universität wien

MASTERARBEIT

Titel der Masterarbeit

„Functional characterization of mammalian Polycomb Response Elements “

verfasst von

Tanja Drexel BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 066 834
Studienrichtung lt. Studienblatt:	Masterstudium Molekulare Biologie
Betreut von:	Dr. Leonie Ringrose

Table of Contents

1 Abbreviations

AmpR	Ampicillin resistance
Ash2L	Ash2 (absent, small, or homeotic)-like
Ash1l	Absent, small, or homeotic-like
bp	Base pairs
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprcipitation-sequencing
dH2O	Distilled water
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide-triphosphate
E.Coli	Escherichia coli
EED	Embryonic ectoderm development
EGFP	Enhanced Green Fluorescent Protein
Esc	Extra sex comb
ESC	Embryonic stem cells
EtOH	Ethanol
Ezh2	Enhancer of ZESTE
HMTase	Histone methyltransferase
Jarid2	Jumonji, AT rich interactive domain 2
kb	Kilobase pairs of DNA
LB	L(uria) B(ertani) Medium
MLL	Myeloid/lymphoid or mixed-lineage leukemia
MLL	Myeloid/lymphoid or mixed-lineage leukemia
NaCL	Sodium Chloride
nm	Nanometer
o/n	Over night
PBS	Phosphate Buffered Saline
PC	Polycomb
PcG	Polycomb-Group
PCR	Polymerase chain reaction
PH	Polyhomeotic
PHO	Pleiohomeotic
PRC	Polycomb repressive complex
PRE	Polycomb Response Element
PSQ	Pipsqueak
qPCR	Quantitative real time PCR
RNA	Ribonucleic Acid
RT	Room temperature
siRNA	Small interfering RNA
SU(Z)12	Supressor of Zeste
Suz12	Suppressor of zeste 12 homolog
TIPs	Transcribed intergenic Polycomb sites
TRE	Trithorax Response Element
TrxG	Trithorax-Group
WT	Wild type

2 Abstract (english)

During development it is the main aim of multicellular organisms to grow not only in size but also to specialize and form different cells and tissues. This is accomplished by changes in the transcriptional profile of the cells, which has to be maintained during several cell divisions. The highly conserved Polycomb (PcG) and Trithorax (TrxG) group proteins constitute a gene regulatory system that is essential for maintaining the correct identity of both stem cells and differentiated cells [1, 2]. The PcG and TrxG proteins act through cis-regulatory elements called Polycomb/Trithorax response elements (PREs) which are crucial for epigenetic inheritance [2-4]. PREs are best characterized in flies, and can be predicted even genome wide via an algorithm specifically designed for *Drosophila melanogaster* [5]. However, although PcG and TrxG proteins are highly conserved, PREs are not, and therefore a computational approach to predict PREs in mammals has previously been lacking. A bioinformatician in the Ringrose group was able to adapt the fly algorithm to the mammalian system and could predict several putative PREs in the mouse genome. Since these predictions had to be confirmed it was my task to revalidate them experimentally. I integrated several PRE reporter constructs stably into ROSA-26 in mouse ESCs, to avoid genomic position effects and measured the luciferase expression. For this purpose I took advantage of an already established luciferase reporter assay system and tested the repressive effect of the PREs on the reporter in transient experiments and stable cell lines. In order to address the question whether the silencing of the reporter is PcG dependent, I furthermore depleted PcG proteins via siRNA knockdowns. In this thesis I could show that several PREs showed silencing of the reporter in transient experiments whereas they behaved differently in Rosa-26 in stable ES cell lines. The PRE Ptch1 was able to silence the reporter even in the stable cell lines. Furthermore I contributed in the improvement of the gene targeting workflow and could show possible limitations of the luciferase reporter assay system, which will be important for future experiments.

3 Zusammenfassung (Deutsch)

Mehrzellige Organismen sind während ihrer Entwicklung bestrebt nicht nur an Größe zuzunehmen, sondern auch spezialisierte Strukturen in Form unterschiedlicher Zellen und Gewebe auszubilden. Dies wird vor allem durch Änderungen im transkriptionellen Profil der Zellen bewerkstelligt welches über viele Zellgenerationen.erhalten bleiben muss. Die hoch konservierten “Polycomb und Trithorax group” (PcG und TrxG) Proteine stellen ein gen-regulatorisches System zur Verfügung, welches nicht nur für die Erhaltung des Status differenzierter Zellen, sondern auch von Stammzellen verantwortlich ist [1, 2]. PcG und TrxG Proteine wirken über Polycomb Response Elements (PREs), welche daher eine entscheidende Rolle in der epigenetischen Vererbung spielen [2-4]. PREs sind am besten in Fliegen beschrieben und können im Genom von *Drosophila Melanogaster* sogar mittels eines für diesen Modellorganismus spezifisch designten Algorithmus bestimmt werden [5]. Denoch, obwohl TrxG und PcG hoch konserviert sind, sind PREs nicht konserviert und aus diesem Grund fehlt bis heute ein Algorithmus für die Bestimmung von PREs im Säugetiergenom. Aus diesem Grund hat eine Bioinformatikerin in der Ringrose Forschungsgruppe den Fliegen-Algorithmus an das Säugetiergenom angepasst und konnte dadurch mehrere PREs bestimmen. Da bioinformatisch getroffene Vorhersagen experimentel bestätigt werden müssen, war es meine Aufgabe die Funktion der PREs zu charakterisieren. Dafür habe ich mich eines, schon teilweise etablierten Luciferase Reporter Systems bedient und habe eine Auswahl verschiedener mutmasslicher PREs in das Reporter Konstrukt kloniert. Diese PRE-Reporter Konstrukte wurden mittels homologer Rekombination stabil in den Rosa-26 locus in embryonale Stammzellen der Maus integriert. Die Wirkung der PREs auf den Luciferase Reporter wurde im transienten Experiment und in den stabilen Zelllinien bestimmt. Dabei zeigte sich dass Ptch1 den Reporter in beiden experimentllen Ansätzen reprimieren konnte. Des weiteren versuchte ich die Frage, ob dieser reprimierende Effekt PcG abhängig ist, zu adressieren indem ich mittels siRNAs geziehlt den Abbau von wichtigen Mitgliedern des “Polycomb Repressive Complexe 2” (PRC2) Komplexes hervorgerufen habe. Des weiteren habe ich dazu beigetragen mögliche Limitationen dieses Systems auf zuzeigen, was für zukünftige Experimente eine wichtige Rolle spielen wird.

4 Introduction

4.1 Epigenetic inheritance during development

Starting from the first totipotent stem cell it is the main aim of the developing organism not only to grow in size, but also to develop special functional units that enable the organism to live in its predestined environment. This is accomplished primarily via an increased number of cell divisions and cell differentiations. The fact that every cell has the same DNA content puts high challenges on the system, since different cell types and tissues can only arise from different gene expression patterns. But that is not sufficient; this pattern has to be maintained during several cell divisions and therefore has to be transmitted from mother to daughter cells. Furthermore this is also relevant for stem cells, which have to maintain their self renewal capacity [6] The highly conserved Polycomb (PcG) and Trithorax (TrxG) group proteins [1, 7] can maintain the transcriptional status of differentiated and stem cell long after the factors, which initiated the status have disappeared. [1, 2, 8-12]. This kind of heritability of the transcriptional status can be termed as epigenetic according to following classical definition: “a change in the state of expression of a gene that does not involve a mutation, but that is nevertheless inherited in the absence of the signal (or event) that initiated the change” [13], (cited in Ringrose and Paro 2007).

4.2 Polycomb (PcG) and Trithorax (TrxG) group proteins

PcG and TrxG where discovered as mediators for epigenetic memory after it was observed that mutations in Polycomb (PC) and Extra Sex Combs (esc) led to a miss-expression of the Hox gene cluster in flies (reviewed in[1]).Hox genes control the correct body plan, and especially the identities of different segments, and must be tightly regulated throughout development [14]. Upon mutation of members of the PcG family a derepression in certain Hox genes and transformations of one body segment into another could be observed [15, 16]. Based on these observations PcG proteins were suggested to be involved in the regulation of the Hox gene expression pattern. Later after this discovery another class of proteins was discovered to be involved in

the regulation of the Hox gene cluster: TrxG proteins were identified to work against the PcG mutations (reviewed in [1]). Furthermore it was not only suggested that PcG proteins are involved in the regulation of the Hox gene expression pattern but also that PcG and TrxG work antagonistically on their target genes [17, 18]

The repressive function was thereby assigned to PcG and the activating function to TrxG proteins. Their mammalian counterparts could be identified later based on the high conservation of PcG and TrxG by using homology studies [19-21]. To date, many more PcG and TrxG targets have been identified, since rapid progress has been made in the genome wide profiling sector [22, 23]. Furthermore several biochemical studies have revealed that PcG and TrxG proteins act in large complexes where they share tasks in order to silence or activate the appropriate target genes [24-26]

The catalytic subunit of each protein complex is responsible for several histone modifications. Histone H3K27 trimethylation placed by PcG complexes is associated with silencing [27-30], whereas H3K4me3 and H3K36me3 are set by TrxG complexes and associated with transcriptional activation [31, 32]. Another histone mark which is associated with active chromatin is hyperacetylation at H3K27, which not only antagonizes H3K27 methylation [33-35], but also influences the intensity of DNA and histone attraction via neutralization of the positive lysine charge of the histone tail [36].This destabilization of the interaction between DNA and histones increases chromatin accessibility and this in turn could facilitate the binding of other large protein complexes like Polymerase II. [37]. Another histone mark, which is added by the PRC1 complex, is the monoubiquitilation of histone H2A [38-40]. This histone mark causes a poised state of RNA polymerase II in ESCs [41, 42] and therefore also contributes to transcriptional silencing. As PcG and TrxG proteins have such an important role for differentiation and maintanance of pluripotency during development, null alleles for many PcG and TrxG members are fatal as they lead to the death of the animal (flies and mice) in early stages [9, 43-48]. Furthermore aberrant expression of these proteins is also associated with many types of cancer.[49, 50]

4.2.1 PcG and TrxG protein complexes in files and mammals

4.2.1.1 Polycomb repressive complex 2 (PRC2)

The PRCII complex includes the core proteins EZH2 (fly E(z)), SUZ12 (fly Su(z)12) and EED (fly Esc). EZH2 (flies E(Z)) is responsible via its methyltransferase activity for H3K27me3 in flies and vertebrates [29, 51-53]. Suz12 is indispensable for embryonic development as mutations in Suz12 lead to a global loss of H3K27me3 [54] and early embryonic death [43], since it is important to enhance EZH2 activity[43] and also for cell differentiation of ESCs [55]. EED (flies: ESC) in turn is another enhancer of the methyltransferase activity[56].

4.2.1.2 Polycomb repressive complex 1 (PRC1)

The H3K27me3 mark which is added by the PRC2 complex via the EZH2 Set domain [29, 51-53] can be recognized by the fly and mammalian PRC1 complex via the chromodomain of the CBX (fly PC) proteins [27, 57] Other important core proteins of the PRC1 complex are BMI-1 (fly PSC), which has been shown to positively regulate the H2K119 ubiquitination activity of RING1B [38]. Monoubiquitilation at this position leads to silencing of target genes via the poising of RNA Polymerase II in ESCs [42, 58]. Other important complex members are RING1A/ RING1B, which play an important role in the monoubiquitination of the H2A histone [38-40] .In addition it has been shown that the PRC1 complex induces compaction of nucleosoms [59] and inhibits the chromatin remodeling which leads in combination to the formation of repressive chromatin [60, 61]

4.2.1.3 DNA binding proteins

In flies several DNA binding proteins have been identified which interact with PcG complexes. Examples are the proteins Zeste [62, 63], GAGA binding factor (GAF)[57] [64], Pipsqueak (Psq) [65], Dorsal switch proteins 1 (Dsp1) [66] and Pleiohomeotic (Pho) [67]. In mammals it has been demonstrated that YYI (flies: Pho) interacts with PcG proteins [68, 69]. In addition, the DNA binding protein JARID2 [70-74] has been reported to be involved in Polycomb recruitment to their target genes. [70-73] Loss of

JARID2 leads to failures in stem cell differentiation and a global loss of H3K27me3 [8, 72]

Taken together, based on these observations one could conclude that the PRC2 complex is responsible for the H3K27me3, which are used as anchoring points for the PRC1 complex. The binding of PRC1 could finally lead to silencing of the PcG target via Ubiquitination of histone H2A, leading to Pol II arrest, and compaction of chromatin. However it has been reported that the PRC1 complex can bind not only independently from H3K27me3 [57, 75], indeed also independently from histones themselves to naked DNA [59]. Furthermore it has been reported that PRC2 can bind directly to H3K27 and that this mechanism is used for the maintenance of silencing from mother to daughter cells after [30]. However recently quite the contrary has been postulated as it has been shown that TrxG and PcG proteins but not methylated histones remain associated with DNA through replication [76]. Thus, most of the data show evidence supporting that H3K27me3 is somehow involved in PcG regulation. Whether this histone mark is the reason, the consequence or both stays unclear, as well as the hierarchic order of PcG recruitment.

4.2.1.4 Trithorax group (TrxG) proteins

ASH1L (fly: Ash1) is required for maintenance of active Hox gene expression [77] and seems to be in general associated with an active state of chromatin and transcription [78]. Ash1L furthermore co-localizes with MLL, which is in turn the SET domain containing histonmethyltransferase. MLL exists in complexes with ASH2L and WDR5, which seems to be important for a functioning methylation of histone 3 at lysine 4 (H3K4me3) [77, 79].

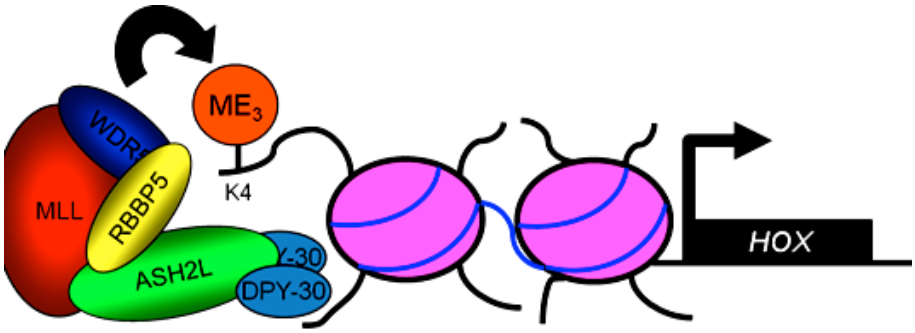


Figure 1| TrxG proteins in the histone methyltransferase complex. Ash2L increases the activity of the MLL complex whereas a knockdown of Ash2l causes a genome wide decrease in H3K4me3 [79]. (figure from http://atlasgeneticsoncology.org/Deep/ASH2LFunctionID20097.html)

4.2.1.5 Transcribed intergenic Polycomb (TIP) sites

Transcribed intergenic Polycomb target sites (TIPs) were recently identified by using a combination of transcriptional profiling and ChIP analysis [80]. TIPs are a novel class of PcG target sites in mammals, which can recruit PcG proteins and are transcribed into ncRNA. In reporter assays based on transient transfection of luciferase constructs, transcribed TIPs could repress a flanking gene. TIPs are found in intergenic positions and were used as training data for the computational prediction of mammalian PREs (see 4.3)

4.2.1.6 Polycomb Response elements (PREs)

The PcG and TrxG proteins act through cis regulatory elements called Polycomb and Trithorax response elements (PRE/TRE), or in this thesis simplified to PRE. In flies PREs respond to PcG and TrxG proteins in order to give epigenetic memory of the active and the silent transcriptional states. PREs are best described in flies as they were discovered in the *Drosophila Bithorax complex* [3, 4, 24]. Very different behaviors of two classes of cis-regulatory elements were observed, and they were separated into initiator elements and maintenance elements on the basis of these properties [16, 81]. Later on, the maintenance elements were called PREs and it was than demonstrated that they can determine an epigenetically inherited state of repression [76, 82]. More fly PREs could be defined experimentally, but a big step

forward in this field was made, when an alignment-independent algorithm was designed. This algorithm was based on the binding motifs of DNA binding proteins (4.2.1.3), which were co- purified or co-localised with the fly PcG proteins. Via this approach the PRE/TREs of the BX-c could be successfully identified, and furthermore on a genome wide level new candidate PREs could be predicted which were also verified experimentally [5]. In mammals, to date only a handful of PREs have been identified, including PREs in the mouse MafB/Kreisler [83], the human HoxD loci [84] and also the mouse En2 and Igf2 loci (R. Heinen et al., manuscript in preparation). Nevertheless the sequence properties of a mammalian PRE have remained highly elusive. This is mainly because of the lack of conservation between flies and mammals regarding the sequences, which underlies a PRE or the DNA binding proteins, which bind PcG proteins in flies [62, 63], 57, [64].

Another hypothesis involves the role of CpG rich sequences in PcG recruitment. A model has been proposed where transcriptionally inactive GC-rich DNA itself is sufficient to recruit Polycomb repressive complex 2 (PRC2) to an ectopic locus in mouse ESCs [85]. Furthermore the importance of unmethylated CGs has been investigated and it has been proposed that in the absence of transcriptional activators PcG recruitment is the default state, at long stretches of unmethylated CGs [86]. CG stretches may play an important role in PcG recruitment. However approximately 70% of promoters contain CpG island [87] whereas it is estimated that only about 7% of genes in ESCs are targeted by PcG [71]. Thus also other sequence feature must play a role in PcG recruitment beneath CpG islands.

Much progress has been made in the generation of genome wide profiling data sets for PcG proteins and H3K27me3 marks [71, 88] [89-93] [94] for several mammalian cell types. However it is not possible to identify PREs simply from PcG binding profiles or the distribution of H3K27me3 marks. This difficulty arises from the fact that PcG proteins spread over large distances, including the promotor region, the PRE itself and even the gene body (Figure 2). Another reason which adds to the difficulty of identifying mammalian PREs is the fact that chromatin tends to form loops between regulatory elements and target structures, which makes it even more difficult to find the source of Polycomb binding (reviewed in [95])

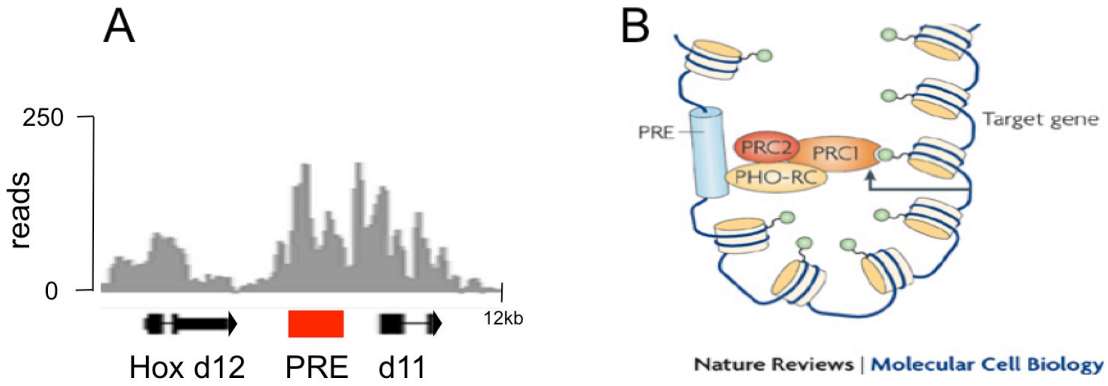


Figure 2| A) PcG proteins spreading. Grey: ChIP-seq data for the PRC2 protein Jarid 2 are shown for the mouse Hox d11 and 12 genes. Red bars show characterised PREs. Hoxd11-12 PRE was identified in the human genome [84] (figure A by Johanna Trupke) B) Looping of Chromatin, between PRE and target sequences (figure B by [95])

In summary, as the sequence principles of the PREs are not conserved, it is very inefficient to determine only experimentally which of the PcG peaks is the PRE and which peak not. Thus the most efficient way to detect mammalian PREs would be a computational prediction algorithm adapted to the mammalian system.

4.3 Computational prediction of mammalian PREs

All bioinformatic procedures, data and figures in this section (computational prediction of mammalian PREs) are performed and created by Johanna Trupke. (Unless otherwise indicated)

In flies an algorithm for computational prediction of PREs was designed based on 3 classes of binding sites for GAF, Pipsqueak (which share a common motif), Zeste and Pho. The role of these proteins in regulating PREs is reviewed in [1]. The binding motifs of these proteins were paired with each other and assigned to a weight. This was done by determination of how often the motifs pairs occurred in a positive compared to a negative dataset. The positive dataset was composed of an already experimentally validated set of PRE/TREs, whereas the negative dataset consisted of non-PRE/TREs, which also contained the motifs in question. Scores for each motif pair were calculated by counting their occurrence for windows of 500 bp multiplied with their weight. The window was moved in 100 bp steps across the sequences of

the positive and negative dataset and the thereby generated values for each window for a given sequence were added together, giving a score. This scoring system was than also applied genome - wide to detect already known and new PREs in flies. A modification of the procedure and generalization to user- entered motifs and training data was developed by the Rehmsmeier group [96]

The main aim of Johanna Trupkes project (a PhD student in the RIngrose lab) was now to find out if the same method could be used in order to understand the sequence principles of Polycomb recruitment in mammals. However, unlike in flies, for this project it was necessary to start further upstream in the prediction pipeline. This is due to the fact that the motifs of the DNA binding proteins are not conserved between the flies and vertebrates. Furthermore another source for a positive and negative training data was required, in order to extract novel DNA motifs and on order to proceed then as in flies. The TIP sites mentioned above (4.2.1.5) in combination with other published genome-wide Polycomb ChIP-seq data, more precisely intergenic peaks for Jarid1a Jarid2, Suz12. in ESCs, were used as positive training data. Non PcG bound intergenic sites (Sox2, Jarid1a) were used as negative training data [71, 80, 97]. The next step was then to perform de novo motif discovery. The training data was analyzed in order to identify motifs specifically enriched in Polycomb binding sites and the negative data sets. Indeed via the usage of different algorithms **[98-100]**, 10 motif clusters could be extracted. Furthermore to ensure that they were specific to Polycomb binding sites, the jPredictor program was applied on the different motifs. **[96]**. As in flies, weights were generated for all possible motif pairs, by training on Jarid1a and TIPs (Equation 1). Furthermore these weights were than used to score the training data. (

Equation 2).

$$w(M) = \ln \frac{f(M|positive)/kb}{f(M|negative)/kb}$$

Equation 1| Formula for calculating the weights of the different motif pairs w (M). Occurrence of Motif pair (M) per Kilobase (kb) in the positive compared to the negative

$$windowscore(S) = \sum_{m} w(M) \times o(M)$$

Equation 2| Formula for scoring sequences. o(M) occurrence of a motif pair in the scoring window (S).

As anticipated the scoring values clearly separated the negative from the positive training data. In order to be able to evaluate scores on a genome wide level, a cutoff was defined by scoring 100 randomized mouse genomes. This cutoff was than used to define regions, which are putative PREs (Table 1, Figure 3). Predicted putative PREs and control sequences were selected to test them experimentally. The selection included PREs of different categories. These categories are defined and composed of different parameters in different combinations (Table 1). Important parameters were whether the sequence was predicted or not in ESCs (band: true/false) or whether regions showed ChiP enrichment for Jarid2 and Suz12 in ESCs (ESC PcG OVERLAP). Furthermore data from the ENCODE project, more precisely H3K27 ChIP enrichment in differentiated cells (ENCODE H3K27me3) was taken into account. These parameters in different combinations resulted in the different categories. Furthermore the occurrence of CpG islands are also mentioned in Table1 since they played an important role in publications about PcG recruitment [86, 91, 101]. During my master thesis project I tested the selection of putative PREs experimentally (see Table1)

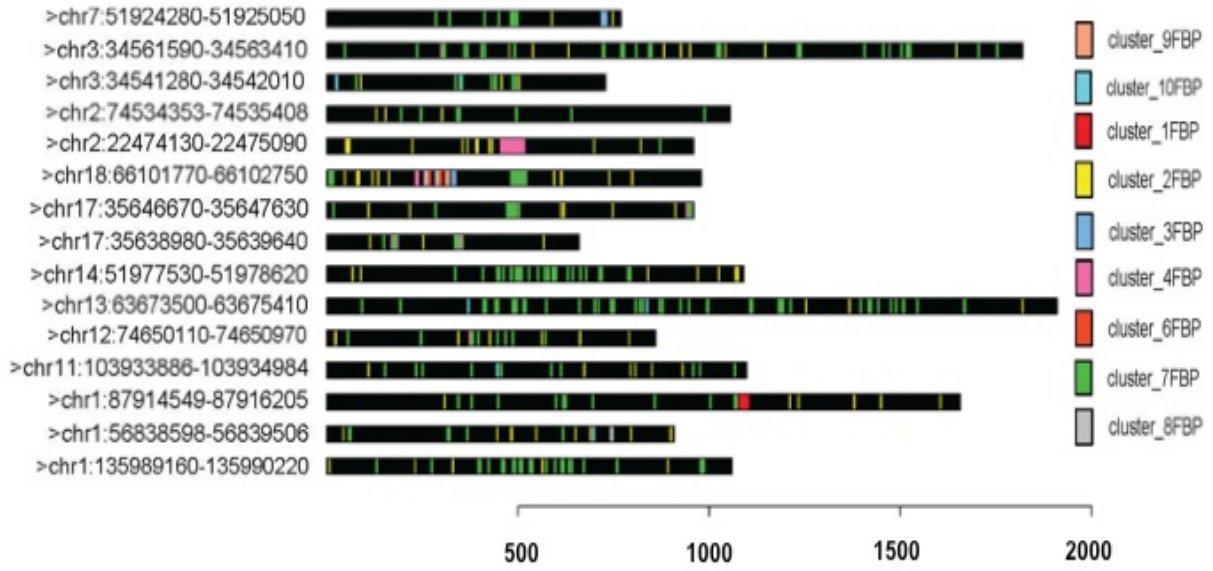


Figure 3| Motif composition of the sequences, which were tested experimentally

Name	Band	ENCODE H3K27me3	ESC PcG OVERLAP	Category	CpG islands
HoxD9-up	FALSE	TRUE	TRUE	Chip-only	FALSE
Spata3 chr1	FALSE	TRUE	TRUE	Chip-only	TRUE
Gene desert chr11	FALSE	TRUE	TRUE	Chip-only	FALSE
Gene desert chr1	FALSE	TRUE	TRUE	Chip-only	FALSE
Ptch1	TRUE	TRUE	TRUE	prediction+ESC+ENCODE	TRUE
Gad2/Myo3A chr2	TRUE	TRUE	FALSE	Differentiation only	FALSE
Tmem chr12	TRUE	TRUE	FALSE	Differentiation only	FALSE
Chit1	TRUE	TRUE	FALSE	Differentiation only	TRUE
Myh14 chr7	TRUE	TRUE	FALSE	Differentiation only	FALSE
Sox2 upstream	TRUE	TRUE	FALSE	Differentiation only	FALSE
Rax chr18	TRUE	TRUE	FALSE	Prediction + ENCODE	FALSE
Oct4-upstream	TRUE	TRUE	FALSE	Prediction + ENCODE	FALSE
Oct4-gene end	TRUE	TRUE	FALSE	Prediction + ENCODE	FALSE

Table 1| Categories of the experimentally tested PREs. Band (prediction) refers to regions with scores above cutoff. ENCODE H3K27 refers to H3K27 ChIP peaks in differentiated cells [94], whereas the column ESC PcG Overlap shows whether there is a ChIP enrichment for Jarid2 and Suz12 in ESC (TRUE) or not (FALSE). CpG island coordinates were downloaded from the USCS genome tables. Category black contains control sequences, which are not predicted to function as a PRE, although they show Jarid2 and Suz12 peaks in ESCs and peaks for H3K27me3 in differentiated cells. The red category contains a PRE, which is predicted to silence in ESCs and differentiated cells. The green category is composed of PREs, which are predicted to silence only in differentiated cells since there is no overlap with Jarid2 or Suz12 peaks in ESCs.

4.4 Luciferase reporter assay system

In order to perform a quantitative analysis of the putative PREs a reporter assay system was used, which was previously designed in the Ringrose lab, to address the following hallmarks of PRE function:

- Silencing of the reporter at an ectopic locus
- Loss of silencing of the reporter upon PcG depletion
- Recruitment of PcG proteins to the candidate PRE at an ectopic locus (for review see [1])

The system is based on following principles: Candidate PREs are cloned upstream of a promoter and luciferase reporter in order to determine possible PcG recruitment to the PRE and associated reporter repression. A minimal promoter was used in previous experiments, which identified the Engrailed (En) -PRE and Insulin-like growth factor 2 (Igf2) -PRE, to drive the expression of the luciferase gene. For my master thesis the minimal promoter was exchanged against the stronger viral TK promoter. The three hallmarks of a PRE can be addressed with following experiments:

4.4.1 Silencing of the reporter at an ectopic locus

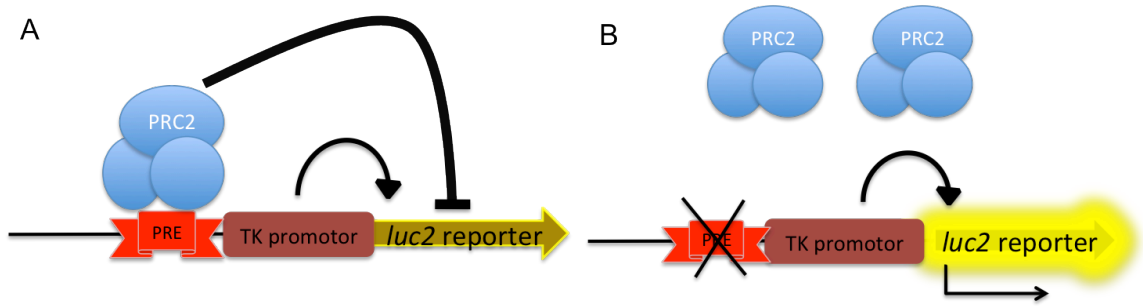


Figure 4| Luciferase reporter assay construct A) PRE recruits PcG, which leads to a silencing of the reporter gene. B) In case of the control construct, or in case that the predicted PRE doesn't fulfill the first hallmark of a PRE, PcG will not be recruited and the reporter will not be silenced

In order to determine if the predicted PREs have the ability to silence their corresponding gene they were inserted upstream of the TK promotor and the luciferase reporter. These constructs were than integrated into the ROSA-26 locus in mouse embryonic stem cells by homologues recombination to avoid genomic position effects and to guarantee a quantitative read out of the luciferase expression.

Reporter constructs carrying a PRE should attract PcG, which leads to silencing of the reporter. In general constructs carrying a PRE should show a lower reporter activity compared to a control line without PRE (see Figure 4).

4.4.2 Loss of silencing of the reporter upon PcG depletion

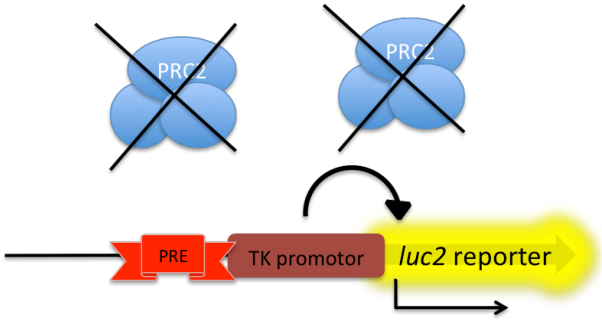


Figure 5| Derepression of the reporter upon depletion of PRC2

The question of whether silencing of the reporter is Polycomb dependent can be addressed by siRNA mediated depletion of Suz12 and Ezh2, important members of the PRC2 complex. In case of a Polycomb dependency the reporter should become de-repressed compared to a control knockdown. (see Figure 5)

4.4.3 Recruitment of PcG proteins to the candidate PRE at an ectopic locus

To evaluate whether direct PcG binding to that locus mediates the PcG dependent repression of the reporter, chromatin immunoprecipitation (ChIP) can be performed on the reporter lines. In case of a Polycomb dependent silencing, enrichment of PcG should be detectable on the transgenic constructs.

5 Aim of this thesis

This thesis addresses the experimental part of a project, which aims to gain further insights into the sequence principles that define mammalian Polycomb Response Elements (PREs). A bioinformatician in the Ringrose group predicted several PREs using an algorithm, which was adapted to the mammalian system. The specific aim of this Masters thesis is to characterize and validate a selection of these candidate PREs with experiments that address the hallmarks of a PRE, namely repression of a reporter gene and derepression upon depletion of Polycomb proteins.

6 Materials and Methods

6.1 Materials

6.1.1 Standard Buffers and Media

Name	Composition
LB- media pH 7.5	Prepared by the IMBA service department: 5g NaCL, 10g Tryptone, 5g Yeast extract, ad 1 L dH_2O- pH adjustment with NaOH
LB-Amp Agar plates pH 7.5	Prepared by the IMBA service department: 16g Agar, 5g NaCL, 2g Tryptone, 5g Yeast extract, – pH adjustment with NaOH
SOB media	Prepared by the IMBA service department: 20 g Tryptone, 5g Yeast Extract, 2ml 5M NaCL, 2.5 ml 1M $MgCl_2$, 10ml 1M MgSO4, ad 1L dH_2O, sterilized by autoclaving
SOC Media	SOB media + 2.5 mM KCL, 20 mM glucose, 10 mM $MgCl_2$
10x TAE buffer pH 8.7	Prepared by the IMBA service department: 48.4g TRIS [0.4M], 3.7g EDTA [0.01M], ad 1L dH_2O – pH adjusted with acetic acid
1x PBS	Prepared by the IMBA service department: 8g NaCl [137mM], 200mg KCl [2.7mM], 1.44g Na_2HPO_4 [10mM], 240mg KH_2PO_4O [2mM], ad 1L dH_2O
5M NaCl	Prepared by the IMBA service department: 292.2g NaCl [5M], ad 1L dH_2O
NaAc pH 5.2.[3M]	Prepared by the IMBA service department: 20.412 g Sodium acetate [= CH3COONa•3H2O, trihydrate],pH adjustment with acetic acid to 50 ml with dH2O
Depurination solution	0.25 M HCL: 25 ml of 37 % HCL ad 1 L
Denaturation solution	0,5M NaOH, (200g NaOH pellets ad 500 ml ddH20), 1,5M NaCl
Neutralization solution	500 ml Tris pH 7.5 (1M) + 300 ml NaCL (5M)
Low stringency buffer	2xSSC, 0.1 %SDS,
Mltra high stringency buffer	0.5 SSC, 0.1 % SDS
20x SSC	Prepared by the IMBA service department
5x passive lysis buffer	Promega
Conventionall cell lysis buffer	0.1 M EDTA (pH 8.0), 0.5% (w/v) SDS, 10 mM Tris-Cl (pH 8.0)

Table 2| Standard buffers and media used for this project

6.1.2 Standard buffers and media ES cell culture

Name	Composition and/or company
DMEM ++: Standard ESC media + LIF	DMEM as basic media (prepared by the IMBA service department) + 10% FCS (Gibco) 1x L-glutamine (Sigma, 100x, 200mM)) 1x Penicillin/Streptomycin (Sigma, 100x) 1x Non-Essential Amino Acids (Sigma, 100x) 1x Sodium Pyruvate (Sigma, 100x, 100mM) 0.1mM β-mercaptoethanol 1000U/ml LIF (Esgro)
DMEM Stop : Standard ESC media - LIF	DMEM (prepared by the IMBA service department) 10% FCS (Gibco) 1x L-glutamine (Sigma, 100x, 200mM)) 1x Penicillin/Streptomycin (Sigma, 100x) 1x Non-Essential Amino Acids (Sigma, 100x) 1x Sodium Pyruvate (Sigma, 100x, 100mM) 0.1mM β-mercaptoethanol 1000U/ml LIF (Esgro)
Freezing media	50% FCS, 40% DMEM ++, 10 % DMSO
1x Trypsin	Diluted from Gibco® 10x Trypsin
Opti-MEM®	Life technologies
0.1 % Gelatin	Diluted from 2% Gelatin in H2O (Sigma)

Table 3| Standard ES cell cμlture related media and reagents

6.1.3 OLIGONUCLEOTIDES

6.1.3.1 PRE amplification primers

Name forward primer	Name reverse primer	forward primer 5'-->3'	reverse primer 5'-->3'	T (a)°C
HoxD9chr2fwd	HoxD9chr2rev	AGCCTCGAGGGGGACGTA TTTTTCTGAGGCTA	ACTGAAGCTTAATAAAAC GGACGGATGGAG	61
Spatachr1fwd	Spatachr1rev	AGCCTCGAGGCTTGAGG CAGAAGGCTGTA	ACTGAAGCTTGCAGTCC TTCATCCTTGCAT	58
Genedeschr11fwd	Genedeschr11rev	ACGAGCTCGCACAGTGTG GAGAGCAAAG	ACTGAAGCTTATGTCAAA GGCCTTCAGTGG	64
Ptchchr13fwd	Ptchchr13rev	AGCCTCGAGGGCATTTTA AACGTGGACTTC	ACTGAAGCTTTCCCTGG ACAGACAACAACA	56
Raxhr18fwd	Raxxhr18rev	AGCCTCGAGCATGTGTAA TGCTACTGCTAGGG	ACTGAAGCTTTAAATGGT GGAAGGCAGAGG	61
GADmyochr2fwd	GADmyochr2rev	AGCCTCGAGGGGGGAGG GTAACAAATTCA	ACTGAAGCTTTCCTCTA GGACTGGTTCCACTT	61
Tmemchr12fwd	Tmemchr12rev	AGCCTCGAGTTCATATCA AACGTGAGTTTTCTCTT	ACTGAAGCTTGGTTGGC ACAGTTAGGGTGA	57
Chit1fwd	Chit1rev	AGCCTCGAGCGGAGCTG ACTCTCTTAAGGT	ACTGAAGCTTTGGTTTGT TTGTTTTGCTTTG	57
Myh14chr7fwd	Myh14chr7rev	AGCCTCGAGAGGTGTCAA CACTGGCCTTT	ACTGAAGCTTACCCCTG TCTCACAGCCTCT	62
Genesechr14fwd	Genesechr14rev	ACGCTCGAGCCCTTCTCC CTCAGGTACAG	ACTGAAGCTTTCAGGAA ACTGAGATTCTAAGAAAG A	56
Sox2upfwd	Sox2uprev	TGGAGCTCGCGAAGGAG TCTTCTCTCCA	TACTCGAGGAAAGAACC CCGAACAAGAA	61
Oct4upfwd	Oct4uprev	ACGCTCGAGTGTGTGTGT GTATCTGAGTGCAA	ACTGAAGCTTAGTCCTC CAGGCCAGCTT	62
Oct4gendfwd	Oct4gendrev	ACGCTCGAGCCCACCAG CCTGACAATG	ACTGAAGCTTCCACCCC TGTTGTGCTTTTA	61

Table 4 List of primer pairs used for the amplification of the predicted PREs from the mouse genome. The list contains also the optimal annealing temperatures determined via gradient PCRs. Sequence parts depicted in green represent the added restriction sites, blue parts show spacer oligonucleotides

6.1.3.2 qPCR primers

Target transcript	forward primer 5'-->3'	reverse primer 5'-->3'	Product size	comments
Suz12	GACAGAAGCCAGAGACGACCT	TTCCTGCATAGGAGCCATCAT	182	intron spanning
Gapdh	TGCGACTTCAACAGCAACTC	CTTGCTCAGTGTCCTTGCTG	200	
Ezh2	GTGACAGAGAAGCAGGGACTG	CACCACTCCACTCCACATTCT	165	intron spanning
MLL	GGCCCTGTGAAAATGATCTG	TCCAAAGGATAGCCATGAGG		Intron spanning
Ash2L	GGCCCTGTGAAAATGATCTG	TCCAAAGGATAGCCATGAGG	179 bp	intron spanning

Table 5| List of primer pairs used for targeting of PcG/TrxG transcripts via qPCR

6.1.3.3 Oligonucleotides for colony PCRs, diagnostic PCRs and the amplification of southern blot probes.

Name	Sequence 5'--> 3'	Purpose of use	Product length
pGl4.26mcsfwd	AGTGCAGGTGCCAGAACATT	Colony PCR: validation of TK promoter insertion into pGL4.26 hygro floxed	900 bp
pGl4.26mcsrev	AACAGTACCGGATTGCCAAG		
pGl4.26mcsfwd	AGTGCAGGTGCCAGAACATT	Colony PCR; validation of insertion of the PRE elements intp pGl4.26 TK hygro floxed	Dependent on the PRE lentgh +150 bP
FlpTK-eff-LCb	CAAACCCTAACCACCGCTTA		
ROSAdiaga	TCTACCAGAGCCTCGTGGAC	Diagnostic PCR for the validation of site directed integration into Rosa-26 in the mouse genom	1563 bp
ROSA-dcloningb	GCTAGTCGACCACCGCCCCA CACTTATT		
ROSA26mcs-fwd	GAGTTCTCTGCTGCCTCCTG	Colony PCR, validation of the insertion of the pGL-PRE construct into the ROSA vector	430 bp
ROSA-PCR-diag a4	CTCGTGGACGGCAACTTCGA		
ROSA26msc-rev	TCTTCTGGGCAGGCTTAAAG	Colony PCR, validation of the insertion of pGL-PRE construct into the ROSA vector	402+ size of PRE element
FlpTK-eff-LCb	CAAACCCTAACCACCGCTTA		
Ro-diag-luc-2a	GGACTTGGACACCGGTAAGA	Amplification of the internal probe for southern blot analysis	439 bp
Ro-diag-luc-2b	GTCCACGAACACAACACCAC		

Ro-diag-mm-s3a	TTTGCCTGGGTATTGCCTAC	Amplification of the external probe for southern blot analysis	467 bp
Ro-diag_mm-S3b	CCTAGGATCTTGGCTTGCAC		
Ro_dTg2a	ACGTTTCCGACTTGAGTTGC	Validation of the removal of the hygromycin cassette in mouse ESCs	707 bp
pGl4.26 Lucpa	AGATCCGCGAGATTCTCATT	Validation of the removal of the hygromycin cassette in mouse ESCs	

Table 6| List of primer pairs used in different tests and experiments

6.1.3.4 Oligos for sequencing

Name	Sequence	Purpose of use
FlpTK-eff-LCb	CAAACCCTAACCACCGCTTA	Verification of the PRE sequences
pGl4.26mcsrev	AACAGTACCGGATTGCCAAG	
Luc2-seq-rev	AAAGCCGCAGATCAAGTAGC	Verification of the luc2 gene sequence
Luc2-seq-fwd	GCTACTTGATCTGCGGCTTT	
ROSA26mcs-fwd	GAGTTCTCTGCTGCCTCCTG	Check for insertion of the pgl4.26 PRE-TK construct into the Rosa-26 targeting vector
ROSA26mcs-rev	TCTTCTGGGCAGGCTTAAAG	

Table 7| primers used for the sequencing of important sections on the different vector constructs

6.1.4 Enzymes

6.1.4.1 Restriction enzymes used for the cloning procedure and restriction analysis

Name PRE	cloning site 5'	cloning site 3'	shuttle site 5'	shuttle site 3'	linearization	Restriction analysis
HoxD9-10 chr2	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
Spata3 chr1	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
GeneDesert chr11	SacI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
Ptch1 chr13	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
GeneDesert chr1	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
Rax chr18	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
Gad2/Myo3A chr2	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
Tmem chr12	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
Chit1	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
Myh14 chr7	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
Sox2 upstream	SacI	XhoI	AscI	AsiSI	PacI	HindIII, AsiSI
Oct4 up	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI
Oct4 geneend	XhoI	HindIII	AscI	AsiSI	PacI	HindIII, AsiSI

Table 8| Restriction sites which were added via PCR to the 5' and 3' ends of the predicted PREs. The corresponding restriction enzymes were used for the cloning of the PREs into pGL 4.26 TKhygrofloxed. Additionally the enzymes, which were used for the shuttling into the targeting vector, are shown, as well as the enzymes, which were used for the linearization of the Rosa-26 targeting vector constructs. All enzymes were purchased from thermo scientific.

6.1.4.2 Additional enzymes

Enzyme	Purpose of use	Company
Quick ligase	Ligation of vector fragments	NEB
Kapa2G Robust PCR Kit	Diagnostic PCR	Peqlab
KapaHiFi PCR Kit	Amplification of the PREs from the mouse genome	Peqlab
Kappa fast ready mix	Colony PCRs	Peqlab
Proteinase K	Protein removal during DNA isolation	Sigma Aldrich

Table 9| other Enzymes used in the various experiments

6.1.5 Kits and other materials

Kit	Company
QIAquick gel extraction /PCR purification kit	Quiagen
QIAprep spin miniprep kit	Quiagen
Pure yield plasmid system	Promega
High Pure RNA isolation Kit	Roche
SYBR. -Green Jump StartTM Taq Ready MixTM for Quantitative PCR	Sigma
TURBO DNA-freeTM	Ambion
Immobilon-Ny+ Membrane, charged Nylon, 0.45 µm, 30 cm x 3.3 m roll	Millipore
Prime-It RmT Random Primer La; Kit;	Stratagene
Whatman Chromatography paper 46*57 cm	Schleicher and Schuell
Illustra™ ProbeQuant G-50 Micro Columns Radiolabeled Probe purification Kit	GE Healthcare
Dual-luciferase ® reporter assay system	Promega
PureYield™ Plasmid Midiprep System	Promega
Luciferase Assay System	Promega

Table 10| Kits and other material used in the different experiments

6.1.6 Competent E-coli strain and WT ESCs

Cell type	Genotype	Company/reference
DH5α	F⁻_80*lacZ*ΔM15 Δ*pho*A gyrA96	IMBA service department
46C mouse ESCs	carries a GFP transgene inserted in the Sox1 locus; GFP is specifically upregulated in neural progenitors; FACS isolation of pure populations after *in vitro* differentiation is possible	[80, 102]

Table 11| Competent E-coli strain used for the different cloning steps;

6.1.7 Plasmids

Plasmid	Feature	reference
pGL4.26TK hygrofloxed	loxP sites, TK promoter, luciferase gene, ampicillin, hygromycin resistance cassette	Promega basic pGL4.26, modified by Robert Heinen, TK promoter added by myself
Rosa26	Homologouss arms for Rosa-26 in the mouse genom, AmpR	Soriano P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet 21: 70-1
pEGFP-N3	EGFP gene, Kanamycin resistance gene	BD Biosciences Clontech
CRE	CRE recombinase, KanR	Vector Biolabs
pRL-TK Vector	Ampicillin resistance gene, renilla luciferase gene	Promega

Table 12| Plasmids used in this project

6.1.8 Chemicals, radioactivity

Name	Company
Agarose	Sigma
SYBR® Safe	Invitrogen
Radioactivity	dCTP, [α-32P]- 3000Ci/mmol 10mCi/ml EasyTide Lead
Perfecthyb™ Plus	Sigma-Aldrich
Lipofectamine® 2000	Life technologies
DMSO	Sigma-Aldrich
6x loading dye	Thermo scientific
Hygromycin B	Invitrogen

6.1.9 . Software

Software	Company
GCK (Gene construction Kit)	Textco version 3.5.3.
Primer3	http://frodo.wi.mit.edu/primer3/ version 0.4.0.
CLC Combined Workbench	CLCbio, Version 6.6.2.
WEbcutter	Version 2.0.
PrimerBlast	NCBI

Table 13| Software used for the analysis and planning of the different experiments

6.2 Methods

6.2.1 Cloning of PCR products

In order to clone the 13 predicted PREs into the pGL4.26TK hygrofloxed vector, they were first amplified via PCR from the genome of SV40 cell culture cells. For this purpose primers were designed that bordered the predicted sequences and contained restriction sites in order to clone the PREs into the target vector (Table 4 and 7.1). The optimal annealing temperatures for the primer pairs were determined individually via gradient PCRs. In general, cells positive for containing the aimed plasmid construct were first detected by performing colony PCRs, followed by the sequencing of the crucial sequence sections by the IMBA sequence department. Furthermore restriction analysis was performed as an additional validation step.

6.2.1.1 Gradient PCR

Gradient PCRs were performed to determine the optimal annealing temperature for the primer pairs, which were designed to amplify the predicted PREs from the mouse genome. For this purpose the gradient program of the Mastercycler gradient S from Eppendorf was used. With this type of program it is possible to test up to 12 different temperatures in a range of 20 °C within one PCR experiment. I tested a gradient of 8 different temperatures from 50 °C up to 65 °C. A test to determine the optimal annealing temperature is often indispensable for obtaining specific PCR products when amplifying form a very complex genome. Following 1x reaction mixture was used for all gradient PCRs and was scaled up to 10x in order to test 8 different annealing temperatures per primer pair.

Component	µl in 25 µl
PCR grade water	Up to 25 µl
KAPPA2G Fast ReadyMix with dye	12.5 µl
Forward primer (50 µM)	0.2 µl
Reverse primer (50um)	0.2 µl
Mouse genomic DNA	200 ng

Table 14| shows the 1x reaction mixture for performing gradient PCRs

And following cycling parameters were used

Initial denaturation	95 **°C**	2 min	
Denaturation	95 **°C**	15 sec	1x
Primer annealing	Gradient program: 55 **°C -65°C**	15 sec	35x
Extension	72 **°C**	15 sec/kb depended on the calcµlated PRE length	
Final extension	72 **°C**	5 min	1x
Cooling	4 **°C**	Unlimited	1x

Table 15| shows the cycling parameters specific for gradient and colony PCRs in combination with the KAPPA2G Fast ready mix

On the basis of the gradient PCR results the annealing temperatures were selected, which showed the thickest and most specific bands after separation of the PCR product on a 1% Agarose gel (comments on 1% Agarose gels →6.2.1.3)

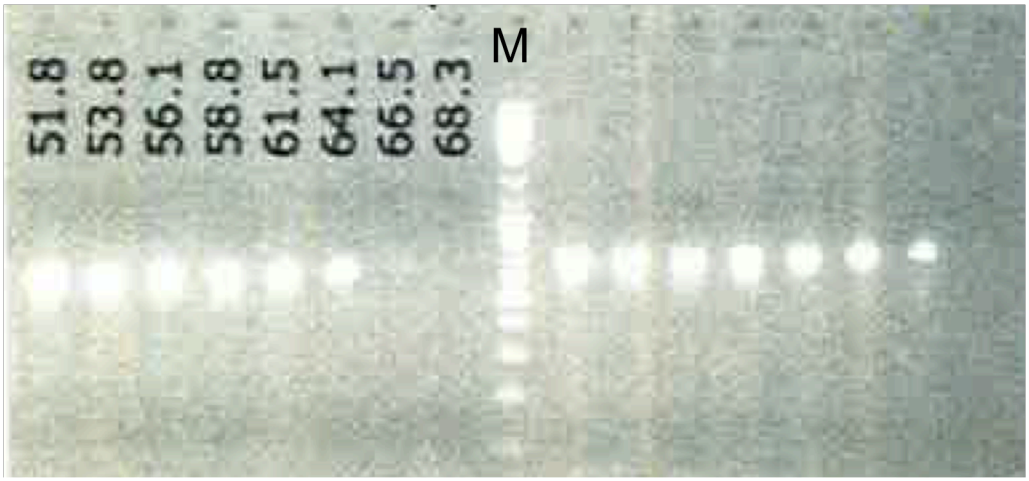


Figure 6| example for the outcome of a gradient PCR for two different primer pairs (to the left and the right of the marker M). A temperature gradient of 8 different annealing temperatures was applied during the amplification with the tested primer pairs

6.2.1.2. Amplification from the mouse genome using a proofreading polymerase

In order to avoid mutations in the sequences of the predicted PREs, a proofreading polymerase was used for the amplification from the mouse genome. All PREs were amplified from the same 46C WT genomic DNA sample (received from Robert Heinen) using the same PCR conditions regarding the reaction mixture and the cycling parameters (with exception of the annealing temperature and the elongation time). The 46c ESC line is adapted to grow under feeder free conditions on gelatin-coated tissue culture dishes maintain pluripotency and grow preferentially in monolayers. 46c cells have the additional advantage that they carry a GFP transgene inserted in the Sox1 locus, thus GFP is specifically upregulated in neural progenitors, enabling FACS isolation of pure populations after *in vitro* differentiation [80, 102] This feature will be used at later stages of the project.

Following reaction mixture was used for the amplification of the predicted PREs:

Component	µl in 50 µl
PCR grade water	Up to 50 µl
5x KAPAHIFI Fidelity Buffer	10 µl
dNTP mix (10mM)	1.5 µl
Forward primer (50 µM)	0.4 µl
Reverse primer (50um)	0.4 µl
Mouse genomic DNA	400 ng
KAPAHIFI DNA Polymerase (1U/µl)	1.0 µl

Following Cycling parameters were applied.

Initial denaturation	95 **°C**	2 min	
Denaturation	98 **°C**	3 min	1x
Primer annealing	Optimal temperature was determined individually for every primer pair via a gradient PCR	15 sec	30x
Extension	72 **°C**	30 sec/kb depended on the calculated PRE length	
Final extension	72 **°C**	5 min	1x
Cooling	4 **°C**	Unlimited	1x

6.2.1.3 Gel electrophoresis

In order to analyze PCR products, agarose gel electrophoresis was performed.
For this purpose a 1% agarose gel in 0.5 x TAE buffer was prepared and 10 µl SYBR® Safe per 100 µl agarose gel was added in order to visualize DNA.
6x-loading dye (thermo scientific) was added to the DNA sample. Thermo Scientific Gene Ruler DNA Ladder Mix, ready-to-use, was used as a size marker in all experiments in order to evaluate the product sizes.

Gel electrophoresis was performed in general for 30 -40 min at 100V.

6.2.1.4 PCR product and backbone vector were cut with the appropriate restriction enzymes

20/50 µl PCR product were separated via gel electrophoresis.
The non-UV transilluminator (Invitrogen) was used in order to visualize the DNA and to avoid UV damage. In the case that only a single band was visible on the gel, which was equivalent to a highly specific amplification of the corresponding PRE, cloning was performed directly with the remaining 30-µl PCR product. The PCR product was purified with PCR purification kit (Quiagen). In case that several bands where visible on the gel, bands at the expected sizes were cut out and transferred into an Eppendorf tube. Furthermore the remaining 30 µl was additionally separated via gel electrophoresis and the cut out gel slices were added together. In both cases the PRE and the pGL4.26 TK hygrofloxed vector were subsequently cut with the appropriate restriction enzyme. In order to determine the plasmid and PCR product concentrations, DNA concentration was measured with the UV/VIS Spectrophotometer Nanodrop® ND-1000 (PEQLAB) at 260nm.
2 ug pGL4.26 TK hygrofloxed backbone vector and 30µl (100-500 ng) PRE were digested for 2 hours at 37 °C with the appropriate restriction enzymes (6.1.4.1)

Water, nuclease-free	**Ad 20 µl**	Water, nuclease-free	**Ad 40 µl**
10X Tango	2 µl	10X Tango	4 µl
Backbone vector DNA	2 ug	PCR product	30 µl
Restriction enzyme 5’	1 µl (10 U)	Restriction enzyme 5’	2 µl (20 U)
Restriction enzyme 3’	1 µl (10 U)	Restriction enzyme 3’	2 µl (20U)

The complete reaction mixtures from the digestion of the backbone vector and the PCR product were separated via a 1% agarose gel.
The backbone vector was cut twice in the multiple cloning site and formed two fragments on the gel. The appropriate band was cut out and eluted. The PCR product was also purified with the Gel extraction kit (Quiagen) to remove short DNA fragments, which could compete with the insert.

6.2.1.5 Ligation

In order to ligate vector and insert DNA the Quick ligase Kit from NEB was used. For this purpose 50 ng vector DNA and a 3 fold molar excess of insert DNA were added to 10 µl Quick ligase buffer and 1 µl Quick ligase enzyme.
The reaction was incubated for 5 min. at RT.

6.2.1.6 Transformation

200 µl of CaCl competent E.coli bacteria (DH5alpha- IMBA service department) were added to 2 µl ligation mixture. Immediately the cells were incubated on ice for 20 min, heat shocked for 1.5 min at 42 °C and afterwards again put on ice for 2 min. 800 µl SOC media were added and the cell suspension was incubated for 1 h shaking (1200 rpm) at 37 °C. 200 µl and 800 µl cell suspension were plated onto LB^{ampR} plates according to the antibiotic resistance on the pGL4.26 TK hygrofloxed basic vector. Plates were incubated over night at 37 °C.

6.2.1.7 Identification of positive clones via colony PCR

In order to identify bacterial colonies carrying the intended vector product, colony PCRs were performed.
For this purpose a bacterial colony was picked with a yellow tip, streaked on a LB-ampR plate and dipped subsequently into a PCR tube containing the PCR reaction mixture. The LB ampR plate functioned as a backup plate. The bacteria streaks that were left behind on the plate were labeled in order to match positive colonies, identified via the colony PCR, to the streak on the plate. About 20-30 colonies were picked per transformation. The KAPPA2G Fast ReadyMix with dye was used for this purpose. The same cycling parameters and same composition of the reaction mixture were used as in 6.2.1.1

6.2.1.8 Sequencing of the amplified PREs

The IMBA service department sequenced all 13 cloned PREs. For this purpose primers were designed that bordered the region in the pGl4.26 TK hygrofloxed vector in which the PREs were inserted (Table 7). The Sequencing data was analyzed and compared to the original database sequences using the software CLCbio, Version 6.6.2. Thereby SNPs and INDELs were detected within the sequences of the amplified PREs. They were identified and recorded with following online database http://phenome.jax.org (see also 7.1.3)

6.2.2 Shuttling of the luciferase reporter construct into the Rosa-26 targeting vector

In order to target Rosa-26 in mouse ES cells the luciferase expression construct had to be cloned into a vector containing the homologouss arms for Rosa-26 (see also 7.1.1 and 7.1.2.). To ensure that, 2 ug Rosa targeting vector and 2 ug of the different pGL4.26 PRE-TK hygrofloxed vectors were cut with Asisi and AscI and incubated at 37 °C. for 1 hour.

Water, nuclease-free	**Ad 20 µl**
10X recommended buffer for restriction enzym	2 µl
Rosa targeting vector or pGL4.26 PRE-TK hygrofloxed	2 ug
Restriction enzyme Asisi'	1 µl (10 U)
Restriction enzyme AscI	1 µl (10 U)

Table 16| restriction digestion reaction for shuttling the luciferase expression constructs into the Rosa-26 vector

The complete volume was separated via gel electrophoresis and the appropriate bands consisting of the PRE-TK luciferase expression construct and the Rosa-26 backbone vector were cut out and eluted from the gel.
The ligation and the transformation were performed as in 6.2.1.5 and 6.2.1.6. .

6.2.3 Preparations of plasmid DNA in order to target Rosa-26 in mouse ESCs

In order to guarantee optimal conditions for a successfµl homologues recombination into Rosa-26 in mouse ES cells, the targeting vectors had to be linearized and purified from enzyme and reagents. Thereby it had to be considered that 20 ug vector construct were intended to be used for the electroporation. Since there is always a loss of in put DNA after phenol-chloroform purification, double the amount of finally needed vector DNA was linearized.

6.2.3.1 Linearization of the Rosa-26 targeting constructs

The Rosa-26 backbone contained a unique restriction site for the linearization of the Rosa-26 targeting vectors.
40 ug plasmid DNA were cut with PacI and the digestion reaction was incubated o/n.

Water, nuclease-free	**Ad 100 µl**
10X PacI Buffer	10 µl
Rosa-26 pGL4.26 PRE-TK hygrofloxed	40 ug
Restriction enzyme PacI	10 µl (100 U)

Table 17| reaction mixture for the linearization of 40 ug of vector DNA

6.2.3.2 Phenol- Chloroform isoamylalcohol (PCIA) purification and ethanol precipitation of vector DNA

The linearized vector DNA was purified via PCIA purification. For this purpose, the digestion reaction was increased to a volume of 500 µl with water.
1 vol. (500 µl) P:C:IA (25:24:1) was added and the reagents were mixed by vortexing. After centrifugation for 5min at 30 000 rpm two phases, an aqueous and an organic phase occurred. Proteins stayed in the lower organic phase whereas the aqueous polar phase contained the negatively charged DNA, which was transferred into a new Eppendorf tube. This purification step was repeated a second time.

For the precipitation of the purified DNA, 1/10 vol. NaAc pH 5.2, and 2.5 vol. 100% ice-cold ethanol were added, vortexed and incubated for 15 min at -80 °C.
After centrifugation for 20 min at 13000 rpm a DNA pellet became visible. The supernatant was discarded and the pellet was washed in 500 µl 70% ice-cold ethanol. The solution was centrifuged again, the supernatant was discarded and the DNA pellet was dried from excess alcohol (open tube at 37°C) The DNA concentration was measured with the UV/VIS Spectrophotometer Nanodrop®. Furthermore 120 ng linearized vector DNA were put on a 1% agarose gel (0.5 xTAE) together with undigested vector DNA as control. This was done on the one hand to determine if the total amount of vector DNA was linearized completely and on the other hand to verify the DNA concentration against the concentration of the marker bands.

6.2.4 General ESC techniques

In general the amounts of the different reagents and media used for the cultivation of the mouse ESCs were adapted to the various cell culture plates and dishes that were used for this study (Table 18).

plate/dish	media	trypsin	gelatin	cm2/well	company
96 well	200 µl	25 µl	50 µl	0.33	nunc
48 well	0.5-1 ml	50-100 µl	150 µl	1	greiner
24 well	1 ml	200 µl	250 µl	1.9	nunc
12 well	1.5-2.2 ml	300 µl	500 µl	3.8	falcon
6 well	3 ml	500 µl	0.5 - 1 ml	9.6	nunc
10 cm	10 ml	1 ml	4 ml	56.7	nunc

Table 18| media distribution for the different cell cµlture plates and dishes

6.2.4.1. Gelatinization of cell culture dishes/plates

The embryonic stem cell line 46C is able to grow independent from feeder cells on gelatinized cells and to maintain pluripotency (ReF 61 proposal)
Therefore cell culture plates and dishes were coated with 0.1 % gelatin for at least 20' until to 24 hours. The gelatin was removed completely and washed either with 1xPBS or dried for 2-5 min in the flow of the cell clture hood with open lid.
The gelatinized plates and dishes were always prepared freshly

6.2.4.2 Thawing cells

Cryotubes, containing 1.0x 10^6 cells were removed from the liquid nitrogen tank and quickly defrosted in the 37° C. water bath. Subsequently they were diluted in 9ml DMEM Stop media and centrifuged for 5 min at 1000 rpm to remove excess DMSO.
The supernatant was removed, pellets were resuspended in 5ml DMEM ++ media and distributed on gelatinized 10 cm dishes already containing 5 ml DMEM++ media.

6.2.4.3 Cell counting

Cells were counted using the Muse™ Cell Analyzer (Millipore).
For this purpose cell pellets from 10 cm dishes or 6 well plates were diluted in 5ml DMEM Stop or DMEM++ media. 450 µl of muse counting reagent were mixed with 50 µl cell suspension and incubated for 5 min.

6.2.4.4 Passaging/expanding ESCs

Cells attached to cell culture plates or dishes, were washed with 1x PBS and detached by the addition of the appropriate amount of 1x Trypsin following by the incubation for 5-10 min at 37° C.
Subsequently the reaction was stopped by the addition of DMEM Stop media. Cells were centrifuged 5 min at 1000 rpm to remove excess Trypsin and the pellet was resuspended in DMEM Stop or DEMEM++ media and either counted or directly distributed on plates/dishes coated with 0.1% gelatin.

6.2.4.5 Freezing ESCs in cryo vials and in whole cell culture plates

1x 10^6 cells were resuspended in 1 ml Freezing media and transferred into a cryo vial. The cryo vial was slowly cooled down using NalgeneTM freezing chambers, giving a

freezing rate of 1° C/ min. The freezing chamber was immediately put on -80 ° C until the next day and the cells were finally stored in a liquid nitrogen tank

In case of cells growing on 96 well plates, they were frozen directly in their wells. Therefore cells were washed with 150 µl PBS and detached with 20 µl 1x Trypsin. Subsequently 150 µl DMEM Stop were added and pipetting up and down singµlarized the cells.

To mimic the slow freezing rate of an isopropanol chamber, the plate was first put on ice for 5 min, afterwards wrapped in 2 paper towels and put for 2 hours on -20 ° C. They were finally stored at – 80 ° C for max. 2 month.

6.2.4.6. DNA isolation from ESCs

Cells at 100 % confluence growing on 6 well plates or 10 cm dishes were lysed with 500 µl lysis buffer containing freshly added ProteinaseK (f.c.1mg/ml). The cell lysat was incubated over night at 55 ° C - 65 ° C. In order to precipitate the purified DNA, 1/10 vol. 5M NaCL and 2.5 vol. 100% ice-cold ethanol were added, vortexed and incubated for 15 min at -80 °C

After centrifugation for 20 min at 13 000 rpm a DNA pellet became visible The supernatant was discarded and the pellet was washed in 500 µl 70% ice cold ethanol. The solution was centrifuged again, the supernatant was discarded and the DNA pellet was dried from excess alcohol (open tube at 37°C) The DNA concentration was measured with the UV/VIS Spectrophotometer Nanodrop®.

6.2.5 Workflow for targeting Rosa-26 in mouse ESCs in detail

6.2.5.1 Electroporation

1 x 10 ^{6} 46C WT cells were thawed on a 10 cm dish and expanded every second day until a cell number of over 10 x 10 ^{6} cells was reached.
10 x 10^{6} cells were harvested and the cell pellet was resuspended in 800 µl 1xPBS and transferred together with 20 µl purified linearized targeting vector into an electroporation cuvette (Gene Pµlser®/MicroPµlser™, 0.4 cm, Biorad). Cells were electroporated using the Bio-Rad Gene Pµlser Xcell in combination with following parameters: Voltage (V) 230, Capacitance (µF) 500, Resistance (Ω) ∞, Cuvette (mm) 4. 1 x 10^{6} cells were seeded per 10 cm dish. 4 dishes were seeded per cell line in order to assure that at least 200 clones will grow on the plates. They regenerated for 2 days until the selection started with hygromycin added to DMEM++ at a concentration of 200 ug/ml.

6.2.5.2 Hygromycin selection

After the electroporation cells that integrated the transgenic construct with the hygromycin resistance cassette into their genome grew for 8-10 days under hygromycin selection and formed cell colonies. The selection process started with a concentration of 200 ug/ml Hygromycin and was reduced stepwise down to ¼ of the starting concentration, as cells died and became less. In more detail, in the beginning of the selection progress, media containing the full hygromycin concentration was changed every day. After 3-4 days media was changed only every second day and the hygromycin concentration was reduced dependent on the amount of cells on the plate.

6.2.5.3 Isolation of clones

After 8- 10 days 192 cell foci were picked per PRE containing the transgene.
For this purpose cell culture media was removed and replaced by 1xPBS.

Clones were zoomed in 1.6x via a stereomicroscope and sucked up in 20 µl PBS with UV light sterilized yellow tips. The cell clones were transferred directly into 96 well PCR plates. Subsequently 20 µl trypsin were added to each well, followed by singularization by pipetting up and down. 20 µl cell suspension were transferred into a gelatinized 96 well cell culture plate, containing already 150 µl DMEM++ media.
The remaining 20 µl were used for the diagnostic PCR in order to identify clones with an integration event in Rosa-26

6.2.5.4 Diagnostic PCR

The 20 µl cell suspension, which remained in the PCR plate was diluted with 1xPBS and centrifuged for 5 min at 1050 RPM. The supernatant was discarded and the cells were treated with 10 µl lysis buffer (ProteinaseK concentration 1mg/ml). Incubation took place in a thermo cycler for 1 hour at 65° C, followed by inactivation at 95 ° C for 10'.
Following 100 x Master mix was prepared for the diagnostic PCR.

Component	**µl in 2000 µl**
PCR grade water	1900 µl
5x KAPPARobust GC Buffer	500 µl
dNTP mix (10mM)	50 µl
diagA1 (50 µM)	20 µl
clonB (50µM)	20 µl
KAPPARobust DNA Polymerase (1U/µl)	10 µl

20 µl PCR reaction mixture were added to 2 µl cell suspension. The plates were sealed with a removable adhesive foil and subsequent the diagnostic PCR was performed using following cycling parameters.

Initial denaturation	95 **°C**	2’	
Denaturation	95 **°C**	20”	1x
Primer annealing	58 **°C**	15 sec	
Extension	72 **°C**	2’	35x
Final extension	72 **°C**	3’	1x
Cooling	4 **°C**	Unlimited	1x

The PCR products were analyzed on a 1% agarose gel (7.3.1).

DNA from clones that produced a ~ 1500 bp fragment, where identified and the corresponding cell clones were further expanded as follows: The 96 well plate, containing the viable cells including the PCR positive clones, was spitted on two freshly gelatinized 96 well plates. One of the two plates served as DNA backup plate, the DNA was isolated and the whole plate was stored at -20. The positive clones from the other 96 well plate were expanded until there were enough cells grown to cµltivate them on 6 well plates.

From the 6 well plate clones were split once more into two 6 well plates. After the cells in the one set of 6 well plates reached a confluence of about 80 % they were frozen in cryo vials (6.2.4.5) as they were intended to be the future cell lines. The corresponding cell clones on the second set of 6 well plates grew until they reached a confluence of 100%. Subsequently the DNA was isolated (6.2.4.6) in order to use it in southern blots for the revalidation of the integration into Rosa-26 and to exclude cell lines with possible double integrations.

6.2.6 Southern Blot analysis

Southern blot analysis is generally used to transfer size separated DNA from a gel to a filter membrane followed by hybridization with a labeled probe, in order to determine the presence of a certain DNA fragment. Different strategies of labeling and visualization of the probes exist. In this study labeling via the isotope ^{32}P was the

method of choice. The southern blot analysis was used to detect possible double integration events of the transgenic construct into the mouse genome, as it is impossible to detect such an event via the diagnostic PCR.

6.2.6.1. Digestion of genomic DNA

7 ug Genomic DNA were each cut in two separate digestion reactions with 2 different restriction enzymes. One reaction mixture was set up with the restriction enzyme HindIII, the other one with BamHI (for a further explanation of the Southern Blot strategy see also 7.3.2). The reaction mixtures were incubated at 37 ° C o/n.

DNA	7 ug	7 ug
Buffer	4 µl buffer red	4 µl buffer BamHI
Enzyme	2 µl	2 µl
Total volume	Ad 40 µl dH_2O	Ad 40 µl dH_2O

Table 19| 1x restriction digestion mixtures. The same DNA sample was cut in two different reactions with either HindIII or BamHI

6.2.6.2 Agarose gel construction, running, staining

A 1% agarose gel was prepared with 1/3 of the suggested amount of cyber green staining solution. 10 µl 6x loading dye were added to each sample.
Additionally 10 µl of a 1:100 dilution of the marker Gene Ruler DNA Ladder Mix, ready-to-use (Thermo scientific), a WT control DNA and a positive control was loaded. The WT DNA consisted of 46C genomic DNA and the positive control was the En- transgenic cell line (4.4), but with the minimal promoter instead of the TK promoter. The gel was run at 120 V for 4 hours. After the electrophoresis the gel was photographed, using the Gel DOC gel imaging system from Biorad, in order to evaluate/confirm the completeness of the restriction digestion,

6.2.6.3 Denaturation of the Agarose gel

The gel was first incubated in the depurination solution for 20 min on a slightly rocking platform, which resulted in single strand breaks and facilitated the transfer of the DNA onto the membrane. Afterwards the gel was put into the denaturation solution for 20

min, followed by incubation in the neutralization solution for 30 min. (for the composition of the solutions see 6.1.1)

6.2.6.4 Blotting: DNA transfer from the agarose gel to the membrane

The transfer membrane was first soaked in ddH20 and subsequently equilibrated for 10 min. in 20x SSC. The loading slots of the gel were removed as well as excess gel that didn't contain DNA samples. The blotting tower was constructed like described in Figure 7. During the over night incubation, the DNA was transferred via capillary forces from the agarose gel onto the nylon membrane. The next day the blotting tower was disassembled and the membrane was dried for 1h at 80 ° C. Subsequently the DNA was cross linked in the Stratagene 1800 crosslinker, by using the Auto cross linking program.

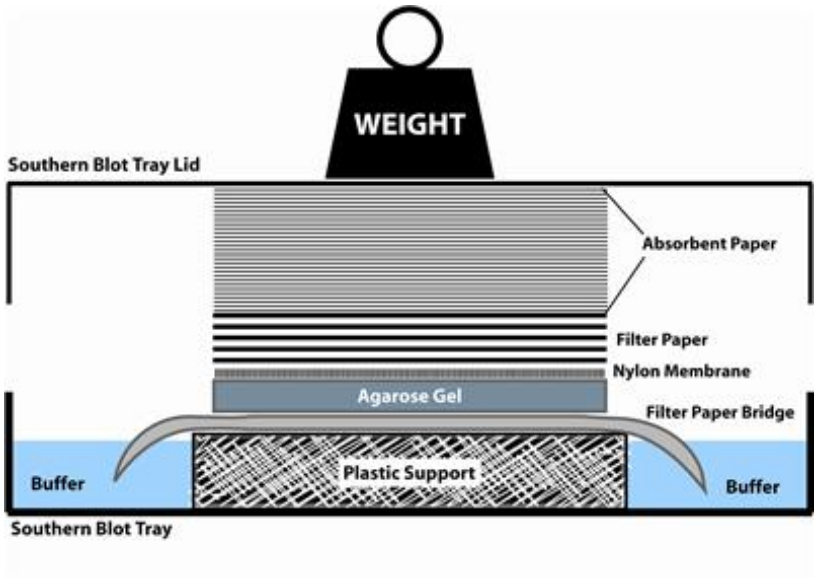


Figure 7| Construction of the "Blotting tower". The bottom layer consisted of 3MM whatman paper and reached into the plastic tray filled with 20X SSC. The gel containing the DNA was put directly onto the 3MM whatman paper layer and covered with the transfer membrane. 3 layers of whatman paper were put on top of the membrane followed by a stack of paper towels. Covering it with a lid and putting a weight on top of the lid stabilized the tower and kept the whole construction under pressure (http://www.gbiosciences.com/EducationalProducts/products.aspx?productid=0423564f-b320-4f8d-99b7-ed6f9fba5861&print=true)

6.2.6.5 Radioactive labeling and purification of the probe

For the detection of the transgenic construct in the genome of mouse ESCs an internal and an external probe were used (more detailed in 7.3.2). For the detection of the probes, they first had to be radioactive labeled.

For this purpose the Prime it Rmt random primer labeling kit (Stratagene) was applied following the according protocol:

50 ng of probe DNA were added to a single use reaction tube and filled up to 42 µl with ddH20. The reaction mixture was than transferred into a 1.5 µl vial with a screw cap.

Following steps took place in the isotope lab:

The probes were boiled for 10 min and were transferred immediately afterwards on ice. 5µl radioactivity were added to each probe followed by 3 µl magenta DNA polymerase (4 U/µl). The reaction mixture was incubated for 10-20 min on 37 °C.

The probes were purified from unincorporated labeled nucleotides using ProbeQuant G-50 Micro Columns (GE healthcare). The sephadex columns were prepared as described in the manual. The probes were directly added onto the column and were eluted by centrifugation. The probes were once more boiled for 10 min and stored on ice.

6.2.6.6 Hybridization

The nylon membranes were put into separate hybridizing tubes and were prehybridized rotating at 65 °C for 30 min. via the addition of prehybridization buffer Perfecthyb™ Plus (Sigma Aldrich).

Subsequently the labeled probes were added directly into the prehybridizing solution without getting in direct contact with the membranes. The membranes were incubated over night rotating at 65 °C.

6.2.6.7 Washing, Exposure and imaging

The prehybridization solution containing the radioactive labeled probes was discarded and the membranes were washed consecutively with the pre-warmed (65 °C) low stringency and the ultra high stringency buffer.
Afterwards the membranes were put on the exposure cassette and remained there for 6 – 24 hours. During the exposition the energy of the beta radiation deriving from the radioactive labeled probes was locally accumulated on the surface of the cassette. This energy was finally read via the Typhoon Phosphoimager (Amersham). The exposure cassette (molecular dynamics) could be reconstituted via light irradiation by a Screen-Eraser (Amersham)

6.2.7 Transient transfections

Lipofectamine 2000 (Life technologies) was used as transfection reagent since it is specifically designed for the transfection of plasmid DNA in combination with RNA. The amount and the composition of the reagents used for the different experiments depended on the scale of the experiment and also on the type of transfection (Plasmid DNA, Plasmid DNA + siRNA)
For this purpose ESCs were seeded 24 hours before transfection on 24-well cell culture plates, containing the standard ESC media, however without the addition of antibiotics. For the transfection Lipofectamine was primarily mixed with Opti-mem® (Life technologies) followed by incubation for 5 min at RT. Subsequently the suspension was added to the plasmid DNA or Plasmid\RNA mixture, which were also diluted with Opti-mem. The DNA/Lipid complex was incubated for 20 min. These complexes were than added to the ESCs, where they grew for additional 48 hours.
The following amounts were used per transfection for the different experiments

Experiment	Cell density ESCs (% plate surface) and media volume (µl) in 24 well plates	Lipofectamine 2000 + x µl Opti-mem	Transfected Vectors and/or siRNAs
Luciferase expression in transient transfected cells	50-70%	1.5 µl+ 48.5 µl	400 ng pEGFP expression vector+ 25ng renilla luciferase normalization vector+ 375ng test vector
Knockdown of Suz12/Ezh 2/Mll/Ash2l	40-50%	2 µl +48 µl	660 ng pEGFP 220 ng siPcG/siTrxG

Table 20| Amounts and volumes of vectors and/or siRNAs, which were applied in the transfections for the different experiments.

6.2.8 Measurement of the luciferase expression in transient transfected cells

46C WT cells were transfected as described in 6.2.7 with the Rosa-26 constructs, an EGFP expression vector and the renilla luciferase normalization vector. Cells were washed with 1x PBS and lysed with 150 µl 1x passive lysis buffer (Promega), for 30 min on a rocking platform at RT. Subsequently the 150µl lysates were transferred into 8-well PCR stripes. An 8-channel multipipette was used to transfer 10 µl/sample into a 96-well clear bottom plate. The reagents for the read out of the renilla and firefly luciferase were prepared as recommended in the manual (Dual Luciferase® Reporter Assay System Technical Manual, Promega)

Following settings were applied for the measurement:

Settings for measuring renilla and firefly luciferase

Temperature Set point: 22 °C
Plate Type: 96 well plate

Well mode:
Dispense 50 µl using dispenser 1
Shake linear for 0:02
Delay for 0:02.00
Read: (L) Lum

Dispense 50 µl using dispenser 2
Shake linear for 0:02
Delay for 0:02.00
Read: Renilla

End mode

6.2.9 Repression assays with stable ESCs

Cells were seeded at equal densities (1.0×10^5) in 24 well plates and were lysed after 56 hours. The procedure to measure the luciferase level was the same as in 6, with the exception that only the firefly luciferase levels were measured.

Settings for measuring only the firefly luciferase

Temperature set point: 22 °C
Plate Type: 96 well plate
Well mode:
Dispense 50 µl using dispenser 1
Shake linear for 0:02
Delay for 0:02.00
Read: (L) Lum
End mode

6.2.10 Derepression assays

6.2.10.1 Reporter assay

The members of the PRC2 complex, Suz12 and Ezh2 and the TrxG members MLL and Ash2L were targeted in two different experiments with siRNAs. SiRNAs against each of the targets were ordered as “Set of 4/upgrade: siGENOME (Dharmacon),

which contained 4 siRNAs against 4 different target sites on the same transcript.In order to use all 4 RNAs within one lipofection, the set of 4 was always mixed together after resuspension in ddH2O. Finally the siRNAs were stored as aliquots (200 ng/µl) at -80 °C.
Cells were seeded at equal densities in 24 well plates and were given 18 hours to settle down until they were transfected with plasmid DNA and siRNAs (Table 20). Single and double knockdowns were performed. For this purpose cells were either transfected only with the siRNA/Plasmid mixture against one target transcript, or with a transfection mixture, which contained the siRNAs against two targets in equal amounts. The cells were lysed 48 hours after transfection. The strategy for cell lysis and luciferase measurement was the same as in 6.

6.2.10.2 Expression analysis

In order to quantitatively evaluate Suz12/Ezh2/Mll/Ash2l transcript levels and to determine the knockdown efficiencies from the siRNA knock down, quantitative real-time PCRs were performed. For this purpose total RNA was isolated and converted into cDNA. All samples were kept on ice during the different preparation steps.

RNA isolation

Total RNA was isolated using the High Pure RNA Isolation Kit (Roche) according the manual. Following steps were performed in advance. In order to detach the cells from the bottom of the 24 well plates, cells were washed with 1x PBS and incubated in 100 µl 1xTrypsin at 37 °C for 5 min. The reaction was stopped with 500 µl DMEM Stop and centrifuged for 5 min at 1000 rpm. The cell pellet was resuspended in 200 µl 1xPBS in order to proceed with step 1 of the manual.
The eluted RNA was additionally treated with Ambion® TURBO™ DNase in order to remove residual DNA that could affect the qPCR results. 0.5 vol. TURBO DNase buffer was added in combination with 1 µl (2U) TURBO DNase. The reaction mixture was incubated for 30 min at 37 °C and stopped after addition of 0.5 Vol. of the deactivation buffer and incubation for 5 min. The Turbo DNase digested RNA was centrifuged in order to remove the inactivation buffer, which concentrated on the

bottom of the reaction tube. The RNA was finally transferred into a new 1.5 ml tube. The concentration was measured via UV/VIS Spectrophotometer Nanodrop® ND-1000 (PEQLAB) at 260nm and aliquots of 1.5 ug RNA were stored at -80°C.

cDNA preparation

The reverse Transcriptor kit (Roche) was used for the preparation of total cDNA. For this purpose 1.5 μg RNA were mixed with 10 μM polyA primer and 2 μl random decamers (Invitrogen). The reaction mixture was filled up to 19 μl and incubated at 65 °C to ensure denaturation of RNA secondary structures. Subsequently 6 μl Transciptor RT buffer and 3 μl of a10 mM dNTP mix were added. 9.5 μl of this mixture were removed in order to provide a no RT (reverse trancriptase) control for the following qPCR.
0.5 μl reverse RT (10U) and 0.5 μl RNAse inibitor (20 U) were added and the reaction mixture was incubated first for 10 min on RT and subsequently for 30 min at 55 °C. The reaction was finally inactivated at 85 °C for 5 min.

Quantitative real-time PCRs

Realtime qPCR was performed in 96 well PCR plates (Eppendorf) using 10 µl per well of the SYBR®Green JumpStart™ Taq REadyMix™ (Sigma).
The final concentration of the primers was 150 nM in a total volume of 20 µl .
3 dilutions in a series of 1:5 dilutions were used for the cDNA templates and the WT cDNA, which served as positive control. In order to raise the cT values of the reference template GAPDH, the corresponding cDNA was additionally diluted 1:100 previous to the 1:5 dilution series. AS negative controls “no RT” as well as the no template control were implied in the experiment. The Plate design and legend are illustrated in
Figure 8, cycling parameters in Table 21

Primers		cDNAs	
1	GAPDH	A	si control
2	EZH2	B	si EZH2
3	SUZ12	C	si SUZ12
		D	wt cDNA, positive control

		1	2	3	1	2	3	1	2	3	1	2	3
cell line #1	no RT 1:1												
	cDNA 1:1	1:100			1:100			1:100			1:100		
	cDNA 1:5		cDNA A			cDNA B			cDNA C			cDNA D	
	cDNA 1:25												
	no RT 1:1		no RT A			noRT B			noRT C			H2O	
cell line #2	cDNA 1:1	1:100			1:100			1:100			1:100		
	cDNA 1:5		cDNA A			cDNA B			cDNA C			H20	
	cDNA 1:25												

Figure 8| plate design qPCR

Table 21| cycling parameters, qPCR

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

7 Results

7.1 Cloning of 13 predicted elements into the targeting vector for Rosa-26 in mouse ESCs

7.1.1 Cloning strategy

In order to apply the luciferase reporter system to the 13 predicted PREs, various consecutive cloning steps were performed (Figure 9). In previous experiments the luciferase reporter system was used in combination with a minimal promoter. In order to obtain more robust levels of luciferase expression, as a first step the TK promoter was added upstream of the luciferase reporter. The pGL 4.26 vector carrying the luciferase expression construct furthermore contained an ampicillin resistance gene for selection in bacteria and a hygromycin cassette, for selection in eukaryotic cells, flanked by loxP sites. Starting from this basic vector the 13 predicted elements were amplified via PCR from the genome of SV40 cell culture cells and cloned upstream of the TK promoter and the luciferase gene into the pGL4.26 TK vector (Figure 10). Subsequently the different PRE reporter constructs were shuttled via two unique restriction sites into the Rosa26 targeting vector. This vector finally carried the homologous arms for site directed integration into the ROSA26 locus in mouse ES cells (Figure 11). For a more detailed description see materials and methods section.

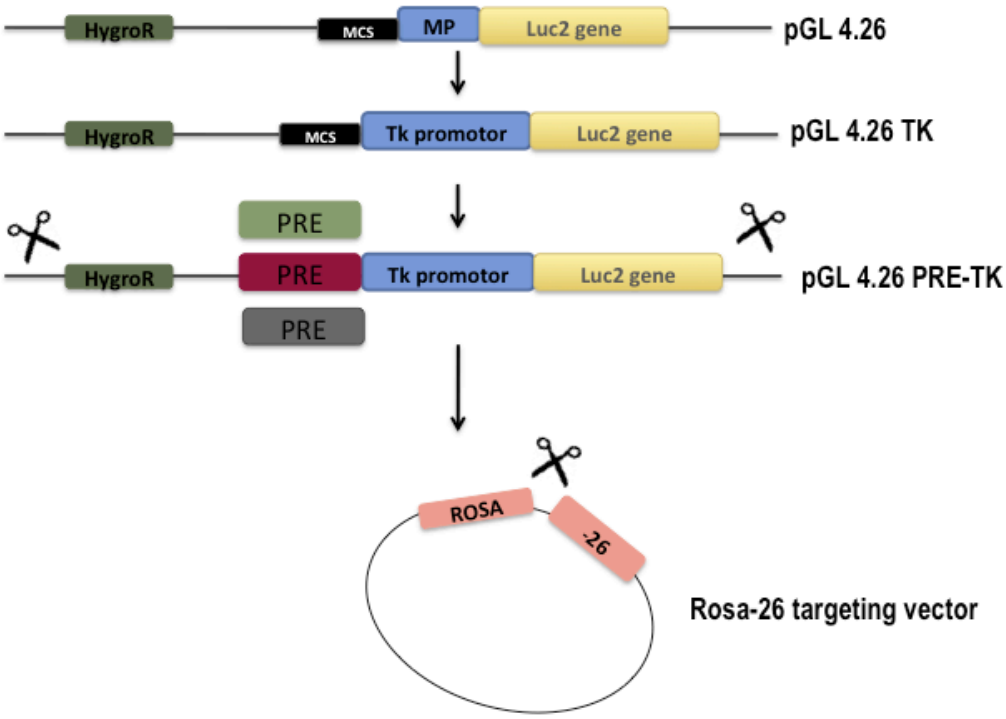


Figure 9| Simplified overview over the cloning concept

Name	Size (bp)	Category
HoxD9-10 chr2	1015	chip-only
Spata3 chr1	1640	chip-only
GeneDesert chr11	1097	chip-only
Ptch1 chr13	1886	prediction+ESC+ENCODE
GeneDesert chr1	894	chip-only
Rax chr18	965	prediction + ENCODE
Gad2/Myo3A chr2	950	prediction + ENCODE
Tmem chr12	860	prediction + ENCODE
Chit1	1050	prediction + ENCODE
Myh14 chr7	769	prediction + ENCODE
sox2 upstream	714	prediction + ENCODE
oct4 up	657	prediction + ENCODE
oct4 geneend	930	prediction + ENCODE

Table 22| Sizes and categories of the PREs used for cloning

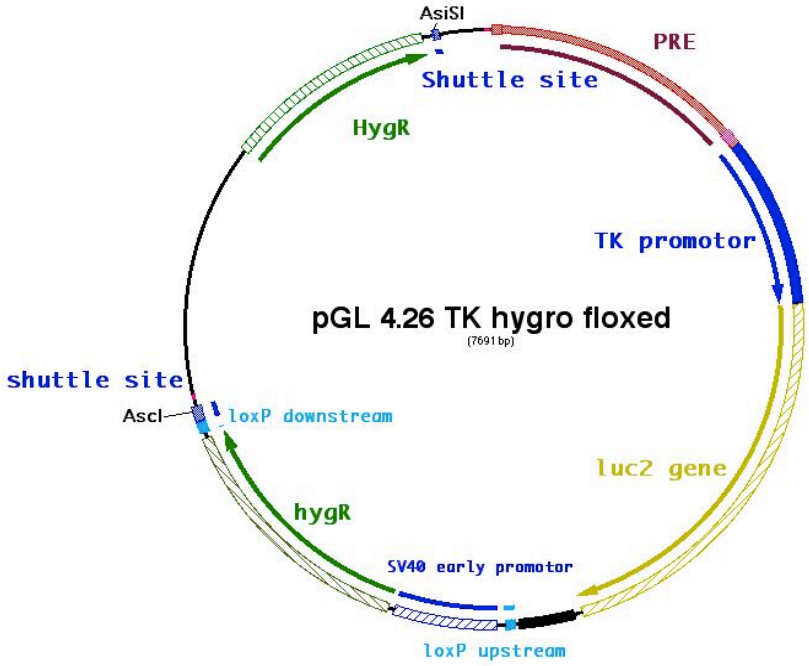


Figure 10| pGL 4.26 basic vector with inserted PRE upstream of the luciferase reporter construct. Shuttle sites for the transfer into the targeting vector for Rosa26 are indicated.

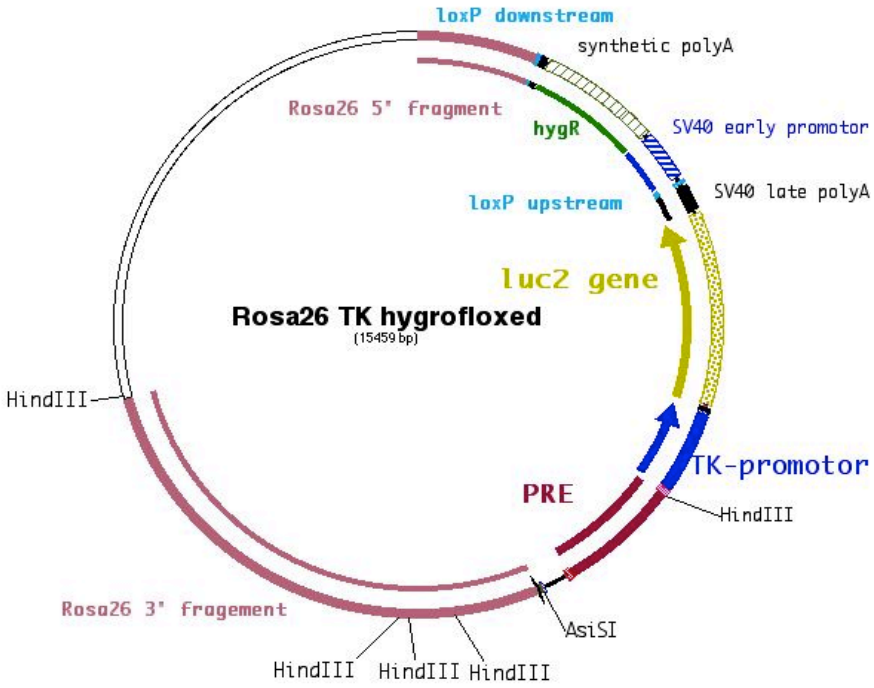


Figure 11| Targeting vector for the integration into Rosa26 .Restriction sites used for the restriction analysis are indicated.

7.1.2 . Verification of constructs via restriction analysis

In order to verify the 13 different Rosa-26 vector constructs a restriction analysis was performed with 2 enzymes that generated various fragment sizes throughout the Rosa26 vector. Furthermore the restriction enzymes were selected in a way that the different vector constructs coμld be distinguished via the different sizes of their inserted PRE. The original pGL 4.26-TK vector and the Rosa-26 targeting vector were additionally digested as controls (**Figure 12**)

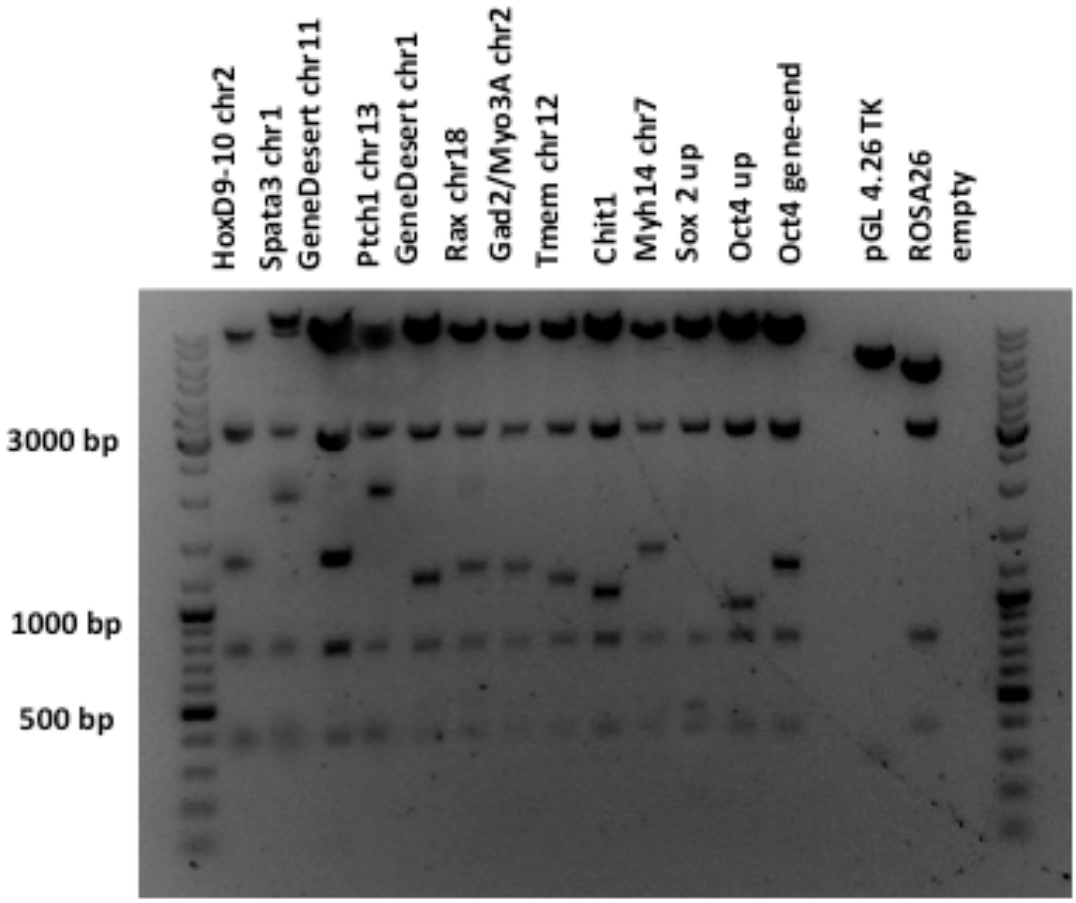


Table 23)

Figure 12| Restriction analysis of the 13 Rosa-26 targeting vectors. All vectors were double digested with the restriction enzymes HindIII and Asisi.

Rosa26 TK-PRE	Vector size	Expected fragment sizes cut with HindIII and AsisI						
HoxD9-10 chr2	15693	10090	3098	1290	743	381	79	
Spata3 chr1	16301	10084	3098	1916	743	381	79	
GeneDesert chr11	15736	10084	3098	1351	743	381	79	
Ptch1 chr13	16547	10084	3098	2162	732	381	79	
GeneDesert chr1	15555	10084	3098	1170	743	381	79	
Rax chr18	15626	10084	3098	1241	743	381	79	
Gad2/Myo3A chr2	15611	10084	3098	1226	743	381	79	
Tmem chr12	15520	10084	3098	1135	743	381	79	
Chit1	15711	10084	3098	1326	743	381	79	
Myh14 chr7	15430	10084	3098	1045	743	381	79	
sox2 upstream	15395	10084	3098	830	743	381	180	79
oct4 up	15318	10084	3098	933	743	381	79	
oct4 geneend	15591	10084	3098	1206	743	381	79	
pGL 4.26 TK	14687	14687						
Rosa26 empty	9971	5670	3098	381	79			

Table 23| Restriction analysis of the Rosa-26 targeting vectors; expected fragment sizes after digestion with HindIII and AsisI

7.1.3 The amplified PRE sequences contain SNPs and INDELs when compared to the database sequences

Since a restriction analysis cannot exclude possible mutations, which can always occur during the PCR amplification from genomic DNA, the PREs were additionally sequenced by the IMBA sequencing facility.

Thereby I detected several discrepancies in the cloned PRE sequences when aligned to the database sequences of the same PREs. These variations appeared to be SNPs and INDELS and occurred since genomes from different mouse strains were compared in the alignment. This resulted from the fact that although the genome from the BL6 mouse strain was used for the computational analysis of the PREs, the genome of SV126up mice was used for their amplification. The SNPs and INDELs were detected by use of http://phenome.jax.org.

All polymorphisms were recorded (Table 24) and the PhD student Johanna Trupke, who also predicted the PRE elements ascertained that the polymorphisms did not significally affect the prediction score.

PRE element	Genomic location	Reference	Observed alternative
HoxD9-10 chr2	74534386	CAGAGAG	CAG
Spata3 chr1	87915560	T	C
	87915560	A	G
	87915626	AACACACACAC	AACACACACACACACACAC
	87915428	T	C
	87915529	A	G
	87915560	A	G
Rax chr18	66102095	GAA	G
	66102323	C	T
	66102717	G	T
	66102702	C	T
Tmem chr12	74650447	C	T
	74650483	ACCCCCCCCCCC	ACCCCCCCCC
	74650541	A	T
	74650555	A	G
	74650560	T	C
	74650561	CCATGT	C
	74650650	C	G
	74650744	AT	A
	74650790	G	A
	74650881	T	C
	74650883	G	A
	74650909	G	T
Myh14 chr7	51924426	T	A
	51924721	GA	G
	51924736	T	C
	51924907	T	C
	51924934	C	T
	51925006	T	G
	51925010	TTCTTTCTTTCTTTCTTTC	T
oct4 up chr17	35639149	ACC	A

Table 24| List of SNPs and INDELs that were detected in the cloned PREs when compared to the database sequences

7.2 PREs in transient experiments show a silencing effect on the luciferase reporter in mouse embryonic stem cells

Two mammalian PREs have previously been identified, at the mouse MafB/Kreisler [83] and human HoxD loci [84], whereby most of the experiments to analyze the human HoxD PRE were performed with transient transfected cells. Therefore we decided to look first at the effect of the PREs on the reporter in transient experiments. The resµlts also influenced the later decision regarding which of the elements were chosen for the stable integration into Rosa26 in the mouse genome. 46C mouse embryonic wild type cells were lipofected with equimolar amounts of each Rosa26 targeting vector construct along with a renilla luciferase and an eGFP expression vector. The lipofection efficiency was determined via GFP fluorescence microscopy in order to estimate the overall occurrence of transfected cells (Figure 13). The firefly luciferase expression levels were measured and normalized to the renilla luciferase expression level, which served as an internal transfection control. Finally the normalized firefly luciferase values of the vector constructs with inserted PRE were compared to the TK control vector lacking a PRE. Since PRE lines of different categories were tested (4.3), different results were expected regarding the degree of silencing of the luciferase reporter.

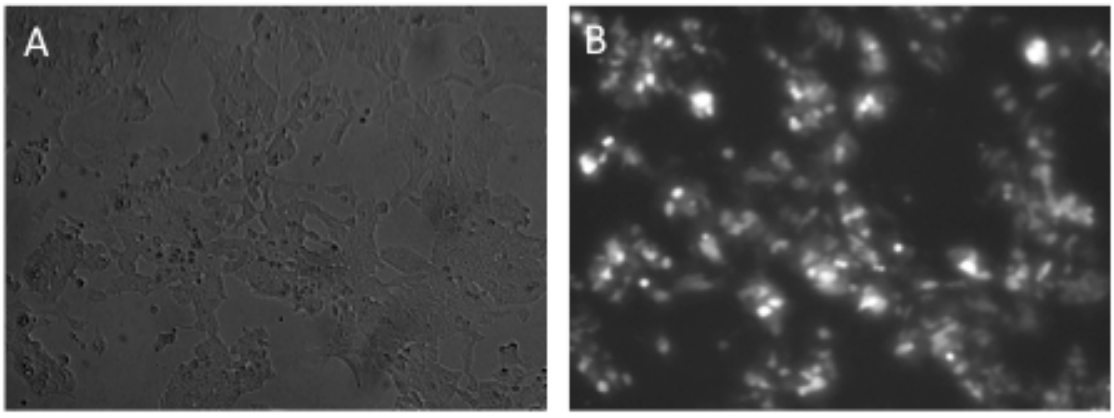


Figure 13| 46C cells 24 hours after lipofection, A) bright field. B) GFP fluorescence microscopy

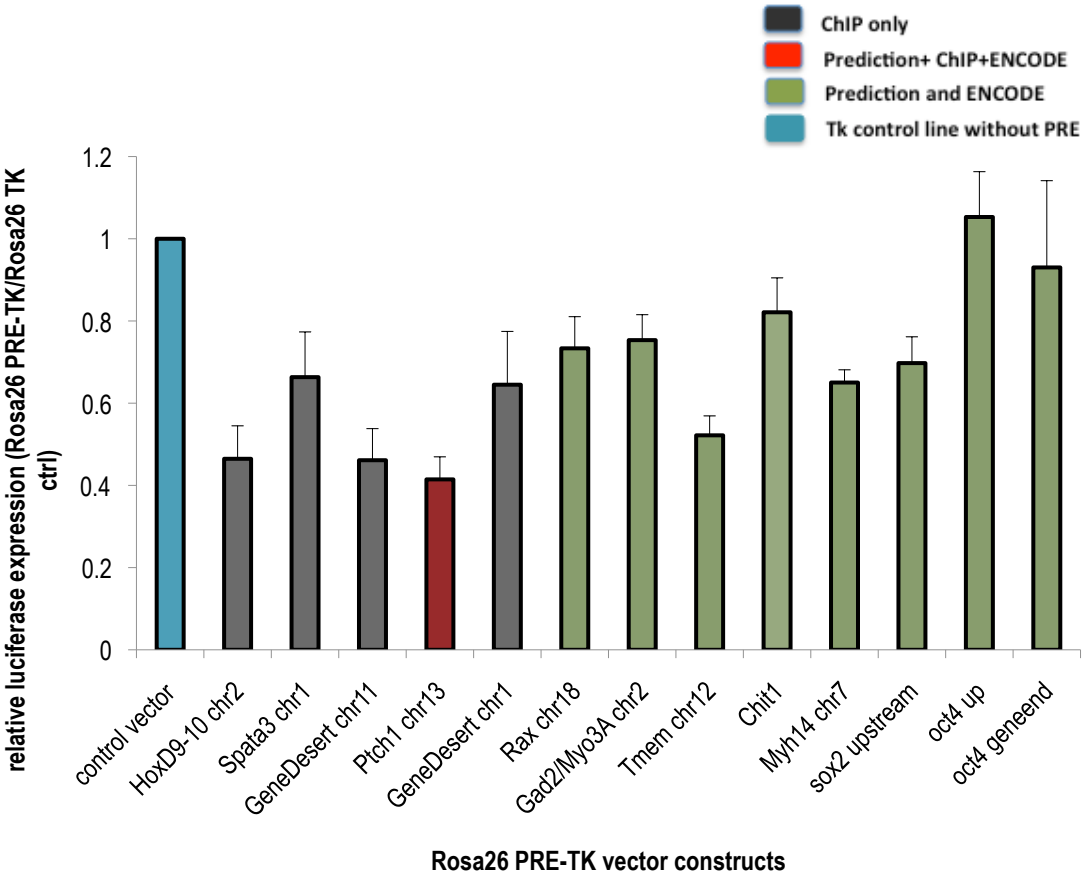


Figure 14| Relative luciferase expression in transient transfected mouse ES WT cells. Luciferase expression was measured 48 hours after transfection. ECS were lipofected with the Rosa26 PRE-TK vector constructs 2 times in duplicates. Error bars show SD.

Most of the elements showed repression of the luciferase reporter compared to the control construct lacking a PRE (Figure 14). Interestingly also constructs that were not expected to silence the reporter in ESCs led to a down regulation of the luciferase expression values (for a further explanation of the different PRE categories see 4.3). Only Oct4 up and oct4 gene-end showed no repressive effect, which is in agreement with the fact that the corresponding gene is highly expressed in embryonic stem cells.

However, the situation on a plasmid does not reflect the natural situation in the genome of embryonic stem cells. Indeed there is some debate regarding the events that occur when plasmid DNA enters a cell, concerning whether it is packed into nucleosoms or not [103-105]. Therefore the stable integration into the

genome of mouse ES cells was indispensable for measuring the effect of the PREs on the reporter. Since the creation of stable stem cells lines is very labor intensive, one PRE from each category (ChiP only/ Prediction, ChIP and ENCODE/ Prediction, ENCODE) and the TK control vector were selected for the site directed integration into mouse ESCs.

7.3 Targeting of PRE reporter constructs via homologous recombination into the ROSA-26 locus in mouse ES cells

In order to target the Rosa-26 locus in mouse ESCs a workflow, previously established by Robert Heinen, a former Postdoc in the Ringrose group, was applied to the selection of PRE luciferase reporter constructs (Table 25). ESCs were electroporated with the Rosa26 targeting vectors to enable the homologous recombination into the mouse genome. Since the expression constructs contained in addition a hygromycin resistance cassette it was possible to screen for transgenic cells via addition of hygromycin to the basic cell culture media. After about 8-10 days single transgenic cells formed foci that could be isolated. About 200 colonies were picked per cell line and grown in 96 well cell culture plates. Half of the cells were used for a diagnostic PCR, which I specifically established for this project, to screen for integration events into the Rosa-26 locus. Cell lines that were positive in the diagnostic PCR were further expanded, aliquots were frozen and the DNA was isolated. Finally double integration events were excluded via southern blotting.

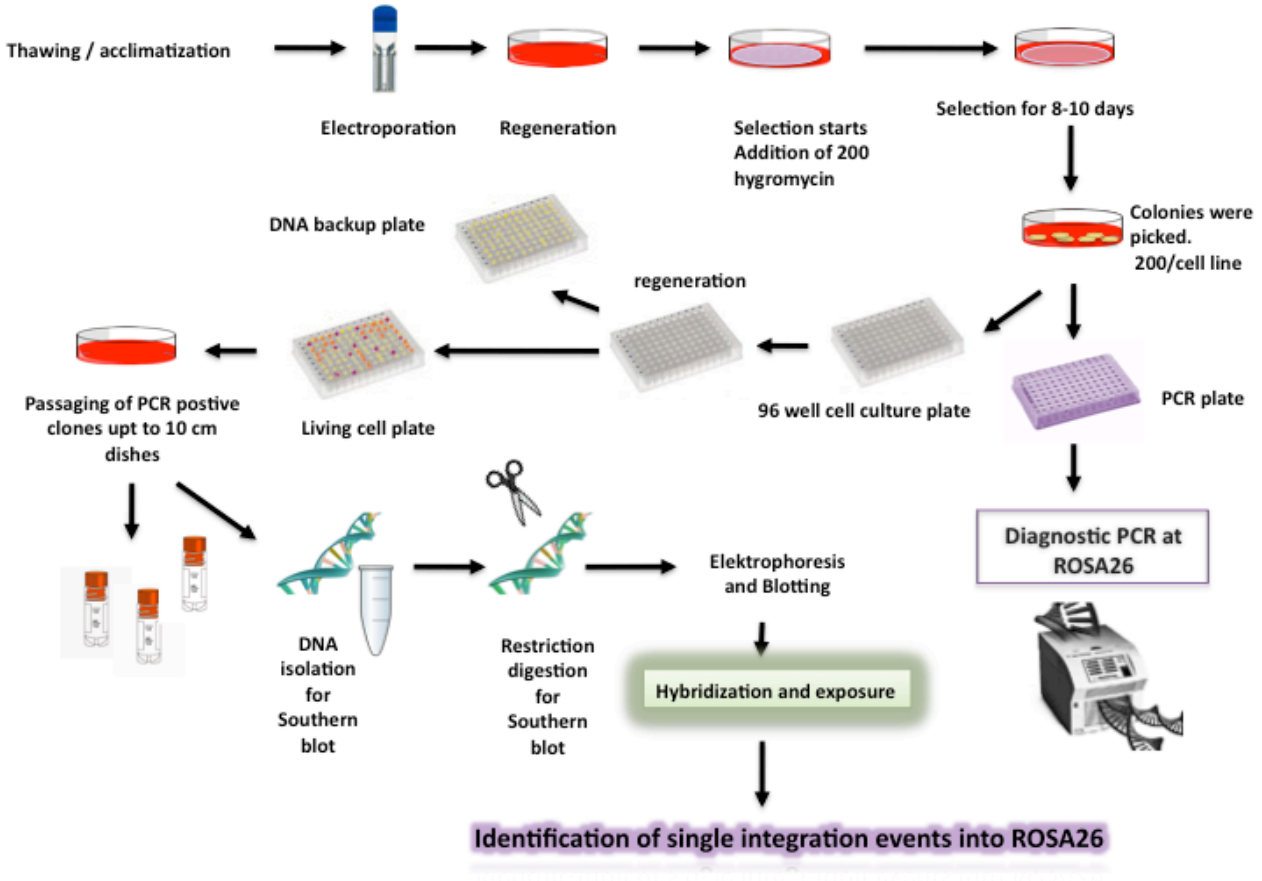


Figure 15| Simplified version of the workflow for targeting Rosa-26 in mouse ESCs

7.3.1 Establishment of specific PCR conditions enables rapid identification of integration events into the ROSA-26 locus

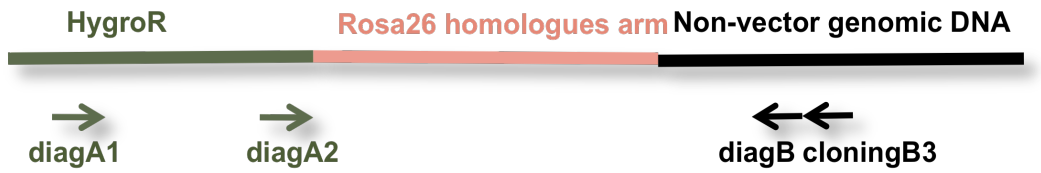
The Rosa 26 locus is reported to be free of PcG binding in mammalian ESCs and known to contain eukaryotic chromatin, and is therefore well-suited for PcG mediated repression studies (42).

However, the specific region in the Rosa26 locus we used for homologous integration contains a high GC content making it difficult to amplify via PCR.

Thus different parameters needed to be optimized in order to use diagnostic PCR for the screening of positive integration events into Rosa-26. The main aim was to make the screening procedure as efficient as possible and to keep at the same time the labor intensity as low as possible. The most effective time point to determine the integration localization is directly after isolation of the cell colonies. Colonies tested positively for the integration into Rosa-26 at this stage of the experiment, can be

directly isolated from the 96 well plate and expanded. In order to establish a diagnostic PCR protocol for Rosa-26, different primer pairs and polymerases from different suppliers were tested, first on purified mouse genomic DNA. The tested primers were designed in a way that the 5' primer always bound the hygromycin cassette and the 3' primer always bound genomic, non-plasmid DNA (Figure 17)

In this combination it is possible to distinguish between cells which harbor the luciferase reporter construct at ROSA26 and cells in which the integration event happened elsewhere in the mouse genome. Genomic DNA, already tested positive for integration into ROSA26, served as positive control and a plasmid containing the primer binding sites, as amplification control. The results showed that only one specific polymerase combined with the primer pair diagA1-clonB was able to produce a PCR product in the expected size.



Combination	Product size (bp)
diagA1-clonB	**1563**
diagA1-diagB	1653
diagA2-diagB3	1400

Figure 16| Primer design for the establishment of a diagnostic PCR at Rosa26.

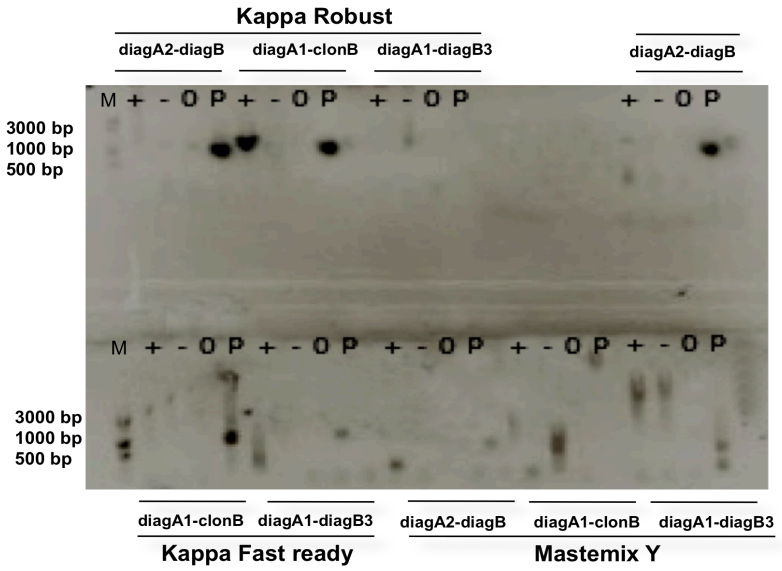


Figure 17| Primer and Polymerase testing. Purified DNA (+), Plasmid DNA (P) and wild type DNA (-) served as controls.

Although these results, showing that the PCR amplified the targeted region on purified DNA, were promising, it was not clear whether the PCR would be as robust under non-optimal conditions, such as those that would prevail during the integration workflow. I addressed this problem by testing different methods of obtaining DNA in different media background. For this purpose, cells were harvested for the diagnostic PCR test directly after the colonies were picked. Thereby they were either isolated in standard cell culture media or PBS. Subsequently they were lysed either by using lysis buffer or by boiling cells either in PBS or in standard PCR buffer. 0.5 µl to 2.0 µl cell lysates were finally used for the PCR test.

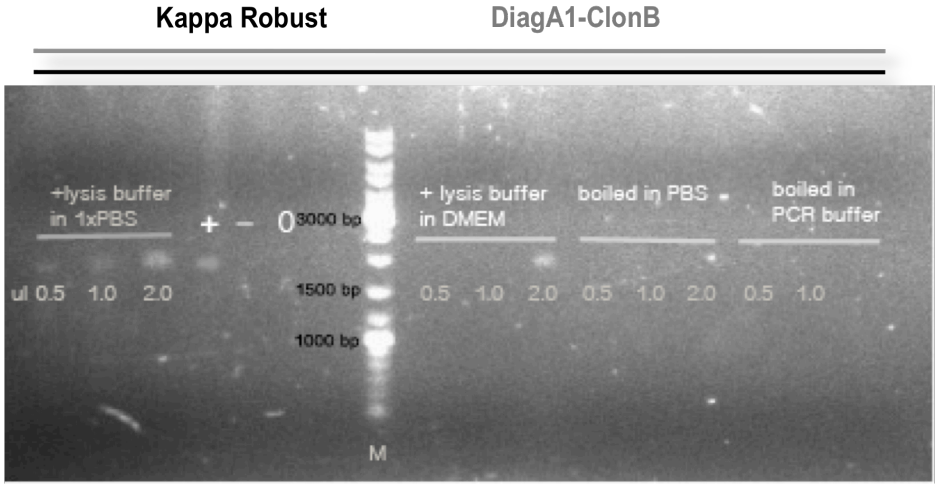


Figure 18| Optimization of the diagnostic PCR at Rosa-26. Cells were harvested directly after picking

Interestingly the result appeared to be very clear (Figure 18). Bands with the highest intensity where obtained when the cells where kept in PBS while they were lysed via lysis buffer and when 2 µl of this lysat where used for the diagnostic PCR. The PCR also showed a result, although less efficient, when the cells where kept in the standard stem cell media and lysed with lysis buffer. Boiling of the cells, without using lysis buffer led neither in PBS nor in PCR buffer to a detectable PCR product.

These results confirmed that the diagnostic PCR could be adjusted to the screening of integration events into ROSA-26 in mouse ESCs. 192 clones were picked per cell line and screened via diagnostic PCR. This analysis identified approximately 5-10 clones positive for integration into ROSA-26 (Figure 19).

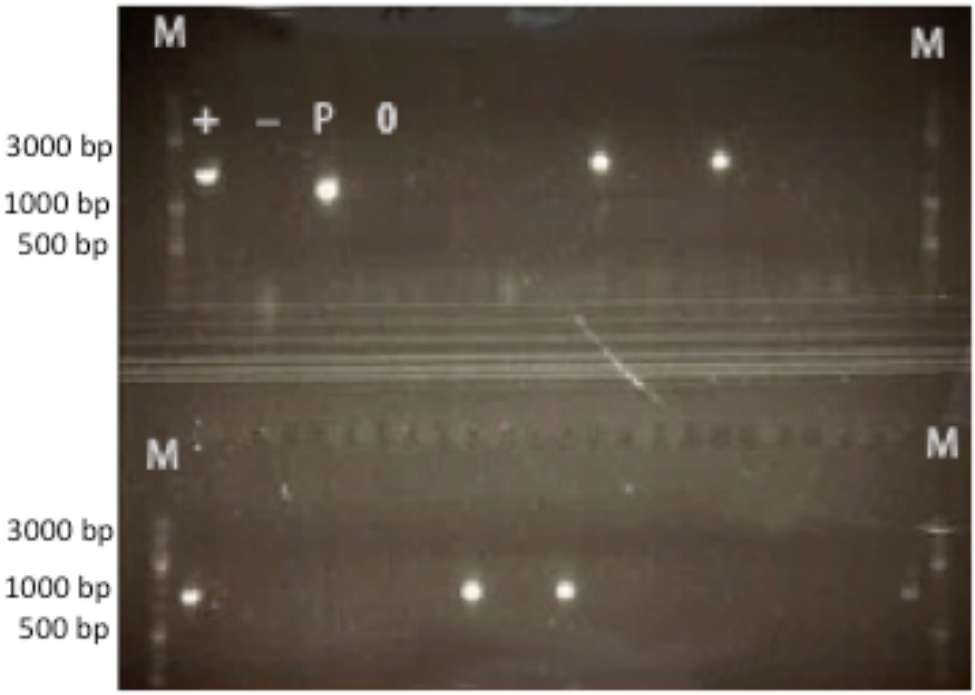


Figure 19| Screening of 50 clones via diagnostic PCR. Transgenic DNA (+), wild type 46C DNA (-), plasmid DNA (P), water sample (0)

7.3.2 Double integration events were excluded via Southern blotting

Although the diagnostic PCR could be used in the initial screening for integration events into Rosa-26, the PCR cannot detect possible double integrations into the mouse genome. Since the aim was a quantitative read out of the luciferase expression, it was absolutely necessary to detect such events and exclude these cell lines from the analysis. In order to exclude cell lines with such a double integration event southern blot analysis was performed (Figure 20). DNA from cells that were tested positive in the diagnostic PCR was isolated and digested either with HindIII or BamHI. The HindIII digested DNA was hybridized to an external radioactive probe that binds genomic non-transgenic DNA and served as a revalidation for the integration into Rosa-26. Wild type DNA and DNA from an already validated transgenic line served as controls. In case of the integration into the correct locus, I expected one band at 7900 bp, which represents the integration of the transgenic construct into Rosa-26. Furthermore I expected also a band at 4400 bp which represents the WT allele. In contrast possible double integration events could be detected via an internal probe binding to the luciferase reporter gene. Furthermore by using BamHI in combination with the internal probe, the different cell lines could be distinguished from each other, via the different sizes of the PREs (Figure 21). About 80 % of the PCR positive tested clones could be revalidated for integration into Rosa-26, whereas only 1 double integration event was detected via Southern blotting.

	E	WT	I
Spata3 chr1	**7903**	**4400**	**3769**
GeneDesert chr11	**7903**	**4400**	**6996**
Ptch1 chr13	**7903**	**4400**	**3335**
Rax chr18	**7903**	**4400**	**3107**
Chit1	**7903**	**4400**	**6955**
Oct4 up	**7903**	**4400**	**6578**
Oct4 gene end	**7903**	**4400**	**6851**
Control vector	**7903**	**4400**	**5942**

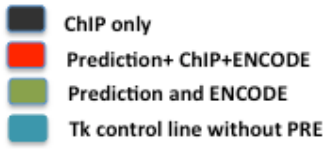


Table 25| Southern Blot; expected fragment sizes, for the external (E) and the internal (I) probe and the WT allel (WT)

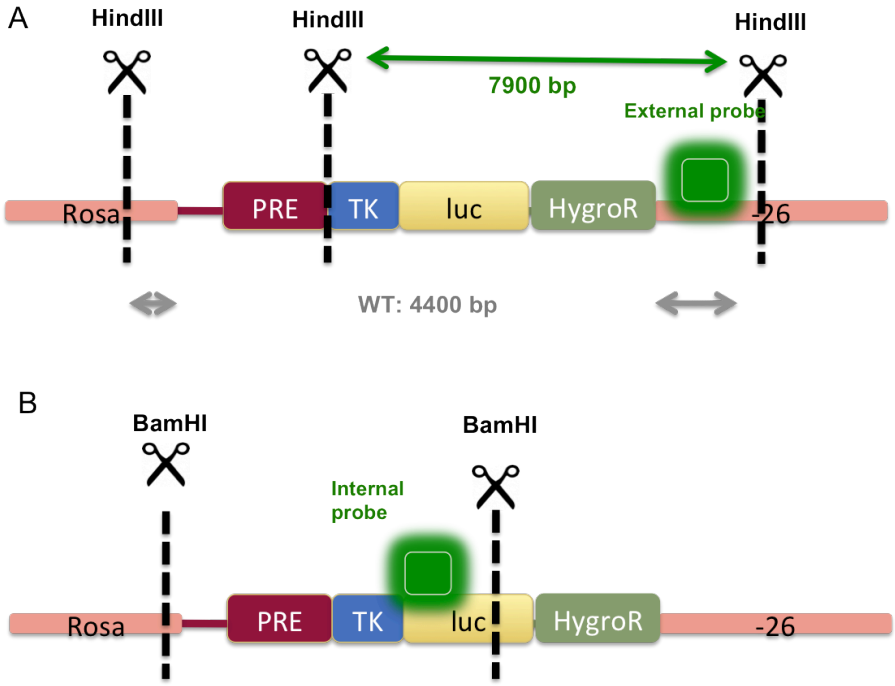


Figure 20| Simplified scheme of the southern blot analysis strategy (A) DNA was digested with HindIII and hybridized with an external probe. (B) DNA digested with BamHI and hybridized with an internal probe to detect possible double integration events

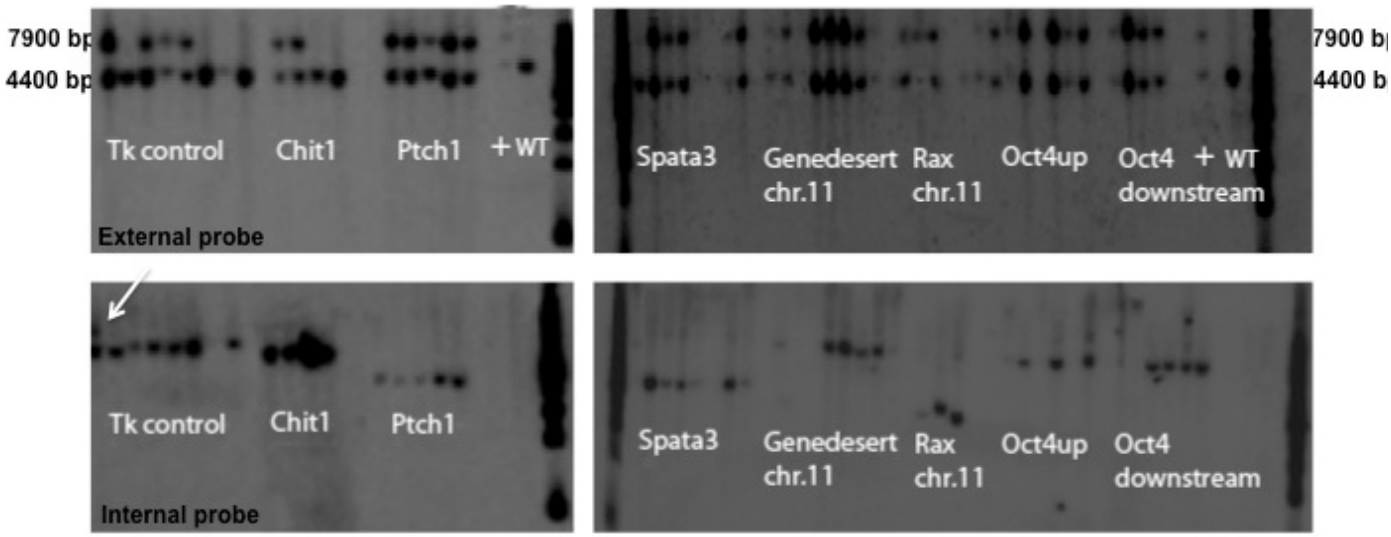


Figure 21| Southern blot analysis of the transgenic cell lines. Upper panel: revalidation for integration into Rosa-26. Lower panel: Check for double integration events. One double integration event could be detected (white arrow)

7.4 Ptch1 and engrailed show silencing of the reporter at ROSA-26 in mouse embryonic stem cells

In order to address the question of whether the putative PREs fulfill one of the hallmarks of a mammalian PRE, namely silencing of a reporter at an ectopic locus, Repression assays were performed. This was achieved by seeding equal amounts of cells from each line on 24-well cell culture plates and by giving them also equal amounts of time to grow. Since uniform growth of the cell lines was required to assure a quantitative read out, the growth constants of all lines were determined (Figure 22). For this purpose, an equal number of cells of each cell line was seeded and passaged on 10 cm cell culture dishes. They were split every second day and cells were counted, before again an equal amount of cells was seeded on the next 10 cm cell culture dish. The cells were passaged over 2 weeks and the cell number was measured 5 times. In addition, the time between the measurements was recorded and was also considered in the calculation. The results show that the growth constants of the different cell lines are essentially identical. This is also clearly visible after averaging the constants of the different measurement days (Figure 23).

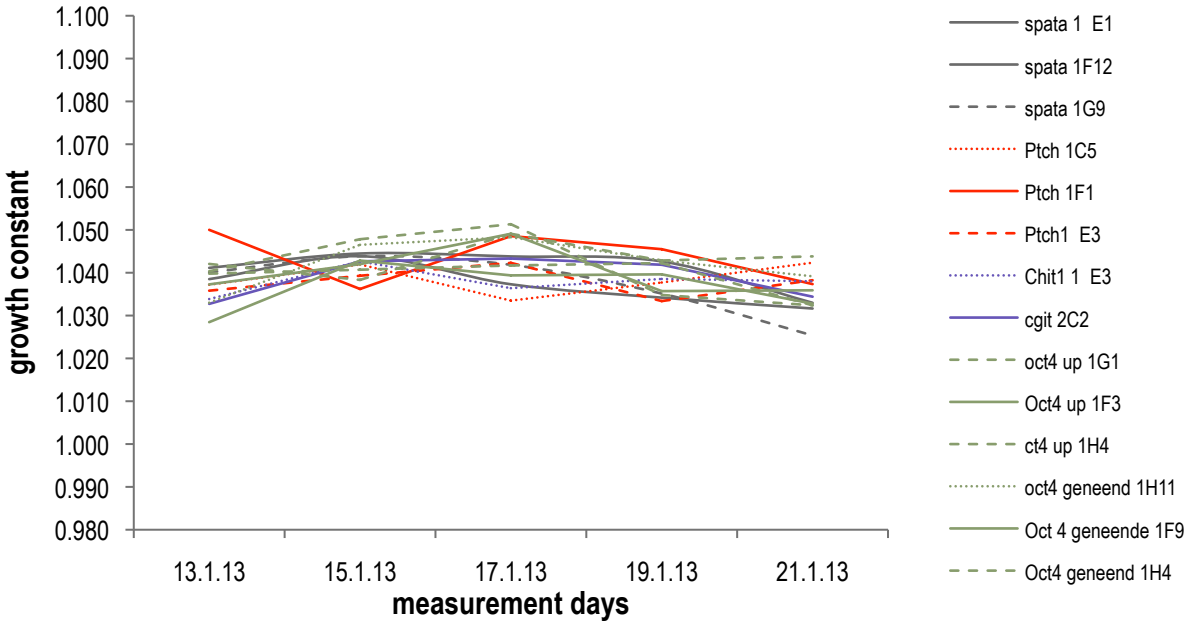


Figure 22| Growth constants from the different measurement days.

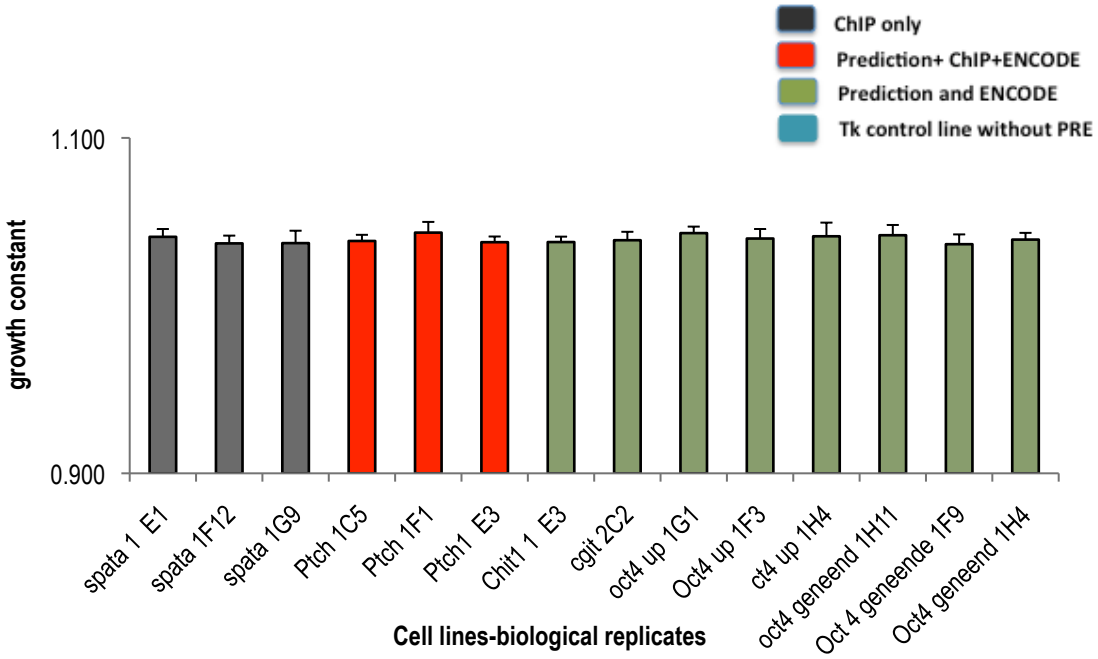


Figure 23| Growth constants from the 5 measurement days averaged. For a further explanation of the color code see 4.3

$$k = \frac{\ln(N_1) - \ln(N_0)}{t_1 - t_0}$$

Equation 3| Formula to calculate the growth constant of the different PRE lines. N…total cell number; $t_{.1}$-$t_{0...}$ time interval between the measurements.

Since the stable cell lines showed a highly similar growth rate, it could be assumed that after seeding an equal amount of cells, after a certain period of time an equal amount of cells could be also harvested for the luciferase assays.

100 000 cells were seeded per cell line on 24 well cell culture plates. After 48 hours cells were lysed, and the luciferase expression level from each PRE line was measured and compared to the expression levels of the TK control line, lacking a PRE. For each line 3 independent clones were assayed (except Chit1 =2 and Tk crtl =2), whereas cells from every clone were seeded in 2 independent wells and this whole procedure was repeated 2 to 3 times.

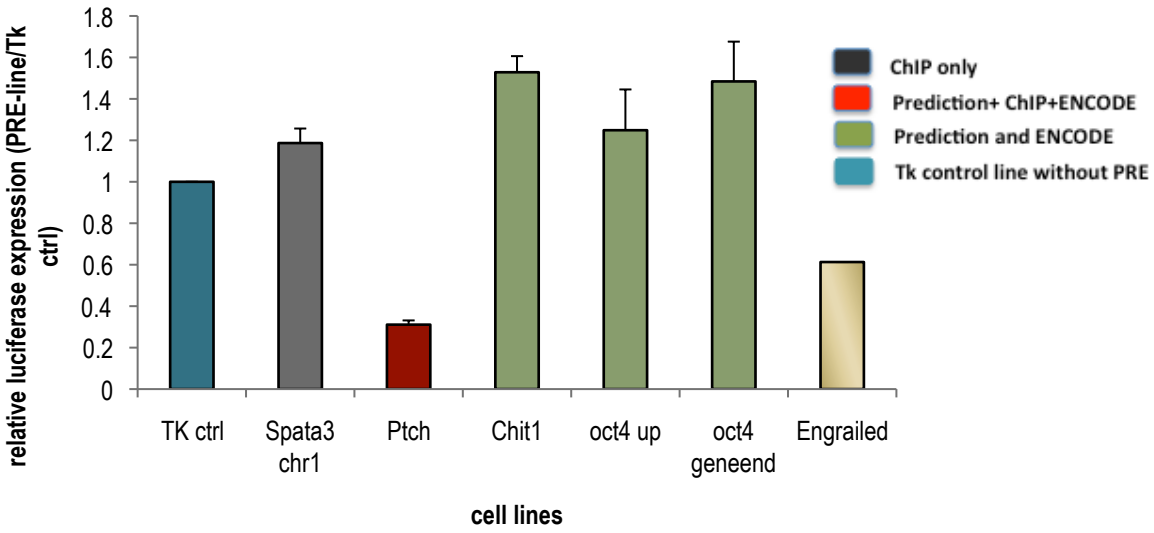


Figure 24| Luciferase expression levels of the transgenic PRE lines relative to the TK- control line and the En-PRE control line. For a further explanation of the color code see 4.3

Interestingly in contrast to the transient repression assay, only one element could mediate repression of the luciferase reporter in stable mouse ESC lines. The Ptch1 line showed a lower luciferase expression compared to the cell line with only the TK promoter (Figure 24). The overall expression was at about 30% of the TK control line, demonstrating the repressive function of this element. A cell line containing the En-PRE, which was recently detected by Robert Heinen a former Postdoc in the Ringrose lab (unpublished data), and the luciferase gene under the control of the TK promoter (cell line created by Melita Ticevic), served as control line. In the former experiments with the En-PRE line, it showed a 4.5 fold repression of the reporter, however under the control of a minimal promoter. The En-PRE line which was used as positive control, is an already in the Ringrose group experimentally validated PRE (R. Heinen et al., manuscript in preparation)

Despite the strong promoter, the En- PRE was able to repress 1.6 fold, demonstrating that it has repressive properties in the context of both the minimal and the TK promoter construct.

7.5 SiRNA knockdown experiments address Polycomb/Trithorax dependency of reporter silencing

In order to evaluate whether PcG proteins mediate reporter repression and in order to determine whether the reporter is derepressed upon PcG depletion, members of the mammalian PRC2 complex were knocked down. This was accomplished via previously established RNAi knockdown strategy of Suz12 and the methyltransferase Ezh2.

Despite the fact that only Ptch1 showed repression of the reporter in mouse ES cells, the de-rerepression assay was performed also on the lines, which did not repress the reporter activity. The level of reporter activity may reflect a balance between PcG mediated silencing and TrxG mediated activation. Thus we reasoned that removal of PcG components may result in activation of the reporter genes above the level of the control construct.

Therefore all transgenic PRE lines were lipofected with Suz12 and Ezh2 siRNAs individually and combined to perform single and double knockdowns of the PcG proteins. Simultaneously, cells were also lipofected with non-targeting control siRNAs in order to reflect the baseline cellular response that occurs when cells are transfected with siRNAs.

7.5.1 Knock down efficiency was validated via Quantitative real time PCR

In order to determine the knock down efficiency of Suz12 and Ezh2, the transcripts were measured in two representative transgenic lines via quantitative real time PCR(qRT-PCR)

For this analysis, primers for the Suz12, Ezh2 and Gapdh transcripts (the latter was used for normalization), were used in turn on the cDNA of Ezh2, Suz12 and siControl transfected ES cells. The Suz12 and Ezh2 levels were normalized to the Gapdh levels within the same cDNA sample and finally compared to the Suz12 and Ezh2 levels of the siRNA control transfection. qRT-PCR confirmed a knockdown efficiency of approximately 5 fold for both Suz12 and Ezh2 on the

mRNA level (Figure 25). Interestingly a cross-regulation could be observed when primers against the Ezh2 transcript were used on cells in which Suz12 was knocked down. In this case Ezh2 was between 1.5 to 2.5-fold up regulated and indeed this cross regulation disappeared after performing also a double knockdown of Suz12 and Ezh2 in ESCs.

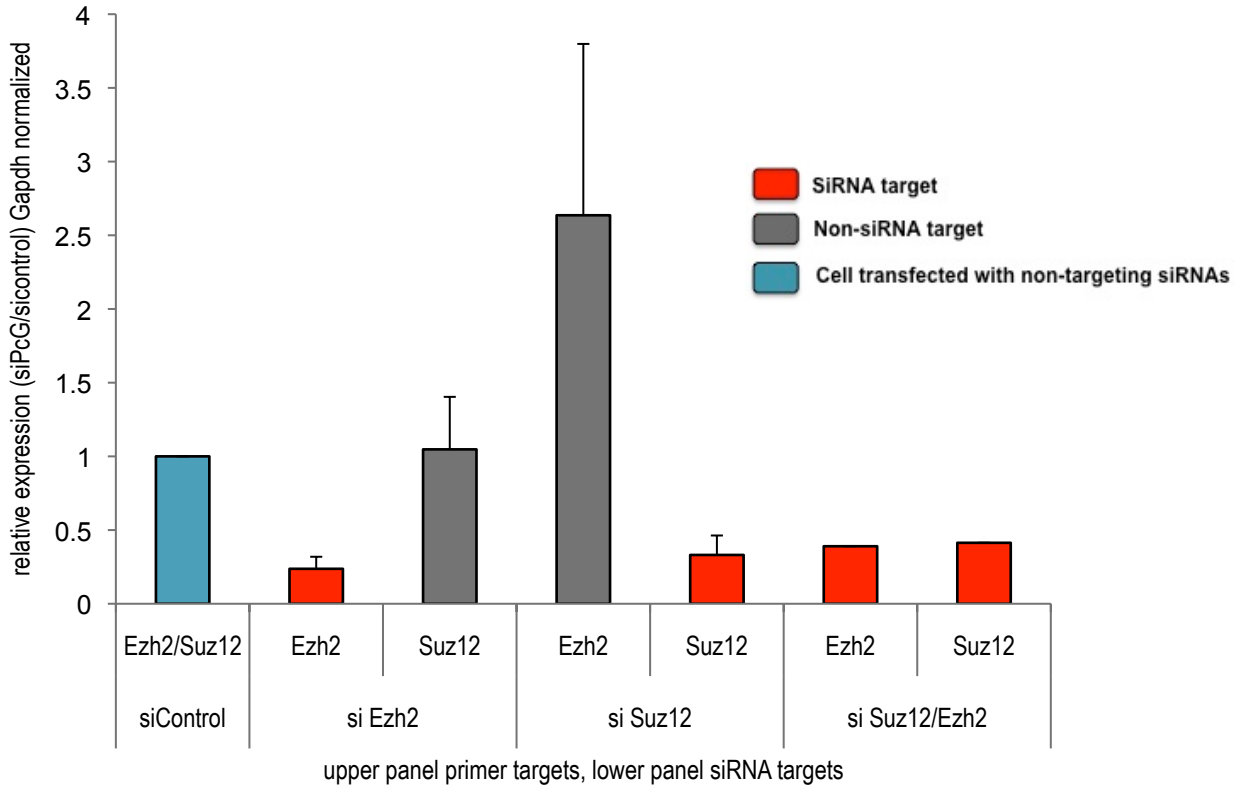


Figure 25| KD efficiency test: qRT-PCR of Ezh2 and Suz12 transcripts in Suz12 and Ezh2 KD ES cells. Red Bars indicate the target transcript level relative to the level in siControl transfected cells (blue bar). Transcript levels were normalized to the GAPDH expression levels within the same cDNA. This was done for cell transfected with siEZH2, siSuz12 and siControl/non targeting. The normalized transcript levels of Ezh2 and Suz12 in siSuz12 and siEzh2 transfected cells were then put relative to the normalized transcript levels of Ezh2 and Suz12 in sicontrol-transfected cells. The averaged values of 2 cell lines are shown in 3 repeats, except double KD, only 1 line, 1 repeat. .

7.5.2 Luciferase expression levels remain unchanged after Knockdown of Suz12 and Ezh2

In order to test the effect of SUZ12 and EZH2 knockdowns on luciferase levels, an equal amount of transgenic ES cells was seeded on 24 well cell culture plates and lipofected with the corresponding siRNAs. 48 hours after transfection the cells were lysed and the reporter activity was measured. The luciferase level of siEzh2 and/or siSuz12 transfected cell lines was compared to the luciferase activity of the same transgenic cell line transfected with the control siRNAs. Values are depicted as relative to the TK control line. For each line 2 independent clones were assayed 3 times, in duplicates, except double Knockdown, 2 times in 2 duplicates.

The results showed that after Suz12 and Ezh2 depletion, disappointingly none of the transgenic PRE lines showed an up regulation of the luciferase expression. Interestingly when Melita Ticevic from the Ringrose group performed derepression assays under the same conditions she could show that the reporter even in the En-PRE line showed no up regulation when combined with the TK promoter. This was surprising since Robert Heinen from the Ringrose group could observe a significant 1.8 fold derepression compared to cells without PcG depletion. This observation, together with the results from the repression assay suggests that via the replacement of the weak MP by the strong TK promoter, repression and derepression can no longer be detected or that the knockdown is not sufficient to detect an effect.

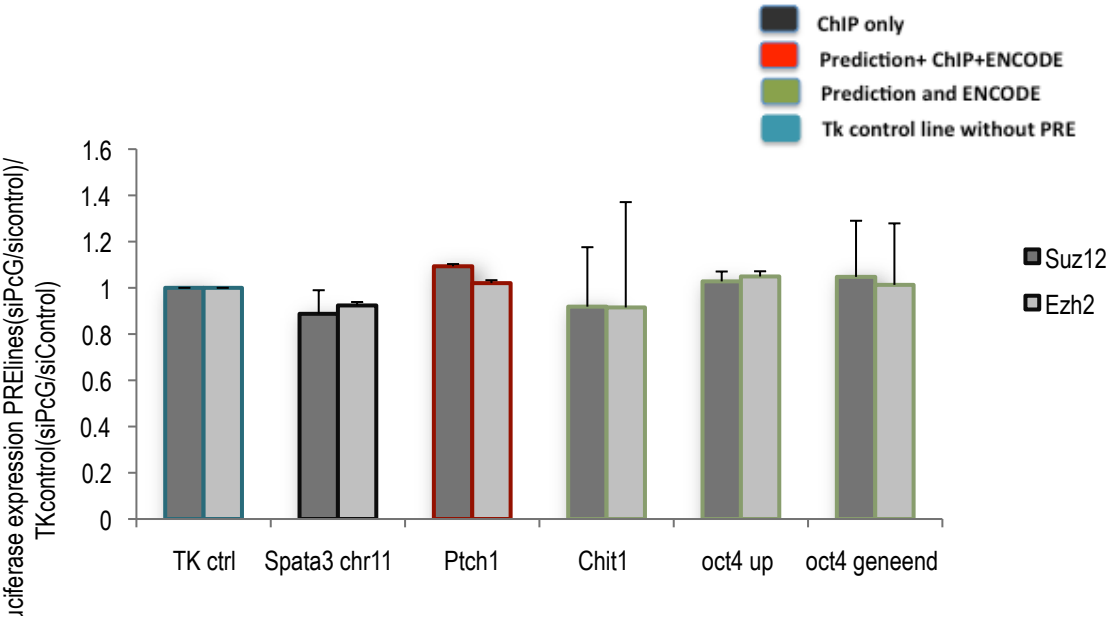


Figure 26| Luciferase expression level of the PRE cell lines compared to TK crtl line after single knock down of Suz12 and Ezh2 via siRNAs, normalized to siCtrl. For a further explanation of the color code see 4.3

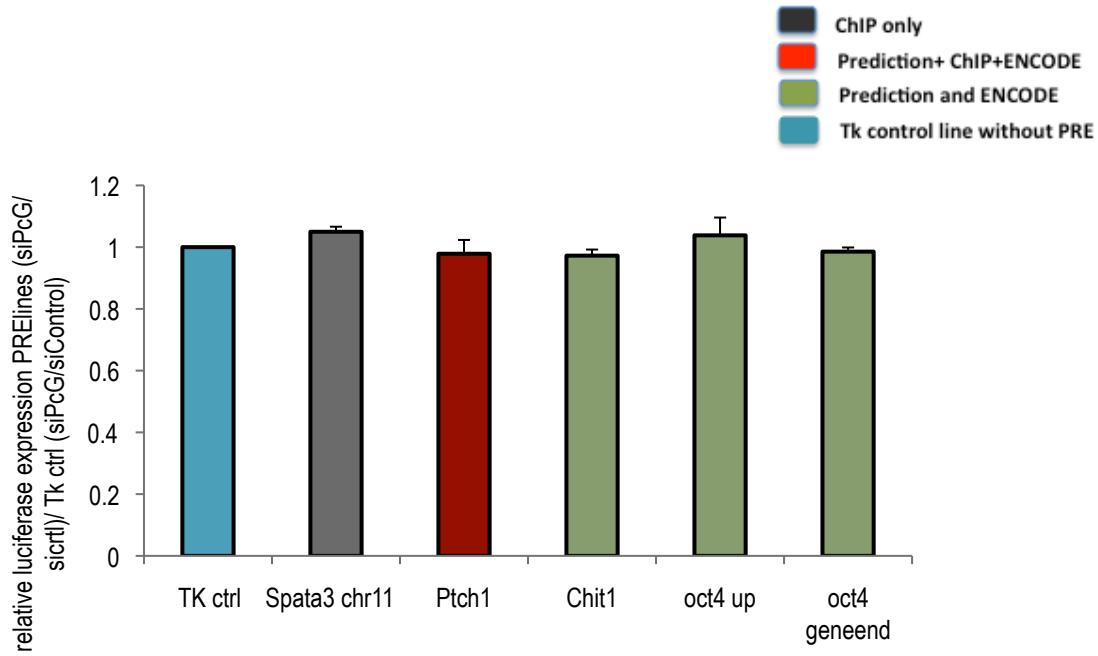


Figure 27| | Results relative luciferase expression compared to TK crtl after double knock down of Suz12 and Ezh2 via siRNAs, normalized to siCtrl. For a further explanation of the color code see 4.3

7.5.3 Knockdown of members of the mammalian MLL complex does not result in up-regulation of the reporter activity

More than 2 decades ago the first evidence that the very same DNA segments can recruit PcG proteins as well as Trithorax group proteins was shown in flies via genetic studies on the Hox gene locus [82]. In addition this evidence was brought via Polytene chromosome stainings where it was clearly shown that PcG and TrxG proteins co-localize at many binding sites. Later it was possible to gain a higher resolution of the DNA segments via ChIP experiments and thereby it was proved that the co-localizing binding sites for PcG and TrxG proteins were indeed already detected fly PREs (reviewed in [1].).

I wanted to address whether the predicted PREs show such a behavior, since there is the possibility that the predicted PREs also contain binding motifs for members of the Trithorax group proteins.

In order to investigate this I knocked down the TrxG proteins Ash2l and MLL. I chose these two proteins since MLL is the methyltransferase, that methylates histone H3K4 and Ash2l, has been reported to be indispensible for the coordination of setting these methylmarks [106]. Furthermore it has been shown that Ash2L can bind directly to DNA via certain motifs and it has been suggested that it can recruit the mammalian Mll complex directly to its target genes [106, 107]. In case of a bi-stable character of the PREs I expected a down regulation of the luciferase expression compared to the TK control line upon depletion of MLL and Ash2l.

The transgenic PRE lines were lipofected with MLL and Ash2l siRNAs individually and combined to perform single and double knockdowns of the TrxG proteins. Simultaneously cells were also lipofected with non-targeting control siRNAs to reflect the baseline cellular response that occurs when cells are transfected with siRNAs. Furthermore the knockdown efficiency was tested via qRT-PCR. The workflow was the same as on page 72, however the knockdown was not as efficient as in the case of Suz12 and Ezh2. Interestingly again, a cross regulation could be observed upon knockdown of one the members of the mammalian MLL complex. In contrast to the observation in the Suz12/ Ezh2 KD experiments, upon depletion of either the MLL or Ash2L transcript I observed also a down regulation of the other transcript.

Subsequently luciferase levels were measured as described in 3.5.2. The results from the luciferase read out showed a slight increase in the reporter expression in case of

the cell lines oct4up and Oct4 gene-end. This observation was against the hypothesis that luciferase levels should decrease after depletion of a putative activator. Nevertheless it is not possible to draw a final conclusion based on these results, since the knockdown efficiency test on protein level is missing. It still can be possible that the knockdown simply was not efficient enough to provoke changes in the luciferase expression levels.

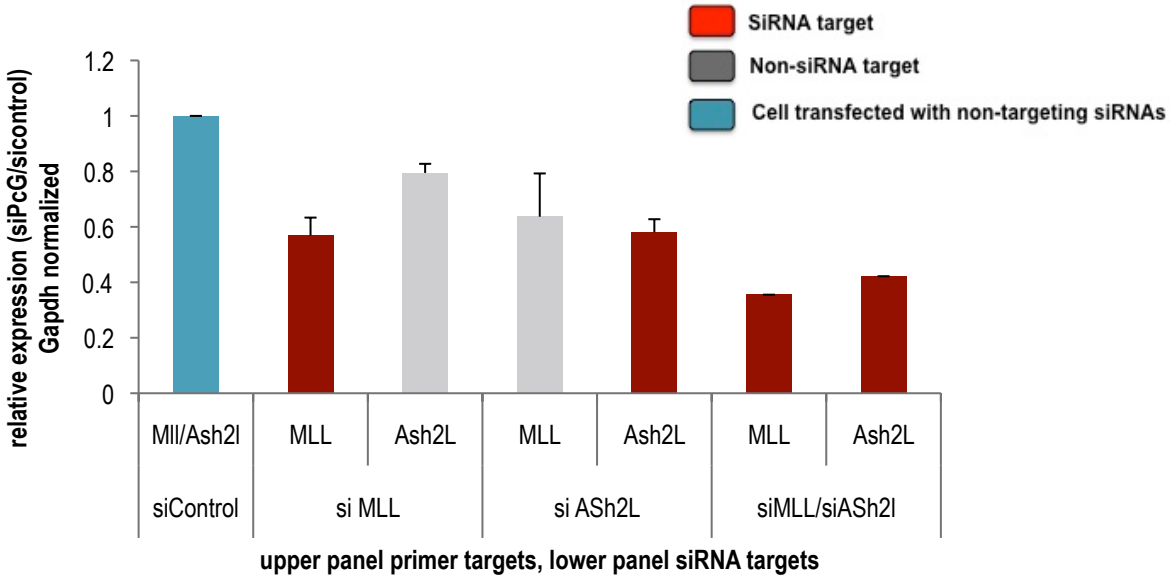


Figure 28| KD efficiency test: qRT-PCR of Ezh2 and Suz12 transcripts in Suz12 and Ezh2 KD cell lines. Red Bars indicate the target transcript levels relative to the level in siControl transfected cells (blue bar), Gapdh normalized. Transcript levels were normalized to the GAPDH expression levels within the same cDNA. This was done for cell transfected with siMLL, siAsh2l and siControl/non targeting. The normalized transcript levels of MLLand Ash2l in siMLL and siAsh2l transfected cells were then put relative to the normalized transcript levels of Ezh2 and Suz12 in sicontrol tranfected cells. The averaged values of 2 cell lines are shown in 2 repeats, except double KD, only 1 line, 1 repeat.

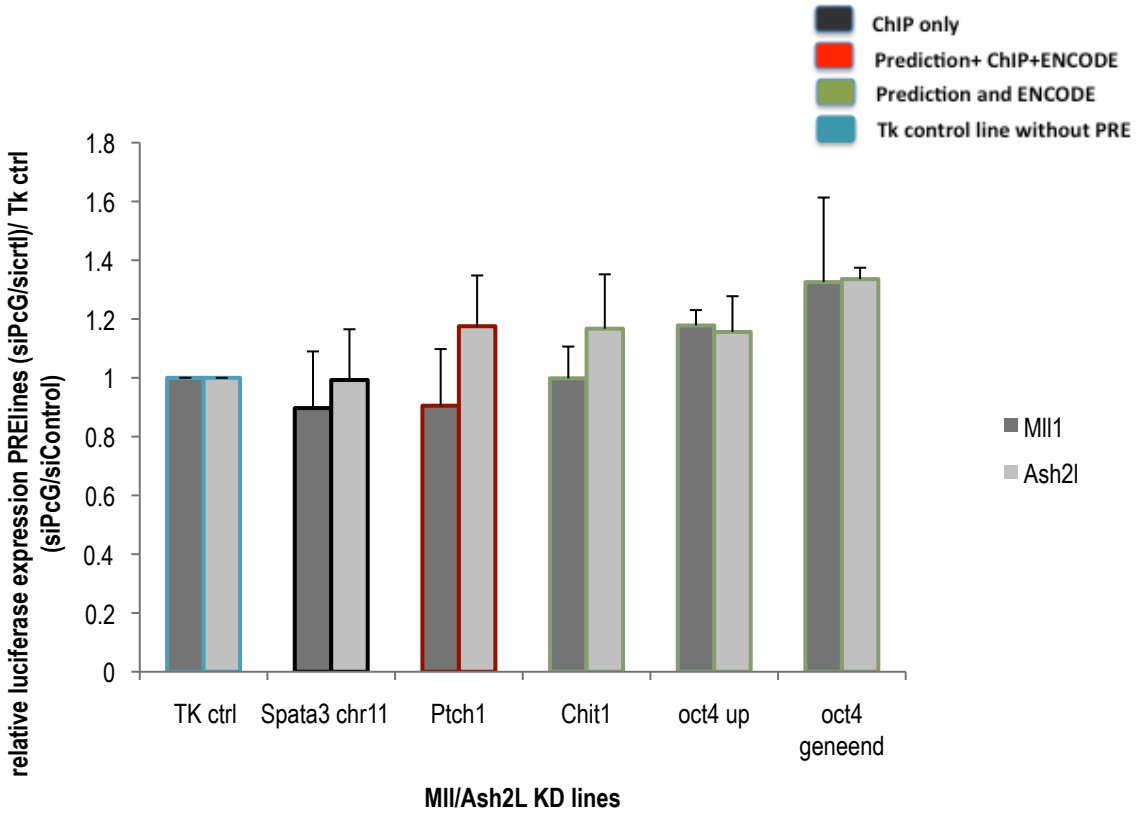


Figure 29| Luciferase expression level of the PRE cell lines compared to TK crtl line after single knock down of MLL and Ash2l via siRNAs, normalized to siCtrl. For a further explanation of the color code see 4.3

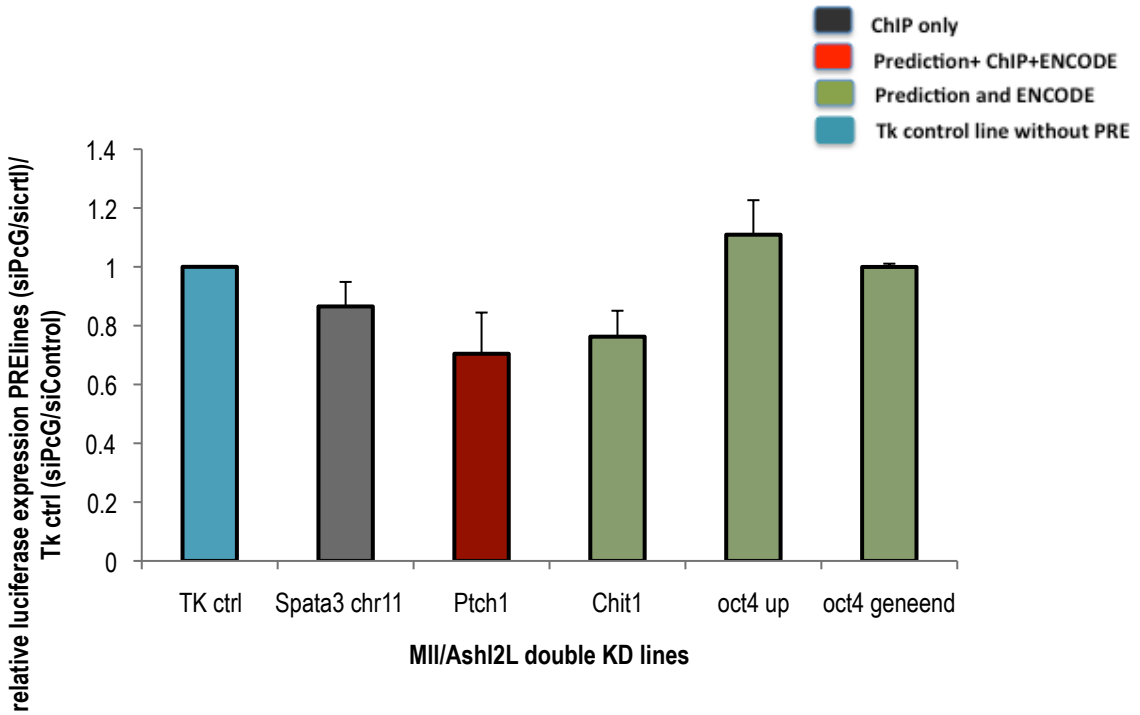


Figure 30| Relative luciferase expression compared to TK crtl after double knock down of MLL and Ash2l via siRNAs, normalized to siCtrl. For a further explanation of the color code see 4.3

7.6 Removal of the hygromycin cassette reduces the level of luciferase expression in stable ESCs

During the workflow of creating stable cell lines, antibiotic resistance genes are normally used in order to discriminate between cells that contain a transgenic construct and wild type cells.

We used for this selection a hygromycin resistance cassette, which is driven by the SV40 early enhancer/ promoter and originates from the pGL4.26 basic vector. LoxP sites for the removal with the Cre-Lox system flanked the whole cassette. Since it is always possible that promoters not only drive the expression of the nearby gene, but also affect the expression of genes further up or downstream of the target gene by inducing an active chromatin environment, I removed the hygromycin cassette via the CRE recombinase to address whether absence of the hygromycin cassette has an influence on the luciferase expression in stable ESCs or not.

For this analysis, one clone of each cell line was transfected with a plasmid carrying the CRE recombinase. 24 hours after transfection the cells were seeded at a clonal level and cell foci were picked at an appropriate size. The removal of the hygromycin cassette was validated via PCR with primers flanking the resistance cassette.

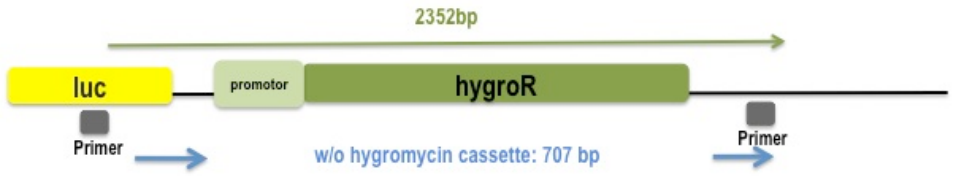


Figure 31| Simplified scheme of the primer design for the validation of hygromycin cassette removal.

The results of the PCR showed that the hygromycin cassette was successfully removed in all tested cell lines, since I expected in its absence a fragment size of 707 bp (Figure 31, Figure 32).

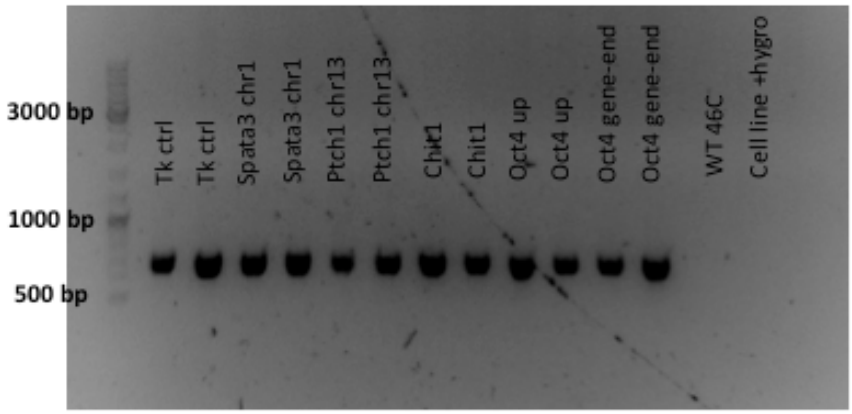


Figure 32|PCR test for removal of the hygromycin cassette

Subsequently repression and de-repression assays were performed with 2 independent clones per cell line, 2 times in duplicates. For the repression assay one clone from each corresponding PRE line, still carrying the hygromycin cassette, was measured in parallel to assure a comparable read out of the absolute luciferase expression values. Interestingly the results demonstrate that the hygromycin cassette had an influence, not only on the absolute expression levels (Figure 33) of the luciferase reporter, but also on the relative values, when compared to the TK control line (Figure 34). The luciferase levels were in general lower, which suggests that the hygromycin promotor has also driven the expression of the luciferase gene. Furthermore the luciferase expression was measured again in the lines lacking the hygromycin cassette after knockdown of Suz12 and Ezh2. Surprisingly the removal of the hygromycin cassette had no effect on the expression of luciferase after knock down of Suz12 and Ezh2 when compared to the results of the lines with the resistance cassette. (Figure 26, Figure 35)

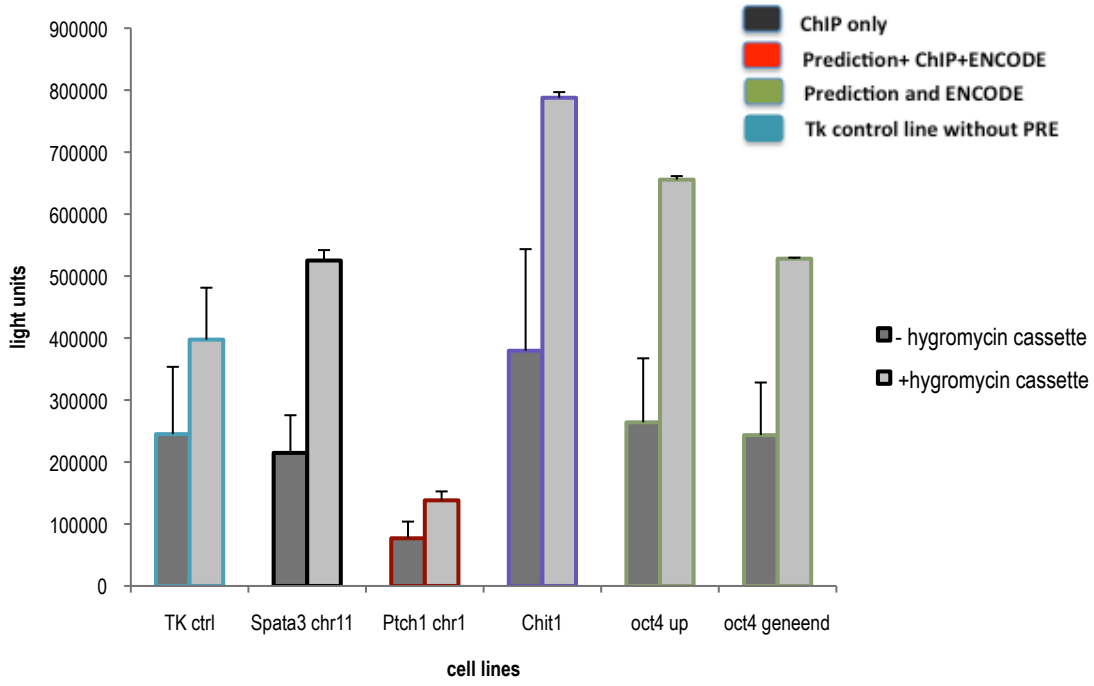


Figure 33| Absolute luciferase expression values of cell lines and without hygromycin cassette in direct comparison. For a further explanation of the color code see 4.3

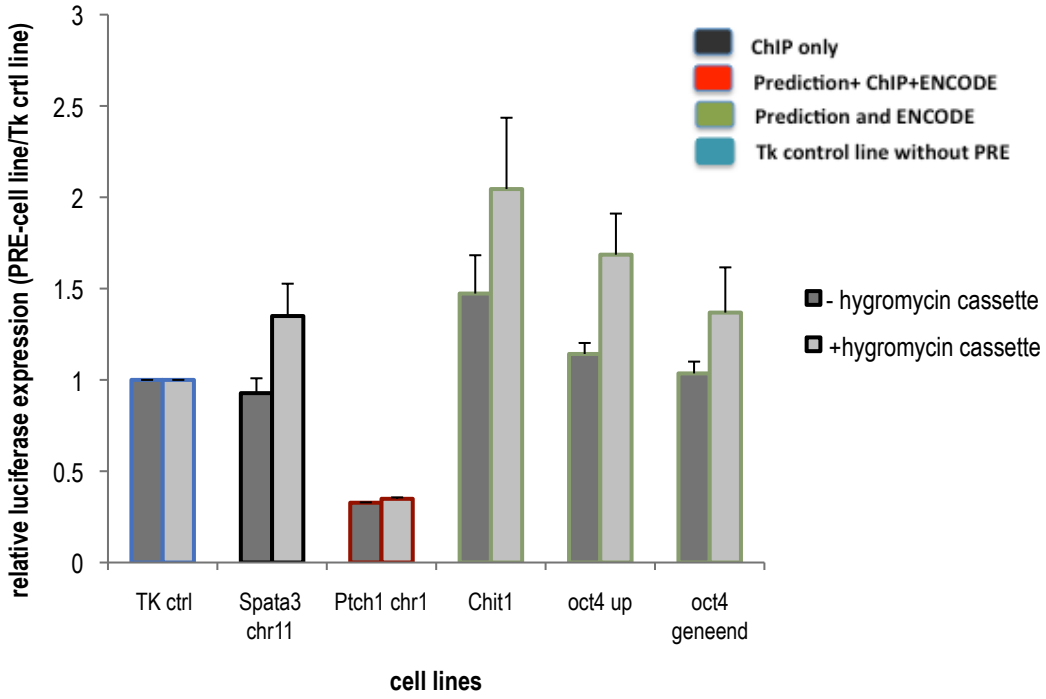


Figure 34| Relative luciferase expression values of cell lines with and without the hygromycin cassette in direct comparison. For a further explanation of the color code see 4.3

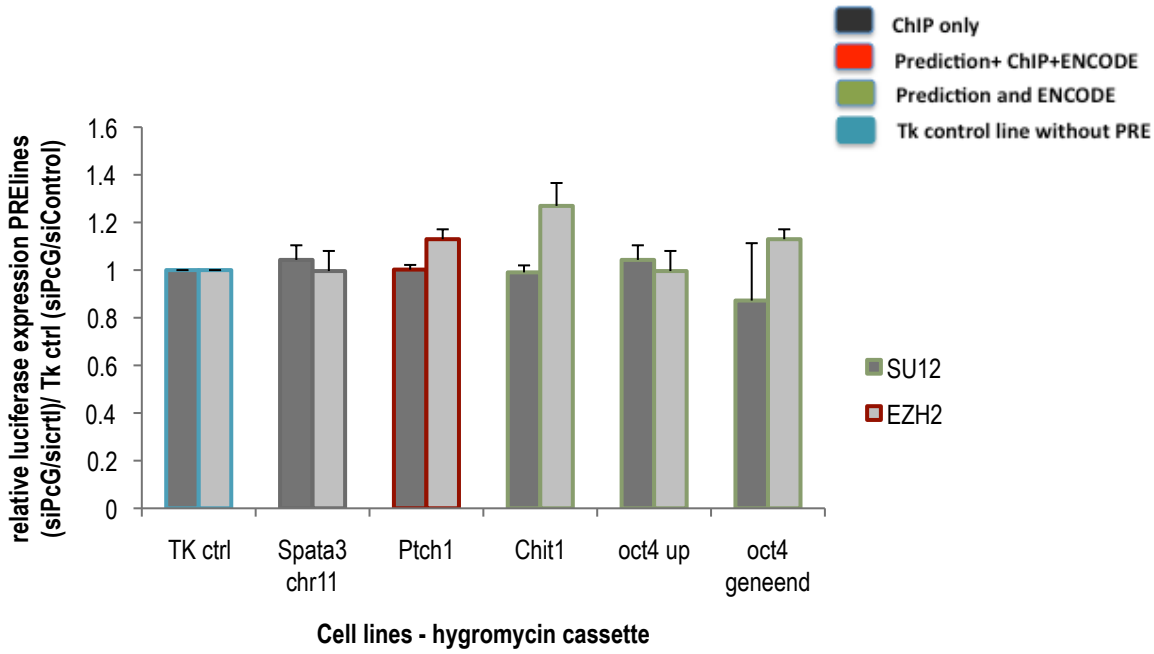


Figure 35| Luciferase expression level of the PRE - hygro cell lines compared to TK –hygro crtl line after Suz12 and Ezh2 single knock down via siRNAs, normalized to siCtrl. For a further explanation of the colour code see 4.3

8 Discussion

In this thesis I tested several computationally predicted PREs experimentally, in order to address the hallmarks of a PRE in the mammalian system (the computational predictions were performed by Johanna Trupke, see also 4.3). The computationally predicted PREs and the control sequences were attributed to different categories, based on the ChiP profile of Jarid2, Suz12 and H3K27me3 marks in ESCs and differentiated cells. The "Red" category is equivalent to a predicted binding region for PcG, contains Suz12 and Jarid2 peaks in ESCs and H3K27me3 peaks in differentiated cells. The "black" category contains regions that are not predicted to function as PREs, but show an overlap with Suz12 and Jarid2 peaks in ESCs and with H3K27me3 in differentiated cells. The "green" category stands for predicted PcG binding sites, for which there are no Suz12 and Jarid2 peaks in ESCs, but which overlap with H3K27me3 enrichments in differentiated cells regarding. Regarding these predictions, PREs of the red category should silence their corresponding gene in ESCs, whereas the green category PREs should have a silencing effect on their target gene in differentiated cells. The black category contains the control sequences, which are not expected to silence the reporter via a Polycomb dependent mechanism. Furthermore the Engrailed (En) PRE line was included in the experiments to act as a positive control. The En-PRE was previously discovered in the Ringrose group and showed repression of the Reporter in ESCs and Derepression upon PcG depletion (R. Heinen et al., manuscript in preparation).
In this chapter the results (7) from the experimental part of a project that addresses the sequence characteristics of mammalian PREs will be discussed.

8.1 Luciferase repression of Ptch 1 in stable stem cell lines and transient experiments

In order to test whether PREs of the red category indeed show a silencing effect on their target gene in ESCs I used a luciferase reporter system, which has already been established in the Ringrose group. Experimental validation of PREs often takes

advantage of reporter systems, since the silencing effect of a PRE can be directly determined via the reporter gene expression level. In flies such a reporter is often the *miniwhite* gene, whose expression has in turn an effect on the eye color of the flies (reviewed in[1]) In mammals, cell culture based systems are frequently used to define PcG recruiting sequences combined with different approaches [22, 83, 84, 86, 108, 109]

In order to evaluate the influence of the predicted PREs on the luciferase reporter I performed luciferase repression assays in transient transfected ES cells and in stable cell lines. Although the study which led to the identification of the first human PRE, performed most of their experiments in transient transfected cells [84], the main goal of this project was always to perform the final experiments with PREs that are stably integrated into the mouse genome, to provide an optimal chromatin environment. Nevertheless, since Woo *et al* [84] performed most of their experiments in transient experiments, we decided to look also at the PRE activity in transient transfected cells. The outcome was that most of the candidate PREs could mediate repression of the reporter, even PREs that were, based on their category, not expected to silence the reporter in ESCs

Table 1) This was surprising but was not enough to challenge the computational prediction, since there exist different opinions about the fate of bacterial DNA when it enters the nucleus of a eukaryotic cell. It is still not completely clear whether the DNA becomes packaged into nucleosoms or to what extent [103-105]. This knowledge is especially important for studies involving PcG or TrxG proteins since they work trough histone modifications. Although PREs are reported at least in flies to be free of nucleosoms [3, 110, 111], their target genes are not [112], and therefore nucleosomal packaging can have a large influence on the outcome of such an experiment.

For this reason I integrated a selection of the PRE reporter constructs into the ROSA-26 locus in mouse ESCs in order to determine whether the predicted PREs are able to silence the reporter at an ectopic locus. We expected different silencing effects, based on the different categories, which were attributed to the PREs. Indeed the outcome was that Ptch1 (red category, expected to silence) and the En-PRE (positive control, expected to silence) mediate repression of the reporter whereas PREs from the black and green categories showed no repression or showed even an activation of the reporter compared to a control line without PRE. The En-PRE cell line was used as a positive control, since the En-PRE was already validated in a previous

study, which used a minimal do drive the expression of the luciferase gene (Robert Heinen, unpublished data) According to these results the reporter construct with the luciferase gene driven by the TK promotor worked very well. Spata3, which was not predicted to function as PRE, did not show silencing of the reporter neither could candidate PREs from the green category silence the reporter. This was also consistent with the predicted category, since they are only predicted to silence in differentiated cells.

Compared to the study of Woo *et al,* [84] who used a similar luciferase reporter assay system the silencing effect of Ptch1 was almost as strong as the effect of the HoxD11-12 PRE characterized in that study, which means in case of the Ptch1 PRE that the luciferase activity was reduced to less than 30 % and in case of the HoxD9-10 PRE a reduction of the luciferase expression to 5%. (7.4 compare to [84]) Rather surprising was on the contrary that Woo *et al* additionally reported that they could not measure any luciferase activity when they used a reporter system consisting only of a TK promotor and the luciferase reporter. This is in complete disagreement with our observations since the luciferase expression levels of the control constructs which consisted exactly out of these two elements showed a very high expression level (for absolute values see Figure 33). One reason for this may be that Woo et al [84] used a different Luminometer, which was not as sensitive as ours, or that they used different settings, which prevented the detection of the expression level driven by only the TK promotor.

Rather unexpected was in addition the outcome of our transient and stable repression assays and the discrepancy in the results regarding the effect of the different PREs on the reporter, namely that although strong repression was observed in the transient assay, very few elements repressed the reporter upon integration into the genome (see above, 7.4). A possible explanation for these differences could be the different chromatin condition for the transgene on the plasmid compared to stable cell lines [103-105]. Nevertheless Woo *et a*l [84] preformed a lot of their experiments with either transient transfected cells or in stable cell lines, but never showed a comparison of the results side by side This encourages my opinion that, if working with chromatin based mechanism its indispensible to work with transgenes integrated into the genome to provide an optimal chromatin environment.

In summary, the effect of the PREs was according to their categories and also the extent of reporter gene silencing given by the Ptch1 element was at the level or even

higher compared to already validated PREs. For a complete picture of PREs from categories, which should only silence in differentiated cells, these elements will have to be tested in differentiated cells. Furthermore the Polycomb dependency of reporter silencing needs to be proven, which can be addressed via depletion of Polycomb in ESCs.

8.2 Possible limitations of the siPcG knockdown assay

I performed SiRNA knockdowns of Suz12 and Ezh2, members of the mammalian PRC2 complex and important for silencing of Polycomb target genes (see 4.2.1.1), in order to address the question of whether the observed silencing effect on the reporter is PcG dependent. A typical siRNA knockdown workflow consist of the transfection of the cells with siRNAs, the incubation time, which allows the siRNAs to target the transcripts, followed by the determination of the knockdown efficiency and the measurement of the related phenotype (for the exact procedure see 6.2.7). Optimal timing of each of these steps is crucial for a successful knockdown rate. Further parameters which can influence the outcome of such an experiment are the cell type, transfection reagent and of course the siRNAs [113].

Derepression assays, in which Suz12 and Ezh2 were knocked down in order to prove PcG dependency of reporter silencing, were already performed successfully in the Ringrose lab and used for the identification of 3 mammalian PREs at the Engrailed, the Insulin like growth factor 2 (Igf2) and the LIM homeobox transcription factor 1, alpha (Lmx1A) locus. Reporter constructs carrying the PREs showed significant derepression within a broad range of derepression levels of up to 3fold (R. Heinen et al., manuscript in preparation)

I tested the knockdown efficiency with quantitative real time PCRs and could determine a reduction of Suz12 and Ezh2 mRNAs of 70-80 % in mouse ES cells. Interestingly a cross regulation could be observed between the two members of the PRC2 complex. Knockdown of Suz12 led to a massive up-regulation of Ezh2 in ESCs. This effect disappeared upon knockdown of both PcG proteins in the same cells (7.5.1) A cross regulation between PcG proteins upon depletion of PcG members has been already observed but on protein not on mRNA level [114]. In order to determine whether this cross-regulation maintains on protein level, western blots with antibodies against Suz12 and EZH2 in Suz12 and Ezh2 knockdown cell would need to be performed.

Disappointingly the luciferase values in cells depleted for Suz12 and Ezh2 compared to cells with normal levels of Ezh2 and Suz12 showed no significant up-regulation (7.5.2). There was also no effect on the En-PRE control line observable which was surprising since the En-PRE cell line with the MP showed in previous experiment up regulation of the luciferase expression after Suz12 and EZh2 knockdown (R. Heinen et al., manuscript in preparation). One reason for this may be an insufficient knockdown of Suz12 and Ezh2, since these are very abundant proteins in ESCs. Furthermore we cannot be sure that a knockdown of 70-80 % on mRNA level equals exactly this efficiency on protein level. The final protein level is dependent on the stability of the mRNA and on the regulation on a translational level. One study, which worked also with Suz12 and Ezh2, has been shown that the knockdown efficiencies of Suz12 and EZH2 on mRNA level equal the knockdown efficiency on protein level. However these experiments were not performed in ESCs [54]. And therefore it could be, that this is still different in ESCs. A Western blot to determine the protein levels after knockdown of Suz12 and Ezh2 is indispensable for the interpretation of the effective knockdown efficiency.

In case that the results from the Western Blot experiment show that Suz12 and Ezh2 are still highly expressed on protein level, the knockdown efficiency could be further improved by changing different parameters (see 8.2 above). For this thesis the knockdown experiments were performed in ESCs on 24 well plates and cells were harvested 48 hours after transfection (more detailed in 6.2.8). Since we cannot change the cell type, which can also influence the transfection efficiency, a starting point for improvements could be the incubation time. 48 hours is the suggested time for the optimal siRNA depletion of mRNAs but not optimal for a knockdown on protein levels, which requires 48-96 hours [113]. However in our case we tried to make a compromise between the experimental set up on 24 well plates and the conditions for an optimal knockdown efficiency. On 24 well plates a cell confluence of 100 % was already reached after 48 hours. Longer incubation would lead to cell death and unequal growth, which is unsuitable for a quantitative luciferase read out. A possible solution for that problem would be to start with a larger surface area or the splitting of cells after 48 hours. This would allow us to incubate the cells for another 48 hours. However such an approach would have an effect on the scale of the experiments and would restrict the number of lines which can be tested in parallel per person. In general there exists also an alternative to siRNA knockdowns, namely the usage of

shRNA delivery by lentiviruses. This would have the advantage to provide long lasting, highly efficient gene silencing [115].

Another possible explanation for the lack of derepression of the luciferase levels besides the knock down efficiency could be the fact that we only targeted members of the PRC2 complex and not members of the PRC1 complex. This could be problematic since it has been shown that the proteins of these complexes have redundant functions [114]. Furthermore it has been shown that PRC1 can bind independently from H3K27 to chromatin [68, 69] and even to DNA independently from chromatin [59].Thus it would be interesting to repeat the knockdown by targeting both members of the PRC1 and the PRC2 complex.

Additional controls for the efficiency of the knockdown could be introduced, such as the effect on endogenous target genes of the PcG by RT-qPCR analysis.

The possibility that the PREs contain also binding sites for members of the TrxG proteins as is the case in flies (reviewed in [4] , was addressed by a knockdown of MLL and Ash2l, important members of the TrxG proteins. Binding of TrxG proteins would also interfere with a derepression of the reporter after PcG knockdown. However, almost no effect could be observed on the expression of the luciferase reporter, when compared to cell lines, which were only transfected with non targeting siRNAs (7.5.3). The knockdown efficiency test showed a 50 % down regulation of Mll and Ash2l transcripts in ESCs. Again it is complicated to estimate how effective the knockdown really was without the results of the knockdown efficiency on the protein level. It can be on the one hand that the knock down was efficient enough to provoke a change in the luciferase expression, which would mean for our experiment that TrxG proteins are not involved in the regulation at these loci in ESCs. On the other hand it could be that with a better knockdown efficiency the effect from TrxG proteins on the reporter could become measurable. Therefore no conclusion can be drawn from the luciferase read out until the knockdown is validated on protein level, and its effect on endogenous target genes of the PcG and TrxG proteins has been validated.

In summary, determination of Suz12 and Ezh2 protein levels, and expression of endogenous target genes after siRNA knockdown will be useful for a further interpretation of the knockdown efficiency tests and the interpretation of the corresponding luciferase read out.

This is also true for the knockdown of MLL and Ash2l, the results of which could provide further important information about a possible bi-stable character of PREs in mammals.

8.3 The hygromycin resistance cassette has a strong influence of on the reporter expression in stable stem cell lines

During the creation of stable cell lines, markers such as antibiotic resistance genes are normally used to identify cell clones that carry a specific transgenic construct. For this project we used a hygromycin resistance cassette for the identification of the transgenic cell lines. A strong viral promotor drives the expression of the hygromycin resistance gene, which is located only about 500 bp downstream of the luciferase reporter (Figure 11). It is always possible that promoters drive not only the expression of the nearby gene, but also affect the expression of genes further up or downstream of the target gene by inducing an active chromatin environment. I removed the hygromycin cassette, to address whether this is also true for our transgenic constructs, by taking advantage of the Cre-LoxP system. Indeed after measuring the luciferase expression in embryonic cell lines with removed hygromycin cassette I could observe not only a difference in the absolute luciferase expression values but also in the relative expression levels when compared to the control construct. The relative repression level of Ptch1 stayed the same whereas the activating effect of the cell lines Spata3, Oct4 up and Oct4 gene end disappeared after removal of the hygromycin cassette. However, the removal of the hygromycin cassette had no influence on the derepression experiment, which is another indication that the knockdown efficiency levels are maybe not sufficient to provoke a derepression of the luciferase values. These results have to be kept in mind for future experiments, since the changes in the chromatin structure provoked by the hygromycin cassette could also affect the influence the PRE can have on the expression of the reporter.

8.4 Conclusion and future perspectives

In this thesis I started with the experimental characterization of computationally predicted PREs in mouse embryonic stem cells. For this purpose I used an already partly established luciferase reporter assay system and integrated PREs in front of a reporter in order to measure the effect on the luciferase expression levels. I created stable embryonic stem cell lines and performed repression assays in transient experiments and with the stable ESC lines. I could show in both approaches that Ptch1, which was predicted to function as PRE in ESCs could silence the reporter activity compared to a control cell line lacking the PRE. Furthermore I could show discrepancies between the results of the transient experiments and the stable cell lines, which suggests that transient experiments are not the optimal choice when dealing with questions that address the chromatin structure. I knocked down Suz12 and Ezh2, important members of the PRC2 complex in order to address the question whether reporter repression is Polycomb dependent or not. I could show possible limitations of the system, regarding the knock down efficiency, which provides important knowledge for future experiments with this system.

Furthermore I could show that the promoter of the hygromycin selection marker has a strong influence on the expression of the reporter, which has to be kept in mind for future experiments.

In general the luciferase reporter assay system offers great potential for the quantitative experimental characterization of PREs in mammals, although some improvement may have to made concerning the de-repression experiments. In order to improve different parameters of the siRNA knock down workflow, the knockdown efficiency levels have to be determined on a proteins level via Western Blot analysis. Furthermore it would be interesting to test other integration sites for the PRE reporter constructs beside Rosa-26. Another target site may substantially affect the ability of a PRE to silence the reporter, since Rosa-26 is an extremely open euchromatic chromatin region.

Another indispensible experiment in order to address PcG dependency of reporter silencing are ChIP experiments in order to determine PcG enrichment at the transgenic construct.

In future it will be important to gain, with this combination of computational prediction and experimental approaches, a better knowledge of the sequence principles, which underlie mammalian PREs.

9 Acknowledgements

First of all I would like to thank my Supervisor Leonie Ringrose for the opportunity to do my Master Thesis project in her research group and for her continuous support and encouragement in any situation. Furthermore I want to thank the whole Ringrose group for the absolutely wonderful year and of course I want to thank Johanna who made the predictions of the PREs. Thanks to Veronika, Robert and Melita who became real friends and were always up for spirited conversations and to my friend Lucia: July is our month! Finally I also want to thank my Mother, my sister Nadja and my sweetheart Juergen for always supporting me in life and during my studies.

In memory of my father who will unfortunately never read this thesis, but who always believed in me and supported me until the end.

10 References

1. 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.
2. Schwartz, Y.B. and V. Pirrotta, *Polycomb silencing mechanisms and the management of genomic programmes.* Nat Rev Genet, 2007. **8**(1): p. 9-22.
3. Muller, J. and J.A. Kassis, *Polycomb response elements and targeting of Polycomb group proteins in Drosophila.* Curr Opin Genet Dev, 2006. **16**(5): p. 476-84.
4. Ringrose, L. and R. Paro, *Polycomb/Trithorax response elements and epigenetic memory of cell identity.* Development, 2007. **134**(2): p. 223-32.
5. Ringrose, L., et al., *Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster.* Dev Cell, 2003. **5**(5): p. 759-71.
6. He, S., D. Nakada, and S.J. Morrison, *Mechanisms of stem cell self-renewal.* Annu Rev Cell Dev Biol, 2009. **25**: p. 377-406.
7. Schuettengruber, B., et al., *Genome regulation by polycomb and trithorax proteins.* Cell, 2007. **128**(4): p. 735-45.
8. Margueron, R. and D. Reinberg, *The Polycomb complex PRC2 and its mark in life.* Nature, 2011. **469**(7330): p. 343-9.
9. Kennison, J.A., *Introduction to Trx-G and Pc-G genes.* Methods Enzymol, 2004. **377**: p. 61-70.
10. Christophersen, N.S. and K. Helin, *Epigenetic control of embryonic stem cell fate.* J Exp Med, 2010. **207**(11): p. 2287-95.
11. Dillon, N., *Factor mediated gene priming in pluripotent stem cells sets the stage for lineage specification.* Bioessays, 2012. **34**(3): p. 194-204.
12. Cavalli, G. and R. Paro, *Epigenetic inheritance of active chromatin after removal of the main transactivator.* Science, 1999. **286**(5441): p. 955-8.
13. Ptashne, M.a.G., A., *Genes and Signals.* Cold Spring Harbor, New
York: Cold Spring Harbor Laboratory Press, 2002.
14. Scott, M.P. and S.B. Carroll, *The segmentation and homeotic gene network in early Drosophila development.* Cell, 1987. **51**(5): p. 689-98.
15. Lewis, E.B., *A gene complex controlling segmentation in Drosophila.* Nature, 1978. **276**(5688): p. 565-70.
16. Maeda, R.K. and F. Karch, *The ABC of the BX-C: the bithorax complex explained.* Development, 2006. **133**(8): p. 1413-22.
17. Klymenko, T. and J. Muller, *The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins.* EMBO Rep, 2004. **5**(4): p. 373-7.
18. Poux, S., et al., *The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing.* Development, 2002. **129**(10): p. 2483-93.
19. van Lohuizen, M., et al., *Sequence similarity between the mammalian bmi-1 proto-oncogene and the Drosophila regulatory genes Psc and Su(z)2.* Nature, 1991. **353**(6342): p. 353-5.
20. Gunster, M.J., et al., *Identification and characterization of interactions between the vertebrate polycomb-group protein BMI1 and human homologs of polyhomeotic.* Mol Cell Biol, 1997. **17**(4): p. 2326-35.

21. Brunk, B.P., E.C. Martin, and P.N. Adler, *Drosophila genes Posterior Sex Combs and Suppressor two of zeste encode proteins with homology to the murine bmi-1 oncogene.* Nature, 1991. **353**(6342): p. 351-3.
22. Mendenhall, E.M. and B.E. Bernstein, *Chromatin state maps: new technologies, new insights.* Curr Opin Genet Dev, 2008. **18**(2): p. 109-15.
23. Ringrose, L., *Polycomb comes of age: genome-wide profiling of target sites.* Curr Opin Cell Biol, 2007. **19**(3): p. 290-7.
24. Schwartz, Y.B. and V. Pirrotta, *Polycomb complexes and epigenetic states.* Curr Opin Cell Biol, 2008. **20**(3): p. 266-73.
25. Breiling, A., L. Sessa, and V. Orlando, *Biology of polycomb and trithorax group proteins.* Int Rev Cytol, 2007. **258**: p. 83-136.
26. Schuettengruber, B., et al., *Trithorax group proteins: switching genes on and keeping them active.* Nat Rev Mol Cell Biol, 2011. **12**(12): p. 799-814.
27. 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.
28. Min, J., Y. Zhang, and R.M. Xu, *Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27.* Genes Dev, 2003. **17**(15): p. 1823-8.
29. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing.* Science, 2002. **298**(5595): p. 1039-43.
30. Hansen, K.H., et al., *A model for transmission of the H3K27me3 epigenetic mark.* Nat Cell Biol, 2008. **10**(11): p. 1291-300.
31. Schmitges, F.W., et al., *Histone methylation by PRC2 is inhibited by active chromatin marks.* Mol Cell, 2011. **42**(3): p. 330-41.
32. Tanaka, Y., et al., *Trithorax-group protein ASH1 methylates histone H3 lysine 36.* Gene, 2007. **397**(1-2): p. 161-8.
33. Tie, F., et al., *CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing.* Development, 2009. **136**(18): p. 3131-41.
34. Schwartz, Y.B., et al., *Alternative epigenetic chromatin states of polycomb target genes.* PLoS Genet, 2010. **6**(1): p. e1000805.
35. Pasini, D., et al., *Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes.* Nucleic Acids Res, 2010. **38**(15): p. 4958-69.
36. Zentner, G.E. and S. Henikoff, *Regulation of nucleosome dynamics by histone modifications.* Nat Struct Mol Biol, 2013. **20**(3): p. 259-66.
37. Gorisch, S.M., et al., *Histone acetylation increases chromatin accessibility.* J Cell Sci, 2005. **118**(Pt 24): p. 5825-34.
38. Cao, R., Y. Tsukada, and Y. Zhang, *Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing.* Mol Cell, 2005. **20**(6): p. 845-54.
39. 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.
40. Wang, H., et al., *Role of histone H2A ubiquitination in Polycomb silencing.* Nature, 2004. **431**(7010): p. 873-8.
41. Brookes, E., et al., *Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs.* Cell Stem Cell, 2012. **10**(2): p. 157-70.

42. 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.
43. Pasini, D., et al., *Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity.* EMBO J, 2004. **23**(20): p. 4061-71.
44. van der Lugt, N.M., et al., *Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene.* Genes Dev, 1994. **8**(7): p. 757-69.
45. Yu, B.D., et al., *Altered Hox expression and segmental identity in Mll-mutant mice.* Nature, 1995. **378**(6556): p. 505-8.
46. Faust, C., et al., *The eed mutation disrupts anterior mesoderm production in mice.* Development, 1995. **121**(2): p. 273-85.
47. Core, N., et al., *Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice.* Development, 1997. **124**(3): p. 721-9.
48. 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.
49. Rowley, J.D., *The critical role of chromosome translocations in human leukemias.* Annu Rev Genet, 1998. **32**: p. 495-519.
50. Valk-Lingbeek, M.E., S.W. Bruggeman, and M. van Lohuizen, *Stem cells and cancer; the polycomb connection.* Cell, 2004. **118**(4): p. 409-18.
51. Czermin, B., et al., *Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites.* Cell, 2002. **111**(2): p. 185-96.
52. Muller, J., et al., *Histone methyltransferase activity of a Drosophila Polycomb group repressor complex.* Cell, 2002. **111**(2): p. 197-208.
53. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.* Genes Dev, 2002. **16**(22): p. 2893-905.
54. 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.
55. Pasini, D., et al., *The polycomb group protein Suz12 is required for embryonic stem cell differentiation.* Mol Cell Biol, 2007. **27**(10): p. 3769-79.
56. Ketel, C.S., et al., *Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes.* Mol Cell Biol, 2005. **25**(16): p. 6857-68.
57. Bernstein, E., et al., *Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin.* Mol Cell Biol, 2006. **26**(7): p. 2560-9.
58. Jung, H.R., et al., *Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36.* Mol Cell Proteomics, 2010. **9**(5): p. 838-50.
59. 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.
60. Shao, Z., et al., *Stabilization of chromatin structure by PRC1, a Polycomb complex.* Cell, 1999. **98**(1): p. 37-46.
61. Levine, S.S., et al., *The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans.* Mol Cell Biol, 2002. **22**(17): p. 6070-8.

62. Mulholland, N.M., I.F. King, and R.E. Kingston, *Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA.* Genes Dev, 2003. **17**(22): p. 2741-6.
63. Saurin, A.J., et al., *A Drosophila Polycomb group complex includes Zeste and dTAFII proteins.* Nature, 2001. **412**(6847): p. 655-60.
64. Schwendemann, A. and M. Lehmann, *Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci.* Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12883-8.
65. Huang, D.H., et al., *pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex.* Mol Cell Biol, 2002. **22**(17): p. 6261-71.
66. Dejardin, J., et al., *Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1.* Nature, 2005. **434**(7032): p. 533-8.
67. Wang, L., et al., *Hierarchical recruitment of polycomb group silencing complexes.* Mol Cell, 2004. **14**(5): p. 637-46.
68. Atchison, L., et al., *Transcription factor YY1 functions as a PcG protein in vivo.* EMBO J, 2003. **22**(6): p. 1347-58.
69. Brown, J.L., et al., *The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1.* Mol Cell, 1998. **1**(7): p. 1057-64.
70. Shen, X., et al., *Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells.* Cell, 2009. **139**(7): p. 1303-14.
71. 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.
72. Pasini, D., et al., *JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells.* Nature, 2010. **464**(7286): p. 306-10.
73. Li, G., et al., *Jarid2 and PRC2, partners in regulating gene expression.* Genes Dev, 2010. **24**(4): p. 368-80.
74. Landeira, D., et al., *Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators.* Nat Cell Biol, 2010. **12**(6): p. 618-24.
75. Vincenz, C. and T.K. Kerppola, *Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16572-7.
76. Petruk, S., et al., *TrxG and PcG proteins but not methylated histones remain associated with DNA through replication.* Cell, 2012. **150**(5): p. 922-33.
77. Tanaka, Y., et al., *Dual function of histone H3 lysine 36 methyltransferase ASH1 in regulation of Hox gene expression.* PLoS One, 2011. **6**(11): p. e28171.
78. Gregory, G.D., et al., *Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes.* Mol Cell Biol, 2007. **27**(24): p. 8466-79.
79. Steward, M.M., et al., *Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes.* Nat Struct Mol Biol, 2006. **13**(9): p. 852-4.
80. Hekimoglu-Balkan, B., et al., *Intergenic Polycomb target sites are dynamically marked by non-coding transcription during lineage commitment.* RNA Biol, 2012. **9**(3): p. 314-25.

81. Simon, J., et al., *Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products.* Dev Biol, 1993. **158**(1): p. 131-44.
82. Chan, C.S., L. Rastelli, and V. Pirrotta, *A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression.* EMBO J, 1994. **13**(11): p. 2553-64.
83. Sing, A., et al., *A vertebrate Polycomb response element governs segmentation of the posterior hindbrain.* Cell, 2009. **138**(5): p. 885-97.
84. Woo, C.J., et al., *A region of the human HOXD cluster that confers polycomb-group responsiveness.* Cell, 2010. **140**(1): p. 99-110.
85. Mendenhall, E.M., et al., *GC-rich sequence elements recruit PRC2 in mammalian ES cells.* PLoS Genet, 2010. **6**(12): p. e1001244.
86. Lynch, M.D., et al., *An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment.* EMBO J, 2012. **31**(2): p. 317-29.
87. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes Dev, 2011. **25**(10): p. 1010-22.
88. Lee, T.I., et al., *Control of developmental regulators by Polycomb in human embryonic stem cells.* Cell, 2006. **125**(2): p. 301-13.
89. Boyer, L.A., et al., *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. **441**(7091): p. 349-53.
90. Bracken, A.P., et al., *Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions.* Genes Dev, 2006. **20**(9): p. 1123-36.
91. Ku, M., et al., *Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains.* PLoS Genet, 2008. **4**(10): p. e1000242.
92. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.
93. Squazzo, S.L., et al., *Suz12 binds to silenced regions of the genome in a cell-type-specific manner.* Genome Res, 2006. **16**(7): p. 890-900.
94. Consortium, E.P., *A user's guide to the encyclopedia of DNA elements (ENCODE).* PLoS Biol, 2011. **9**(4): p. e1001046.
95. 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.
96. Fiedler, T. and M. Rehmsmeier, *jPREdictor: a versatile tool for the prediction of cis-regulatory elements.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W546-50.
97. Chen, X., et al., *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.
98. Hughes, J.D., et al., *Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae.* J Mol Biol, 2000. **296**(5): p. 1205-14.
99. Pavesi, G., et al., *Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W199-203.
100. Bailey, T.L., et al., *MEME: discovering and analyzing DNA and protein sequence motifs.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W369-73.
101. Tanay, A., et al., *Hyperconserved CpG domains underlie Polycomb-binding sites.* Proc Natl Acad Sci U S A, 2007. **104**(13): p. 5521-6.
102. 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.

103. Lechardeur, D. and G.L. Lukacs, *Nucleocytoplasmic transport of plasmid DNA: a perilous journey from the cytoplasm to the nucleus.* Hum Gene Ther, 2006. **17**(9): p. 882-9.
104. Reeves, R., C.M. Gorman, and B. Howard, *Minichromosome assembly of non-integrated plasmid DNA transfected into mammalian cells.* Nucleic Acids Res, 1985. **13**(10): p. 3599-615.
105. V. Mladenova, E.M., G. Russev, *ORGANIZATION OF PLASMID DNA INTO NUCLEOSOME-LIKE*
STRUCTURES AFTER TRANSFECTION IN EUKARYOTIC CELLS. Biotechnol. & Biotechnol, 2009. **Eq. 23/2009/1**.
106. Wan, M., et al., *The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells.* J Biol Chem, 2013. **288**(7): p. 5039-48.
107. Sarvan, S., et al., *Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain.* Nat Struct Mol Biol, 2011. **18**(7): p. 857-9.
108. 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.
109. Arnold, P., et al., *Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting.* Genome Res, 2013. **23**(1): p. 60-73.
110. Mohd-Sarip, A., et al., *Architecture of a polycomb nucleoprotein complex.* Mol Cell, 2006. **24**(1): p. 91-100.
111. Papp, B. and J. Muller, *Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins.* Genes Dev, 2006. **20**(15): p. 2041-54.
112. Schwartz, Y.B., et al., *Genome-wide analysis of Polycomb targets in Drosophila melanogaster.* Nat Genet, 2006. **38**(6): p. 700-5.
113. *siRNA-Induced mRNA Knockdown and Phenotype: How to Measure and When to Measure.*
114. Leeb, M., et al., *Polycomb complexes act redundantly to repress genomic repeats and genes.* Genes Dev, 2010. **24**(3): p. 265-76.
115. Rao, D.D., et al., *siRNA vs. shRNA: similarities and differences.* Adv Drug Deliv Rev, 2009. **61**(9): p. 746-59.

11 Curriculum Vitae

Personal

Second, first name	Drexel Tanja, BSc
Address	Akaziengasse 44-46/65/15 1230 Wien 0660/4839466
E-Mail	tanjadrexel@gmail.com
Date of birth	23.01.1984
Place of birth	6800 Feldkirch
Family status	Unmarried, no children

Education

1994-2003	"Bundesgymnasium Feldkirch"(grammar school)
06/2003	"Matura" (General qualification for university entrance)
10/2003	Human Medicine at the Medical University of Vienna
01/2005- 03/2010	Interruption of my studies and start as an employee at Libro Gmbh
11/2008	Start Bachelor studies in Biology
01/2012	Bachelor of Science degree
02/2012 –07/2013	Master studies in Molecular Biology at the University of Vienna
06/2012- 07/2013	Start of my Master Thesis project at the Institute for molecular Biotechnology (IMBA)

Additional Qualifications

Language skills	German	Mother tongue
	English	Written, oral fluent skills
	Spanish	Written good, oral basic skills
	Latin	Latin proficiency certificate

Computer skills	
Operating system	Windows (XP, Vista, 7), Mac OS
EDV	Microsoft Office, Apple iWork Adobe Illustrator, Photoshop, R
Cloning and Sequencing programs	CLC workbench, gene constructor kit, serial cloner
Online databases	NCBI, UCSC, Ensemble

Research Experience

Period	Project	Methods
09/2011 Bachelor thesis Mag. Dr. Armin Resch Department of Microbiology, Immunobiology & Genetics Dr. Bohr-Gasse 9/4 1030 - Vienna, Austria	The influence of the small non-coding RNA Dicf on the translation of *miaA in* E-coli	PCR/colony PCR, β-Galactosidase assays; cloning of PCR products; production of competent E-coli cells, transformations, DNA preparation, Expression studies: +/- IPTG
03/ 2012 Internship Ass. Prof. Dr. Ferdinand Steinböck Medical University of Vienna, Institute of cancer research Cellular and molecular tumorbiology	The influence of epicatechin on the mutation frequency in stationary yeast cells	Western Blot, FACs, SDS Page, forward mutation assays

05/2012 Training (1 week) A.o.Univ.Prof.Dr. Klaus Holzmann Medical University of Vienna, Institute of cancer research Development of Experimental Cancer Therapies	Training in Cell culture (cancer cells) and visualization of cells.	Real Time qPCR, passaging and microscopic visualisation of cells
04/ 2012 Internship Univ. Prof. Dr. Karl Kuchler Medical University Vienna, Max F. Perutz Laboratories, Department of Medical Biochemistry, Christian Doppler Laboratory	Characterization of KO--strains of putative Pdr12 interactors in yeast	ATPase assays, liquid growth assays, spotting assays
06/2012 – 07/2013 Masters project Leonie Ringrose PhD Institute of Molecular Biotechnology Dr. Bohr-Gasse 3 1030 Vienna, Austria	Functional characterization of mammalian Polycomb Response Elements	Cloning of PCR products, creation of stable cell lines in mouse ES cells, Southern Blots (radioactive labeling), diagnostic PCRs, transient transfections, luciferase reporter assays in stable cell lines and transient transfected cells, siRNA knockdown, qPCR, ChiP analysis, usage of the Cre-Lox system, Western Blot

Meetings attended

October 2012	Epigenetics and the Environment workshop in Munich
November 2012	Biomimetics Symposium

Place/Date
Vienna, 7th June 2013

Signature