmid DNA containing the Nkx2-9 TIP site, Utf1 TIP site and the

En2 PRE was linearized using the enzyme PacI. The linearized plasmid DNA was isolated by Phenol-Chloroform Isoamylalcohol (25:24:1) purification.

1 x 10^7 46c ES cells were trypsinized and resuspended in 800 µl 1x PBS. Using an electroporation cuvette (Gene Pulser, 0.4cm, Biorad), cells were electroporated with 20 µg of linearized plasmid DNA using the GenePulser Xcell (Biorad) under following conditions: Voltage (V) 230, Capacitance (µF) 500, Resistance (Ω) ∞, Cuvette (mm) 4.

After electroporation, 200 µl of 1x PBS were added to the cell suspension. 100 µl of the cell suspension (corresponds to 1 x 10^6 cells) were seeded on a 10 cm plate with +/+ media.
36 hours after seeding, cells were selected on +/+ media with hygromycin (200µg/ml) until the cell colonies were ready to be picked and expanded (app. 7 – 10 days).

Verification of 46c ES cells positive for the single, site-specific integration of the TIP or PRE construct was performed by diagnostic PCR and southern blot.

3.2.5 Diagnostic PCR

46c ES cells in a 96 well pate were lysed with 10 µl lysis buffer containing proteinase K (1mg/ml) and incubated for 1 hour at 65°C. Proteinase K inactivation took place for 10 minutes at 95°C.
A 100-fold mastermix was prepared as follows:

PCR grade water	1900 µl
5x Kappa2G robust GC buffer	500 µl
10 mM dNTP mix	50 µl
50 µM Rosa diagA primer	20 µl
50 µM Rosa dcloningB primer	20 µl
Kappa2G robust DNA polymerase (1U/ml)	10 µl

20 µl of the mastermix were added to the cell suspension into each well. PCR was performed under following conditions: 2min at 95°C, 20sec at 95°C, 35 cycles of 15sec at 58°C/2min at 72°C, 3min at 72°C.
The DNA was analysed using agarose gel electrophoresis.

3.2.6 Southern Blot

In order to verify correct integration at Rosa26, and to exclude multiple integrations of the constructs of interest into the genome of murine 46c ES cells, a southern blot was performed using the isotope ^{32}P.

Firstly, the DNA was isolated by lysis of 46c ES cells using lysis buffer and proteinase K (1mg/ml). DNA concentration was determined using the UV/VIS Spectrophotometer Nanodrop ND-1000 (Peqlab) at 260 nm. 7 µg of the genomic DNA were digested with the restriction enzyme HindIII and - in a separate reaction set-up - with the restriction enzyme BamHI. The genomic DNA was loaded on a 1% agarose gel along with the marker Gene Ruler DNA Ladder mix ready to use (Thermo Scientific).
The gel was run for 4 hours at 120 V.

After the gel was rinsed in dH_2O, the depurination solution was added and the gel was incubated for 20 minutes, shaking gently. The gel was rinsed in dH_2O and the denaturation solution was added. The gel was incubated for 30 minutes and then washed again with dH_2O. Finally, neutralization solution was added and incubated for another 30 minutes, shaking gently.

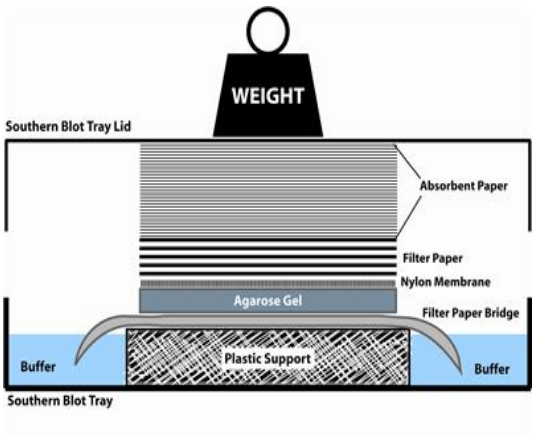


Figure 8 Schematic view of the platform for the southern blot
(http://www.gibthai.com/services/technical_detail.php?ID=17).

A platform for the transfer was constructed (for details see figure 8). The gel was placed upside down on the platform and covered with the Immobilon-Ny+ Membrane (Millipore). Three layers of Whatman Chromatography Paper (Schleicher and Schuell) were cut to match the size of the gel, wetted with 20x SSC buffer and placed on the gel. A stack of absorbing paper towels was placed on the 3 layers of Whatman paper. Additionally, a glass plate was placed on the paper towels and finally, a 0.5 kg weight was put on top of it. Blotting took place overnight at RT.

The nylon membrane was dried for 1 hour at 80°C and the DNA was fixed by cross linking using the Stratagene 1800 crosslinker.
The internal and external probes were labeled radioactively by using the Prime-It RmT Random Primer Labeling Kit (Agilent Technologies) following the manufacturer's protocol. The probes were labeled using (α-^{32}P)- dCTP (Perkin Elmer). The radio-labeled probes were purified using the illustra ProbeQuant G-50 Micro Columns (GE Healthcare).
The Perfecthyb Plus (Sigma-Aldrich) prehybridization buffer was added to the nylon membrane and incubated for 30 minutes at 65°C. Radioactively labeled probes were added to the prehybridization solution and incubated overnight at 65°C.
The nylon membranes were washed firstly with the low-stringency buffer and then with the ultra high-stringency buffer. The membranes were wrapped into Saran plastic film and placed into an exposure cassette overnight. Visualization was performed using the Typhoon Phosphoimager (Amersham).

Once stable cell lines were generated and site-specific insertion of the constructs of interest was validated by diagnostic PCR and southern blot, expression of the luc2 reporter gene was determined using the Luciferase Reporter Assay System (Promega).

3.2.7 cDNA Preparation and qPCR analysis

Total RNA from 46c murine ES cells was isolated using the High Pure RNA Isolation Kit (Roche) and after isolation treated with Turbo DNase (Ambion). RNA concentration was measured using the UV/VIS Spectrophotometer Nanodrop ND-1000 (Peqlab) at 260 nm. cDNA preparation was performed using the Transcriptor Reverse Transcriptase (Roche).
qPCR analysis was performed using SYBR Green Jump Start Taq Ready Mix for Quantitative PCR (Sigma) and Realplex mastercycler (Eppendorf) with the following program:

95°C	3min.	1x
95°C	15 sec.	
60°C	45 sec.	40x
72°C	30 sec.	
95°C	15 sec.	
60°C	15 sec.	1x
Dissociation analysis*	20 min.	
95°C	15 sec.	

4 RESULTS

4.1 CLONING OF TIP SITES INTO THE TARGETING VECTOR OF THE MURINE ROSA26 LOCUS

The initial pGL4.26 vector contains a TIP site or the En2 PRE, a TK promoter, the luciferase reporter gene (luc2) and a hygromycin resistance gene (hyg^R) for positive selection of ES cells. The hyg^R gene is driven by the SV40 promoter and flanked by loxP sites. The latter can be useful for the excision of the hyg^R gene once ES cells have been selected positively for the site-specific insertion of the transgene. Additionally the vector contains an ampicillin resistance gene (Amp^R). For the insertion or excision of the TIP site or the PRE, FRT-sites were cloned into the vector.

In order to exclude any effect from the FRT-sites on the regulation of the TIP sites or the En2 PRE, the TIP sites and the PRE were excised using restriction enzymes as described above (Table 5) and inserted into a pGL4.26 vector, resembling the initial vector. However, the new pGL4.26 vector does not contain FRT-sites (Figure 9B).

The Gt(ROSA)26Sor (ROSA26) locus in mice is 9 kb long and ubiquitously expressed in embryonic and adult tissues [74]. Over the past decade 130 mouse knock-in lines and the stable expression of single-copy transgenes in murine ES cells have proven this locus to be optimal for site-specific integration of transgenes. To achieve successful homologous recombination in ES cells, the constructs of interest were cloned into the corresponding Rosa26 targeting vector (TV).
The Rosa26 locus is an transcriptionally active locus and expressed moderately in ES cells [73]. However, we reasoned that the strength of the Rosa26 promoter might be too weak to drive the expression of the luc2 reporter gene [74]. To overcome this problem the viral TK promoter was inserted downstream of the TIP site or the PRE.

All constructs of interest containing a TIP site or the En2 PRE, the TK promoter, hyg^R gene and the luc2 reporter gene, as well as the Rosa26 targeting vector were digested with AscI and AsiSI restriction enzymes. The constructs of interest were ligated into the Rosa26 TV in between of two genomic Rosa26 fragments (Rosa26 5' – and 3' fragment), which drive the homologous recombination process (Figure 9C).

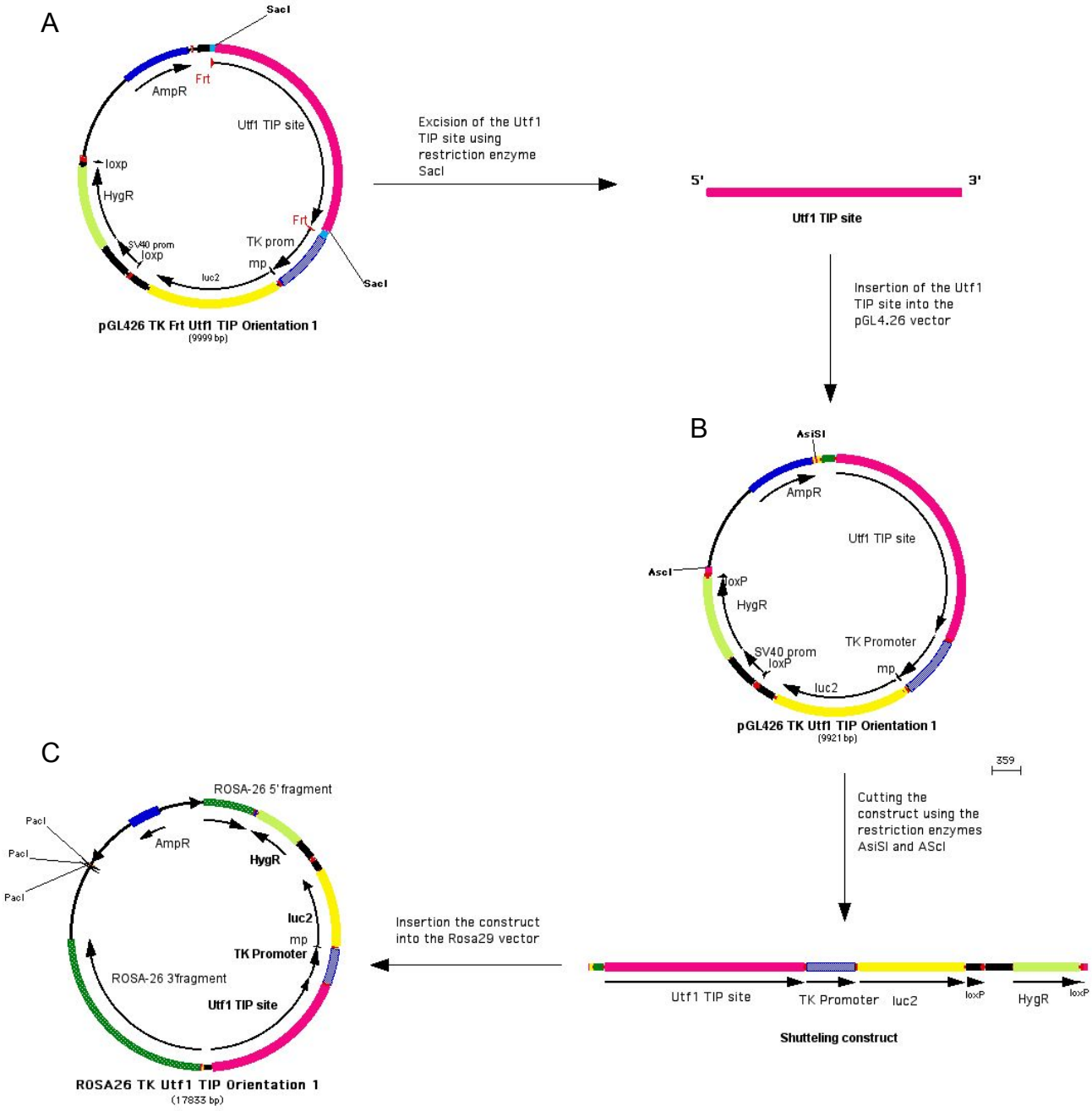


Figure 9 Cloning of the TIP site into the Rosa26 targeting vector.

The Utf1 TIP site is shown here as an example for the cloning of all TIPS sites and the En2 PRE. (A) The pGL4.26 vector with FRT-sites. (B) The pGL4.26 vector without FRT-sites. (C) The Rosa26 targeting vector.

4.2 TRANSIENT TRANSFECTIONS OF TIP SITES AND THE EN2 PRE SHOW A REPRESSIVE EFFECT ON THE LUCIFERASE REPORTER GENE IN MURINE ES CELLS

Previously it has been shown that the expression of the luc2 reporter gene, driven by the viral TK promoter, was down-regulated upon transient transfection of murine 46c ES cells with pGL4.26 constructs containing a TIP site and flanked by FRT-sites [1]. Repression was dependent on the presence of the TIP site. Furthermore, Robert Heinen investigated the repressive effect of the En2 PRE on the luc2 gene (Heinen *et al.*, manuscript in preparation). However, the constructs used by Robert Heinen contained a minimal promoter instead of the viral TK promoter upstream of the PRE. Since the repressive effect from the En2 PRE on the luc2 gene was known we used this PRE in combination with the TK promoter as a positive control for the luciferase assays. Furthermore, we wanted to investigate whether the En2 PRE will be able to down-regulate the expression of the reporter gene even though it is driven by the TK promoter.

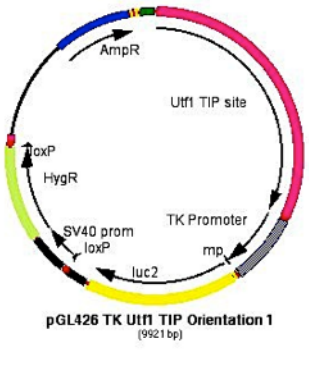


Figure 10 The Utf1 TIP site construct as an example for the vectors used for transient transfections of 46c ES cells.

In order to quantify the effect of the TIP site or the En2 PRE on the expression of the firefly luc2 gene, 46c ES cells were transfected transiently with the pGL4.26 vectors containing the TIP site or the PRE driven by the TK promoter (Figure 10). Concomitantly as a control for the transfection efficiency the Renilla luciferase was transfected. As an additional control constructs lacking the TIP site or the PRE were transfected.

Constructs containing the Nkx2-9 or the Utf1 TIP sites were able to repress the expression of the luc2 reporter gene in 46c ES cells upon transient transfection (Figure 11). In contrast, the Meis 1 TIP site did not repress the reporter gene. These results are consistent with previous studies [1].

Furthermore, the En2 PRE was able to repress the reporter driven by the TK promoter in this assay, confirming that the En2 PRE has a repressive effect even in combination

with this promoter, which drives high-level expression of the Luciferase reporter gene (Heinen *et al.*, in preparation).

Overall these data indicate that the expression of the luc2 gene is repressed upon transfection with TIP sites or the En2 PRE construct, whereas the Meis1 TIP site does not show a repressive effect.

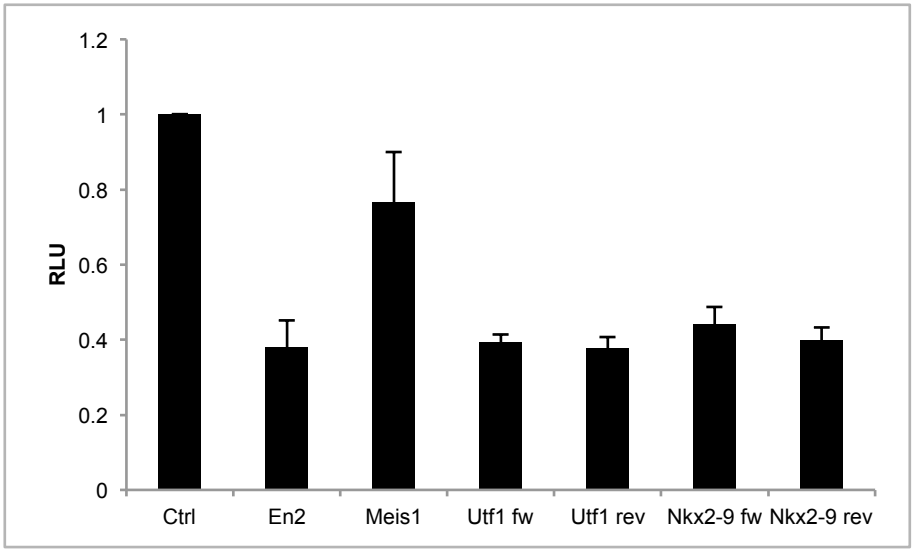


Figure 11 Expression levels of the firefly Luciferase reporter gene in 46c ES cells transiently transfected with pGL4.26 vectors containing constructs of interest.

Normalization of the data: Absolute values of the renilla luciferase were normalized on the pGL4.26 vector (Ctrl). The resulting product was multiplied with the absolute firefly luciferase values. RLU = Relative light units. (n=2)

4.3 VERIFICATION OF SITE-SPECIFIC, SINGLE INTEGRATION OF TIP SITES INTO THE MURINE ROSA26 LOCUS BY DIAGNOSTIC PCR AND SOUTHERN BLOT

In order to examine the effect of the TIP sites on the reporter upon stable integration into the genome, the constructs were targeted to the Rosa26 locus in murine ES cells.
After the electroporation of 46c ES cells, picked cell colonies were expanded and cultured for later analysis. In order to test whether these colonies contained the construct of interest at the right genomic locus (Rosa26) and if the insertion took place once as desired or multiple times, a diagnostic PCR and a southern blot were performed.
The diagnostic PCR was performed to preselect cell colonies, which were presumably positive for the site-specific insertion of the transgene in 46c ES cells. For this purpose primers were designed (blue arrowheads in Figure 12, sequences in Table 1), which span over a region of 1.8 kb (blue solid line in Figure 12). This region contains the transgenic hyg^R gene, a part of the Rosa26 recombination site as well as a section of the genomic DNA flanking the recombination site.

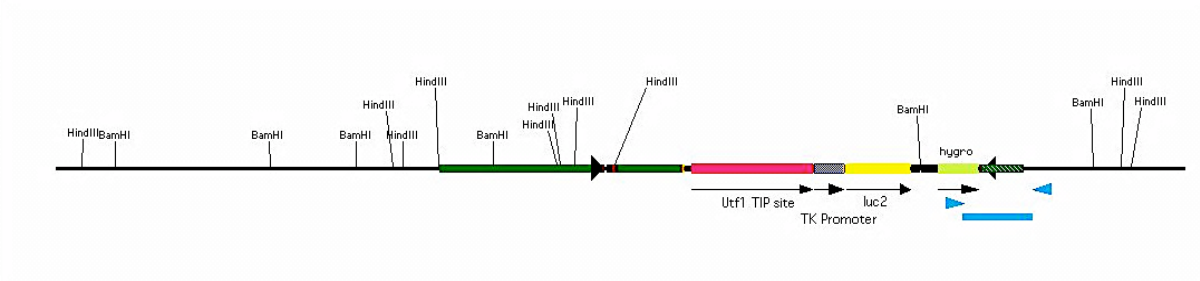


Figure 12 Genomic Rosa26 locus with integrated Utf1 TIP site transgene.
In green are depicted the Rosa26 recombination sites flanking the transgenic construct. In black is shown the genomic DNA. Blue arrowheads illustrate the primers used for the diagnostic PCR. The PCR product is shown as a solid blue line.

The PCR product was separated on an agarose gel as can be seen in figure 13. The PCR product has a size of 1.8 kb and contains the transgene on the desired genomic locus in 46c ES cells.
Additionally, the site-specific integration of the transgene was validated by southern blot. For this purpose the DNA of the cell colonies, which were preselected using diagnostic PCR, was cut with the restriction enzyme HindIII.
An external probe was used, which was hybridized to a complementary region on the genomic DNA flanking the Rosa26 locus (Figure 14, dark blue solid bar).

If the transgene was inserted site-specifically at the Rosa26 locus, the region amplified and radiolabelled will have a size of 6.8 kb. In heterozygous cells, the transgene is inserted only once at the Rosa26 locus. On the other allele the transgene must be missing and a region of 4.5 kb will be detected (Figure 14B). Due to this, two bands will be detectable on the blot after autoradiography: One band will be at the height of 6.8 kb and represent the transgene at the Rosa26 locus in 46c ES cell (Figure 14A). The other band will be at the height of 4.5 kb and will represent the wt Rosa26 locus without the transgene in 46c ES cells (Figure 14B).

For the transgenes Nkx2-9 (Figure 16, lane 1 – 6), Utf1 (Figure 16, lane 7 – 12) and En2 (Figure 16, lane 17 – 22) six cell lines were preselected by diagnostic PCR and verified by southern blot for site-specific integration.

In the case of the transgenes Utf1 and En2, all six cell lines were positive for a successful integration at the Rosa26 locus. However, two cell lines expected to contain the Nkx2-9 transgene were not positive for site-specific integration (Figure 16, lane 2 and 5).

Four cell lines containing the Meis1 transgene were preselected and confirmed positive for the site-specific integration in the Rosa26 locus by southern blot (Figure 16, lane 13 – 16).

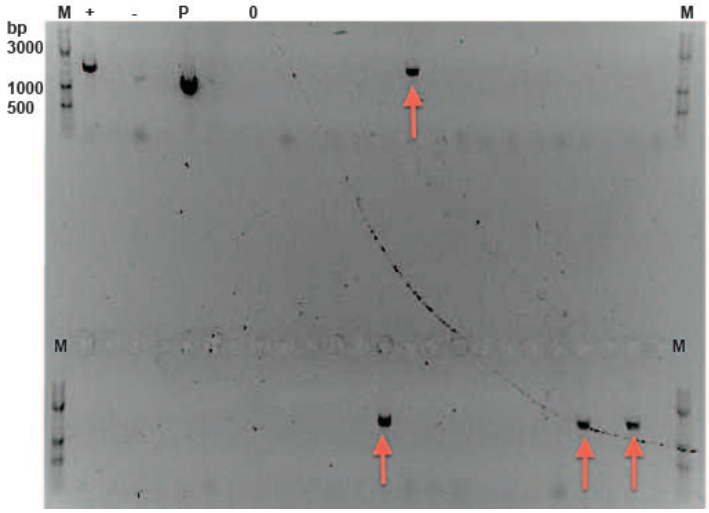


Figure 13 Diagnostic PCR.
Screening for cell lines containing the Utf1 transgene. Arrows indicate cell lines positive for the insertion of the transgene. M = Marker, + = positive control, - = negative control, P = Plasmid (additional positive control), 0 = only dH2O (additional negative control).

To validate whether the integration of the transgene occurred once or multiple times, the DNA of preselected cell lines was cut in a separate restriction digest with the restriction enzyme BamHI. Due to the hybridization position of the internal probe, which is located in the luc2 reporter gene (Figure 15, red solid bar), the amplified and radiolabelled product will vary in in size among the transgenic cell lines. The reason for

the unequal length of the products is the differing length of each TIP site or PRE (for more details see table 5).

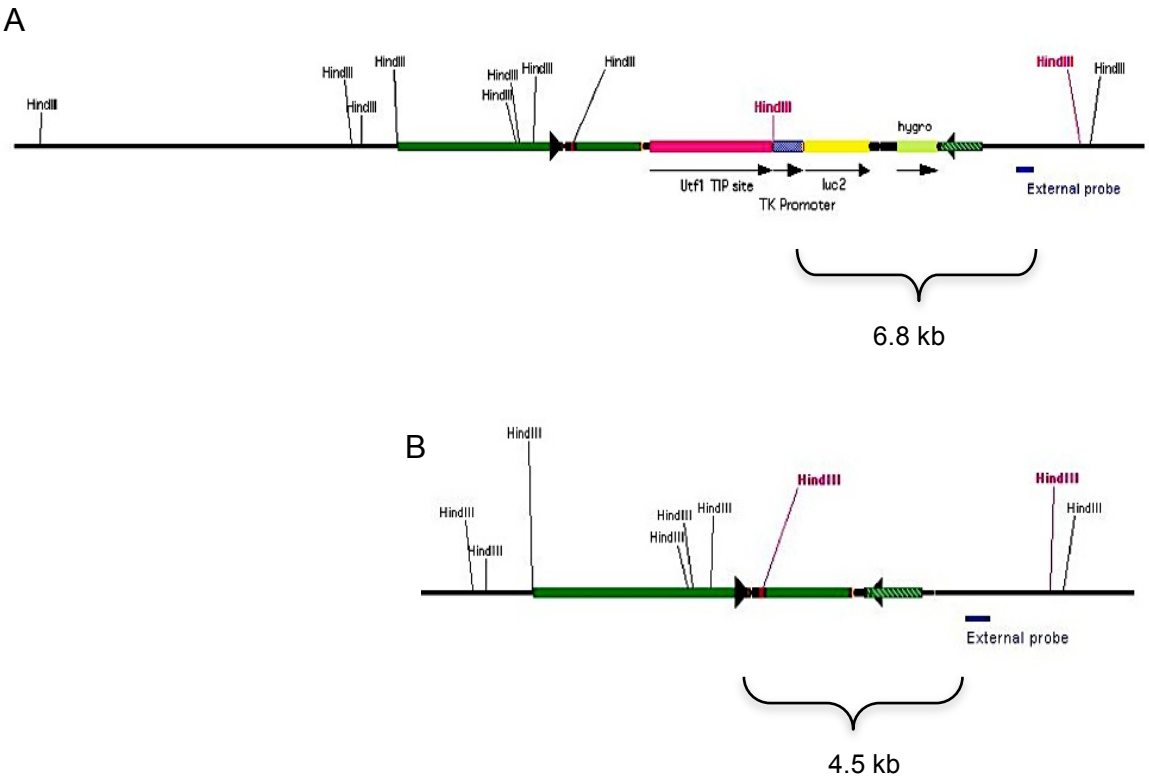


Figure 14 Murine Rosa26 locus with integrated Utf1 transgene.

Blue solid bar represents the hybridization region of the external probe. (A) Construct detected in this region is 6.8 kb long if the transgene is inserted. (B) Construct detected in the murine wt Rosa26 locus (for example if the transgene was not inserted) is 4.5 kb long.

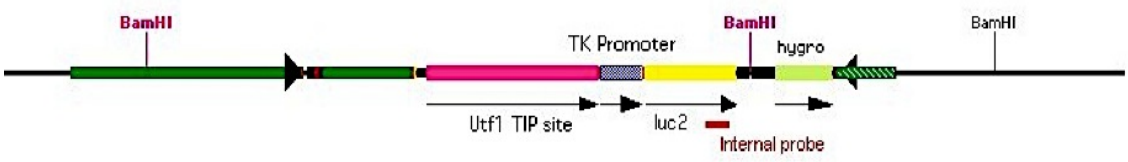


Figure 15 Rosa26 locus with integrated Utf1 transgene.

Red solid bar represents the hybridization region of the internal probe. The size of the constructs detected using the internal probe depends on the size of the integrated TIP site or the En2 PRE in this locus.

As shown in figure 17, all six transgenic cell lines contain the Utf1 TIP site or the En2 PRE and the insertion of these occurred once. In the case of the Nkx2-9 TIP site, single insertion of the transgene took place in all cell lines except for one (Figure 17, lane 2). As for the Meis1 TIP site, single insertion of the transgene was identified in two cell lines (Figure 17, lane 13 and 14) whereas in the other two cell lines possible double integration of the TIP site might have occurred (Figure 17, lane 15 and 16).

Taken together, the results of the diagnostic PCR (Figure 13) and the southern blot (Figure 16 and figure 17) demonstrate that for each integrated TIP site (Nkx2-9, Utf1 and Meis1) and for the En2 PRE there are at least two independent lines, which can be used as biological replicates. Single insertions of the transgenes were validated by southern blot using an internal probe. The site-specific integration was confirmed by diagnostic PCR and southern blot using an external probe.

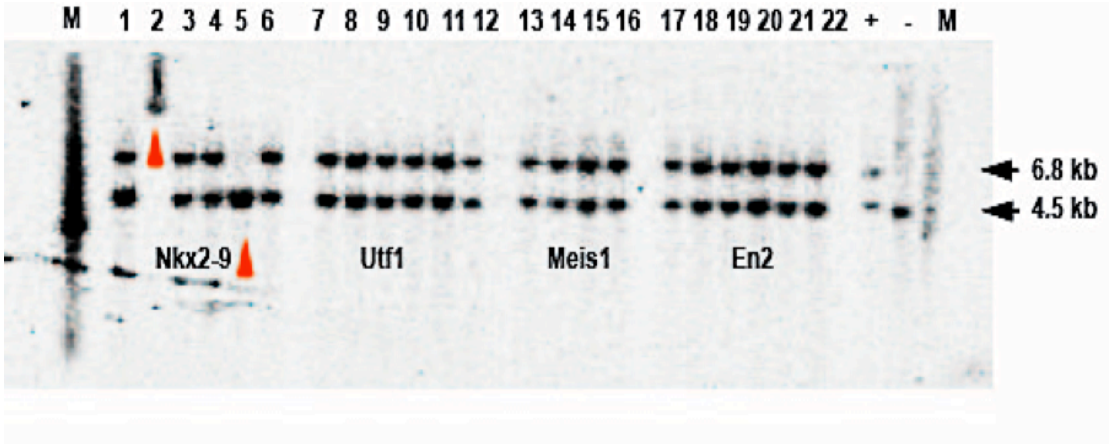


Figure 16 Southern blot: External probe.

Lane 1 – 6: Nkx2-9 preselected cell lines. Lane 7 – 12: Utf1 preselected cell lines. Lane 13 – 16: Meis1 preselected cell lines. Lane 17 – 22: En2 preselected cell lines. Black arrows represent the region containing the transgene at the Rosa26 locus (6.8 kb) or the wt Rosa26 locus without the transgene (4.5 kb). Red arrowheads represent cell lines not positive for integration of the transgenes at the Rosa26 locus (lane 2 and 5). M = Marker, + = positive control, - = negative control.

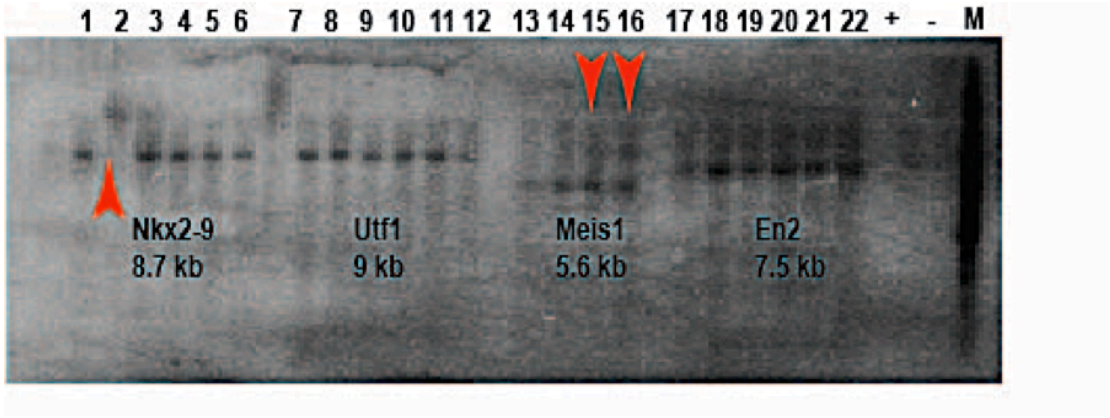


Figure 17 Southern blot: Internal probe.

Lane 1 – 6: Nkx2-9 preselected cell lines, size of the amplified region: 8.7 kb. Lane 7 – 12: Utf1 preselected cell lines, size of the amplified region: 9 kb. Lane 13 – 16: Meis1 preselected cell lines, size of the amplified region: 5.6 kb. Lane 17 – 22: En2 preselected cell lines, size of the amplified region: 7.5 kb. Red arrowheads represent cell lines not positive for single integration of the transgenes (lane 2) or double integration of the transgene (lane 15 and 16). M = Marker, + = positive control, - = negative control.

4.4 STABLE INTEGRATION OF TIP SITES AND THE EN2 PRE INTO THE MURINE ROSA26 LOCUS SHOW SILENCING OF THE LUCIFERASE REPORTER GENE

In previous as well as in this study it has been shown, that the expression of the luc2 reporter gene was repressed upon transient transfection of murine 46c ES cells with the selected TIP sites or the En2 PRE. Due to this we addressed whether this will also be the case once the TIP sites and the En2 PRE are integrated into the murine Rosa26 locus. The murine Rosa26 locus is an active gene region, which is transcribed at a moderate level and free of PcG protein binding [74, 75]. We have decided to target this region and insert the transgenes into the Rosa26 locus because it is a well-characterized locus that enables efficient integration.
Once it was verified which cell lines contained the transgenes of interest at the murine Rosa26 locus in 46c ES cells, the cells were lysed and a luciferase assay was performed to determine the expression levels of the luc2 reporter gene. For the transgenes Nkx2-9, Utf1 and En2 three independent cell lines were measured as biological replicates. For the control (Ctrl; contains neither a TIP site nor a PRE) and the Meis1 transgene two independent cell lines were used for the luciferase assay as biological replicates.

The Meis1 transgene does not repress the expression of the reporter gene as can be seen in figure 18. This is consistent with the results we obtained from transient transfection assays (Figure 11) and with previous data [1]. In contrast, the Nkx2-9 transgene does repress the expression of the luc2 gene also in accordance with previously shown data. The expression of the reporter gene is down regulated to only 27% compared to the expression of the reporter gene in Ctrl cell lines. Moreover, the expression of the luc2 gene is even lower in Nkx2-9 transgenic cell lines compared to the results obtained from the transient transfection in which the reporter was repressed to 44% of control levels.
However, the data for Utf1 and En2 transgenic lines do not show repression of the reporter gene as strongly as was the case in transient assays. In Utf1 cells the expression of the luc2 gene is 84% of the control level, whereas in the transient assays it was 40%. The same is true for En2 cell lines, repressing the reporter gene only to 66% compared to 37% in the transient assay.

In general the data indicate that the Nkx2-9 transgene represses the expression of the luc2 reporter gene in 46c murine ES cells. Furthermore, the Meis1 transgene has been shown not to repress the reporter gene in transient assays (Figure 11) as well as in

stable cell lines (Figure 18). However, the repressive effect of the Utf1 TIP site as well as of the En2 PRE on the luc2 reporter gene were not as strong as in transient assays.

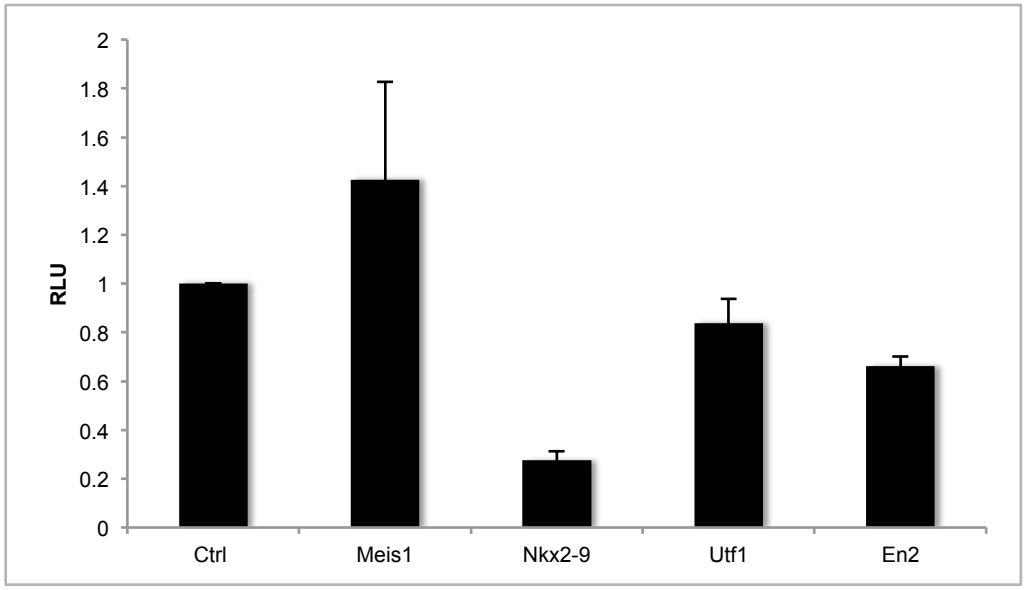


Figure 18 Expression levels of the firefly luciferase reporter gene in transgenic 46c ES cells.
Normalization of the data: Absolute values of the firefly luciferase were normalized on the absolute expression levels of the luciferase in control cell lines (Ctrl). Error bars represent standard deviations resulting from four different measurements and the comparison between their relative expression of the luc2 gene. RLU = Relative light units.

4.5 KNOCKDOWN OF EZH2 AND SUZ12 HAS NO EFFECT ON THE EXPRESSION LEVEL OF THE LUCIFERASE REPORTER GENE IN TRANSGENIC 46C ES CELLS

In previous studies it was shown that TIP sites, like PREs, are PcG target sites. In order to evaluate whether TIP sites represent a class of vertebrate PREs three major aspects of a PRE must be tested: Silencing of a reporter gene at an ectopic locus, loss of silencing of the reporter gene upon genetic PcG depletion, and recruitment of PcG to the PRE at the ectopic locus [5, 52].

As shown before, TIP sites and the En2 PRE have the ability to repress transcription of the luc2 reporter gene in murine 46c ES cells and thereby they fulfill the first hallmark of a PRE (Figure 18). Going further, we wanted to know whether expression of the reporter gene was up regulated upon knockdown of PcG proteins.

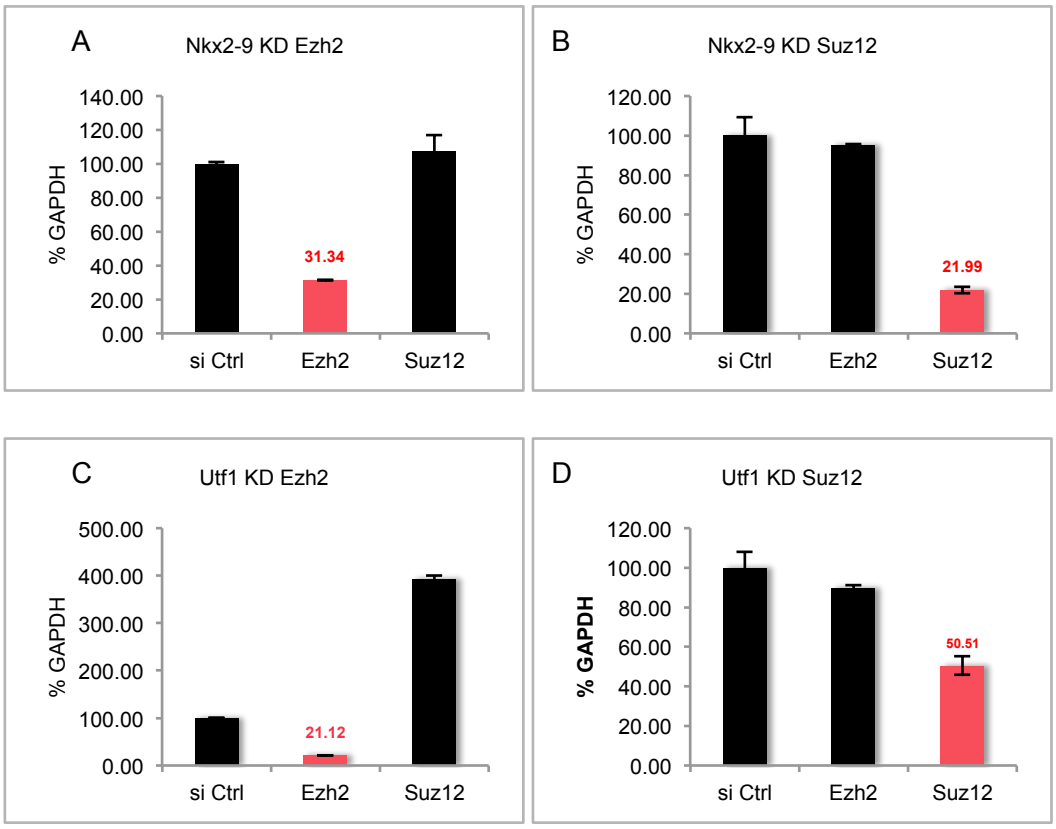


Figure 19 qPCR to determine PcG knockdown efficiency using siRNA targeting Ezh2 or Suz12 transcripts.

Si Ctrl represents cells transfected with non-targeting siRNAs. Values were normalized to *GAPDH*. Error bars show standard deviation. (A) KD of the Ezh2 transcript in Nkx2-9 transgenic cell lines. (B) KD of the Suz12 transcript in the Nkx2-9 transgenic cell lines. (C) KD of the Ezh2 transcript in the Utf1 transgenic cell lines. (D) KD of the Suz12 transcript in Utf1 transgenic cell lines.

For this purpose transgenic 46c ES cells were transfected with siRNAs targeting either Ezh2 or Suz12 transcripts (more information in table 3, RNA interference). To detect the knockdown of the Ezh2 or Suz12 transcript qPCR was performed (Figure 19, primer sequences in table 2).

The knockdown of both transcripts, Ezh2 and Suz12, was efficient to varying degrees (Figure 19). In Nkx2-9 transgenic cells the expression of the Ezh2 transcript was down regulated to 31% (Figure 19A) and the expression of the Suz12 transcript was lower than 22% (Figure 19B). The knockdown of the expression of Ezh2 in Utf1 transgenic cells was efficient as well (21%) whereas the knockdown of the Suz12 transcript in Utf1 transgenic cells was not as efficient (50.5 %) but still detectable (Figure 19C and D).

Overall the data show that the knockdown of either the Ezh2 or the Suz12 transcript was efficient.

Interestingly, an increased expression of the luc2 reporter gene upon knockdown of PcG proteins was not detected (Figure 20). In each cell line containing a TIP site or the En2 PRE as well as in cell lines not containing any TIP site or PRE (Ctrl) no change in luc2 gene expression was observed even though the knockdown itself was effective (Figure 19). This suggests either that the TIP sites are not subject to PcG regulation, or that the knockdown strategy was not sufficient to cause misregulation.

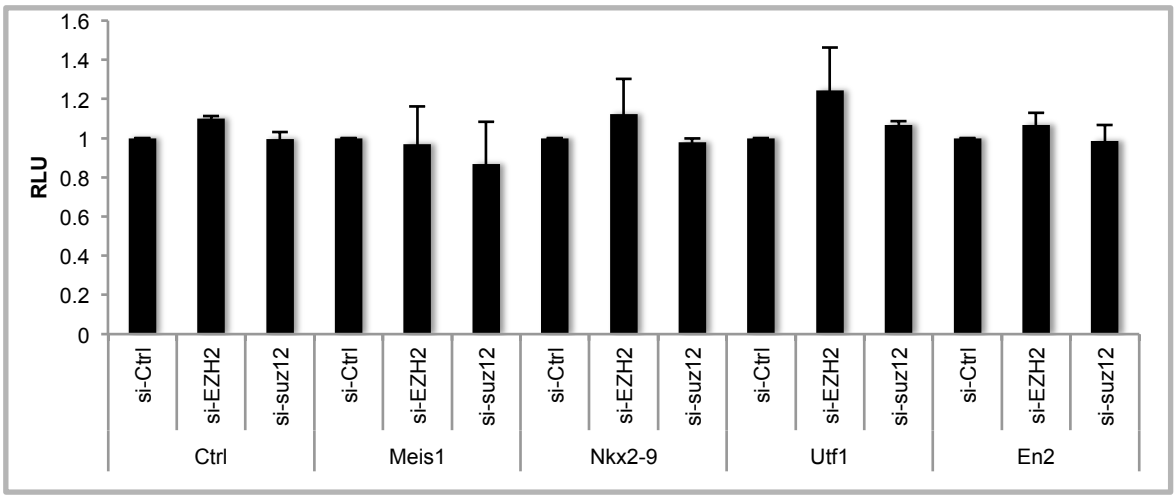


Figure 20 Detection of the expression levels of the luc2 reporter gene in murine 46c ES cells upon knockdown of PcG proteins Ezh2 and Suz12.

Ctrl = Control cell lines, which do not contain a TIP site or a PRE. Si-Ctrl = non targeting siRNAs. siEZH2 = siRNA targeting the Ezh2 transcript. siSuz12 = siRNA targeting the Suz12 transcript Error bars show standard deviations from seven independent measurements (n=1 except for Ctrl and En2 cell lines n=2). RLU = Relative light units.

4.6 Excision of the Hygromycin cassette using Cre recombinase leads to reduced expression of the luciferase reporter gene in some transgenic cell lines

The simian virus 40 (SV40) promoter is a promoter used for high-level expression (Invitrogen). In the constructs we used to integrate into 46c ES cells this promoter lies upstream of the hygromycin resistance gene and drives its expression (Figure 9C). Due to its ability to drive high-level expression we reasoned that it might have an influence also on the expression of the luc2 reporter gene.
To examine this possibility, the hyg^R gene, which is flanked by loxP sites, was cut out using the Cre recombinase. Primers flanking the region which was cut out were used to perform a diagnostic PCR and to confirm cell lines positive for the excision of the hyg^R gene (Figure 21). In the case of successful excision of the hygromycin resistance gene, the construct detected by diagnostic PCR has a length of 707 bp. The results of the diagnostic PCR confirm that the excision of the hygromycin resistance gene using Cre recombinase was successful (Figure 21A and 21B).

Next, a luciferase assay was performed to determine whether the SV40 promoter has an impact on the expression of the luc2 reporter gene (Figure 22). Intriguingly, the expression levels of the luc2 reporter gene vary highly upon excision of the hyg^R gene in cell lines containing either of the TIP sites at the murine Rosa26 locus (Figure 22). Meis1+ represents a cell line containing the Meis1 TIP site and the hyg^R gene at the Rosa26 genomic locus in murine 46c ES cells. The expression level of the *luc2* reporter gene in Meis1 transgenic cell lines is relatively high compared to the expression level of the reporter gene in the Ctrl cell line, also containing the hyg^R gene (Ctrl+). This has already been observed before (Figure 18). Meis1A and Meis1B represent cell lines positive for the excision of the hyg^R gene. The expression of the luc2 reporter gene in Meis1A cell lines is twice as high as the expression of the reporter gene in Meis1B cell lines. This is very striking since these two cell lines were thought to be biological replicates. This variation pattern can be observed in cell lines containing the Nkx2-9 TIP site or the Utf1 TIP site as well.

In contrast to these data, control cell lines containing either the En2 PRE or no TIP site or PRE at all, do not show such extreme variations in the expression of the reporter gene.

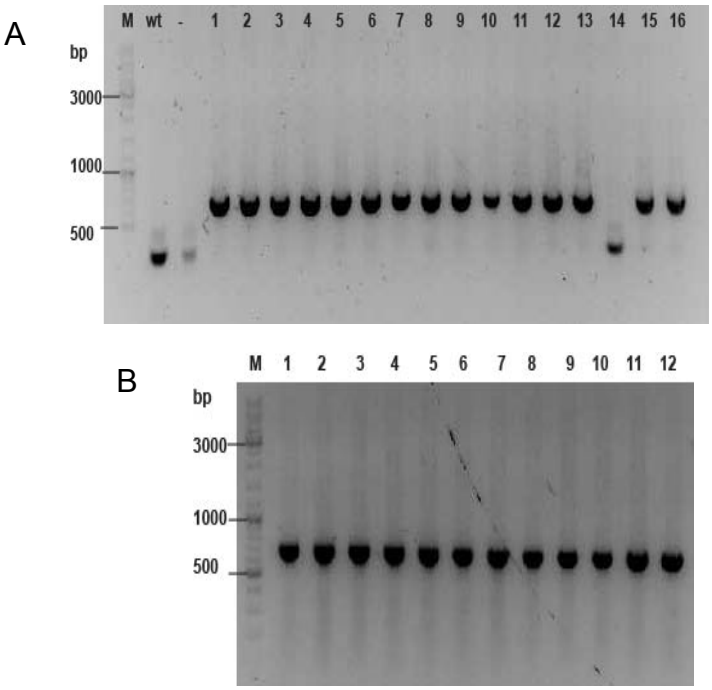


Figure 21 End-point PCR performed to identify cell lines positive for the excision of the hyg^R gene.
M = Marker. "Wt" represents wt 46c ES cells, which do not contain any kind of transgene. - = additional H_2O negative control. (A) Lane 1 – 6: independent cell lines containing the En2 PRE. Lane 7 – 11: independent cell lines containing no TIP site or PRE (Ctrl). Lane 12 – 16: independent cell lines containing the Nkx2-9 TIP site. (B) Lane 1 – 6: independent cell lines containing the Meis1 TIP site. Lane 7 – 12: independent cell lines containing the Utf1 TIP site.

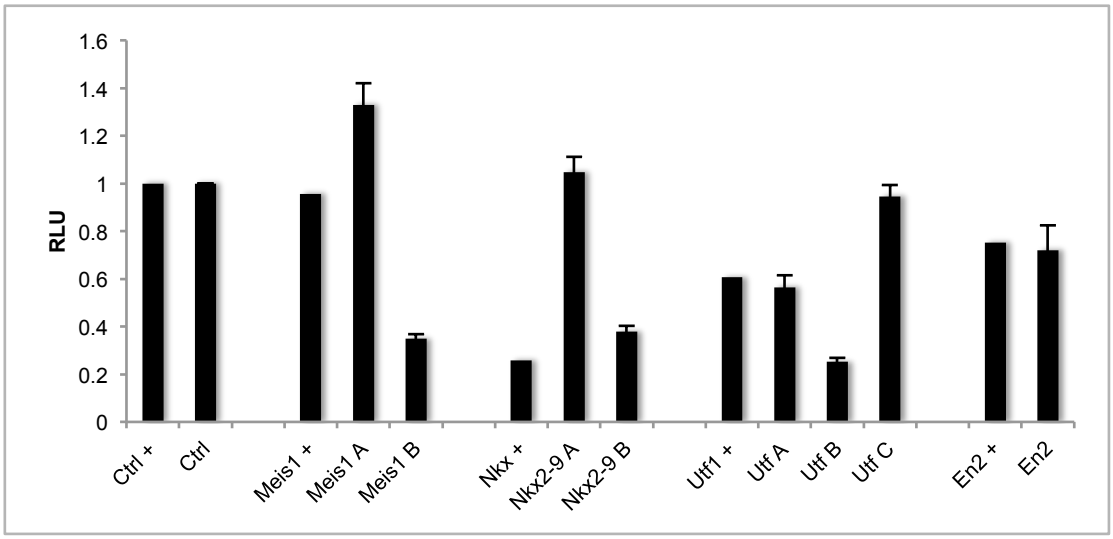


Figure 22 Expression levels of the luc2 reporter gene in murine 46c ES cells after hyg^R gene excision.
(Ctrl) represents cell lines, which do not contain any TIP site or PRE. (+) represents cell lines, which still contain the hyg^R gene. Error bars show standard deviations between four independent measurements. RLU = Relative light units.

Overall, the results of the luciferase assay are ambiguous for cell lines containing the TIP sites Meis1, Nkx2-9 and Utf1 transgenes compared to the controls. From these data it is not possible to conclude, what effect the hyg^R gene has on the reporter gene expression.

4.7 NcRNA ANALYSIS AT TIP SITES AND CORRESPONDING GENES

The involvement of ncRNAs in PcG targeting and recruitment has been observed previously by several authors [42, 63-65]. In genome-wide studies it has been demonstrated that TIP sites are transcribed during neural differentiation of ES cells [1]. Moreover, by performing a reporter assay the authors were able to detect one transcript, the Nkx2-9 TIP site transcript, which was required for full repression of the reporter gene [1].

Due to this, expression studies were performed to determine whether TIP sites are transcribed into ncRNA once they are integrated into chromatin. Non-coding transcripts from each TIP site were measured in transgenic cell lines as well as in the control lines without the transgenes and in 46c wild-type ES cells (Figure 23). Previously it has been shown in transient assays that the Meis1 TIP site was highly transcribed compared to the other TIP sites [1]. However, we find that in the transgenic cell lines, transcription from the Utf1 TIP site is substantially higher than in the other two TIP sites (Figure 23A). Moderate transcription from the Nkx2-9 TIP site was detected with an expression level two fold higher in the transgenic line compared to the control line without the TIP site (Figure 23B). Expression of the Meis1 TIP site was detected in the transgenic cell line and endogenous levels were very low (0.2 % GAPDH in the control line without the transgene) or not detectable at all (in wild-type ES cells, Figure 23B).
We were also interested in the expression of the transcripts from the corresponding TIP site genes. To ask whether the overexpression of the TIP site transcript from the transgene would affect the transcription of the corresponding endogenous gene, we performed qPCR to detect Utf1, Nkx2-9 and Meis1 mRNAs in the presence and absence of the transgenic construct. We find that the Utf1 gene is transcribed highly in the transgenic cell line compared to the control without the transgene (Figure 23C). However, expression of the Utf1 gene in wild-type cells was also elevated, thus it is not clear whether this is a cell line variation. The Nkx2-9 gene was expressed highly above endogenous background in the Nkx2-9 transgenic cell line, whereas expression of the Meis1 gene in Meis1 transgenic cell lines was not detectable above background (Figure 23D).

To determine whether the non-coding transcripts of the transgenic TIP sites have an influence on the repression of the reporter gene, knockdown experiments were performed using Locked nucleic acids (LNAs). These targeted either the respective TIP site transcript or were non-specific towards a target (Ctrl). Surprisingly, expression of

the Nkx2-9 TIP site for one biological replicate is reduced and for the other this is not the case (Figure 24A). The same is true in Nkx2-9 transgenic cells, where the TIP transcript was knocked-down and the expression of the corresponding gene was detected (Figure 24C).
For the other biological replicate of Nkx2-9 as well as for the Utf1 TIP sites the knockdown did not work (Figure 24B and 24D). Furthermore, no changes in reporter gene expression were detected upon knockdown of the TIP site transcripts (Figure 24E and 24F).

In summary, transcription from the TIP sites and from the corresponding genes was detectable above background in all transgenic cell lines except for the Meis1 gene. The data also indicate that there is massive transcription from the Utf1 TIP site and the Utf1 gene. Upon knockdown of the TIP site transcripts, no changes in TIP site expression or TIP site corresponding genes have been observed except for one biological replicate of Nkx2-9. In addition to that, no reduction in reporter gene expression was detected.

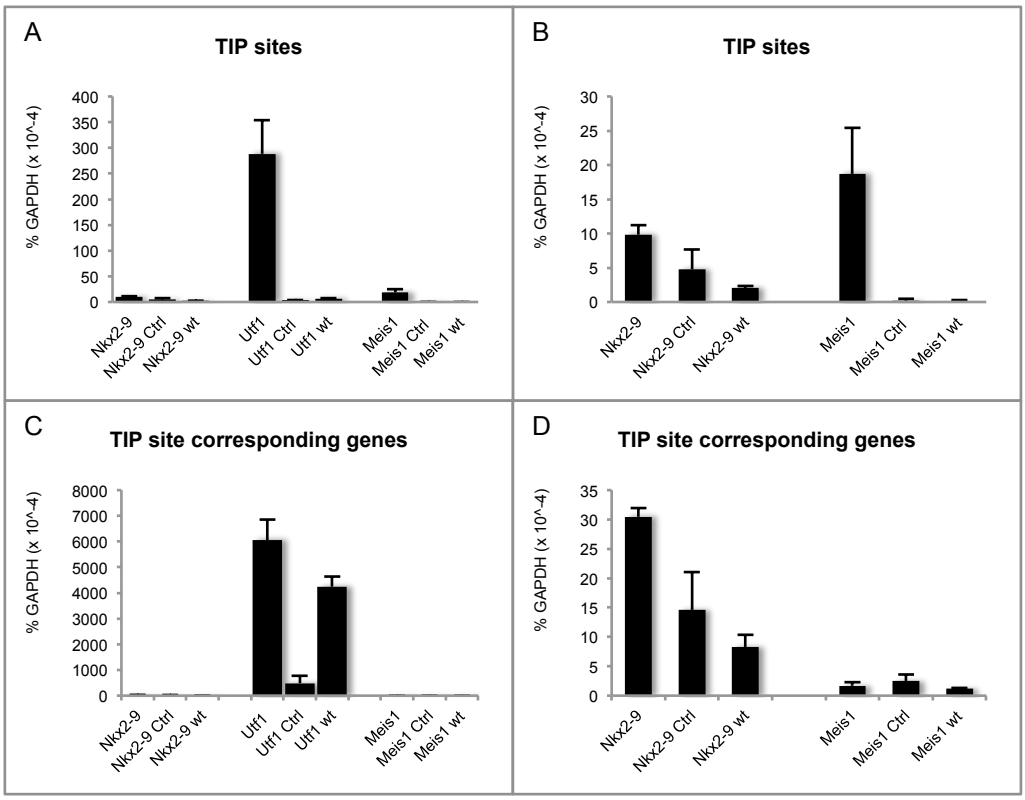


Figure 23 Detection of TIP site transcripts and transcripts from corresponding TIP site genes.
(A) Detection of the TIP site transcripts in transgenic cell lines (Nkx2-9, Utf1 and Meis1), in control cell lines without integrated TIP site (Nkx2-9 Ctrl, Utf1 Ctrl and Meis1 Ctrl) and in 46c wild-type ES cells (Nkx2-9 wt, Utf1 wt and Meis1wt). (B) Detection of the Nkx2-9 and Utf1 TIP site transcripts in transgenic cell line, control cell lines without integrated TIP site and in 46c wild-type ES cells. (C) Expression levels of corresponding TIP site genes in transgenic cell lines, control cell lines and 46c wild-type ES cells. (D) Expression levels of corresponding Nkx2-9 and Utf1TIP site genes in transgenic cell lines, control cell lines and 46c wild-type ES cells.

X-Axis: In “Nkx2-9” expression of the transgenic and endogenous Nkx2-9 TIP site or gene was measured in the transgenic cell line containing the Nkx2-9 TIP site. In “Nkx2-9 Ctrl” the expression of the endogenous TIP site or gene transcript was measured in transgenic cell lines, which contain the same transgene as Nkx2-9 transgenic cell except for the TIP site itself. In “Nkx2-9 wt” the transcript of the endogenous Nkx2-9 TIP site or gene was detected in 46c wild-type ES cells.

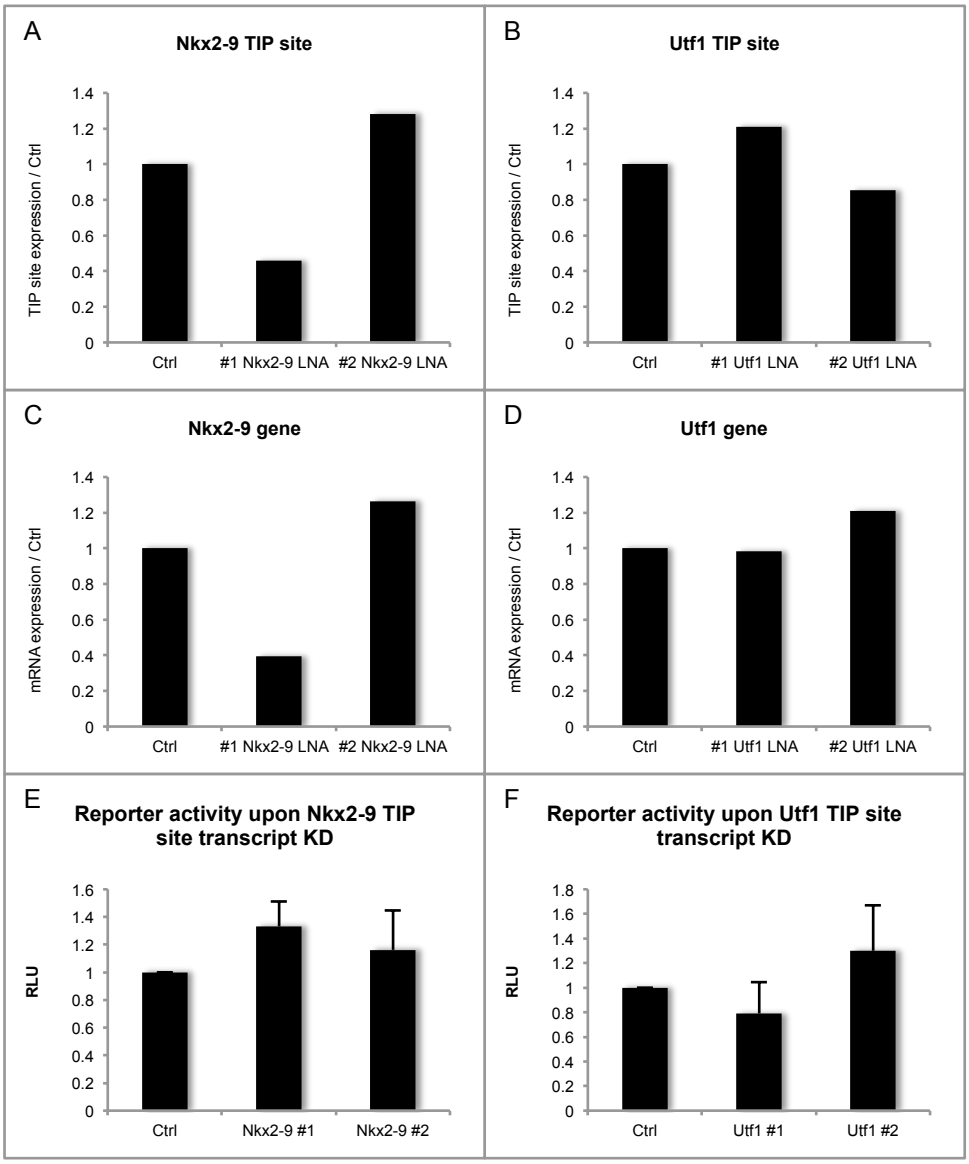


Figure 24 Determination of TIP site ncRNA involvement in PcG dependent repression of the reporter gene.

(A) Detection of Nkx2-9 and (B) Utf1 TIP site transcripts in transgenic cell lines containing the respective TIP site upon TIP site transcript knockdown with scrambled LNAs (Ctrl) and LNAs targeting the particular TIP site transcript. (C) Expression levels of the corresponding Nkx2-9 and (D) Utf1 TIP site genes upon knockdown of the TIP site transcripts in the respective transgenic cell lines. (E) Relative values for the expression of the Luciferase reporter gene in transgenic cell lines carrying the Nkx2-9 TIP site or (F) the Utf1 TIP site upon knockdown with scrambled LNAs (Ctrl) and LNAs targeting the particular TIP site transcript. RLU = Relative light units. "#1 Nkx2-9 LNA" and "#2 Nkx2-9 LNA" represent two biological replicates, as well as "#1 Utf1 LNA" and "#2 Utf1 LNA".

5 DISCUSSION

5.1 ANALYSIS OF REPORTER GENE REPRESSION IN TRANSIENT AND TRANSGENIC ASSAYS

In the course of this study transgenic cell lines were generated carrying either transcribed intergenic Polycomb (TIP) sites or the En2 PRE. As a control, transgenic cell lines carrying the same constructs but without a TIP site or the En2 PRE were used. Reporter gene assays were performed to determine whether TIP sites have the potential to repress transcription from the Luciferase reporter gene.

In previous studies (Heinen *et al*., manuscript in preparation) the En2 PRE was identified and shown to fulfill the hallmarks of a PRE, namely silencing of a reporter gene at an ectopic locus, loss of silencing of the reporter upon PcG protein depletion and recruitment of PcG proteins to the PRE at the ectopic locus [5, 52]. In that study a minimal promoter was used to drive moderate expression of the Luciferase reporter gene. In our experiments we used the En2 PRE as a positive control, since we knew what to expect from the reporter gene assays for this PRE. In contrast to the study mentioned above, a viral TK promoter was used to drive expression of the reporter gene instead of a minimal promoter. The TK promoter was also used in the constructs that carried a TIP site, and has previously been shown to be repressed by specific TIP sites in transient [1]. The choice of the TK promoter for the integrated assays described here, instead for the minimal promoter used in Heinen *et al.*, was intended to drive high-level expression of the *Luciferase* reporter gene, giving a high dynamic range within which repression could occur. Thus in the case of reporter gene repression, this would provide evidence that the TIP site or the PRE have the ability to silence expression of the reporter gene from a strong promoter.

In transient assays performed during this study 46c ES cells transfected with constructs carrying the TIP site or the En2 PRE show reduced transcription from the *Luciferase* reporter gene down to 40% (Figure 11). This is true for all TIP sites and the En2 PRE except for the Meis1 TIP site. These results agree with transient assays performed previously by Hekimoglu-Balkan and colleagues [1]. Additionally, here it was shown for the first time that the En2 PRE in combination with the strong viral TK promoter has a repressive effect on the reporter gene. This confirms that the En2 PRE has strong repressive function in gene regulation due to its ability to reduce

transcription from the reporter gene. This is also true for the Nkx2-9 and the Utf1 TIP sites.

To investigate the function of TIP sites and the En2 PRE upon stable integration into chromatin of 46c ES cells, transgenic cell lines were generated in which each construct was integrated at the same genomic location. Using these lines we performed reporter assays, demonstrating that repression of reporter transcription changed compared to the Luciferase read-out in transient assays (Figure 18). The En2 PRE lost more than 20% in repression compared to transient assays. The Utf1 TIP site repressed the reporter gene only to 80% whereas it showed twice as much repression in the transient assays. Surprisingly, the Nkx2-9 TIP site showed stronger repression in the transgenic assays and repressed the reporter gene more than threefold compared to the control cell line without the TIP site and almost twice as much as in transient assays. The data for the Meis1 TIP site remained unaltered, showing no repression of the reporter gene.

The source of these differences in repression profiles might be due to alterations in chromatin environment upon stable integration of the TIP sites into genomic DNA. To date it is still unclear whether bacterial plasmid DNA becomes packaged into chromatin upon transient transfections in mammalian cells [76, 77]. The function of regulatory elements such as PREs and TIP sites is dependent on their chromatin context. From studies performed in *Drosophila* it is known, that PREs act together to enhance gene silencing. Furthermore it has been reported, that PREs can act via higher-order chromatin interactions [52, 78]. Assuming that this is also true for vertebrate PREs, the chromatin context in which the PREs and TIP sites are inserted will influence their activity. If however no such chromatin context can be established when inserting the TIP site into the cell using bacterial plasmid DNA, then the activity of these TIP sites and PREs might be altered.

Furthermore, it has to be taken into consideration that we have used the Rosa26 locus as our targeting region and as mentioned before, this locus is euchromatic. This means, it has an open chromatin structure and is actively transcribed [74, 75]. This circumstance might be problematic insofar as it may affect recruitment of PcG proteins in a negative way. *In vitro* experiments show that PRC2 is recruited more effectively to dense chromatin compared to more dispersed chromatin regions [79]. Hence, this would result in less PcG protein recruitment and less reporter gene repression at the euchromatic Rosa26 locus. However, following these thoughts it is difficult to reconcile the fact that the Nkx2-9 TIP site repressed the reporter gene more efficiently in transgenic cell lines than in transient assays. Nonetheless it would be necessary for

future experiments to consider targeting other genomic regions, which are not as active as the Rosa26 locus.

Another possibility for such drastic changes in reporter gene repression between the transient and the integrated assay could be the herpes simplex virus (HSV) TK promoter, which lies upstream of the Luciferase reporter gene and drives its expression. The TK promoter drives high-level expression of the reporter gene and was used in this study exactly because of this feature to obtain a broader spectrum of reporter gene expression. On the other hand one can assume that the TK promoter drives *Luciferase* expression at a high level and by this interferes with the repressive function of the TIP sites. This way the detected expression levels of the reporter gene would be falsified in the case of the En2 PRE, the Meis1 TIP site and the Utf1 TIP site (Figure 18). One future possibility of overcoming this problem is to generate transgenic cell lines driven by a minimal promoter, which drives moderate expression of the reporter gene. The drawback of this promoter is a reduced detection spectrum of reporter gene expression. However, the Nkx2-9 TIP site represses *Luciferase* expression in transgenic cell lines (Figure 18). These data argue against the interpretation, that the TK promoter interferes with the repressive mechanism of the TIP sites.

Each transgenic cell line contained besides the TIP site or the En2 PRE also a hygromycin resistance gene, which is also driven by a viral promoter, namely the SV40 promoter. This promoter drives high-level expression of the hyg^R gene. To exclude the possibility of strong interference resulting from the SV40 promoter, for each TIP site and the En2 PRE cell lines were generated without the hyg^R gene (Figure 21). Luciferase reporter assays were performed and we expected to obtain an overall reduced expression of the reporter gene upon hyg^R excision. Strikingly, biological replicates of transgenic cells containing the TIP sites Utf1, Meis1 and Nkx2-9 show drastic variation in reporter gene expression (Figure 22). Since the source for this diverse data points might be a contamination among the cell lines, genotyping of each replicate will be necessary to confirm their identities.

5.2 KNOCKDOWN OF PRC2

To determine whether repression of the reporter gene transcript was dependent on PcG proteins, we have used siRNAs to knockdown transcripts from SUZ12 and EZH2, two core components of the PRC2 complex. Quantitative PCR was performed to confirm successful knockdown of these transcripts (Figure 19).

Furthermore we performed assays to determine the expression of the Luciferase reporter gene upon knockdown of PRC2. Against our expectations, expression of the reporter gene was not elevated (Figure 20). Interestingly, reporter gene expression in the En2 PRE transgenic cell lines did not increase upon knockdown of the PRC2 components, although cell lines carrying this PRE cell lines responded to the same knockdown strategy in Heinen *et al*. The quality of the siRNAs was considered intact because of the efficient knockdown (Figure 19). Nonetheless, an additional method detecting protein expression of the transcript, which was knocked down, e.g. a western blot should be considered for future work.
Furthermore, the knockdown method we have used may not be the optimal approach for this system. After the delivery of the siRNAs to each cell lines, the cells were incubated at 37°C for 48 hours as described in the transfection protocol (Lipofectamin 2000 reagent, Invitrogen). Using this protocol, the cells may not have had sufficient time to degrade the PRC2 components fully in only 48 hours. An extension of the incubation period might lead to degradation of the siRNAs. Furthermore, since the cells are dividing actively, the space provided in a 24-well plate may not have been sufficient. Furthermore, since it has been reported that H3K27me3 is a stable mark with a low turn-over [47], it is possible that the cells require more time to dispose of this histone modification after the knockdown. Due to this, subsequent experiments will focus on an shRNA-mediated knockdown of PRC2. Following this technique, the shRNAs are delivered into the cells using a viral vector and integrated into the genome. The shRNA is then transcribed by Pol II, resulting in a pri-microRNA that can be further processed firstly by Drosha and then by Dicer until it is loaded into the RNA-induced silencing complex (RISC). The passenger strand will be degraded whereas the guide strand will lead to mRNA cleavage of the targeted transcript or to repression of its translation, in both cases leading to gene silencing. This will allow to knockdown components of PRC2 over a longer period of time [80].

Finally, since repression from the reporter gene did not respond to PRC2 knockdown, and transcript levels vary so drastically between transient and transgenic cell lines, it

would be interesting to know whether H3K27me3 enrichment profiles are distributed over the TIP site in transgenic cell lines as they are on the endogenous TIP sites. To gain better insight into this, it would be necessary to perform ChIP experiments with the transgenic cell lines and compare these to the transient data.

5.3 NcRNA ANALYSIS

In previous studies it was observed that TIP sites are transcribed into ncRNA in ES cells transiently transfected with constructs carrying the TIP sites, demonstrating that the TIP sites themselves contain promoters [1]. Moreover, the authors showed that in specific cases the transcripts were required for reporter gene repression. We wished to address whether the TIP promoters would drive transcripts from these sites once they were stably integrated into the genomic DNA of ES cells, and whether the transcripts would be required for reporter gene repression.

We were able to detect transcription from all TIP sites above background levels (Figure 23). Meis1 TIP site transcription was approximately 20-fold above endogenous levels. In contrast, transcription from the Meis1 TIP site observed by our colleagues was 166-fold above endogenous levels in transient assays [1]. Moreover, the authors observed that the transcription from the Meis1 TIP site was relatively higher compared to the transcription levels of the other TIP sites. The Nkx2-9 TIP site was detected 5-fold and the Utf1 TIP site 3-fold above endogenous levels [1]. In our study, we show that transcript levels from the Utf1 TIP site were drastically increased compared to the other TIP site transcripts (Figure 23A). Moreover, transcription levels from the corresponding Utf1 gene were augmented as well compared to the other TIP site genes (Figure 23C). However, expression of the Utf1 gene in wild-type cells was also elevated. Interestingly we find, that transcription of the Utf1 gene in the Utf1 TIP site transgenic cell line is 1.5-fold higher than in the “Utf1 wt” cell line whereas it is 12-fold higher compared to the “Utf1 Ctrl” line. Since both cell lines, the “Utf1 wt” and the “Utf1 Ctrl” cell line contain the same number of Utf1 TIP sites, we would have had expected that the mRNA levels of the Utf1 gene in these cell lines would be similar. This difference in transcript levels between the wild type and control cell line might be due to contaminations that happened in the cell culture. If the repetition of this experiment shows that the transcription of the Utf1 gene is indeed elevated in those cell lines, which contain the transgenic Utf1 TIP site, then this would suggest that the Utf1 TIP site transcript is directly involved in the regulation of the Utf1 gene expression.

Additionally, we observed that the transcription level of the Nkx2-9 TIP site was 2-fold higher compared to the control line without the transgenic Nkx2-9 TIP site and 5-fold higher compared to the wild type cells. Similar differences in mRNA expression of the Nkx2-9 gene were observed in the respective cell lines. The mRNA level of the Nkx2-9 gene was 2-fold higher in the transgenic cell line compared to the control and 4-fold higher compared to wild type cells. To conclude whether the Nkx2-9 TIP site has regulatory function on the expression of the corresponding gene a repetition of this experiment is required as it is for the Utf1 TIP site.

Furthermore, we wanted to know if repression from the Luciferase reporter gene depends on transcripts from the TIP sites. To evaluate this, we used Locked Nucleic Acids (LNAs) to knockdown the TIP site transcripts from the Nkx2-9 and Utf1 TIP site (Figure 24A and 24B), because these two TIP sites showed repression of the reporter gene (Figure 18). In previous studies it has been shown that the knockdown of the Utf1 TIP site transcript does not affect expression of the Luciferase reporter gene whereas the knockdown of the Nkx2-9 TIP site led to a significant increase in Luciferase reporter activity [1]. Due to this the authors concluded that the transcript of the Nkx2-9 TIP site might be involved in the repression of the reporter gene. At the same time we have also examined the corresponding genes in those cell lines in which the respective TIP site transcript was knocked down (Figure 24C and 24D).

Unfortunately, Utf1 TIP site transcripts did not decrease upon transcript knockdown using LNAs (Figure 24B), demonstrating that the knockdown was not successful. Surprisingly, it was observed that the Nkx2-9 TIP site transcript was reduced to 50% for one biological replicate (Figure 24A, #1 Nkx2-9 LNA). Nonetheless, these data could not be reproduced with the other biological replicate (Figure 24A, #2 Nkx2-9 LNA). However, we observed the same pattern in transcript knockdown when we examined the corresponding Nkx2-9 gene. It was interesting to observe that transcription of the Nkx2-9 coding gene decreased upon knockdown of the Nkx2-9 TIP site transcript (Figure 24C). This result might indicate that the non-coding transcript from the Nkx2-9 TIP site has a regulatory effect on its corresponding gene. This would be consistent with the apparent up-regulation of the Nkx-2-9 gene in the presence of the TIP transgene. However these data must be considered as preliminary because the method needs to be improved in order to generate reproducible data.

Previously it has been reported that reporter gene repression was dependent on the non-coding transcript from the Nkx2-9 TIP site [1]. To ask whether this is also the case in the integrated system, luciferase assays were performed to detect expression from

the Luciferase reporter gene upon knockdown of the TIP site transcripts. However, no reduction in reporter gene was observed even though the transcript knockdown was apparently effective for one replicate carrying the Nkx2-9 TIP site (Figure 24E and 24F).

Although the knockdown was performed using the same methods and conditions as previously reported [1, 71], massive cell death was observed upon LNA transfection. Furthermore, it has to be taken into account that due to time reasons the LNA transfections were performed only once. Considering these facts, interpretation of the data set is currently not possible. Optimization of the knockdown procedure as well as several repetitions of this experiment will be required to evaluate these data. Due to this, we can only conclude that there is potential for improving the knockdown while at the same time minimizing cell death to obtain reproducible data. Once this has been achieved, it will be interesting to re-examine the data reported here.

6 ABBREVIATIONS

^{32}P	Phosphorus-32
Amp^R	Ampicillin resistance
ANT-C	Antennapedia complex
AP	Alkaline phosphatase
BMI1	B lymphoma Mo-MLV insertion region 1 homolog
bp	Basepairs of DNA
BSA	Bovine serum albumin
BX-C	Bithorax complex
C	Celsius
CBX	Chromobox
cDNA	Complementary DNA
CG-rich	Cystidine-Guanine-rich
cHC	Constitutive Heterochromatin
ChIP	Chromatin immunoprecipitation
Ctrl	Control
dCTP	Deoxycytidine triphosphate
ddH20	Double distilled water
dH20	Distilled water
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EED	Embryonic ectoderm development
En2	Engrailed
ES cells	Embryonic stem cells
ESC	Extra sex comb
EZH2	Enhancer of Zeste homolog 2
FCS	Fetal calf serum
fHC	Facultative Heterochromatin
fw	Forward
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GCK	Gene Construction Kit
H2AK119ub1	Histone 2A Lysine 119 monoubiquitination
H3	Histone 3
H3.3	Histone 3.3
H3K27ac	Histone 3 Lysine 27 acetylation
H3K27me	Histone 3 Lysine 27 monomethylation
H3K27me2	Histone 3 Lysine 27 dimethylation
H3K27me3	Histone 3 Lysine 27 trimethylation
H3K36me3	Histone 3 Lysine 36 trimethylation
H3K4me2	Histone 3 Lysine 4 dimethylation
H3K4me3	Histone 3 Lysine 4 trimethylation
HCl	Hydrogen chloride
HMT	Histonemethyltransferase
HSV	Herpes simplex virus
Hyg^R	Hygromycin resistance
IMBA	Institute of Molecular Biotechnology
Kan^R	Kanamycin resistance
Kb	Kilobase pairs of DNA
KCl	Potassium chloride
KD	Knockdown
kg	Kilogram
KO	Knockout
LB media	Lysogeny Broth media
LIF	Leukemia inhibitory factor
lincRNA	Large intergenic non-coding RNA
LNA	Locked nucleic acid
luc2	Luciferase
Meis1	Meis homeobox 1
MgCl2	Magnesium chloride
MgSO4	Magnesium sulfate
min	Minute
ml	Milliliter
MP	Minimal promoter
mPH	Mammalian Polyhomeotic
NaCl	Sodium chloride
ncRNA	Non-coding RNA
ng	nanogram
Nkx2-9	NK2 transcription factor related, locus 9
nm	Nanometer
Oligo	Oligonucleotide
PBS	Phosphate buffered saline
PC	Polycomb

PcG	Polycomb group
PCR	Polymerase chain reaction
PH	Polyhomeotic
Pol	Polimerase
PRC	Polycomb repressive complex
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PRE	Polycomb response element
qPCR	quantitative PCR
RbAP	Retinoblastoma-associated protein
RING1A	Ring finger protein 1
RLU	Relative light units
RNA	Ribonucleic acid
RNAi	RNA Interference
RNS	Ribonukelinsäure
RT	Room temperature
rv	Reverse
SDS	Sodium dodecyl sulfate
sec	Second
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
SSC	Saline sodium citrate
SUZ12	Suppressor of Zeste 12
SV40	Simian virus 40
TIP site	Transcribed intergenic Polycomb target site
TK	Thymidine kinase
TrxG	Trithorax group
TV	Targeting vector
Utf1	Undifferentiated embryonic cell transcription factor 1
UV/VIS	Ultraviolet/visible spectroscopy
V	Volt
wt	Wildtype
Xi	Inactive X-chromosome
µg	Microgram
µl	Microliter
µM	Micromolar

7 List of Figures

8 LIST OF TABLES

9 REFERENCES

1. Hekimoglu-Balkan, B., et al., *Intergenic Polycomb target sites are dynamically marked by non-coding transcription during lineage commitment.pdf.* RNA Biology, 2012. **9**(3): p. 1-12.

2. Azuara, V., et al., *Chromatin signatures of pluripotent cell lines.* Nature cell biology, 2006. **8**(5): p. 532-542.

3. Christophersen, N.S. and K. Helin, *Epigenetic control of embryonic stem cell fate.* J Exp Med, 2010. **207**(11): p. 2287-95.

4. Kennison, J.A., *Introduction to Trx-G and Pc-G Genes.* Methods in Enzymology, 2004. **377**.

5. Ringrose, L. and R. Paro, *Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins.* Annu Rev Genet, 2004. **38**: p. 413-43.

6. Cavalli, G. and R. Paro, *The Drosophila Fab-7 Chromosomal Element Conveys Epigenetic Inheritance during Mitosis and Meiosis.* Cell, 1998. **93**: p. 505-518.

7. Rank, G., M. Prestel, and R. Paro, *Transcription through Intergenic Chromosomal Memory Elements of the Drosophila Bithorax Complex Correlates with an Epigenetic Switch.* Molecular and Cellular Biology, 2002. **22**(22): p. 8026-8034.

8. Lewis, E.B., *A gene complex controlling segmentation in Drosophila.* Nature, 1978. **276**.

9. Struhl, G., *A gene product required for correct initiation of segmental determination in Drosophila* Nature, 1981. **293**.

10. Voncken, J.W., et al., *Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition.* Proc Natl Acad Sci U S A, 2003. **100**(5): p. 2468-73.

11. Ho, L. and G.R. Crabtree, *An EZ mark to miss.* Cell Stem Cell, 2008. **3**(6): p. 577-8.

12. Morey, L. and K. Helin, *Polycomb group protein-mediated repression of transcription.* Trends Biochem Sci, 2010. **35**(6): p. 323-32.

13. Pasini, D., et al., *Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity.* The Embo Journal, 2004. **23**: p. 4061–4071.

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

15. Schuettengruber, B. and G. Cavalli, *Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice.* Development, 2009. **136**(21): p. 3531-42.

16. Simon, J.A. and R.E. Kingston, *Mechanisms of polycomb gene silencing: knowns and unknowns.* Nat Rev Mol Cell Biol, 2009. **10**(10): p. 697-708.

17. Schwartz, Y.B. and V. Pirrotta, *Polycomb silencing mechanisms and the management of genomic programmes.* Nat Rev Genet, 2007. **8**(1): p. 9-22.

18. O'Carroll, D., et al., *The polycomb-group gene Ezh2 is required for early mouse development.* Mol Cell Biol, 2001. **21**(13): p. 4330-6.

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

20. Kirmizis, A., et al., *Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27.* Genes Dev, 2004. **18**: p. 1592-1605.

21. Faust, C., et al., *The eed mutation disrupts anterior mesoderm production in mice.* Development, 1995. **121**: p. 273-285.

22. Tie, F., et al., *The N terminus of Drosophila ESC binds directly to histone H3 and is required for E(Z)-dependent trimethylation of H3 lysine 27.* Mol Cell Biol, 2007. **27**(6): p. 2014-26.

23. Margueron, R., et al., *Role of the polycomb protein EED in the propagation of repressive histone marks.* Nature, 2009. **461**(7265): p. 762-7.

24. Hansen, K.H., et al., *A model for transmission of the H3K27me3 epigenetic mark.* Nat Cell Biol, 2008. **10**(11): p. 1291-300.

25. Alkema, M.J., et al., *Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex.* Genes & Development, 1997. **11**(2): p. 226-240.

26. Francis, N.J., et al., *Reconstitution of a Functional Core Polycomb Repressive Complex.* Molucular Cell, 2001. **8**: p. 545-556.

27. Gunster, M.J., et al., *Identification and Characterization of Interactions between the Vertebrate Polycomb-Group Protein BMI1 and Human Homologs of Polyhomeotic.* Molecular and Cellular Biology, 1997. **17**(4): p. 2326-2335.

28. Satijn, D.P.E., et al., *RING1 Is Associated with the Polycomb Group Protein Complex and Acts as a Transcriptional Repressor.* Molecular and Cellular Biology, 1997. **17**(7): p. 4105-4113.

29. Shao, Z., et al., *Stabilization of Chromatin Structure by PRC1, a Polycomb Complex.* Cell, 1999. **98**(1): p. 37-46.

30. Franke, A., et al., *Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster.* The Embo Journal, 1992. **11**(8): p. 2941-2950.

31. Simon, J.A. and R.E. Kingston, *Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put.* Mol Cell, 2013. **49**(5): p. 808-24.

32. Simon, J.A. and J.W. Tamkun, *Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes.* Current Opinion in Genetics & Development, 2002. **12**: p. 210-218.

33. Schoeftner, S., et al., *Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing.* The Embo Journal, 2006. **25**: p. 3110–3122.

34. de Napoles, M., et al., *Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation.* Dev Cell, 2004. **7**(5): p. 663-76.

35. Stock, J.K., et al., *Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells.* Nat Cell Biol, 2007. **9**(12): p. 1428-35.

36. Zhou, W., et al., *Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation.* Mol Cell, 2008. **29**(1): p. 69-80.

37. Francis, N.J., R.E. Kingston, and C.L. Woodcock, *Chromatin compaction by a polycomb group protein complex.* Science, 2004. **306**(5701): p. 1574-7.

38. Levine, S.S., et al., *The Core of the Polycomb Repressive Complex Is Compositionally and Functionally Conserved in Flies and Humans.* Molecular and Cellular Biology, 2002. **22**(17): p. 6070-6078.

39. Aloia, L., B. Di Stefano, and L. Di Croce, *Polycomb complexes in stem cells and embryonic development.* Development, 2013. **140**(12): p. 2525-34.

40. Zentner, G.E. and S. Henikoff, *Regulation of nucleosome dynamics by histone modifications.* Nat Struct Mol Biol, 2013. **20**(3): p. 259-66.

41. Goll, M.G. and T.H. Bestor, *Histone modification and replacement in chromatin activation.* Genes Dev, 2002. **16**(14): p. 1739-42.

42. Plath, K., et al., *Role of histone H3 lysine 27 methylation in X inactivation.* Science, 2003. **300**(5616): p. 131-5.

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

44. Cui, K., et al., *Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation.* Cell Stem Cell, 2009. **4**(1): p. 80-93.

45. Grewal, S.I. and S. Jia, *Heterochromatin revisited.* Nat Rev Genet, 2007. **8**(1): p. 35-46.

46. Grewal, S.I. and S.C. Elgin, *Transcription and RNA interference in the formation of heterochromatin.* Nature, 2007. **447**(7143): p. 399-406.

47. Zee, B.M., et al., *In vivo residue-specific histone methylation dynamics.* J Biol Chem, 2010. **285**(5): p. 3341-50.

48. Fischle, W., et al., *Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains.* Genes Dev, 2003. **17**(15): p. 1870-81.

49. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.

50. Ringrose, L., et al., *Genome-Wide Prediction of Polycomb/Trithorax Response Elements in Drosophila melanogaster.* Developmental Cell, 2003. **5**: p. 759-771.

51. Cavalli, G., *Epigenetic Inheritance of Active Chromatin After Removal of the Main Transactivator.* Science, 1999. **286**(5441): p. 955-958.

52. Kassis, J.A. and J.L. Brown, *Polycomb group response elements in Drosophila and vertebrates.* Adv Genet, 2013. **81**: p. 83-118.

53. Peng, J.C., et al., *Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells.* Cell, 2009. **139**(7): p. 1290-302.

54. Sing, A., et al., *A vertebrate Polycomb response element governs segmentation of the posterior hindbrain.* Cell, 2009. **138**(5): p. 885-97.

55. Woo, C.J., et al., *A region of the human HOXD cluster that confers polycomb-group responsiveness.* Cell, 2010. **140**(1): p. 99-110.

56. Meng, S., et al., *Identification and characterization of Bmi-1-responding element within the human p16 promoter.* J Biol Chem, 2010. **285**(43): p. 33219-29.

57. Wang, L., et al., *Hierarchical recruitment of polycomb group silencing complexes.* Mol Cell, 2004. **14**(5): p. 637-46.

58. Atchison L., et al., *Transcription factor YY1 functions as a PcG protein in vivo.* The Embo Journal, 2003. **22**(6): p. 1347 - 1358.

59. Mendenhall, E.M., et al., *GC-rich sequence elements recruit PRC2 in mammalian ES cells.* PLoS Genet, 2010. **6**(12): p. e1001244.

60. Cuddapah, S., et al., *A Novel Human Polycomb Binding Site Acts As a Functional Polycomb Response Element in Drosophila.* PLoS One, 2012. **7**(5).

61. Hekimoglu, B. and L. Ringrose, *Non-coding RNAs in polycomb/trithorax regulation.pdf.* RNA Biology, 2009.

62. Zhao, J., et al., *Genome-wide identification of polycomb-associated RNAs by RIP-seq.* Mol Cell, 2010. **40**(6): p. 939-53.

63. Kanhere, A., et al., *Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2.* Mol Cell, 2010. **38**(5): p. 675-88.

64. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.* Science, 2008. **322**(5902): p. 750-6.

65. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.* Nature, 2010. **464**(7291): p. 1071-6.

66. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes.* Science, 2010. **329**(5992): p. 689-93.

67. Ku, M., et al., *Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains.* PLoS Genet, 2008. **4**(10): p. e1000242.

68. Pandey, R.R., et al., *Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation.* Mol Cell, 2008. **32**(2): p. 232-46.

69. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.

70. Lempradl, A. and L. Ringrose, *How does noncoding transcription regulate Hox genes?* Bioessays, 2008. **30**(2): p. 110-21.

71. Hekimoglu, B., *Analysis of Transcription and Chromatin Profiles during Mammalian Differentiation Defines Transcribed Intergenic Polycomb Target Sites.* 2012.

72. Aubert, J., et al., *Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice.* Proc Natl Acad Sci U S A, 2003. **100 Suppl 1**: p. 11836-41.

73. Soriano, P., *Generalized lacZ expression with the ROSA26 Cre reporter strain.* Nature genetics, 1999. **21**.

74. Casola, S., *Mouse Models for miRNA Expression: The ROSA26 Locus.* MicroRNAs and the Immune System: Methods and Protocols, Methods in Molecular Biology, 2010. **667**.

75. Boyer, L.A., et al., *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. **441**(7091): p. 349-53.

76. Lechardeur, D. and G.L. Lukacs, *Nucleocytoplasmic Transport of Plasmid DNA: A Perilous Journey from the Cytoplasm to the Nucleus.* Human Gene Therapy, 2006. **17**: p. 882-889.

77. Reeves, R., C.M. Gorman, and B. Howard, *Minichromosome assembly of non-integrated plasmid DNA transfected into mammalian cels.* Nucleic Acids Res, 1985. **13**.

78. Okulski, H., et al., *Quantitative analysis of polycomb response elements (PREs) at identical genomic locations distinguishes contributions of PRE sequence and genomic environment.* Epigenetics Chromatin, 2011. **4**: p. 4.

79. Yuan, W., et al., *Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation.* Science, 2012. **337**(6097): p. 971-5.

80. Dow, L.E., et al., *A pipeline for the generation of shRNA transgenic mice.* Nat Protoc, 2012. **7**(2): p. 374-93.

10 ACKNOWLEDGEMENTS

The last year was an experience, I will not forget so easily. I was very lucky because I have gone through a lot with people I got to appreciate so much. I have found very good colleagues and friends in the Ringrose lab. First of all I want to thank Leonie who gave me the opportunity to do my Master thesis under her supervision. It was a joy working for and with you! You have always approached me with a lot of sensitiveness and respect and your positive attitude has always encouraged me to go on and enjoy every day I spent here. I want to thank Betül who has started this study and by this built the grounds for this thesis. Thank you for being so patient and answering to all of my e-mails. Thank you, Robert, Tanja, Veronika, Johanna, Philipp, Joao, Christina, Helena, Christine, Eva and Moritz. Thank you for your help, for your feedback, for inspiring discussions and for the great time I had being a part of this group.

Although some of you might know how hard it can be to deal with me, there is only one person who actually wants to do this every day. Thank you Matija for being such a good friend. Living this routine madness day in and day out has significantly proven to be extraordinarily joyful only with you.

Finally I want to thank my parents…
Hvala mama i tata. Bez vas ništa od ovoga nikada nebih bilo pisano. Ja znam, ovaj rad neće ništa na svijetu promjeniti ali za mene ovaj rad puno znači. On znači, da ste vi meni dali mogućnost da idem na fakultet, da diplomiram. Ja znam, da vam nikada nije bilo lahko. Ali nadam se, da će ovaj rad mozda I vama pokazati, koliko ste uspjeli, koliko ste postigli. Tata, ti si uvijek bio moj leksikon. Od kad sam malena znam, ako nesto nerazumijem, ti ćes mi objasniti. Sjećam se, kako si mi pomagao sa zadaćom. Imam osjećaj, iako onda nisam puno razumjela, sta si mi objašnjavao, toliko puno sam ipak naučila od tebe. Naprimjer, da za sve jedan uspjeh u životu treba disciplina I strpljenje. Mama, ti si mi uvijek bila jaka podrška. Znam da sebi sama nisi nikada mogla dati toliko puno snage koliko si meni dala svaki dan. Često si rekla, koliko ti je zao, sto mi nisi kupila Schultüte za prvi školski dan. Mama, ovo je moja Schultüte. Volim vas.

Vienna, 17.10.2013

11 CURRICULUM VITAE

MELITA TICEVIC

address
Ungargasse 71/2/9
1030 Vienna, Austria

mail melita.ticevic@imba.oeaw.ac.at
ticevic33@gmail.com
phone +43 660 542 11 15

Profile

Nationality: Austrian
Birthdate: July 17th 1985
Birthplace: Ključ, Bosnia

Experience

2013 Research technician, IMBA Vienna, Austria

2012-2013 Master thesis, IMBA Vienna, Austria
Project on "Characterization of transcribed intergenic Polycomb sites in mammalian embryonic stem cells" in Dr. Leonie Ringrose's laboratory

2011 Bachelor thesis, AKH Vienna, Austria
Project on "Immune activity of secretion products from irradiated apoptotic peripheral blood mononuclear cells" in Dr. Hendrik Jan Ankersmit's laboratory

Education

2011 – 2013 Master studies in Genetics and Developmental Biology, University of Vienna, Austria

2008 – 2011 Bachelor studies in Molecular Biotechnology, FH Campus Wien, Vienna, Austria

2006 – 2008 Trainee in wholesaling, Bosch GmbH, Vienna, Austria

2004 – 2006 Undergraduate studies in Medicine, Medical University of Vienna, Austria

Additional Education

2013 R for biologists, IMBA, Vienna, Austria

2012 Epigenesys workshop on Epigenetics and Environment, Munich, Germany

2010 Dialog Gentechnik "Science or Fiction?", Vienna, Austria

2010 2nd EACTS Meeting on Cardiac and Pulmonary Regeneration, Vienna, Austria