Poised enhancers are key cis-regulatory elements during ESC differentiation whose activity is facilitated by Polycomb repressive complex 2

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Sara de la Cruz Molina

aus Tordesillas (Spanien)

Köln 2017

The Doctoral Thesis “Poised enhancers are key cis-regulatory elements during ESC differentiation whose activity is facilitated by Polycomb repressive complex 2” was performed at the Center for Molecular Medicine of Cologne in collaboration with the Cluster of Excellence of Cologne, both institutions belong to the University of Cologne, from October 2013 to June 2017.

Berichterstatter:
Prof. Dr. Björn Schumacher
Prof. Dr. Mirka Uhlirova
Prof. Dr. Siegfried Roth

Tag der mündlichen Prüfung: 05.10.17

Para mis padres y hermano

Acknowledgements

First of all, I would like to thank Alvaro Rada-Iglesias for the excellent supervision and support during my PhD-thesis and for giving me the opportunity to work on this amazing project. Furthermore, I would like to thank him for his enthusiasm, his motivating discussions and encouragement. I greatly appreciate his dedication.

I would like to thank Prof. Dr. Björn Schumacher, Prof. Dr. Mirka Uhlirova and Prof. Dr. Roth for agreeing to build my thesis committee.

Of course a big “thanks” to all the former and present members of the AG Rada-Iglesias for the nice time in the lab. Special thanks to Patricia for her good advises inside and outside lab, Christina for the uncountable colony work and Magdalena for her scientific and personal care.

Moreover, I would like to thank AG Papantonis for their help with the 4Cs, especially to Akis for his suggestions and advices, Lili for her patience and support and Theo for the technical recommendations.

Thanks to the Mensa lunch group for making every day more fun.

This list wouldn’t be completed without my expat colleagues, the ones who are still working and to the ones who have already finished. I am especially thankful to Marta for her constant support, to Borja for being a great friend in every sense, to Laura for the scientific and not scientific discussions and to Lara for being up to do any plan in any moment. Thanks also to Alexander for the translations.

Thanks a lot to my parents for their support, love and patience. To Carmen for being the first one who wanted to read my paper. And finally, I would like to especially thank Iñigo, who supported me before I started my PhD. When things were going wrong; I always thought in the strength, kindness and generosity he had with me, and these nice feelings helped me to go on.

TABLE OF CONTENTS

Abstract

The developmental transitions occurring during embryogenesis involve spatial and temporal changes in gene expression patterns, which are largely dependent on a group of regulatory elements known as enhancers. Enhancers are short DNA sequences able to positively control the expression of their target genes in a distance and orientation independent manner. Previous studies uncovered a unique set of enhancers in pluripotent embryonic stem cells (ESCs), named “Poised Enhancers”. Poised enhancers are marked by repressive histone marks, like the trimethylation of histone three lysine 27, which is deposited by Polycomb repressive complex 2 (PRC2), and also are bound by co-activators like P300. The fact that poised enhancers display both activating and repressing features as well as evidences indicating that they are associated with genes involved in early organogenesis led us to suggest that these regulatory elements are already bookmarked in ESCs and thus primed for their future activation once the differentiation process starts. However, the functional relevance of poised enhancers and the role of their unique chromatin features remain unknown.

To gain insights into these major questions, first, poised enhancers were identified in mouse embryonic stem cells (mESCs) and their activation was evaluated during the establishment of anterior neural identity. Second, using CRISPR/Cas9 technology, poised enhancer candidates were deleted in mESCs, which were then differentiated into anterior neural progenitors. In general, the poised enhancer deletions resulted in severely reduced induction of the poised enhancer’s target genes. Furthermore, circularized chromosome conformation capture coupled to sequencing experiments revealed that poised enhancers and their target genes physically interacted already in ESCs, thus preceding the activation of poised enhancers and their target genes. Interestingly, these poised enhancer-target gene interactions observed in mESCs were found to be PRC2 dependent. Additionally, it was demonstrated that while PRC2 was not necessary to maintain poised enhancers in an inactive state in mESCs, was required for the induction of the poised enhancers’ target genes upon anterior neural differentiation. Finally, poised

enhancers were found to frequently reside within a high CpG-dinucleotide genomic context that can directly mediate the recruitment of PRC2.

Overall, these findings demonstrate that poised enhancers are essential for the proper expression of the anterior neural developmental program. Importantly, our work illuminates an unexpected function for PRC2 in promoting neural induction. Our data demonstrates that both poised enhancers and their target genes display intrinsic sequence features that can directly mediate PRC2 recruitment. Consequently, PRC2 can bring poised enhancers and their targets into physical proximity already in mESCs, thus providing a permissive regulatory topology that we propose can facilitate the future induction of mayor anterior neural genes.

Zusammenfassung

Die Entwicklungsphasen während der Embryogenese werden durch räumliche und zeitliche Veränderungen in den Genexpressionsmustern bestimmt, die stark von bestimmten Enhancern reguliert werden. Enhancer sind kurze DNA Sequenzen, die in der Lage sind die Expression von Genen unabhängig von Distanz und Orientierung positiv zu beeinflussen. Einige Studien haben bereits eine spezielle Gruppe von Enhancern in pluripotenten embryonischen Stammzellen (ESC) identifiziert, die "Poised Enhancers" (PE) genannt werden. PE sind durch repressive Histone-Markierungen, wie z.B., die Trimethylatisierung von Histone-3-Lysine-27 (H3K27me3) gekennzeichnet, welche durch den Polycomb-Repressive-Complex-2 (PRC2) angebracht werden. Gleichzeitig werden sie aber auch durch Co-Aktivatoren wie P300 markiert. Da die PE‘s gleichzeitig aktivierende und repressive Eigenschaften aufweisen und mit bestimmten Genen assoziiert sind, die in der frühen Embryogenese aktiv sind, liegt die Vermutung nahe, dass sie in den ESC vormarkiert werden, um für eine bevorstehende Aktivierung bei einer anstehenden Differenzierung bereit zu sein. Die funktionelle Relevanz der PE‘s und die Rolle ihrer einzigartigen Chromatin Eigenschaften sind bisher noch ungeklärt.

Um diese wichtigen Fragen beantworten zu können, wurden in dieser Arbeit zuerst geeignete PE-Kandidaten in Maus-ESC‘s (mESC‘s) identifiziert und ihre Aktivierung während der frühen neuronalen Entwicklung evaluiert. Zweitens, wurden exemplarisch einige PE-Kandidaten mit Hilfe des CRISPR-Cas9-Systems in mESC entfernt, welche anschließend in anteriore neuronale Vorläuferzellen differenziert wurden (AntNPC). Alle hier erzielten Deletionen der PE führten zu einer stark reduzierten Induktion der PE-Zielgene in den AntNPC. Darüber hinaus zeigten 4C-seq-Experimente, dass PE und ihre Zielgene nicht nur in AntNPC, sondern auch in ESC, d.h. vor der Aktivierung der PE und ihrer Zielgene physikalisch interagieren. Interessanterweise konnte auch nachgewiesen werden, dass die Interaktionen der PE mit ihren Zielgenen von PRC2 abhängig sind. Außerdem wurde hier festgestellt, dass PRC2 nicht

notwendig ist, um die PE‘s in den mESC inaktiv zu halten, jedoch für die Aktivierung des PE-Zielgens während der AntNPC Differenzierung notwendig ist. Anschließend wurde gezeigt, dass PE‘s oft mit einem hohen CpG-dinucleotide Genkontext auftreten, die mit der Rekrutierung von PRC2 in Verbindung stehen.

Zusammenfassend zeigen die hier gezeigten Ergebnisse, dass die PE‘s essenziell für eine normale Expression des vorangehenden neuralen Entwicklungsprogramms sind. Die Daten demonstrieren auch, dass beide, die PE und ihre Zielgene, intrinsische Sequenzeigenschaften besitzen, welche die PRC2-Rekrutierung beeinflussen. Dementsprechend kann PRC2 die PE‘s und ihre Zielgene bereits in mESC in physikalische Nähe zueinander bringen und so eine permissive regulatorische Topologie bereitstellen, welche die zukünftige Induktion der anterioren neuronalen Vorläufergene erlaubt.

1- INTRODUCTION

1.1- *In vitro* modeling of *in vivo* embryogenesis

Embryogenesis is the process of cell division and cellular differentiation of the embryo from the one-cell state or zygote to an adult being. In order to study this process, ESCs derived from the epiblast of mammalian embryos can be used as an *in vitro* highly tractable model. ESCs possess two fundamental characteristics: self-renewal and pluripotency. Self-renewal is the capability of replication of a cell into the same cell state and pluripotency is the ability of a single cell to generate all the cell lineages. In the case of mouse embryonic stem cells (mESCs), pluripotency can be captured in culture under different conditions (see methods 3.32 and 3.33) and each condition represents a different pluripotency state. Naïve pluripotency, which characterizes *in vivo* the epiblast cells of the mouse pre-implantational blastocyst, can be recapitulated *in vitro* by the addition of MEK and GSK3 inhibitors (i.e. 2i conditions) (Hackett and Surani 2014). Formative pluripotency represents a more advance stage distinctive of the post-implantaton epiblast, represented *in vitro* by epiblast-like cells (EpiLC) (Hayashi et al. 2011). Finally, as development proceeds and gastrulation starts, the epiblast cells aquire a primed pluripotent state in which somatic lineage specifiers start to be expressed and that is recapitulated *in vitro* by epiblast stem cells (EpiSC) (Smith 2017; Kalkan and Smith 2014).

Stem cell fate is determined by the equilibrium between self-renewal and differentiation signals. ESCs are able to respond to the signaling environment and either maintain the core pluripotency program or to exit the pluripotent state when differentiation signals appear. To initiate differentiation, ESCs have to escape from self-renewal, the transcriptional networks providing stem cell identity need to be extinguised and dififferentiation towards a specific lineage has to be specified. ESC differentiation involves the silencing of *Nanog*, *Pou5F1* and other major pluripotency regulators. Moreover, lineage specific transcription factors (TFs) need to be induced in order to initiate the expression of a specific differentiation program (Young 2011). Ultimately, all these

processes are controlled at the chromatin level where transcriptional regulation is mediated by cis-regulatory elements like enhancers, promoters, silencers and insulators (Calo and Wysocka 2013; Signolet and Hendrich 2015). Enhancers in particular, play a key role during the exit of pluripotency and subsequent differentiation by controlling the specific set of genes that any given cell type expresses (de Laat and Duboule 2013).

Lastly, during the exit of pluripotency, Polycomb group proteins (PcG) play also an important role. These proteins can act as transcriptional activators in certain cellular and genomic contexts as well as mediators of both short and long-range genomic interactions that contribute to global genome architecture (Bantignies et al. 2011; Creppe et al. 2014; Denholtz et al. 2013; Entrevan et al. 2016; Joshi et al. 2015; Kondo et al. 2014; Kondo et al. 2016; Lanzuolo et al. 2007; Morey et al. 2015; Schoenfelder et al. 2015).

This work is focused on the role of a unique group of enhancers, called poised enhancers as well as the function of PRC2 during mESCs differentiation.

1.2- Enhancers regulate cell type transitions

Enhancers are cis-regulatory elements that positively control gene expression over long distances and in an orientation independent manner (Spitz and Furlong 2012). They are compact DNA sequences (200-500 bp) that act as a binding platforms for TFs (Figure 1.1) and play an important role in driving cell-type specific gene expression (Bulger and Groudine 2011; Yáñez-Cuna et al. 2013). TFs can bind to enhancers in a combinatorial and modular manner, allowing the establishment of different spatiotemporal gene expression programs (Figure 1.1) (Spitz and Furlong 2012). The importance of enhancers during embryogenesis is well illustrated in several studies where deletions of specific enhancers in animal models cause developmental abnormalities and affects gene expression patterns (Sagai et al. 2005).

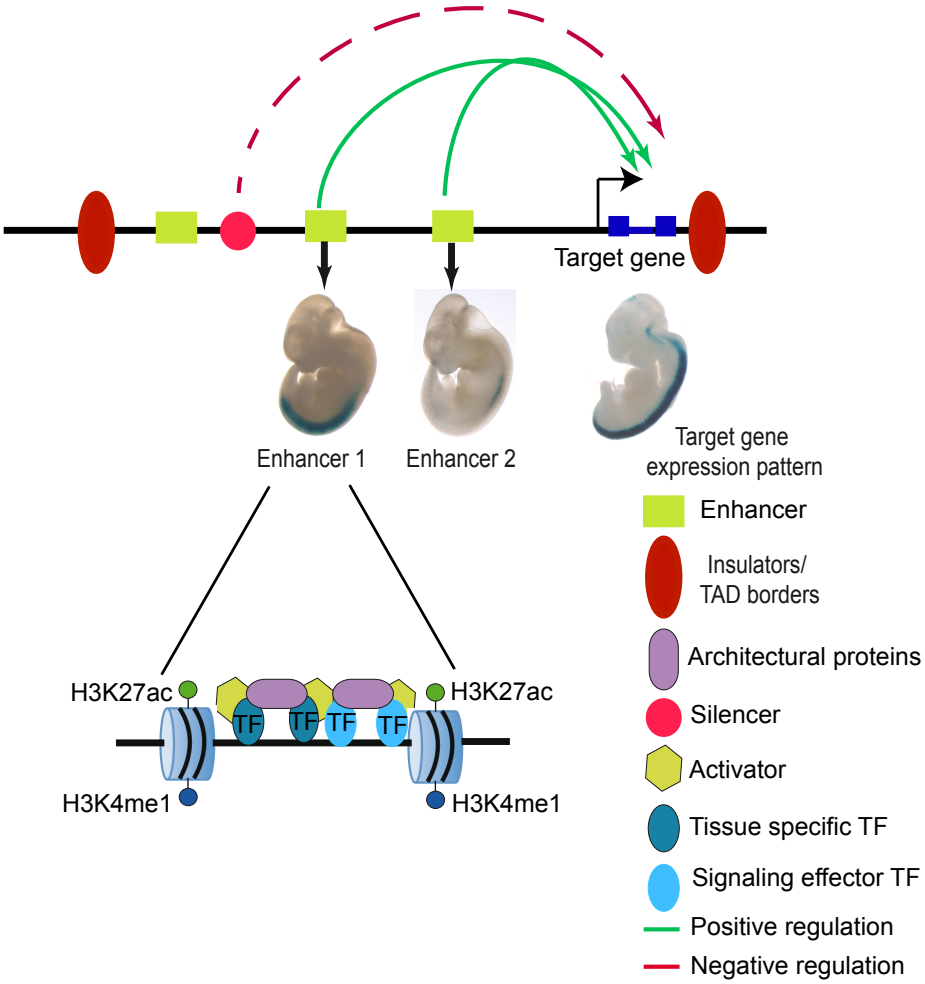

Figure 1.1: Genetic and epigenetic features of enhancers. Enhancers are able to confer cell type and developmental-stage specific spatiotemporal gene expression patterns by serving as integrating platforms of different types of developmental information (tissue-specific and signaling-dependent TFs, epigenetic modifications, recruitment of co-activators and architectural proteins). The regulatory influence of enhancers on the expression of nearby genes is delimited by insultators. (Adapted from Noonan, 2009).

1.2.1- Identification and characterization of enhancers

Despite their relevance, enhancers were difficult to identify due to the fact that they can drive transcription in a distance and orientation independent manner with respect to their putative genes, they do not present a stereotypic sequence composition and, in addition, they are not placed at a fixed position or distance with respect to their target gene promoters. However, this dramatically changed with the emergence of epigenomic approaches that based on the presence of particular chromatin signatures, allowed the global

identification of enhancers in a sequence conservation and cell type independent manner. Recent advances in genomic techniques like chromatin immunoprecipitation coupled to sequencing (ChIP-seq), (see methods 3.5.1) have enabled the global mapping of TFs and histone modification binding profiles. Importantly, enhancers can be identified using a variety of chromatin features, including the presence of certain histone marks, binding of transcription factors or chromatin accessibility (Figure 1.1) (Boroviak et al. 2015; Kalkan and Smith 2014; Ying et al. 2008; Buenrostro et al. 2015; Menno P. Creyghton et al. 2010; Heintzman et al. 2009; Rada-Iglesias et al. 2011).

Classically, the target genes of enhancers have been inferred based on correlative observations like proximity. However, enhancers can control the expression of genes located at great distances, which tipically involves the physical proximity between enhancers and their target genes (Calo and Wysocka 2013). Therefore, enhancers do not necessarily control the expression of their nearest genes. Chromatin Conformation Capture (3C) technologies and their derivatives allow evaluating the physical interaction between different genomic loci (de Wit and de Laat 2012; Duan et al. 2012; Simonis, Kooren, and de Laat 2007; Splinter et al. 2012; van de Werken et al. 2012). For example, by using Circular Chromatin Conformation Capture coupled to sequencing (4C-seq), the genes that physically interact with the enhancers of interest and that are likely to represent their target genes can be identified (Figure 1.2).

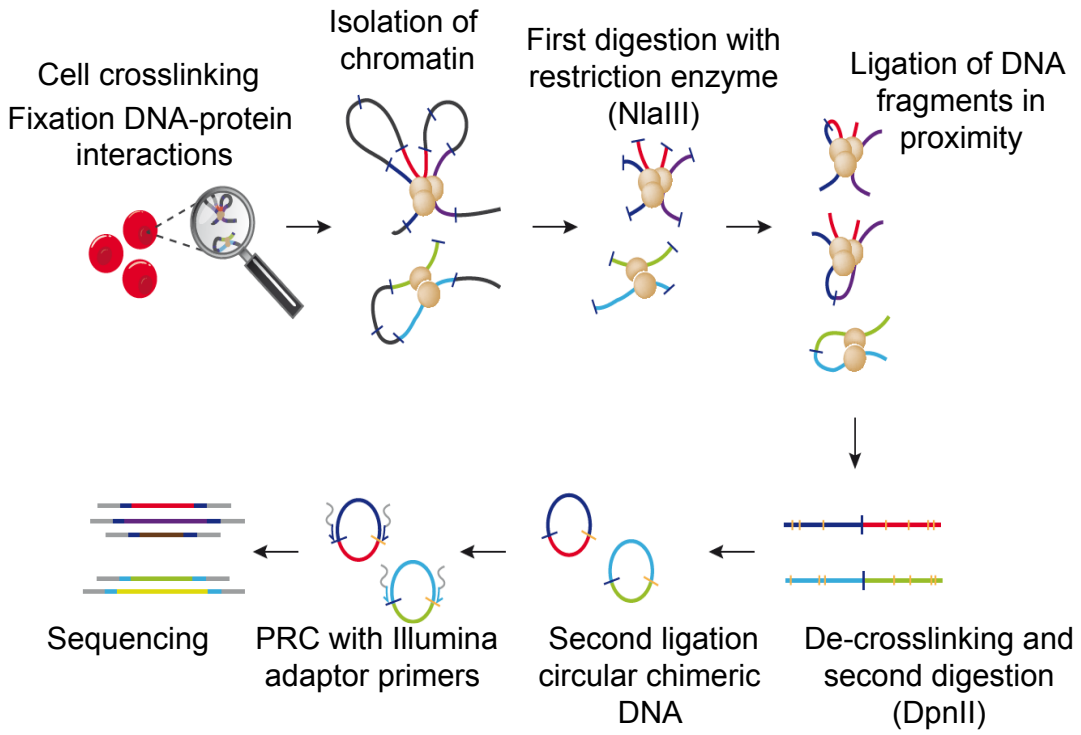

Figure 1.2: 4C-seq methodology (adapted from Stadhouders et al. 2013). Cells are crosslinked in order to fix the interactions between DNA and proteins. After nuclei isolation, chromatin is digested and then ligated in order to evaluate the contacts of DNA fragments that are in proximity. Following reverse crosslinking, a second round of digestion-ligation is performed. Finally chimeric circular DNA molecules are obtained representing the contacts between different loci which can be amplified by inverse PCR using primers located within a locus of interest (bait or viewpoint) and subsequently sequenced (see methods 3.7).

Enhancers were originally characterized using transient reporter gene assays in cultured cell lines, in which enhancer activity is tested in an exogenous and artificial genomic context (Banerji, Rusconi, and Schaffner 1981; Bulger and Groudine 2011). Since then, new methodologies, such as CRISPR-Cas9, have arised to characterize and determine the function of enhancers within their endogenous genetic and chromatin context. Briefly, this technique consists of gRNA molecules that guide the Cas9 nuclease to specific loci to introduce double-stranded DNA breaks at the desired location. Consequently, these DNA breaks can be repaired by non-homologous end-joining repair (NHEJ), which can result in the deletion of the sequence of interest (e.g. enhancer), whose function can then be evaluated.

1.2.2- Enhancer states

These epigenomic methods also revealed that enhancers can exist in various regulatory states (e.g. active, primed, poised), which can be distinguished based on unique chromatin signatures and that display distinct regulatory properties (Table 1.1) (Calo and Wysocka 2013; Menno P. Creyghton et al. 2010; Rada-Iglesias et al. 2011; Zentner et al. 2011).

Active enhancers present low nucleosomal density, are bound by TFs and co-activators (e.g. p300/CBP, CHD7, BRG1), are flanked by nucleosomes marked with histone three lysine four monomethylation (H3K4me1) and histone three lysine 27 acetylation (H3K27ac) and they are associated with highly expressed genes (Menno P. Creyghton et al. 2010; Rada-Iglesias et al. 2011; Zentner et al. 2011).

Primed enhancers are marked with H3K4me1 but not H3K27ac. They are linked to genes that exhibit moderate expression levels and are implicated in a broad range of biological processes (Menno P. Creyghton et al. 2010; Zentner, et al. 2011).

Poised enhancers (PEs) were defined originally in human and mouse embryonic stem cells (hESC, mESC) as a small group of highly conserved regulatory elements. Like active enhancers, they are also bound by TFs and co-activators (p300/CBP, CHD7, BRG1) and display low nucleosomal density. However, in contrast to active enhancers, they are not marked with H3K27ac but with histone three lysine 27 trimethylation (H3K27me3) instead, a repressive histone modification mediated by Polycomb Repressive Complex two (PRC2) (Rada-Iglesias et al. 2011; Zentner et al. 2011). Importantly, PEs are in proximity of genes with major cellular identity and developmental functions that are inactive in ESC and become expressed upon somatic differentiation. Since PEs are already bound in ESC by co-activators like P300, it has been proposed that they represent developmental enhancers that are bookmarked for their future activation once the appropriate differentiation signals are received (Rada-Iglesias et al. 2011; Spitz and Furlong 2012).

Additionally, it has also been proposed that the presence of H3K27me3 and PRC2 might keep PEs in an inactive state, preventing their spurious activation until the relevant differentiation cues become available (Tie et al. 2016).

Table 1.1: Types of enhancers and their epigenetic features

Properties	Active enhancers	Primed enhancers	Poised enhancers
DNAse I sensitivity	Yes	Not reported	Yes
H3K4me1	Yes	Yes	Yes
H3K27ac	Yes	No	In differentiated cells
H3K27me3	No	No	In ESC
Activator TFs	Yes	No	Yes
Associated genes	High expressed genes	Non specific	Developmental regulators

1.2.3- Mechanism of action of enhancers

Since the discovery of enhancers, the dominant model to explain their mechanism of action involves the physical interaction of enhancers with their target promoters through the formation of chromatin loops. This “looping model” (Bulger and Groudine 2011; Calo and Wysocka 2013; de Laat and Duboule 2013; Vernimmen and Bickmore 2015) has received extensive experimental support from chromosome conformation capture related technologies (e.g. 4C-seq, Hi-C, ChIA-PET) that enable the detection of physical interactions between distally located loci. Interestingly, most enhancer-promoter interactions occur within Topologically Associated Domains (TADs), which represent self-interacting Mb-scale domains that physically restrain enhancer activity to genes located within the same TAD (Figure 1.3) (Dixon et al. 2012). Although enhancer-promoter communication was originally proposed to occur once the target genes become active (Vernimmen et al.

2007), recent observations indicate that enhancer-promoter contacts can precede gene activation (Ghavi-Helm et al. 2014; Jin et al. 2013). At least in some cases, these pre-formed contacts might confer a permissive chromatin topology that poises genes for future induction (de Laat and Duboule 2013).

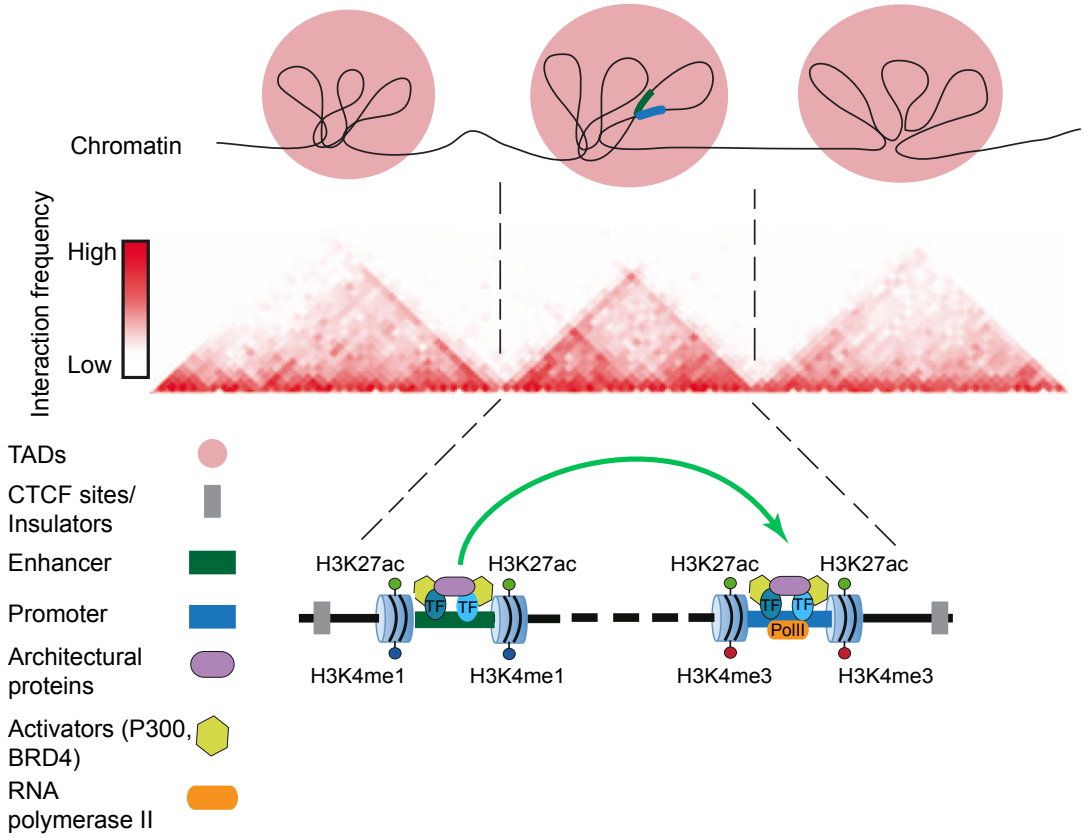

Figure 1.3: Chromatin landscape and mechanism of action of enhancers. The three-dimensional nuclear configuration of mammalian genomes is organized in topological associated domains (TADs). A TAD represents a region of DNA where physical interactions occur relatively frequently, whereas interactions across a TAD boundary occur relatively infrequently. Enhancer-promoter interactions occur within the same TAD, and they are facilitated by the numerous TFs, activators and architectural proteins that are associated to these regions. Figure adapted from (Dixon et al. 2012).

The activation of gene promoters requires that many transcriptional components come together to assemble the pre-initiation complex (PIC), initiate transcription, overcome the paused RNA polymerase II (PolII) and eventually drive transcriptional elongation. By forming loops, enhancers come in close proximity with their target promoters and facilitate the aforementioned process by increasing the concentration of TFs and co-activators around the

promoter regions. Some of these factors include architectural proteins such as cohesin as well as mediator complexes and other factors (e.g. P300, BRD4) involved in releasing paused PolII and initiation of transcriptional elongation (Heinz et al. 2015) (Figure 1.3). Interestingly, previous studies have demonstrated that removal of the promoter does not affect the recruitment of the PolII to the enhancer region, while the enhancer deletion does affect the binding of PolII to the promoter (Vernimmen et al. 2007). It is worth mention that enhancers are also transcribed (enhancer RNAs or eRNAs), although at low levels, and their expression correlates positively with the mRNA levels at their target genes (Vernimmen and Bickmore 2015).

It has been described that PEs already communicate with their putative target genes in undifferentiated ESC (Schoenfelder et al. 2015). However, the molecular basis and functional relevance of these preformed PE-promoter contacts is currently unknown. Similarly, it is not clear yet if and how PEs control the expression of their target genes once activated.

1.2.4- Poised enhancers during pluripotency and differentiation

Previous reports (M. P. Creyghton et al. 2010; Rada-Iglesias et al. 2011) have determined that poised enhancers are developmental regulatory elements that are found in a silent but primed state in undifferentiated ESC. As mentioned above, one major distinctive feature of PEs is that they are bound by PRC2 and, hence, marked with H3K27me3. Upon ESC differentiation, PEs get activated, lose H3K27me3 and gain H3K27ac (Rada-Iglesias et al. 2011; Zentner et al. 2011). These two histone modifications are mutually exclusive, implying that, despite being bound by co-activators with histone acetyl-transferase (HAT) activity (e.g. P300, CBP) (Rada-Iglesias et al. 2011), the presence of H3K27me3 at PEs might avoid the accumulation of H3K27ac and the spurious activation of these regulatory elements (Pasini et al. 2010; Reynolds et al. 2012; Tie et al. 2016). Interestingly, this data is in concordance with the classically view of Polycomb group proteins as major epigenetic repressors playing important roles in differentiation and maintenance of cellular identity (Di Croce and Helin 2013). Additionally, some studies reported that PEs were also found in non-pluripotent cells and therefore, their relevance was

questioned since they were proposed to display a repressed rather than poised state or to even act as silencers (Bonn et al. 2012; Entrevan et al. 2016; Kondo et al. 2016; Spitz and Furlong 2012). However, in these and other studies the identification of PEs in ESC was based on the presence of H3K4me1 and/or H3K27me3 but the binding of TFs/co-activators was not required (Bonn et al. 2012; Menno P. Creyghton et al. 2010; Schoenfelder et al. 2015).

Due to the association of PEs with major cell identity and developmental regulators, it has been previously suggested that PEs might play important roles during the execution of somatic differentiation programs (Heintzman et al. 2009; Bulger and Groudine 2011; Rada-Iglesias et al. 2012). However, we still do not know which mechanisms lead to lineage-specific activation of poised enhancers. Most importantly, the functional relevance of PEs is only supported by correlative observations, including reporter assays in zebrafish embryos that demonstrated that human poised enhancers sequences can act as developmental enhancers *in vivo* (Rada-Iglesias et al. 2011).

In conclusion, the importance of PEs for the induction of their cognate genes and the execution of somatic differentiation programs has not been formally demonstrated. Correspondingly, it is also unknown if and how the PEs chromatin features actually facilitate the future activation of the PE target genes. (Rada-Iglesias et al. 2011; Zentner et al. 2011). Therefore, additional research is needed in order to elucidate these important questions.

1.3- Polycomb group proteins

Polycomb group proteins (PcG) are a large, conserved and diverse family of epigenetic regulators of transcription. Historically they have been considered as repressors of homeotic (HOX) genes, but more recently it has been revealed that they have broader roles in stem-cell identity, differentiation and disease (Chittock et al. 2017; Di Croce and Helin 2013; Pasini et al. 2010; Schoenfelder et al. 2015).

PcG proteins are part of transcriptional-repressive complexes, termed Polycomb Repressive Complexes (PRCs). Two major PRCs have been identified in most metazoan species, PRC1 and PRC2. In mammals, the canonical PRC1 contains two core subunits, RING1 which monoubiquitinates histone H2A on lysine 119 (H2AK119ub) and PcG ring finger protein (PCGF), which interacts with specific binding partners. PRC1 also contains a chromobox protein (CBX) which binds the H3K27me3 mark and a Polyhomeotic (PH) homolog protein (PHC). PHC proteins might play an important role in higher order chromatin organization by forming long-range contacts between remote PRC-bound sites in the genome (Figure 1.4) (Chittock et al. 2017; Di Croce and Helin 2013; Entrevan et al.2016).

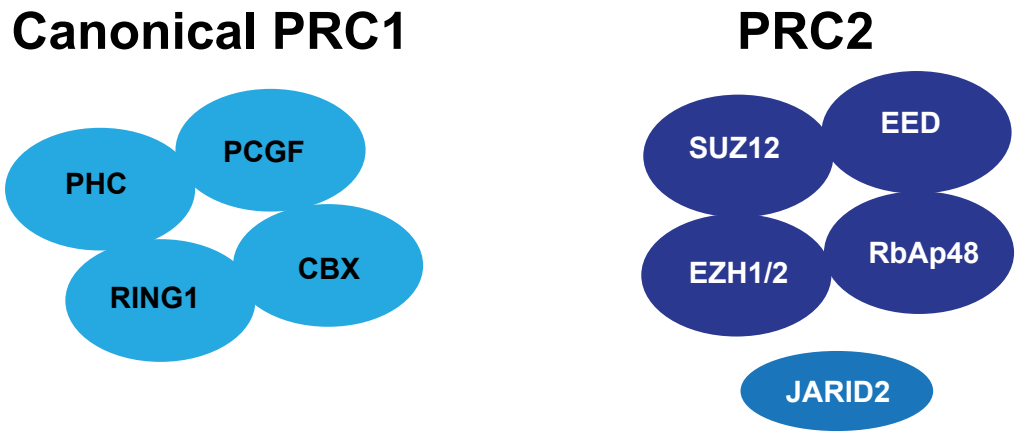

Figure 1.4: Composition of Polycomb group proteins. PcG consist of two main complexes, PRC1 and PRC2. PRC1 contains RING1, PHC, CBX and PCGF subunits (light-blue) and is responsible of ubiquitinilation of H2AK119, recognition of H3K27me3, chromating compactation and PolII inhibition. PRC2 consist of Suz12, EED and EZH the core subunit responsible of the methylation of H3K27. In addition, PRC2 also contains RbAP48, a protein which stabilizes the whole complex and Jarid2, a cofactor that facilitates the recruitment of PRC2 to its genomic targets (Entrevan et al. 2016; Blackledge, Rose, and Klose 2015; Di Croce and Helin 2013; Sanulli et al. 2015; Chittock et al. 2017).

The PRC2 core complex has three main proteins: EZH1/2, EED and SUZ12 The SET (enhancer of zeste) domain-containing proteins responsible of methylation of H3K27, EZH1 and EZH2, are mutually exclusive and differentially expressed in proliferating and non dividing tissues (Entrevan et al. 2016). EED (Embryonic ectoderm development) binds to H3K27me3 and contributes to the propagation of this repressive mark. SUZ12 (suppressor of zeste) is also required for the histone methyltransferase catalytic activity of the PRC2 complex. In addition, PRC2 also contains other subunits such as RbAp48, which stabilizes the complex, and JARID2, a co-factor which recruits PRC2 to its genomic targets (Figure 1.4) (Chittock et al. 2017; Di Croce and Helin 2013; Entrevan et al. 2016; Peng et al. 2009).

1.3.1- Role of PcG in gene regulation

Polycomb proteins are present at repressed genes with developmental functions. PcG repressive function entails chromatin compaction, mediated by PRC1, block of HATs (Histone Acetyltransferases) and interference with RNA Polymerase two (PolII) activity. PRC1 chromatin compaction activity involves the ligase activity of RING1 (Bantignies et al. 2011; Francis 2004) as well as the function of other PRC1 canonical subunits (e.g. CBX, PH) which can lead to the formation of subnuclear compartments, known as Polycomb bodies (Creppe et al. 2014; Schoenfelder et al. 2015, 1; Isono et al. 2005).

On the other hand, PcG binding at gene promoters can interfere with transcription by “holding” PolII at the transcription start sites (TSS). This results in a “paused” PolII, unable to elongate. In agreement with this function, the deletion of RING1 results in PolII phosphorylation at serine2 (S2) and increased transcriptional elongation (Stock et al. 2007) Moreover, this PolII pausing function of PcG complexes might also explain the existence of so called bivalent promoters, which are enriched in H3K27me3 and histone three lysine four (H3K4me3). These bivalent promoters are typical of developmental

genes and particularly abundant in ESC where they are proposed to bookmark genes for subsequent activation upon differentiation.

1.3.2- PcG in pluripotency and differentiation

PcG proteins are essential for embryonic development and, mice lacking EZH2, EED or SUZ12 die soon after implantation (Faust et al. 1998; O'Carroll et al. 2001; Pasini et al. 2004). On the other hand, PRC2 activity is not required for mESCs self renewal although it appears necessary for proper differentiation. Nevertheless, $EED^{-/-}$ mESCs, which display very low PRC2 activity and H3K27me3 levels, are able to contribute to all embryonic tissues when injected into mouse blastocyst, indicating that PRC2 null mESCs are to some extent pluripotent (Chamberlain et al. 2008).

On the contrary, the role of PRC1 in ESC self-renewal and differentiation is not as clear due to the large variety of PRC1 complexes and associated subunits. It has been reported that, with the exception of RING1B, the deletion of most PRC1 subunits leads to late developmental defects during mouse embryogenesis. However, embryos with deletion of at least two PRC1 subunits (two different ring finger subunits (Akasaka et al. 2001) or ZFP144 ring finger protein and RING1B (Isono et al. 2005)) failed to pass midgestation.

Taken together, the data suggest that *in vitro*, PcG proteins are not required for mESC self renewal but can affect pluripotency and differentiation potential (Riising et al. 2014).

1.3.3- Recruitment of PcG

Previous studies reported that the recruitment of canonical PRC1 complexes depends on the presence of the H3K27me3 mark deposited by PRC2, which is recognized by the chromodomain of the CBX subunits (Min et al. 2003). On the other hand, other studies described that PRC1 complexes are targeted to CpG islands (CGIs) by the KDM2B histone demethylase. PRC1 then catalyzes H2A ubiquitination, which in turn can recruit PRC2 (Blackledge et al. 2014). Although PcG recruitment mechanisms are not fully understood, CGIs have

recently emerged as potential Polycomb response elements (PREs) in vertebrates, similar to those previously described in Drosophila (Entrevan et al. 2016). Different mechanisms have been suggested for the recruitment of PRC2 to CGIs. JARID2, a PRC2 co-factor, has a DNA binding domain that preferentially binds to CpG-rich sequences and could thus help in the recruitment of PRC2 to CpG islands (CGIs). Accordingly, the loss of JARID2 leads to a decrease in PRC2 binding to their target genes (Di Croce and Helin 2013; Peng et al. 2009). More recently, non-canonical PRC1 complexes containing RYBP (RING1A and YY1 binding protein) and KDM2B were shown to bind to CGIs in the absence of H3K27me3 (van den Boom et al. 2016). For instance, the histone lysine demethylase KDM2B, has been reported to specifically recognize non-methylated DNA in CGIs and recruit PRC1 (Farcas et al. 2012). RYBP and its mouse homolog YAF2 are responsible for PcG recruitment to DNA, which is mediated by YY1 TF DNA binding (Basu et al. 2014).

Recent studies indicate that the transcriptional status at CGIs can determine whether PcG complexes are recruited to their target genes, since global inhibition of transcription lead to ectopic recruitment of PcG proteins to all CGIs (Riising et al. 2014). In addition to CGI, sequence-specific TFs like SNAIL, REST or PLFZ can regulate the recruitment of PRC2 to their target genes although this does not seem to be a general mechanism for PcG recruitment, and TFs may only transiently associate with PcG (Dietrich et al. 2012). Lastly, several studies propose that noncoding RNAs have a role in gene silencing and PcG-protein recruitment. RNA-immunoprecipitation techniques have identified several thousand of RNAs associated with PRC2 and furthermore, it has been demonstrated that the long noncoding Xist transcript can interact with PRC2 during the inactivation of the X chromosome (Plath et al. 2003). However, most recently, RNA binding to PcG proteins has been suggested to facilitate the eviction rather than the recruitment of these complexes to their chromatin targets (Aranda et al. 2015).

In summary, different models of recruitment of PcG proteins have been proposed, although most likely several of the suggested mechanisms might act

in a combinatorial manner to confer robustness to the binding of these important epigenetic repressors to their cognate genomic targets.

1.3.4- PcG as major organizers of nuclear architecture

It is known that cell fate transitions are accompanied by alterations in chromatin structure. The action of PcG on chromatin leads not only to compaction, but also to specific 3D chromatin folding (Bantignies et al. 2011). Due to the appearance of novel approaches to study genome nuclear organization, such as 3C and derivative techniques (e.g. 4C-seq, Hi-C), it is now possible to elucidate the topological organization of chromatin within the nucleus.

PcG are starting to emerge as major organizers of nuclear architecture. For example, it has been demonstrated that the repressive chromatin hubs found at the homeotic bithorax complex (BX-C) locus in *Drosophila* are composed of multiple chromatin loops. In these loops, all major interacting elements, including core promoters and PREs, are bound by PcG. As a result, a topologically complex structure referred to as a PcG body is formed and required for the silencing of BX-C (Lanzuolo et al. 2007). Furthermore, it has been also shown that switches in the BX-C transcriptional status are accompanied by major rearrangements in the high order chromatin structures (Lanzuolo et al. 2007). On the other hand, in mECSs it has been revealed that EED, one of the PRC2 subunits, is required for the maintenance of interactions between PcG regions separated by tens of megabases or even located on different chromosomes (Denholtz et al. 2013). These interactions were reduced when mESCs were grown under naïve pluripotency conditions, most likely due the lower genomic levels of PcG complexes and associated histone marks (e.g. H3K27me3) present under those conditions (Tee et al. 2014). Similarly, it has been recently reported that RING1 depletion in mESC leads to the lost of long-range interactions between PcG target genes, while interactions between active pluripotency genes remained unaffected (Schoenfelder et al. 2015). Moreover, PcG-bound promoters also interacted with poised enhancers marked by H3K4me1 and H3K27me3. Remarkably,

loss of RING1 led to activation of poised enhancers, whereas poised enhancer–promoter contacts were overall maintained. Lastly, it has been also proposed that during mouse brain development, the Meis2 gene promoter is initially associated with a silencing element via RING1B. Interestingly, this PcG-bound element topologicaly facilitates the subsequent interaction between Meis2 and a specific enhancer, ultimately leading to gene activation. When RING1B is not present, the enhancer can no longer contact the promoter region, resulting in reduced Meis2 expression (Kondo et al. 2014).

Since TADs were discovered, research efforts have been invested in understanding the link between chromatin 3D structure and gene expression. In Drosophila, TADs can be broadly classified into active, inactive and PcG-repressed states. PcG TADs are the most compact, show the highest inter-TAD interactions and display minimal contacts with other domains. On the other hand, active and transcriptionally inactive TADs show higher levels of inter-domain mixing. It has been illustrated that PH, a component of Polycomb PRC1, plays an important role in these interactions and without PH, PcG TADs are lost (Boettiger et al. 2016). In agrrement with this study, it was shown that the disruption of the polymerization activity of the PH sterile alpha motif (SAM) domain leads to a dispersion of PcG clusters and chromatin interactions. Moreover, similar findings were recently reported in mouse cells (Isono et al. 2013). Overall, these recent reports indicate that PH SAM domains may mediate the organization of PcG proteins into Polycomb bodies (Figure 1.5) (Wani et al. 2016).

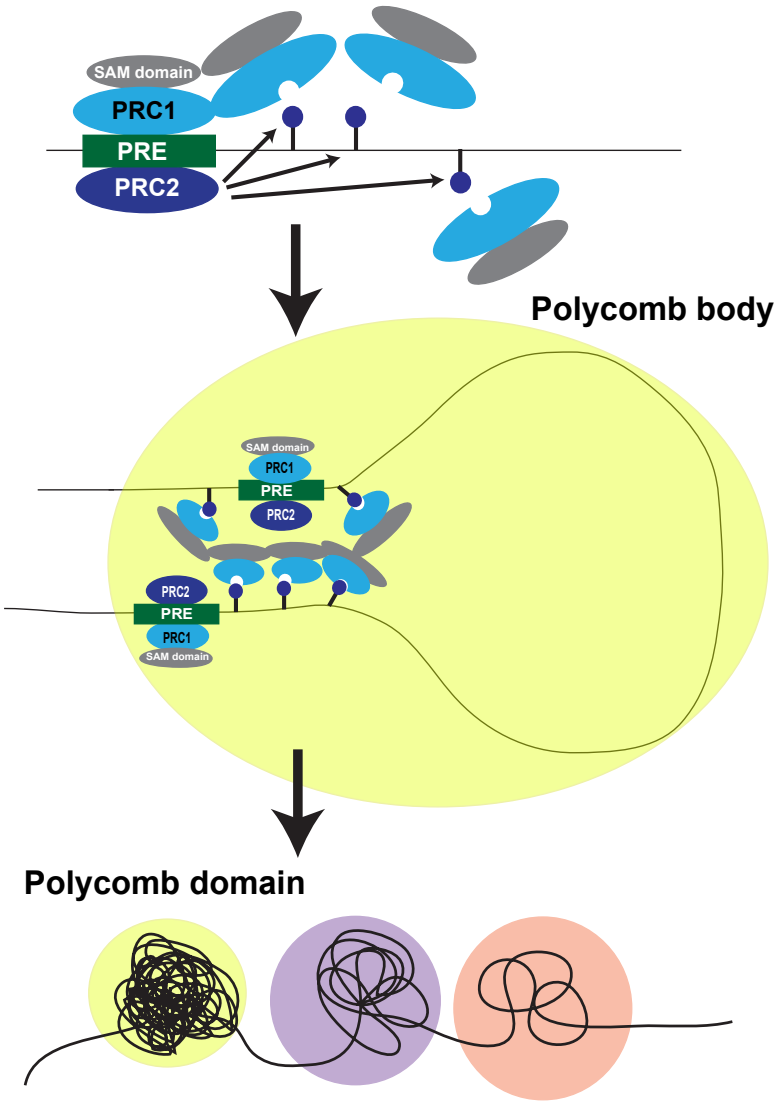

Figure 1.5: Proposed model for Polyhomeotic (PH) SAM domains in the formation of Polycomb bodies. PcG complexes are recruited to their target sites via PREs and PcG polymerization occurs through SAM-mediated interactions. PcG-mediated looping might involve the recognition of H3K27me3 (dark blue lollipop) by PRC1 and the PRC1 polymerization mediated by PH SAM domains. These PRC1/PRC2 clusters correspond to microscopically visible subnuclear domains known as Polycomb bodies. These Polycomb bodies or Polycomb domains (yellow circle) are significantly more compacted than transcriptionally inactive (purple) or active (pink) domains. Adapted from (Entrevan et al. 2016).

The role of PRC2 in regulating global nuclear architecture is still not very clear. As it was mentioned previously, the loss of EED in mESCs leads to reduced interactions between Polycomb-bound regions (Denholtz et al. 2013; Joshi et al. 2015). However, it was also reported that the loss of PRC2 did not affect TAD boundaries either globally or across the X chromosome (Nora et al. 2012). Nevertheless, it has been demonstrated that induced recruitment of

EZH2 to a particular locus can lead to relocation into a new nuclear compartment occupied by other Polycomb-bound regions. This study suggests that specific nuclear subcompartments, including PcG domains, are formed as a consequence of affinities between chromatin-associated proteins or modifications (Wijchers et al. 2016).

In summary, PcG proteins can mediate looping interactions between regulatory elements, define global nuclear architecture and ultimately regulate gene expression. However, the role of PcG upon PEs activation and lineage commitment needs to be further characterized.

2- AIM

The knowledge about the establishment of gene expression programs during mammalian developmental transitions is far from being completed. We hypothesize that poised enhancers are important regulatory elements that, in ESCs, display a silent but primed state sustained by PcG. Moreover, we also hypothesize that, once activated, poised enhancers might be important for the induction of their target genes during the earliest steps of somatic differentiation. Therefore, the major goals of this project are to examine the functional and developmental relevance of poised enhancers and to investigate how PcG influences poised enhancer's regulatory activity.

In order to fulfill these objectives, we first characterized poised enhancer candidates with well-defined activation dynamics during mESCs differentiation into anterior neural progenitors. Secondly, we identifed the putative target genes of poised enhancers based on their physical interactions. Then, we engineered poised enhancer deletions to assess the functional importance of these regulatory elements during mESCs differentiation. Lastly, we examined the role of PcG, specially the PRC2 complex as a regulator of poised enhancers activity and function.

3. MATERIAL AND METHODS

3.1 – Equipment

Table 3.1: Equipment

Equipment	Company	Catalog number
Bacteria Incubator	Infors HT Ecotron	s00120638
Cell culture Incubator	Sanyo	8070263
Microscopes		
Inverted microscope	Leica DMILED	376977
Fluorescent microscope	Olympus IX2-UCB	9A01123
Microscope ECLIPSE TS100	Nikon	Kurian's Lab
Electrophoresis		
Chambers	Biorad	1704502 and 1704406
Power supply	Biorad	041BR110323
Western Blot		
Chamber	Biorad	1658004
Power supply	Biorad	043BR500041
Centrifuges		
Cell culture	Hermle	311001101
Big 4° for falcon tubes	Hermle	31130026
Small 4° for eppendorf tubes	Hermle	61150064
For bacterial liquid culture	Labscience	7.601.314.101
Cell culture bath	Memmert	325741
Tissue Culture hood	Kojair	22198
Bioruptor plus	Diagenode	Brng130125
PRC Thermal cycler		
q-PCR LightCycler 480 II	Roche	5662
c1000 Touch	Biorad	ct024292
Electic balance	Kern	WB1251009
Regular balance	Kern	WB1300143
Spectrophotometer Nanodrop	Thermo Scientific	F673
pH meter	Sartorius	29053099
Automated cell counter	Biorad	508BR05586
Shaker	Skyline	12DE117
Thermo Block	Ditabis	980052301
ChemiDoc MP	Biorad	731br01726

3.2- Chemicals and reagents

3.2.1- Chemicals

Table 3.2: Chemicals

Reagent	Company	Catalog number
2-Propanol	Roth	9866.5
25:24:1 Phenol- Chlorophorm Isoamyl Alcohol	Sigma Aldrich	P2069
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Roth	HN78.2
30% Acrylamide/ Bis solution	Sigma Aldrich	161-0156
Adenosine triphosphate (ATP)	Sigma Aldrich	A2383
Agarose	Life Technologies	16500100
Bovine Serum Albumina (BSA)	Roth, Germany	3737.2
Bromophenol Blue	Sigma Aldrich	B0126
Chloroform	Sigma Aldrich	366919
DAPI	Sigma Aldrich	D9542-5MG
Dithiothreitol (DTT)	Roth	6908.3
Ethanol	Roth	5054.3
ethylendiaminetetraacetic acid (EDTA)	Roth	8043.1
ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid) EGTA	Roth	3054.2
Formaldehide solution 37%	Sigma Aldrich	252549
Glycerin	Sigma Aldrich	68898
Gycerol	Roth	3783.2
Hydrochloric acid (HCl)	Roth	281.1
Litium Chloride (LiCl)	Roth	3739.2
Magnesium Chloride (MgCl2)	Roth	KK36.2
Methanol	Sigma Aldrich	494437
Mounting medium	Southern Biotech	0100-01
N-Lauroylsarcosine	Sigma Aldrich	61743
Na-Deoxycholate	Sigma Aldrich	D678
Non-Fat Milk	Hartestein	CM35
NP-40	Sigma Aldrich	I3021
Parafolmaldehide	Sigma Aldrich	158127
Phosphate buffered saline (PBS)	Sigma Aldrich	D8537
Potasium Chloride (KCl)	Roth	HN02.3
Sodium Acetate (C2H3NaO2)	Roth	6773.2
Sodium Chloride (NaCl)	Roth	3957.2

Sodium dodecyl sulfate (SDS)	Roth	183.3
Triton X-100	Roth	3051.4
Trizma Base	Sigma Aldrich	T1503
Trizol (Tripure)	Roche	11667157001
Tween-20	Roth	9127.2

3.2.2- Buffers and solutions

Table 3.3: Buffers

Buffer/ Solution	Composition
Blocking Solution ChIP	PBS (1x) 0.5% BSA (w/v)
Lysis Buffer 1 (LB1)	50 mM Hepes (pH 7,5) 140 mM NaCl 1 mM EDTA 10% glycerol 0.5% NP-40 0.25% TX-100 dH_20
Lysis Buffer 2 (LB2)	10 mM Tris pH 8 200 mM NaCl 1 mM EDTA 0.5 mM EGTA dH_20
Lysis Buffer 3 (LB3)	10 mM Tris pH 8 100 mM NaCl 1 mM EDTA 0.5 mM EGTA 0.1% Na-Deoxycholate 0.5% N-lauroylsarcosine dH_20
RIPA Wash Buffer for ChIP	50 mM HepespH 7,5 500 mM LiCl 1 mM EDTA 1% NP-40 0.7% Na-Deoxycholate dH_2O
Elution Buffer	50 mM Tris pH 8 10 mM EDTA 1% SDS dH_2O

Dilution Buffer (Sequential ChIP)	20 mMTris-HCl pH 8 2 mMEDTA 1% Triton X-100 150 mM NaCl dH_2O
RIPA Wash Buffer for Western Blot	50 mM Tris, pH 8.0 150 mM NaCl 1.0 % NP-40 0.5% sodium deoxycholate 0.1% SDS dH_2O
Laemmli buffer	375 mM Tris pH 6.8 60% glycerol 600 mM DTT 0.06% bromphenol blue 12% SDS dH_2O
Running buffer	25 mM Tris pH 8,3 190 mM glycine 0.1% SDS dH_2O
Transfer buffer	25mM Tris 190 mM glycine 20% Methanol 0.1 % SDS dH_2O
PBST washing buffer	PBS 1% Tween
Blocking Solution for Western Blot	5% milk 1% PBST
Permeabilizing solution IF	0.3% Triton X-100 dH_2O
Lysis Buffer for 4C	50 mM Tris–HCl pH 7.5 150 mM NaCl 5 mM EDTA 0.5% NP-40 1% TX-100 1X protease inhibitors dH_2O
Ligation Buffer for 4C	50 mM Tris-HCl pH 7.6 10 mM MgCl2 1 mM ATP 1mM DTT dH_2O

TE Buffer	10 mM Tris pH 8.0 1 mM EDTA dH_2O
TAE buffer (1X)	40mM Tris pH 8.6 20mM Acetate 1mM EDTA dH_2O

3.2.3- Kits and commercial assays

Table 3.4: Kits

Kits	Company	Catalog number
Innuprep RNA mini kit	Analytic Jena	845-KS-2040250
ProtoScript II First Strand cDNA Synthesis Kit	NEB	E6560L
PCR purification column	QIAgen	28104
SuperScript® VILO™ cDNA Synthesis Kit	Thermo Fisher	11754-050
Nucleospin Plasmid MiniPrep, 250 rxn	Macherey-Nagel	740588-250

3.2.4 – Cell culture reagents

Table 3.5: Cell culture medium

Medium	Components	Company	Catalog number
Serum +LIF	15% Fetal Bovine Serum (FBS)	Life Technologies	16141-061
	1 % Antimycotic/antibiotic	Hyclone	SV30079.01
	0.02% Beta-mercaptoethanol 55 mM	Life Technologies	21985-023
	1 %Glutamax	Life Technologies	35050-038
	1 % MEM NEAA	Life Technologies	11140-035
	Knock-out DMEM	Life Technologies	10829-018
	LIF	N/A	N/A
Freezing medium Serum plus LIF (2x)	40% Serum plus LIF medium		
	40% FBS	Life Technologies	16141-061
	10% DMSO	Sigma Aldrich	D2650
N2B27	50% DMEM/F-12	Invitrogen	21041-025
	0.5 % N2 suplement	Invitrogen	17502-048

	25 µg/m Insulin	Sigma Aldrich	I-1882
	1% penicillin-streptomycin	Invitrogen	15070
	2 mM l-glutamine	Invitrogen	25030-081
	50% Neurobasal medium	Invitrogen	12348-017
	1 % B27 suplement	Invitrogen	12587-010
	0.1 mM β-mercaptoethanol	Sigma Aldrich	M6250
2i medium	N2B27 medium		
	3 µM CHIR 99021	Amsbio	1677-5
	0.4 µM PD 325901	Miltenyi Biotech	130-103-923
	LIF	Housemade	
	0.3 % BSA	Invitrogen	15260-037
Freezing medium 2i (1x)	50% N2B27		
	40% FBS	Invitrogen	16141-061
	10% DMSO	Sigma Aldrich	D2650
Tryple Wash medium	DMEM/F-12	Invitrogen	21041-025
	10 % BSA	Invitrogen	15260-037
Mesodermal precursors basal medium	75% IMDM	Thermo Scientific	21980032
	25% Ham's F12	Thermo Scientific	11765054
	2mM GlutaMax	Life Technologies	35050-038
	0.5X B27 supplement	Invitrogen	12587-010
	0.5X N2 supplement	Invitrogen	17502-048
	50 µg/ml Ascorbic acid	Pan Biotech	P04-0070K
	4.5x10-4 M Monothioglycerol	Sigma Aldrich	M6145
	0.05 % BSA	Invitrogen	15260-037
Mesodermal precursors differentiation medium	Basal mesodermal medium		
	5 ng/ml human VEGF	Pan Biotech	P04-0070K
	4 ng/ml human Activin A	Peprotech	120-14-10
	0.2 ng/ml human BMP4	Miltenyi Biotech	130-098-786

Table 3.6: Cell culture reagents

Reagent	Company	Catalog number
Trypsine	Thermo Fisher	25200072
Tryple Express	Thermo Fisher	12604-021
Xav939	Sigma Aldrich	x3004
Puromycin	Thermo Fisher	A2856,0025

Optimem medium for transfection	Thermo Fisher	51985026
X-tremeGene transfection reagent	Roche	6366236001
b-FGF	Peprotech	100-18B
Neomycin	AppliChem	A6798,0020

3.2.5 – Molecular biology reagents

Table 3.7: Molecular biology reagents

Reagent	Company	Catalog number
Proteinase K	Sigma Aldrich	P2308
RNAse A	Peqlab	12-RA-03
Turbo DNAse	Thermo Fischer	AM1907M
Proteinase Inhibitors	Roche	5892953001
Glycogen	Peqlab	37-1810
Lumi-Light Western Blotting Substrate	Roche	1.2015196E
QuickExtract DNA Extraction Solution	Biozym	QE09050
TRIzol® Reagent /tripure	Roche	11667157001
dNTPS	Promega	U1515
SYBR® Safe DNA gel stain	Invitrogen	S33102
Dynabeads Protein G	Thermo Fisher	1004D
Protease inhibitor cocktail	Roche	05892953001
GeneRuler 1 kb Plus DNA Ladder, 75-20,000 bp	Peqlab	25-2240

Table 3.8: Enzymes

Enzyme	Company	Catalog number
NlaIII	NEB	R0125L
DpnII	NEB	R0543M
BsaI	NEB	R0539L
T4-Ligase (1U/µl)	Invitrogen	15224-041
T4-Ligase	NEB	M0202M
Expand long template PCR system	Roche	11681842001
Platinum Taq polymerase	Life technologies	10966034
ORA qPCR Green ROX L Mix	HighQu	QPD0105
One Taq DNA polymerase	NEB	M0509S

3.2.6 – Bacterial culture reagents

Table 3.9: Bacteria strains

Bacteria strains	Source	Identifier
CopyCutter EPI40 Electrocompetent *E. coli*	Epicentre Biotechnologies	C400EL10
Top10 *E.Coli*	Kurian Lab	N/A

Table 3.10: Plasmids

Plasmid	Source	Identifier
PB enhancer GFP Neo	Buecker et al., 2014	N/A
pX330-U6-Chimeric_BB-CBh-hSpCas9	Addgene	42230
pX330-hCas9-long-chimeric-grna-g2p	Kurian Lab	N/A
Super PiggyBac Transposase	System Biosciences	PB210PA-1

3.3 -Cell culture procedures

Cell culture work was performed under sterile conditions in order to avoid contamination. Sterile conditions were guaranteed by laminar flow cell culture hoods as well as sterile solutions, materials and medium supplemented with antibiotics and antimycotics. Cells were kept in an incubator at 37° in a humid atmosphere containing five per cent CO2.

3.3.1- Cell lines

For this study, we used three different types of mESCs lines and we generated seven PE knockout lines:

Table 3.11: mESC lines used in culture

Cell line	Reference
WT (E14) mESC	Wysocka Lab
$EED^{-/-}$ mESC	Schoeftner et al., 2006
$SUZ12^{-/-}$ mESC	Pasini et al., 2007
PE Lhx5 $(-109)^{-/-}$	N/A
PE Six3 $(-133)^{+/-}$	N/A
PE Six3 $(-133)^{-/-}$	N/A

PE Wnt8b (+21)$^{+/-}$	N/A
PE Sox1 (+35)$^{-/-}$	N/A
PE Sox21 (+3,5)$^{-/-}$	N/A

3.3.2- Culture of mESCs in serum plus LIF conditions

WT (E14) mESC, EED-/- mESC (Schoeftner et al. 2006) and mESC lines with PE deletions were grown on gelatin-coated plates using Knock-out DMEM (KO-DMEM, Life Technologies) supplemented with 15% FBS (Life Technologies) and LIF in order to promote a primed pluripotency state (Table 3.5).

3.3.3- Culture of mESCs in "2i" conditions

For inducing the "naïve ground state of pluripotency" (see introduction 1.1), WT (E14) mESC were grown in 2i plus LIF medium, as it has been previously reported (Respuela et al. 2016). To reach this state, mESCs growing in serum plus LIF were disaggregated using Tryple Express and washed with Tryple wash medium (Table 3.5). Afterwards, cells were splitted 1:8-1:10 and cultured in N2B27 medium supplemented with MEK inhibitor PD0325901 at 0.4 μM, GSK3 inhibitor CHIR99021 at 3 μM, and LIF in cell culture dishes pre-treated with gelatin during at least four passages.

3.3.4- Differentiation into Anterior Neural Progenitors (AntNPCs)

E14 mESC were plated at a density of 10000 cells/cm^2 on gelatin-coated plates and grown for three days in N2B27 medium supplemented with 10 ng/ml b-Fgf (Table 3.5) without serum or LIF, following a previously described protocol (Gouti et al. 2014) with slight modifications. Subsequently, cells were grown for another two days in N2B27 medium without b-Fgf (D3–D5). Moreover, to improve the homogeneity of the differentiation, from D2-D5 the N2B27 medium was also supplemented with 5μM Xav939, a potent WNT inhibitor (see results 4.2) (Matsuda and Kondoh 2014).

3.3.5- Differentiation into mesodermal progenitors

E14 mESCs were plated at 75.000 cells/ml in a mixture of 75% IMDM (and 25% Ham's F12, supplemented with GlutaMax, B27, N2, Ascorbic acid 4.5x10-4 M, Monothioglycerol and BSA in order to produce embryonic bodies. After 48h, embryonic bodies were dissociated and re-aggregated for two days in the presence of 5 ng/ml human VEGF, 4 ng/ml human Activin A and 0.2 ng/ml human BMP4 following the instructions of a reported protocol (Wamstad et al., 2012).

3.4- Molecular biology methods

3.4.1- Genomic DNA isolation

DNA was isolated using QuickExtract DNA Extraction Solution 1.0 (Table 3.7) or an already established isolation protocol. Briefly, in this protocol, cells were suspended at 107 cells /ml in 10 mM Tris (pH 8.0), 10 mM EDTA. Then SDS and Proteinase K were added to a final concentration of 0.5% and 200 µg/ml, respectively. The mixture was incubated at 55 °C for 2 h. After that, NaCl was added to a final concentration of 0.2 M. The extraction was done twice with equal volumes of phenol/chloroform/isoamyl alcohol (25:24:1) and once with chloroform. RNase A was added to a final concentration of 25 µg/ml and incubated for 1 h at 37 °C. Another extraction with phenol/ chloroform/isoamyl alcohol (25:24:1) was done followed by a second one with only chloroform. DNA was precipitated with 1.5 volumes of ethanol followed by a centrifugation at 10000g for 5 minutes to pellet the DNA. The DNA pellet was washed with 70% ethanol twice and centrifuged at 10000g for 5 minutes. Finally the DNA pellet was resuspended in water.

3.4.2- RNA isolation

Total RNA was isolated using innuprep RNA mini kit (Table 3.4) according to the manufacturer's instructions or Tripure reagent (Table 3.7). For the Tripure isolation, this reactive was added directly to the wells after washing with PBS (500 µL reagent per 1x106 cells), following a two minutes incubation at RT. Afterwards, cells were resuspended, transferred to an eppendorf and a five minutes additional incubation at RT was performed to ensure complete

dissociation of nucleoprotein complexes. Then, chloroform was added (100 µL per 500 µL Tripure), followed by vigorous mixing and a posterior RT incubation during 10 min. Samples were centrifuged subsequently at 12.000g for 15 minutes at four degrees and the upper aqueous phase was isolated. RNA was precipitated by adding isopropanol to the aqueous phase (250 µL isopropanol per 500 µL Tripure), followed by intensive mixing and posterior incubation at RT during 10 min. Subsequently, samples were centrifuged at 7500g for 5 min at four degrees and pellet representing RNA was resuspended in RNAse free water. RNA concentration was measured by using Qubit fluorometer.

3.4.3- cDNA synthesis

To perform expression analyses, RNA was reversely transcribed into cDNA since RNA cannot serve as a template for RT-qPCR, cDNA was generated using ProtoScript II First Strand cDNA Synthesis Kit (Table 3.4). Exclusive transcription of mRNAs was warranted using an oligo-dT Primer binding to the mRNA specific poly-A tail.

3.5- Immunological methods

3.5.1- ChIP (Chromatin Immunoprecipitation)

A previously described protocol (Boyer et al. 2005) was followed with slight modifications. Briefly, 50 million of cells for a P300 ChIP or 10 million for histones ChIP were crosslinked with 1% formaldehyde for 10 minutes at RT and then quenched with 0,125M glycine for another 10 min. Posteriorly, cells were rinsed with PBS and resuspended sequentially in three different lysis buffers (Table 3.3) to isolate chromatin. Chromatin was then sonicated for 18 cycles (30 sec on 45 sec off) for P300 ChIP and 23 cycles for histones ChIP using the Bioruptor Plus. After sonication the material was centrifuged during 10 min at 16000g and 4°C. Afterwards, the chromatin from the supernatant was divided in different aliquots that were incubated overnight at 4 °C with 3 µg antibody for histones and 10 µg for P300 (Table 3.12). One of the aliquots was not incubated with the antibody, representing total input control for the ChIP reactions. Next day 100 µl of protein G magnetic beads were added to the

P300 ChIP reactions and 75µl to the histones ChIP. After four hours incubation at 4 °C, magnetic beads were washed with RIPA buffer and chromatin eluted, followed by reversal of the crosslinking and DNA purification. Briefly, for DNA purification, two extractions were performed, one with phenol/ chloroform/isoamyl alcohol (25:24:1), followed by a second chloroform extraction. The aqueous phase was isolated and DNA was precipitated during 30 minutes at -80°C by adding 1/10 of the volume of Sodium acetate, 1 µl glycogen (as internal carrier) and 3 volumes of 100% Ethanol. Afterwards, samples were centrifuged during additional 30 minutes at 4°C with DNA representing pellet. DNA was eluted in water. All antibodies used have been previously reported as suitable for ChIP (see Table 3.12). ChIP samples were analyzed by q-PCR using the primers shown in Table 3.25.

Table 3.12: Antibodies for ChIP

Antibodies	Company	Reference
P300	Santa Cruz	sc-585
H3K4me1	Active motif	39297
H3K4me3	Active motif	39159
H3K27me3	Active motif	39155
H3K27ac	Active motif	39133
RNA Pol2 8WG16	Covance	MMS-126R-500
S5p RNA Pol2	Abcam	ab5131

3.5.2- Sequential ChIP

A previously described protocol (Furlan-Magaril et al., 2009; Rada-Iglesias et al., 2011) was followed. Chromatin was prepared as described above for ChIP, after addition of the first antibody (H3K27me3, 3 µg) the corresponding washes were performed. Magnetic beads were then re-suspended in 75 ml TE/10 mM DTT solution. DNA samples were diluted 20 times with dilution buffer and second antibody (10 µg of P300 or GFP (Mock) antibodies) was added. Finally, beads were washed, crosslinking reversed, DNA purified and dissolved in water.

3.5.3- Western blot

Proteins from WT (E14), PE Lhx5$^{-/-}$ (-109) mESC and their AntNPCs were extracted using RIPA buffer supplemented with protease inhibitor cocktail (Table 3.3). After 20 minutes of incubation on ice, protein extracts were recovered by centrifugation (20 minutes at 14000g). For western blot 30 µg of protein were mixed with Laemmli buffer, heated to 95º and then separated in a 12% SDS-PAGE gel in running buffer (Table 3.3). Proteins were transferred to a PVDF (polyvinylidene difluoride) membrane using transfer buffer for 90 minutes at 80V. The membranes were blocked for 1 h with blocking solution and incubated afterwards with primary antibody overnight at 4°C (Table 3.13). After 3 x 5 min washes with PBST washing buffer, the membranes were incubated with the secondary antibody for 1 h at room temperature (Table 3.14). Horseradish peroxidase coupled anti-IgG antibody was detected using a chemiluminescence substrate. Antibodies used are listed on the tables:

Table 3.13: Primary Antibodies for western blot

Antigen	Host	Dilution	Company	Reference
LHX5	Goat	1:1000	R&D Systems	AF6290
TUBULIN	Mouse	1:10000	Millipore	MAB3408
ZIC2	Rabbit	1:1000	Abcam	ab150404

Table 3.14: Secondary Antibodies for western blot

Species	Conjugated	Dilution	Company	Reference
Anti-mouse	HRP	1:10000	Invitrogen	616520
Anti-goat	HRP	1:10000	Invitrogen	611620
Anti-rabbit	HRP	1:5000	Invitrogen	656120

3.5.4- Immunofluorescence staining

mESCs and AntNPCs growing in 12 well plates were rinsed with PBS and then fixed for 12 min in 4% paraformaldehyde (PFA) in PBS at room temperature (RT). PFA was removed and cells were rinsed with PBS. Cells were permeabilized during seven minutes at RT, followed by incubation with

blocking solution for 30 min at RT (Table 3.3). Incubation with primary antibodies was done in blocking solution overnight (12–18 h) at 4 °C with gently shaking (Table 3.15). Cells were rinsed three times with PBS and incubated with secondary antibodies in blocking solution for one hour at room temperature and then rinsed again with PBS (Table 3.16). Finally, cell nuclei were stained with DAPI (1µg/ml) during 10 min at RT and then mounted with anti-fading mounting medium.

Table 3.15: Primary Antibodies for immunofluorescence

Antigen	Host	Dilution	Company	Reference
POU5F1	Goat	1:500	Santa Cruz Biotecnhology	sc-8628
NESTIN	Rabbit	1:500	Covance	PRB-315C
PAX6	Mouse	1:100	Hybridoma Bank	PAX6
LHX5	Mouse	1:500	Santa Cruz Biotecnhology	sc-130469

Table 3.16: Secondary Antibodies for immunofluorescence

Species	Conjugated	Dilution	Company	Reference
Anti-mouse	488	1:500	Life Technologies	A11001
Anti-goat	594	1:500	Life Technologies	A11058
Anti-rabbit	488	1:500	Life Technologies	A11008

3.6 - Genetic deletions using clustered regulatory interspaced short palindromic repeats (CRISPR-Cas9)

3.6.1. Design of guide RNA (sgRNA)

In order to generate deletions of selected poised enhancers, pairs of sgRNA were designed flanking each candidate poised enhancer according to the instructions of the genome engineering toolbox from Zang lab (http://www.genome-engineering.org/crispr/). To minimize potential off-targets, all selected gRNAs displayed on-target scores >80 and potential off-targets with scores <7 (in most cases off-target scores were <1). Furthermore, the sequence was truncated to be 19 bp long and a G was added at the 5' end since that was reported to help the specificity of the gRNAs (Fu et al. 2014). For each selected sgRNA, two oligonucleotides were synthesized carrying the ends of BbsI restriction enzyme (Table 3.17). Complementary oligos were annealed by incubation at 95°C for five minutes and subsequent cooling to 25°C at a cooling rate of 5°C/min. The next annealing reaction was performed:

1 µl oligo 1 (100µM)
1 µl oligo 2 (100µM)
1 µl 10X T4 Ligation Buffer (NEB)
7 µl ddH 2 O
10 ul total

Table 3.17: gRNAs used for CRISPR-Cas9 and oligos for vector cloning

Name	gRNA with PAM domain	Oligo F (5'-3')	Oligo R (5'-3')
PE_Six3_ left	CCTGACTTCTCGTAGACTCC-TGG	caccGTGACTTCTCGTAGACTCC	aaacGGAGTCTACGAGAAGTCAC
PE_Six3_ right	GGACCTTTCTGGACGCTGAT-AGG	caccGACCTTTCTGGACGCTGAT	aaacATCAGCGTCCAGAAAGGTC
PE_Sox1_ left	ATTGCCTCCTGCGCGTGGCA-TGG	caccGTGCCTCCTGCGCGTGGCA	aaacTGCCACGCGCAGGAGGCAC
PE_Sox1_ right	GGCTTCTGAAGTGGGGCGTC-TGG	caccGCTTCTGAAGTGGGGCGTC	aaacGACGCCCCACTTCAGAAGC
PE_Lhx5_ left	TCTTTCAGCGATGATCCGGG-AGG	caccGTTTCAGCGATGATCCGGG	aaacCCCGGATCATCGCTGAAAC
PE_Lhx5_ right	AGCGCGCTTAACAAGCATTA-CGG	caccGCGCGCTTAACAAGCATTA	aaacTAATGCTTGTTAAGCGCGC
PE_Sox21_ left	TGCGCACAGATCCCGACGCT-GGG	caccGCGCACAGATCCCGACGCT	aaacAGCGTCGGGATCTGTGCGC
PE_Sox21_ right	GTCCAAGGAAGTAACCGCAA-GGG	caccGCCAAGGAAGTAACCGCAA	aaacTTGCGGTTACTTCCTTGGC
PE_Wnt8b_ left	TATGTGTAGGCTCACCAGCC-AGG	caccGTGTGTAGGCTCACCAGCC	aaacGGCTGGTGAGCCTACACAC
PE_Wnt8b_ right	CACTAATAGGTGTCCTCGTC-CGG	caccGCTAATAGGTGTCCTCGTC	aaacGACGAGGACACCTATTAGC

3.6.2. Generation of guide RNA (sgRNA) Cas9 vector

The CRISPR-Cas9 expression vector pX330-U6-Chimeric_BB-CBh-hSpCas9 or pX330-hCas9-long-chimeric-grna-g2p was digested with BbsI and column purified.

1 µg vector
1 µl BbsI
2 µl 2.1 Buffer
X µl ddH2O
µl total

A dilution 1:200 of the pair of annealed oligos and 50 ng of the digested vector were ligated overnight at 16°C using T4 ligase. For heat shock transformation reaction, 30µl of chemically competent E.coli (TOP10) (Table 3.9) bacteria were thaw on ice and mixed with 2.5 µl of the ligation reaction. After tapping the tube three times, bacteria solution was incubated 1 minute on ice, followed of heat shock for 1 minute at 37°C and another 1 minute incubation on ice. Afterwards, solution was transferred to 1ml TB Medium and incubated 1 hour at 37 °C with shaking. 200 µl of this liquid culture were plated in LB ampilicine plates in order to obtain bacteria colonies that were analized by PCR to determine if they carried the plasmid of interest (3.9.3).

3.6.3 Miniprep

The gRNA-Cas9 expression vectors were purified following the instructions of a comertial kit (Table 3.4) and sequenced to confirm that the gRNAs were correctly cloned.

3.6.4 Transfection of mESCs

mESC were grown on gelatin-coated 12 well or 6 well plates with standard mESCs cell culture conditions. Cells were transfected using X-tremeGENE HP DNA transfection reagent according to the manufacturer's instructions with two sgRNA Cas9 vectors carrying gRNAs flanking the PE regions for posterior deletion. Transfection efficiency was checked by GFP signal 16 hours later.

When transfection efficiency was between 40-70%, puromycin selection was performed for 48 hours at 2µg/ml concentration. Afterwards, surviving cells were isolated in 96-well plates by serial dilution and, after expansion; clones with the chosen deletions were identified by PCR (see methods 3.9.4). Using the primers listed in table 3.18. 20-40, mESC clones were investigated by PCR for each PE candidate. Sanger sequencing in all the $PE^{-/-}$ or PE $^{+/-}$ clonal mESC lines corroborated PE deletions. In order to explore how the PE deletions affected the induction of their target genes towards AntNPC differentiation, mESC lines with PE deletions were compared to WT mESC in differentiations performed in parallel. This comparison corrected for technical bias between differentiations.

Table 3.18: Primers used to detect deletions generated by CRISPR-Cas9

Primer name	Sequence	Product size (bp)	Deletion size (bp)	Outcome size (bp) Homozygous deletion
PE_Lhx5_del_F	CCACCTTAGGGCTGATCAAA	795	555	240
PE_Lhx5_del_R	TGTGCCCTGGATTTCTCTTC			
PE_Sox21_del_1F	ATGGGCAAGCAAAGAGAAGA		1400	
PE_Sox21_del_2R	GCCAGAGTTGGAAAGTGAGC	(1+2) 2324		924
PE_Sox21_del_3R	GCCTCGGGGTGTTTACAGAA	(1+3) 713		No band
PE_Sox1_del_1F	CAAGAATTCCACCCTCATCC		5500	
PE_Sox1_del_2R	TCTGTGAAGGGAGCTGAGGT	(1+2) 5994		494
PE_Sox1_del_3R	TGGAATGTATCGGAGGGGAC	(1+3) 666		No band
PE_Six3_del_1F	TTGAATCCTTTGGCCTCATC		5200	
PE_Six3_del_2R	ACCACAAGAACCCACCAGAG	(1+2) 6318		1118
PE_Six3_del_3R	ATTAAAAACAGCAGTGCCCCA	(1+3) 938		No band
PE_Wnt8b_del_F	GGGGTCCTGAGAAGTGACAA	1433	1200	233
PE_Wnt8b_del_R	GTGTGGCTTCCTGCACTGTA			

3.7- Circular chromatin conformation capture (4C)

3.7.1- 4C library generation

Circular Chromatin Conformation Capture assays were performed as previously reported (Splinter et al. 2012; Stadhouders et al. 2013) with minor modifications. 107 mESC or AntNPCs were crosslinked with 1% formaldehyde during 20 minutes and quenched with 0.125M glycine (final concentration) for additional 10 min. Cells were rinsed with PBS and proteinase inhibitors and resuspended in lysis buffer (Table 3.3) during 10 min on ice. After 5 min centrifugation at 650g (4 °C), chromatin was re-suspended in 0.5 ml of 1.2X restriction buffer with 0.3% SDS and incubated at 37 °C and 900 rpm for 1h. Then, Triton X-100 was added to a final concentration of 2% followed by incubation at 37 °C and shaking at 900 rpm during 1h. Subsequently, chromatin was digested at 37 °C overnight while shaking (900 rpm) with 400 U of NlaIII. The inactivation of the enzyme was performed by adding SDS to a final concentration of 1.6% followed by incubation during 20 min at 65 °C and 900 rpm. The digested chromatin was transferred to 50-ml tubes and 6.125 ml of 1.15X ligation buffer (Table 3.3) was added. After addition of Triton X-100 (1% final concentration) chromatin was incubated for 1 h at 37 °C (shaking gently). The choice of the restriction enzymes depends on the distance between target and viewpoint, in this case, target gene and poised enhancer. Usually, the fragment should be preferentially located within the viewpoint and have a size around 500 bp (Splinter et al. 2012). To get a better resolution map and sequence depth (broad range of contacts), two four base pair cutters (NlaIII and DpnII) were chosen. After Triton incubation, digested chromatin was ligated with 100 U of T4 DNA ligase for 8 h at 16 °C. RNAse A was added (300 µg) for 45 min at 37 °C and afterwards, chromatin was de-crosslinked with 300 µg of Proteinase K and overnight incubation at 65 °C. Phenol/chloroform extraction followed by ethanol precipitation was used to purify DNA followed by re-suspension in 100 µl of water. Digestion and ligation efficiencies were evaluated by running a small fraction of the purified DNAs on an agarose gel (Figure 3.1). DNA obtained from the previous step was digested with 50U of

was digested with 50U of DpnII in 500 µl of 1x NEB DpnII buffer at 37 °C overnight. Afterwards, DNA was purified by phenol/chlorophorm extraction, followed by ethanol precipitation. Samples were re-suspended in 500 µl of water and a second ligation was performed. In order to increase the contacts in proximity, ligation was carried out in a final volume of 14 ml of water supplemented with 1X ligation buffer (Table 3.3) and 200 U of T4 DNA Ligase were added. After overnight incubation at 16 °C, DNA samples were again purified by phenol/chlorophorm extraction and ethanol precipitation, re-suspended in 100 µl of water and finally isolated with a QIAgen PCR purification column (Table 3.4). The efficiencies of the second digestion and ligation were also evaluated by agarose electrophoresis (Figure 3.1).

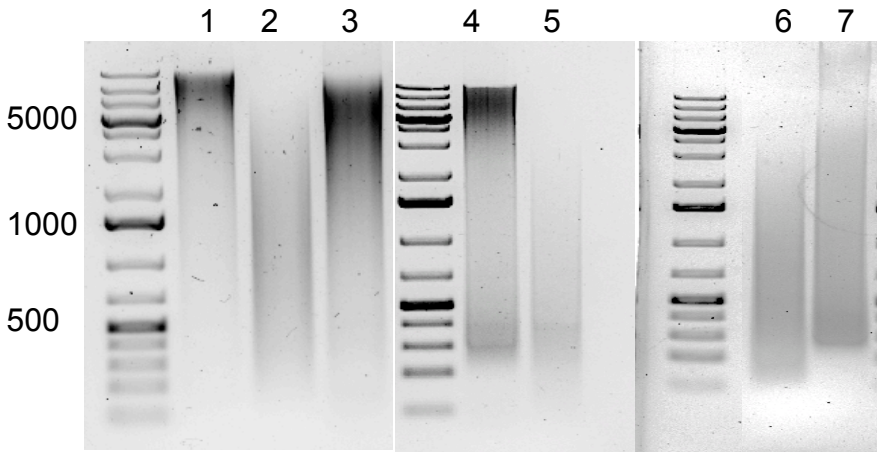

Figure 3.1: Analysis of digestion and ligation of the 4C protocol. A, 1,5% agarose electrophoresis gel. **1:** undigested DNA**. 2**: digested DNA with NlaIII **3,4** ligated DNA after digestion with Nla III. **5, 6:** digested DNA with DpnII. **7**: Ligated DNA after digestion with DpnII.

3.7.2- 4C -seq

The resulting 4C DNA products were amplified by inverse PCR (3.9.5) using primers described in table 3.19, and run into a 1,5 % agarose gel. When the pattern was the expected (one prominent band product of self ligation and a broad smear, (Figure 3.2), samples were sent for sequencing to the Erasmus Medical Center in Rotterdam (Netherlands). 4C libraries were prepared according to Illumina protocol and the resulting reads from the sequencing facility were then mapped and analyzed with R3C-seq (Thongjuea et al. 2013)

to identify contacts between our viewpoints (poised enhancers) and other genomic loci.

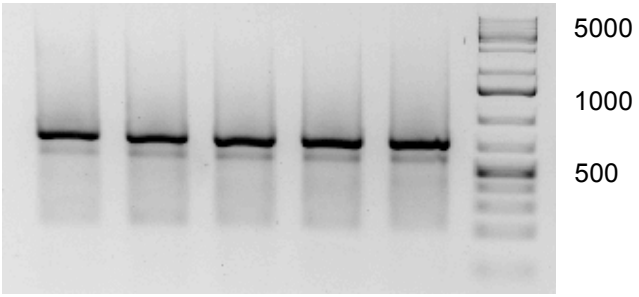

Figure 3.2: Pool of 4C-PCR reactions: 1,5% agarose gel for a pool of five 4C reactions to be further sequenced. The pattern shows a prominent band, product of the self-ligation, a secondary band, which could be unspecific products or re ligated products and a broad smear representing the 4C contacts.

3.7.3 - 3C validation of 4C-seq results.

3C (DNA samples obtained after first ligation of the 4C protocol) and 4C samples were prepared as previously described, using NlaIII and NlaIII+DpnII as restriction enzymes, respectively. Fragments located within the PEs and gene promoter were used to design 3C primers in NlaIII areas of interest, typically around 100 bp away from a NlaIII site and without DpnII sites in between (Table 3.20). This strategy warranted that the same primers could be used to analyze both 3C and 4C samples. As a loading control, a set of PCR primers which amplify a 200 bp fragment without NlaIII sites at the Tbp locus were also designed. Primer sequences and the expected product sizes can be found in table 3.20.

To generate the 3C control library, two ~2000 bp fragments per viewpoint locus were amplified by PCR. Some of these 2000 bp fragments contained our PE sequences (i.e. the 4C-seq viewpoint) and the others, the promoters of the PEs cognate genes (Table 3.21). The 2000 bp fragments contained several NlaIII sites. In addition, we also included the ~200 bp loading control *Tbp*, which was not digested by NlaIII. PCR products were digested with NlaIII overnight

followed by a round of phenol/chlorophorm purification. Afterwards, they were mixed in equimolar concentration (nanomolar range), ligated and purified. This control library covered all the chimeric products expected to be obtained with the 3C primers in high concentration. The control library was used to test the specificity and efficiency of the 3C primers by performing serial dilutions PCRs (see methods 3.9.6).

Table 3.19: Primers used for 4C-seq

	Viewpoint	Illumina adaptor	Barcode	Primer 5' --> 3'	Complete sequence
p5 (1):	PE Six3	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT	TGACCA	GGGGCGCTCCTTCT TTTT	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGA CGCTCTTCCGATCTTGACCAGGGGCGCTCCTTCTTTTT
p5 (2):	PE Six3	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT	ACATGT	GGGGCGCTCCTTCT TTTT	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGA CGCTCTTCCGATCTACATGTGGGGCGCTCCTTCTTTTT
p7:	PE Six3	CAAGCAGAAGACGGCATACGA		CGGCTTGTTTACCG CTTTAAT	CAAGCAGAAGACGGCATACGACGGCTTGTTTACCGCTTT AAT
p5 (1):	PE Lhx5	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT	GCCAAT	CGGGGGTCTGATTA AAGGTC	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGA CGCTCTTCCGATCTGCCAATCGGGGGTCTGATTAAAGGT C
p5 (2):	PE Lhx5	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT	CTAGCT	CGGGGGTCTGATTA AAGGTC	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGA CGCTCTTCCGATCTCTAGCTCGGGGGTCTGATTAAAGGT C
p7:	PE Lhx5	CAAGCAGAAGACGGCATACGA		GGGAGGGGGTTTAC AAACTT	CAAGCAGAAGACGGCATACGAGGGAGGGGGTTTACAAA CTT
p5 (1):	PE Sox1	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT	CTTGTA	AAACTGGCCGGCTG AATAG	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGA CGCTCTTCCGATCTCTTGTAAAACTGGCCGGCTGAATAG
p5 (2):	PE Sox1	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT	GCCAAT	AAACTGGCCGGCTG AATAG	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGA CGCTCTTCCGATCTGCCAATAAACTGGCCGGCTGAATAG
p7:	PE Sox1	CAAGCAGAAGACGGCATACGA		GGGGAATACTGGCT GGTAGC	CAAGCAGAAGACGGCATACGAGGGGAATACTGGCTGGT AGC
p5 (1):	PE Lmx1b	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT	ATCACG	TTCGCAATGCAAAG CATCTA	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGA CGCTCTTCCGATCTATCACGTTCGCAATGCAAAGCATCT A
p5 (2):	PE Lmx1b	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT	TGACCA	TTCGCAATGCAAAG CATCTA	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGA CGCTCTTCCGATCTTGACCATTCGCAATGCAAAGCATCT A
p7:	PE Lmx1b	CAAGCAGAAGACGGCATACGA		CCTCACACAGGGAG ACCACT	CAAGCAGAAGACGGCATACGACCTCACACAGGGAGACC ACT

Table 3.20: 3C Primers used for 3C validation of 4C

Primer name	Sequence	Distance to NlaIII site (bp)	Expected size product (bp)
3Cval_Six3_enh	CCTGCTTAACGGGGAAATTG	104	192
3Cval_Six3_pr	TCGCAAGTCATCTTCAATCG	88	
3Cval_Sox1_enh	CATTAATCTGGGAACAAACAGCTAA	115	195
3Cval_Sox1_pr	CAGACAGACAAACTTTCTCCATTTT	80	
3Cval_Lmx1b_enh	GGTGACAGTTACCCGCAGTT	111	176
3Cval_Lmx1b_pr	GACCACAGCCTTGGATTCAT	65	
3Cval_Lhx5_enh	CCAGCTCTCCAGCTCATTTT	62	181
3Cval_Lhx5_pr	TAGCTCGGCAAAGGAGAAAG	119	
NEGATIVE CONTROL			
3Cval_Sox1_enh	CATTAATCTGGGAACAAACAGCTAA	115	203
3Cval_Six3_pr	TCGCAAGTCATCTTCAATCG	88	
LOADING CONTROL			
Tbp_F	ACAACAGCAGGCAGTAGCAA	NA	193
Tbp_R	TGGTGTGGCAGGAGTGATAG	NA	

Table 3.21: Template primers used for 3C validation of 4C

Primer name	
4cval_temp_six3_Pr_F	CTCCAGTTGGGTTCACGTTT
4cval_temp_six3_Pr_R	GGAGGAGGAAGGACGTAAGG
4cval_temp_six3_Enh_F	ATGGCTATTGCTCCACCAAG
4cval_temp_six3_Enh_R	AAGCAGGCTGCCTTGAAATA
4cval_temp_Lmx1b_Pr_F	TAAAACAACTCCGCGCTTCT
4cval_temp_Lmx1b_Pr_R	CGTTGAAGTGGCTGACTGAA
4cval_temp_Lmx1b_Enh_F	GGTGACAGTTACCCGCAGTT
4cval_temp_Lmx1b_Enh_R	CAGCCTTCAGGAGAGAGAGC
4cval_temp_Sox1_Pr_F	TGCCAGGGAATGTAAACACA
4cval_temp_Sox1_Pr_R	GTCCCCAGAACATGAGTGCT
4cval_temp_Sox1_Enh_F	CCAGCCTCTTAATCCAGTGC
4cval_temp_Sox1_Enh_R	CAGCTTTGTGAAGCCATTGA
4cval_temp_Lhx5_Pr_F	CACCACGTCACCAAATCTGA
4cval_temp_Lhx5_Pr_R	AGCAGAAGAATGGGAGCAGA
4cval_temp_Lhx5_Enh_F	CTGCCAGAGGAGGAAATGAG
4cval_temp_Lhx5_Enh_R	AATGAGGAAGCTGGGGTTCT

3.8 - Enhancer reporter assays

These experiments were performed using a previously described system based on PiggyBac transposon (Respuela et al. 2016; Cruz-Molina et al. 2017). Basically, this method consisted in amplifying by PCR the PE sequences of interest (PE Lhx5 (-109) and PE Wnt8b (+21)) (Table 3.22) and clone them afterwards in front of a minimal TK promoter driving GFP expression. Then, WT mESC were transfected with Super PiggyBac Transposase vector and the resulting PE PiggyBac reporter vectors (Table 3.10) using X-tremeGENE HP DNA transfection reagent according to the manufacturer's instructions. 48 hours after transfection, neomycin (0.2 mg/ml) was added to the medium to select cells that incorporated the reporter vector in a stable way. After one week of neomycin selection, the surviving cells were expanded and differentiated afterwards into AntNPC. The negative control for this experiment was a mESC line, also transfected with an empty PiggyBac reporter vector. The levels of GFP were evaluated using a Nikon Microscope ECLIPSE TS100 coupled with Nikon intense light C-HGFI for florescence illumination (Table 3.1).

Table 3.22: Primers used for amplification of PEs in reporter assays.

Primer name	Sequence	Product size(bp)
Wnt8b_PE_BamhI_5	AAAAAAGGATCCAGAGAAAGGCTCCTCCTTGG	1208
Wnt8b_PE_EcoRI_3	AAAAAAGAATTCAACTGAACCTGCGTCCTTTG	
Lhx5_PE_BamhI_5	AAAAAAGGATCCCTGGCTTTTCAAGGAAGCTG	896
Lhx5_PE_EcoRI_3	AAAAAAGAATTCCCTTGTCTTGTCTGGTGCAC	

3.9- Polymerase Chain Reaction (PCR)

3.9.1- RT-qPCR

In order to determine the transcript levels of our cells, RNA was extracted and cDNA was produced as described above (3.4.2 and 3.4.3). Prior to RT-qPCR, samples were diluted 1:5 in water and 0,5 µl of this dilution were used per 10 µl reaction (master mix per sample: 5 µl ORA qPCR Green ROX L Mix (2x),

0,125.µl 20µM primer mix and 4,375 µl dH20). PCR reactions were performed in 396 well plates on the Light Cycler 480II (Roche) with the program “SYBR Green I 384 II” (Table 3.23) using *Eef1a1* and *Tbp* as housekeeping genes (Table 3.24). Analysis of the resulting amplification curves was performed with the second derivative maximum method implemented in the LightCycler software. All the measurements were performed as triplicates and standard deviations were represented as error bars.

Table 3.23: RT-qPCR program

Step	Temp	Time
Initial Denaturation	95°C	5 minutes
45 Cycles	95°C 60°C 72°C	10 seconds 10 seconds 10 seconds
Melting curve	95°C 65°C	5 minutes 1 minute
Hold	4-10°C	

Table 3.24: RT-qPCR primers

Primer name	Sequence
Eef1a1_cDNA_F	TAGACGAGGCAATGTTGCTG
Eef1a1_cDNA_R	AGCGTAGCCAGCACTGATTT
Tbp_cDNA_F	TTCGTGCAAGAAATGCTGAA
Tbp_cDNA_R	TCCTGTGCACACCATTTTTC
Zfp42_cDNA_F	GCGGTGTGTACTGTGGTGTC
Zfp42_cDNA_R	GACAAGCATGTGCTTCCTCA
Six3_cDNA_F	CCTCACCCCCACACAAGTAG
Six3_cDNA_R	CTGATGCTGGAGCCTGTTCT
Wnt8b_cDNA_F	ACTCCCGAAATGGACAACTG
Wnt8b_cDNA_R	CTGCTTGGAAATTGCCTCTC
Lxh5_cDNA-F	CGGGAAGGCAAGCTATACTG
Lhx5_cDNA-R	CAGGTCGCTCGGAGAGATAC
Sox1_cDNA_F	TCGAGCCCTTCTCACTTGTT
Sox1_cDNA_R	CAACCCAAAAGAGCGGTAAC
Sox21_cDNA_F	CTCTCCTCTCCTGTGCCAAA

Sox21_cDNA_R	GGAACCCCCAATCCTGTAGT
Lmx1b_cDNA_F	GGGATCGGAAACTGTACTGC
Lmx1b_cDNA_R	GTAGGGGCGATCTTCTCCAT
Emx2_cDNA_F	GCACGCTTTTGAGAAGAACC
Emx2_cDNA_R	GTTCTCCGGTTCTGAAACCA
T_cDNA_F	TGCTGCAGTCCCATGATAAC
T_cDNA_R	CCATTGCTCACAGACCAGAG
Six3_cDNA_AS_F	GCTGTACTGGCCTCTTCTGG
Six3_cDNA_AS_R	AAAGGGCCAAGGAATGATCT
Lmx1b_cDNA_AS_F	GACCACAGCCTTGGATTCAT
Lmx1b_cDNA_AS_R	CTTCGGTTCTTGAGGCAGAG
Lhx5_cDNA_AS_F	GATAATTCAAGGCGGCAGTG
Lhx5_cDNA_AS_R	TGTCTGGTCAGCAAGCAATC
Sox21_cDNA_AS_F	AGCAGTGGCATGTTAAGTGG
Sox21_cDNA_AS_R	GCAACAGGAGAGCAAAAACA
Sdsl_cDNA_F	CTCCGTGCAGGTGGTGAG
Sdsl_cDNA_R	ACATTGGCTTCATCCCAGAC
Camkmt_Six3_cDNA_F	GAGCCCATTTCAAACTTCCA
Camkmt_Six3_cDNA_R	AAGCAGAAGTGGGTAGTGCAG
Spaca7_Sox1_cDNA_F	TCTGTCTTCCTGCTGTGCTG
Spaca7_Sox1_cDNA_R	TGCTTCGGGTATGTTTTCTG
Sec31b_wnt8b_cDNA_F	GAAGCACACTGGAGCTGTCA
Sec31b_wnt8b_cDNA_R	GGTGGTTCAAATCCCAAATG
Gpr180_sox21_cDNA_F	GGTCATCACAGTGGAGAGGAG
Gpr180_sox21_cDNA_R	TCATTAAAAGCGACAGCAATG
GSC_cDNA_F	CAGGAGACGAAGTACCCAGAC
GSC_cDNA_R	AGGAGGATCGCTTCTGTCG
Fez2f_cDNA_F	GGCTACAAGCCCTTCGTCT
Fez2f_cDNA_R	TGTGGGTGAGCTTGTGATTC
Mixl1_cDNA_F	CGACAGACCATGTACCCAGA
Mixl1_cDNA_R	GGCTGAAATGACTTCCCACT
mPax6_cDNA_F	CTTGGGAAATCCGAGACAGA
mPax6_cDNA_R	CTAGCCAGGTTGCGAAGAAC
Tubgcp3_cDNA_ SOX1_F	CTCACTGACATCCGGAAAGG
Tubgcp3_cDNA_ SOX1_R	ATCTCTGAGGCCAGGATGTG

Atp11a_cDNA_ SOX1_F	CCACCGTCTTTATGCTTTCC
Atp11a_cDNA_ SOX1_R	GGGGTGAATTCTGAAATGCT
Ndufb8_cDNA_Wnt8b_F	GGGACCACTCAGAACTCAGG
Ndufb8_cDNA_Wnt8b_R	CATGACATCCCAGGACACAG
Hif1an_cDNA_Wnt8b_F	TTGAGAAGATGCTTGGAGAGG
Hif1an_cDNA_Wnt8b_R	CGGCCTTTAATCATGGTGTT
Prep1_cDNA_Six3_F	GGAACCACGTGATTGAGGAT
Prep1_cDNA_Six3_R	GCATCAGTGCTGTCAAGTCC
Six2_cDNA_Six3_F	GCAAGTCAGCAACTGGTTCA
Six2_cDNA_Six3_R	TGGAGTTCTCGCTGTTCTCC
Gm9366_cDNA_Sox21_F	CATCACTCAGCGTCGAAGAA
Gm9366_cDNA_Sox21_R	CACAGGACGTTTTCCAATCC
Tgds_cDNA_Sox21_F	TAGGATGGAAGCCCAAAGTG
Tgds_cDNA_Sox21_R	TGGACTGGAAAGGGCTCTAA

3.9.2- ChIP-qPCR

ChIP DNA samples were analyzed by q-PCR to detect the level of enrichment of the histone marks and P300. Input samples were diluted 1:10 and histone ChIP samples 1:3 prior to q-PCR. PCR conditions, program and analysis of the curves were the same than above (3.9.1, Table 3.23). ChIP-qPCR signals were calculated as percentage of input. Standard deviations were measured from the technical triplicate reactions and were represented as error bars. Each ChIP sample was normalized to the signal average obtained in the same sample when two different negative control regions were used (chr2_neg, chr6_neg) (Table 3.25). Standard deviations were calculated from technical triplicate reactions and represented as error bars.

Table 3.25: Primers used in ChIP-qPCR

Primer name	Sequence
Chr6_neg_F	CTGGACTGAGGACCTTCTGC
Chr6_neg_R	AGGAAGGCAGATGAGGGATT
Chr2_neg_F	CCTGAGGCTGGAAGTTTCTG

Chr2_neg_R	CTCCTGGGATTAAAGGCACA
Sox2_prom_F	CGCCCAGTAGACTGCACAT
Sox2_prom_R	CCCTCCCAATTCCCTTGTAT
Zfp42_prom_Pol2_F	TGCTTTCACCATTGACTTTCC
Zfp42_prom_Pol2_R	CACTGGGGATACACGCATTA
Six3_prom_H3K27ac_F	ACCACATCCTGGAGAACCAC
Six3_prom_H3K27ac_R	CTTCTCGGCCTCCTGGTAGT
Six3_PE_H3K27ac&me3&Pol2_F	GCTCACTGAAGCAGCATTTG
Six3_PE_H3K27ac&me3&Pol2_R	CCTAAGGATTGTGGGCTCAT
Six3_prom_Pol2_F	CCTCCTTTCCATCTCCTCCT
Six3_prom_Pol2_R	CGGAATACCATGGACTGACC
Six3_enh147kb_H3K27ac_F	GAGGAGGCTCAGTGTCAAGG
Six3_enh147kb_H3K27ac_R	TTGGTCACCTTTGGTTCACA
Six3_enh7kb_H3K27ac_F	ACTTACCCAGCCACCTGTTG
Six3_enh_7kb_H3K27ac_R	TTGGCTTTTGCTTTTGTGTG
Lhx5_prom_Pol2&K27ac_F	CGGGAAGGCAAGCTATACTG
Lhx5_prom_Pol2&K27ac_R	TCCCAAGGAGCCAGAGTAGA
Lhx5_PE_H3K27ac&me3&Pol2_F	ATGCCTCTCTGGCTTTTCAA
Lhx5_PE_H3K27ac&me3&Pol2_R	ATCAGCCCTAAGGTGGCTTT
Lhx5_enh144kb_H3K27ac_F	CCCCTCCCTCAATCTCTTCT
Lhx5_enh144kb_H3K27ac_R	CCAGGAAAGCCTTGGTAAAA
Lhx5_enh61kb_H3K27ac_F	GACTCCCTTGCTGTGAGACC
Lhx5_enh61kb_H3K27ac_R	GTAGCCCGCTTGGACATAAA
Wnt8b_PE_Pol2_F	GAGTGAAATGCCCCTGACAT
Wnt8b_PE_Pol2_R	AGGGAAGTATGTCCCCTGCT
Wnt8b_prom_Pol2_F	GGGATCGCTTACACACCAAG
Wnt8b_prom_Pol2_R	GGGTGAGCGACTCAAACATT
Sox1_PE_Pol2_F	CTCAAGAGCATTTTGCCACA
Sox1_PE_Pol2_R	GGGCTTGCTTCATTCTTTGA
Sox1_prom_Pol2_F	CTAGAAGTTGCGGTCCCAGA
Sox1_prom_Pol2_R	GTTCGCTGCCTCCTCTTGT
Brpf3_2556_F	GATCTCTCCTTTGGGCCTCT
Brpf3_2556_R	GGCTGGCCTTAAAATGACAA
Calcoco2_2382_F	TCAAATTCGGGGTCAGTTTC
Calcoco2_2382_R	GCCTTTCAACAATGCTCCTC

Chr_17neg_R	GAACAGGAATGGCCACAGAT
Chr_17neg_F	CTCCTGCCCTGTGGTATGTT
Lhx5_seqchip_p300_F	GGCCCAAAACCCCTAAAATA
Lhx5_seqchip_p300_R	GGCTTCATGTGGGACCTTTA
Six3_seqchip_p300_F	GTCAGGGCTGTTGTCCAATC
Six3_seqchip_p300_R	TCTGCATCCACATTTGTTCC
Sox1_seqchip_p300_F	CCTGAAAATGATGCTGCTGA
Sox1_seqchip_p300_R	GGAGTAGCTGTGGGTGTGGT

3.9.3- Colony- PCR

In order to determine whether the bacterial colonies carry the plasmid of interest, PCRs were performed using One Taq DNA polymerase (Table 3.26 and 3.27). Bacterial colonies were resuspended in 30 µl of dH20 and 20 of these were heated at 95°C to lysate the cells. Three microliters of the lysate was used per PCR reaction. The 5'-3' oligo cloned in the vector (Table 3.17) was used as a forward primer and the px330-R primer from the vector as a reverse (ggaaagtccctattggcgtt). If PCR works, a band of 300 bp is obtained.

Table 3.26: Colony- PCR master mix

Component	25 µl reaction	Final Concentration
5X One*Taq* Standard Reaction Buffer	5 µl	1X
10 mM dNTPs	0.5 µl	200 µM
10 µM Forward Primer	0.5 µl	0.2 µM
10 µM Reverse Primer	0.5 µl	0.2 µM
One*Taq* DNA Polymerase	0.125 µl	1.25 units/ 50µl PCR
Template DNA	3 µl	< 1,000 ng
Nuclease-free water	to 25 µl	

Table 3.27: Colony- PCR program

Step	Temp	Time
Initial Denaturation	94°C	30 seconds
30 Cycles	94°C 63°C	30 seconds 30 seconds

	68°C	1 minute/kb
Final Extension	68°C	5 minutes
Hold	4-10°C	

3.9.4- Clonal genotyping PCR

To define the genotype of the clones obtained after transfection (3.6.4), DNA was isolated from them using quick DNA extract solution (Table 3.7) and PCR was subsequently performed (Table 3.28, 3.29). Platinum PCR was chosen due to its robustness (Table 3.8).

Table 3.28: Clonal genotyping PCR master mix

Component	25 µl reaction	Final Concentration
10X PCR Buffer without Mg	2.5 µl	1X
50 mM $MgCl_2$	0.75 µL	1.5 mM
10 mM dNTPs	0.5 µl	200 µM
10 µM Forward Primer	0.5 µl	0.2 µM
10 µM Reverse Primer	0.5 µl	0.2 µM
Platinum™ *Taq* DNA Polymerase	0.1 µl	2 U/reaction **
Template DNA	1 µl	< 500 ng
Nuclease-free water	to 25 µl	

Table 3.29: Clonal genotyping PCR program used

Step	Temp	Time
Initial Denaturation	94°C	2 minutes
39 Cycles	94°C 60°C 72°C	30 seconds 30 seconds 1 minute/kb
Hold	4-10°C	

3.9.5- 4C –PCR

4C samples were amplified by inverse PCR using expand long template PCR system and primers located within selected poised enhancers, which were designed as previously described (Stadhouders et al. 2013) (Table 3.30, 3.31). The primer requirements are similar to those used in standard PCR

reactions, with a preferred length of the oligos between 17 and 24 nt to facilitate efficient amplification, and annealing temperatures are generally around 60°C. Primers were design with Primer 3 software (http://bioinfo.ut.ee/primer3-0.4.0/). First, primers without Illumina adaptors were tested to determine if 4C products present the expected amplification pattern after PCR on a gel (Table 3.19). If pattern is correct, PCR is performed using primers with Illumina adaptors and 1/5 of the PCR product was run into a gel (Figure 3.2). After checking that a shift of the size of the adaptors is produced in the gel pattern, the rest of the sample is purified using Quiagen PCR purification columns and sent for sequencing.

Table 3.30: 4C –PCR master mix

Component	25 µl reaction	50 µl for Ilumina seq	Final concentration
10X PCR buffer 1 (17,5mM $MgCl_2$)	2.5 µl	5 µl	1X
10 mM dNTPs	0.5 µl	1 µl	200 µM
35 pmol/µl Forward Primer	1 µl	2µl	35 pmol
35 pmol/µl Reverse Primer	1 µl	2µl	35 pmol
Expand Long Template Polymerase	0.35 µl	0,7 µl	3.75 U
Template DNA	Depends on concentration	Depends on concentration	150 ng
Nuclease-free water	to 25 µl		

Table 3.31: 4C –PCR Program used

Step	Temp	Time
Initial Denaturation	94°C	2 minutes
30 Cycles	94°C 60°C 68°C	10 seconds 1 minute 3 minutes
Final extension	68°C	5 minutes
Hold	4-10°C	

3.9.6 – 3C-PCR validation of 4C

Control template library and 3C PCR amplifications were performed with the Platinum Taq polymerase using the following components and program (Table 3.32, 3.33):

Table 3.32 3C- PCR validation of 4C master mix

Component	25 µl reaction	Final Concentration
10X PCR Buffer without Mg	2.5 µl	1X
50 mM $MgCl_2$	0.75 µL	1.5 mM
10 mM dNTPs	0.5 µl	200 µM
10 µM Forward Primer	0.5 µl	0.2 µM
10 µM Reverse Primer	0.5 µl	0.2 µM
Platinum™ *Taq* DNA Polymerase	0.1 µl	2 U/reaction **
Template DNA	Depends on concentration	< 500 ng
Nuclease-free water	to 25 µl	

Table 3.33 3C-PCR validation of 4C Program used

Step	Temp	Time
Initial Denaturation	94°C	2 minutes
4-0 Cycles	94°C 60°C 72°C	30 seconds 30 seconds 30 seconds
Hold	4-10°C	

*For generating control template library, the extension time was increased to two minutes since the amplification fragments were 2000 bp long.

3.9.7- Agarose gel

To determine presence and size of the PCR products, PCR samples were run at 85 V for 45 min in 1X TAE electrophoresis buffer using a 1-2% agarose gel, depending on the size of the product, and containing SYBR Safe DNA gel dye. Size of the PCR product was determined by loading 7µl of DNA gene ladder.

3.10 - Statistical Analysis

3.10.1- RT-qPCR statistical analysis

The 2ΔCt method was used to calculate relative gene expression levels. Standard deviations were calculated from technical triplicate reactions and were represented as error bars.

3.10.2- ChIP-qPCR statistical analyses

Technical triplicates were used to calculate ChIP signals as percentage of input. Each ChIP sample was normalized to the average signals obtained in the same sample when using different control regions (see figures for more details). Standard deviations were calculated from technical triplicate reactions and represented as error bars.

3.10.3- RNA-seq analysis

RNA-seq data were analyzed using a high-throughput next-generation sequencing analysis pipeline (Wagle, Nikolić, and Frommolt 2015) (see recent publication for more details (Cruz-Molina et al. 2017)). In order to measure the genes differentially expressed between WT, PE Lhx5$^{-/-}$ and EED$^{-/-}$ AntNPCs, four biological replicates for WT, two replicates for two different PE Lhx5 (-109)$^{-/-}$ clones (clone one and two) and four biological replicates for EED$^{-/-}$ were used.

3.10.4- ChIP-seq analysis

ChIP-seq sequencing experiments were performed for P300 and various histone modifications in mESCs grown serum + LIF, AntNPCs and 2i mESCs. ChIP sequencing reads were mapped to the mouse genome (mm10 assembly) using BWA (Li et al. 2013). Afterwards, a bioinformatics pipeline was applied in order to identify genomic regions significantly enriched in the investigated histone marks and proteins in comparison to the total genomic input DNA (see (Cruz-Molina et al. 2017) for further details).

3.10.5- 4C-seq analysis

4C- PCR samples were sequenced at Erasmus Medical Center (Rotterdam) using HiSeq 2500 system. The reads were aligned to the mouse genome (mm9 assembly) using Bowtie (Langmead et al. 2009). To generate reads per million (RPM), the resulting mapped reads were evaluated with R3C-seq (Thongjuea et al. 2013). Information about 4C-seq library generation, analysis and quantification of 4C signals has been previously described (Cruz-Molina et al. 2017).

3.10.6- Public datasets and additional bioinformatics analysis

The following publically available ChIP-seq datasets were used in this study (Attanasio et al. 2014; Kagey et al. 2010; Nord et al. 2013; Chen et al. 2008; Ku et al. 2008; Ma and Wong 2011; Marson et al. 2008).

- In mESC: OCT4/POU5F1 (GSM307137), SOX2 (GSM307138, GSM3071389), NANOG (GSM307140, GSM307141), PRDM14 (GSM623989), RING1B (GSM656523), SUZ12 (GSM288360), EZH2 (GSM327668), MED1 (GSM560347), MED12 (GSM560345), SMC1 (GSM560341), SMC3 (GSM560343), CTCF (GSM560352). From E11.5.
- In E14.5 mouse embryos: E11.5 forebrain H3K27ac (GSM1264352), E14.5 forebrain H3K27ac (GSM1264356), E11.5 liver H3K27ac (GSM1264386), E14.5 liver H3K27ac (GSM1264388), E11.5 limb H3K27ac (GSM1371056).

To evaluate the *in vivo* enhancer activity of poised enhancers identified in mESCs, we compared our list of mESCs poised enhancers with highly conserved sequences included in the Vista Enhancer Browser (Visel et al. 2009) whose *in vivo* enhancer activity was previously evaluated in E11.5 mouse embryos. (See (Cruz-Molina et al. 2017) for additional details).

GREAT (McLean et al. 2010) was used to perform the *in-silico* functional annotation of different groups of enhancers .

Further details regarding the computational analyses used to identify AntNPCs superenhancers, identify motifs overrepresented among poised enhancers, determine the overlaps between CpGs and PEs, etc can be found in (Cruz-Molina et al. 2017).

3.10.7 -Data and Software Availability

The ChIP-seq, RNA-seq, and 4C-seq datasets reported in this dissertation have been deposited in the Gene Expression Omnibus (GEO) repository and are publically available under the accession number GSE89211.

3.10.8 – Software and algorithms

Table 3.34: Software and algorithms

Software	Source	Identifier
BWA	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
MACS2	Zhang et al., 2008	http://liulab.dfci.harvard.edu/MACS/
Galaxy	Afgan et al., 2016	http://main.g2.bx.psu.edu/
deep Tools	Ramírez et al., 2014	http://deeptools.ie-freiburg.mpg.de/
GREAT	McLean et al., 2010	http://bejerano.stanford.edu/great/public/html/
HOMER	Heinz et al., 2010	http://homer.salk.edu/homer/motif/
DAVID	Huang et al., 2009	https://david.ncifcrf.gov/
SeqPos motif tool	Liu et al., 2011	http://cistrome.dfci.harvard.edu/ap/
AME (Analysis of Motif Enrichment)	McLeay and Bailey, 2010	http://meme-suite.org/tools/ame
R3C-seq	Thongjuea et al., 2013	http://bioconductor.org/packages/release/bioc/html/r3Cseq.html
Bowtie	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/
seqPattern	Bionconductor	http://bioconductor.org

4- RESULTS

4.1- Identification and characterization of PEs in mESCs

Different classes of enhancers can be identified based on unique chromatin signatures. As first step to identify active, primed and poised enhancers in mESCs, we performed ChIP-seq experiments in mESCs grown under serum plus LIF for H3K4me1, H3K4me3, H3K27me3, H3K27ac and P300 (see methods 3.5.1). Using the criteria described in Table 1.1, (Cruz-Molina et al. 2017; Rada-Iglesias et al. 2011), three non overlapping set of enhancers were defined: 12142 active enhancers, 19793 primed enhancers and 1015 poised enhancers (Figure 4.1).

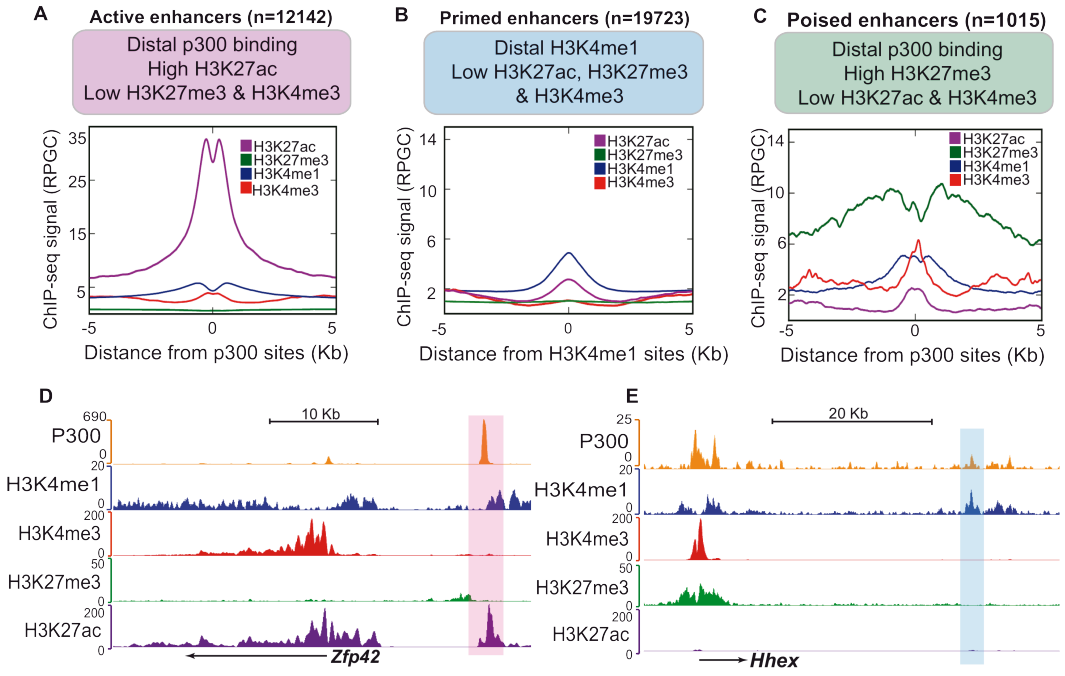

Figure 4.1: Identification of active, primed and poised enhancers in mESCs grown under serum plus LIF. (A-C) Average ChIP-seq signals for H3K27me3, H3K27ac, H3K4me1 and H3K4me3 are showed around **(A)** 12142 active enhancers **(B)** 19723 primed enhancers and **(C)** 1015 poised enhancers. (**D-E)** Genome browser representation of P300, H3K4me1, H3K4me3, H3K27me3 and H3K27ac ChIP-seq profiles in mESC are shown around **(D)** a representative active enhancer (covered in pink) located five prime of *Zfp42* and (**E)** a primed enhancer located downstream of the *Hhex* gene covered in blue.

4.1.1- PE signature in pluripotency

In order to identify PEs in ESC, mESCs were used as a model due to the fact of their high genetic tractability and clonogenicity. In mESCs several pluripotency states have been defined. Epiblast cells within the mouse pre-implantation embryo display naïve or ground state pluripotency, which can be maintained *in vitro* using the so-called 2i conditions (mixture of GSK3 and MEK inhibitors) (see methods 3.32 and 3.33). Naïve pluripotency is characterized by an homogeneus gene expression profile, DNA hypomethylation and low levels of H3K27me3 (Galonska et al. 2015; Guo et al. 2016; Marks et al. 2012). Upon implantation, the epiblast cells progress towards pluripotent states with increasing signs of developmental priming and incipient expression of somatic lineage specifiers. *In vitro*, these primed pluripotent states are represented by epiblast-like cells (EPiLCs) and epiblast stem cells (EpiSCs) (Smith 2017; Hackett and Surani 2014). mESCs grown under serum+LIF conditions represent a more metastable state in which cell can switch between naïve and primed pluripotency and, consequently, display heterogenety in the gene expression profiles (Boroviak et al. 2015; Kalkan and Smith 2014; Ying et al. 2008).

In principle, this heterogeneity found in serum+LIF conditions could contribute to the emergence of the poised enhancer signature in which both active and repressive chromatin features are found at the same genomic loci. Hence, to exclude this possibility, ChIP-seq experiments for P300, H3K4me1, H3K27me3, H3K27ac and H3K4me3 were performed in 2i mESCs to determine if PEs identified in serum+LIF mESCs show any differences in their chromatin features under naïve pluripotency conditions. Importantly, PEs identified in serum+LIF mESCs showed a similar chromatin signature in the naïve state, with somehow decreased P300 and H3K27me3 levels (Figure 4.2A, 2B and 2C). Additionally, using a more relaxed peak-calling criteria to analyze the 2i mESCs ChIP-seq data resulted in 54%, (550/1015) of all serum+LIF mESCs PEs being also identified in 2i mESC, with 85%of them (857/1015) being enriched in H3K27me3 (Cruz-Molina et al. 2017). These

results suggest that PEs in serum+LIF mESCs are not the result of cellular heterogenety.

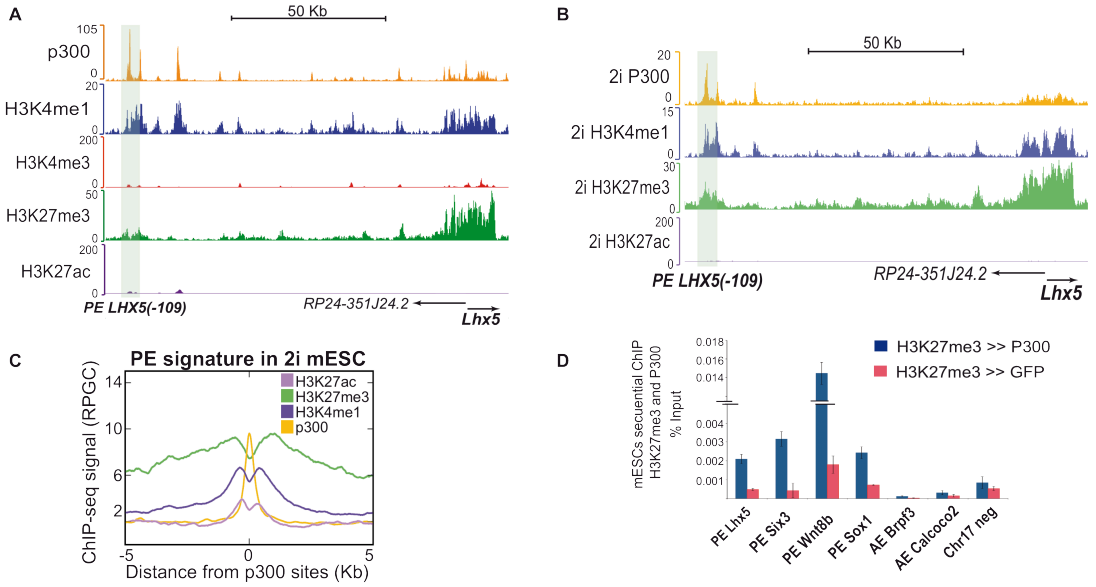

Figure 4.2: Chromatin signature of PEs in naïve mESCs growing in 2i medium. (A-B) P300, H3K4me1, H3K4me3, H3K27me3 and H3K27ac ChIP-seq profiles are shown around a representative poised enhancer (shaded in green) for mESCs grown under serum+LIF conditions **(A)** and mESCs grown in 2i medium (**B). (C)** Average ChIP-seq signals for P300, H3K27me3, H3K27ac and H3K4me1 in 2i mESCs around the PEs identified in serum+LIF mESCs. (**D)** qPCR analysis of sequential ChIP experiments performed in mESCs. Anti-H3K27me3 was used as a primary antibody, followed by immunoprecipitation with anti-P300 (blue) or anti-GFP (red) as secondary antibodies. Four selected PEs (PE *Lhx5*(-109), PE *Six3(*-133), PE *Sox1*(+35), PE *Wnt8b*(+21)), two active enhancers (AE Calcoco2, AE BRPF3) and an intergenic negative control region ("chr17_neg") were analyzed . The error bars represent standard deviation from three technical replicates.

To further demonstrate that the PE chromatin signature does not arise as a consequence of the cellular heterogeneity, sequential ChIP-qPCR was performed, in serum+LIF mESCs using the H3K27me3 and P300 antibodies. These experiments demonstrated that P300 and H3K27me3 simultaneously bound the analyzed PEs candidates in serum+LIF mESC and therefore these marks do not appear as a product of heterogeneity (Figure 4.2D).

4.1.2- Distinctive features of mESCs primed, poised and active enhancers in mESCs

In silico functional annotation of active, primed and poised enhancers was performed using a software called GREAT as it was previously described (McLean et al. 2010; Cruz-Molina et al. 2017). GREAT associates non-coding genomic regions with nearby genes and then statistically evaluates the over-representation of such genes with respect to different gene ontologies and functional categories. These *in silico* analyses confirmed previous observations (Rada-Iglesias et al. 2011; Zentner et al. 2011), each enhancer class associates with a different set of genes. Active enhancers in mESCs were associated to pluripotency genes and also genes expressed during mouse pre-implantation development (Figure 4.3A). Primed enhancers were linked with genes expressed and involved in a broad and unrelated set of developmental stages and biological processes (Figure 4.3B). In contrast, PEs were strongly linked with genes involved in early organogenesis and the establishment of the basic body plan, with a clear bias towards genes expressed and implicated in the development of the future brain (Figure 4.3C).

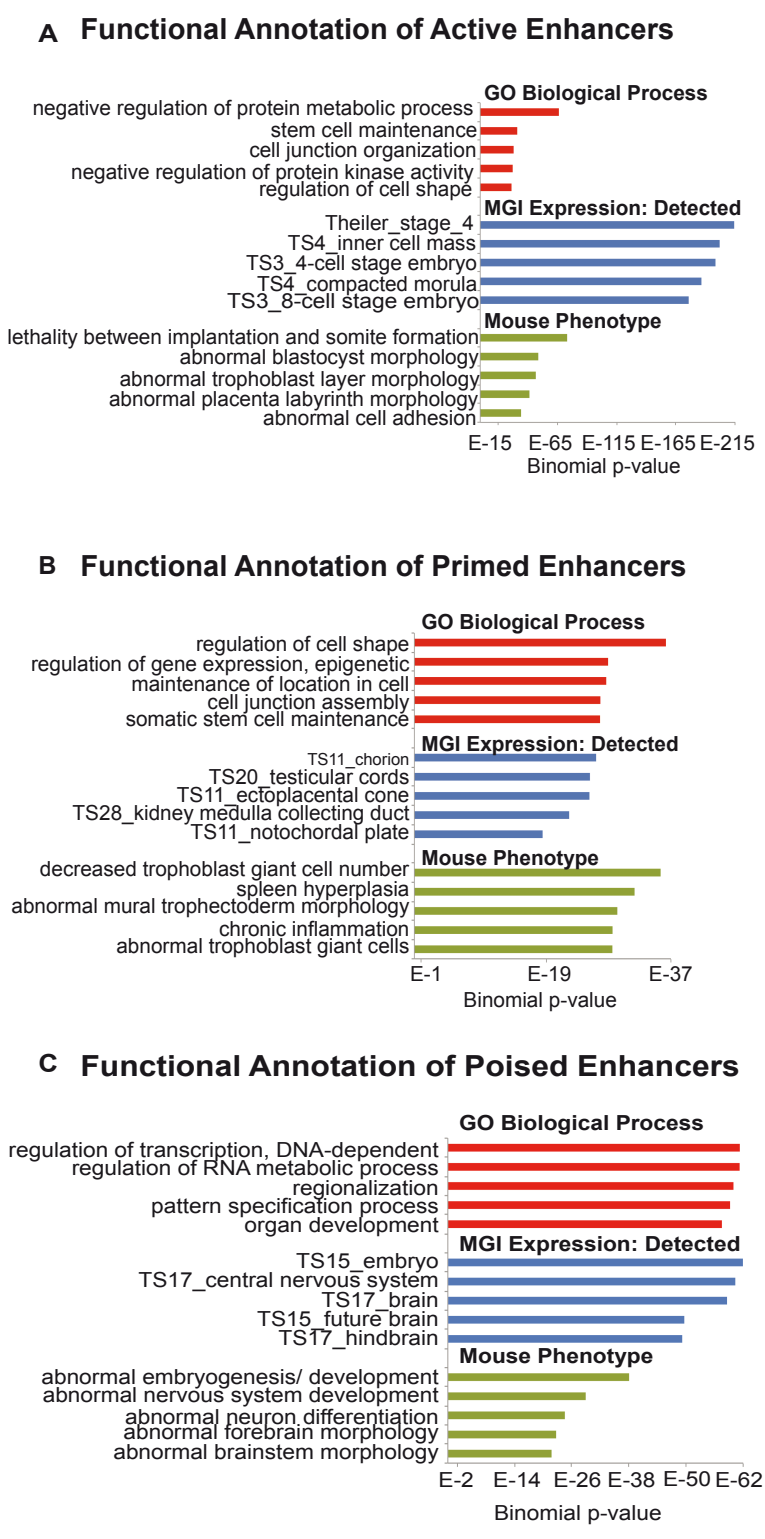

Figure 4.3: Functional annotation of active primed and poised enhancers in mESCs. Active **(A)**, primed **(B)** and poised **(C)** enhancers were functionally annotated *in silico* using GREAT (McLean et al. 2010). From the ten most represented terms of the following gene ontologies (GO), five terminologies are shown: Mouse Phenotype (green), Mouse Genome Informatics (MGI) expression (blue) and Biological Process (red).

The previous bias towards anterior neural identity genes was not exclusive to mESCs PEs, since similar results were obtained when hESCs PEs were re-annotated with updated versions of GREAT (Rada-Iglesias et al. 2011). To test wether this preferential association of PEs with genes involved in brain

development was not simply the result of an *in silico* artifact, mESCs PEs were compared to H3K27ac peaks identified in mouse forebrain, limb and liver tissues isolated from mouse embryos at embryonic days 11.5 (E11.5) and 14.5 (E14.5) (Attanasio et al. 2014; Nord et al. 2013). These H3K27ac peaks should roughly reflect the location of active enhancers in these tissues. Almost 50% of the PEs overlapped with H3K27ac peaks present in at least one embryonic mouse tissue and, remarkably, 70% of them were only found in the embryonic brain (Figure 4.4A), which was in full agreement with the previous in silico observations. Additionally, in order to check if our PE sequences have enhancer activity, they were compared with the Vista enhancer database, which contains a set of highly conserved sequences whose *in vivo* enhancer activity has been investigated in mouse embryos at E11.5. 25 of our PEs overlapped with sequences found in the Vista enhancer database and 88% of them displayed enhancer activity in E11.5 mouse embryos, mostly within the developing brain (Figure 4.4B). Next, sequence conservation analysis revealed that mESC PEs present higher conservation levels among vertebrates than active or primed enhancer sequences. These results were in agreement with previous reports in hESCs (Rada-Iglesias et al. 2011) (Figure 4.4C) and with the potential relevance of PEs during the arrangement of differentiation gene expression programs.

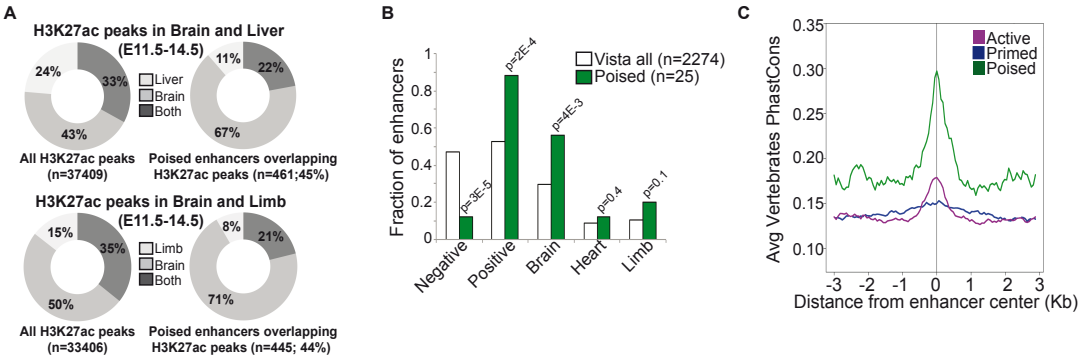

Figure 4.4: Poised enhancers are conserved elements that preferentially become active during brain development (A) H3K27ac peaks identified in mouse embryonic tissues at E11.5 and E14.5 (Attanasio et al. 2014; Nord et al. 2013). A combination of H3K27ac peaks identified in brain and limb (bottom panels) or in brain and liver (top panels) is shown for all the peaks (left) and peaks coinciding with PEs (right). **(B)** Graphic presenting the fraction of sequences showing *in vivo* enhancer activity in mouse E11.5 embryos according to five Vista Enhancer database

categories defined based on the tissues where enhancer activity was detected. The relative abundance of all Vista enhancer sequences (n = 2,274, white) within each category was compared to Vista enhancer sequences overlapping with PEs (n = 25, green). Hypergeometric tests were used to calculate p values. **(C)** Average vertebrate PhastCons signal profiles were created around the central position of mESCs active (purple), primed (blue) and poised (green) enhancers.

Additionally, RNA-seq experiments in mESCs grown under serum+LIF were performed. The expression levels of the genes associated with active, primed and poised enhancers were in accordance with the in silico functional annotation results described above (Figure 4.3). Genes linked to PEs showed lower expression in mESC than genes associated to either active or primed enhancers (Figure 4.5A). Using publically available ChIP-seq data for H3K27me3 in mESCs (Bernstein et al. 2006), it was found that around 40% of the PEs putative target genes were enriched in this histone mark, which also agrees with the overall low expression of this set of genes (Figure 4.5B). Moreover, it has been reported that PEs could be bound by TFs in mESCs (Rada-Iglesias et al. 2011; Zentner et al. 2011). In agreement with these studies, we found that PEs and active enhancers in mESCs were both bound to a similar extent and by pluripotency TFs while primed enhancers were not (Figure 4.5C). In summary, the data demonstrate that PEs are a set of higly conserved regulatory elements preferentially associated with genes involved in the establishment of anterior neural identity.

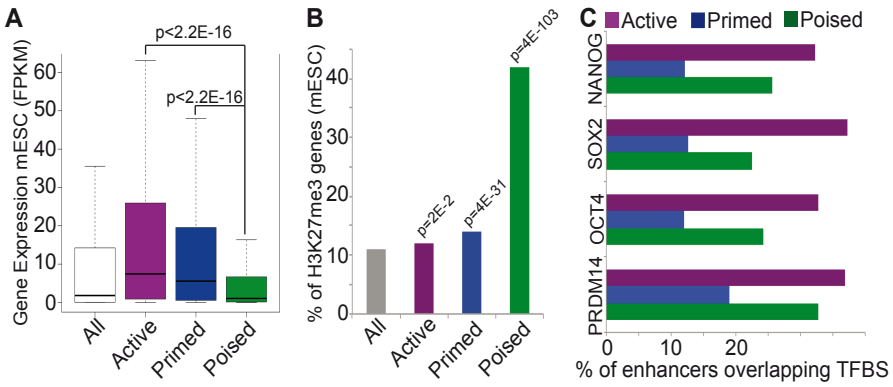

Figure 4.5: Poised enhancers in mESCs are bound by pluripotency TFs and are also linked to inactive developmental genes. (**A)** RNA-seq data generated in serum+LIF mESCs was used to calculate gene expression values (measured as FPKMs) and presented as boxplots for (All) all genes, (Active) genes linked to active enhancers, (Primed) genes linked to primed enhancers and (Poised) genes linked to

poised enhancers. Unpaired Wilcoxon tests were used to calculate *P*-values. (**B)** For the gene categories described in A, the percentage of genes whose promoters are marked by H3K27me3 according to (Bernstein et al. 2006) is shown. (**C)** Percentage of Active, Primed or Poised enhancers bound by the indicated pluripotency transcription factors according to publically available ChIP-seq data (see Methods for references).

4.2- Establishment of a differentiation model for mESCs

Previous studies and our own recent data suggest that poised enhancers might control the induction of genes responsible of the earliest steps of somatic differentiation especially towards the anterior neural fate (Zentner et al. 2011; Cruz-Molina et al. 2017; Rada-Iglesias et al. 2011) (Figure 4.3C, 4.4A and 4.4B). Therefore, we examined the functional relevance of PEs upon differentiation of mESC into anterior neural progenitors (AntNPCs). This proved to be quite challenging, since in our hands, previously reported differentiation protocols were inefficient and resulted in heterogeneous cell populations that included AntNPCs and non-neural cells (Gaspard et al. 2009). Finally, we adapted a recently described differentiation protocol (Gouti et al. 2014) that with slight modifications, enabled us to obtain AntNPCs in five days (Methods 3.3.4). Briefly, mESCs were seeded at 10000 cells/cm^2 on plates coated with 0,1% gelatin and grown for five days in N2B27medium. The first three days, N2B27 medium was supplemented with 10 ng/ml of b-FGF, which helps inducing a postimplantation epiblast-like state. In order to improve the purity of the differentiation, 5μM Xav939 was added to the medium from day two to day five. This compound is a powerful WNT signaling inhibitor which was reported to facilitate the anterior neural differentiation (Matsuda and Kondoh, 2014) (Figure 4.6).

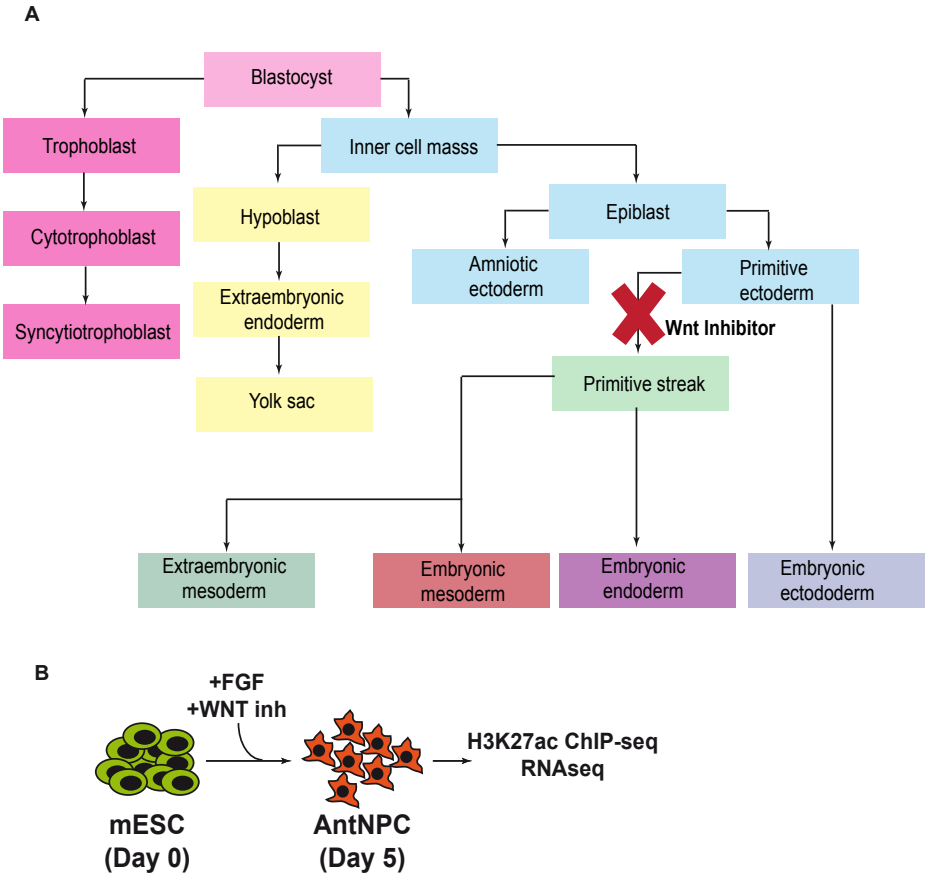

Figure 4.6: Major embryonic and extra-embryonic lineages during mammalian development. (A) In order to differentiate mESCs, which resemble the pluripotent epiblast, into AntNPCs, we used XAV939, a potent inhibitor of WNT signaling. WNT signaling is known to promote the formation of the primitive streak *in vivo* and thus necessary for the differentiation of epiblast cells into the mesodermal and endodermal lineages (adapted from (Tan, Gilmore, and Baris 2013). **(B)** Differentiation protocol followed to identify enhancers that become active in AntNPCs.

To test if the new differentiation protocol was efficient, the expression levels of anterior and posterior neural genes, pluripotency markers and markers from non-neural lineages were measured by RT-qPCR and RNA-seq (Figure 4.7A). In addition, protein levels of pluripotency (OCT4) and anterior neural markers (LHX5, NESTIN and PAX6) were also analyzed by immunofluorescence (Figure 4.7B). All these experiments conclusively showed that our new differentiation protocol resulted in high expression of anterior neural genes, while the pluripotency and non-neural markers were barely expressed. Therefore we concluded that this new protocol efficiently derives AntNPCs.

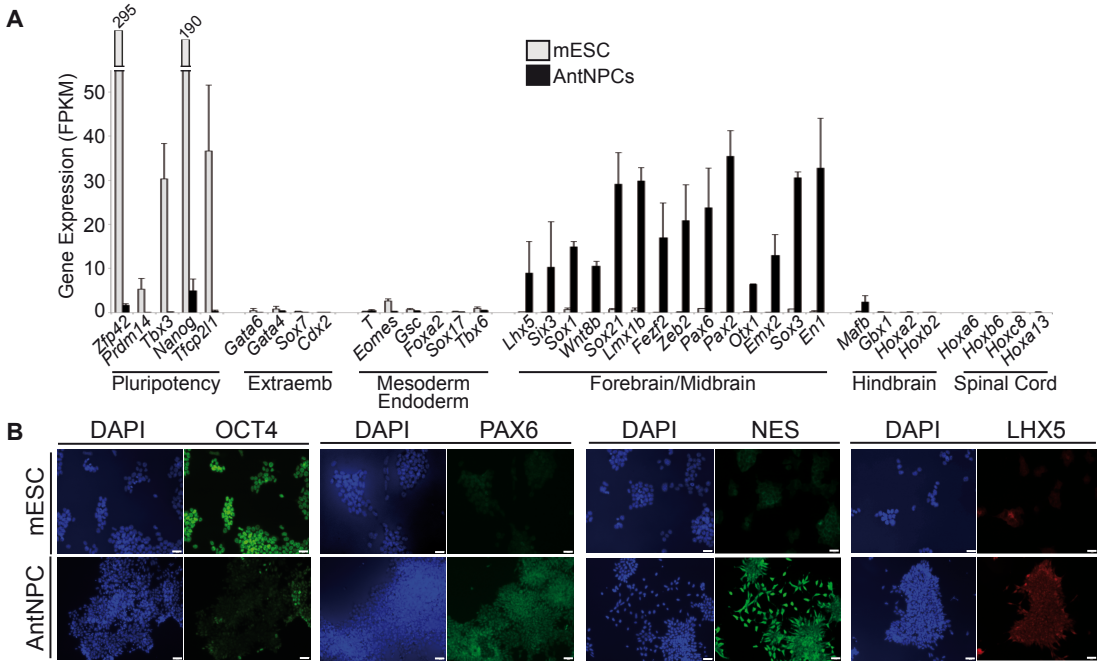

Figure 4.7: Characterization of AntNPCs. (A) Gene expression levels (as FPKMs) were obtained by by RNA-seq in mESC and AntNPCs. Upon differentiation, there was a decrease in the expression of pluripotency markers, while anterior neural markers (forebrain/ midbrain) increased. Posterior neural genes (hindbrain, spinal cord) and markers of non-neural lineages (extraembryonic tissues, mesoderm, endoderm) were barely expressed in both conditions. **(B)** Representative images of immunofluorescence experiments performed for OCT4, PAX6, NESTIN and LHX5 in mESCs and AntNPCs. White bar represents 20 μm scale.

4.3 – Identification of PEs that become active during anterior neural differentiation

In order to identify mESCs PEs that become active in AntNPCs, H3K27ac, H3K27me3 and P300 ChIP-seq experiments were performed in AntNPCs (Figure 4.8). 228 out of the 1015 PEs identified in mESCs gained H3K27ac in AntNPCs and were considered as active in these cells (PoiAct enhancers). The remaining PEs that did not gain H3K27ac, retained high levels of H3K27me3 and lost P300. We speculate that most of these PEs, which we termed as PoiNoAct enhancers, transitioned to an inactive state or remained poised for activation at a later time point or different embryonic context.

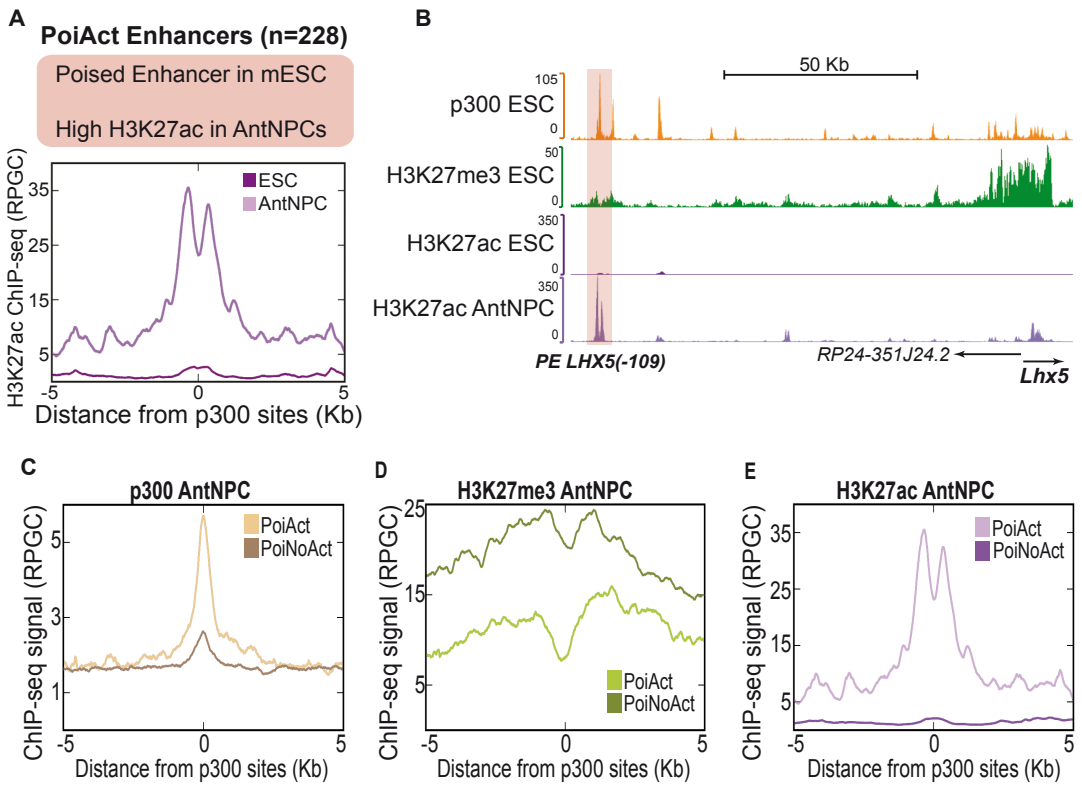

Figure 4.8: PEs become active during anterior neural differentiation. (**A)** Average H3K27ac ChIP-seq signals in mESCs and AntNPCs around PoiAct enhancers. **(B)** Genome browser view of a representative PE (shaded in pink) showing P300, H3K27me3 and H3K27ac levels in mESCs and H3K27ac levels in AntNPCs. Both, the PE and its putative target gene (*Lhx5*) gained H3K27ac in AntNPCs. **(C, D, E)** Average ChIP-seq signal profiles for P300 **(C),** H3K27me3 **(D)** and H3K27ac **(E)** in AntNPCs around PoiAct and PoiNoAct enhancers.

4.3.1- Functional annotation and characterization of PoiAct enhancers

To validate the relevance of PEs in the establishment of anterior neural gene expression programs, the previous set of PoiAct enhancers was subject to additional characterization. First, PoiAct enhancers were functionally annotated using GREAT showing that genes associated with these enhancers are mostly involved in early brain development (Figure 4.9A). Then, PoiAct enhancers were compared with H3K27ac peaks identified in different mouse embryo tissues, which revealed that the majority (approx. 70%) of these enhancers overlapped H3K27ac peaks found in the developing mouse brain in comparison to the other considered tissues (Figure 4.9B). Afterwards, the expression of the PoiAct enhancer's closest genes was also evaluated in mESCs and AntNPCs, which showed a significant up-regulation of these

genes in AntNPCs (Figure 4.9C). Finally, to confirm that PoiAct sequences possess enhancer activity, they were compared with the Vista enhancer database. Importantly, all PoiAct elements found in the Vista Enhancer database displayed enhancer activity *in vivo*, mostly within the developing brain (Figure 4.9D and E). Lastly, the enhancer activity of PoiAct enhancers in anterior neural cells was also confirmed using *in vitro* reporter assays, where the enhancer sequence was cloned in a vector under the presence of a GFP driving promotor (see methods 3.8). The outcome of these experiments showed clear levels of fluorescence in AntNPCs for the PE secuences studied (Figure 4.9F).

From the list of 228 PoiAct enhancers, six candidates were selected for additional functional dissection based on their relative proximity to major anterior neural regulators, high levels of sequence conservation and presence of H3K27ac peaks in both AntNPCs and the embryonic mouse brain. PEs were named based on the relative distance to their closest genes downstream of the TSS (+) or upstream of TSS (-) (Table 4.1).

Table 4.1: PEs selected for further characterization.

PE identification	Closest gene	Distance TSS (Kbp)
PE *Lhx5* (-109)	*Lhx5*	-109
PE *Six3* (-133)	*Six3*	-133
PE *Sox1* (+35)	*Sox1*	+35
PE *Wnt8b* (+21)	*Wnt8b*	+21
PE *Sox 21* (+6.5)	*Sox21*	+6.5
PE *Lmx1b* (+59)	*Lmx1b*	+59

Taken together, all these experiments visibly showed that the investigated PoiAct sequences had considerably higher enhancer activity in AntNPCs than in mESCs.

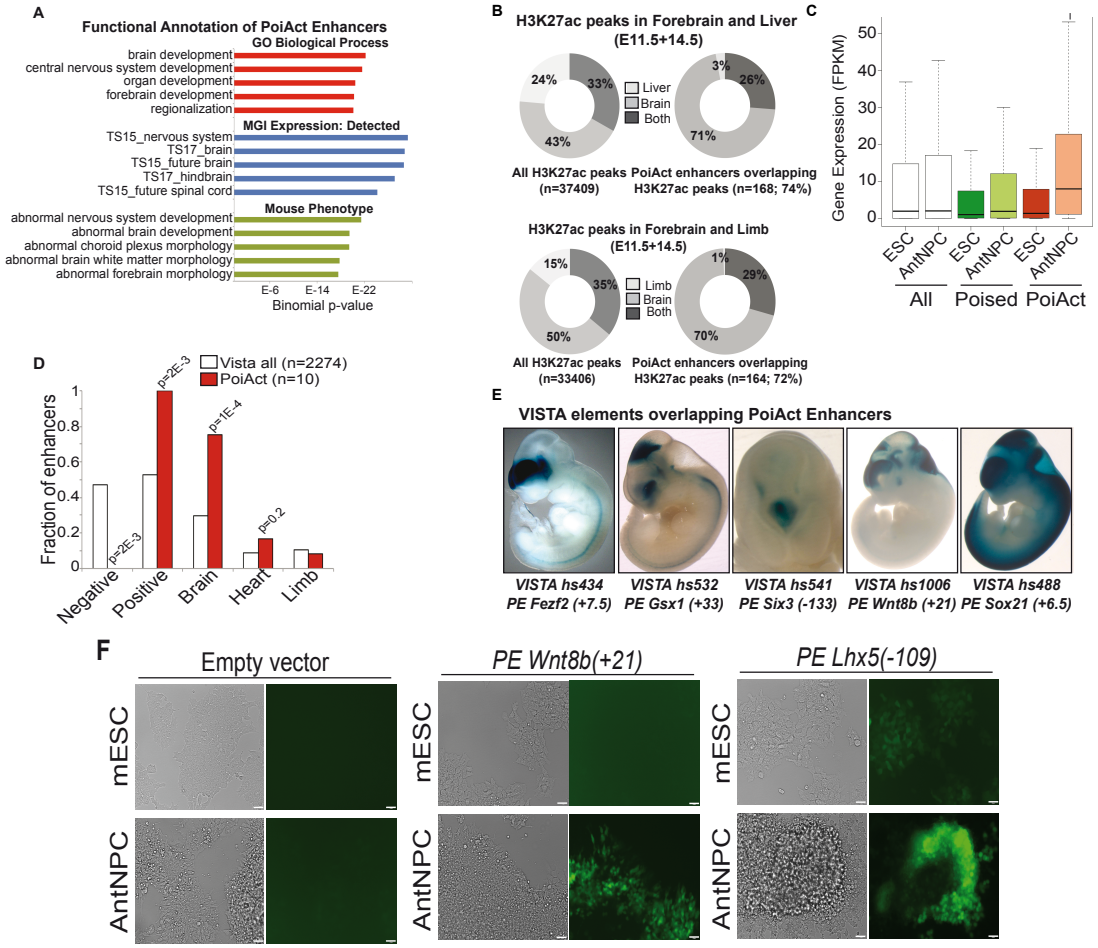

Figure 4.9: PoiAct enhancers are preferentially associated with anterior neural development and display *in vivo* and *in vitro* enhancer activity. (**A**) Functional annotation of PoiAct enhancers using GREAT was performed as described in figure 4.3. **(B)** H3K27ac peaks identified in mouse embryonic forebrain, limb and liver at E11.5 and E14.5 were combined. The percentages of H3K27ac peaks observed in the indicated tissues are shown for all combined H3K27ac peaks (left) or for H3K27ac peaks overlapping with PoiAct enhancers (right). **(C)** Gene expression values (as FPKM) were measured by RNA-seq in mESCs and AntNPC for all genes (All), genes linked to PEs (Poised), and genes linked to PoiAct enhancers. P values were obtained by Paired Wilcoxon tests. **(D)** Sequences from the Vista Enhancer browser were categorized into five different sets (explained in figure 4.4B). The relative abundance of all Vista Enhancer sequences (n= 2.274, white) within each category was compared to Vista Enhancer sequences overlapping with PoiAct enhancers (n=10, red). P-values were calculated by hypergeometric tests. **(E)** Representative examples of the *in vivo* enhancer activity of Vista enhancer sequences overlapping PoiAct enhancers in E11.5 mouse embryos (Visel, Rubin, and Pennacchio, 2009).**(F)** Reporter assays performed in mESCs for two selected PEs regions, PE *Lhx5* (-109) and PE *Wnt8b* (+21) and for a control, an empty vector which did not carry any PE sequence. GFP fluorescence levels were measured in mESCs and in AntNPCs. White bars represent 20µm scale.

4.4- Identification of the putative target genes of selected PEs based on physical interactions

To identify the putative target genes of the PE candidates, 4C-seq experiments were performed using PEs as viewpoints (see introduction 1.2.1 and methods 3.7). From the six PE candidates (Table 4.1), 4C-seq experiments were only performed for those four located more distally from their closest gene (i.e. > 30 Kb), since interactions between closely located loci can not be accurately quantified by 4C-seq. Therefore we analyzed PE *Lhx5* (109), PE *Six3* (-133), PE *Sox1* (+35) and PE *Lmx1b* (+59) (Figure 4.10).

As expected, 4C-seq results showed that PEs contacted their predicted target genes in AntNPCs when they were already active as it is shown by the pink peaks representing 4C contacts (Figure 4.10). Remarkably, such contacts were also observed in mESCs as it is ilustrated by the brown peaks (Figure 4.10). Therefore the contacts precede the activation of both PEs and their target genes which suggests the existence of a pre-formed regulatory topology at PE loci. Furthermore, neither short-range interactions with other flanking genes nor long-range interactions with more distally located genes were observed in mESCs or AntNPCs, as it is shown by the lack of contacts in these regions (Figure 4.10).

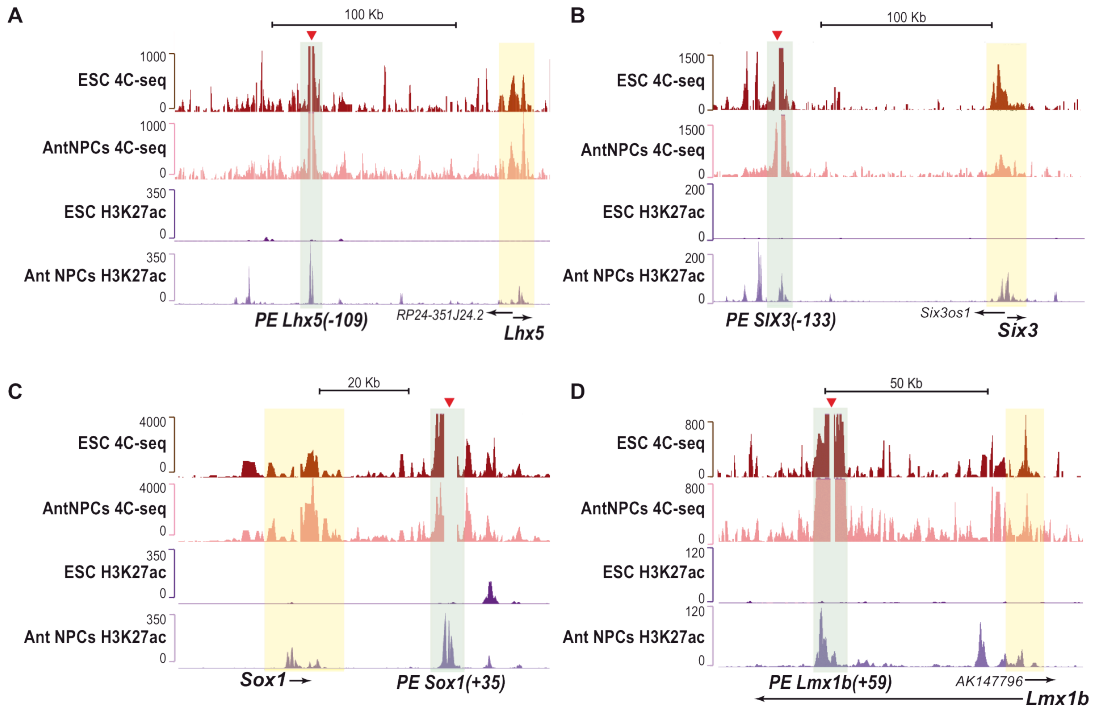

Figure 4.10 Poised enhancers already contact their target genes in mESCs. 4C-seq and H3K27ac ChIP-seq profiles created in mESCs and AntNPCs are shown around PE *Lhx5* (-109) **(A),** PE *Six3* (-133) **(B)**, PE *Sox1* (+35) **(C)**, and PE *Lmx1b* (+59) **(D).** PEs are shaded in green while their putative target genes are in yellow. PEs. Were used as viewpoints (red triangules) in all 4C-seq experiments.

4.5- PEs are essential for the proper induction of major anterior neural genes

In order to determine whether or not PEs are important for proper induction of their putative target genes, we decided to create PE deletions in mESCs using CRISPR-Cas9 (see introduction 1.1.2 and methods 3.6). Briefly, two gRNAs flanking the PE candidates were generated in order to recruit the Cas9 nuclease and generate double-stranded DNA breaks at these locations. Subsequently, these DNA breaks can be repaired by non-homologous end-joining repair (NHEJ), which can frequently lead to deletion of the intervening DNA sequences (i.e the PE candidate). mESCs were co-transfected with vectors co-expresing Cas9 and the pair of gRNAs flanking each PE candidate. Afer 48 hours of drug selection with puromycin, mESCs were seeded at clonal density, expanded and genotyped by PCR in order to identify

mESCs clonal lines with the desired PE deletions (Figure 4.11). At least one mESC clonal line with a homozygous deletion (PE $^{-/-}$) of each PE candidate was obtained, with the exception of PE *Lmx1b*, which was no longer investigated. Moreover, a second mESC clonal line with a homozygous PE deletion was obtained for PE *Lhx5* (-109), PE *Sox1* (+ 35), PE *Sox 21* (+6,5)). For PE *Six3* (-133) and *Wnt8b* (+21) an additional line with a heterozygous PE deletion (PE $^{+/-}$) was generated (Table 4.2).

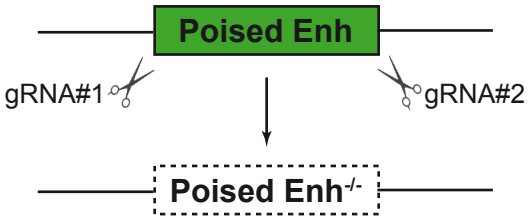

Figure 4.11: Strategy followed to delete PE regions by CRISPR-Cas9

Table 4.2: List of generated mESC lines with PE deletions.

PE identification	Deletion Size (Kbp)	Homozygous clones	Heterozygous clones
PE *Lhx5* (-109)	0.5	2	0
PE *Six3* (-133)	5.2	1	1
PE *Sox1* (+35)	5.5	2	0
PE *Wnt8b* (+21)	1.2	1	1
PE *Sox 21* (+6.5)	1.4	2	0

In order to confirm that the selected mESC clonal lines carried the intended PE deletions, PCRs-based genotyping was used followed by Sanger sequencing (Figure 4.12, see methods 3.9.4). The PE *Lhx5* (-109) and PE *Wnt8b* (+21) deletions were small in size (< 2000bp) and a single pair of primers could be used to detect both the WT and the deletion alleles (Figure 4.12A). On the other hand, the PE *Six3* (-133), PE *Sox1* (+35) and PE *Sox21*

(+6,5) deletions were larger and two different primer pair combinations were used to genotype the corresponding mESC clonal lines (Figure 4.12B).

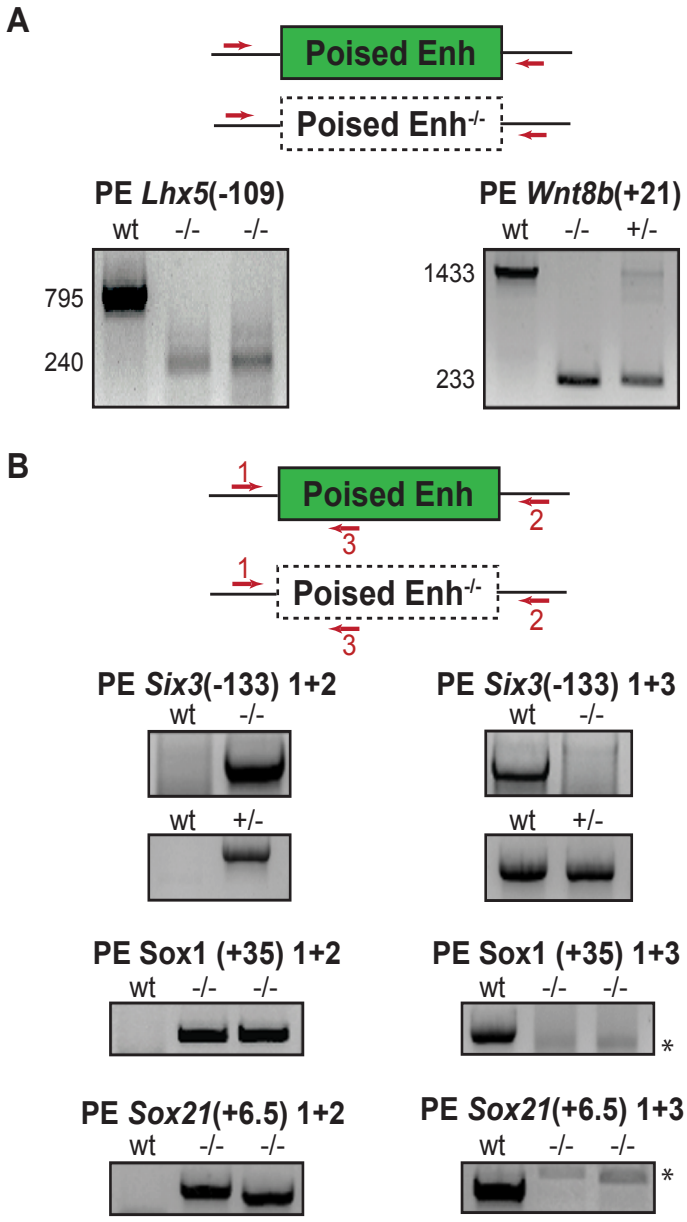

Figure 4.12: Genotyping of mESC clonal lines with PE deletions: (A) For PE *Lhx5* (-109) and PE *Wnt8b* (+21), primer pairs flanking the deleted regions were used. For PE *Lhx5* (-109), the PCR results show two mESC lines with homozygous PE deletions (-/-). For PE *Wnt8b* (+21), two mESC lines with an homozygous (-/-) or heterozygous(+/-) deletion of the PE are shown. **(B)** Two sets of primer pairs were used (1+2, 1+3) to genotype mESC lines with deletions for for PE *Six3* (-133), PE *Sox1* (+35) and PE *Sox21* (+6.5). Depending on the primer combinations used, DNA products specific to sequences that were either WT (1+3) or contained PE deletions (1+2) were amplified. For PE *Sox1* (+35) and PE *Sox21* (+6.5), the PCR results obtained for two mESC lines with homozygous PE deletions (-/-) are shown. For PE *Six3* (-133), two mESC lines with a homozygous (-/-) or heterozygous (+/-) deletion of the PE are shown. * indicates unspecific PCR products.

Afterwards, all the mESC lines with validated PE deletions were differentiated into AnNPCs and the expression of their target genes was evaluated by RT-qPCR and compared to wild type (WT) parental cells (Figure 4.13). In undifferentiated mESC the expression of the PEs target genes was low and almost the same in wild-type mESCs and mESC lines with PE deletions (Figure 4.13). Importantly, these results indicate that PEs are not necessary to keep the expression of their target genes in an inactive state and, therefore,

they do not act as a silencers in mESCs as it has been previously proposed (Bonn et al. 2012; Entrevan et al. 2016; Spitz and Furlong 2012; Kondo et al. 2014). In contrast, a major decrease in the expression of the PE target genes was observed upon differentiation of the mESC with PE deletions into AntNPC (Figure 4.13). For the mESC lines with, homozygous PE deletions, 96% reduced induction for *Lhx5*, 84% for *Sox1*, 78% for *Six3*, 91% for *Wnt8b* and 27% for *Sox21* was observed with respect to their parental WT cells (Figure 4.13F, G, H, I and J). For each $PE^{-/-}$ mESC line, the expression of its target gene was measured in at least two independent AntNPC differentiations and it was compared with WT AntNPCs differentiated in parallel. As an additional validation of the previous results, a second $PE^{-/-}$ mESC clonal line was also differentiated into AntNPC, which confirmed the strong effects of the PE deletions on the induction of their target genes (Figure 4.13). For those PEs for which we could not generate a second mESC line with an homozygous deletion (i.e PE *Wnt8b* (+25) and PE *Six3* (-133)), we used mESC lines with heterozygous PE deletions instead. Upon AntNPC differentiation, these $PE^{+/-}$ mESC lines showed reduced expression of their target genes by roughly 50% compared to WT cells (44% reduction for *Wnt8b* and 52% of reduction for *Six3* in PE Wnt8b $(+25)^{+/-}$ and PE *Six3* $(-133)^{+/-}$ respectively) (Figure 4.13 G and J).

To corroborate that the effects of the PE deletions on the expression of their target genes were not due to general differentiation defects, the expression of *Zfp42*, a pluripotency gene was always evaluated in WT and $PE^{-/-}$ cells. Downregulation of *Zfp42*, was continuously observed upon differentiation of both WT and $PE^{-/-}$ mESCS, proving that the decreased induction of the PE target genes was caused to the specific PE deletions (Figure 4.13 F, G, H, I and J).

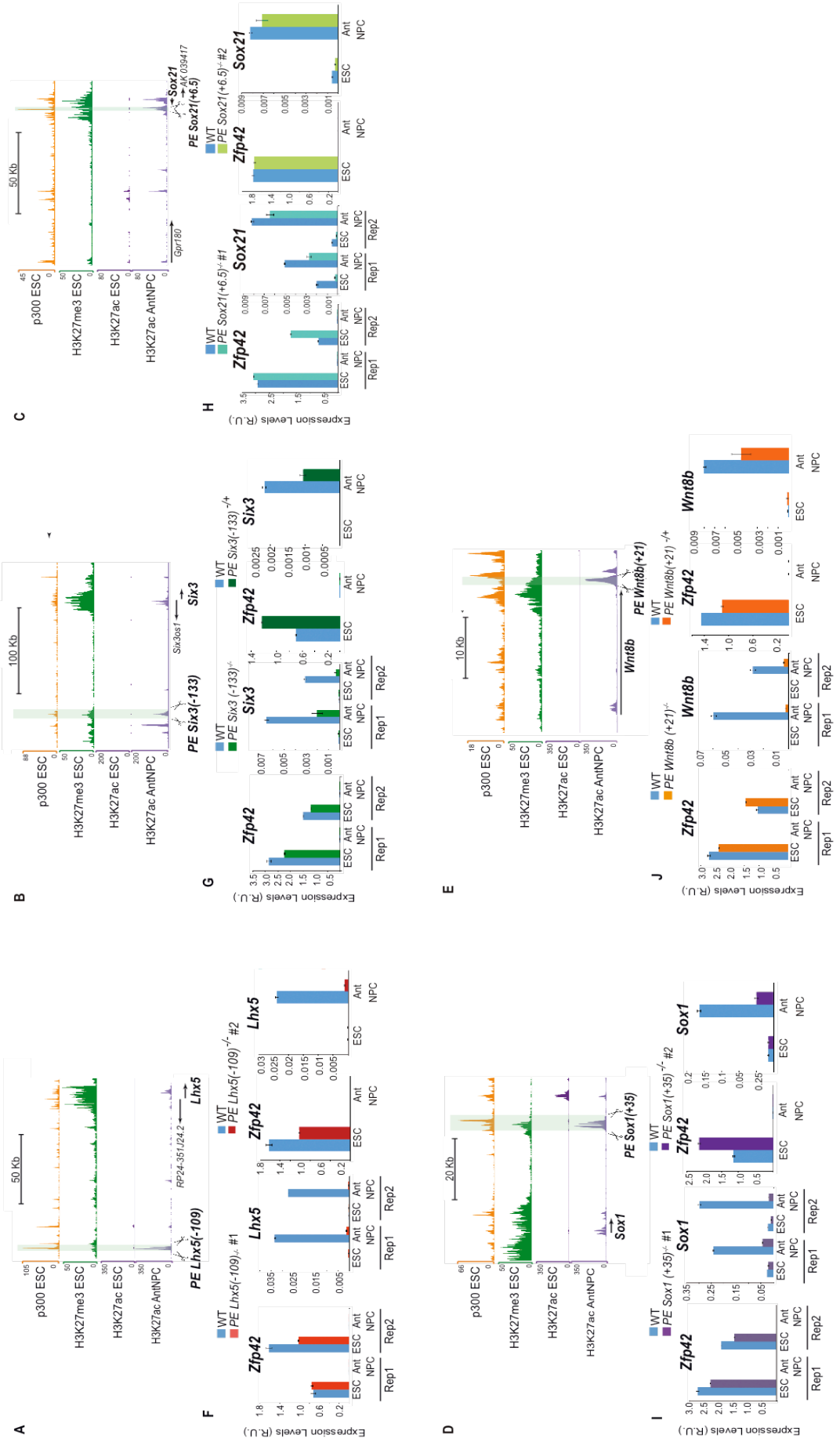
A
B
C
D
E
F
G
H
I
J
p300 ESC
H3K27me3 ESC
H3K27ac ESC
H3K27ac AntNPC
50 Kb
100 Kb
20 Kb
10 Kb
PE Lhx5(-109)
PE Six3(-133)
PE Sox21(+6.5)
PE Sox1(+35)
PE Wnt8b(+21)
Lhx5
Six3
Sox21
Sox1
Wnt8b
Zfp42
Gpr180
Six3os1
RP24-351J24.2
AK039417
WT
PE Lhx5(-109)-/- #1
PE Lhx5(-109)-/- #2
PE Six3(-133)-/+
PE Six3 (-133)-/-
PE Sox21(+6.5)-/- #1
PE Sox21(+6.5)-/- #2
PE Sox1 (+35)-/- #1
PE Sox1(+35)-/- #2
PE Wnt8b (+21)-/-
PE Wnt8b(+21)-/+
ESC
Ant
NPC
Rep1
Rep2
Expression Levels (R.U.)

Figure 4.13: PEs are necessary for the induction of their target genes. (A, B, C, D and E) Genome browser captures of p300, H3K27me3, and H3K27ac ChIP-seq profiles in mESCs and of H3K27ac in AntNPCs are shown around PE *Lhx5* (-109) **(A**), PE *Six3* (-133) **(B)**, PE *Sox21* (+6,5) **(C)**, PE *Sox1* (+35) **(D)**, and PE *Wnt8b* (+21) **(E)**. The deleted regions are shaded in green. **(F, G, H, I and J)** The expression of the putative PE target genes (*Lhx5* in **F** *Six3* in **G**, *Sox21* in **H**, *Sox1* in **I** and *Wnt8b* in **J)**. and of the *Zfp42* pluripotency marker was measured by qRT-PCR in mESCs and AntNPCs that were either WT (blue), homozygous or heterozygous for a deletion of the indicated PEs. For each panel and from left to right, the first two graphs show the results obtained using mESC lines with PE homozygous deletions in two biologically independent differentiation experiments (Rep1 and Rep2). The last two graphs show the results obtained with a second mESC line for each PE candidate and that contained either homozygous PE *Lhx*5 (-109) **(F)**, PE *Sox21* (+6,5) **(H)** and PE *Sox1* (+35) **(I)** or heterozygous PE *Six3* (-133) (**G)** and PE *Wnt8b* (+21) **(J)** PE deletions. The values of expression were normalized to two housekeeping genes (*Eef1a1* and *Tbp*), and standard deviations (SDs) from technical triplicates are represented as error bars.

The previously presented 4C-seq results in AntNPC and mESC (Figure 4.10) showed that PEs preferentially contacted nearby major anterior neural genes, suggesting that other genes were not likely to be directly controlled by these regulator elements. To demonstrate that PEs preferentially controlled the expression of the nearest anterior neural genes, the expression of three additional genes located around each PE was also evaluated by RT-qPCR. In general, the expression levels of these additional PE adjacent genes were not consistently and/or markedly affected by the PE deletion (Figure 4.14).

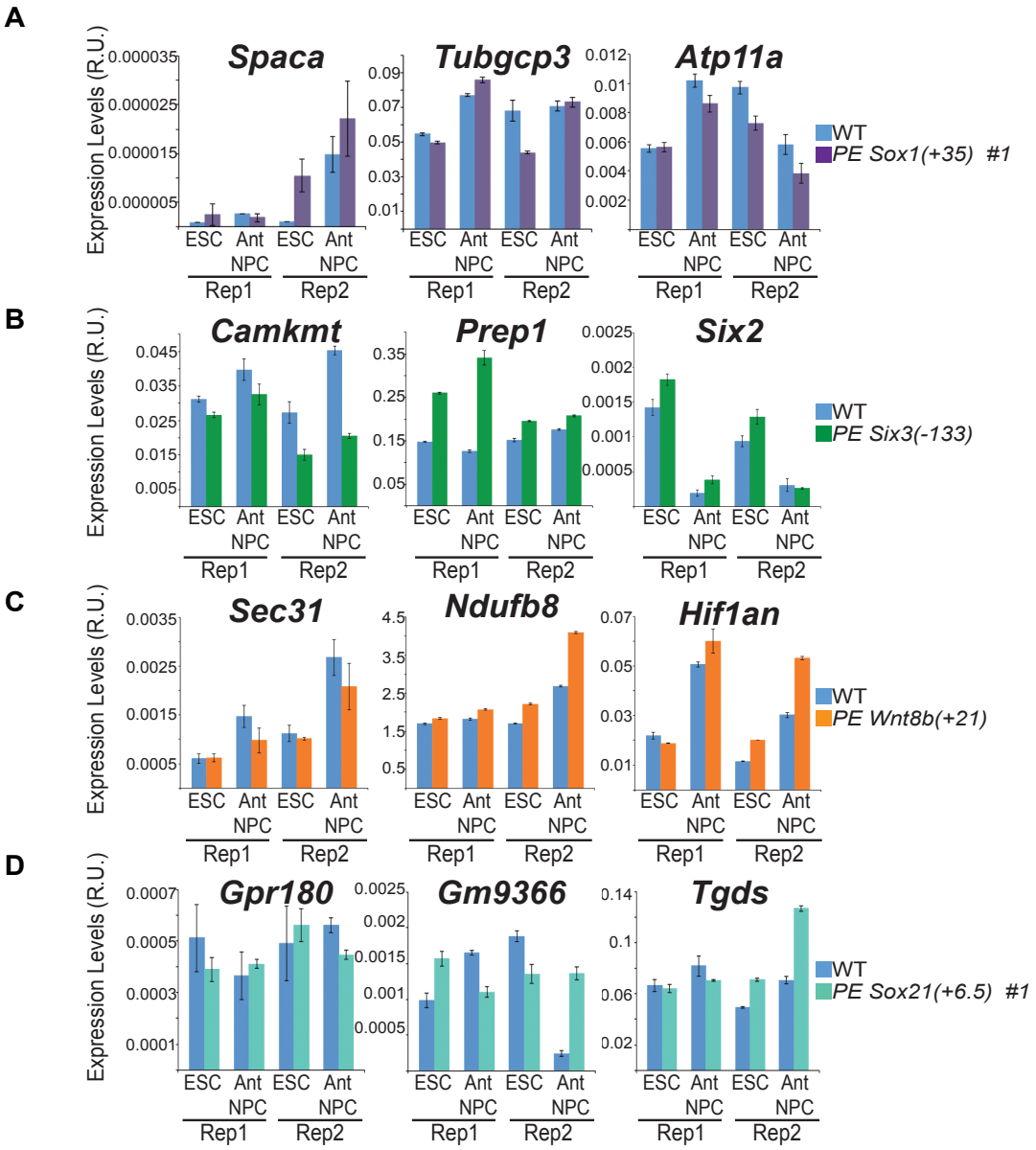

Figure 4.14: PEs neighboring genes are not affected by PEs deletions. The expression of three additional neighboring genes for PE *Sox1* (+35), **(A)** PE *Six3* (-133), **(B)** PE *Wnt8b* (+21) **(C)** and PE *Sox21* (+6.5) **(D)**, was measured by RT-qPCR. For each gene, the results obtained using WT mESC and mESC lines with PE homozygous deletions in two biologically independent differentiation experiments (Rep1 and Rep2) are shown. Expression values were normalized to two housekeeping genes (*Eef1a1* and *Tbp*) and the error bars represent standard deviations from technical triplicates.

Overall, the previous experiments demonstrated that PEs are critical regulatory elements necessary for the proper induction of genes with major roles during the establishment of anterior neural identity.

4.5.1- Additional functional characterization of PE *Lhx5* (-109)

PEs are required for the induction of major anterior neural gene regulators but still it is not clear how. In order to gain more insights into this question, one of the PE candidates, PE *Lhx5* (-109) was more extensively characterized. LHX5 is an important transcription factor during forebrain development. In mice, this protein is essential for the regulation of precursor cell proliferation and the control of neuronal differentiation and migration during hippocampal development (Andrews et al. 2003; Y. Zhao et al. 1999). This protein is also involved in learning and motor functions in adult mice.

As shown above (Figure 4.13F), the PE *Lhx5* (-109) deletion leads to a 96 % decrease in the expression of *Lhx5* in AntNPCs. To test whether a similar effect was observed at the protein level, immunofluorescence and western blot experiments were performed in mESC and AntNPCs that were either WT or homozygous for the PE deletions. Both experimental approaches confirmed that the PE *Lhx5* (-109) deletion caused a dramatic reduction of LHX5 protein levels in AntNPCs. LHX5 western blot band was almost absent in PE *Lhx5* (-109) $^{-/-}$ AntNPCs compared with WT AntNPCs (Figure 4.15A). Similarly, the LHX5 immunofluorescence staining levels for PE *Lhx5* (-109)$^{-/-}$ AntNPCs were very low, almost compared with the levels in mESCs where LHX5 is not expressed (Figure 4.15B).

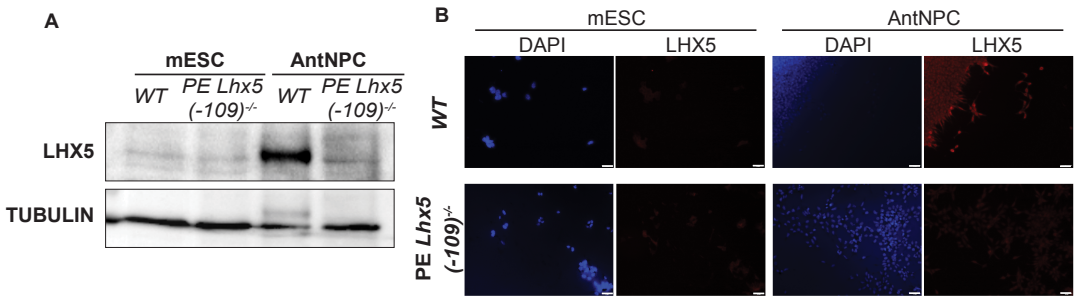

Figure 4.15: PEs Lhx5 (-109) deletion reduces LHX5 protein levels. (A) Western blot for LHX5 and TUBULIN as a loading control were performed in mESCs and AntNPCs for WT and PE Lhx5 (-109)$^{-/-}$ cells. The deletion induces a considerably lost in LHX5 levels in AntNPCs. **(B)** Immunofluorescence images in mESCs and AntNPCs where LHX5 protein levels were investigated in WT (top) or homozygous for a deletion of the PE *Lhx5* (-109) (bottom). White bar represents 20 µm scale bar.

Afterwards, RNA-seq experiments were performed in AntNPCs derived from either WT or PE *Lhx5* (-109)$^{-/-}$ mESCs to detect gene expression differences caused by the PE deletion. As expected, *Lhx5* was the most significantly down regulated gene in PE *Lhx5* (-109)$^{-/-}$ AntNPCs (Figure 4.16A). Interestingly, we observed that 376 additional genes were differentially expressed between WT and PE *Lhx5* (-109) $^{-/-}$ AntNPCs. Among them, 228 genes became up regulated and 148 genes were down regulated (Figure 4.16B). Transcriptions factors involved in the development of the brain were particulary abundant amongst the downregulated set of genes, (Figure 4.16B). We found *Fezf2*, which promotes neural differentiation during forebrain development (Zhang et al. 2014), *Pax6*, which regulates specification of the ventral neuron subtypes by establishing the correct progenitor domain (Takahashi and Osumi 2002) and *Dlx2*, which regulates differentiation of dopaminergic neurons of the ventral thalamus (Andrews et al. 2003) among the downregulated genes. On the other hand, the genes that resulted upregulated seemed to have a role in the neural repairing pathways. For instance, we found WNT signaling genes like *Wnt4* and *Wnt9* as well as genes involved in the regulation of calcium levels and synaptic plasticity. As examples, *Gper1* which induces protection against oxygen glucose deprivation in hippocampal neurons (T.-Z. Zhao et al. 2016) or *Grin2a*, which has important roles in synaptogenesis and synaptic plasticity (Marwick et al. 2015). A particularly interesting gene was *Nnat*, which was highly upregulated in PE *Lhx5* (-109)$^{-/-}$ AntNPCs. Neuronactin, the protein encoded by *Nnat*, participates in the maintenance of the hindbrain segment identity in during development, and it is involved in the maintenance of the overall structure of the nervous system. It has also been reported that high levels of Neuronactin in neurons can produce pathological consequences, as observed in Lafora disease (Sharma et al. 2013). It is worth mentioning, that the previous RNA-seq expression data revealed that the expression of the *Lhx5* neighboring genes was not affected by the PE *Lhx5* (-109) deletion (Figure 4.16C), which was in agreement with previous RT-qPCR experiments performed in ther other PE knockout lines (Figure 4.14). Finally a functional annotation of the genes significantly downregulated in PE *Lhx5* (-109)$^{-/-}$ mESCs was performed, giving as a result an overrepresentation of genes involved in neuronal

differentiation and forebrain development (Figure 4.16D). Taken together, this data further validates that PEs are necessary for the proper induction of their target genes. Since these genes include major anterior neural regulators, PEs can be considered as more generally required for the establishment of anterior neural identity.

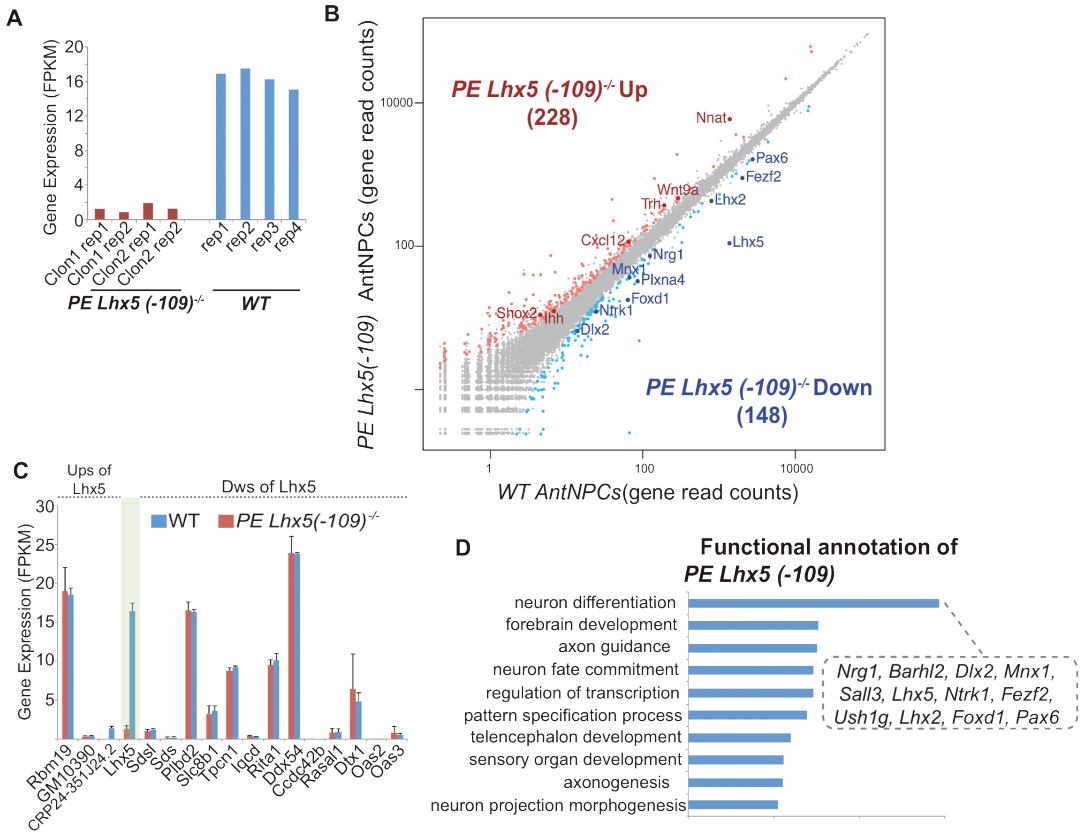

Figure 4.16: Functional characterization of PE *Lhx5* (-109) during AntNPC differentiation. (A) *Lhx5* expression levels (as FPKMs) were measured by RNA-seq in WT AntNPC and Ant NPCs derived from two different PE *Lhx5* (-109)$^{-/-}$ mESC clonal lines (Clon 1 and Clon2) **(B)** Plot representing RNA-seq data generated in WT AntNPCs or PE *Lhx5* (-109)$^{-/-}$ AntNPC. Mouse genes were plotted according to the average normalized RNA-seq read counts in WT AntNPCs and PE *Lhx5* (-109)$^{-/-}$ AntNPCs. Genes considered as significantly up-or down regulated in PE *Lhx5* (-109)$^{-/-}$ AntNPCs are shown in red and blue respectively. **(C)** Expression levels generated by RNA-seq as FPKMs in WT AntNPC and PE *Lhx5* (-109)$^{-/-}$ AntNPC are shown for genes located immediately downstream (Dws) or upstream (Ups) of *Lhx5,* which is shaded in green. Error bars represent standard deviations from four independent biological replicates. **(D)** Functional annotation according to Gene Ontology Biological Process terms was performed for genes significantly downregulated in PE *Lhx5* (-109)$^{-/-}$ AntNPCs. Ten out of the 20 most significantly overrepresented terms are shown.

4.5.2- PEs control the expression of lncRNAs divergently transcribed from the promoters of major anterior neural genes

During the evaluation of the previous RNA-seq data, it was observed that one of the most down-regulated genes in AntNPCs originated from PE *Lhx5* (-109) mESCs knock out lines was *RP24-351J24.2* (Figure 4.17A). This is a long non-coding RNA (lncRNA) divergently transcribed from the same promoter as *Lhx5.* Divergent lncRNAS were recently reported to facilitate the expression of essential developmental genes in cis due to the fact that they are in close proximity to them (S. Luo et al. 2016).

Another two PEs, PE *Six3* (-133) and PE *Sox21* (+6.5), also presented divergent lncRNAs at their target promoters. As in the case of *RP24-351J24.2*, the expression of these lncRNAs (i.e Six3os1 and AK 039417), measured by qPCR, was also reduced in the AnNPCs derived from mESC lines with deletions of the corresponding PEs (Figure 4.17 B and C). These results suggest that PE-dependent induction of major anterior neural genes might also entail divergent lncRNAs (S. Luo et al. 2016).

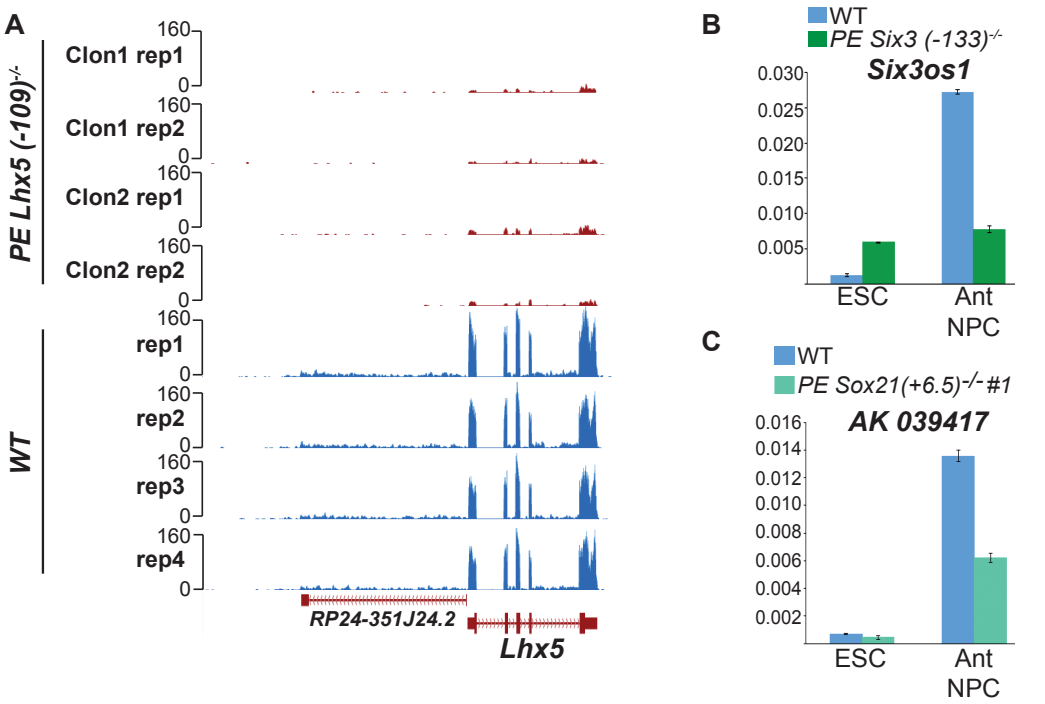

Figure 4.17: PEs also regulate the expression of divergent lncRNAS. (A) RNA-seq data obtained in AntNPC derived from WT mESC (four biological replicates) or from mESC with homozygous deletions of the PE *Lhx5* (-109) is shown for *Lhx5* and *RP24-351J24.*2. **(B and C)** Gene expression values were measured by RT-qPCR of *Six3 (Six3os1*) **(B)** and *Sox21 (AK 039417)* **(C)** lncRNAS in mESCs and AntNPCs that were either WT or homozygous for the indicated PE deletions.

4.6- PEs control the induction of anterior neural genes in a non-redundant manner involving the recruitment of RNA PolII to promoter regions

4.6.1- PEs are non redundant regulatory elements

Results presented in this study demonstrated that PEs regulate the expression of their target genes upon anterior neural differentiation. We have also uncovered that the target genes of PEs are usually anterior neural master regulators. Similarly to other major cell identity genes, PE target genes frequently displayed complex regulatory landscapes in AntNPCs, as illustrated by the presence of multiple H3K27ac peaks representing putative active enhancers in these cells and that can potentially contribute to the expression of these genes. To further illustrate this regulatory complexity, several PE candidates were part of broader super-enhancer (SE) identified in AntNPCs using H3K27ac ChIP-seq data (Cruz-Molina et al. 2017; Heinz et al. 2015) (e.g PE *Six 3*(-133), PE *Lhx5* (-109) and PE *Sox1* (+35)) (Figure 4.18).

Super-enhancers (SEs) differ from typical enhancers in content, TFs density, ability to activate transcription, topology and especially in size. SEs are clusters of enhancers densely occupied by major cell identity genes and mediator proteins. In principle, this regulatory complexity could confer gene expression robustness and the multiple enhancers controlling the same target genes could be redundant (Whyte et al. 2013; Shin et al. 2016). However, our results showed that, despite this complexity the deletion of a single PE has dramatic consequences in the expression of its target gene. This indicates that an enhancer hierarchy might regulate the induction of major anterior

neural genes with PEs being the most important and non-redundant elements of this system.

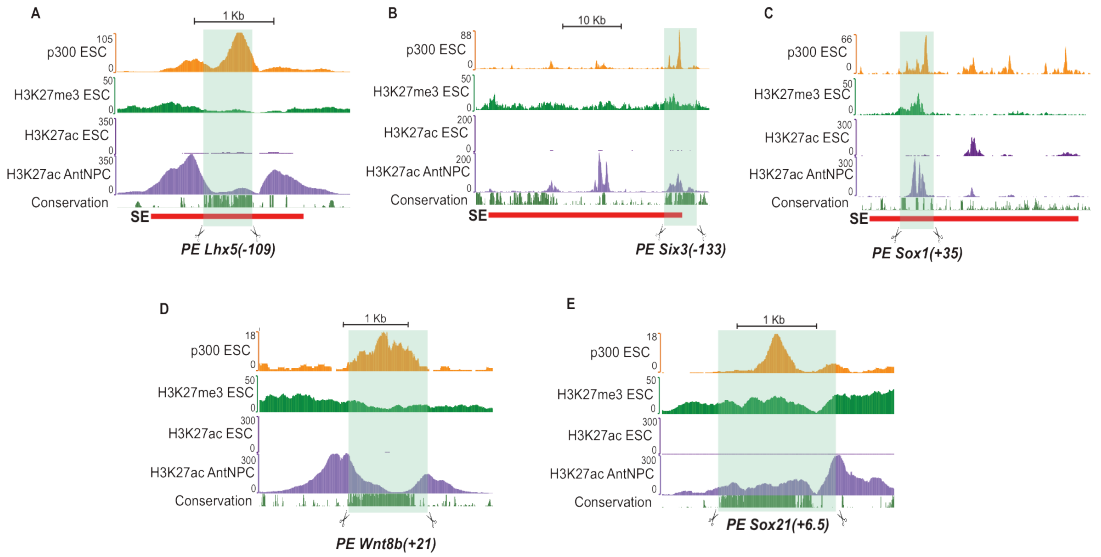

Figure 4.18: PEs can be found within larger AntNPCs super-enhancers. (A-E) Vertebrate conservation (as phastcons) and ChIP-seq profiles for p300, H3K27me3 and H3K27ac in mESCs as well as for H3K27ac in AntNPC are represented as genome browser pictures around five PE candidate regions. The engineered PE deletions are disguised in green. For three PE candidates, PE *Lhx5* (-109) **(A)**, PE *Six3* (-133) **(B)** and PE *Sox1* (+35) **(C)**, the PEs were located within larger SEs identified in AntNPC based on the H3K27ac ChIP-seq signal (red rectangle).

In order to investigate whether the induction of PE target genes is regulated in a hierarchical manner with PE playing an essential and non-redundant function, we examined how the deletion of individual PEs (PE *Lhx5* (-109), PE *Six3* (-133)) affected the activation, measured as gain of H3K27ac, of their target promoters as well as other enhancers located within the PE loci and presumably controlling the same genes in AntNPCs (Figure 4.19). ChIPs for H3K27ac were performed in mESCs and AntNPCs that were either WT or contained PE homozygous deletions (PE *Lhx5* $(-109)^{-/-}$, PE *Six3* $(-133)^{-/-}$) and H3K27ac levels were investigated by qPCR around the PE candidates, the corresponding PE target gene promoters and other putative enhancers located within the same regulatory domain. As expected, the deletions of the PEs lead to a loss in H3K27ac in AntNPCs around the deleted area and at the PE target gene promoters, which was in perfect agreement with the defective induction of PE target genes in AntNPCs. Most interestingly, PE deletions

also reduced H3K27ac levels in AntNPCs around the additional putative enhancers analyzed within each PE locus (Figure 4.19 B and C).

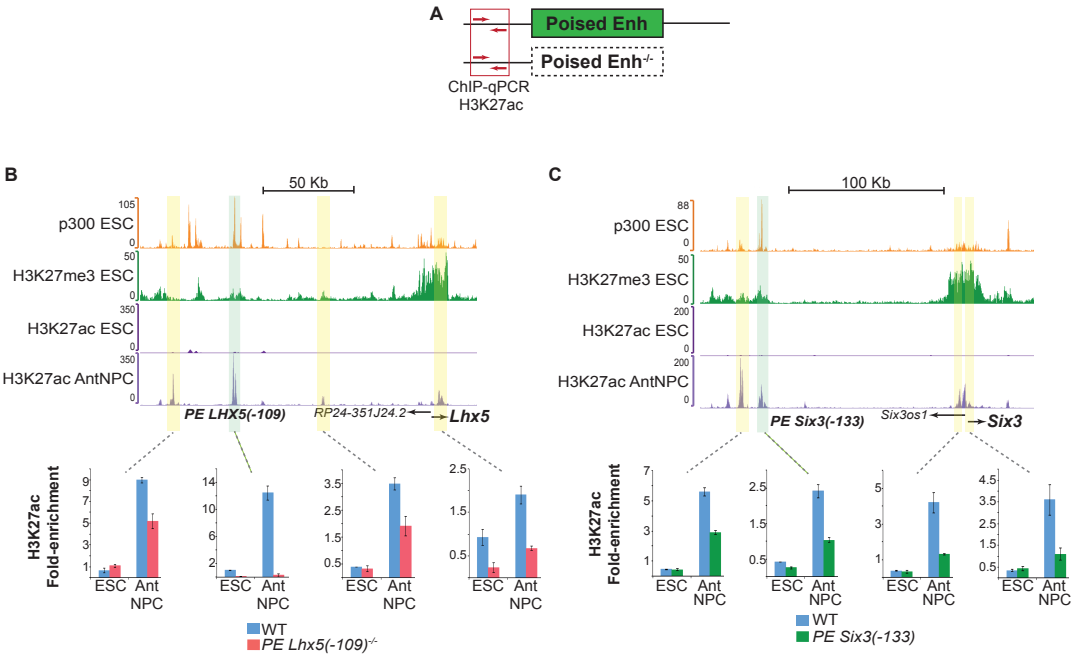

Figure 4.19: PEs are non redundant regulatory elements during the induction of major anterior neural genes. (A), Representation of the design of the ChIP-q-PCR primers used to detect H3K27ac levels around the deleted PEs. Genome browser representation of the *Lhx5* **(B)** and *Six3* **(C)** loci. PEs deletions are covered in green and additional putative enhancers as well as the PE target gene promoters are shaded in yellow. On the bottom, ChIP-qPCR graphics show the levels of H3K27ac in the selected areas for mESCs and AnNPCs that were either WT (blue) or contained homozygous deletions of the corresponding PEs.

4.6.2- PEs facilitate the recruitment of the RNA pol II complex to the promoters of their target genes

It has been reported, that in order to control the expression of their target genes during differentiation, enhancers can act at different steps of the transcriptional cycle, including the recruitment of the PolII pre-initiation complex (PIC) to the promoter regions of their target genes or facilitating the transcriptional initiation or elongation (Vernimmen and Bickmore 2015). To investigate at which of these steps PEs might preferentially act, we used ChIP-qPCR to measure the levels of pre-initiating (unphosphorylated PolII, 8WG16) and initiating (Ser5 phosphorylated PolII) PolII at PEs and their target

gene promoters during AntNPC differentiation. As a control, we also analyzed the promoter region of a pluripotency gene, *Zfp42*, where the two PolII variants were expected to be bound in mESCs but not in AntNPCs (Figure 4.20).

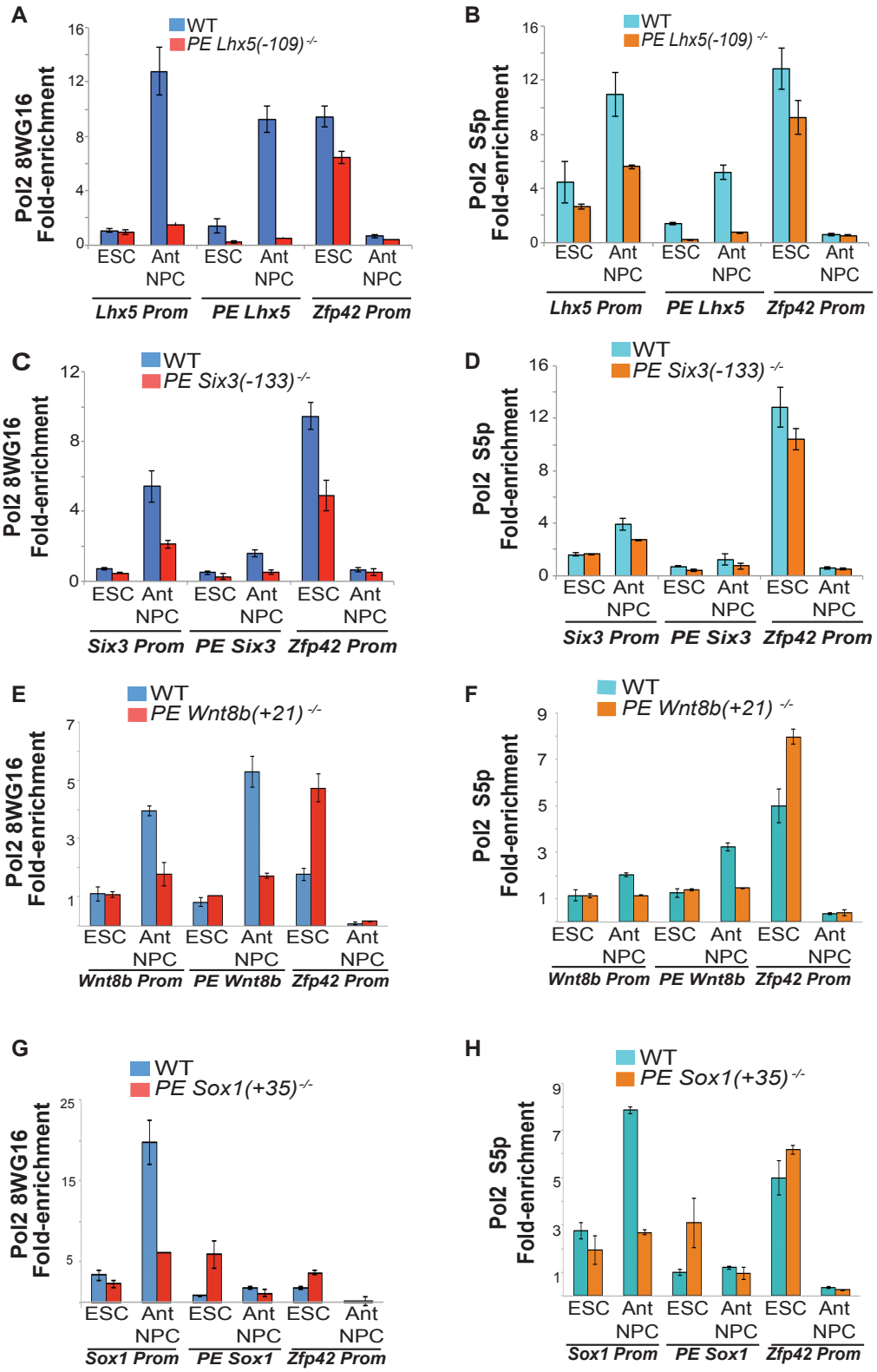

Figure 4.20: PEs mediate the recruitment of the PolII pre-initiation complex to the promoters of their target genes. (A), ChIP-qPCR was performed to investigate the levels of pre-initiating (PolII 8WG16) (**A, C, E** and **G)** and initiating (PolII S5p) (**B, D, F and H)** PolII in mESCs and AntNPCs that were either WT (blue) or homozygous for four different PE deletions. For each PE candidate, PE *Lxh5* (-109) **(A and B)**, PE *Six3* (-133) **(C and D)**, PE *Wnt8b* (+21) **(E and F)** and PE *Sox1* (+35) **(G and H)** PolII levels were measured around the deleted PE (Figure 4.20A), the corresponding PE target gene promoter and the *Zfp42* promoter that was used as a control and should be not significantly affected by the PE deletions. ChIP-qPCRs were performed as technical triplicates and standard deviations are represented as error bars.

The results of these experiments showed that PE deletions caused a strong reduction in the levels of pre-initiating PolII that were detected at the PEs and the PE target gene promoters upon AntNPC differentiation. We also observed similar reductions, although lower in magnitude, for the initiating PolII. In conclusion, PEs seem to preferentially control the induction of their target genes by mediating the recruitment of the pre-initiating PolII PIC to the promoter regions.

4.7- PRC2 as topological facilitator of PEs regulatory activity.

4.7.1- PEs-target gene contacts are PRC2 dependent

In our previous 4C-seq experiments we uncovered that PEs were already contacting their target promoters in mESCs. These pre-formed contacts might facilitate robust and timely gene induction during differentiation and thus support the previously proposed poised enhancer state (Rada-Iglesias et al. 2011) (Figure 4.10). It is worth mentioning that when 4C-seq profiles were evaluated in mESCs, the interaction signals of the PEs with their target genes mirrored the distribution of H3K27me3 around those genes. Since the Polycomb complex, and particularly its PRC2 subunit, mediates the deposition of H3K27me3 (Di Croce and Helin 2013), we proposed that such PE-target gene interaction could be PRC2 dependent. In order to corroborate this idea, we used mESC lines null for EED, a major PCR2 subunit, and thus defective for PRC2 activity. Then we performed 4C-seq experiments in $EED^{-/-}$ mESCs with the same PEs viewpoints that we used in WT mESCs. We found that in $EED^{-/-}$ lines the interactions between the PEs and their target genes were

dramatically reduced compared to WT mESCs at all four investigated loci (Figure 4.21A-D). In order to quantify the 4C signals, two intervals were created (R1 and R2) and contacts were analyzed (Cruz-Molina et al. 2017) (Figure 4.21E). R1 and R2 are 20Kb windows. R1 is centered on the transcription start site (TSS) of each PE target gene and R2 is located in an intermediary position between PE and target gene. The quantification gave as a result a significative difference in the 4C-signal in the target gene area (R1) while this tendency was not observed for the area not occupied by PRC2 (R2). Therefore, we concluded that PRC2 is necessary for the formation of the PEs – target gene contacts in mESCs.

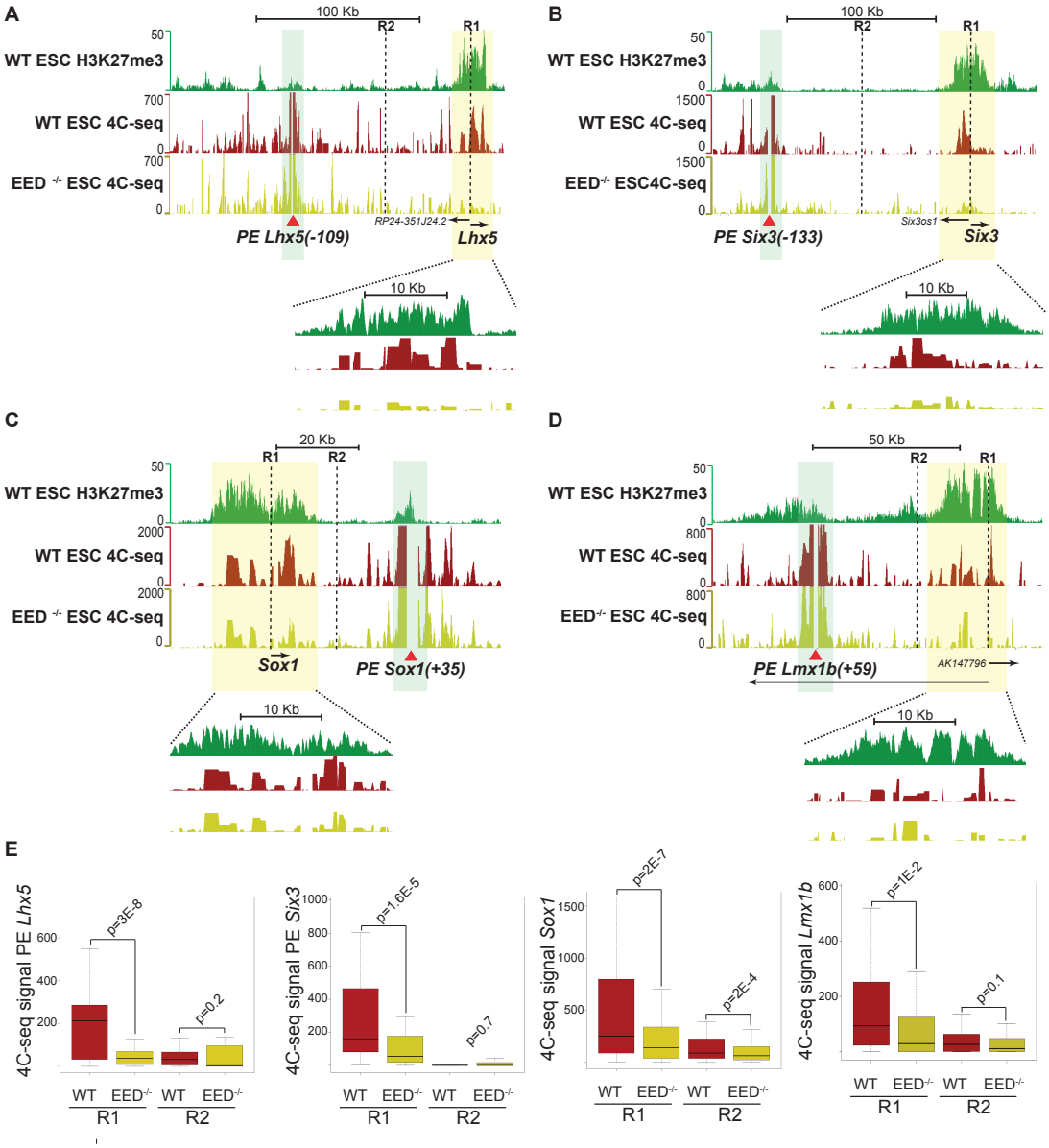

Figure 4.21: Contacts between PEs and their target genes in mESCs are PRC2 dependent. 4C-seq profiles generated in WT and EED$^{-/-}$ mESCs as well as

H3K27me3 ChIP-seq profiles generated in WT mESCs are shown around PE *Lhx5* (-109) **(A),** PE *Six3* (-133) **(B)**, PE *Sox1* (+35) **(C),** and PE *Lmx1b* (+59) **(D).** The H3K27me3-marked regions that cover the PEs target genes are shaded in yellow and the PEs in green. The red triangles represent the 4C-seq viewpoints. **(E)** For each locus showing in (A-D), 4C signals were quantified in two regions R1 and R2, where R1 is a 20 Kb window centered on the transcription start site (TSS) of each PE target gene and R2 is another 20 Kb window located in an intermediate position between each PE and its corresponding target gene. P-values were obtained by using Wilxon tests.

To validate the previous results, 3C-PCR experiments were performed using 4C and 3C libraries generated in WT and EED$^{-/-}$ mESCs as templates. In these experiments, specific primers are used to amplify chimeric DNA fragments representing PE-target gene interactions (see methods 4.7.3, 4.9.6). The outcome of these experiments confirmed our previous results, as PCR products were either absent or considerably weaker in EED$^{-/-}$ mESCs compared to WT mESCs, therefore PRC2 is needed to create this interaction (Figure 4.22).

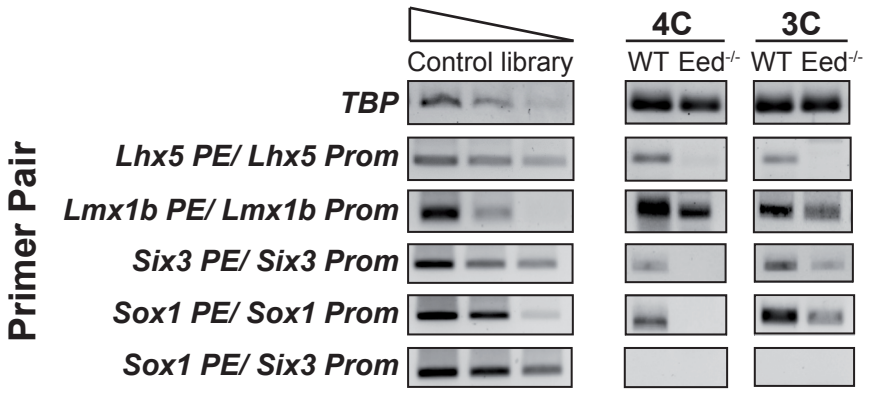

Figure 4.22: PCR validation of the PE-target gene contacts. 3C and 4C libraries generated for WT and EED$^{-/-}$ mESCs lines were analyzed by PCR using primers placed in the PE and the target promoter locus of each PE candidate studied. The *Tbp* amplicon was used as a positive control and a pair of primers located at *Sox1* PE and *Six3* promoter respectively, two regions that according to 4C-seq experiments were not predicted to physically interact, were combined and used as a negative control.

In the literature, it has been previously reported that enhancer-gene interactions are frequently mediated by architectural proteins, such as CTCF, Mediator or Cohesin (Kagey et al. 2010; Phillips-Cremins et al. 2013). To determine whether these architectural proteins also contribute to PE-target gene interactions in mESCs, we compare our mESC PEs with publically

available ChIP-seq data generated in mESCs for these proteins (Kagey et al. 2010; Cruz-Molina et al. 2017). These comparisons showed that only a small fraction of our PEs and none of the PEs investigated by 4C-seq were occupied by these architectural proteins (Figure 4.23A). On the other hand, it has been recently described that PcG bound genes can engage into long-range interactions (trans or inter-TAD interactions) in mESC and such interactions are mediated by PRC1 and PRC2. Moreover, these long-range interactions are particularly prominent at the four *Hox* genes clusters (Denholtz et al. 2013; Joshi et al. 2015; Schoenfelder et al. 2015). However, very limited long-range interactions in mESC were showed in our 4C-seq experiments, which were centered on PEs, our data showed that most interactions occurred in *cis* (intra-TAD) (Figure 4.23B).

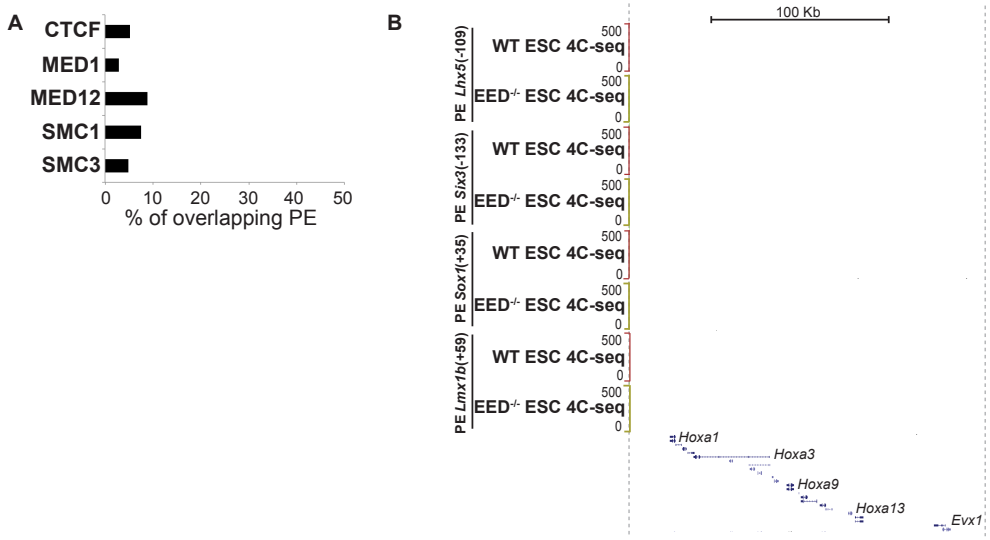

Figure 4.23: PRC2 is the main protein responsible of mediating contacts between PE and their target genes. (A). Percentage of PEs in mESCs overlapping CTCF, Mediator (MED1 and MED12) and Cohesin subunits (SMC1 and SMC3). **(B)** WT and EED$^{-/-}$ mESC 4C-seq profiles generated at the *HoxA* gene locus for four different PEs as viewpoints (from bottom to top: PE *Lmx1b* (+59), PE *Sox1* (+35), PE *Six3* (-133) and PE *Lhx5* (-109)).

4.7.2- Poised enhancer regulatory activity is facilitated by PRC2

The presence of H3K27me3 is one of the defining features of ESC PEs. This histone mark is deposited by PRC2 and it has been usually considered to be associated with transcriptional repression. Consequently, it was previously proposed that the presence of H3K27me3 and PRC2 at PEs could prevent

the spontaneous activation of these elements in absence of the relevant activating signals (Bonn et al. 2012; Spitz and Furlong 2012). However, our previous results have convincingly demonstrated that PRC2 enables the physical interaction between PEs an their target genes in mESCs, thus potentially creating a permissive regulatory topology that could facilitate the induction of the PE target genes during AntNPCs differentiation. In agreement with this hypothesis, recent studies reported that PRC2 is not required for transcriptional repression in mESCs (Riising et al. 2014) indicating that PRC2 is not necessary for maintaining PEs in an inactive state and also that the role of PRC2 in pluripotent cells might be different from what was previously anticipated.

To start evaluating whether PRC2 could facilitate PEs' regulatory activity, we performed H3K27me3 and H3K27ac ChIP-seq experiments in $EED^{-/-}$ mESCs as well as in AntNPCs generated from the $EED^{-/-}$ lines. As it was previously described (Leeb et al. 2010; Schoeftner et al. 2006), we observed that the lost of PRC2 lead to an almost total absence of H3K27me3 in $EED^{-/-}$ mESCs globally around PEs and around PoiAct enhancers (Figure 4.24A, B and D). However, no enrichment in H3K27ac around PEs was observed, especially compared with the H3K27ac levels at PoiAct regions in WT AntNPCS (Figure 4.24A, C and E). If PRC2 was indeed needed for maintaining PEs in an inactive state, we should have observed an increase in H3K27ac in the $EED^{-/-}$ lines. Since that was not the case, these results further suggest that PRC2 is not acting as a repressor in the context of PEs.

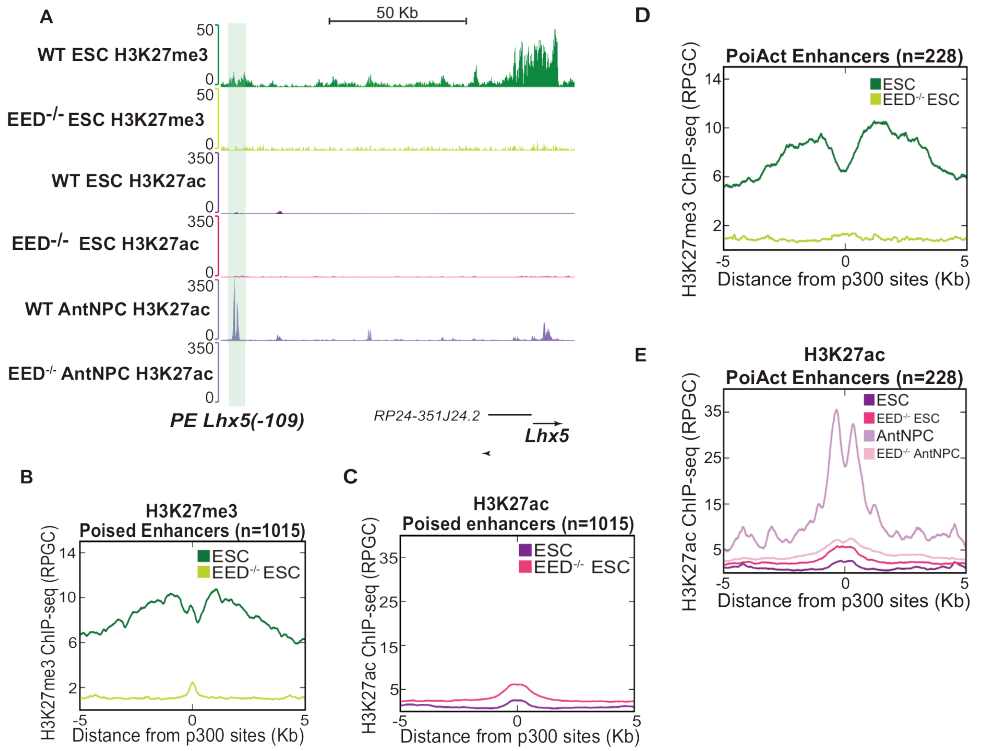

Figure 4.24: PRC2 is not necessary to maintain PEs in an inactive state. (A). Genome browser representation around PE *Lhx5* (-109) locus of H3K27me3 and H3K27ac ChIP-seq profiles generated in WT and EED$^{-/-}$ mESCs as well as H3K27ac ChIP-seq profiles in WT and EED$^{-/-}$ AntNPCs. **(B** and **C)** Average H3K27me3 **(B)** and H3K27ac **(C)** ChIP-seq signals in WT and EED$^{-/-}$ mESC around all mESCs PEs. **(D** and **E)** Average H3K27me3 **(D)** and H3K27ac **(E)** ChIP-seq profiles around PoiAct enhancers are shown for WT and EED$^{-/-}$ mESCs as well as WT and EED$^{-/-}$ AntNPCs.

We then asked how the loss of PRC2 activity could affect the expression of the PE target genes during AntNPC differentiation. Therefore, we generated RNA-seq profiles for EED$^{-/-}$ mESCs and AntNPCs and compared them with similar expression profiles previously generated in WT cells. In agreement with the minimal activation of PEs in EED$^{-/-}$ mESCs, the expression of the genes linked to PEs was barely increased in EED$^{-/-}$ mESCs compared with WT mESCs. Moreover and in concordance with previous reports, pluripotency genes remained highly expressed in EED$^{-/-}$ mESCs, while the expression of other non neural somatic markers like *Gata4* and *Sox17* showed a more evident de-repression (Leeb et al. 2010). Most interestingly, upon differentiation into AntNPCs, major anterior neural genes linked to PEs (e.g. *Lhx5, Six3, Fezf2*) were induced at considerably lower levels in EED$^{-/-}$ AntNPCs compared with WT AntNPCs (Figure 4.25 A-B). In contrast, when genes induced during the differentiation of mESCs into AntNPCs but not

linked to PEs were considered (indAll), the expression differences between WT and EED$^{-/-}$ AntNPCs were minimal (Figure 4.25C-D). These results indicate that the defective induction of neural genes upon differentiation of EED$^{-/-}$ mESCs seems to be specific to anterior neural genes linked to PEs. Consequently, our results indicate that, in mESCs, PRC2 might facilitate the future activation of major anterior neural identity genes.

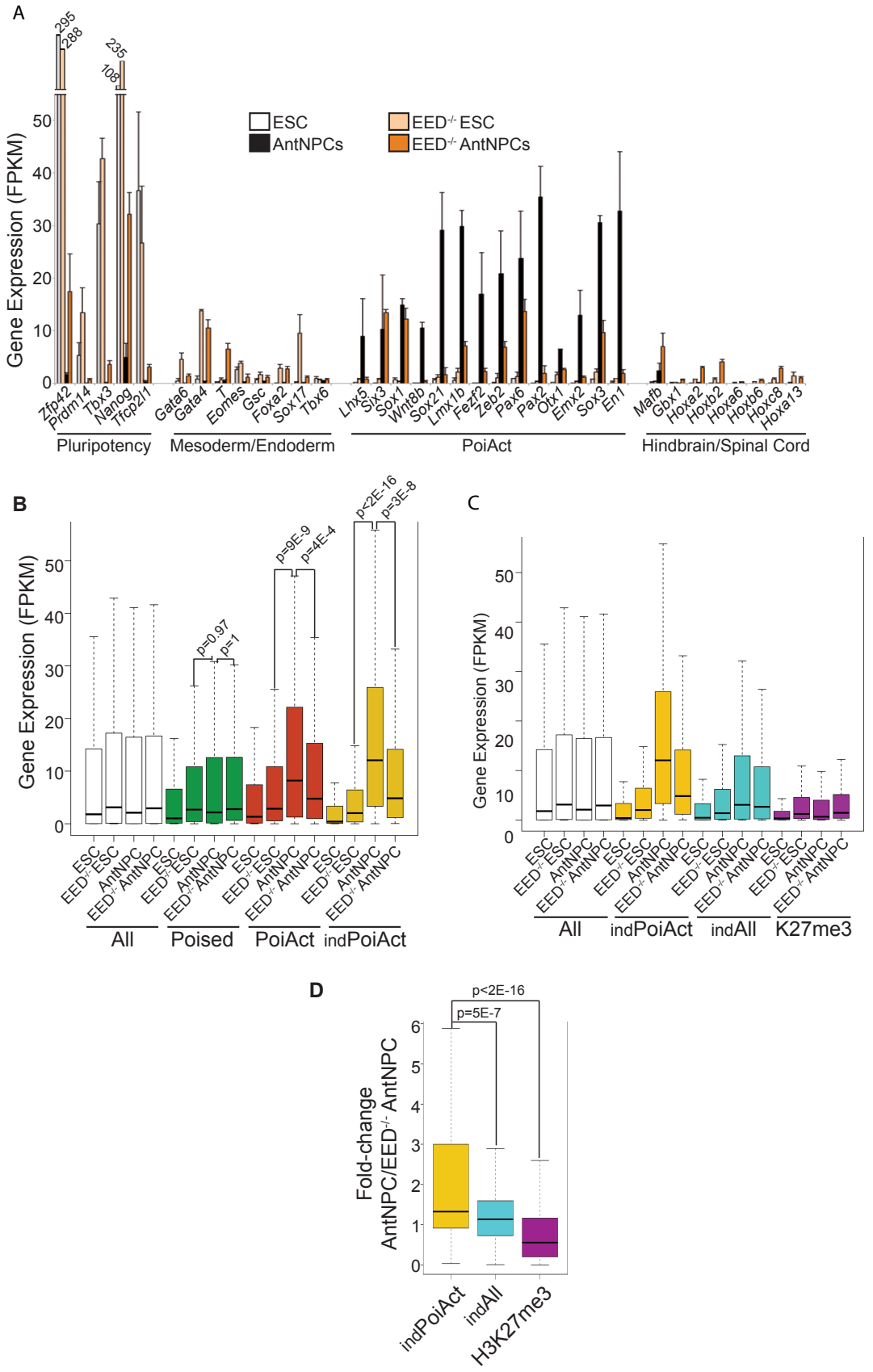

Figure 4.25: PRC2 as facilitator of the induction of anterior neural genes. (A) Gene expression was measured by RNA-seq in WT and EED$^{-/-}$ mESCs and AntNPCs. Expression levels (as FPKMs) are shown for pluripotency, mesodermal,

endodermal and posterior neural gene markers as well as for genes linked to PoiAct enhancers. **(B)** RNA-seq data from WT and EED$^{-/-}$ mESCs as well as their AntNPCs, were used to calculate the expression levels of all mouse genes (All), genes linked to PEs (Poised), genes linked to PoiAct enhancers (PoiAct), and genes linked to PoiAct enhancers that were induced ≥2-fold in WT AntNPCs compared to WT ESCs (indPoiAct). **(C)** RNA-seq data from WT mESC, WT AntNPC, EED $^{-/-}$ mESCs and EED $^{-/-}$ AntNPCs were used to calculate the expression levels of All genes, indPoiAct genes, all genes induced ≥2-fold in WT AntNPC compared to WT ESC after excluding indPoiAct genes (indAll) and genes whose promoters are marked by H3K27me3 in mESC according to (Bernstein et al. 2006) (H3K27me3). **(D)** Fold-change expression differences between WT AntNPCs and EED$^{-/-}$ AntNPCs are shown for indPoiAct genes, indAll genes, and H3K27me3 genes. P values for B, C and D were calculated using unpaired Wilcoxon tests.

In order to validate our previous results and confirm that PRC2 can facilitate PEs regulatory activity we used mESCs null for Suz12, another PRC2 subunit, which are also defective for PRC2 activity (Suz12 $^{-/-}$ mESCs). Suz12$^{-/-}$ and WT mESCs were differentiated into AntNPCs and the expression of some selected genes was measured by RT-qPCR. The results were very similar to the ones we obtained upon differentiation of EED$^{-/-}$ mESCs and major anterior neural genes linked to PEs failed to be induced in Suz12$^{-/-}$ mESCs (Figure 4.26).

Finally, to test whether PRC2 preferentially acts as a facilitator during the induction of genes involved in anterior neural development, we differentiated WT and EED$^{-/-}$ mESCs into mesodermal progenitors used a previously described protocol (see methods 3.3.5) as an example of a non-neural somatic lineage. Importantly, mesodermal markers were induced at normal or even higher levels in EED$^{-/-}$ mesodermal precursors than in their WT counterparts (Figure 4.27). Overall, together with recent observations in hESC (Collinson et al. 2016), our results suggest that, in agreement with the classical view of PcG complexes, PRC2 can act as a repressor of genes involved in mesodermal and endodermal development while playing a novel role as a facilitator during the activation of major anterior neural genes.

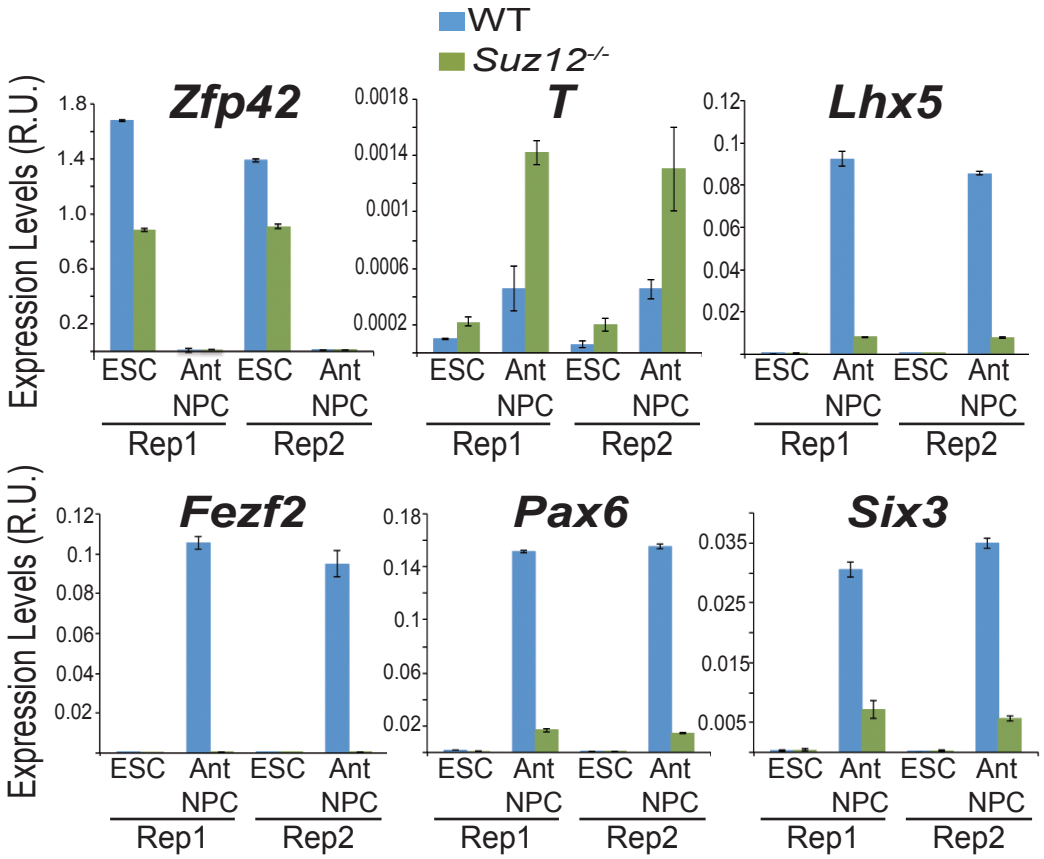

Figure 4.26: Suz12$^{-/-}$ mESCs show an impaired induction of major anterior neural genes during AntNPCs differentiation, which resembles EED$^{-/-}$ mESCs expression profile. RT-qPCRs were performed to measure gene expression in mESC and AntNPCs that were either WT (Blue) or Suz12$^{-/-}$ (green). Gene expression levels are shown for anterior neural genes linked to PEs (*Lhx5, Fezf2, Pax6 and Six3*), pluripotency markers (*Zfp42)* and mesodermal genes (*T*). Experiments were performed as two independent biological replicates (Rep1 and Rep2) and error bars represent the standard deviations of technical triplicates. *Tbp* and *Eef1a* were used as housekeeping genes.

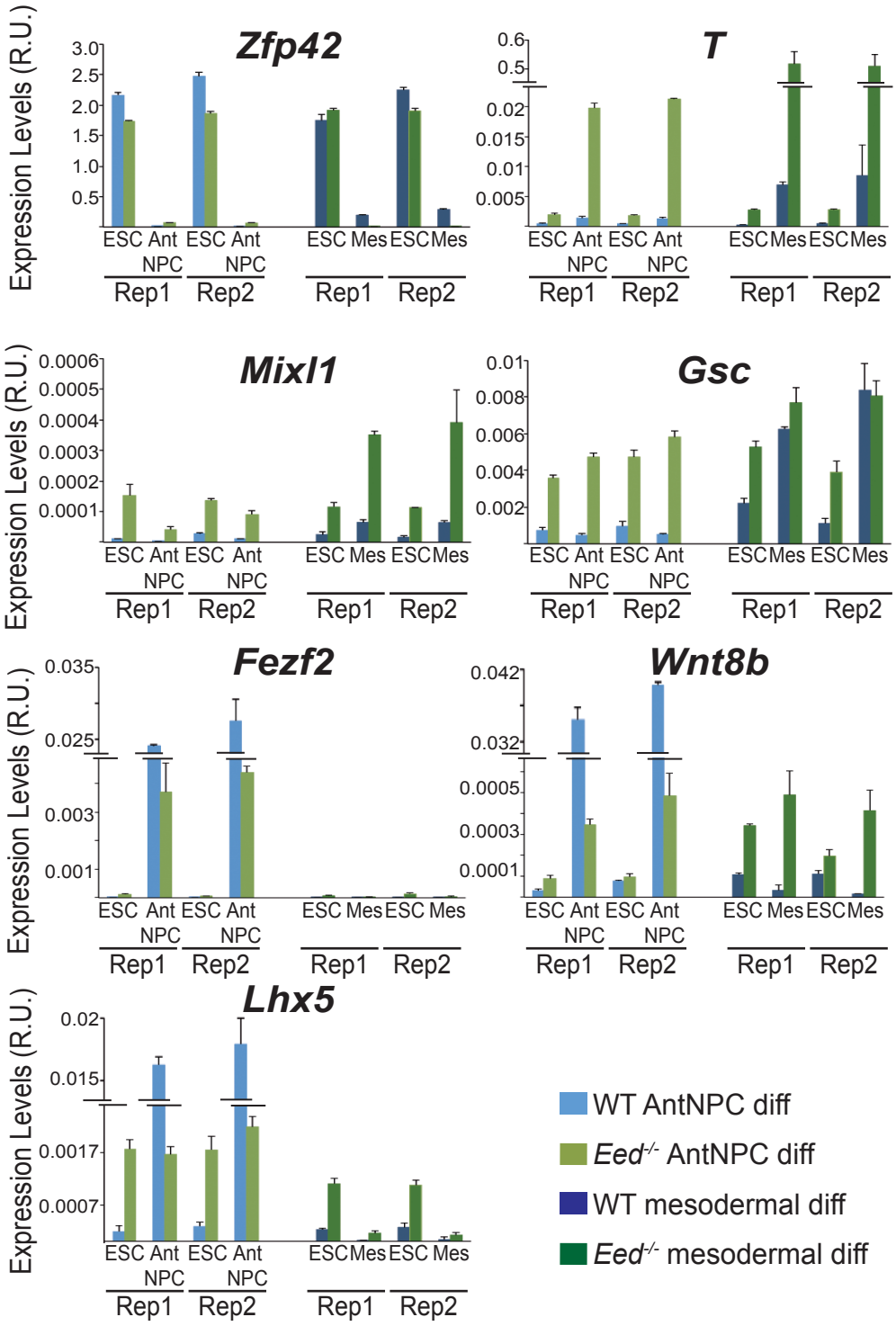

Figure 4.27: PRC2 acts as a facilitator for the induction of anterior neural genes and as a repressor for mesodermal genes. The expression of anterior neural (*Fezf2, Wnt8b, Lhx5*), pluripotency (*Zfp42*) and mesodermal (*T, Mixl1, Gsc*) genes was evaluated by RT-qPCR in WT and EED$^{-/-}$ mESC upon AntNPCs or mesodermal (Mes) differentiation. Experiments were performed as biological duplicates (Rep1 and Rep2). *Eef1a1* and *Tbp* were used as housekeeping genes to normalize the gene expression values and the standard deviations are represented as error bars from technical triplicates.

4.7.3- H3K27me3 deposition is enabled by the intrinsic genetic features of the poised enhancer regions.

In this work, it has been intensively demonstrated that the contacts between PEs and target genes in mESCs are PRC2 dependent and also that these contacts seem to be necessary for the proper induction of major anterior neural gene regulators. These novel and relevant findings made us wonder how PRC2 is initially recruited to PEs. We considered two main possibilities:

1- PEs recruit PRC2 in a direct manner and independently of their target gene promoters.
2- The presence of PRC2 at PEs is indirect and a consequence of the physical proximity to genes heavily bound by PRC2.

In order to gain some initial insights into this question, we first used mESCs public ChIP-seq data (Livyatan et al. 2015; Ku et al. 2008) to confirm that PRC1 and PRC2, and not only H3K27me3 were significantly enriched in the PEs regions. We evaluated the presence of RING1, a subunit of PRC1 as well as EZH2 and Suz12, core components of PRC2 in two PE loci (PE *Six3* (-133) and PE *Sox1* (+35)). As it was expected, PRC1 and PRC2 subunis were enriched in the PE loci as well as in their target genes (Figure 4.28A-B). In addition, proteins present in mESCs were ranking using the enrichment in the whole genome as a background showing a remarkably presence of the PcG subunits in mESCs (Figure 4.28C) (Chen et al. 2008).

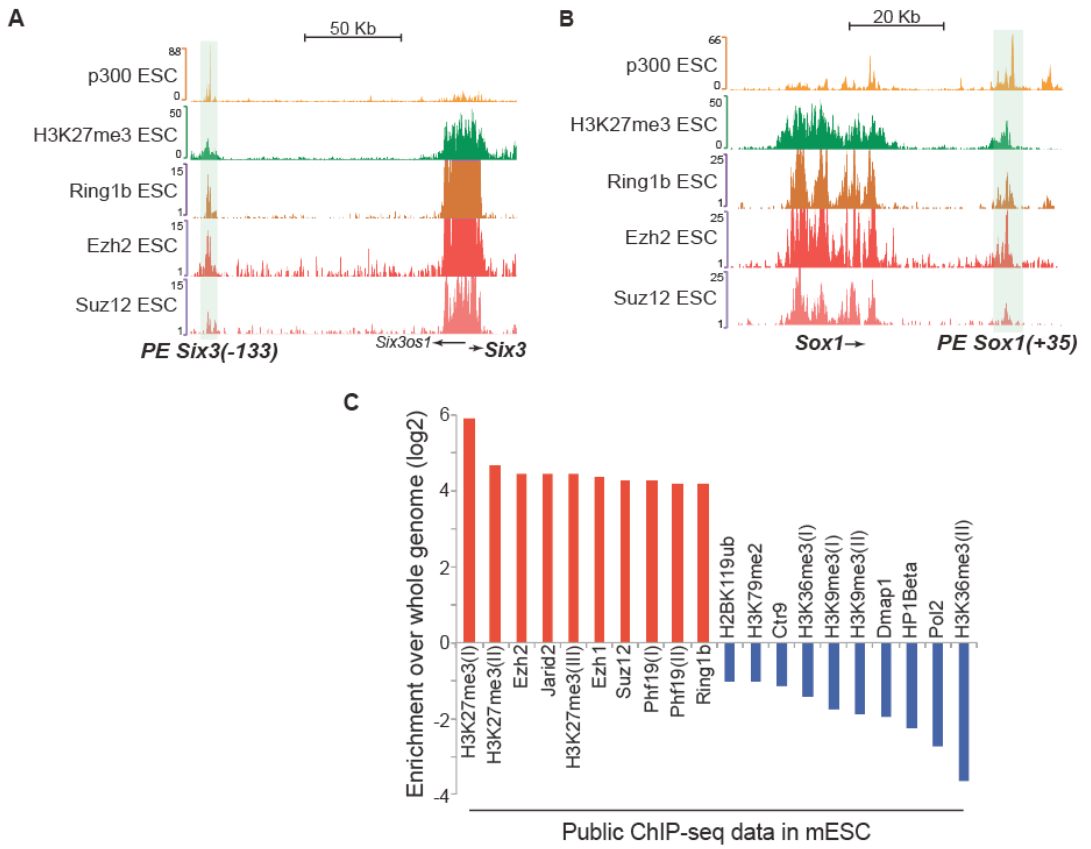

Figure 4.28: PEs are bound by PRC1 and PRC2 in mESCs. (A-B) P300, H3K27me3, Ring1b, Ezh2 and Suz12 ChIP-seq signal profiles in mESC are shown around **(A)** PE *Six3* (-133) and **(B)** PE *Sox1* (+35). **(C)** Proteins investigated by ChIP-seq in mESC were ranked according to their enrichment levels in mESCs with respect to the whole genome (used as background). The top 10 most and least enriched proteins are shown in red and blue, respectively. Enrichment analysis was performed using BindDB (Livyatan et al. 2015). PRC2 subunits: Ezh2, Jarid2, Ezh1, Suz12 and PHF19; PRC1 subunits: Ring1b.

Afterwards, a differential motif analysis was performed between PEs and active enhancers in order to identify DNA sequences that could be potentially involved in the recruitment of PRC2 to PEs. These analyses revealed that the most overrepresented motifs in all PEs presented a high CpG dinucleotide content (Figure 4.29A). Accordingly, PEs displayed much higher CpG dinucleotide frequency than active enhancers and frequently overlapped or were close to CpG islands (Figure 4.29 B). Further analysis divided the CpG regions between “weak” and “strong” depending on the CG content and determined that 74% of our PEs were within a weak CpG island (Cruz-Molina et al. 2017) (Figure 4.29C).

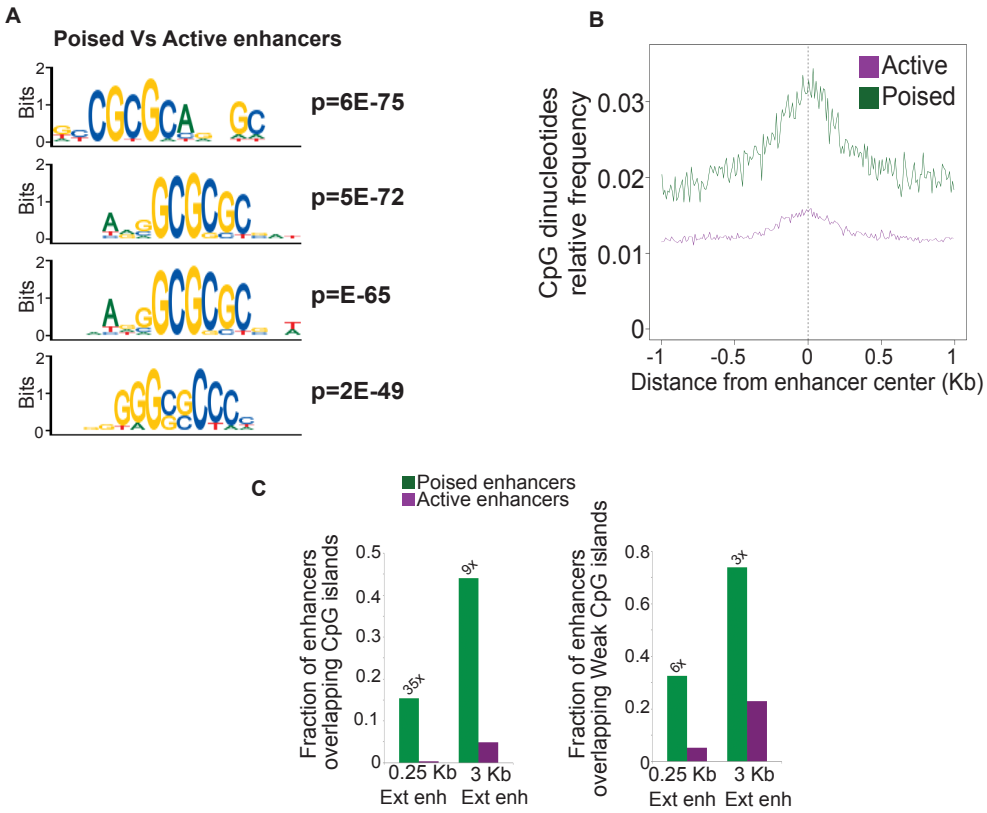

Figure 4.29: PEs display a high CpG dinucleotide content and frequently overlap with CpG islands. (A) Identification of motifs significantly enriched in PEs defined in mESC with respect to active enhancers (Cruz-Molina et al. 2017; McLean et al. 2010). **(B)** CpG dinucleotide frequencies for active (purple) and poised (green) enhancers are shown from 1 kb upstream (-1 kb) to 1 kb downstream (+1 kb) of the enhancers' mid position **(C)** Fraction of active (purple) and poised (green) enhancers overlapping classical (left) or weak (right) CpG islands. Statistical analysis (Cruz-Molina et al. 2017).

Several recent publications reported that CpG islands close to inactive gene promoters can act as a Polycomb response elements (PREs), that can directly recruit PcG complexes (Lynch et al. 2012; Mendenhall et al. 2010).Therefore, the majority of our PEs are within or close to a CpG island, we hypothesize that the PEs are found within a genomic context with a sequence composition which can directly mediate the recruitment of PRC2.

In order to evaluate this hypothesis, we used the mESC lines that we generated with homozygous deletions of different PE candidates. In two of those mESC lines, the PE deletions included a CpG island (PE *Six3* (-133) and PE *Sox1* (+35)), while for another two mESC lines they did not (PE *Wnt8b* (+21) and PE *Lhx5* (-109)). Then, in these mESC lines we measured

H3K27me3 by ChIP-qPCR in the regions immediately flanking the PE deletions as readout of PRC2 binding. Remarkably, H3K27me3 was significantly lost only in the mESC lines in which the PE deletions within a CpG island (Figure 4.30 B and D), while that was not the case for the PE deletions non-overlapping a CpG island (Figure 4.30 A and C). Moreover, ChIP-seq experiments showed that this lost of H3K27me3 was not global but rather restricted to the regions flanking the deleted PEs (e.g. PE *Sox1* (+35)). Therefore, PEs frequently reside in a genomic context with a high CpG dinucleotide content that can directly recruit PRC2 to these regulatory elements. We propose that the presence of PRC2 (and H3K27me3) at the PEs as well as at their target genes might lead to the physical contacts between both regions and therefore create a permissive regulatory topology, which facilitates the robust and timely induction of the anterior neural gene expression program during mESCs differentiation

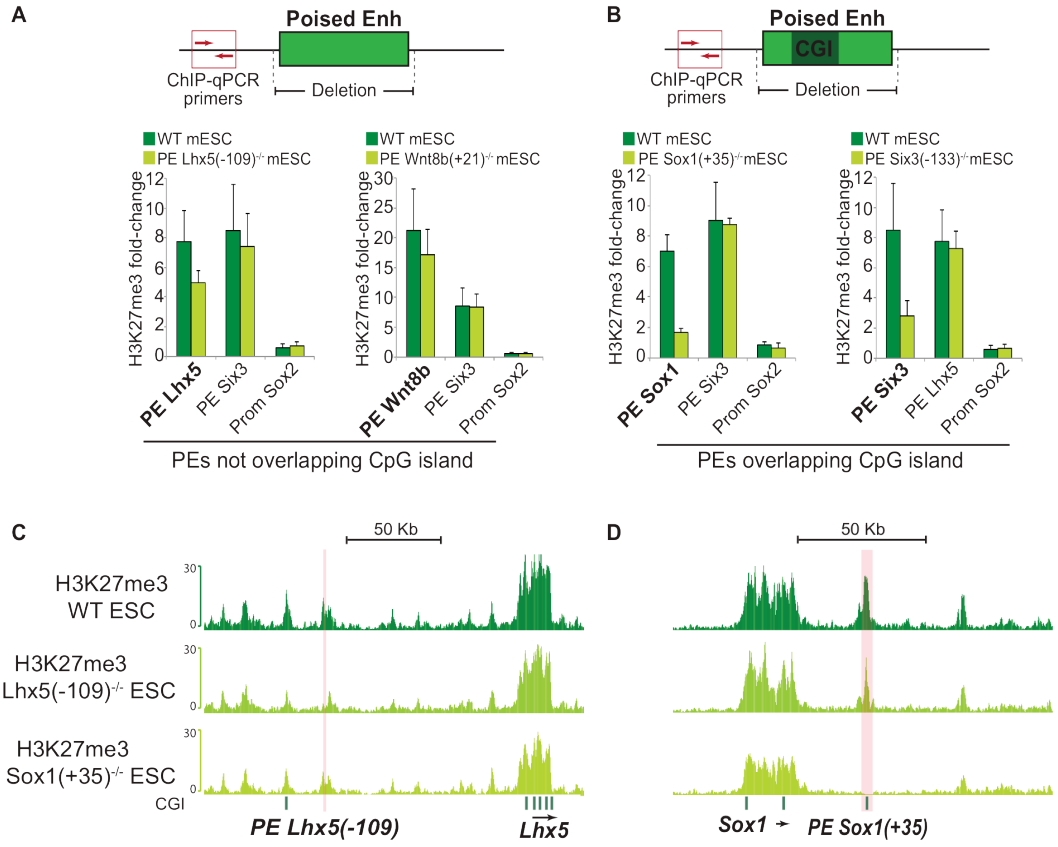

Figure 4.30: PEs overlapping a CpG island can act as PREs. (A-B) To investigate how the PE deletions affect H3K27me3 deposition; ChIP-qPCR primers were designed in close proximity to the deleted PE regions. H3K27me3 ChIPs were

performed in WT mESCs and mESCs with homozygous PE deletions in which the deletion did not include **(A)** or included **(B)** a CpG island. ChIP-qPCR signals were normalized using two different control regions (''Chr6_neg'' and ''Sox2_prom''). The error bars represent the standard deviations of six different ChIP-qPCR measurements (two ChIP biological replicate samples and three technical qPCR replicates for each sample). **(C** and **D)** H3K27me3 ChIP-seq profiles generated in WT, PE *Sox1* $(+35)^{-/-}$ and PE *Lhx5* $(-109)^{-/-,}$ mESCs are shown around PE *Lhx5* (-109) (not overlapping a CpG island) **(C)** and PE *Sox1* (+35) (overlapping a CpG island). **(D)**. Green rectangles represent CpG islands (CGI).

4.8-. Identification of *Zic2* as a potential activator of PEs during AntNPCs differentiation.

The differential motif analysis between poised and active enhancers described above revealed that, in addition to CpG-dinucleotide rich motifs, the binding sites of the Zic TF family was also highly enriched among PEs (Figure 4.31A). Zic TFs are evolutionary conserved regulators of neuroectodermal development, indeed they are involved in development of neural crest and another neural tissues, in somite development and in left-right axis patterning (Merzdorf 2007).

In order to determine if this family of TFs could be involved in the activation of PEs upon AntNPCs differentiation, we first evaluated by RT-qPCR the gene expression profile of the three main Zic family members (*Zic1, Zic2* and *Zic3*) during the differentiation of mESCs into AntNPCs. The expression pattern of *Zic2* was particularly interesting, as it was already expressed in mESCs, became upregulated by differentiation day three and remained highly expressed at days five and seven (Figure 4.31B). On the other hand, *Zic1* was not expressed in mESCs and it was specifically induced at late stages of differentiation (Figure 4.31B). In contrast, *Zic3* was highly expressed in mESCs and posteriorly down-regulated during AntNPCs differentiation, which is in agreement with *Zic3* being required for maintaining pluripotency in mESCs (Lim et al. 2007) (Figure 4.31B).

Afterwards an analysis of public ChIP-seq data for *Zic2* and *Zic3* in mESCs demonstrated that only *Zic2* is frequently bound to PEs in general and to PEs

that become active in AntNPCs differentiation (i.e PoiAct enhancers) (Ishiguro et al. 2007; Z. Luo et al. 2015) (Figure 4.31C and D).

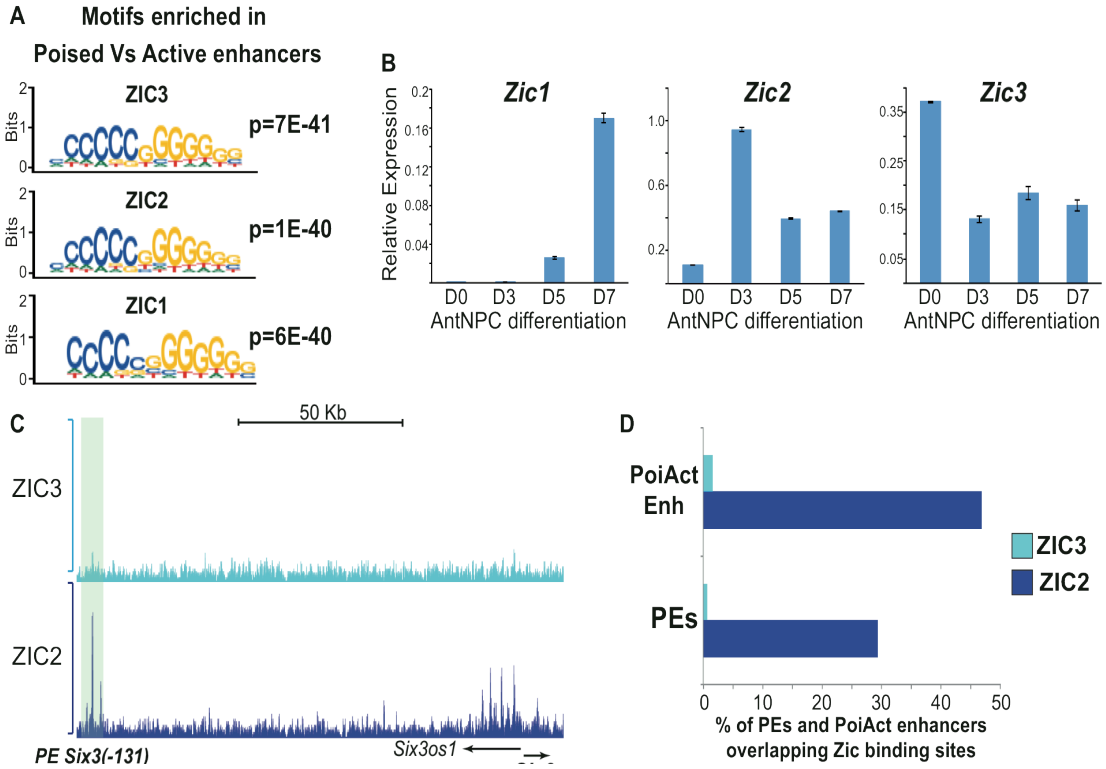

Figure 4.31: *Zic2* might be involved in the activation of PEs. (**A)** The analysis of *Motif Enrichment (AME)* software (McLeay and Bailey, 2010) was used to identify motifs significantly enriched in mESC PEs with respect to mESC active enhancers. The binding sites for Zic TFs were among the most overrepresented motifs in PEs. **(B)** RT-qPCR experiments showing the expression dynamics of *Zic1*, *Zic2* and *Zic3* during the differentiation of mESC (D0) into AntNPCs (Days 3-7, D3-D7). **(C)** Genome browser view of ZIC3 and ZIC2 ChIP-seq profiles in mESC around PE *Six3* (-133). **(D)** Percentage of PEs and PoiAct enhancers bound by ZIC2 (dark blue) and ZIC3 (light blue) in mESC.

Altogether, these preliminary observations suggest that ZIC2 might be involved in PEs activation and anterior neural induction. As more extensively described in the perspectives section, this possibility is now being experimentally tested.

5- DISCUSSION

5.1- Poised enhancers are regulatory sequences associated with the pluripotent cellular state

Poised enhancers were originally described in hESCs based on the simultaneous presence of activating (TFs, P300) and repressive (H3K27me3, PcG) features (M. P. Creyghton et al. 2010; Rada-Iglesias et al. 2011). As these enhancers were shown to become active upon somatic differentiation, they were proposed to reside in an inactive but poised state that could facilitate their timely and robust activation during ESCs differentiation (Rada-Iglesias et al. 2011) (Figures 4.1, 4.4 and 4.9). However, alternative models were subsequently proposed, according to which the presence of H3K27me3 at enhancer sequences indicated an inactive or repressive state that could negatively influence the expression of putative target genes (Bonn et al. 2012; Spitz and Furlong 2012). In these studies, the original definition of PEs was not followed and PEs were identified as genomic regions enriched in H3K4me1 and H3K27me3, without taking into account the presence of co-activators (e.g. p300, CBP) and/or TFs. Nonetheless, the data presented in this thesis indicates that the presence of these TFs and regulatory proteins is an essential and distinctive component of PEs in pluripotent cells. In contrast to H3K4me1 (Bonn et al. 2012; Cheng et al. 2014; Koenecke et al. 2016) these proteins can truly represent activating features that bookmark PEs for future activation. As a matter of fact, our previous work demonstrated that, if the original chromatin signature is used, very few PEs are identified outside the pluripotent state, mostly due to the absence of co-activators or TF binding within a H3K27me3 chromatin environment (Rada-Iglesias et al. 2012). Although the molecular reasons behind the scarcity of PEs outside pluripotency remains unknown, it is possible that chromatin enriched in PcG in ESC displays unique features that make it more accessible to TFs and other regulatory proteins (Bernstein et al. 2006; Menno P. Creyghton et al. 2008; Meshorer et al. 2006). Interestingly, it was recently reported that PEs might be also common in neural crest cells *in vivo* (Minoux et al. 2017). Since similarly

to ESCs, neural crest cells also have broad differentiation potential, it is tempting to speculate that PEs might be found within particularly plastic stem cell populations.

5.2- PEs are essential and non-redundant regulatory elements during the induction of major anterior genes

The functional relevance of PEs was clasically supported by correlative observations like those provided by reporter assays (Rada-Iglesias et al. 2011; Zentner et al. 2011). Using precise genetic deletions, our data shows that these *cis*-regulatory elements are necessary for the induction of their target genes upon mESCs differentiation. Moreover, PE deletions did not have as a consequence the de-repression of their target genes in mESCs, arguing against these sequences having silencer activity (Figure 4.13) (Entrevan et al. 2016; Kondo et al. 2014; Kondo et al. 2016).

The target genes of PEs typically reside within complex regulatory landscapes in which multiple and somehow redundant enhancers are believed to confer transcriptional robustness and fidelity (Frankel et al. 2010; Whyte et al. 2013). Nevertheless, our results demonstrated that the deletion of single PEs was sufficient to dramatically compromise the induction of their cognate genes. In the same way, it was also demonstrated, that individual PE deletions lead to a lost of H3K27ac at other enhancers presumably controlling the same target genes. This indicates that PE deletions can lead to a general decrease in enhancer activity within the regulatory domains of the PE target genes (Figure 4.19). Therefore, an enhancer hierarchy might exist during the induction of major anterior neural genes, with PEs playing a critical and non-redundant role. We hypothesize that other enhancers within these loci might be required for the maintenance and/or fine-tuning of gene expression patterns.

5.3- Poised enhancers as part of the "default" induction of anterior neural identity in ESC.

The first decisive step of neural development, the neural induction, is commonly thought to result from a 'default' embryonic pathway. Traditional developmental models suggest that epiblast cells *in vivo* and ESCs *in vitro* are predetermined by "default" (i.e. without extrinsic signals) towards the neural lineage (Gaspard and Vanderhaeghen 2010; Levine and Brivanlou 2007). Moreover, this "default" model also proposes that, to form non-neural tissues like mesoderm or endoderm as well as more posterior neural tissues (e.g. spinal cord), the epiblast cells *in vivo* and the ESCs *in vitro* need to be exposed to instructive signals (e.g. WNT, BMP, RA) (Levine and Brivanlou 2007; Stern 2005)

However, there is a main open question, in this "default" model of neural induction, why do pluripotent cells show this preference towards the anterior neural fate? Our data shows that PEs are preferentially associated with major anterior neural gene regulators (e.g. *Six3, Lhx5, Pax6, Sox1, Wnt8b, Fezf2, Pax2, Emx2*) (Figure 4.10) and also have an essential function during their induction. Taken together, these results indicate that PEs could represent a novel and important component of the "neural default" model. We propose a model where genetic (TF binding sites, CpG richness) and epigenetic (H3K27me3, p300) features are involved. We suggest that PEs might confer to anterior neural loci with a permissive regulatory topology that gives cells competency for neural induction once pluripotency signals (i.e. BMP, WNT) disappear and neural induction signals (i.e. FGF) arise (de Laat and Duboule 2013; Levine and Brivanlou 2007; Stern 2005). On the other hand, the master regulators of non-neural (i.e. mesoderm and endoderm) and posterior neural tissues, which are not as frequently linked with PEs, might be often found within another kind of regulatory domains in which enhancers and enhancer-promoter contacts are established *de novo* in response to extrinsic signals (de Laat and Duboule 2013).

5.4- PcG proteins as mediators of short and long-range regulatory interactions

4C-seq experiments revealed that in mESCs, PEs physically interact with their target genes in a PRC2 dependent manner. These findings are not in agreement with a recent study (Schoenfelder et al. 2015) where interactions were measured by using promoter capture Hi-C (CHi-C). In that study it was reported that long-range (inter-TAD) interactions between PcG bound promoters but not short-range (intra-TAD) interactions between PEs and their putative target genes were PRC1 dependent in mESCs. There could be several reasons for these discrepancies:

1- In Schoenfelder et al. PEs were defined only by the presence of H3K4me1 and H3K27me3, without requiring the binding TF and/or co-activators. Therefore, a broader set of genomic regions with potentially different regulatory and chromatin properties compared to our stringent set of PEs might have been considered.

2- In Schoenfelder et al. the PRC1 dependency of the PE-promoter contacts was analyzed using an “all-or-nothing” approach, considering whether a contact was either maintained or lost. Nevertheless, it is plausible that loss of PRC1 or PRC2 might lead to a partial, rather than total, loss of contacts between PEs and their target genes, as our 4C-seq data and the 3C and 4C validation experiments suggests (Figure 4.10, 4.21 and 4.22). Lastly, some, but not all, of the PE-target gene contacts identified by Schoenfelder *et al.* were maintained in PRC1 null mESCs.

3- Schoenfelder *et al.* used a 6-bp cutter enzyme in their CHi-C experiments. This kind of enzymes have less spatial resolution than the double 4bp-cutter approach used in our 4C-seq experiments and,

consequently, they are less suitable to investigate short-range interactions like the ones established between PEs and target genes.

It has been previously reported that in mESCs and *Drosphila*, PcG complexes may be involved in both short and long-range genomic interactions (Bantignies et al. 2011; Denholtz et al. 2013; Entrevan et al.2016; Joshi et al. 2015; Lanzuolo et al. 2007; Schoenfelder et al. 2015). Such interactions are probably dependent on the biochemical compatibility between different PRC1/2 subunits and their associated histone modifications (H3K27me3, H2AK119ub) (Hansen et al. 2008; Kyoichi Isono et al. 2013). Notably, our data indicates that poised enhancers mostly establish local interactions with target genes located in the same TAD in contrast with the long-range inter-TAD interactions between PcG bound genes that have been reported in several recent studies (Denholtz et al. 2013; Joshi et al. 2015; Splinter et al. 2012; Vernimmen et al. 2007). Major developmental genes usually display higher and broader PRC1/2 and H3K27me3 occupancy than PEs. We speculate that this fact might offer a biochemical explanation for the capacity of developmental genes to engage into long-range inter-TAD interactions (Hansen et al. 2008; Kyoichi Isono et al. 2013).

Overall, together with a number of recent reports, our data support a major role for PcG proteins in regulating genome architecture through a combination of long and short-range physical interactions (Entrevan et al. 2016).

5.5- PcG complexes as facilitators of gene induction during mESC differentiation

It has been previously shown that PRC2 null mESCs present differentiation defects due to insufficient silencing of pluripotency genes and unspecific de-repression of somatic regulators, which is in full agreement with the general view of PcG proteins as major epigenetic repressors, (Boyer et al. 2006; Obier et al. 2015; Pasini et al. 2007). Therefore, it was proposed that, in the context of PEs, PcG proteins could prevent these regulatory elements from premature or spurious activation or even confer them with silencer activity (Di Croce and

Helin 2013; Entrevan et al. 2016; Kondo et al. 2016). However the results here shown demosntrated that loss of PRC2 activity and H3K27me3 depletion did not lead to the activation of PEs, as these regulatory sequences only marginally gained H3K27ac in EED knockout mESCs (Figure 4.24). Moreover, the deletion of PEs did not result in a significant de-repression of their target genes in $EED^{-/-}$ mESCs, arguing against the previously proposed silencer activity of these regulatory elements (Figure 4.25). Most interestingly, the differentiation of $EED^{-/-}$ mESCs towards AntNPC demonstrated that the loss of PRC2 heavily compromised, rather than increased, the induction of major anterior neural genes linked to PEs. Together with the PRC2-dependent contacts that PEs and their target genes already establish in mESCs, these results made us propose a novel role for PcG-complexes in pluripotent cells, whereby PRC2 bestows major anterior neural loci with a permissive regulatory topology that makes them competent to become robustly activated once the appropriate differentiation cues become available (Figure 5.1).

This model is in full agreement with recent reports demonstrating that PRC2 is not required for gene silencing in mESCs as well as with classical *in vivo* and *in vitro* studies in which PRC2 was shown to be necessary for somatic differentiation but not for the maintenance of pluripotency (Chamberlain, Yee, and Magnuson 2008; Faust et al. 1998; Leeb et al. 2010; Pasini et al. 2007; Riising et al. 2014). We postulate that the differentiation defects observed in PRC2 knockout mESC might involve a dual role of PcG proteins as epigenetic repressors of mesodermal and endodermal genes and topological facilitators of anterior neural genes (Figure 4.27).

The role of PcG complexes as topological facilitators of gene induction has a single precedent in vertebrates (Kondo et al. 2014). According to this study, a PRC1-bound distal element located at the 3' end of *Meis2* acts as a silencer of this gene in most tissues during mouse embryogenesis. However, this same element facilitates the communication of *Meis2* promoter with a hindbrain-specific enhancer and thus the following activation of *Meis2* in the hindbrain. Interestingly, this PRC1-bound element is among the PEs we

identified in mESCs and, according to H3K27ac ChIP-seq data generated in mouse E11.5 forebrain (Nord et al. 2013), it represents an active enhancer in the forebrain, a tissue displaying high *Meis2* expression (Kondo et al. 2014). Thus, we hypothesize that the PRC1-bound element described in by Kondo et. al represents a PE in pluripotent cells that becomes active in the forebrain while acting as a structural scaffold in the hindbrain and ultimately facilitating *Meis2* induction in these two tissues.

Finally, we also discovered that a significant fraction of PEs reside in a genomic context with high CpG content, which endows these regulatory elements with the innate capacity to recruit PcG-complexes (Lynch et al. 2012; Mendenhall et al. 2010). These sequence features might allow PEs to remain marked by H3K27me3 and bound by PRC2 once pluripotent cells differentiate into non-anterior neural lineages. Since PRC2 repressive function is most apparent upon non neural differentiation (Riising et al. 2014), this might ensure that PEs do not become active in the wrong developmental and cellular contexts.

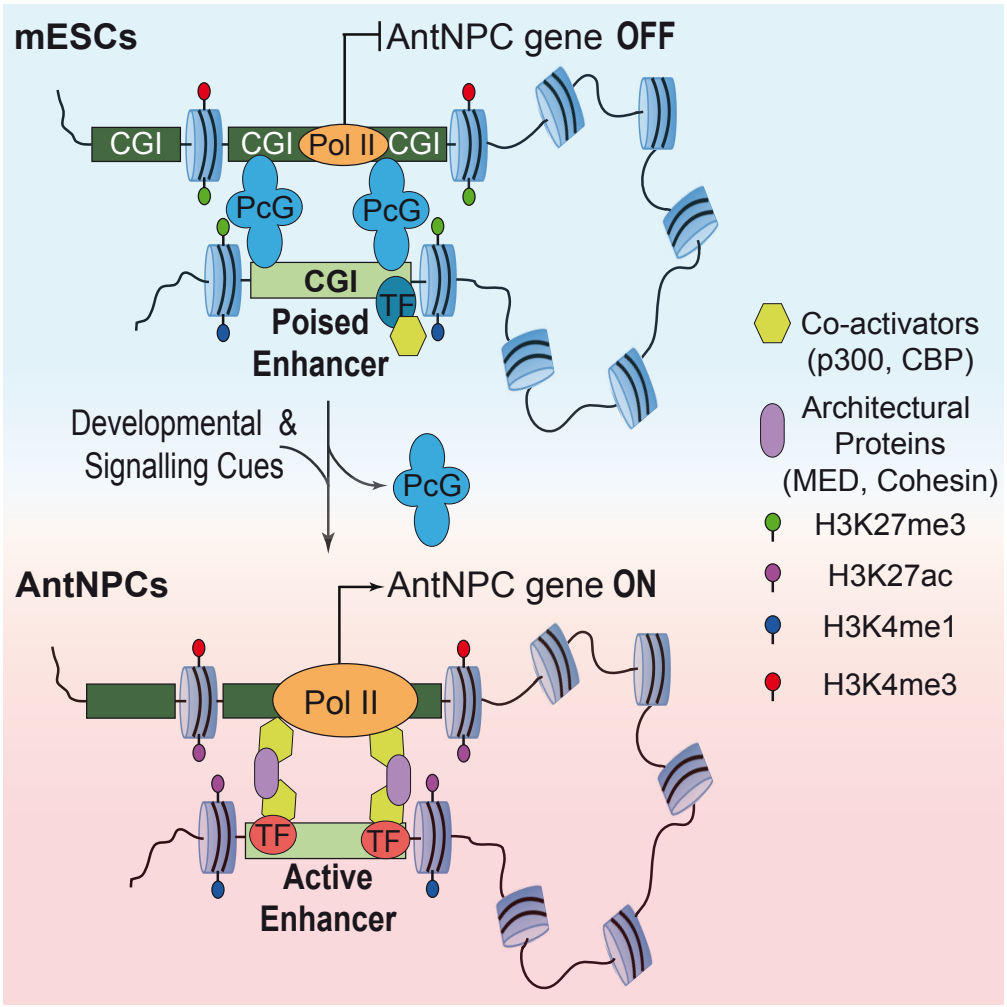

Figure 5.1: Proposed model for PEs function. In mESCs PEs are already contacting their target genes in a PRC2 dependent manner. We propose that these contacts create a permissive regulatory topology that upon appearance of the relevant differentiation cues and subsequent binding of TFs, co-activators, RNA PolII to the PEs, facilitates the induction of their congnate genes.

6- PERSPECTIVES

6.1- *In vivo* relevance of PEs.

In this work, we have demonstrated the functional relevance of pluripotency-associated PEs *in vitro*, however, their importance has not been evaluated *in vivo* yet, although preliminary observations in zebrafish embryos at least support their existence (Bogdanovic et al. 2012; Kaaij et al. 2016). Hence, a major goal in the future should be to determine if the PE chromatin signature is evolutionary conserved among pluripotent cells in vertebrates and whether this signature also poises the anterior neural fate *in vivo*.

mESCs grown under serum+LIF conditions, that we use as *in vitro* model resemble the epiblast cells of mouse E4-5-E6.5 embryos *in vivo*. Therefore, in order to determine whether PEs exist *in vivo* based on the presence of their unique chromatin signature, we propose to perform ChIP-seq for co-activators and histone marks in epiblast cells directly isolated from mouse embryos. However, for a TF or a co-activator like p300, ChIP-seq requires millions of cells, which would be very difficult to obtain for the mouse epiblast. In order to solve this problem, Assay for Transposase-Accessible Chromatin (ATAC-seq) (Buenrostro et al. 2015) could be performed instead. This technique allows to identify low nucleosome occupancy regions in the genome, thus reflecting binding of TFs and/or co-activators, but requires much less material ($<5x10^4$ cells). In addition to being bound by TFs and co-activators, in ESCs, PEs are defined by low H3K27ac and high H3K27me3 levels. ChIP-seq for histone modifications, as it happen with ATAC-seq, can be also performed with lower cell numbers. Importantly, we recently implemented a ChIP-seq protocol that enables the generation of histone modification profiles from low abundance embryonic tissues ($5x10^4$-$5x10^5$ cells) (Rehimi et al. 2016). Therefore, we will generate ATAC-seq, H3K27ac ChIP-seq and H3K27me3 ChIP-seq data in epiblast cells isolated from E6.5 mouse embryos to test whether PEs identified in mESCs display a similar chromatin signature in pluripotent cells *in vivo.*

If these experiments show that the chromatin signature of PEs is present *in vivo*, the next step would be to evaluate if the contacts between PEs and target genes are already established in the epiblast and then maintained in the developing forebrain in a similar way to what we observed during mESCs differentiation. To investigate the topological organization of PE loci *in vivo*, 4C-seq experiments in E6.5 epiblast and E11.5 forebrains will be performed using some of the previously investigated PEs as a viewpoints (PE *Lhx5* (-109), PE *Six3*(-133), PE *Sox1*(+35)).

In addition to all the foregoing, to evaluate the functional relevance of PEs *in vivo*, these sequences should be deleted in the mouse embryo. To perform these experiments, the Alt-R™ CRISPR-Cas9 System (IDT) seems the most appropriate, as in principle it should enable us to create PE deletions in mouse zygotes by injecting a pair of crRNA oligos flanking the selected PE, a tracrRNA oligo and recombinant Cas9 protein. Our goal is to generate a mouse line with heterozygous deletions ($PE^{+/-}$ mice) for a couple of PE candidates (PE *Lhx5* (-109) and PE *Wnt8b* (+21)). Then, male and female $PE^{+/-}$ mice will be crossed and the resulting embryos will be genotyped and analyzed at E10.5, a developmental stage at which the target genes of the selected PEs are strongly expressed in the forebrain. The expression levels of the PE target genes will be evaluated by RNA *in situ* hybridization as well as by RT-qPCR. If these experiments confirm that mouse embryos with $PE^{-/-}$ show reduced expression of the PEs target genes in comparison to WT embryos, then this fact would strongly support the *in vivo* relevance of PEs in the establishment of anterior neural identity during mouse embryogenesis.

6.2- Mechanism of activation of PEs

In this thesis it has been demonstrated that PEs are crucial regulatory elements for the proper induction of the anterior neural differentiation program. However, the mechanism by which PEs become active and perform their function is not clear yet. ChIP-seq data and motif analyses have shown that *Zic2* might have a role during PEs activation. Mechanistically, ZIC2 can

act as a transcriptional activator that, at least in some cellular contexts, binds to and is required for the activation of enhancers (Frank et al. 2015; Ishiguro et al. 2007). Therefore, we hypothesize that, based on its expression pattern during AntNPC differentiation, binding profile in mESCs and known *in vivo* functions, ZIC2 might be involved in PEs activation and anterior neural induction.

To test this hypothesis we will generate *Zic2* homozygous knockout mESCs lines and differentiate them into AntNPCs. RNA samples will be taken at day 0, day 3 and day 5 of differentiation and the expression of the PEs target genes will be evaluated and compared with WT cells. If our hypothesis is correct, and ZIC2 is necessary for the activation of PEs, then the induction of at least some of these genes should be compromised upon differentiation of *Zic2*$^{-/-}$ mESCs. RNA-seq experiments would also allow us to know if the induction of PoiAct enhancer target genes is globally affected in Zic2$^{-/-}$ cells and therefore more globally assess the importance of ZIC2 as a regulator of anterior neural induction.

If the previous set of experiments support the theory of ZIC2 as an important activator of PEs and major anterior neural genes, ZIC2, H3K27ac and H3K27me3 binding profiles in WT cells would be evaluated. Briefly, we would use ChIP-seq for the previous proteins in WT mESCs, day 3 WT AntNPCs and day 5 WT AntNPCs. These experiments would determine whether ZIC2 binding is observed and even increased at PEs that become specifically activated during AntNPCs differentiation. If so, *Zic2*$^{-/-}$ mESCs would be similarly used to generate H3K27ac and H3K27me3 ChIP-seq profiles on day 0, day 3 and day 5 of AntNPC differentiation. A comparison between both cell lines would reveal whether the loss of ZIC2 affects the chromatin features of PEs in mESCs and, most importantly, the subsequent activation of a subset of PEs during AntNPC differentiation.

In order to specifically demonstrate whether ZIC2 is necessary for the activation of PEs and their associated AntNPC genes, more experiments need to be done. One alternative would be to delete or disrupt the ZIC2

binding sites within selected PEs. Once mESC clones carrying these deletions are generated, they would be differentiated into AntNPCs together with WT mESCs. ChIP-qPCR and RT-qPCR experiments would be performed in order to show whether the deletion of ZIC2 binding sites leads to a loss of ZIC2 binding and reduce activation of the PE candidates as well as the induction of the associated target genes.

Finally, to gain mechanistic insights into how ZIC2 might activate PEs during AntNPC differentiation, ZIC2 immunoprecipitations (IPs) in mESC, day 3 AntNPC and day 5 AntNPC followed by Mass Spectrometry (MS) could be performed to identify the proteins interacting with ZIC2 in both mESC and AntNPC. Afterwards, and focusing on co-activators interacting specifically with ZIC2 in AntNPCs, ChIPs could be performed in WT mESC and WT AntNPC to determine whether such co-activators bind to PEs activated by ZIC2. Moreover, similar experiments could also be done in *Zic2*$^{-/-}$ mESCs and *Zic2*$^{-/-}$ AntNPCs, which would reveal whether the recruitment of the co-activators to PEs is ZIC2 dependent. Depending on the results obtained from the previous set of experiments, additional insights into the mechanisms leading to PEs activation might be gained and experimentally tested.

All in all, this project has conclusively demonstrated that PEs play a fundamental role in the induction of major anterior neural genes towards mESC differentiation. In addition, we have uncovered a new function for PRC2 as a facilitator of PEs regulatory activity. Lastly, we have proposed a series of experiments in order to evaluate if PEs are also important neural gene regulators *in vivo* as well as to deterimine their mechanism of action.

7. REFERENCES

Akasaka, T., M. van Lohuizen, N. van der Lugt, Y. Mizutani-Koseki, M. Kanno, M. Taniguchi, M. Vidal, M. Alkema, A. Berns, and H. Koseki. 2001. “Mice Doubly Deficient for the Polycomb Group Genes Mel18 and Bmi1 Reveal Synergy and Requirement for Maintenance but Not Initiation of Hox Gene Expression.” *Development (Cambridge, England)* 128 (9): 1587–97.

Andrews, Gracie L., Kyuson Yun, John L. R. Rubenstein, and Grant S. Mastick. 2003. “Dlx Transcription Factors Regulate Differentiation of Dopaminergic Neurons of the Ventral Thalamus.” *Molecular and Cellular Neurosciences* 23 (1): 107–20.

Aranda, S., G. Mas, and L. Di Croce. 2015. “Regulation of Gene Transcription by Polycomb Proteins.” *Science Advances* 1 (11): e1500737–e1500737. doi:10.1126/sciadv.1500737.

Attanasio, C., A. S. Nord, Y. Zhu, M. J. Blow, S. C. Biddie, E. M. Mendenhall, J. Dixon, et al. 2014. “Tissue-Specific SMARCA4 Binding at Active and Repressed Regulatory Elements during Embryogenesis.” *Genome Research* 24 (6): 920–29. doi:10.1101/gr.168930.113.

Banerji, J., S. Rusconi, and W. Schaffner. 1981. “Expression of a Beta-Globin Gene Is Enhanced by Remote SV40 DNA Sequences.” *Cell* 27 (2 Pt 1): 299–308.

Bantignies, Frédéric, and Giacomo Cavalli. 2011. “Polycomb Group Proteins: Repression in 3D.” *Trends in Genetics* 27 (11): 454–64. doi:10.1016/j.tig.2011.06.008.

Bantignies, Frédéric, Virginie Roure, Itys Comet, Benjamin Leblanc, Bernd Schuettengruber, Jérôme Bonnet, Vanessa Tixier, André Mas, and Giacomo Cavalli. 2011. “Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in Drosophila.” *Cell* 144 (2): 214–26. doi:10.1016/j.cell.2010.12.026.

Basu, Arindam, Frank H. Wilkinson, Kristen Colavita, Colin Fennelly, and Michael L. Atchison. 2014. “YY1 DNA Binding and Interaction with YAF2 Is Essential for Polycomb Recruitment.” *Nucleic Acids Research* 42 (4): 2208–23. doi:10.1093/nar/gkt1187.

Bernstein, Bradley E., Tarjei S. Mikkelsen, Xiaohui Xie, Michael Kamal, Dana J. Huebert, James Cuff, Ben Fry, et al. 2006. “A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells.” *Cell* 125 (2): 315–26. doi:10.1016/j.cell.2006.02.041.

Blackledge, Neil P., Nathan R. Rose, and Robert J. Klose. 2015. “Targeting Polycomb Systems to Regulate Gene Expression: Modifications to a Complex Story.” *Nature Reviews Molecular Cell Biology* 16 (11): 643–49. doi:10.1038/nrm4067.

Blackledge, Neil P., Anca M. Farcas, Takashi Kondo, Hamish W. King, Joanna F. McGouran, Lars L.P. Hanssen, Shinsuke Ito, et al. 2014. “Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation.” *Cell* 157 (6): 1445–59. doi:10.1016/j.cell.2014.05.004.

Boettiger, Alistair N., Bogdan Bintu, Jeffrey R. Moffitt, Siyuan Wang, Brian J. Beliveau, Geoffrey Fudenberg, Maxim Imakaev, Leonid A. Mirny, Chao-ting Wu, and Xiaowei Zhuang. 2016. “Super-Resolution Imaging Reveals Distinct Chromatin Folding for Different Epigenetic States.” *Nature* 529 (7586): 418–22. doi:10.1038/nature16496.

Bogdanovic, Ozren, Ana Fernandez-Miñán, Juan J. Tena, Elisa de la Calle-Mustienes, Carmen Hidalgo, Ila van Kruysbergen, Simon J. van Heeringen, Gert Jan C. Veenstra, and José Luis Gómez-Skarmeta. 2012. “Dynamics of Enhancer Chromatin Signatures Mark the Transition from Pluripotency to Cell Specification during Embryogenesis.” *Genome Research* 22 (10): 2043–53. doi:10.1101/gr.134833.111.

Bonn, Stefan, Robert P. Zinzen, Charles Girardot, E. Hilary Gustafson, Alexis Perez-Gonzalez, Nicolas Delhomme, Yad Ghavi-Helm, Bartek Wilczyński, Andrew Riddell, and Eileen E. M. Furlong. 2012. “Tissue-Specific Analysis of Chromatin State Identifies Temporal Signatures of Enhancer Activity during Embryonic Development.” *Nature Genetics* 44 (2): 148–56. doi:10.1038/ng.1064.

Boom, Vincent van den, Henny Maat, Marjan Geugien, Aida Rodríguez López, Ana M. Sotoca, Jennifer Jaques, Annet Z. Brouwers-Vos, et al. 2016. “Non-Canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis.” *Cell Reports* 14 (2): 332–46. doi:10.1016/j.celrep.2015.12.034.

Boroviak, Thorsten, Remco Loos, Patrick Lombard, Junko Okahara, Rüdiger Behr, Erika Sasaki, Jennifer Nichols, Austin Smith, and Paul Bertone. 2015. “Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis.” *Developmental Cell* 35 (3): 366–82. doi:10.1016/j.devcel.2015.10.011.

Boyer, Laurie A., Tong Ihn Lee, Megan F. Cole, Sarah E. Johnstone, Stuart S. Levine, Jacob P. Zucker, Matthew G. Guenther, et al. 2005. “Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells.” *Cell* 122 (6): 947–56. doi:10.1016/j.cell.2005.08.020.

Boyer, Laurie A., Kathrin Plath, Julia Zeitlinger, Tobias Brambrink, Lea A. Medeiros, Tong Ihn Lee, Stuart S. Levine, et al. 2006. “Polycomb Complexes Repress Developmental Regulators in Murine Embryonic Stem Cells.” *Nature* 441 (7091): 349–53. doi:10.1038/nature04733.

Buenrostro, Jason D., Beijing Wu, Howard Y. Chang, and William J. Greenleaf. 2015. “ATAC-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide: ATAC-Seq for Assaying Chromatin Accessibility.” In *Current Protocols in Molecular Biology*, edited by Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, and Kevin

Struhl, 21.29.1-21.29.9. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/0471142727.mb2129s109.

Bulger, Michael, and Mark Groudine. 2011. “Functional and Mechanistic Diversity of Distal Transcription Enhancers.” *Cell* 144 (3): 327–39. doi:10.1016/j.cell.2011.01.024.

Calo, Eliezer, and Joanna Wysocka. 2013. “Modification of Enhancer Chromatin: What, How, and Why?” *Molecular Cell* 49 (5): 825–37. doi:10.1016/j.molcel.2013.01.038.

Chamberlain, Stormy J., Della Yee, and Terry Magnuson. 2008. “Polycomb Repressive Complex 2 Is Dispensable for Maintenance of Embryonic Stem Cell Pluripotency.” *Stem Cells (Dayton, Ohio)* 26 (6): 1496–1505. doi:10.1634/stemcells.2008-0102.

Chen, Xi, Han Xu, Ping Yuan, Fang Fang, Mikael Huss, Vinsensius B. Vega, Eleanor Wong, et al. 2008. “Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells.” *Cell* 133 (6): 1106–17. doi:10.1016/j.cell.2008.04.043.

Cheng, Jemmie, Roy Blum, Christopher Bowman, Deqing Hu, Ali Shilatifard, Steven Shen, and Brian D. Dynlacht. 2014. “A Role for H3K4 Monomethylation in Gene Repression and Partitioning of Chromatin Readers.” *Molecular Cell* 53 (6): 979–92. doi:10.1016/j.molcel.2014.02.032.

Chittock, Emily C., Sebastian Latwiel, Thomas C.R. Miller, and Christoph W. Müller. 2017. “Molecular Architecture of Polycomb Repressive Complexes.” *Biochemical Society Transactions* 45 (1): 193–205. doi:10.1042/BST20160173.

Collinson, Adam, Amanda J. Collier, Natasha P. Morgan, Arnold R. Sienerth, Tamir Chandra, Simon Andrews, and Peter J. Rugg-Gunn. 2016. “Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells.” *Cell Reports* 17 (10): 2700–2714. doi:10.1016/j.celrep.2016.11.032.

Cong, Le, F Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D Hsu, et al. 2013. “Multiplex Genome Engineering Using CRISPR/Cas Systems.” *Science (New York, N.Y.)* 339 (6121): 819–23. doi:10.1126/science.1231143.

Creppe, Catherine, Anna Palau, Roberto Malinverni, Vanesa Valero, and Marcus Buschbeck. 2014. “A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation.” *PLoS Genetics* 10 (12): e1004851. doi:10.1371/journal.pgen.1004851.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, et al. 2010. “Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State.” *Proceedings of the National Academy of Sciences* 107 (50): 21931–36. doi:10.1073/pnas.1016071107.

Creyghton, Menno P., Styliani Markoulaki, Stuart S. Levine, Jacob Hanna, Michael A. Lodato, Ky Sha, Richard A. Young, Rudolf Jaenisch, and Laurie A. Boyer. 2008. “H2AZ Is Enriched at Polycomb Complex Target Genes in ES Cells and Is Necessary for Lineage Commitment.” *Cell* 135 (4): 649–61. doi:10.1016/j.cell.2008.09.056.

Cruz-Molina, Sara, Patricia Respuela, Christina Tebartz, Petros Kolovos, Milos Nikolic, Raquel Fueyo, Wilfred F.J. van Ijcken, et al. 2017. “PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation.” *Cell Stem Cell*, March. doi:10.1016/j.stem.2017.02.004.

Denholtz, Matthew, Giancarlo Bonora, Constantinos Chronis, Erik Splinter, Wouter de Laat, Jason Ernst, Matteo Pellegrini, and Kathrin Plath. 2013. “Long-Range Chromatin Contacts in Embryonic Stem Cells Reveal a Role for Pluripotency Factors and Polycomb Proteins in Genome Organization.” *Cell Stem Cell* 13 (5): 602–16. doi:10.1016/j.stem.2013.08.013.

Di Croce, Luciano, and Kristian Helin. 2013. “Transcriptional Regulation by Polycomb Group Proteins.” *Nature Structural & Molecular Biology* 20 (10): 1147–55. doi:10.1038/nsmb.2669.

Dietrich, Nikolaj, Mads Lerdrup, Eskild Landt, Shuchi Agrawal-Singh, Mads Bak, Niels Tommerup, Juri Rappsilber, Erik Södersten, and Klaus Hansen. 2012. “REST–Mediated Recruitment of Polycomb Repressor Complexes in Mammalian Cells.” Edited by Hiten D. Madhani. *PLoS Genetics* 8 (3): e1002494. doi:10.1371/journal.pgen.1002494.

Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. “Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions.” *Nature* 485 (7398): 376–80. doi:10.1038/nature11082.

Duan, Zhijun, Mirela Andronescu, Kevin Schutz, Choli Lee, Jay Shendure, Stanley Fields, William S. Noble, and C. Anthony Blau. 2012. “A Genome-Wide 3C-Method for Characterizing the Three-Dimensional Architectures of Genomes.” *Methods* 58 (3): 277–88. doi:10.1016/j.ymeth.2012.06.018.

Entrevan, Marianne, Bernd Schuettengruber, and Giacomo Cavalli. 2016. “Regulation of Genome Architecture and Function by Polycomb Proteins.” *Trends in Cell Biology* 26 (7): 511–25. doi:10.1016/j.tcb.2016.04.009.

Farcas, Anca M, Neil P Blackledge, Ian Sudbery, Hannah K Long, Joanna F McGouran, Nathan R Rose, Sheena Lee, et al. 2012. “KDM2B Links the Polycomb Repressive Complex 1 (PRC1) to Recognition of CpG Islands.” *ELife* 1 (December). doi:10.7554/eLife.00205.

Faust, C., K. A. Lawson, N. J. Schork, B. Thiel, and T. Magnuson. 1998. “The Polycomb-Group Gene Eed Is Required for Normal Morphogenetic Movements during Gastrulation in the Mouse Embryo.” *Development (Cambridge, England)* 125 (22): 4495–4506.

Francis, N. J. 2004. “Chromatin Compaction by a Polycomb Group Protein Complex.” *Science* 306 (5701): 1574–77. doi:10.1126/science.1100576.

Frank, Christopher L., Fang Liu, Ranjula Wijayatunge, Lingyun Song, Matthew T. Biegler, Marty G. Yang, Christopher M. Vockley, et al. 2015. “Regulation of Chromatin Accessibility and Zic Binding at Enhancers in the Developing Cerebellum.” *Nature Neuroscience* 18 (5): 647–56. doi:10.1038/nn.3995.

Frankel, Nicolás, Gregory K. Davis, Diego Vargas, Shu Wang, François Payre, and David L. Stern. 2010. “Phenotypic Robustness Conferred by Apparently Redundant Transcriptional Enhancers.” *Nature* 466 (7305): 490–93. doi:10.1038/nature09158.

Fu, Yanfang, Jeffry D Sander, Deepak Reyon, Vincent M Cascio, and J Keith Joung. 2014. “Improving CRISPR-Cas Nuclease Specificity Using Truncated Guide RNAs.” *Nature Biotechnology* 32 (3): 279–84. doi:10.1038/nbt.2808.

Galonska, Christina, Michael J. Ziller, Rahul Karnik, and Alexander Meissner. 2015. “Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming.” *Cell Stem Cell* 17 (4): 462–70. doi:10.1016/j.stem.2015.07.005.

Gaspard, Nicolas, Tristan Bouschet, Adèle Herpoel, Gilles Naeije, Jelle van den Ameele, and Pierre Vanderhaeghen. 2009. “Generation of Cortical Neurons from Mouse Embryonic Stem Cells.” *Nature Protocols* 4 (10): 1454–63. doi:10.1038/nprot.2009.157.

Gaspard, Nicolas, and Pierre Vanderhaeghen. 2010. “Mechanisms of Neural Specification from Embryonic Stem Cells.” *Current Opinion in Neurobiology* 20 (1): 37–43. doi:10.1016/j.conb.2009.12.001.

Ghavi-Helm, Yad, Felix A. Klein, Tibor Pakozdi, Lucia Ciglar, Daan Noordermeer, Wolfgang Huber, and Eileen E. M. Furlong. 2014. “Enhancer Loops Appear Stable during Development and Are Associated with Paused Polymerase.” *Nature* 512 (7512): 96–100. doi:10.1038/nature13417.

Gouti, Mina, Anestis Tsakiridis, Filip J. Wymeersch, Yali Huang, Jens Kleinjung, Valerie Wilson, and James Briscoe. 2014. “In Vitro Generation of Neuromesodermal Progenitors Reveals Distinct Roles for Wnt Signalling in

the Specification of Spinal Cord and Paraxial Mesoderm Identity." Edited by Roel Nusse. *PLoS Biology* 12 (8): e1001937. doi:10.1371/journal.pbio.1001937.

Guo, Ge, Ferdinand von Meyenn, Fatima Santos, Yaoyao Chen, Wolf Reik, Paul Bertone, Austin Smith, and Jennifer Nichols. 2016. "Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass." *Stem Cell Reports* 6 (4): 437–46. doi:10.1016/j.stemcr.2016.02.005.

Hackett, Jamie A., and M. Azim Surani. 2014. "Regulatory Principles of Pluripotency: From the Ground State Up." *Cell Stem Cell* 15 (4): 416–30. doi:10.1016/j.stem.2014.09.015.

Hansen, Klaus H., Adrian P. Bracken, Diego Pasini, Nikolaj Dietrich, Simmi S. Gehani, Astrid Monrad, Juri Rappsilber, Mads Lerdrup, and Kristian Helin. 2008. "A Model for Transmission of the H3K27me3 Epigenetic Mark." *Nature Cell Biology* 10 (11): 1291–1300. doi:10.1038/ncb1787.

Hayashi, Katsuhiko, Hiroshi Ohta, Kazuki Kurimoto, Shinya Aramaki, and Mitinori Saitou. 2011. "Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells." *Cell* 146 (4): 519–32. doi:10.1016/j.cell.2011.06.052.

Heintzman, Nathaniel D., Gary C. Hon, R. David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F. Harp, Zhen Ye, et al. 2009. "Histone Modifications at Human Enhancers Reflect Global Cell-Type-Specific Gene Expression." *Nature* 459 (7243): 108–12. doi:10.1038/nature07829.

Heinz, Sven, Casey E. Romanoski, Christopher Benner, and Christopher K. Glass. 2015. "The Selection and Function of Cell Type-Specific Enhancers." *Nature Reviews Molecular Cell Biology* 16 (3): 144–54. doi:10.1038/nrm3949.

Ishiguro, Akira, Maki Ideta, Katsuhiko Mikoshiba, David J. Chen, and Jun Aruga. 2007. "ZIC2-Dependent Transcriptional Regulation Is Mediated by DNA-Dependent Protein Kinase, Poly(ADP-Ribose) Polymerase, and RNA

Helicase A." *The Journal of Biological Chemistry* 282 (13): 9983–95. doi:10.1074/jbc.M610821200.

Isono, K. 2005. "Mammalian Polycomb-Mediated Repression of Hox Genes Requires the Essential Spliceosomal Protein Sf3b1." *Genes & Development* 19 (5): 536–41. doi:10.1101/gad.1284605.

Isono, K.-i., Y.-i. Fujimura, J. Shinga, M. Yamaki, J. O-Wang, Y. Takihara, Y. Murahashi, Y. Takada, Y. Mizutani-Koseki, and H. Koseki. 2005. "Mammalian Polyhomeotic Homologues Phc2 and Phc1 Act in Synergy To Mediate Polycomb Repression of Hox Genes." *Molecular and Cellular Biology* 25 (15): 6694–6706. doi:10.1128/MCB.25.15.6694-6706.2005.

Isono, Kyoichi, Takaho A. Endo, Manching Ku, Daisuke Yamada, Rie Suzuki, Jafar Sharif, Tomoyuki Ishikura, Tetsuro Toyoda, Bradley E. Bernstein, and Haruhiko Koseki. 2013. "SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing." *Developmental Cell* 26 (6): 565–77. doi:10.1016/j.devcel.2013.08.016.

Jin, Fulai, Yan Li, Jesse R. Dixon, Siddarth Selvaraj, Zhen Ye, Ah Young Lee, Chia-An Yen, Anthony D. Schmitt, Celso A. Espinoza, and Bing Ren. 2013. "A High-Resolution Map of the Three-Dimensional Chromatin Interactome in Human Cells." *Nature* 503 (7475): 290–94. doi:10.1038/nature12644.

Joshi, Onkar, Shuang-Yin Wang, Tatyana Kuznetsova, Yaser Atlasi, Tianran Peng, Pierre J. Fabre, Ehsan Habibi, et al. 2015. "Dynamic Reorganization of Extremely Long-Range Promoter-Promoter Interactions between Two States of Pluripotency." *Cell Stem Cell* 17 (6): 748–57. doi:10.1016/j.stem.2015.11.010.

Kaaij, Lucas J. T., Michal Mokry, Meng Zhou, Michael Musheev, Geert Geeven, Adrien S. J. Melquiond, António M. de Jesus Domingues, et al. 2016. "Enhancers Reside in a Unique Epigenetic Environment during Early Zebrafish Development." *Genome Biology* 17 (1): 146. doi:10.1186/s13059-016-1013-1.

Kagey, Michael H., Jamie J. Newman, Steve Bilodeau, Ye Zhan, David A. Orlando, Nynke L. van Berkum, Christopher C. Ebmeier, et al. 2010. “Mediator and Cohesin Connect Gene Expression and Chromatin Architecture.” *Nature* 467 (7314): 430–35. doi:10.1038/nature09380.

Kalkan, T., and A. Smith. 2014. “Mapping the Route from Naive Pluripotency to Lineage Specification.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1657): 20130540–20130540. doi:10.1098/rstb.2013.0540.

Koenecke, Nina, Jeff Johnston, Qiye He, Samuel Meier, and Julia Zeitlinger. 2016. “*Drosophila* Poised Enhancers Are Generated during Tissue Patterning with the Help of Repression.” *Genome Research*, November. doi:10.1101/gr.209486.116.

Kondo, Takashi, Kyoichi Isono, Kaori Kondo, Takaho A. Endo, Shigeyoshi Itohara, Miguel Vidal, and Haruhiko Koseki. 2014. “Polycomb Potentiates Meis2 Activation in Midbrain by Mediating Interaction of the Promoter with a Tissue-Specific Enhancer.” *Developmental Cell* 28 (1): 94–101. doi:10.1016/j.devcel.2013.11.021.

Kondo, Takashi, Shinsuke Ito, and Haruhiko Koseki. 2016. “Polycomb in Transcriptional Phase Transition of Developmental Genes.” *Trends in Biochemical Sciences* 41 (1): 9–19. doi:10.1016/j.tibs.2015.11.005.

Ku, Manching, Richard P. Koche, Esther Rheinbay, Eric M. Mendenhall, Mitsuhiro Endoh, Tarjei S. Mikkelsen, Aviva Presser, et al. 2008. “Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains.” *PLoS Genetics* 4 (10): e1000242. doi:10.1371/journal.pgen.1000242.

Laat, Wouter de, and Denis Duboule. 2013. “Topology of Mammalian Developmental Enhancers and Their Regulatory Landscapes.” *Nature* 502 (7472): 499–506. doi:10.1038/nature12753.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biology* 10 (3): R25. doi:10.1186/gb-2009-10-3-r25.

Lanzuolo, Chiara, Virginie Roure, Job Dekker, Frédéric Bantignies, and Valerio Orlando. 2007. “Polycomb Response Elements Mediate the Formation of Chromosome Higher-Order Structures in the Bithorax Complex.” *Nature Cell Biology* 9 (10): 1167–74. doi:10.1038/ncb1637.

Leeb, Martin, Diego Pasini, Maria Novatchkova, Markus Jaritz, Kristian Helin, and Anton Wutz. 2010. “Polycomb Complexes Act Redundantly to Repress Genomic Repeats and Genes.” *Genes & Development* 24 (3): 265–76. doi:10.1101/gad.544410.

Levine, Ariel J., and Ali H. Brivanlou. 2007. “Proposal of a Model of Mammalian Neural Induction.” *Developmental Biology* 308 (2): 247–56. doi:10.1016/j.ydbio.2007.05.036.

Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” *Eprint ArXiv:1303.3997*, ARXIV, , March. http://arxiv.org/abs/1303.3997.

Lim, Linda Shushan, Yuin-Han Loh, Weiwei Zhang, Yixun Li, Xi Chen, Yinan Wang, Manjiri Bakre, Huck-Hui Ng, and Lawrence W. Stanton. 2007. “Zic3 Is Required for Maintenance of Pluripotency in Embryonic Stem Cells.” *Molecular Biology of the Cell* 18 (4): 1348–58. doi:10.1091/mbc.E06-07-0624.

Livyatan, Ilana, Yair Aaronson, David Gokhman, Ran Ashkenazi, and Eran Meshorer. 2015. “BindDB: An Integrated Database and Webtool Platform for ‘Reverse-ChIP’ Epigenomic Analysis.” *Cell Stem Cell* 17 (6): 647–48. doi:10.1016/j.stem.2015.11.015.

Luo, Sai, J. Yuyang Lu, Lichao Liu, Yafei Yin, Chunyan Chen, Xue Han, Bohou Wu, et al. 2016. “Divergent LncRNAs Regulate Gene Expression and

Lineage Differentiation in Pluripotent Cells." *Cell Stem Cell* 18 (5): 637–52. doi:10.1016/j.stem.2016.01.024.

Luo, Zhuojuan, Xin Gao, Chengqi Lin, Edwin R. Smith, Stacy A. Marshall, Selene K. Swanson, Laurence Florens, Michael P. Washburn, and Ali Shilatifard. 2015. "Zic2 Is an Enhancer-Binding Factor Required for Embryonic Stem Cell Specification." *Molecular Cell* 57 (4): 685–94. doi:10.1016/j.molcel.2015.01.007.

Lynch, Magnus D., Andrew J. H. Smith, Marco De Gobbi, Maria Flenley, Jim R. Hughes, Douglas Vernimmen, Helena Ayyub, et al. 2012. "An Interspecies Analysis Reveals a Key Role for Unmethylated CpG Dinucleotides in Vertebrate Polycomb Complex Recruitment." *The EMBO Journal* 31 (2): 317–29. doi:10.1038/emboj.2011.399.

Ma, Wenxiu, and Wing Hung Wong. 2011. "The Analysis of ChIP-Seq Data." In *Methods in Enzymology*, 497:51–73. Elsevier. doi:10.1016/B978-0-12-385075-1.00003-2.

Marks, Hendrik, Tüzer Kalkan, Roberta Menafra, Sergey Denissov, Kenneth Jones, Helmut Hofemeister, Jennifer Nichols, et al. 2012. "The Transcriptional and Epigenomic Foundations of Ground State Pluripotency." *Cell* 149 (3): 590–604. doi:10.1016/j.cell.2012.03.026.

Marson, Alexander, Stuart S. Levine, Megan F. Cole, Garrett M. Frampton, Tobias Brambrink, Sarah Johnstone, Matthew G. Guenther, et al. 2008. "Connecting MicroRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells." *Cell* 134 (3): 521–33. doi:10.1016/j.cell.2008.07.020.

Marwick, Katie, Paul Skehel, Giles Hardingham, and David Wyllie. 2015. "Effect of a GRIN2A de Novo Mutation Associated with Epilepsy and Intellectual Disability on NMDA Receptor Currents and Mg(2+) Block in Cultured Primary Cortical Neurons." *Lancet (London, England)* 385 Suppl 1 (February): S65. doi:10.1016/S0140-6736(15)60380-4.

Matsuda, Kazunari, and Hisato Kondoh. 2014. “Dkk1-Dependent Inhibition of Wnt Signaling Activates *Hesx1* Expression through Its 5′ Enhancer and Directs Forebrain Precursor Development.” *Genes to Cells* 19 (5): 374–85. doi:10.1111/gtc.12136.

McLean, Cory Y., Dave Bristor, Michael Hiller, Shoa L. Clarke, Bruce T. Schaar, Craig B. Lowe, Aaron M. Wenger, and Gill Bejerano. 2010. “GREAT Improves Functional Interpretation of Cis-Regulatory Regions.” *Nature Biotechnology* 28 (5): 495–501. doi:10.1038/nbt.1630.

Mendenhall, Eric M., Richard P. Koche, Thanh Truong, Vicky W. Zhou, Biju Issac, Andrew S. Chi, Manching Ku, and Bradley E. Bernstein. 2010. “GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells.” *PLoS Genetics* 6 (12): e1001244. doi:10.1371/journal.pgen.1001244.

Merzdorf, Christa S. 2007. “Emerging Roles for Zic Genes in Early Development.” *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 236 (4): 922–40. doi:10.1002/dvdy.21098.

Meshorer, Eran, Dhananjay Yellajoshula, Eric George, Peter J. Scambler, David T. Brown, and Tom Misteli. 2006. “Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells.” *Developmental Cell* 10 (1): 105–16. doi:10.1016/j.devcel.2005.10.017.

Min, J. 2003. “Structural Basis for Specific Binding of Polycomb Chromodomain to Histone H3 Methylated at Lys 27.” *Genes & Development* 17 (15): 1823–28. doi:10.1101/gad.269603.

Minoux, Maryline, Sjoerd Holwerda, Antonio Vitobello, Taro Kitazawa, Hubertus Kohler, Michael B. Stadler, and Filippo M. Rijli. 2017. “Gene Bivalency at Polycomb Domains Regulates Cranial Neural Crest Positional Identity.” *Science* 355 (6332): eaal2913. doi:10.1126/science.aal2913.

Morey, Lluis, Alexandra Santanach, Enrique Blanco, Luigi Aloia, Elphège P. Nora, Benoit G. Bruneau, and Luciano Di Croce. 2015. “Polycomb Regulates

Mesoderm Cell Fate-Specification in Embryonic Stem Cells through Activation and Repression Mechanisms." *Cell Stem Cell* 17 (3): 300–315. doi:10.1016/j.stem.2015.08.009.

Nora, Elphège P., Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, et al. 2012. "Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre." *Nature* 485 (7398): 381–85. doi:10.1038/nature11049.

Nord, Alex S., Matthew J. Blow, Catia Attanasio, Jennifer A. Akiyama, Amy Holt, Roya Hosseini, Sengthavy Phouanenavong, et al. 2013. "Rapid and Pervasive Changes in Genome-Wide Enhancer Usage during Mammalian Development." *Cell* 155 (7): 1521–31. doi:10.1016/j.cell.2013.11.033.

Obier, Nadine, Qiong Lin, Pierre Cauchy, Vroni Hornich, Martin Zenke, Matthias Becker, and Albrecht M. Müller. 2015. "Polycomb Protein EED Is Required for Silencing of Pluripotency Genes upon ESC Differentiation." *Stem Cell Reviews* 11 (1): 50–61. doi:10.1007/s12015-014-9550-z.

O'Carroll, D., S. Erhardt, M. Pagani, S. C. Barton, M. A. Surani, and T. Jenuwein. 2001. "The Polycomb-Group Gene Ezh2 Is Required for Early Mouse Development." *Molecular and Cellular Biology* 21 (13): 4330–36. doi:10.1128/MCB.21.13.4330-4336.2001.

Pasini, Diego, Adrian P. Bracken, Jacob B. Hansen, Manuela Capillo, and Kristian Helin. 2007. "The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation." *Molecular and Cellular Biology* 27 (10): 3769–79. doi:10.1128/MCB.01432-06.

Pasini, Diego, Adrian P Bracken, Michael R Jensen, Eros Lazzerini Denchi, and Kristian Helin. 2004. "Suz12 Is Essential for Mouse Development and for EZH2 Histone Methyltransferase Activity." *The EMBO Journal* 23 (20): 4061–71. doi:10.1038/sj.emboj.7600402.

Pasini, Diego, Martina Malatesta, Hye Ryung Jung, Julian Walfridsson, Anton Willer, Linda Olsson, Julie Skotte, et al. 2010. "Characterization of an

Antagonistic Switch between Histone H3 Lysine 27 Methylation and Acetylation in the Transcriptional Regulation of Polycomb Group Target Genes." *Nucleic Acids Research* 38 (15): 4958–69. doi:10.1093/nar/gkq244.

Peng, Jamy C., Anton Valouev, Tomek Swigut, Junmei Zhang, Yingming Zhao, Arend Sidow, and Joanna Wysocka. 2009. "Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells." *Cell* 139 (7): 1290–1302. doi:10.1016/j.cell.2009.12.002.

Phillips-Cremins, Jennifer E., Michael E.G. Sauria, Amartya Sanyal, Tatiana I. Gerasimova, Bryan R. Lajoie, Joshua S.K. Bell, Chin-Tong Ong, et al. 2013. "Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment." *Cell* 153 (6): 1281–95. doi:10.1016/j.cell.2013.04.053.

Plath, Kathrin, Jia Fang, Susanna K. Mlynarczyk-Evans, Ru Cao, Kathleen A. Worringer, Hengbin Wang, Cecile C. de la Cruz, Arie P. Otte, Barbara Panning, and Yi Zhang. 2003. "Role of Histone H3 Lysine 27 Methylation in X Inactivation." *Science (New York, N.Y.)* 300 (5616): 131–35. doi:10.1126/science.1084274.

Rada-Iglesias, Alvaro, Ruchi Bajpai, Sara Prescott, Samantha A. Brugmann, Tomek Swigut, and Joanna Wysocka. 2012. "Epigenomic Annotation of Enhancers Predicts Transcriptional Regulators of Human Neural Crest." *Cell Stem Cell* 11 (5): 633–48. doi:10.1016/j.stem.2012.07.006.

Rada-Iglesias, Alvaro, Ruchi Bajpai, Tomek Swigut, Samantha A Brugmann, Ryan A Flynn, and Joanna Wysocka. 2011. "A Unique Chromatin Signature Uncovers Early Developmental Enhancers in Humans." *Nature* 470 (7333): 279–83. doi:10.1038/nature09692.

Rehimi, Rizwan, Milos Nikolic, Sara Cruz-Molina, Christina Tebartz, Peter Frommolt, Esther Mahabir, Mathieu Clément-Ziza, and Alvaro Rada-Iglesias. 2016. "Epigenomics-Based Identification of Major Cell Identity Regulators within Heterogeneous Cell Populations." *Cell Reports* 17 (11): 3062–76. doi:10.1016/j.celrep.2016.11.046.

Respuela, Patricia, Miloš Nikolić, Minjia Tan, Peter Frommolt, Yingming Zhao, Joanna Wysocka, and Alvaro Rada-Iglesias. 2016. “Foxd3 Promotes Exit from Naive Pluripotency through Enhancer Decommissioning and Inhibits Germline Specification.” *Cell Stem Cell* 18 (1): 118–33. doi:10.1016/j.stem.2015.09.010.

Reynolds, Nicola, Mali Salmon-Divon, Heidi Dvinge, Antony Hynes-Allen, Gayan Balasooriya, Donna Leaford, Axel Behrens, Paul Bertone, and Brian Hendrich. 2012. “NuRD-Mediated Deacetylation of H3K27 Facilitates Recruitment of Polycomb Repressive Complex 2 to Direct Gene Repression.” *The EMBO Journal* 31 (3): 593–605. doi:10.1038/emboj.2011.431.

Riising, Eva Madi, Itys Comet, Benjamin Leblanc, Xudong Wu, Jens Vilstrup Johansen, and Kristian Helin. 2014. “Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide.” *Molecular Cell* 55 (3): 347–60. doi:10.1016/j.molcel.2014.06.005.

Sagai, Tomoko, Masaki Hosoya, Youichi Mizushina, Masaru Tamura, and Toshihiko Shiroishi. 2005. “Elimination of a Long-Range Cis-Regulatory Module Causes Complete Loss of Limb-Specific Shh Expression and Truncation of the Mouse Limb.” *Development (Cambridge, England)* 132 (4): 797–803. doi:10.1242/dev.01613.

Sanulli, Serena, Neil Justin, Aurélie Teissandier, Katia Ancelin, Manuela Portoso, Matthieu Caron, Audrey Michaud, et al. 2015. “Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation.” *Molecular Cell* 57 (5): 769–83. doi:10.1016/j.molcel.2014.12.020.

Schoeftner, Stefan, Aditya K. Sengupta, Stefan Kubicek, Karl Mechtler, Laura Spahn, Haruhiko Koseki, Thomas Jenuwein, and Anton Wutz. 2006. “Recruitment of PRC1 Function at the Initiation of X Inactivation Independent of PRC2 and Silencing.” *The EMBO Journal* 25 (13): 3110–22. doi:10.1038/sj.emboj.7601187.

Schoenfelder, Stefan, Robert Sugar, Andrew Dimond, Biola-Maria Javierre, Harry Armstrong, Borbala Mifsud, Emilia Dimitrova, et al. 2015. “Polycomb Repressive Complex PRC1 Spatially Constrains the Mouse Embryonic Stem Cell Genome.” *Nature Genetics* 47 (10): 1179–86. doi:10.1038/ng.3393.

Sharma, Jaiprakash, Diptendu Mukherjee, Sudheendra N. R. Rao, Soumya Iyengar, Susarla Krishna Shankar, Parthasarathy Satishchandra, and Nihar Ranjan Jana. 2013. “Neuronatin-Mediated Aberrant Calcium Signaling and Endoplasmic Reticulum Stress Underlie Neuropathology in Lafora Disease.” *The Journal of Biological Chemistry* 288 (13): 9482–90. doi:10.1074/jbc.M112.416180.

Shin, Ha Youn, Michaela Willi, Kyung Hyun Yoo, Xianke Zeng, Chaochen Wang, Gil Metser, and Lothar Hennighausen. 2016. “Hierarchy within the Mammary STAT5-Driven Wap Super-Enhancer.” *Nature Genetics* 48 (8): 904–11. doi:10.1038/ng.3606.

Signolet, Jason, and Brian Hendrich. 2015. “The Function of Chromatin Modifiers in Lineage Commitment and Cell Fate Specification.” *FEBS Journal* 282 (9): 1692–1702. doi:10.1111/febs.13132.

Simonis, Marieke, Jurgen Kooren, and Wouter de Laat. 2007. “An Evaluation of 3C-Based Methods to Capture DNA Interactions.” *Nature Methods* 4 (11): 895–901. doi:10.1038/nmeth1114.

Smith, Austin. 2017. “Formative Pluripotency: The Executive Phase in a Developmental Continuum.” *Development* 144 (3): 365–73. doi:10.1242/dev.142679.

Spitz, François, and Eileen E. M. Furlong. 2012. “Transcription Factors: From Enhancer Binding to Developmental Control.” *Nature Reviews Genetics* 13 (9): 613–26. doi:10.1038/nrg3207.

Splinter, Erik, Elzo de Wit, Harmen J.G. van de Werken, Petra Klous, and Wouter de Laat. 2012. “Determining Long-Range Chromatin Interactions for Selected Genomic Sites Using 4C-Seq Technology: From Fixation to

Computation.” *3D Chromatin Architecture* 58 (3): 221–30. doi:10.1016/j.ymeth.2012.04.009.

Stadhouders, Ralph, Petros Kolovos, Rutger Brouwer, Jessica Zuin, Anita van den Heuvel, Christel Kockx, Robert-Jan Palstra, et al. 2013. “Multiplexed Chromosome Conformation Capture Sequencing for Rapid Genome-Scale High-Resolution Detection of Long-Range Chromatin Interactions.” *Nat. Protocols* 8 (3): 509–24. doi:10.1038/nprot.2013.018.

Stern, Claudio D. 2005. “Neural Induction: Old Problem, New Findings, yet More Questions.” *Development (Cambridge, England)* 132 (9): 2007–21. doi:10.1242/dev.01794.

Stock, Julie K., Sara Giadrossi, Miguel Casanova, Emily Brookes, Miguel Vidal, Haruhiko Koseki, Neil Brockdorff, Amanda G. Fisher, and Ana Pombo. 2007. “Ring1-Mediated Ubiquitination of H2A Restrains Poised RNA Polymerase II at Bivalent Genes in Mouse ES Cells.” *Nature Cell Biology* 9 (12): 1428–35. doi:10.1038/ncb1663.

Takahashi, Masanori, and Noriko Osumi. 2002. “Pax6 Regulates Specification of Ventral Neurone Subtypes in the Hindbrain by Establishing Progenitor Domains.” *Development (Cambridge, England)* 129 (6): 1327–38.

Tan, Wen-Hann, Edward C. Gilmore, and Hagit N. Baris. 2013. “Human Developmental Genetics.” In *Emery and Rimoin’s Principles and Practice of Medical Genetics*, 1–63. Elsevier. doi:10.1016/B978-0-12-383834-6.00018-5.

Tee, Wee-Wei, Steven S. Shen, Ozgur Oksuz, Varun Narendra, and Danny Reinberg. 2014. “Erk1/2 Activity Promotes Chromatin Features and RNAPII Phosphorylation at Developmental Promoters in Mouse ESCs.” *Cell* 156 (4): 678–90. doi:10.1016/j.cell.2014.01.009.

Thongjuea, S., R. Stadhouders, F. G. Grosveld, E. Soler, and B. Lenhard. 2013. “R3Cseq: An R/Bioconductor Package for the Discovery of Long-Range Genomic Interactions from Chromosome Conformation Capture and next-

Generation Sequencing Data." *Nucleic Acids Research* 41 (13): e132–e132. doi:10.1093/nar/gkt373.

Tie, Feng, Rakhee Banerjee, Chen Fu, Carl A. Stratton, Ming Fang, and Peter J. Harte. 2016. "Polycomb Inhibits Histone Acetylation by CBP by Binding Directly to Its Catalytic Domain." *Proceedings of the National Academy of Sciences of the United States of America* 113 (6): E744-753. doi:10.1073/pnas.1515465113.

Vernimmen, Douglas, and Wendy A. Bickmore. 2015. "The Hierarchy of Transcriptional Activation: From Enhancer to Promoter." *Trends in Genetics: TIG* 31 (12): 696–708. doi:10.1016/j.tig.2015.10.004.

Vernimmen, Douglas, Marco De Gobbi, Jacqueline A. Sloane-Stanley, William G. Wood, and Douglas R. Higgs. 2007. "Long-Range Chromosomal Interactions Regulate the Timing of the Transition between Poised and Active Gene Expression." *The EMBO Journal* 26 (8): 2041–51. doi:10.1038/sj.emboj.7601654.

Visel, Axel, Edward M. Rubin, and Len A. Pennacchio. 2009. "Genomic Views of Distant-Acting Enhancers." *Nature* 461 (7261): 199–205. doi:10.1038/nature08451.

Wagle, Prerana, Miloš Nikolić, and Peter Frommolt. 2015. "QuickNGS Elevates Next-Generation Sequencing Data Analysis to a New Level of Automation." *BMC Genomics* 16 (1). doi:10.1186/s12864-015-1695-x.

Wani, Ajazul H., Alistair N. Boettiger, Patrick Schorderet, Ayla Ergun, Christine Münger, Ruslan I. Sadreyev, Xiaowei Zhuang, Robert E. Kingston, and Nicole J. Francis. 2016. "Chromatin Topology Is Coupled to Polycomb Group Protein Subnuclear Organization." *Nature Communications* 7 (January): 10291. doi:10.1038/ncomms10291.

Werken, Harmen J. G. van de, Gilad Landan, Sjoerd J. B. Holwerda, Michael Hoichman, Petra Klous, Ran Chachik, Erik Splinter, et al. 2012. "Robust 4C-

Seq Data Analysis to Screen for Regulatory DNA Interactions." *Nature Methods* 9 (10): 969–72. doi:10.1038/nmeth.2173.

Whyte, Warren A., David A. Orlando, Denes Hnisz, Brian J. Abraham, Charles Y. Lin, Michael H. Kagey, Peter B. Rahl, Tong Ihn Lee, and Richard A. Young. 2013. "Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes." *Cell* 153 (2): 307–19. doi:10.1016/j.cell.2013.03.035.

Wijchers, Patrick J., Peter H.L. Krijger, Geert Geeven, Yun Zhu, Annette Denker, Marjon J.A.M. Verstegen, Christian Valdes-Quezada, et al. 2016. "Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments." *Molecular Cell* 61 (3): 461–73. doi:10.1016/j.molcel.2016.01.001.

Wit, Elzo de, and Wouter de Laat. 2012. "A Decade of 3C Technologies: Insights into Nuclear Organization." *Genes & Development* 26 (1): 11–24. doi:10.1101/gad.179804.111.

Yáñez-Cuna, J. Omar, Evgeny Z. Kvon, and Alexander Stark. 2013. "Deciphering the Transcriptional Cis-Regulatory Code." *Trends in Genetics* 29 (1): 11–22. doi:10.1016/j.tig.2012.09.007.

Ying, Qi-Long, Jason Wray, Jennifer Nichols, Laura Batlle-Morera, Bradley Doble, James Woodgett, Philip Cohen, and Austin Smith. 2008. "The Ground State of Embryonic Stem Cell Self-Renewal." *Nature* 453 (7194): 519–23. doi:10.1038/nature06968.

Young, Richard A. 2011. "Control of the Embryonic Stem Cell State." *Cell* 144 (6): 940–54. doi:10.1016/j.cell.2011.01.032.

Zentner, Gabriel E., Paul J. Tesar, and Peter C. Scacheri. 2011. "Epigenetic Signatures Distinguish Multiple Classes of Enhancers with Distinct Cellular Functions." *Genome Research* 21 (8): 1273–83. doi:10.1101/gr.122382.111.

Zhang, Siwei, Jingjing Li, Robert Lea, Kris Vleminckx, and Enrique Amaya. 2014. “Fezf2 Promotes Neuronal Differentiation through Localised Activation of Wnt/β-Catenin Signalling during Forebrain Development.” *Development (Cambridge, England)* 141 (24): 4794–4805. doi:10.1242/dev.115691.

Zhao, Tian-Zhi, Fei Shi, Jun Hu, Shi-Ming He, Qian Ding, and Lian-Ting Ma. 2016. “GPER1 Mediates Estrogen-Induced Neuroprotection against Oxygen-Glucose Deprivation in the Primary Hippocampal Neurons.” *Neuroscience* 328 (July): 117–26. doi:10.1016/j.neuroscience.2016.04.026.

Zhao, Y., H. Z. Sheng, R. Amini, A. Grinberg, E. Lee, S. Huang, M. Taira, and H. Westphal. 1999. “Control of Hippocampal Morphogenesis and Neuronal Differentiation by the LIM Homeobox Gene Lhx5.” *Science (New York, N.Y.)* 284 (5417): 1155–58.

FIGURE INDEX

TABLE INDEX

LIST OF ABBREVIATIONS

AntNPCs	anterior neural progenitor cells
bp	base pair
b-FGF	fibroblast grow factor
BMP	bone morphogenetic protein
BSA	bovine serum albumin
BXC	bitorax complex
cDNA	complementary DNA
CBX	chromobox/ chromo domain protein
CGIs	CpG islands
ChIP	chromatin immunoprecipitation
CHIR	N^6-[2-[[4-(2,4-Dichlorophenyl)-5-(1*H*-imidazol-2-yl)-2-pyrimidinyl]amino]ethyl]-3-nitro-2,6-pyridinediamine
CO_2	carbon dioxide
DNA	desoxiribonucleic acid
DSB	double-strand-break
EDTA	ethylenediaminetetraacetic acid
EED	embryonic ectoderm development (PcG subunit)
EGTA	ethylenediaminetetraacetic acid
ESCs	embryonic stem cells
EZH	enhancer of zeste (PcG subunit)
FBS	fetal bovine serum
For	forward
g	Earth's gravitational force, relative centrifuge force
GSK-3β	glycogen synthase kinase 3β
h	hour
H2AK119ub	Histone two A lysine one hundred nineteen ubiquitunation
H3K4me1	histone three lysine four monomethylation
H3K4me3	histone three lysine four trimethylation
H3K27ac	histone three lysine twenty-seven acetylation

H3K27me3	histone three lysine twenty-seven trimethylation
HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1piperazineethanesulfonic acid
hESCs	human embryonic stem cells
Hox	homeobox genes
IndAll	all genes induced
IndPoitoAct	induced genes that are associated with poised enhancers which become active upon differentiation
IP	immunoprecipitation
LIF	leukemia inhibitory factor
M	molar
MEK	Mitogen-activated kinase
mESCs	mouse embryonic stem cells
$MgCl_2$	magnesium chloride
min	minute
NaCl	sodium chloride
NHEJ	non-homologous end joining
P	p-value
PBS	phosphate buffered saline
PBST	phosphate buffered saline plus tween
PCGF	Polycomb group ring fingers
PCR	polymerase chain reaction
PD	N-[(2R)-2,3-Dihydroxypropoxy]-3,4-difluoro-2-[(2-fluoro -4-iodophenyl)amino]-benzamide
PE	poised enhancer
PFA	paraformaldehyde
PH	polyhomeotic domain (PcG subunit)
PIC	pre-initiation complex
PolII	RNA polymerase II
PoiAct	Poised enhancers that become active
PoinoAct	Poised enhancers that do not become active
PRCs	Polycomb repressive complexes
PRC1	Polycomb repressive complex one

PRC2	Polycomb repressive complexes
PREs	Polycomb response elements
rev	reverse
RNA	ribonucleic acid
RNA polII	RNA polymerase II
rpm	rounds per minute
RT	reverse transcriptase/room temperature
RYBP	RING1 And YY1 Binding Protein
S2	serine two
SAM	sterile alpha motifs
SD	standard deviation
SDS	sodiumdodecylsulfate
sec	second
TAD	topological associated domain
TAE	Tris-acetic acid-EDTA-buffer
TF	transcription factor
TSS	transcription start site
Tween	polyoxethylene-sorbitan-monolaureate
U	unit
V	volt
VEGF	vascular endothelial grow factor
WT	wild-type
YY1	yingyang1 protein
2i	two inhibitors medium
5’	five prime-end of DNA sequences
3’	three prime-end of DNA sequences
3C	chromatin conformation capture
4C	circular chromatin conformation capture

Eidesstattliche Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstaÅNndig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit −einschliesslich Tabellen, Karten und Abbildungen −, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie − abgesehen von unten angegebenen Teilpublikationen − noch nicht veroÅNffentlicht worden ist sowie, dass ich eine solche Veroöffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Herrn Prof. Dr. Björn Schumacher betreut worden.

Köln in Oktober 2017 Sara de la Cruz Molina

Nachfolgende Teilpublikationen liegen vor:

Cruz-Molina, Sara, Patricia Respuela, Christina Tebartz, Petros Kolovos, Milos Nikolic, Raquel Fueyo, Wilfred F.J. van Ijcken, et al. 2017. “PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation.” *Cell Stem Cell*, March. doi:10.1016/j.stem.2017.02.004.