Department of Gene Regulation, Stem Cells and Cancer
Centre for Genomic Regulation
Doctoral Thesis 2016
Universitat Pompeu Fabra

The Polycomb group protein Cbx6 is an essential regulator of embryonic stem cell identity

Dissertation presented by
Alexandra Santanach Buxaderas
For the degree of Doctor in Biomedical Research

Work carried out under the supervision of Dr. Luciano Di Croce in the Epigenetic Events in Cancer Group in the Gene Regulation, Stem Cells and Cancer Program, in the Centre for Genomic Regulation (CRG)

“Dans la vie, rien n'est à craindre,
tout est à comprendre.”

Marie Curie

Abstract

Polycomb group (PcG) proteins are transcriptional repressors that control cell identity and development. In mammals, at least five different CBX (Chromobox) proteins (CBX2, CBX4, CBX6, CBX7, and CBX8) are believed to be associated with the core Polycomb repressive complex 1 (PRC1). CBX6 and CBX7 are the Pc (Polycomb) orthologs expressed in embryonic stem (ES) cells, yet little is known about their function in these cells. Here we have characterized the function of CBX6 in ES cell pluripotency and differentiation. Using a comparative gene expression profile between control and CBX6-depleted ES cells, we show that CBX6 is an essential regulator of ES cell identity, such that ablation of CBX6 function destabilizes the pluripotency network by perturbing the WNT and MAPK signaling pathways, and triggers differentiation towards the ectoderm fate. CBX6 mass spectrometry data and genome-wide chromatin occupancy reveals that CBX6 is unequivocally linked to the canonical PRC1 (cPRC1) complex. However, contrary to expectations, our results indicate that CBX6 also has a non-canonical PRC1 function. Taken together, our findings reveal that CBX6 is an essential component for ES cell biology and that it contributes to the structural and functional complexity of the PRC1 complex.

Les proteïnes del grup Polycomb (PcG, en anglès) són repressors transcripcionals que controlen la identitat cel·lular i el desenvolupament. En mamífers hi ha cinc proteïnes CBX diferents (CBX2, CBX4, CBX6, CBX7 i CBX8) que estan relacionades amb el Complex Repressiu de Polycomb 1 (PRC1, en anglès). CBX6 i CBX7 són les proteïnes ortòlogues de Pc més expressades a les cèl·lules mare embrionàries (ES, en anglès), però tot i així, no se sap la funció de CBX6 en aquestes cèl·lules. Aquí, hem caracteritzat la funció de CBX6 en l'estat de pluripotència i de diferenciació de les cèl·lules mare. Un anàlisi comparatiu de l'expressió gènica entre cèl·lules control i cèl·lules sense CBX6, ens ha permès mostrar que CBX6 es una proteïna reguladora essencial per la identitat cel·lular de les cèl·lules mare, ja que l'abolició de la funció de CBX6 desestabilitza la xarxa de pluripotència, a través de la pertorbació de les vies de senyalització cel·lular WNT i MAPK, la qual cosa estimula la diferenciació cel·lular cap a ectoderm. L'espectrometria de masses de CBX6, així com la seva distribució genòmica a la cromatina, revela que CBX6 està estretament enllaçada amb el complex canònic de PRC1 (cPRC1), tot i així, sorprenentment, els nostres resultats indiquen que també té una funció no canònica (ncPRC1). En conjunt, els nostres resultats revelen que CBX6 és un component essencial per a la biologia de les cèl·lules mare, i contribueix a la complexitat estructural i funcional del complex PRC1.

Table of contents

Results

Discussion

Materials and methods

Figure index

Introduction

Results

Discussion

Table index

Introduction

Results

Materials and methods

INTRODUCTION

1. Mouse embryonic stem (ES) cells

1.1 Origin of ES cells

Mammalian embryonic development starts with fertilization when a single sperm cell penetrates the cell membrane of the oocyte and delivers its haploid genome into the oocyte (**Figure I.1**). After three consecutive rounds of divisions (8-cell state), each of the cells (called blastomeres) in this spherical cell aggregate (also known as morula) undergoes a process of compaction, in which cell adhesion between blastomeres increases and morphology changes. Around the 16-cell stage (E3.0, three days after fertilization), cavitation starts. During this process, a cavity (blastocel) of gradually increasing size is molded, and fluid accumulates within it. The embryo at this stage (E3.5) is now called a blastocyst (32-cell stage). Asymmetrical cell divisions generate two visibly distinct subpopulations of cells, which are clustered on one side of this cavity: the smaller inner cells will give rise to the so-called inner cell mass (ICM), and the outermost cell layer will become the trophoectoderm (TE) lineage[6,7]. In the mid-late blastocyst (E3.75), soon after the ICM and TE lineages have been specified, a second cell fate decision choice results in the formation of two new layers of specialized cells: the epiblast (EPI) and the primitive endoderm (PE)[7].

Therefore, before implantation, the embryo already contains three distinct lineage-restricted subpopulations: the TE and the PE develop into extra-embryonic tissues, such as the placenta, the EPI retains pluripotency and gives rise to the somatic tissues and

the germline of the proper embryo[8,9]. ES cells are derived from the EPI[10,11].

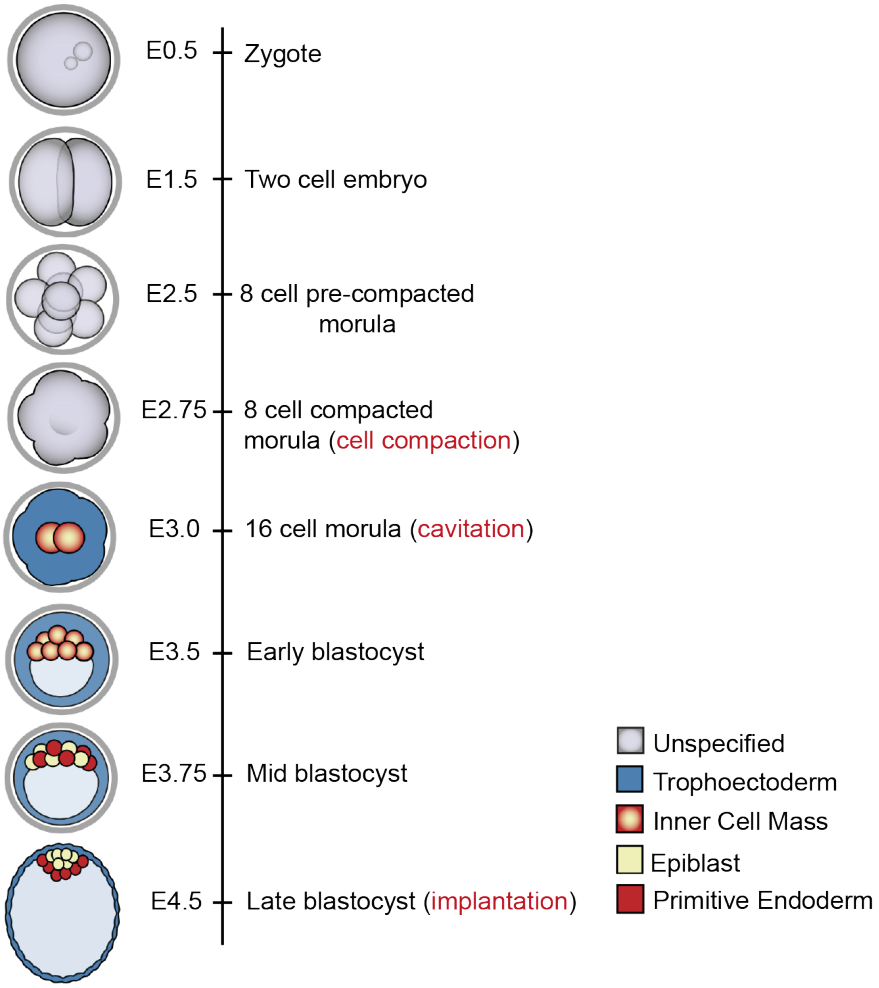


Figure I.1. Developmental stages before implantation. Cell fate lineages are color-coded according to the key at the bottom of the figure.

Two transcription factors, OCT4 (Octamer-binding transcription factor 4) and CDX2 (Caudal type homeobox 2), direct the first cell decision between the TE and the ICM in the embryo[12]. A reciprocal inhibition of the transcriptional activities between the two factors results in a mutually antagonistic expression pattern: CDX2 represses *Oct4* expression in the TE compartment, while OCT4 represses *Cdx2* in the ICM. This mutual silencing is achieved by epigenetic mechanisms: in TE cells, the chromatin remodeling protein BRG1 (BRG1-Associated Factor 190A) cooperates with CDX2 to ensure *Oct4* repression[13],while in ICM cells, the H3K9 histone methyltransferase SETBB1 (SET Domain Bifurcated 1) interacts with OCT4 to silence the expression of *Cdx2* and other TE-associated genes[14].

Similar to the first cell fate decision, this second cell fate decision is also guided by transcription factors. Before implantation, *Nanog* (Nanog Homeobox) and *Gata6* (GATA Binding Protein 6) are expressed in a mutually exclusive manner, which has been described as a "salt-and-pepper" expression pattern, as they inhibit each other's expression[15]. Moreover, NANOG depletion in ES cells induces expression of PE markers, suggesting that only repressing *Nanog* is sufficient to induce PE differentiation, and that forcing expression of *Gata6* causes *Nanog* downregulation, concomitant with an aberrant differentiation of the ICM towards PE cells. The FGF/RTK signaling pathway plays an essential role in arbitrating this battle between NANOG and GATA6, as all the ICM cells express NANOG and mutants of *Grb2* (an essential protein for the transduction in this signaling pathway) fail to generate PE cells. Furthermore, NANOG-expressing cells induce

full PE differentiation in a non–cell-autonomous manner by activating the expression and secretion of FGF4 to the extracellular medium, which in turn triggers the expression of later markers of PE in *Gata6*-expressing cells, thereby reinforcing PE cell identity[16,17].

1.2 Features that define ES cells

Mouse ES cells are stable cell lines derived from the inner cell mass (ICM) of the pre-implantation embryo at the blastocyst stage. ES cells have three defining properties: self-renewal, pluripotency, and primary chimera formation. Pluripotency is the capacity to differentiate into all the cell lineages of the developing organism (endoderm, mesoderm, and ectoderm). Self-renewal is the ability to proliferate indefinitely in the same state. ES cells were firstly isolated by Evans in 1981[10] and since then have been extensively studied. After sustained expansion in culture, ES cells retain full responsiveness to differentiation stimuli and show no intrinsic biases in the generation of germline/somatic lineages when re-introduced into a receptor blastocyst, being able to give rise to a complete organism. Understanding the molecular mechanisms that govern the ES cell state is of great interest not only for basic research—for instance, ES cells represent an impeccable system to study cellular differentiation *in vitro*—but also for their potential implications in human health—these mechanisms are likewise involved in cancer progression and could be exploited for regenerative medicine as well.

1.3 ES cell transcriptional regulatory network

1.3.1 Transcription factors maintaining pluripotency

In ES cells, the pluripotent state is largely governed by the core transcription factors OCT4, SOX2, NANOG, and KLF4.
The transcription factor OCT4 (Octamer-binding transcription factor 4) was one of the first factors identified as a master regulator of ES pluripotency[18,19]. It is highly expressed in the blastocyst ICM[20], and its expression rapidly decreases during embryo development; indeed, OCT4 is widely used as a reporter marker to assess differentiation[21]. Accordingly, *Oct4* null embryos fail to form a pluripotent ICM and do not develop beyond the blastocyst stage[20]. *Oct4* expression is tightly regulated, and reduction by only 50% causes ES cells to spontaneously differentiate towards the trophoectoderm (TE) lineage[22].

Another master regulator of ES cell pluripotency is SOX2 (Sry-box–containing gene 2)[23]. Similar to *Oct4*, *Sox2* is also highly expressed in ICM, where it has the essential role of retaining the maximum pluripotency capacity of the ES cells[24]. Changes of *Sox2* expression in ES cells also trigger differentiation: overexpression induces differentiation towards neuroectodermal cells[25], whereas SOX2 deletion results in TE differentiation. Furthermore, targeted disruption of *Sox2 in vivo* results in peri-implantation lethality due to a strong impairment of the ICM[23]. Importantly, *Sox2* expression is not restricted to ES cells: in adult mice, *Sox2* expression is maintained in many adult stem/progenitor stem cells[23,26].

NANOG is another transcription factor involved in the ES cells pluripotency network[27]. *NANOG* is highly expressed not only in the ICM but also in the epiblast cells of the embryo[28]. Similarly to *OCT4* and *SOX2*, *NANOG*-null embryos fail to develop. However, in contrast to those from *OCT4*- and *SOX2* null mutants, ES cells from *NANOG* mutant embryos can be derived and maintained in culture, indicating that *NANOG* is dispensable for ES cell self-renewal[29].

KLF4 (Kruppel-Like Factor 4) has also been shown to participate in the ES cell biology, although its expression is not restricted in the early embryo, as it has been shown to be transcribed in adult tissues, especially in the epithelium of some organs (such as in intestine and epidermis). Overexpression of KLF4 is sufficient to maintain the undifferentiated state of mouse ES cells in the absence of LIF; however, KLF4 is not essential for ES cell self-renewal, as other *Klf* gene family members play a redundant role with KLF4. Triple knockout of *Klf2*, *Klf4* and *Klf5* resulted in defective self-renewal phenotype[30].

OCT4, SOX2, NANOG, and KLF4 (shorted to OSNK) do not work independently but rather are involved in an intricate regulatory circuitry, which also involves other transcription factors[31,32] (**Figure I.2**). Chromatin immunoprecipitation sequencing (ChIP-seq) experiments revealed that pluripotency factors co-occupy gene regulatory elements in a large spectrum of genes, in what it is called multiple transcription factor–binding loci (MTL)[33]. Chen and colleagues resolved 3,583 MTLs, 43.4% of which were co-occupied by OCT4, SOX2, and NANOG, indicating a functional cooperation in gene regulation[34]. Moreover, these pluripotency

factors positively regulate their expression by binding to their own promoters, co-occupying and activating expression of other genes essential to maintain ES cell pluripotency, and cooperating to repress lineage-specific transcription factors, which must be silenced to prevent exit from pluripotency and spontaneous differentiation.

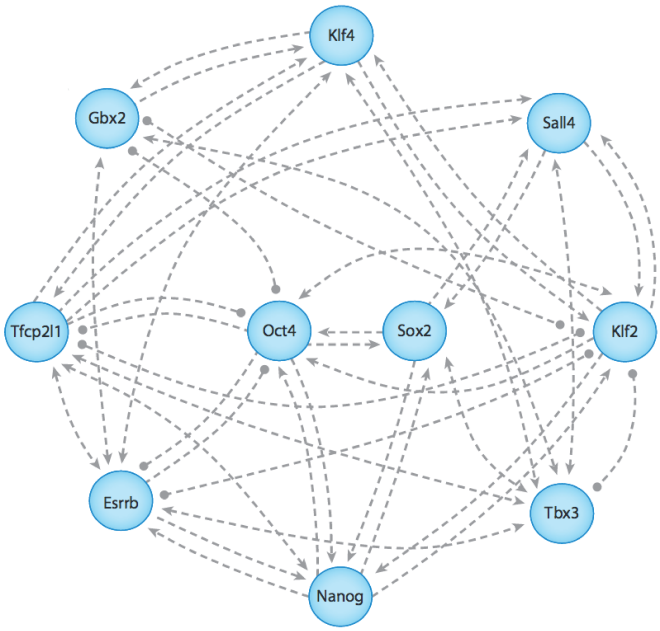


Figure I.2. Pluripotency transcriptional network. Nodes of the pluripotency circuitry are depicted in blue. Note that pluripotency factors OCT4, SOX2, KLF4, and NANOG regulate each other. Dashed lines indicate potential interactions between nodes, inferred from correlated expression[1].

The ability of OSNK to positively or negatively regulate gene expression relies on their ability to interact with specific transcription factors and epigenetic machineries. Much effort has

been made recently to characterize the interactome of these factors[21,35,36]. Despite a few differences between those reports, all of them converge in identifying numerous associated proteins, including nucleosome-remodeling complexes (such as SWI/SNF and NURD[37]), histone methyltransferases (i.e. SETDB1 and WDR5[14]), enhancer-associated factors (i.e. Mediator[38]), and pluripotency factors[35,37,39]. For instance, it has been shown that Oct4 and Nanog associate with proteins of the NURD complex, MTA1/2 and HDAC1/2, to compose a unique complex (termed NODE) that has a deacetylation activity comparable to that of the NURD complex, connecting OCT4 and NANOG to repressor functions[37]. In contrast, OSNK have also been reported to have a strong participation in actively transcribed regions in the genome. Indeed, OCT4, SOX2, and NANOG were shown to recruit Mediator, and therefore RNA polymerase II, to activate transcription of many genes that ultimately characterize ES cell biology[38].

1.3.2 Extrinsic signaling in mouse ES cells

A central question in stem cell biology is to understand how extrinsic signaling pathways modulate ES cell pluripotency and differentiation. Since ES cells were firstly derived in 1981[10], a large effort has been made to develop laboratory cell culture conditions that resemble ICM conditions. This did not occur until 1988, when the function of the cytokine LIF (Leukemia inhibitory factor) was discovered[40]. LIF is secreted by fibroblasts that are co-cultured with the ES cells (the so-called feeder cells) and is indispensable for the proliferation of ES cells cultured in media supplemented with fetal bovine serum (FBS)[40,41]. LIF binds to the

LIFR/gp130 receptor, which drives activation of the JAK-STAT pathway and expression of STAT3 (Signal Transducer And Activator Of Transcription 3)[42]. STAT3 forms a dimer and regulates the expression of pluripotency factors, like KLF4, NANOG, and TFCP2L1 (Transcription Factor CP2-Like 1)[43-45]. Although LIF is essential to derive and maintain ES cells, LIF $^{-/-}$ embryos are viable and fertile[46].

The fact that LIF can only sustain self-renewal when ES cells are cultured with FBS suggested that the serum contains essential components to prevent exit of pluripotency. Indeed, Smith and colleagues discovered that the BMP4-mediated (Bone Morphogenetic Protein 4) signaling pathway supports pluripotency in the presence of LIF. The role of BMP4 has been extensively studied during development and was found to induce expression of Id proteins in pluripotent ES cells, which in turn indirectly inhibit the expression of the bHLH (basic helix-loop-helix) factors involved in differentiation[47]. In agreement with the functional link between Id proteins and BMP4, overexpression of Id proteins is sufficient to keep pluripotency in cells cultured with only LIF and without BMP4 or serum[48].

Other soluble factors, like WNT (Wingless-Type MMTV Integration Site), signal stem cells to continue self-renewal. The activation of canonical WNT is mediated through the cytoplasmic pool of β-Catenin (Cadherin-Associated Protein Beta 1) molecules, which is determined by the presence of WNT ligands. Briefly, in the absence of WNT ligands, β-catenin is phosphorylated by GSK3B (Glycogen synthase kinase 3b)

(destruction complex), which leads to β-Catenin ubiquitination and degradation. The presence of exocrine or autocrine WNT ligands leads to the interaction between the ligand and its receptor. This interaction then results in the inhibition of the destruction complex and the consequent increase in the cytoplasmic pool of functional β-catenin, which translocates into the nucleus to acts as a coactivator for transcription.

Another signaling pathway that regulates lineage specification in the embryo is the FGF pathway. FGF4 is the main FGF factor expressed in the early stages of the embryo and in ES cells[49]. This stem cell–specific expression pattern relies on the presence of a distant enhancer in the *Fgf4* promoter that is under the control of the pluripotency factors OCT4 and SOX2, which positively activate its transcription[49]. The function of FGF4 in ES cells has also been extensively studied. FGF4, through its activation of the ERK1/2 signaling cascade, acts as an auto-inductive stimuli for ES cells to exit self-renewal and to enter into the differentiation program. *Fgf4*-null cells can be cultured and do not show any apparent defect, but they do not differentiate when they are cultured in media that induces neuronal or mesodermal differentiation. This suggests that FGF4 signaling, acting through the ERK1/2 pathway, may act as one of the primary signals to direct cell fate commitment[50]. Accordingly, ERK1/2 inhibitors have extensively been used to improve ES cell derivation efficiency[51]. ES cells maintained in serum and LIF are often morphologically heterogeneous and express the pluripotency factors in a heterogeneous manner, indicating that the pluripotent state is unstable[33,34]. To overcome this limitation, Ying and colleagues defined specific culture conditions in which ES cells are more

protected from differentiation stimuli, thereby resembling transcriptional status of the ES cells in the blastocyst[52].

This serum-free medium contains LIF and inhibitors of two kinases, PD0325901 and CHIR99021. PD0325901 targets the MEK (Mitogen-activated protein kinase), which is upstream of the ERK1/2 kinases, and inhibits the autoinductive stimulation by the FGF4 factor. CHIR99021 inhibits GSK3. As mentioned, inhibition of GSK3 alleviates transcriptional repression of the pluripotency network and increases ES cell resistance to differentiation[53]. Combination of these molecules helps to maintain a naïve ground state of ES cells (**Figure I.3**)[52].

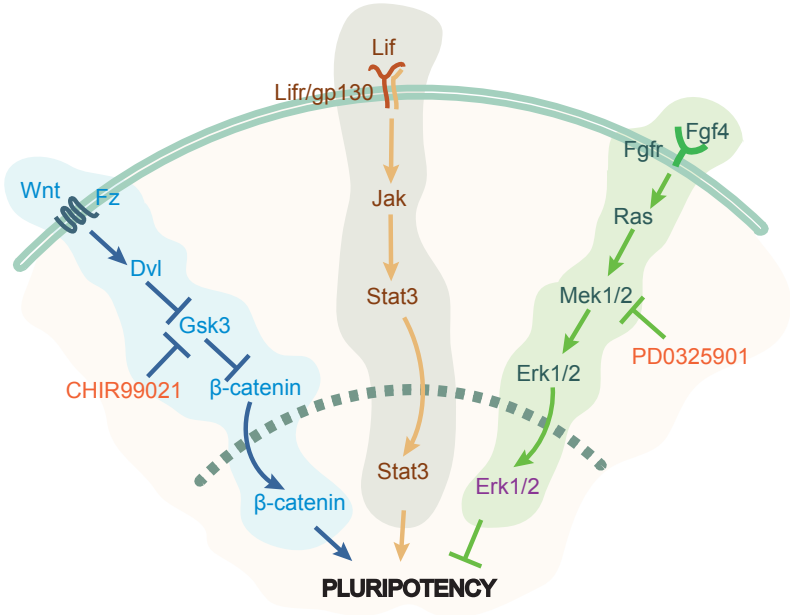


Figure I.3. Extrinsic signaling pathways govern ES cell pluripotency. ES cell biology is under the regulation of extrinsic stimuli like WNT, LIF, and FGF4. Each of these molecules activates signaling pathways that have an effect on transcription. The balance between these pathways will ultimately determine ES cell commitment.

2. Chromatin organization and epigenetics in transcription

2.1 Chromatin structure

Mammalian cells contain 1.7 meters of DNA that needs to be packed into a 5-micrometre nucleus. To achieve this, DNA is highly condensed in a nucleoprotein complex named chromatin. The basic repeat element of chromatin is the nucleosome, each of which wraps 147 bp of DNA around a histone octamer. These octamers consist of the histones H3, H4, and two H2A-H2B dimers[54]. Nucleosomes are connected by the internucleosomal linker DNA that is bound by histone H1, which is involved in establishing higher-order structures[55]. Chromatin can then be further condensed by supercoiling DNA. Chromatin is a dynamic structure that not only helps to package the entire eukaryotic genome into the nucleus but also allows cell division, supports chromosome integrity, and regulates the accessibility of DNA for gene transcription, DNA replication and repair.

Chromatin can be classified according to its compaction state and transcriptional potential. Heterochromatin was originally defined as regions of nuclei that stained strongly with basic dyes during cell cycle division, whereas euchromatic regions changed their degree of condensation throughout the cell cycle[56]. Today, the term heterochromatin designates a highly condensed chromatin conformation, inaccessible for transcription factors and thus transcriptionally less active than its counterpart euchromatin. Euchromatin corresponds to a rather open and transcriptionally active chromatin conformation[57]. These regions of open chromatin permit transcription factors and the basal transcription

machinery to access DNA, thereby allowing gene transcription to occur[58,59].

2.2 Epigenetics

The term epigenetics refers to the changes in chromatin that impact gene expression or chromatin conformation without affecting the underlying genetic sequence. Epigenetics include processes such as DNA methylation, histone variants, or histone modifications[60].

2.2.1 Histone post-transcriptional modifications

Histones can be modified through the action of specific enzymes that can deposit different chemical groups on the N-terminal histone tail. Methylation, acetylation, phosphorylation, ubiquitination, and sumoylation are some of the post-translational modifications (PTM) found in the chromatin (**Figure I.4**)[61]. The combination of these marks (also known as histone code), confers chromatin with particular features that have an specific outcome at the level of chromatin accessibility and the transcriptional status of a gene during cellular events, such as development and differentiation[5]. Some of these modifications are related to active transcription, such as trimethylation of lysine 4 in histone H3 (H3K4me3) or acetylation of lysine 27 of histone H3 (H3K27ac). Other modifications are associated with gene silencing, such as H3K27me3 and H3K9me3.

Histone modifications can influence chromatin conformation directly, as some of these modifications confer histones with a positive charge, which affects to the interaction with the

negatively-charged DNA, causing repulsion. Acetylation is the best-known modification that alters chromatin conformation[62]. Acetylated regions are opened and actively transcribed regions, as the positive charge of acetyl groups induces unfolding of the chromatin. Histone modifications can also exert their effects by binding effector proteins. Different modifications are recognized by specific domains of the effector molecules. The recruited proteins can then mediate their function at chromatin, such as catalyzing other histone modifications, activating or repressing gene transcription, or repairing DNA, among others.

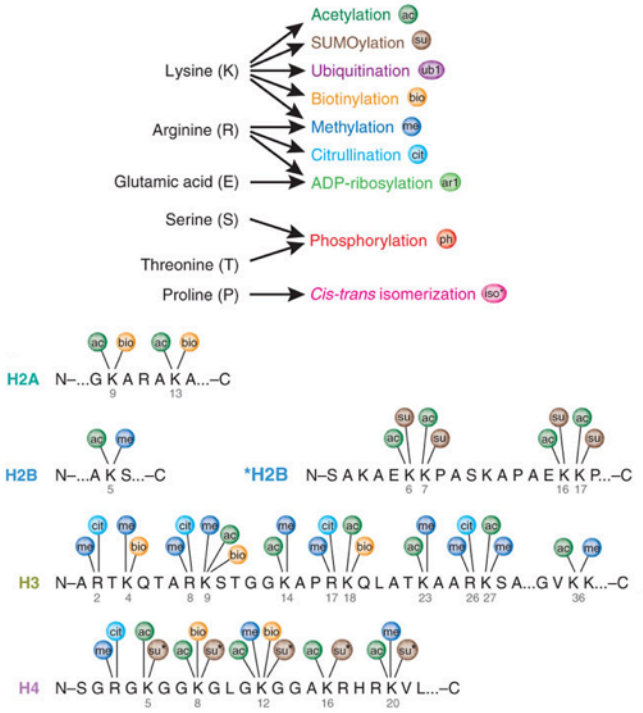


Figure I.4. Known post-translational modifications and the amino acid residues they modify. Histone residues can be modified through different enzymes. Each modification inhibits subsequent modifications. ac, acetylation; bio, biotinylation; cit, citrullination; me, methylation; su, SUMOylation[5].

2.3 Chromatin structure and chromatin regulators in ES cells

ES cells possess a unique nuclear architecture and display specific chromatin features, which confer them with enough plasticity to be capable of differentiating into any of the cell types present in the adult organism.

ES cell nuclei are larger, a looser chromatin organization, than those from differentiated cells, which allows a rapid response to any differentiation trigger[2]. The idea that the chromatin of ES cells is more "open" is somewhat reflected by the ratio between euchromatin and heterochromatin in ES cells, which is higher than in differentiating cells[63]. Much work supported this evidence, including that of Park and colleagues, who were able to visualize by electron microscopy that heterochromatin was a minor fraction in ES cells but prevalent in differentiated cells[2,64]. Importantly, this was demonstrated *in vivo* in cells from the ICM of mouse blastocysts at E3.5: using electron spectroscopic imaging (ESI), it was shown that there were changes in chromatin structure in the early embryo, demonstrating that pluripotent cells have a less condensed chromatin than lineage-committed cells[65] (**Figure I.5**). Moreover, light microscopy observations revealed marked changes in the number, size, and distribution of heterochromatin foci[63].

Another distinctive ES cell chromatin feature is that they have relatively low nucleosome repeat length (the spacing between two adjacent nucleosomes), which is one of the determining factors for chromatin structure[55].

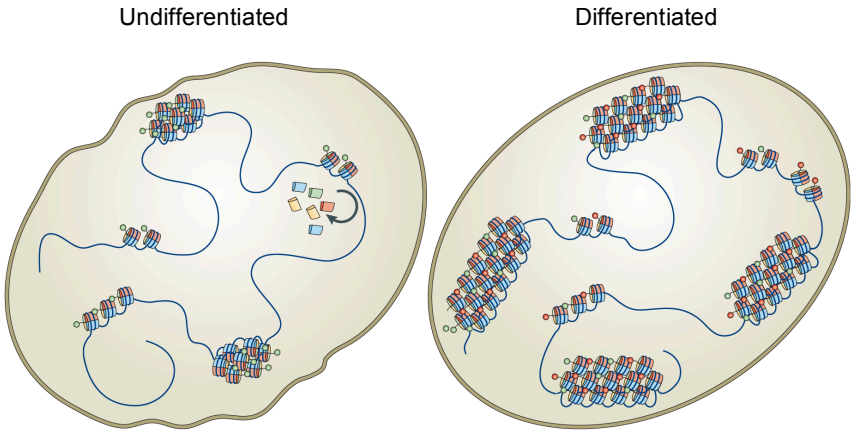


Figure I.5. Chromatin structure during ES cell differentiation. In pluripotent ES cells (left), chromatin is globally decondensed and contains proteins loosely bound and is enriched in active histone marks (green circular tags). As the cells differentiate (right), chromatin is more compacted (heterochromatin), repressive histone marks accumulate (red circular tags), and chromatin proteins become stably associated[2].

Chromatin structure itself and the position of nucleosomes can be altered both globally and at the level of specific gene loci by ATP-dependent nucleosome remodeling complexes, which are multiprotein complexes of variable compositions[61]. Using energy from ATP hydrolysis, these enzymes relocate nucleosomes through sliding mechanisms and nucleosome eviction, induce changes in nucleosomes conformation and favoring the interchange of canonical histones by histone variants[66-68]. Genomic disruption of chromatin remodeling proteins results in premature embryonic death at the blastocyst stage prior to implantation[69-72]. Interestingly, the expression of many of these ATP-dependent chromatin remodeling enzymes is significantly enriched in undifferentiated ES cells. They cooperate with

pluripotency factors in gene expression regulation to actively maintain chromatin in an open state through the disassembly of nucleosomes and the unwinding of higher-order chromatin structures[73]. As mentioned above, ATP-dependent nucleosome remodeling factors are also implicated in core histone replacement by histone variants, which can also contribute to maintaining open chromatin[55,70].

2.4 Histone modifications and DNA methylation in ES cells

Undifferentiated ES cells have a unique epigenetic landscape. ES cells have a globally open chromatin structure with abundant levels of histone modifications related to active transcription, such as histone H3K4me3 and global levels of acetylated histones H3 and H4[74,75]. In contrast, the H3K9me3 mark, which is associated with a repressive chromatin state, is maintained at low levels in ES cells[63]. ChIP-seq studies showed that H3K9me3 expands from around 4% genome coverage in ES cells to 12% in differentiated cells, and a similar pattern is observed for the H3K27me3 mark[76]. Therefore, differentiation is accompanied by a transition towards a less permissive chromatin.

ES cells have the ability to respond quickly to differentiation stimuli. This capacity is in part attributed to a very specific epigenetic trait known as bivalency[77]. Bivalent chromatin is enriched for histone H3K27me3 and H3K4me3, which are marks associated with transcriptionally inactive and active chromatin, respectively. The coexistence of these modifications with opposite roles is thought to poise bivalent genes (expressed at very low levels in ES cells), which allows the rapid activation of lineage

specification genes through loss of H3K27me3 when differentiation is induced.

DNA methylation is a general indicator of silenced genome regions and is a major epigenetic factor, influencing gene regulation and chromatin structure. In mammals, nearly all DNA methylation occurs at cytosine residues within CpG dinucleotides. Immediately after fertilization, the genome undergoes a dramatic demethylation process that reaches its minimum at the early blastocyst stage (E3.5) (from where ES cell are extracted). ES cells freshly derived from ICM contain a globally hypomethylated genome when first cultured, but gradually acquire methylation marks during prolonged passaging. Methylation is resumed following implantation of the embryo, as differentiation starts.

3 Polycomb repressive complexes

3.1 Polycomb repressive complexes in transcriptional regulation

The open and permissive chromatin structure in ES cells requires the action of chromatin repressors, in order to prevent untimely transcription of differentiation genes, as well as for the orchestration of gene expression during differentiation. The Polycomb group proteins (PcG) have now been determined to play essential roles in this regulation. PcGs were originally identified in *Drosophila melanogaster* as transcriptional repressors required for the correct spatiotemporal expression of developmental regulators along the body axis, as mutant flies develop homeotic transformations[78]. Characterization of mutants showing similar phenotypes enabled the identification of 18 PcG genes in *Drosophila*. The number of PcG ortholog genes expanded remarkably during the metazoan evolution, from 18 to 37 members in mammals, most probable due to duplication events [79].

The different PcG proteins associate in complexes that are classified in two major functional groups, the Polycomb repressive complex 1 (PRC1) and the Polycomb repressive complex 2 (PRC2)[80,81], both of which are catalytically active. PRC2 contains a histone methyltransferase enzyme that deposits di- and tri-methylation on lysine 27 of histone H3 (H3K27me2/3)[82]. PRC1 contains an E3-ligase subunit, which monoubiquitinates histone H2A at lysine 119 (H2AK119ub).

The molecular mechanisms of Polycomb-mediated gene repression are constantly under examination and debate by the

scientific community. The classical mode of action of Polycomb complexes is the hierarchical model, in which PRC2 complex is initially recruited to chromatin by co-factors and/or RNA to deposit H3K27me3[83]. Once H3K27me3 is established, the PRC1 complex is recruited to chromatin by binding this mark, where it ubiquitinates histone H2A on lysine 119 (**Figure I.6**).

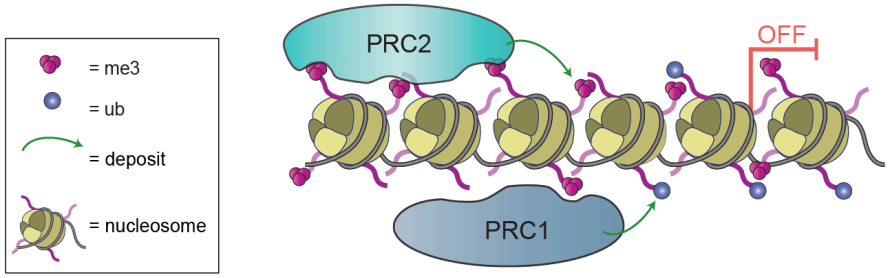


Figure I.6. Polycomb-mediated gene repression is a sequencial mechanism. PRC2 is recruited to its target genes, where it trimethylates histone H3 on lysine 27. H3K27me3 provides a docking site for PRC1, which then monoubiquitinates H2A on lysine 119.

While this model is widely established, other Polycomb recruitment mechanisms should exist, as the hierarchical model does not explain for instance why a functional PRC1 complex is still recruited to the inactivated X chromosome in differentiating female ES cells even in the absence of PRC2[84], or why H2AK119ub levels remain unaffected in ES cells that lack a functional PRC2 complex[84]. Recently, it has been proposed that a non-canonical PRC1 complex, containing RYBP, is able to deposit H2AK119ub independently on the PRC2 complex[85].

Based on a *de novo* targeting system in ES cells, it has been

proposed that the PRC2 complex can be recruited to chromatin in a PRC1-dependent manner, in which PRC1 complexes are first recruited to DNA, where they monoubiquitate H2AK119 to drive PRC2 recruitment[86]

Much evidence supports that a biological function of H2AK119ub is necessary for gene repression and chromatin compaction, yet the exact mechanism of action is still poorly understood. Koseki and colleagues showed that H2A ubiquitination activity of PRC1 is dispensable for its target binding and its activity to compact chromatin at the *Hox* loci, but that it is indispensable for efficient repression of target genes—and thereby ES cell[87]. *In vitro* studies show that Polycomb complexes induce condensation of nucleosomal arrays[88]. Moreover, H2AK119ub has also been implicated in restraining RNAPII (RNA polymerase II) at bivalent regions. This would suggest that the poised RNAPII configuration is enforced by PRC1 complex–mediated ubiquitination of H2A, as conditional deletion of PRC1 complex leads to the sequential loss of H2AK119 ubiquitination and subsequent gene derepression[89]. H2AK119ub also serves as a recruitment platform for ZRF1 (Zuotin-related factor 1). Upon differentiation of human teratocarcinoma and ES cells, ZRF1 binds to H2AK119ub and displaces the PRC1 complex from chromatin, resulting in the upregulation of lineage specification genes[90,91]. Finally, *in vitro* experiments suggest that the presence of H2AK119ub specifically prevents H3K4 methylation[92]. These data collectively demonstrate that multiple effector mechanisms exist by which H2A ubiquitination triggers chromatin compaction.

3.2 PRC1

The initial mammalian PRC1 complex purified from HeLa cells contained four subunits: one of the chromobox proteins (Cbx2,4,6-8) subunit, homologous to Pc (Polycomb) in *Drosophila*; one enzyme similar to dRING (Ring finger protein), of either RING1A or RING1B; one of the three PH (Polyhomeotic homolog) homologs (PHC1–3); and one of the six PCGF1–6 (Polycomb group of ring finger) PSC (Protein posterior sex combs) homologs. The mammalian PRC1 family is currently subdivided into two subfamilies: the canonical PRC1 complex (cPRC1) and the non-canonical PRC1 complex (ncPRC1). This classification is mainly based on the presence of one Cbx protein in the cPRC1 complexes, and the presence of RYBP (Yy1-binding protein) in the ncPRC1 complexes. The core components of both cPRC1 and ncPRC1 include the E3 ubiquitin ligase Ring1A/B and one PCGF subunit.

With the intention to elucidate and better understand the PRC1 complex classification, Gao and collaborators defined six major groups of PRC1 complexes according to the Pcgf subunit associated to it: PRC1.1, PRC1.2, PRC1.3, PRC1.4, PRC1.5, and PRC1.6[93]. Thus, each complex contains a distinct Pcgf subunit, a RING1A/B ubiquitin ligase, and a unique set of associated polypeptides (**Figure I.7**).

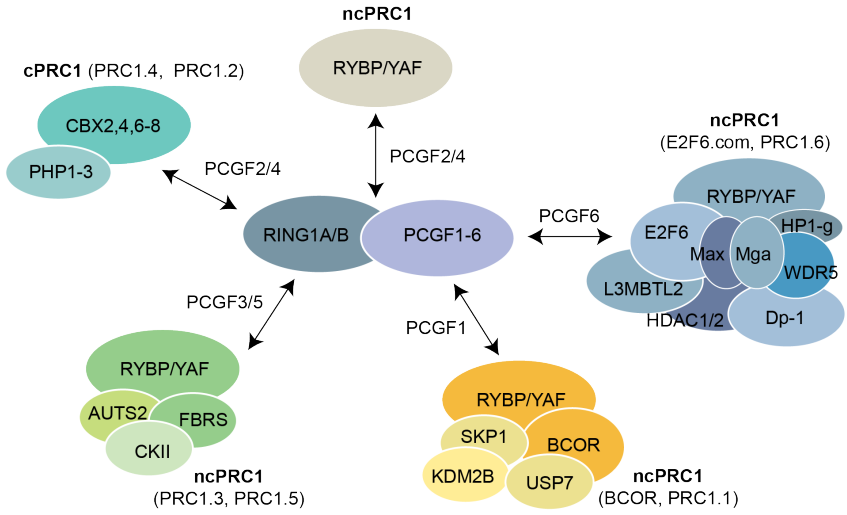


Figure I.7. Composition of the PRC1 complexes. Core PRC1 complex comprises Ring1A/B and one Pcgf subunit. The PRC1 family is subdivided into two subfamilies of canonical PRC1 (cPRC1) and non-canonical PRC1 (ncPRC1), which include an heterogenous group of complexes with distinct functions (adapted from [4]) .

The CBX2,4,6-8 proteins determine the nature of the cPRC1 complex (according to the Pcgf classification, these complexes are PRC1.2 and PRC1.4). CBX proteins contain a chromodomain that recognizes the histone H3K27me3 site, which renders CBX proteins essential for recruiting the cPRC1 complex to its target sites. In ES cells, CBX7 is the main PRC1-associated CBX protein and is necessary to safeguard the pluripotency capacity, as CBX7-containing PRC1 represses a large number of genes involved in lineage commitment. Interestingly, OCT4 binds to the *Cbx7* promoter in pluripotent ES cells[38], thereby putting CBX7 under direct regulation of the pluripotency network. cPRC1 and

the CBX family of proteins will be extensively discussed below (Section 4 in the Introduction).

The PRC1.1 complex, also named BCOR (BCL6 Corepressor), has been recently characterized in ES cells[94]. This complex contains FBXL10 (F-Box And Leucine-Rich Repeat Protein 10) a H3K36me3 lysine demethylase, RYBP, PCGF1 and RING1B. Interestingly, FBXL10 is a H3K36me3 demethylase and contains a CXXC domain. Chromatin recruitment of this ncPRC1 complex depends on the CXXC domain of FBXL10, which drives this ncPRC1 to CpG islands. Importantly, neither ncPRC1 recruitment nor its activity depends on FBXL10 histone demethylase activity [95,96].

The RYBP-PRC1 complex is similar to the cPRC1 complex but contains RYBP, or its homolog YAF2, rather than a CBX protein. In ES cells, although there is some overlap between Cbx7-PRC1 and RYBP-PRC1 regarding target genes, these two complex regulate two different subsets of genes in substantially different ways. RYBP-PRC1 is mostly implicated in modulating the expression of metabolic and cell cycle related genes, whereas CBX7-PRC1 represses classical Polycomb genes, such as lineage specification genes. Another difference is that RYBP-PRC1 allows promoter activity, as these are co-occupied by elongating RNAPII at its target genes; however CBX7-PRC1 target genes are fully repressed.

The PRC1.6 complex is also known as the E2F6 complex due to the presence of this transcriptional repressor factor. This complex also contains the oncoprotein L3MBTL2 (Lmbt-Like 2 Isoform 2), which has recently been described to be indispensable to compact the chromatin in a histone modification–independent manner. Moreover, L3MBTL2 depletion resulted in increased

expression of target genes[97]. The complex is completed by its association with HP1γ, the histone deacetylases HDAC1/2, the transcription factors MAX (MYC Associated Factor X), MGA (MAX Dimerization Protein), and Dp-1 (DRTF1-Polypeptide 1), the H3K9 methyltransferase G9A, the H3K4 demethylase JARID1C, and WDR5 (WD Repeat Domain 5).

The PRC1.5 complex has been shown in neurons to associate with CK2 (Casein kinase 2) and AUTS2 (Autism susceptibility candidate 2). Phosphorylation of RING1B by CK2 results in the catalytic inhibition of RING1B. Interestingly, this complex is associated to actively transcribed genes[97].

3.3 PRC2

The PRC2 complex is composed of three core protein subunits: SUZ12 (Suppressor of zeste 12), which contains a Zinc finger domain; EED (Embryonic ectoderm development), which contains a WD40 that recognizes trimethylated peptides; and EZH1/2 (Enhancer of zeste 1 or 2), which contains the SET domain responsible for the di/trimethylation for the histone H3K27 (**Figure I.8**). The PRC2 complex architecture is not as heterogeneous as the PRC1 complex, yet several co-factors have recently been identified to associate with PRC2 members in ES cells[98]. These partners modulate the enzymatic activity of the complex and/or regulate its recruitment to chromatin. RBBP4/7 (retinoblastoma binding protein 4 and 7) display affinity for histones and probably contribute to PRC2 binding to chromatin[99]. AEBP2 (Adipocyte enhancer-binding protein) enhances enzymatic activity of PRC2, and its DNA binding motif may be necessary to target the complex to specific genomic loci[100]. In 2010, the Jumanji protein

JARID2 (Jumonji AT-rich interactive domain 2) was identified as the first protein to interact directly, via its interaction with SUZ12, with the core PRC2 complex in ES cells. JARID2 co-occupies a large set of PRC2 target genes, and its depletion greatly reduces PRC2 occupancy at promoters. Whether JARID2 positively or negatively regulates PRC2 activity in ES cells is under debate[101,102]. JARID2 contains an AT-rich interaction domain and was therefore postulated to be a recruiter to PRC2, yet to date it has not been demonstrated that JARID2 can directly bind to DNA. Thus, it seems that JARID2 is a component, but not a recruitment factor, of a specific PRC2 complex in ES cells.

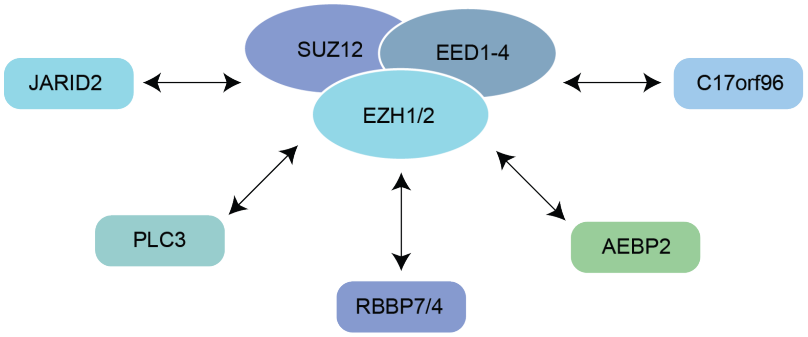


Figure I.8. Composition of the PRC2 complex. The core PRC2 complex contains either Ezh1 or Ezh2, Suz12 and Eed. The core complex can associate with other factors that will affect PRC2 stability and function (Jarid2, Pcl3, Rbbp7/4, Aebp2 and C17orf96) (adapted from[4]).

In *Drosophila*, the Polycomb-like protein (PCL) interacts with Su(z)[102], and PCL mutant flies have reduced levels of H3K27me3. In mammals, there are three PCL orthologous: PCL1, PCL2/MTF2, and Pcl3/PHF19[80]. Consistently with the fly phenotype, PHF19 facilitates PRC2 recruitment to chromatin and

therefore H3K27me3 deposition[103,104]. Surprisingly, although MTF2-depleted ES cells contain increased global levels of H3K27me3, PRC2 recruitment and H3K27me3 levels are reduced at some PRC2 target genes[105]. PCL proteins contain a Tudor domain, which recognizes H3K36me3[103,104].

Mechanistically, it has been demonstrated that, through its binding to H3K36me3, PHF19 recruits H3K36me3 demethylases to active genes, thus favoring PRC2 recruitment and resulting in gene repression[103,104]. More recently, the mammalian-specific protein C17ORF96 (Chromosome 17 Open Reading Frame 96) has been also identified as a new PRC2-associated protein in a complex also containing JARID2 and PCL2. Although the C17ORF96 function in ES cells has not yet been interrogated genome-wide, C17ORF96 might be required for PRC2 stability and enzymatic activity[106].

In sum, six polypeptides—RBBP4/7, AEBP2, JARID2, PHF19, MTF2, and C17ORF96—have been recently identified as new co-factors that sub-stoichiometrically associate with different PRC2 complexes in ES cells. A comprehensive analysis of JARID2, PHF19, MTF2, and C17ORF96 target genes and gene expression profile in ES cells depleted by these co-factors will shed light on the functional regulation of these PRC2 sub-complexes.

3.4 Polycomb repressive complexes in ES cell self-renewal

The role of PRC1 in ES cell self-renewal and differentiation is challenging to study due to the numerous homologous subunits

that may be able to compensate for each other. In agreement with this, knockout mice for the single subunits that have been tested (i.e., for *Ring1A*, *Mel18*, *Bmi1*, *Cbx2*, or *Phc1*)[83] are viable (with the exception of *Ring1B*) yet have late developmental defects. *Ring1B* mutant mice display the most severe phenotype, as mice lacking *Ring1B* are embryonic lethal due to gastrulation defects. *Cbx7* mutant mice are viable, show increased body length, and develop lung and liver tumors[107].

In cultured ES cells, *Ring1A*- or *Ring1B*-null cells proliferate and self-renew normally[108]. While single *Ring1A* depletion does not have a major impact on gene regulation, *Ring1B* is required for gene repression maintenance of lineage-specific genes[109,110]. In contrast to PRC2-null ES cells, complete impairment of all of the PRC1 complexes (Ring1A/B double knock-out ES cells) strongly affects self-renewal[111], suggesting that PRC2 and PRC1 complexes are not completely functionally redundant. Depleting CBX7, RYBP, or FBXL10 in ES cells does not compromise self-renewal, but RYBP depletion decreases their proliferation[93,112,113].

Mutant mice for all core PRC2 components display embryonic lethality at different developmental stages: *Ezh2*-null mice are embryonic lethal at day E8.5, *Eed*-null mice, at E9.5, and *Suz12*-null mice, between days E8.5 and E10.5[83]. *Eed*-, *Suz12*-, or *Ezh2*-null ES cells self-renew and proliferate normally, yet their differentiation capacity is strongly compromised[114,115]. Surprisingly, the precise contribution of each of the PRC2 subunits during ES cell lineage specification is still poorly understood. It has been shown that *Eed*-null ES cells lose their cellular identity and are not able to differentiate towards any specific cell lineage, yet

another report indicates that *Eed*$^{-/-}$ ES cells are capable of developing teratomas that contain all three cell lineages but that also have an over-representation of ectoderm and mesoderm tissues[116]. Suz12-mutant ES cells display defects during neuroectodermal and EB differentiation, and ES cells lacking *Ezh2* fail to differentiate towards the mesoendoderm lineage[117].

Regarding the PRC2-associated proteins in ESCs, JARID2 is also essential during embryogenesis[118]. While JARID2 and C17ORF96 are also dispensable for ES cell self-renewal, MTF2-depleted ES cells self-renew faster[105], and PHF19-depleted cells display a certain degree of spontaneous differentiation[104]. JARID2, MTF2, and PHF19 are essential regulators of ES cell differentiation, although they are rapidly downregulated during differentiation[106,119].

4. Chromobox proteins (CBX)

4.1 Conservation and diversification of Pc homologues

At least eight members of the CBX family of proteins exist in the mammalian genome. All share a particular domain in the N-terminal part of the protein, the chromodomain, which is a well-characterized methyl-lysine reader module. The CBX family is further divided into two groups: (I) CBX1, CBX3, and CBX5 are the homologs of heterochromatin protein 1 (HP1) in *Drosophila;* and (II) CBX2, CBX4, CBX6, CBX7, and CBX8, which are homologs of the Pc gene. The CBX family of genes represents a good example of gene expansion and divergence. While there is only a Pc gene representative in insects, vertebrates have up to five Pc homologues, most probably due to gene duplication events during evolution. The significance of having more Pc genes is of great interest and remains under continuous investigation.

Apart from the chromodomain, CBX2, -4, and -6–8 share a C-terminal domain, the Polycomb repressor box (PcR box). The PcR box concedes the repressor potential to the CBX proteins by mediating binding to Ring1B, and thereby to the whole PRC1 complex (**Figure I.9**).

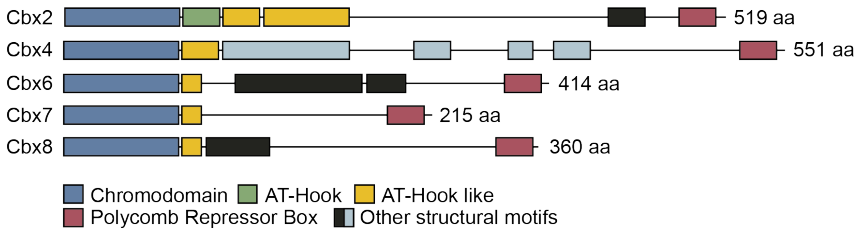


Figure I.9. Structure of the CBX proteins. All CBX proteins share the chromodomain and the Polycomb repressive (PcR) box, conserved across evolution. Other structural motifs are specific for each CBX protein.

The chromodomain is a three beta-stranded anti-parallel sheet flanked by an alpha helix–containing domain. Chromatin recruitment of the PRC1 complex is substantially mediated through the chromodomain of the CBX proteins, which is able to recognize the histone H3K27me3 modification, previously deposited by the PRC2 complex. However, it is believed that this is not the only mechanism necessary to recruit and stabilize the cPRC1 complex to its target genes; for instance, RNA could also play a role (discussed below). Unlike the fly Pc that strictly recognizes H3K27me3, its mammalian counterparts display different affinities for different histone modifications, as determined by Bernstein and colleagues in two independent assays[120]. In the first assay, recombinant chromodomains (CD) of all CBX proteins were incubated with fluorescently-labeled, histone-modified peptides, and binding affinities were measured through a specific device. The distinct CBX-CDs had different binding affinities for the histone peptides tested (H3 and H4 unmodified, H3K9me1/2/3, H3K27me1/2/3 and H4K20me1/2/3):

CBX2 and CBX7 were able to bind both H3K9me3 and H3K27me3 modifications with a similar dissociation constant to those of *Drosophila* HP1 for H3K9me3 and Pc for H3K27me3; CBX4 showed a significant preference towards the H3K9me3 peptide; and CBX6 and CBX8 were not able to bind to any. These results were confirmed by a classical peptide pulldown assay in which recombinant GST-CDs were incubated with biotinylated histone-tail peptides; again, CBX6-CD did not recognize any of the histone modifications tested. In this same publication, a more *in vivo* analysis studied whether (full-length) CBX-EGFP co-localized with the inactive X chromosome (Xi) in female ES cells, which is known to be highly enriched in H3K27me3. Interestingly, all CBX proteins except for CBX4 co-localized with Xi. The authors hypothesized that CBX6, which did not bind H3K27me3 *in vit*ro, may be recognizing another histone modification present in the Xi, or that alternatively, recruitment is mediated through RNA, as Xi is entirely coated by the RNA Xist, involved in chromosome X inactivation. In fact, all CBX-CDs except for that of CBX2 could bind RNA in a non–sequence-specific manner. This ability depended on the secondary structure of the chromatin, as heat-denatured recombinant CDs no longer bound to RNA in the RNA gel shift assay.

Within the chromodomain, the aromatic cage pocket that is responsible for recognizing the methylated peptide is essentially indistinguishable between CBX paralogues. However, the amino acid sequences surrounding the aromatic cage have subtle variations. Further, the CBX proteins have significant variations outside the chromodomain and the PcR box domain, such as in their total lengths and due to the presence of other

uncharacterized domains, which may explain their differential functions[121]. For instance, it is known that the tryptophans (at positions W33 and W36) and tyrosine (Y40) are involved in the recognition of and stabilization to H3K27me3, respectively. While W33 and W36 are conserved in all Pc homologues, but Y40 (a polar, uncharged amino acid) is substituted by a histidine (a polar, charged amino acid) in CBX6 and CBX8. Adjacent to the aromatic cage of the chromodomain, there is an hydrophobic pocket with three amino acids involved: valine (V11), phenylalanine (F13), and alanine (A13). In CBX6, V11 is substituted by an isoleucine, whereas in CBX2, A13 is replace by an aspartic acid. These subtle changes in the amino acid sequence could explain different binding preferences of CBX paralogues. Milosevich and coworkers gained insights in the CBX structure by attempting to develop small molecule inhibitors for the CBX paralogues: they realized that the amino acid residue at position 14 in the chromodomain was valine for CBX4/7 and alanine for CBX2/6/8[122]. This residue is important for determining the shape of the small hydrophobic pocket adjacent to the aromatic cage. They developed a peptidomimetic agent that specifically inhibits CBX6 but not the other CBX paralogues, providing evidence that the CBX6 structure is exceptionally different from the other Pc homologues. However, the particular CBX6 selectivity for this agent is not simply explained by the Val/Ala difference at position 14, as CBX2 and -8 are not affected by this inhibitor. According to their results, the alanine present in the ARKS motive of the trimethylated histone peptides (H3K9me3 and H3K27me3) is not able to fit into the hydrophobic pocket of CBX6, suggesting that CBX6 might be reading another, yet-unidentified histone modification[122].

The second main domain of CBX proteins, the PcR box domain, is C-terminal and allows the interaction with Ring1B. Deletion of this domain in *Drosophila* abrogates Pc function and results in homeotic transformation similar to Pc knockout flies[76]. The PcR box shows a high degree of conservation in all homologs and is unique to Pc proteins. However, some of its amino acids are not conserved but rather specific to certain CBX proteins, indicating that these residues may play an essential role in differential interactions of Pc and Ring homologs.

4.2 Dynamic interplay between CBX proteins

The presence of distinct CBX subunits suggested that different PRC1 complexes might exist, and that this structural heterogeneity could indicate diverse functional roles. However, in 2012, our laboratory addressed this question in the context of pluripotent ES cells and differentiation[112]. This study showed that in undifferentiated mouse ES cells, CBX7 and CBX6 are the Pc homologues that are most expressed, both at the RNA and protein level, as compared to the other CBX proteins. Nonetheless, co-immunoprecipitation experiments in ES cells revealed that the cPRC1 mainly contains CBX7 together with PHC1, MEL18/PCGF2, and RING1B, but not CBX6. Importantly, CBX7 is necessary for PRC1-mediated transcriptional repression, as CBX7 depletion strongly reduces chromatin binding of RING1B and MEL18. In turn, CBX7 recruitment is fully dependent on H3K27me3, suggesting that the cPRC1 recruitment follows the hierarchical model in ES cells.

CBX7 is dispensable for ES cell proliferation, as pluripotent markers were not affected upon CBX7 depletion, and colonies

were positive for alkaline phosphatase staining. However, CBX7 is required to maintain gene repression of a large amount of PRC1 target genes. The aberrant upregulation of lineage-specific genes resulted in defects during differentiation of the ES cells into embryoid bodies (EBs), and more specifically, towards the neuroectoderm lineage.

We additionally identified an auto-regulatory mechanism in which PRC1 and PRC2 repress other Polycomb genes, both in the pluripotency and differentiation states. In ES cells, PRC1-CBX7 represses the CBX2/4/8 subunits. During differentiation, CBX7 is strongly downregulated, and CBX2/4/8 expression is upregulated. These CBX proteins are assembled into new cPRC1 complexes during EB differentiation that are involved in CBX7 gene repression, among others. Interestingly, depletion of CBX2 and CBX4 resulted in lineage-specific phenotypes, indicating not only that CBX proteins have non-overlapping functions in pluripotent and differentiating ES cells, but also that the expression of *Cbx2* and *Cbx4* is important to drive the newly assembled cPRC1 complexes to specific and *de novo* loci during differentiation.

Other studies in hematopoietic stem cells (HSC) have shown that CBX7 is the primary CBX protein incorporated in the PRC1 and is required for the proper self-renewal of HSC, as shown *in vitro* by colony formation assay and cobblestone-area-forming cell assay, and *in vivo* by bone marrow transplantation[121]. Ectopic expression of *Cbx2*, *Cbx4*, or *Cbx8* induced differentiation, and these cells failed to contribute to bone marrow reconstitution. The authors showed that CBX proteins could coexist and function in the same cell type. In HSC, *Cbx7* and *Cbx8* are expressed at the same level and share most target genes. However, CBX8 unique target

genes resulted to be highly expressed in ES cells, which become fully repressed upon differentiation, suggesting that this pre-deposited CBX8-PRC1 complex is marking future target genes for their repression.

Finally, Pemberton and colleagues showed in human fibroblasts that multiple PRC1 complexes with different CBX proteins co-localize genome-wide, suggesting that distinct CBX-containing PRC1 complexes overlap in the same promoters forming oligomers to ensure transcriptional repression[123].

4.3 Cbx6: the known and unknown

CBX6 function has been poorly studied to date, so that the number of scientific publications and the knowledge gleaned about the function of CBX6 is very low. In this section, I summarize the information that is known about CBX6 so far, in a biochemical context as well as in stem cells and cancer.

It is generally believed that different CBX proteins may have different functions; still, the properties of each CBX protein have not been systematically analyzed under the same experimental conditions. Vandamme and colleagues used a tandem–mass spectrometry (MS) approach to analyze the interactors of each CBX protein in the same cell type (HeLa cells). To this end, cell lines overexpressing similar levels of TAP-tagged CBX proteins were generated and then subjected to affinity purification followed by TAP-MS/MS analysis. Although CBX6 recovery was efficient, only few PRC1 members were identified in the MS results—in fact, they could barely see interactions with PHC2 and RING1A/B and

did not detect any of the PCGF family members. In contrast, binding assays based on GST-pulldown experiments showed that CBX6-GST could interact with PCGF-FLAG family proteins *in vitro*. Interestingly, size-exclusion chromatography showed that TAP-tagged CBX6 is involved in high–molecular weight protein complexes, suggesting that CBX6 indeed forms part of a functional complex i*n vivo*, although it seems to preferentially bind complexes different from the canonical PRC1. The cellular localization of CBX-GFP fused proteins was found to be dispersed within the nucleus in U2-OS cells. Similarly, CBX6 sub-nuclear localization was shown in a bimolecular fluorescence complementation (BiFC) assay, in which live cell imaging allows visualization of the proteins in their native environments[124].

The group of Gang Li is one of the few that have attempted to understand the function of CBX6 in the context of cancer[125]. Studying the epigenetic alterations involved in the development of glioblastoma multiforme (GBM), a primary brain tumor of the astrocytic lineage, they found that CBX6 was downregulated in GBM cell lines as well as in clinical samples from patients. Overexpression of CBX6 in GBM cell lines leads to cell growth arrest, inhibiting glioblastoma cell proliferation, suggesting that misregulation of CBX6 expression potentially promotes GBM progression.

It is generally assumed that the different variants of the PcGs form distinct complexes to carry out diverse biological functions. However, unexpectedly, various Pc homologues were recently shown to have overlapping ChIP-seq profiles in human fibroblasts[123]. In this scenario, CBX6, CBX7, and CBX8 would

share the same target genes and would collaborate cooperatively to suppress these genes. Interestingly, the authors showed by sequential ChIP that the proteins are associated within the same DNA .

As mentioned above, *in vitro* studies showed that CBX6-CD could not bind significantly to histone H3K27me3 or to any other modifications tested[120]. Surprisingly, the EGFP fusion protein for CBX6 (full-length) could associate with the Xi in 3-day differentiated female ES cells, a point at which H3K27me3 is highly enriched. In these studies, they also reported that CBX6-CD binds RNA in a non–sequence-specific manner, and that a CBX6-CD intact secondary structure was required for this binding[120].

4.4 Cbx6 versus Cbx7 in ES cells

Cbx6 and *Cbx7*, the two most highly expressed Pc homologous in ES cells, share part of their structure due to the presence of a conserved chromodomain and PcR box. However, several pieces of evidence suggesting that these two proteins have independent—but possibly complementary—functions (**Table I.1**).

	Cbx6	Cbx7
Evolutionary emergence	Vertebrate lineage	Mammalian lineage
Conservation of the chromodomain	Very conserved	Very conserved
Binding to H3K27me3?	No	Yes
Conservation of the PcR box	Very conserved	Very conserved
Affinity for RNA?	Yes. Non-–sequence-specific binding	Yes. Sequence specific binding
Expressed in ES cells?	Yes	Yes
Repressed upon ES cell differentiation	Partial downregulation	Rapid downregulation Complete repression
cPRC1/ncPRC1 mMass Spectometryspectrometry evidences of interaction in ES cells?	No	Yes

Table I.1. Comparison between CBX6 and CBX7.
Some of the similar and dissimilar features between CBX6 and CBX7 are summarized. Text in red refers to CBX6-specific features, and blue, to CBX7-specific features.

It is intriguing to note the different appearance of both proteins in the evolutionary context, as nicely depicted in a phylogenetic representation of PcG homologues in a recent review by Whitcomb and colleagues. CBX6 appears at the level of vertebrates, whereas CBX7 appears—probably due to a gene duplication event—in the mammalian lineage[79].

Regarding their biochemical properties, both proteins are able to bind RNA. CBX7 has been demonstrated to specifically bind to the antisense non-coding RNA ANRIL, and this interaction is required for epigenetic transcriptional repression of the *Ink4/Arf* locus[126]. CBX6 can bind RNA in a non–sequence-specific manner[120]; further studies are needed to understand the fine-tuning of this interaction.

Proteins containing chromodomains possess the inherent characteristic to bind to methylated lysines in the tail region of histone H3. It is widely accepted that CBX7 can bind to both H3K9me3 and H3K27me3 histone modifications, *in vitro* as well as *in vivo*[120,127]. However, very little is known regarding the preferences for CBX6 for methylated histone. *In vitro* studies demonstrated that the CBX6 chromodomain does not display affinity for either H3K27me3 or H3K9me3. Although further characterization to uncover CBX6 binding properties is needed, initial results from the laboratory of Dr. Fraser Hof indicate that CBX6 has intrinsic different characteristics to the other CBX proteins, as they could develop a small molecule agent specific for CBX6 structure[122]

Although both *Cbx6* and *Cbx7* are expressed in ES cells, their expression patterns differs: *Cbx7* expression is rapidly downregulated when cells differentiate, so that it is no longer expressed in differentiated cells; in contrast, *Cbx6* expression is gradually downregulated, but *Cbx6* is still expressed in differentiated cells. Thus, CBX6 is found in adult stem cells (such as intestinal stem cells)[128] as well as in terminally-differentiated cells (such as keratinocytes)[129].

Importantly for this project, CBX6 was not previously found to be present in any immunoprecipitation experiments followed by mass spectrometry analyses, performed against core components of the cPRC1 and ncPRC1 subunits in ES cells[85].

OBJECTIVES

Research in our group focuses on the investigation of the transcriptional and epigenetic mechanisms involved in ES cell biology. Particularly, we are interested in understanding the role of PcGs in the context of ES cells as well as during ES cell differentiation. Our laboratory has recently unveiled a dynamic interplay between the different subunits of the CBX family of proteins, but very little is known about CBX6.

The general objective of this thesis was thus to describe the role of CBX6 in the context of ES cells and in the transition towards differentiated cells. We thus focused on the following main aims:

I. Investigate the outcome of CBX6 depletion in ES cell pluripotency and self-renewal
II. Identify CBX6 target genes genome-wide
III. Characterize CBX6-associated proteins
IV. Identify binding preferences of CBX6 for histone modifications

RESULTS

1. CBX6-depleted ES cells spontaneously differentiate

Previous studies in our laboratory indicated that not all the Cbx proteins are expressed in ES cells, with Cbx6 and Cbx7 the most highly expressed Cbx proteins in pluripotent ES cells[112] (**Figure R.1**).

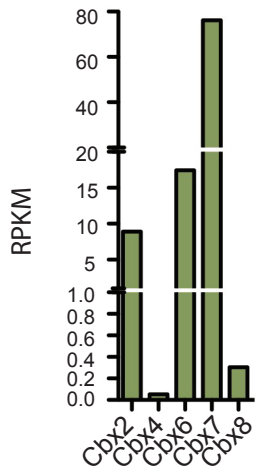


Figure R.1. Gene expression profile of *Cbx* paralogs in ES cells. RPKM levels of *Cbx* proteins in ES cells.

To explore the functional role of CBX6 in ES cells, we stably depleted CBX6 in the mouse ES cell line, E14Tg2A. We used two independent shRNA constructs that efficiently reduced CBX6 mRNA and protein levels (**Figure R.2**). In contrast to CBX7-depleted ES cells, CBX6-depleted ES cells displayed a cellular morphology that resembled spontaneous differentiation (**Figure R.3**). Although CBX6 knockdown ES cells still formed colonies, these were flatter and not

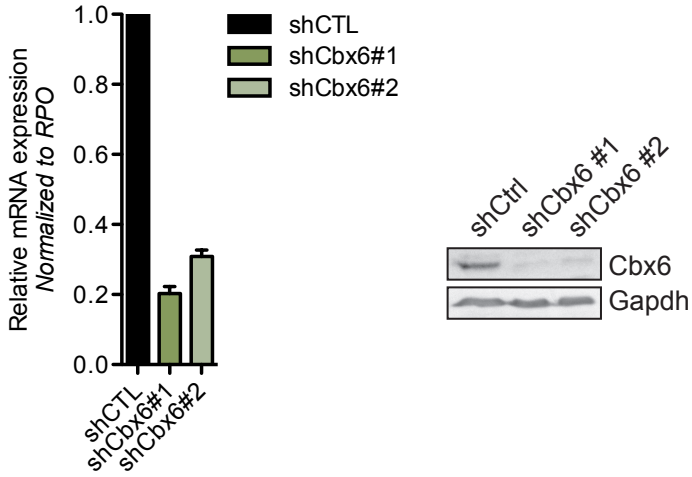


Figure R.2. CBX6 depletion in ES cells. Left, RT-qPCR analysis of control and CBC6-depleted ES cells. Results are shown relative to shCTL, and normalized to the housekeeping gene *Rpo*. Error bars represent standard deviation (SD) of three independent experiments. Right, Western blot analysis of CBX6 in control (shCTL) and CBX6-depleted (shCbx6#1 and #2) ES cells. GAPDH was used as a protein-loading control.

round-shaped. Moreover cells with fibroblastic-like morphology surrounding the colonies appeared only in CBX6-depleted ES cells, suggesting that CBX6 was required for ES cell pluripotency. Importantly, as we previously reported, ES cells depleted for CBX7 did not spontaneously differentiate (**Figure R.3**).

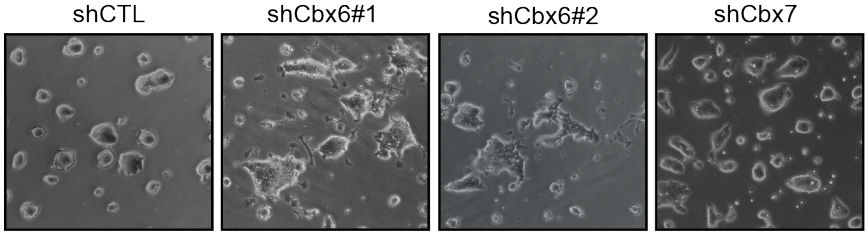


Figure R.3. ES cell morphology of CBX6- and CBX7-depleted ES cells. Phase contrast image of shCTL, shCbx6, and shCbx7 cell lines.

2. CBX6 depletion impairs ES cell self-renewal and is required for pluripotency

To assess whether CBX6 was required for ES cell self-renewal, we performed proliferation assays. Although CBX6-depleted ES cells could be sustained in culture for numerous passages, they were less proliferative than the control cells (**Figure R.4**). These results indicated that self-renewal is compromised in the absence of CBX6.

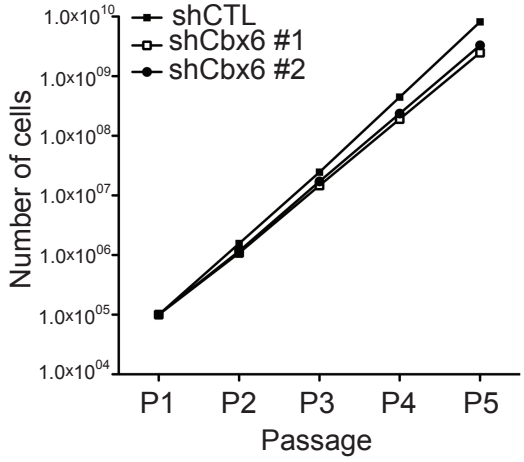


Figure R.4. Self-renewal is affected in CBX6-depleted ES cells. Accumulative proliferation curve of shCTL and shCbx6.

Importantly, knockdown levels were maintained along the passages (**Figure R.5**).

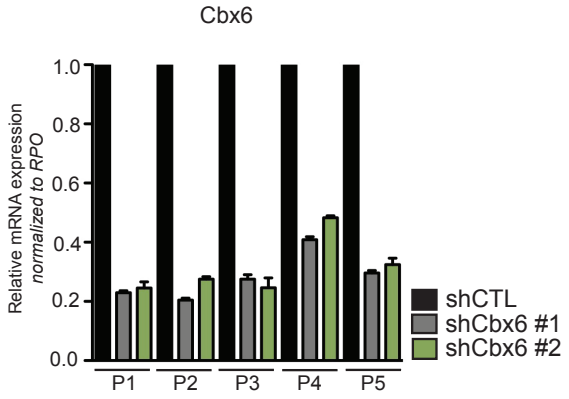


Figure R.5. RT-qPCR analysis of control and CBX6-depleted ES cells. Results are shown relative to shCTL and normalized to Rpo. Error bars represent standard deviation (SD) of two independent experiments (P, passage).

We next studied the pluripotency capacity of the CBX6-depleted cells, using several approaches. First, we evaluated the expression of the factors that sustain ES cell pluripotency. Most of the genes (such as *Rex1*, *Klf4*, *Nanog*, and *Esrrb*) were strongly downregulated upon CBX6 depletion (**Figure R.6**).

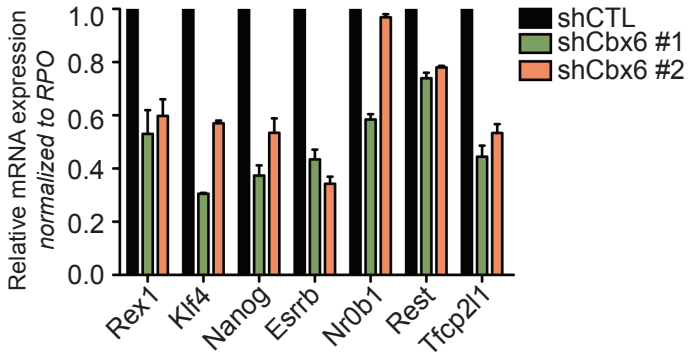


Figure R.6. Pluripotency factors are aberrantly downregulated upon CBX6 depletion. RT-qPCR analysis of control and CBX6-depleted ES cells. Results are shown relative to shCTL, and normalized by RPO. Error bars represent SD of three independent experiments.

We then tested for alkaline phosphatase (AP) staining, which assesses pluripotency, as AP positively stains colonies that retain pluripotent potential. Interestingly, CBX6-depleted cultures contained colonies poorly stained by AP (**Figure R.7**). Quantification of the staining indicated that 52% of the colonies were differentiating, 26% were not clearly stained (which we refer to as "mix"), and 22% were positively stained and thus in the pluripotent state (**Figure R.7)**. Overall, these results clearly indicated that CBX6 contributed to the maintenance of the pluripotency network in ES cells.

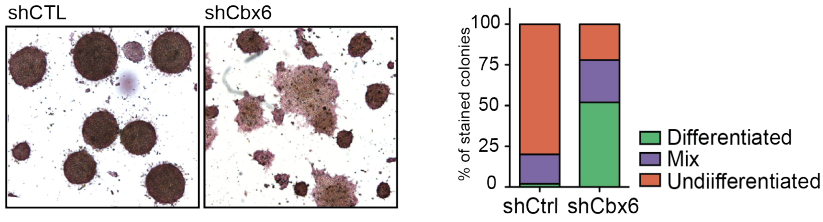


Figure R.7. CBX6 is required for ES cell pluripotency. Left, phase contrast image of AP staining performed on shCtrl and shCbx6. Right, quantification of the AP staining assays, representing the mean of three independent experiments in which around 40 random colonies were counted.

We asked whether CBX6 was also required for ES cell pluripotency when cultured in naïve culture conditions (e.g., 2i medium), which is less permissive for ES cells to spontaneous differentiate. To this end, we stably knocked down CBX6 in 2i-cultured ES cells. Surprisingly, CBX6 depletion did not resulted in a phenotype that was as obvious as that observed for cells cultured with serum and LIF, with no apparent differences

between control and CBX6-depleted ES cells after culturing in 2i media (**Figure R.8**).

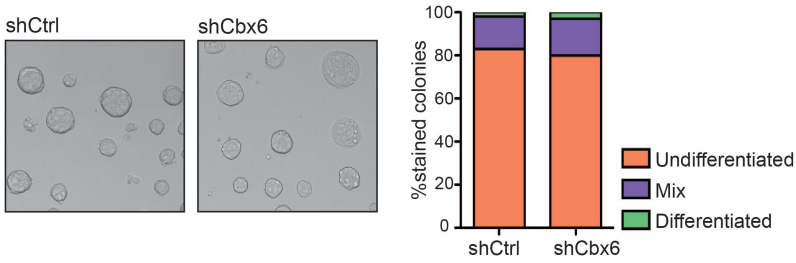


Figure R.8. CBX6-depleted cells cultured in 2i medium do not differentiate. Left, phase contrast image of shCtrl and shCbx6 cells cultured in 2i medium. Right, quantification of AP staining. assay.

As pluripotency-associated genes had been downregulated, we checked the expression of differentiation-related genes, and in particular for specific marker genes for any of the three germ layers (endoderm, mesoderm, and ectoderm). Endoderm genes, although expressed at very low levels in ES cells, were globally less expressed in CBX6-depleted genes. In contrast, the expression of marker genes for mesoderm and ectoderm germ layers did not follow a clear pattern. In both cases, some genes were upregulated (such as *Brachyury* in mesoderm and *Pax6* in endoderm), some gene expression levels remained unchanged, and some were downregulated, upon CBX6 depletion (**Figure R.9**).

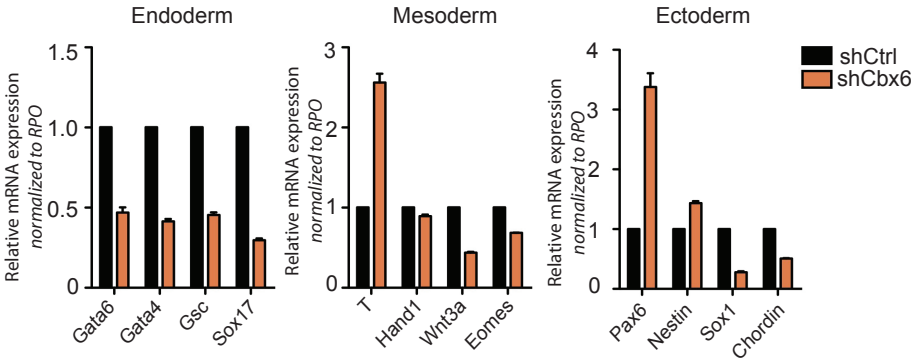


Figure R.9. CBX6 depletion induces defects in differentiation markers. qRT-PCR analysis of control and CBX6-depleted ES cells. Results are shown relative to shCTL and normalized to RPO. Error bars represent SD of three independent experiments.

3. Global transcriptome is affected by CBX6 depletion

To better understand the molecular mechanisms underlying the effects of CBX6 in ES cell self-renewal and differentiation, we next performed RNA sequencing (RNA-seq) in CBX6-depleted and control ES cells. Using a cut-off of two-fold, we found 703 and 783 genes upregulated and downregulated, respectively, in CBX6-depleted ES cells (**Figure 10**).

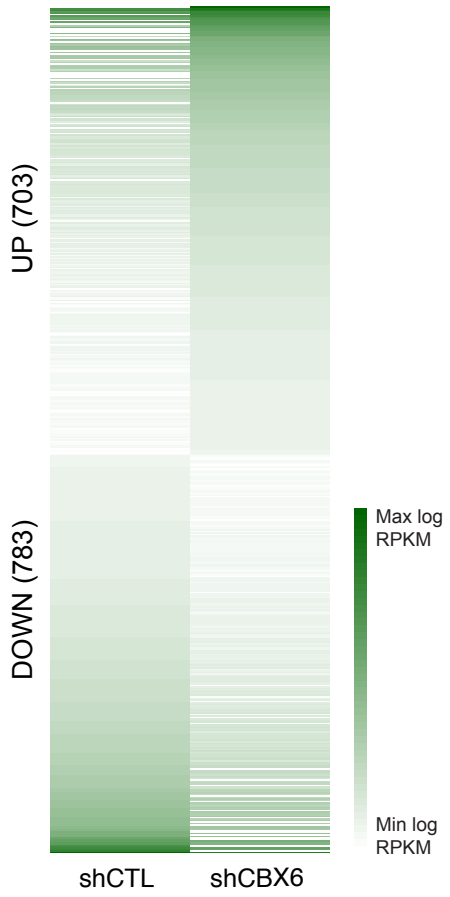


Figure R.10. RNAseq heat-map of up- and downregulated genes in CBX6-depleted cells as compared to control cells. Only genes up- or downregulated by at least 2-fold as compared to control cells are shown.

Using the Enrichr software, we determined whether any biological processes and/or transcriptional networks were significantly deregulated in CBX6-depleted cells. We focused our attention on the top one-hundred deregulated genes.

Analyses of upregulated genes

Interestingly, gene ontology (GO) analyses indicated that upregulated genes scored as genes involved in cell-cell adhesion, regulation of Fibroblast growth factor receptor (FGFr) signaling pathway, neuron differentiation, and cell fate commitment (**Figure R.11**).

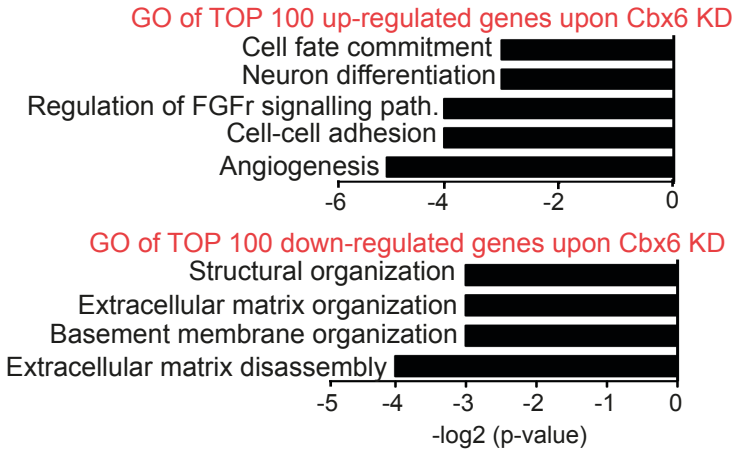


Figure R.11. GO analysis of genes deregulated following CBX6 knockdown; p values are plotted in –log.

Importantly, we observed a strong upregulation of the *Pax6*, *Otx2*, and *Wnt8a* genes, which are involved in neuroectodermal differentiation. This indicates that CBX6-depleted ES cells spontaneously differentiate towards the ectoderm lineage. Moreover, we also found genes important for membrane receptor

pathways, such as *Fgfr2*, which is involved in the FGF signaling pathway, and the receptors *Fzd2* and *Lef1*, which have been extensively associated with the Wnt pathway. Finally, three members from the Claudin family of genes (*Cldn3/6/7*), which produce essential proteins involved in the formation of tight junctions in the epithelial tissues, were aberrantly upregulated upon CBX6 depletion. These data strengthen our observation that CBX6 depletion induced differentiation.

Analyses of downregulated genes

GO analyses of downregulated genes upon CBX6 depletion scored as genes related to homeostasis of the extracellular matrix (basement membrane organization and extracellular matrix disassembly) (**Figure R.11**). These categories were enriched in genes involved in collagen polymerization, such as *Col4a1*, *Col4a2*, and *Col13A1*, and extracellular matrix glycoproteins, such as *Lama1*.

Comparison of the gene expression profiles in CBX6- and CBX7-depleted ES cells

CBX7 and CBX6 are most highly expressed Cbx paralogs in ES cells. Interestingly, CBX7 depleted cells do not spontaneously differentiate. Therefore, we next sought to investigate the gene expression profile of ES cells depleted of CBX6 and CBX7. To this end, we compared our RNA-seq experiments from CBX6-depleted cells with our published gene expression profile from CBX7-depleted ES cells[112]. We observed that more genes were affected in the CBX7-depleted ES cells than in the CBX6-depleted ES cells, with 1,096 up- and 2,565 downregulated as compared to 703 up- and 783 downregulated, respectively.

Therefore, we concluded that CBX7 had a greater impact on gene regulation than CBX6.

We next overlapped up- and downregulated genes from CBX6- and CBX7-depleted ES cells. We found that 241 genes were commonly upregulated, and 432 genes were commonly downregulated, in CBX6- and CBX7-depleted ES cells (**Figure R.12**). In contrast, 462 and 855 genes were exclusively upregulated in CBX6- and CBX7-depleted ES cells, respectively, while 351 and 2,133 genes were exclusively downregulated in CBX6- and CBX7-depleted ES cells, respectively (**Figure R.12**).

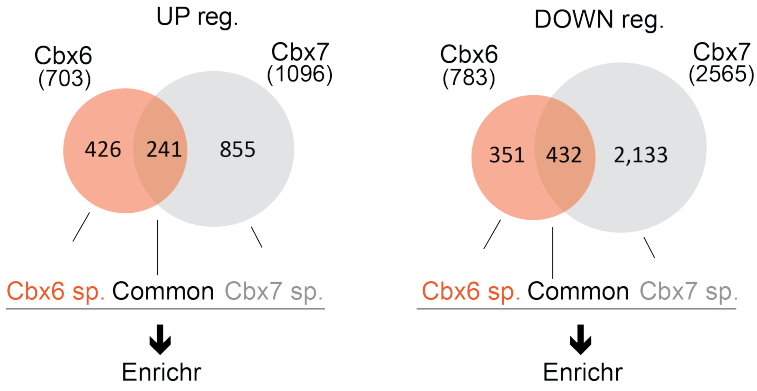


Figure R.12. CBX6-depletion transcriptional changes partially overlap with CBX7-depletion derived changes. Venn diagrams showing the degree of overlap between the up- and downregulated genes in CBX6- and CBX7-depleted ES cells. Sp, specific.

We further analysed each gene list (with the categories Cbx6-specific, common, or CBX7-specific for up- or downregulated genes) using the Enrichr software. We next aimed to identify a subset of genes differentially affected between CBX6- and CBX7-

depleted ES cells that could explain the dissimilar phenotypic characteristics between both cell lines. Unfortunately, after performing many analyses and checking all genes involved in the different lists, we did not observe any trend that allowed us to understand the different phenotypes.

4. CBX6 is indispensable for ES cell differentiation

We performed embryo body (EB) differentiation assays to study the effects of CBX6 depletion on cell differentiation. EBs recapitulate embryo development *in vitro*. During EB differentiation, ES cells spontaneously differentiate into cell types corresponding to the three germ layers in a temporal manner. To assess the capability of CBX6-depleted ES cells to differentiate, we generated EBs from control and shCbx6 cells and followed them until day 9 of differentiation. Early differentiation genes are upregulated during the first 6 days, and late differentiation genes are sequentially upregulated, with a concomitant downregulation of pluripotency genes, from days 7 to 9. Strikingly, EBs derived from CBX6-depleted ES cells did not show significant morphological changes, suggesting that ES cell differentiation is not totally blocked in cells lacking CBX6 (**Figure R.13**). These results are in contrast with EBs derived from CBX7-depleted ES cells, which are smaller than wild-type EBs.

To better understand whether CBX6 is required for correct differentiation, we performed a genome-wide expression analysis (microarray) in CBX6-depleted and control cells. We extracted RNA from 0-, 6-, or 9-day-old EBs. Using a cut-off of 1.5-fold, we found that the expression of numerous genes was altered at each timepoint as compared to control EBs: 1962 genes were deregulated at day 0 (839 downregulated, 1123 upregulated), 561 genes at day 6 (258 downregulated, 303 upregulated), and 633 genes at day 9 (433 downregulated, 200 upregulated) (**Figure R.14**).

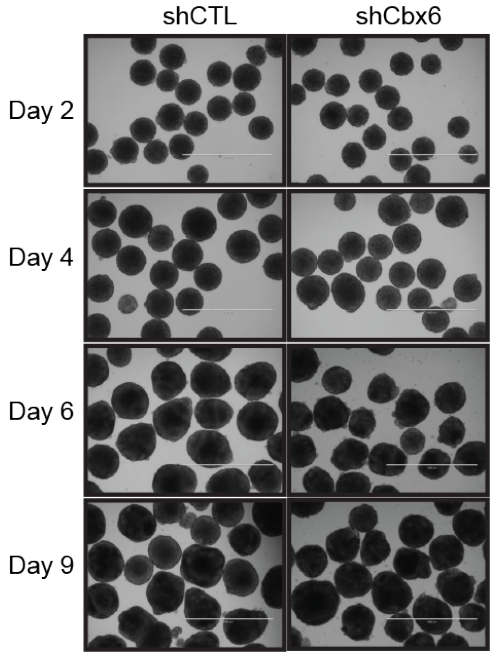


Figure R.13. EBs derived from CBX6-depleted ES cells had no significant changes in their morphology. Phase contrast images of EBs from shCtrl and shCbx6 ES cells, collected during differentiation at the indicated timepoints.

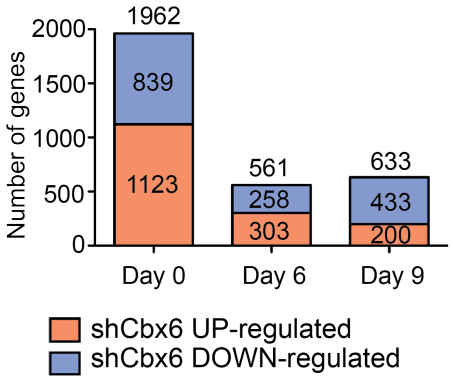


Figure R.14. Transcriptomic changes in EBs from shCbx6 ES cells. Number of genes deregulated in CBX6-depleted ES cells and EBs as compared to the respective control cells. Upregulated and downregulated genes are represented in orange and blue, respectively. Each experiment was performed in triplicate.

GO analyses of up- and downregulated genes did not score for any specific differentiation pathway. Nonetheless, analyses of selected pluripotency and differentiation genes from the three embryonic germ layers indicated that CBX6-depleted EBs were not differentiating properly (**Figure R.15**).

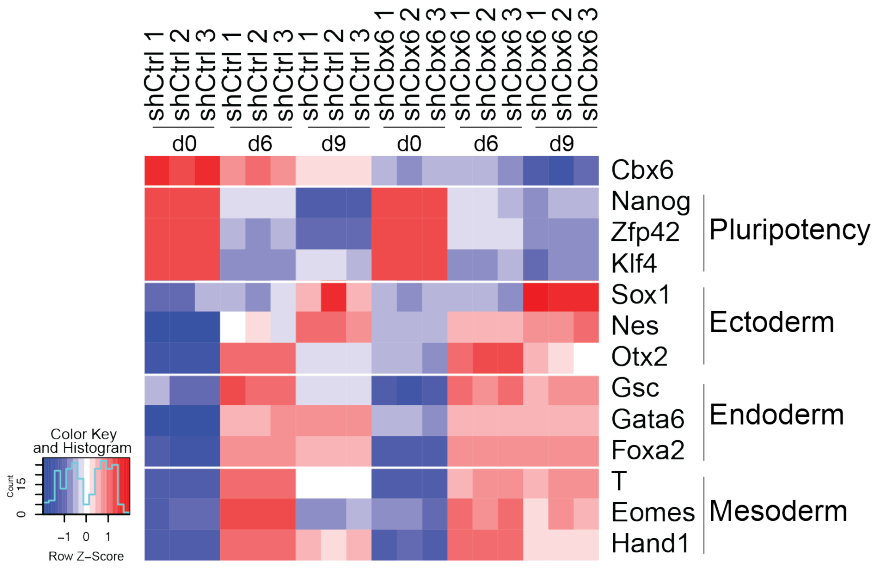


Figure R.15. CBX6-depleted EBs do not differentiate correctly. Relative expression levels of pluripotent, ecto-, endo- and mesodermal markers were evaluated in wild-type versus CBX6-depleted cells. Three biological replicas were analysed in each case. Data were extracted from the microarray experiments. Blue, down-regulated; red, up-regulated; white, not regulated.

Interestingly, in ES cells, *Cbx6* is expressed throughout differentiation, although its expression decreases gradually upon differentiation. In contrast, *Cbx7* is strongly downregulated once differentiation is triggered[112].

Pluripotency genes are expressed at lower levels in the CBX6-depleted ES cells at day 0 (although the intense color in the heatmap does not allow this to be appreciated). Interestingly, although the pluripotency genes were completely repressed in control cells, they were relatively higher expressed in CBX6-depleted cells by day 9. This perhaps indicates that, in the absence of CBX6, the cells could not fully silence these genes and therefore remained in a more “stem” state. Ectoderm, endoderm, and mesoderm genes were aberrantly deregulated in different directions along EB formation.

Together, these results indicated that CBX6 is required for the correctly-timed repression of differentiation-specific genes in the pluripotent state of ES cells, and that the loss of CBX6 produces an imbalance in the ES cell fate commitment toward an endodermal, ectodermal, or mesodermal lineage.

5. The chromodomain and the PcR box are essential to mediate CBX6-specific function

To exclude potential off-target shRNA effects, we generated stable cell lines transfected with a CBX6-expressing construct containing a point mutation that conferred resistance to the shRNA ($Cbx6^{WT}$), to perform a set of rescue experiments. CBX proteins contain two very well-conserved domains: the chromodomain in the N-terminal part of the protein, and the PcR box domain in the C-terminal end of the protein. We were also interested in studying whether these domains are essential to mediate CBX6 function. To this end, we generated Cbx6 constructs that were not only resistant to the shRNA but that also either lacked the 47 aa corresponding to the PcR box ($Cbx6^{\Delta PcR}$)[130] or with a mutation in the two tryptophan residues that are essential for an appropriate aromatic cage conformation ($Cbx6^{AA}$) to give a non-functional chromodomain[131] (**Figure R.16**).

Figure R.16. Schematic representation of the Cbx6 constructs used in the rescue experiment. Note that all the constructs contained a silent mutation (not depicted) that conferred resistance to the Cbx6 shRNA.

To prevent irreversible spontaneous differentiation, we first stably overexpressed Cbx6^{WT}, Cbx6^{AA}, or Cbx6$^{\Delta PcR}$ and subsequently depleted the endogenous *Cbx6* by lentiviral infection. Transfected ES cells with an empty plasmid (Empty) were used as a control. Importantly, the three cell lines overexpressed similar levels of the exogenous construct as analysed by Western blot (**Figure R.17**). Upon infection, only *Cbx6* endogenous mRNA was downregulated (as assessed by specific primers targeting the 3′-UTR) (**Figure R.17**).

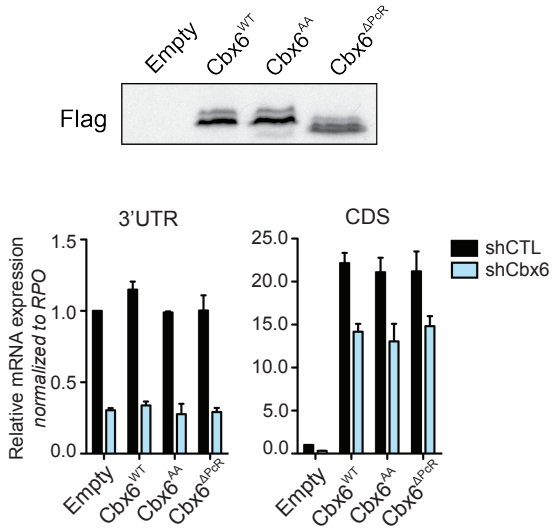


Figure R.17. Control of CBX6 ectopic and endogenous expression levels. Top, Western blot analysis of overexpressing *Cbx6*-Flag constructs. Bottom, RT-qPCR mRNA expression analysis of endogenous *Cbx6* (3′-UTR) or total *Cbx6* (CDS, coding region) in control (shCTL) and CBX6-depleted ES cells (shCbx6) overexpressing Cbx6^{WT}, Cbx6^{AA,} Cbx6$^{\Delta PcR}$ or nothing (Empty). Results are shown relative to shCTL, and normalized to RPO. Error bars represent SD of two independent experiments.

We then monitored whether overexpression of CBX6 would affect the protein levels of the PRC1 and PRC2 subunits or the levels of H2A/H2Aub levels. Western blot analysis indicates that no alterations were observed (**Figure R.18**).

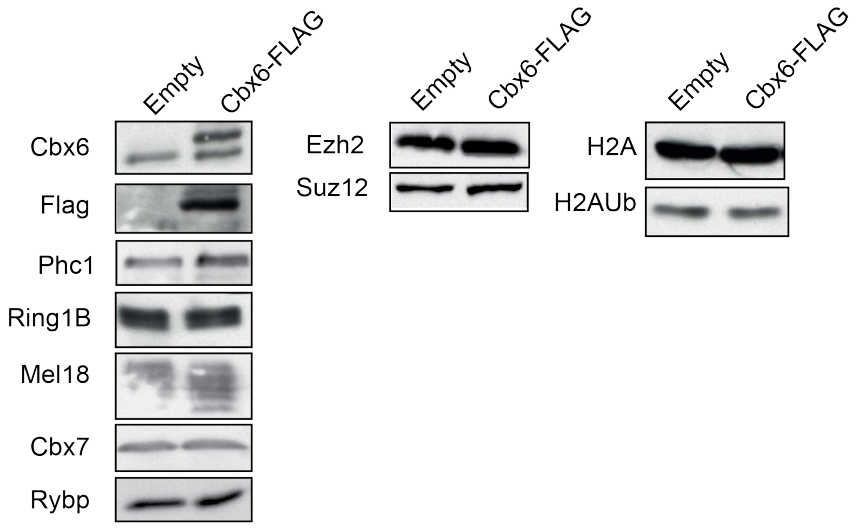


Figure R.18. PRC1/2 and H2AUb are not affected by CBX6 overexpression. Western blot analysis of PRC1 subunits (PHC1, RING1B, MEL18, CBX7, and RYBP), PRC2 (EZH2 and SUZ12) and H2A/H2AUb in control (Empty) and CBX6-FLAG overexpressing cell lines.

Interestingly, only $Cbx6^{WT}$ overexpression prevented the differentiation phenotype induced upon CBX6 depletion (**Figure R.19**). Colonies from the $Cbx6^{WT}$+shCbx6 condition were less flat and less differentiated than the Empty+shCbx6 cell line, indicating that the differentiation features observed upon CBX6 depletion exclusively rely on CBX6. Remarkably, neither the CBX6 construct lacking a functional chromodomain nor the one missing

the PcR box domain restored the morphological appearance derived from CBX6 depletion, as compared to control conditions. Moreover, expression of the mutated constructs in shCtrl-infected cells resulted in notable cell differentiation, suggesting that $Cbx6^{AA}$ and $Cbx6^{\Delta PcR}$ were outcompeting with the endogenous CBX6 protein thus causing a dominant negative effect.

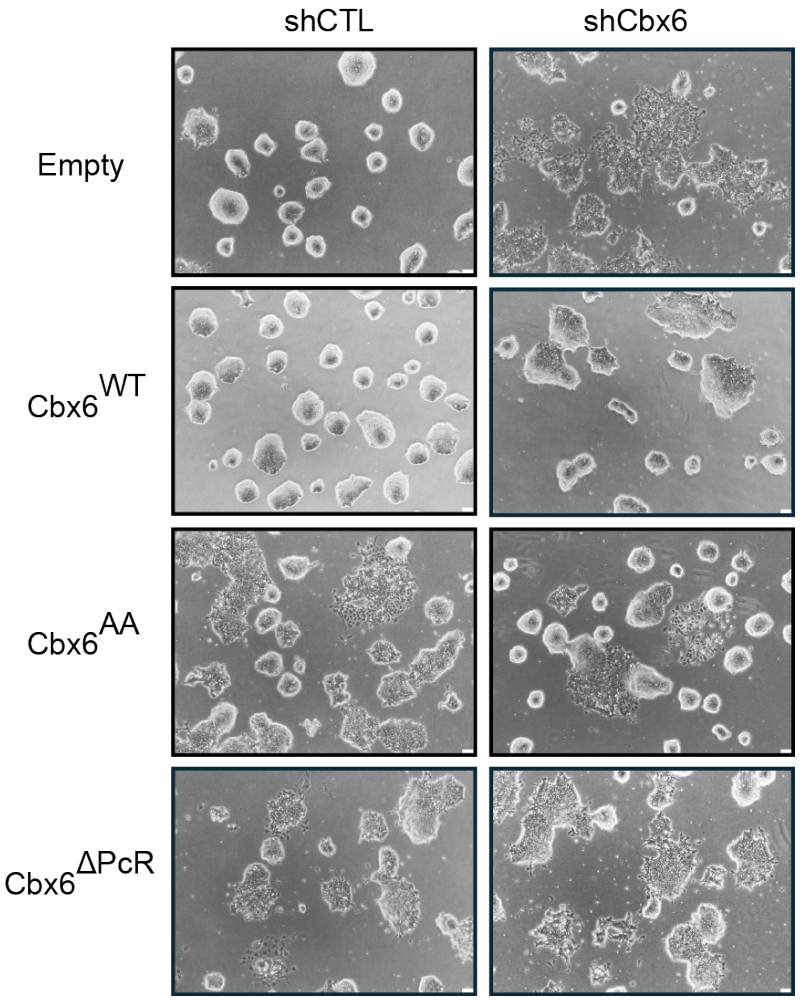


Figure R.19. The phenotype due to CBX6 depletion is rescued by overexpressing $Cbx6^{WT}$ but not Cbx6-mutated constructs. Phase contrast images of different cell-lines overexpressing an Empty, $Cbx6^{WT}$, $Cbx6^{AA}$ or $Cbx6^{\Delta PcR}$ construct, additionally infected with shCTL or shCbx6 lentiviral particles.

To understand whether the preventive effect also occurred at a molecular level, we checked expression levels of genes that are downregulated upon CBX6 depletion (*Klf4*, *Nanog*, *Rex1*, *Tfcp2l1* and *Tet2*). We observed that, in the Cbx6^{WT}+shCbx6 transfected cells, these genes almost reached normal levels as compared to the control cells (**Figure R.20**). Cbx6^{AA} and Cbx6$^{\Delta PcR}$ did not revert the phenotype at the molecular level, as expression of pluripotency-related genes remained downregulated. Additionally, we observed that the mutant Cbx6^{AA} and Cbx6$^{\Delta PcR}$ likewise had a dominant-negative effect at this level (**Figure R.20**).

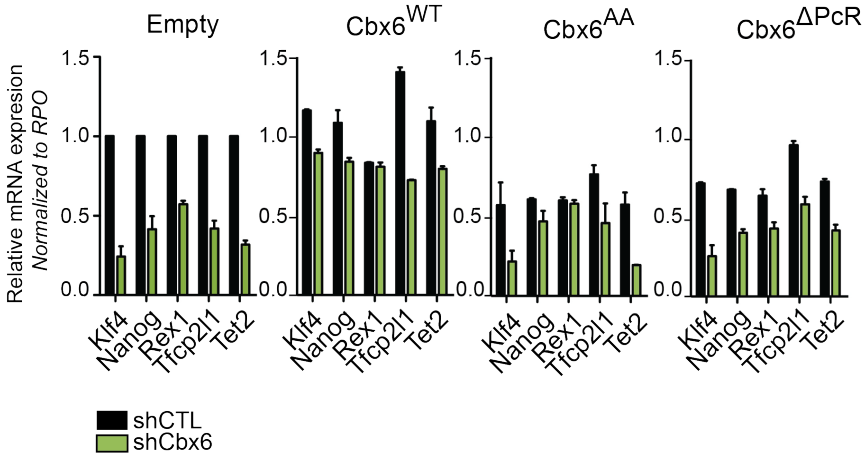


Figure R.20. Cbx6^{WT} rescues gene expression. qRT-PCR analysis. Results are shown relative to Empty-shCTL, and normalized to RPO. Error bars represent SD of two independent experiments.

6. CBX6 histone-binding preferences

CBX proteins have been previously shown to localize in the nucleus and to be associated with chromatin[120,124]. Cell fractionation experiments followed by Western blot analysis confirmed that CBX6 is bound to chromatin (**Figure R.21**).

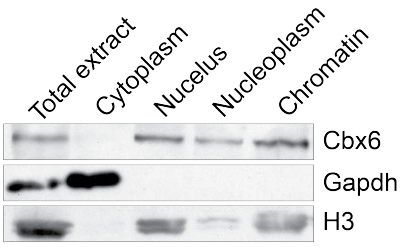


Figure R.21. CBX6 localizes in the nucleus and binds to chromatin. Western blot analysis of CBX6, GAPDH (cytoplasmic marker), and histone H3 (chromatin marker) in the indicated cellular fractions.

Previous studies have shown that the chromo-domain of CBX6 does not bind histone modifications *in vitro*[77]. Specifically, the CBX6 chromo-domain is not able to bind H3K27me3 peptides, while that of CBX7 can.

However, two CBX6 properties suggest that it can potentially recognize a histone modification: *i*) the presence of the chromo-domain, which is well-known to be able to recognize methylated residues; and *ii*) the enrichment of CBX6 in the chromatin fraction in ES cells. Importantly, we already determined that the presence of the chromo-domain is essential for CBX6 function. Thus, we aimed to identify the histone modifications recognized by CBX6.

For this purpose, we took advantage of a histone peptide array (EpiTitan™ Histone Peptide Array, from EpiCypher) that allowed us to monitor the binding of CBX6 to numerous histone-modified peptides in a single experiment. The histone peptide arrays contain around 265 biotinylated histone peptides (20 amino acids in length or more) immobilized on a streptavidin-coated glass slide. Peptides are spotted in two identical sub-arrays. Within each sub-array, each peptide is spotted twice in groups of 3, for a total of 12 spots per peptide on each array (**Figure R.22**, array schema).

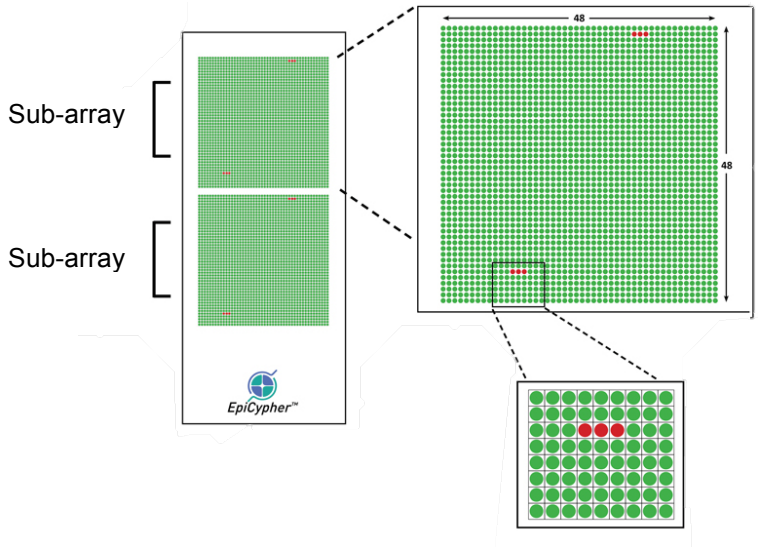


Figure R.22. EpiTitan™ histone peptide array design. Peptides are spotted as two identical sub-arrays, labeled A and B. Within each sub-array, each peptide is spotted twice in groups of 3 (red dots), for a total of 12 spots per peptide on each array.

We expressed GST, CBX7-GST, and CBX6-GST full-length recombinant proteins in bacteria. 2uM solutions of each protein

were hybridized onto the histone peptide arrays. After incubating the array with a primary antibody (αGST) and a secondary antibody (fluorescently labeled), we proceeded to the detection step by scanning the array with an Agilent scanner. The scanned images are shown in **Figure 23**.

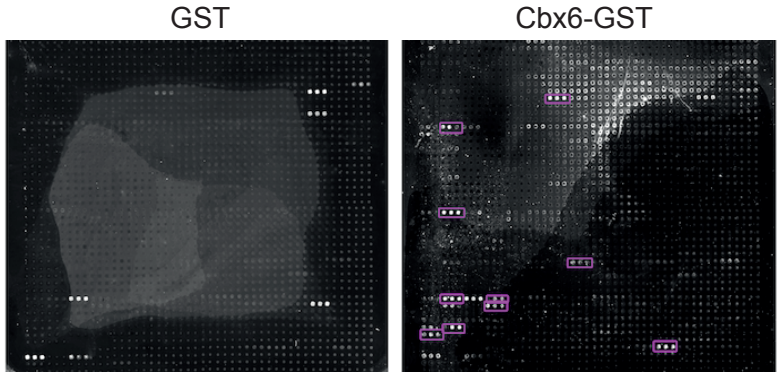


Figure R.23. Representative images of the arrays scanned using an Agilent scanner. Recombinant GST (left panel) and CBX6-GST (right panel) were hybridized onto the arrays. Pink boxes highlight putative CBX6 binding peptides.

To identify peptides with positive binding signals, we overlaid the peptide distribution grid with our scanned images. As peptides are spotted in series of three adjacent spots, we considered a result to be positive when all three peptide spots were lighted up in the same line (such that a single or double spot in a set of three was not considered a true positive result). Ideally, since the same peptide is spotted twice in the same sub-array, there should be six peptide spots lighted up in each sub-array. As expected, the peptide array hybridized with GST showed positive signals for the negative control peptides, and this allowed us to orientate the grid and evaluate false positive signals (**Figure 23**). Since the

background noise signal was relatively high in the array hybridized with CBX6-GST (**Figure R.23**), it was difficult to assess positive hits in some areas of the array. For this reason, we considered a positive binding result when the same peptide (three peptide spots) was spotted at least once in the same sub-array.

We incubated one sub-array for GST, one for CBX7-GST, and three for CBX6-GST. A summary of the binding preferences is presented in **Table R.1**.

Histone peptides recognized by CBX7	Histone peptides recognized by CBX6
H3K9me2	H3R17me2
H3K9me3	H4 (11-27)
H3K27me3	H4K5ac
H3K9me2 + H3K27me2	H4K8ac
H3R17me3	H4K12ac
H3K4acK9me3K16acK18ac	H4K12acK16acK20me3 (1–25)
H4K12acK16acK20me3	H4K16ac
H4K16ac	H4K20ac
H4K20me3	H4K20me3

Table R.1. Summary of the results using histone peptide arrays showing the binding preferences of GST-tagged CBX7 or CBX6. H3 peptides are coloured in green, H4 peptides, in blue.

CBX7 displayed preferential binding to H3K27me3 and H3K9me2/3 peptides, as expected. We also observed affinity of

CBX7 for H4K20me3 and peptides containing other modifications combined with the abovementioned methylation marks, such as H3K9me2 + H3K27me2, H4K12acK16acK20me3, H3K4acK9me3K16ac- K18ac, and other unexpected histone peptides, such as H3R17me3 and H4K16ac.

The CBX6 binding scenario was different from CBX7. Surprisingly, most of the bound peptides were H4-derived peptides. In particular, CBX6 preferentially bound unmethylated histone H4 over methylated histones, such as H4K20me3 and H3R17me2, or acetylated histones, such as H4K5ac, H4K8ac, H4K12ac, H4K16ac, and H4K20ac. Notably, in our arrays, CBX6 did not bind to H3K27me3.

To confirm the array results, we performed histone peptide pull-downs and systematically compared the affinity of full-length recombinant CBX6 to a series of methylated, acetylated, and unmethylated peptides. We utilized biotinylated histone tail peptides conjugated to avidin beads as bait for the bacterial recombinant proteins (GST, CBX7-GST, and CBX6-GST). CBX6 recombinant protein was laborious and difficult to obtain, as the CBX6 protein was rapidly degraded even at the low temperatures used for bacteria cultures. Moreover, the protein precipitated if frozen, so we had to freshly generate recombinant protein for every peptide pull-down experiment that we performed. Representative Western blot analysis of the peptide pull-down experiments are presented in **Figure R.24**. As an internal control of the experiment, we used CBX7-GST, which specifically recognized the H3K27me3 peptide.

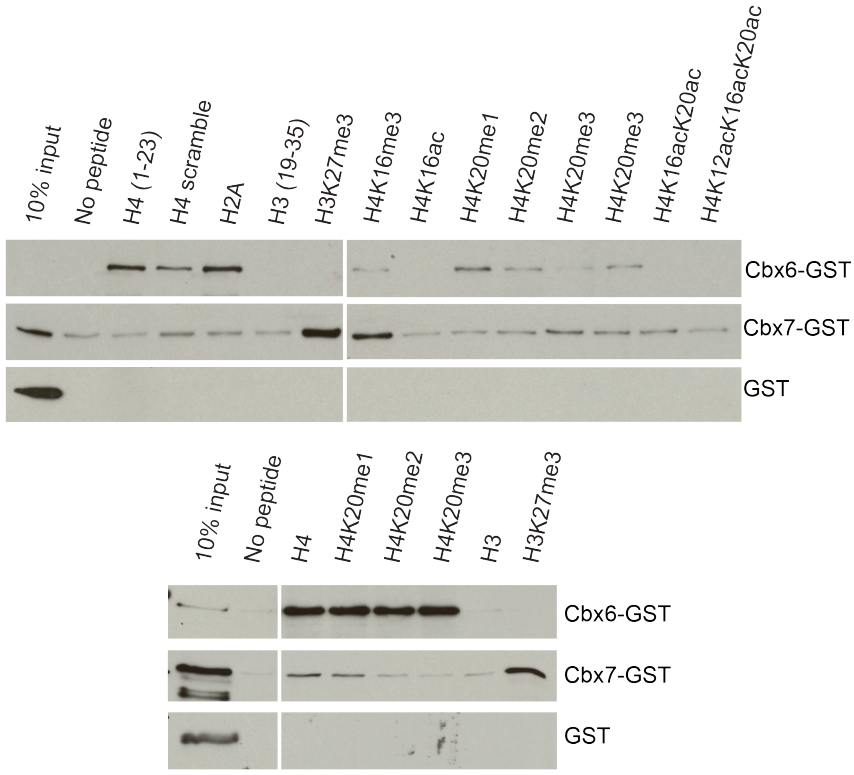


Figure R.24. CBX6 and CBX7 histone binding preferences. Streptavidin pull-down assay interaction analysis of the recombinant proteins GST, CBX6-GST, and CBX7-GST with biotinylated histone peptides. Bound proteins were separated and immunoblotted with anti-GST antibody.

Confirming our histone peptide array results, CBX6 was able to bind to histone H4 unmodified peptide. CBX7 also showed affinity for H4, which was however much lower than that for H3K27me3. Interestingly, CBX6 was able to bind to H2A peptides, possibly due to the similar amino acid sequence of the H4 and H2A histone tails. To determine whether binding of CBX6 to H4 was sequence specific, we designed a scrambled peptide sequence for H4. We observed that CBX6 bound less efficiently to this

peptide, although its binding was not completely disrupted, suggesting that CBX6 binding to H4 partially depends on the amino acid sequence. In agreement to the histone peptide array and to previously reported data, our CBX6 recombinant protein did not bind to H3K27me3. For histone H4 methylated and acetylated peptides, we observed that acetylation prevented CBX6 binding, and that CBX6 preferentially bound to methylated peptides, such as H4K16me3 and H4K20me1/me2/me3. Although the CBX6 binding pattern for H4K20me1/me2/me3 was always positive, the binding affinity gradually decreased with the addition of more methyl groups to the peptide (**Figure R.24**, upper panel). For other peptides, we observed comparable binding properties between the three peptides (**Figure R.24**, bottom panel).

We were also interested in understanding if the chromodomain was indispensable for binding to these histone tails. We thus expressed the Cbx6 construct mutated in the chromodomain (Cbx6-GST^{AA}) as a GST fusion protein for peptide pulldown experiments. We observed that CBX6-GST^{WT} and CBX6-GST^{AA} binding affinities were very similar, suggesting that the chromodomain, at least *in vitro*, is dispensable for histone binding (**Figure R.25**).

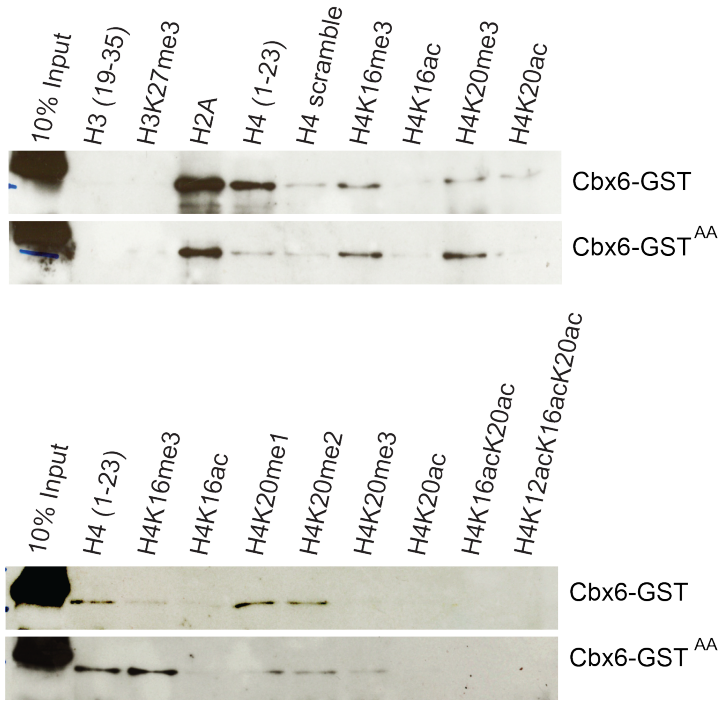


Figure R.25. CBX6 and CBX6-GST^{AA} histone binding preferences. Streptavidin pull-down assay interaction analysis of the recombinant proteins CBX6-GST and CBX6-GST^{AA} with biotinylated histone peptides. Bound proteins were separated and immunoblotted with anti-GST antibody.

Overall, the experiments consistently showed a preference of recombinant CBX6 for unmodified and methylated histone H4 peptides. This binding was reduced when H4 was acetylated (in contrast to what we observed in the histone peptide array). The H4 scramble still showed some binding, indicating that CBX6 affinity for H4 is partially influenced by its sequence as well as its charge.

7. CBX6 interacts with canonical and non-canonical PRC1 proteins

The CBX6 interactome has not yet been studied in ES cells. Likewise, no peptides corresponding to CBX6 have been described in published MS data from PRC1 subunit immunoprecipitation in ES cells. Vandamme and colleagues used a TAP-MS/MS strategy to analyse the CBX6 interactors in HeLa cells[124]. In these cells, they observed that only a low amount of PRC1 protein members (RING1B and PHC2) were identified in the CBX6 pulldown. However, *in vitro* binding experiments indicated that CBX6 has the ability to interact with other PRC1 subunits (in particular, PCGF orthologs) that were not detected in the previous experiment. Moreover, size-exclusion chromatography in this study showed that CBX6 was involved in high–molecular weight protein complexes, suggesting that CBX6 is part of a functional complex *in vivo*.

To gain insight into the CBX6 complex composition in ES cells, we generated stable cell lines expressing CBX6-Flag and used a label-free quantification (LFQ) strategy, of liquid chromatography coupled to tandem–mass spectrometry (LC-MS/MS), to identify the proteins that co-purify with CBX6-Flag on anti-Flag agarose beads. Cryomilled material from control and CBX6-Flag cells was generated and subjected to affinity protein complex purification using an astringent extraction/binding buffer (20 mM HEPES, pH7.4, 400 mM NaCl, 0.5% Triton X-100 v/v). The affinity purified protein complexes were resolved on SDS-PAGE and Coomassie stained. Bands were excised from the gel, and proteins were digested with trypsin and then identified by peptide sequence

determination. Purifications were performed in triplicates. Proteins were filtered with the constraint that at least one group (control or CBX6-Flag) contained at least three valid values. A two sample Student's t-test was performed with an arbitrary fold-change of 2 required for significance and a false discovery rate (FDR) of 0.05. Purification of CBX6-Flag identified ten significant interactors (**Figure R.26**).

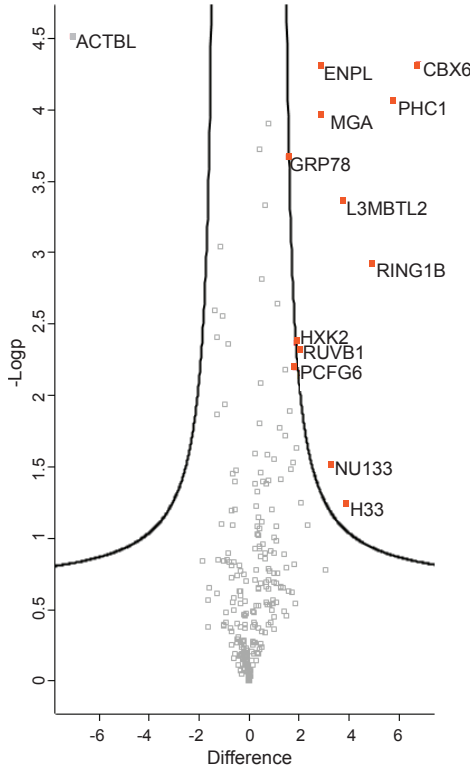


Figure R.26. Statistically-enriched proteins in the FLAG IP identified by permutation-based FDR-corrected t-test.
The plot shows log2 (difference) ratios of averaged peptide intensities of the FLAG pull-down over the control, plotted against the –log 10 (*P* value). The hyperbolic line indicates the permutation-based FDR threshold. The proteins in the upper right corner represent the bait and its interactors (marked in red).

These results showed that CBX6 pulled down members of the PRC1.2 complex (canonical PRC1 complex), such as RING1B and PHC1. Interestingly, we also detected members of the PRC1.6 or E2F6 complex (non-canonical PRC1 complex), such as RING1B, L3MBTL2, MGA, and PCGF6. Other non-Polycomb related proteins were also identified, such as ENPL (a chaperone), HXK2 (kinase involved in glucose metabolism), and UD5 (a nucleoporin).

We were especially interested in understanding the binding of CBX6 to the canonical and non-canonical PRC1 complexes. CBX6 interacting partners were confirmed by co-immunoprecipitation followed by immunoblot (**Figure R.27**).

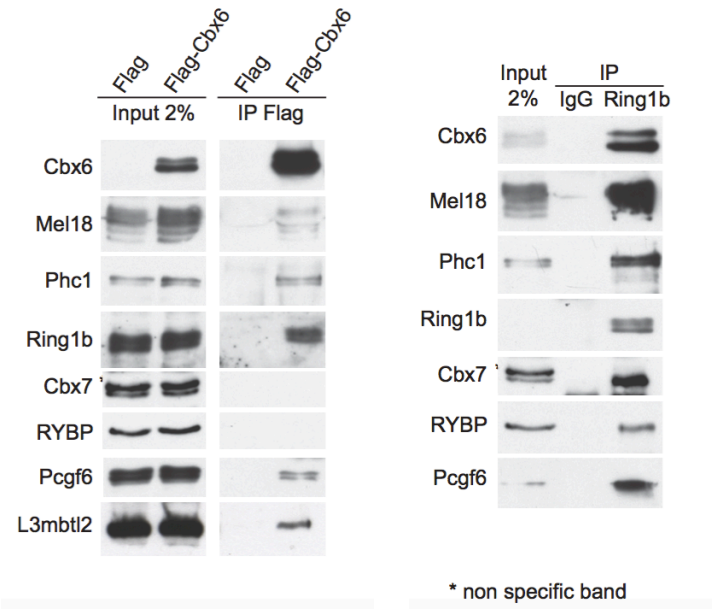


Figure R.27. Analysis of CBX6 interactions in ES cells. Co-IPs from total ES cell extracts using antibodies against RING1B (left) or FLAG (right). Western blots of different proteins are shown.

We also observed an interaction between CBX6 and PCGF2 (MEL18), a member of the cPRC1 complex, although this protein was not detected in the MS analysis. As expected, we saw no CBX6 interactions with CBX7[124] or RYBP, a core component in different non-canonical PRC1 complexes, such as the PRC1.6 complex. Note that CBX proteins and RYBP are mutually exclusive within the PRC1 complex, as they compete for the same binding region of RING1B[85]. Interestingly, an interaction between CBX6 and RING1B was also observed in a reverse immunoprecipitation experiment (**Figure R.27**).

These data suggested that CBX6 could be part of at least one, yet-unidentified PRC1 complex that contains the above-mentioned subunits. Further experiments are required to understand the composition and stoichiometry of such CBX6-containing complex/es.

8. Difficulties of mapping CBX6 binding sites genome-wide

Previous data in our laboratory reported the genome-wide distribution of CBX6[112]. However, after many attempts to validate the results, we determined that the antibody used was unspecific. Thus, achieving a CBX6-binding profile in a genome-wide manner has been highly challenging. In this section, I summarize the different approaches that we have used. A summary of the strategies followed is shown in **Table R.1**.

As no commercial antibodies against CBX6 were suitable for ChIP experiments, we therefore decided to generate an anti-CBX6 antibody to use for determining the genome-wide CBX6 distribution in ES cells. We produced the antibody using the full-length CBX6 recombinant protein (CBX6-GST) as bait to immunize the rabbits. After purification and concentration, we used the antisera for ChIP-sequencing in either control ES cells or those depleted for CBX6. However, this antibody also proved to be unspecific, as there was no decrease in the ChIP-seq signal in CBX6-depleted ES cells.

We then decided to map the CBX6 chromatin occupancy using a FLAG-tag. For this purpose, we overexpressed CBX6-3×FLAG in ES cells and set up a protocol for FLAG-ChIP. As we did not know any target gene of CBX6, we took advantage of an established cell line expressing RING1B-3×FLAG as a positive control for the FLAG-ChIP protocol. Unfortunately, ChIP-seq results from FLAG-chip experiments resulted in no peaks.

By this time, the ChIP-seq binding profile for CBX6 in human fibroblasts (HF) had been reported[123]. In this study, the authors used a home-made antibody, which Dr. Gordon Peters kindly provided to us. We thus performed ChIP experiments of CBX6 in shCtrl/shCbx6 ES cells and, as a positive control, in human fibroblasts. Although our ChIP worked in human fibroblast cells, sequencing results showed no enrichment of CBX6 in ES cells. We optimized the protocol by using an additional crosslinker (DSP), considering the possibility that CBX6 was unstably bound to chromatin, or by overexpressing CBX6. However, the genome-wide sequencing results using these optimized conditions were also negative.

We next tried the HA (hemagglutinin) tag, which has been extensively used elsewhere for ChIP assays. We established CBX6-3×HA– and RING1B-3×HA–overexpressing ES cell lines. We used either the standard protocol or a commercial kit (ChIP-IT high sensitivity kit, Active Motif), which is optimized to give superior results when performing ChIP using antibodies against low abundance proteins or antibodies with sub-optimal binding affinities. As expected, RING1B-ChIP worked using both methods. Importantly, the ChIP-seq for CBX6 worked when used the ChIP-IT high sensitivity kit.

Antibody	Description / Conditions	Sequencing result
αCBX6 home-made	Endogenous conditions Standard protocol	✗
αFLAG	Overxpressed conditions FLAG protocol	✗
αCBX6 Pemberton	Endogenous conditions Standard protocol	✗
αCBX6 Pemberton	Endogenous conditions Double crosslink protocol	✗
αCBX6 Pemberton	Overxpressed conditions Standard protocol	✗
αHA	Endogenous conditions Standard protocol	✗
αHA	Endogenous/Overxpressed ChIP-IT kit	✓

Table R.2. Summary of CBX6 ChIP-sequencing attempts. Red cross, negative sequencing results; green checkmark, positive sequencing results.

To study the CBX6 binding profile in a more biological context, we performed genomic editing using CRISPR/Cas-9 technology. For this, we established an ES cell line expressing *Cbx6*-3×HA at endogenous levels. Importantly, the *Cbx6*-3×HA cell-line behaved as the control cell line (**Figure R.28**).

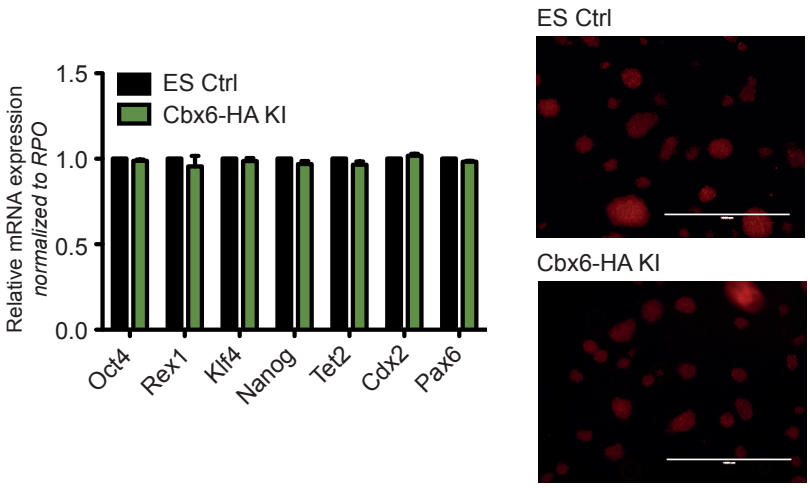


Figure R.28. Knockin (KI) cells for Cbx6 behave normally. Left, qRT-PCR analysis of control and Cbx6-HA knockin ES cells. Results are shown relative to control and normalized to RPO. Error bars represent SD of two independent experiments. Right, alkaline phosphatase staining of Ctrl and Cbx6-HA knockin ES cells.

These conditions were then used to obtain the CBX6 ChIP-seq profile and determine its features.

9. CBX6 genome-wide distribution

Using the endogenous tagged CBX6-3×HA and empty control cell line, we then performed ChIP-sequencing experiments, taking advantage of the above-mentioned ChIP-IT high sensitivity kit. The results of the ChIP-seq are summarized below (**Table R.3**)

	CBX6	Empty
Mapped/Total Reads*	50,973,597/74,999,037	47,504,266/69,215,258
Peaks	5,103	
Genes**	3,730	

Table R.3. Summary of mapped reads, peaks, and target genes identified for CBX6 using ChIP-seq. *Multi-locus reads were removed. NR, non-redundant. **, peaks in the region (from 2.5 kb upstream of the gene until its end)

We identified 3,730 CBX6 target genes. CBX6 was predominantly found at the transcription start site (TSS) of the genes (36.1%), including both proximal and distal regions. Some of the CBX6 peaks were associated with coding sequences (17.6%) and intronic regions (14.9%). Finally, a small percentage associated with intergenic regions (8.4%) (**Figure R.29**).

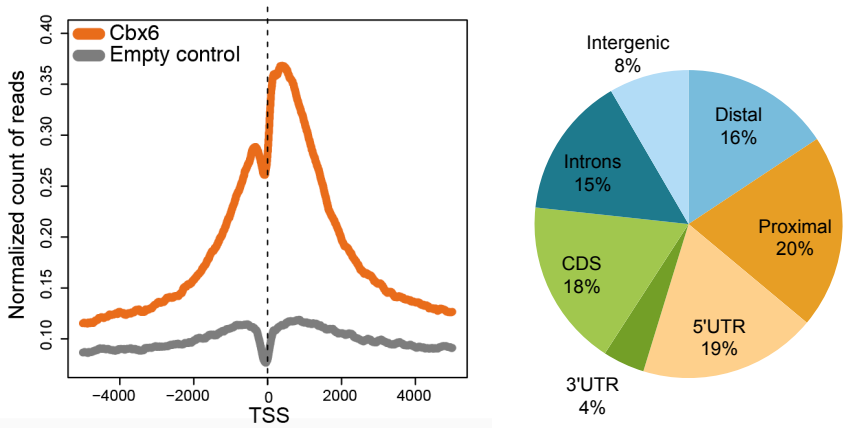


Figure R.29. CBX6 associates preferentially with the TSS. Left, graphical distribution of normalized count of reads 5 kb upstream and downstream of TSS for CBX6 target genes. Right, pie chart showing peak distribution of CBX6 across genomic features.

We confirmed the CBX6 ChIP-seq results by ChIP-qPCR experiments (**Figure R.30**).

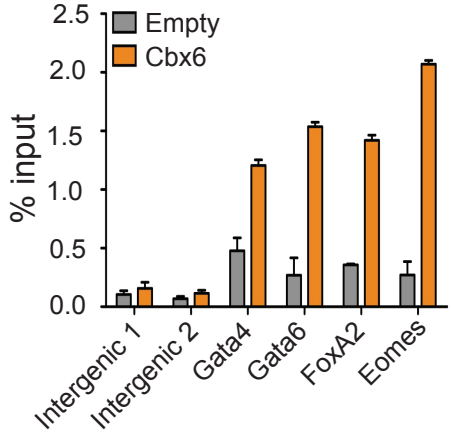


Figure R.30. CBX6 peaks are specific. ChIP-qPCR validation of target genes of CBX6 in control and Cbx6-3×HA expressing cells. Results are shown relative to the input percentage. Error bars represent SD of three biological replicates.

To better understand the function of CBX6 target genes, we next performed GO analysis using the Enrichr software (**Figure R.31**).

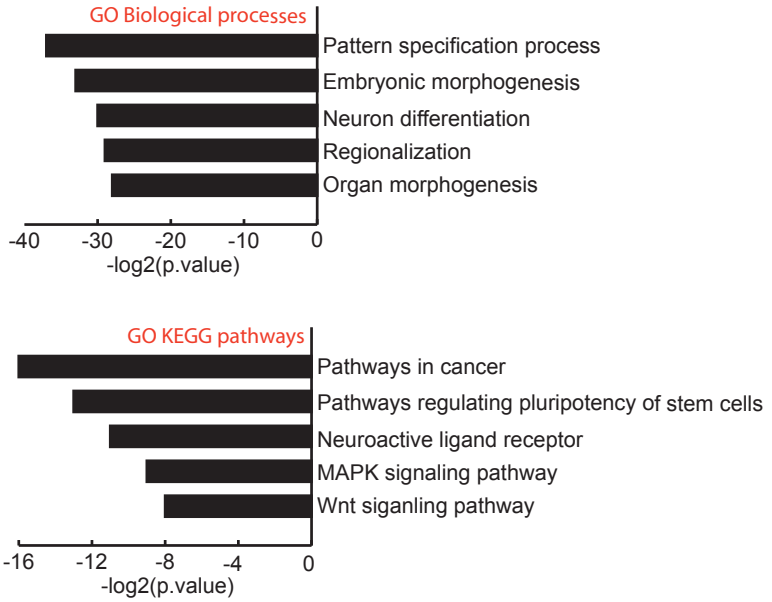


Figure R.31. GO (Enrichr) analysis of Cbx6 target genes. Cbx6 target genes are enriched in developmental biological processes, cancer related pathways, and pathways regulating the pluripotency network.

CBX6 target genes were involved in biological processes related to embryonic development, such as pattern specification processes, embryonic morphogenesis, neuron differentiation, and regionalization. These processes appeared recurrently when analyzing target genes from the Polycomb group family of proteins, as these proteins are required for the correct spatiotemporal expression of genes during embryo development[132]. Interestingly, CBX6 target genes included in these categories were the *Hox* clusters, *Sox17*, *Otx2*, *Gata4*/6,

Tbx20, *Cdx2,* and *Pax6*, all known to be required for proper embryonic development.
GO terms derived from the KEGG pathway database revealed that CBX6 target genes were enriched in genes involved in (I) pathways in cancer, (II) signalling pathways regulating pluripotency of stem cells, (III) MAPK signalling pathways, and (IV) WNT signalling pathways.

Polycomb proteins have been previously associated with carcinogenesis when deregulated[133].
CBX6 in particular has been shown to be involved in gliobastoma multiforme progression[123]. In ES cells, CBX6 targeted genes, such as *Jun* and *Fos*, are well characterized to be proto-oncogenes.

Importantly, Cbx6 target genes were also enriched in pluripotency control-related pathways, such as the MAPK (*Map3k1, Mapk11/12, Fgf2-5/8-10/12/16/18/20*, and *Fgfr1/2/3/4*) and WNT (*Wnt1/2/3/3a/4*) signalling pathways. As mentioned in the Introduction section, extrinsic signalling through different molecules can affect ES cell commitment either positively (by reinforcing the pluripotency state) or negatively (by triggering differentiation). Therefore, CBX6 directly regulates the delicate balance between factors that support and those that prevent differentiation. In sum, GO results correlated with the observation that CBX6-depleted ES cells spontaneously differentiate.

We further checked the expression levels of CBX6 target genes by RNA-seq. Similarly to other PRC1 subunits, CBX6 decorated genes were repressed or slightly expressed, in contrast with

H3K36me3 containing genes, which were highly transcribed (**Figure R.32**).

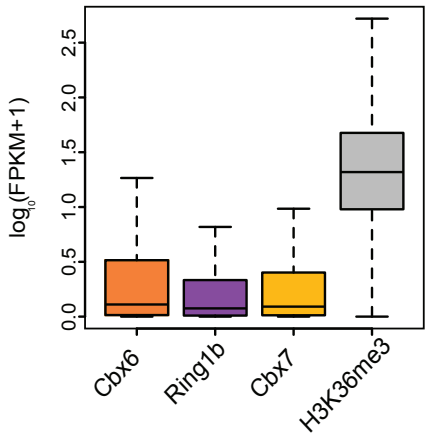


Figure R.32. CBX6 target genes are repressed. Box plots showing expression of CBX6, RING1B, CBX7, and H3K36me3 target genes.

We next studied the connection between CBX6 genome-wide distribution and the transcriptional changes upon CBX6 depletion. From the 703 upregulated genes in CBX6-depleted cells, 189 genes were occupied by CBX6 (26.8%); from the 783 downregulated genes, 206 genes were occupied by CBX6 (26.3%) (**Figure X**). Thus, the expression of 395 CBX6 target genes (10.6%) were deregulated upon CBX6 depletion. These results indicated that CBX6 is an essential regulator of gene transcription.

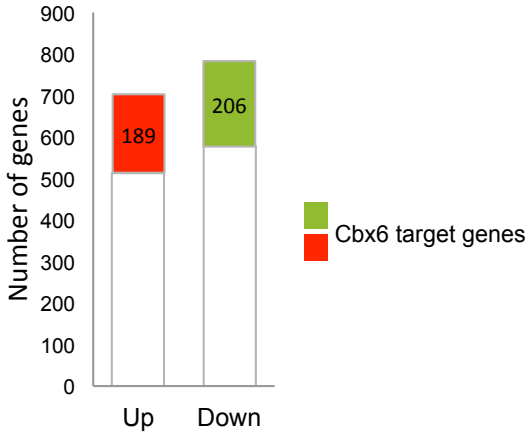


Figure R.33. Cbx6 has a dual function. Number of genes, and target genes, de-regulated upon Cbx6 knockdown.

To determine whether genes transcriptionally affected in CBX6-depleted cells displayed any interesting feature, we performed GO analysis for both gene sets using Enrichr. A summary of the results is presented in **Table R.4**.

		Name	*P*-value	Genes involved
UP	Biological process	Embryonic digestive tract morphology	4×10^{-6}	*Id2, Sox11, Pitx2, Ovol1, Fgfr2*
		Epithelial cell differentiation	6.5×10^{-6}	*Cebpb, Fzd2, Dmrt1, Id2, Pou3f1, Evpl*
		Epithelium development	2.6×10^{-5}	*Fgfr2, Epcam, Egfr, Dll1, Spry1, Sox11*
	KEGG	MAPK signaling pathway	7.8×10^{-4}	*Dusp4/5, Fgf5, Flnc, Mapk12, Egf, Gfgr2,*
		Transcriptional misregulation in cancer	3.6×10^{-4}	*Cebpb, MycN, Wt1, Prom1, RxRg*

		Pathways regulating pluripotency	2.6×10^{-3}	*Fzd2, Id2, Wnt8a, Pax6, Mapk12, Fgfr2*
DOWN	Biological process	Regulation of cell development	4.2×10^{-8}	*Lyn, Sema7A, Smad3, Tgfb1, VegfA, Bmp4*
		Morph. involved in differentiation	1.1×10^{-8}	*Ptprd, Smad3, Islr2, Mapt, Tgfb1, Amigo1,*
		Regulation of neuron projection development	6.2×10^{-9}	*Lyn, Sema7A, CamK1D, Amigo1, Lpar1, Iptka*
	KEGG	Pathways in cancer	1.8×10^{-4}	*Stat5B, Tgfb1, Fgf10, Smad3, Bmp4, RxrA,*
		Pathways regulating pluripotency	3.5×10^{-3}	*Bmp4, Smad3, Fzd4, InhBB, Bmi1, Tbx3*
		PI3K-Akt signaling pathway	7.8×10^{-3}	*RxrA, PPP2R2c, Ccnd2, Flt4, Fgf10*

Table R.4. GO analysis of CBX6 target genes that changed expression in CBX6-depleted ES cells. Analyses were performed with the Enrichr software. UP, upregulated genes; DOWN, downregulated genes upon CBX6 depletion.

Surprisingly, both sets of genes scored similar categories. Biological processes were involved in morphogenesis, development, and differentiation. Affected pathways were related to cancer and pluripotency. Importantly, many genes from the MAPK pathway were upregulated, suggesting again that the deregulation of crucial factors involved in pluripotency was responsible for the spontaneous differentiation observed upon CBX6 depletion.

10. CBX6 genome-wide distribution overlaps with PRC1

Proteomic analysis of CBX6 revealed that CBX6 interacted with both canonical and non-canonical PRC1 subunits. Thus, we compared the genome-wide distribution of CBX6 with cPRC1 and ncPRC1 target genes.

We generated new ChIP-seq data for RING1B and CBX7 using the same experimental procedure as for CBX6, in order to compare the results without technical biases (**Figure x**). We identified 4,055 target genes for CBX7 and 11,920 target genes of RING1B. Most of CBX7 and RING1B peaks were associated to the transcription start site (TSS) of the genes (**Figure R.34**).

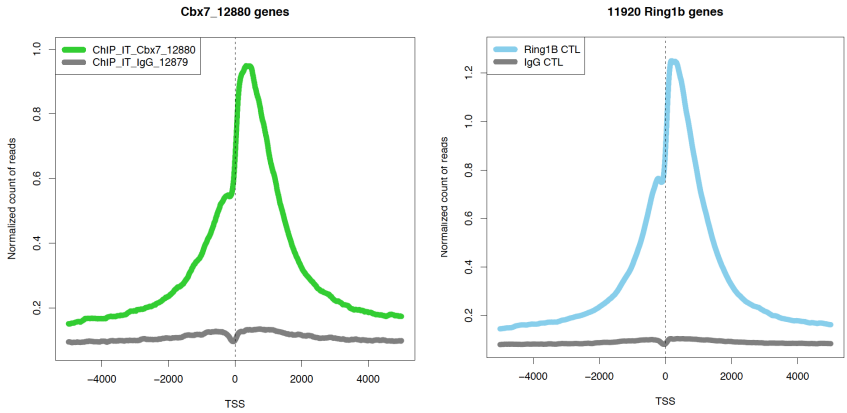


Figure R.34. CBX7 and RING1B peaks associate with the TSS. Graphical distribution of normalized count of reads 5 kb upstream and downstream of TSS for CBX7 and RING1B target genes.

We next overlapped CBX6 target genes with RING1B and CBX7 targets, together with other ChIPseq data generated in the

laboratory for MEL18, H2Aub, and SUZ12, as well as with published ChIP-seq data for L3MBTL2[134] and PCGF6[135].

Heatmaps of CBX6 ChIP-seq signals revealed that CBX6 targets are strongly co-occupied by RING1B, moderately occupied by CBX7, MEL18, and SUZ12, and decorated with H2AK119ub (**Figure R.35**).

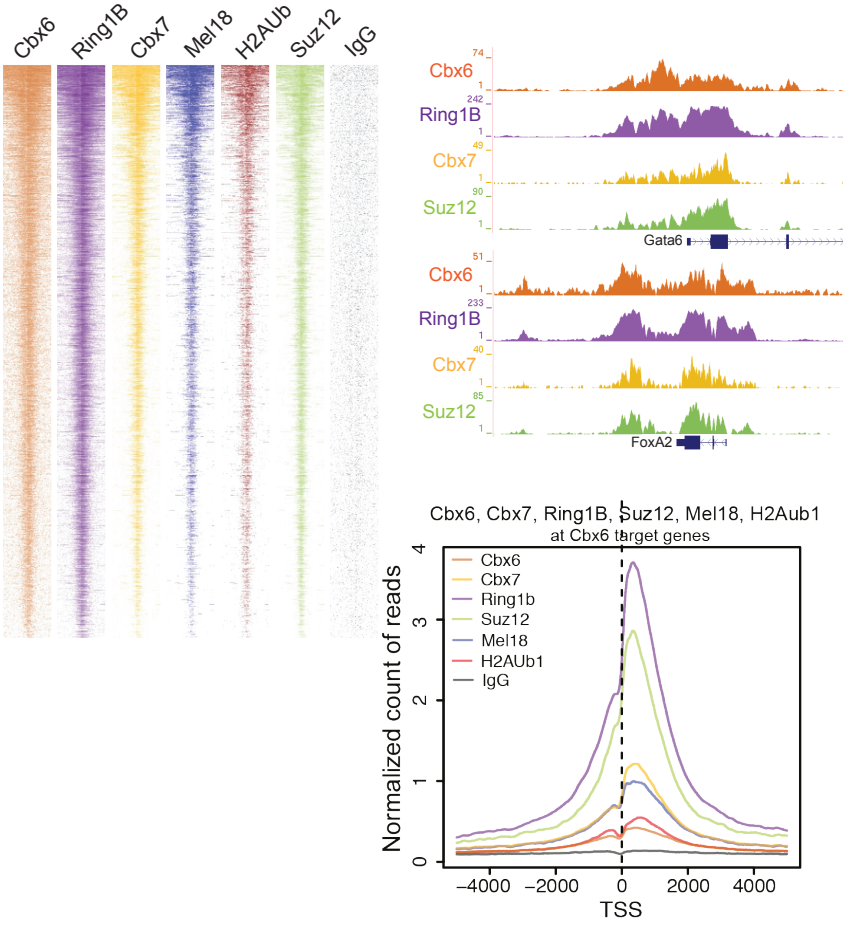


Figure R.35. Cbx6 occupies canonical PRC1 cites. Left, ChIP-seq heatmaps of Cbx6, Ring1B, Cbx7, Mel18, H2AK119ub, Suz12, and IgG. Target genes were defined by the presence of one or more Cbx6 peaks within ±2.5 kb of a given transcriptional start site (TSS). Right, UCSC screenshots of Cbx6, Ring1B, Cbx7, and Suz12 ChIP-seq profiles for selected genes. Numbers at the top left of each graph represent the number of reads for each ChIP-seq in a given gene. Below, normalized signals relative to all TSSs for Cbx6, Ring1B, Cbx7, Mel18, H2AK119ub, Suz12, and IgG ChIP-seq.

Further analysis of CBX6, and RING1B, CBX7, MEL18, SUZ12 and H2AK119ub ChIP-seq signals at the 3,730 CBX6 target genes corroborated that most of CBX6 targets were cPRC1 target genes (**Figure R.36**).

We also overlapped CBX6 genome-wide distribution with the non-canonical subunits that came out in the CBX6 interactome. Around 20% of CBX6 target genes overlapped with L3MBTL2, whereas the overlap with PCGF6 was very minor (**Figure R.36**).

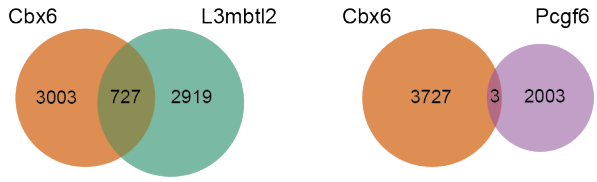


Figure R.36. CBX6 does not overlap with ncPRC1. Venn diagrams showing the overlap of CBX6 target genes with either L3MBTL2 or PCGF6 target genes.

These results suggest that CBX6 occupancy is strongly correlated with cPRC1 complex, but not ncPRC1, target genes.

11. cPRC1 distribution is not affected in CBX6-depleted ES cells

To address whether CBX6 contributes to cPRC1 stability and recruitment in ES cells, we performed ChIP-qPCR experiments in control and CBX6-depleted ES cells at two different promoter cohorts: (i) CBX6 target genes downregulated upon CBX6 depletion (such as *Esrrb*, *Prdm14*, *Gsc*, and *Tbx3*), and (ii) CBX6 target genes upregulated upon CBX6 knockdown (such as *Pkdcc*, *Mmp25*, *Pou3f1*, *Nestin*, and *Sox3*). In contrast to CBX7 depletion, RING1B and CBX7 occupancy was not reduced in shCbx6 cells, except for at some genes (**Figure R.37**).

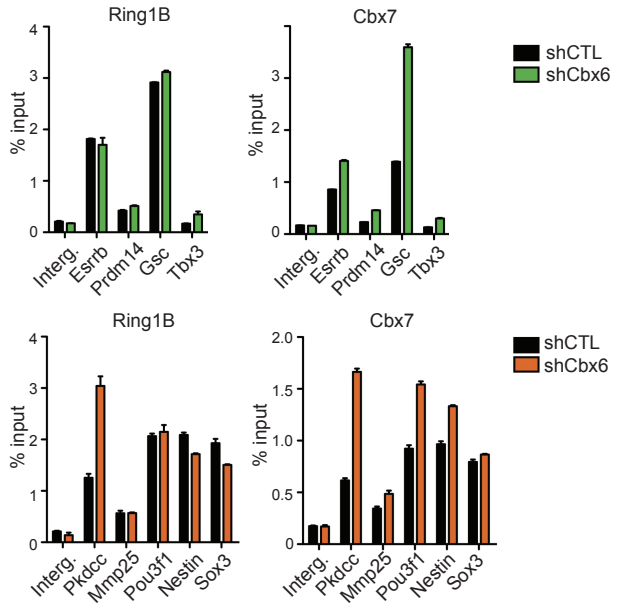


Figure R.37. CBX6 is not essential for cPRC1 stability. ChIP-qPCR of CBX7 and RING1B in shCTL or shCbx6 ES cells. An intergenic region was used as a negative control gene. Green bars depict downregulated genes, and red bars, upregulated genes. Results are shown relative to % of input. Error bars represent SD of three biological replicates.

To exclude the possibility that there was a genome-wide redistribution of RING1B that we were not able to capture within the analysed promoters, we performed ChIP-seq of RING1B in shCtrl and shCbx6 ES cells using the regular ChIP-seq protocol. We obtained 3,633 RING1B target genes in shCtrl ES cells, and 3,814 in shCbx6 ES cells, 3,411 of which were common to both.

In accordance with the ChIP-qPCR results (**Figure R.37**), we observed an increase in the RING1B occupancy in CBX6-depleted cells (**Figure R.38**).

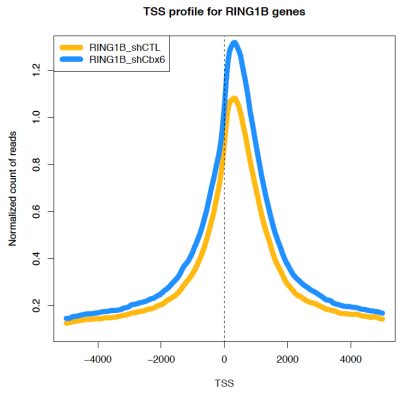


Figure R.38. Gain of RING1B after CBX6 depletion (I). Graphical distribution of normalized count of reads 5 kb upstream and downstream of TSS for RING1B target genes in shCTL or shCbx6 ES cells.

To elaborate a more quantitative approach of this observation, we measured the ChIP signal levels in the promoters of the full set of genes in the genome to identify directly those presenting a significant gain (or loss) after CBX6 depletion. In particular, we selected those genes in which the ChIP-seq level 500 bp around the TSS of one sample over the other one was above 1.75-fold

change (FC). We identified 116 genes that significantly exhibited higher RING1B ChIP-seq levels, and 351 genes that lost RING1B. As genes losing RING1B already had very low levels of RING1B in their promoters, we considered them to be non-specific; we thus focused in the cohort of genes that gained RING1B (**Figure R.39**, examples UCSC).

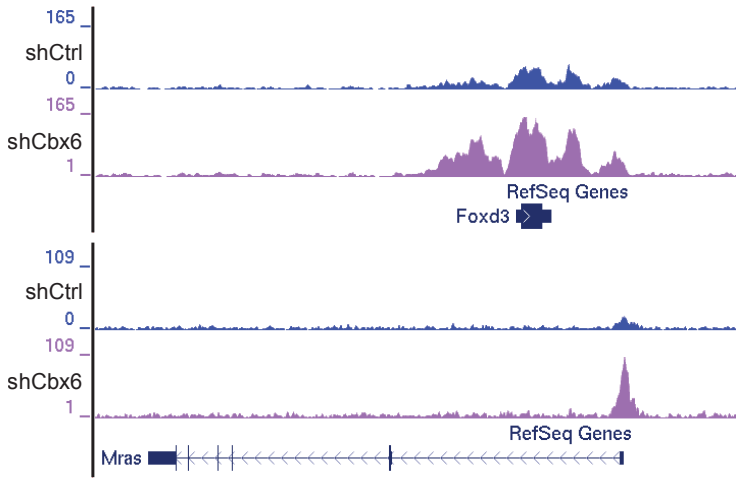


Figure R.39. Gain of RING1B after CBX6 depletion (II). UCSC screenshots of RING1B ChIP-seq profile for selected genes in shCTL (blue) or shCbx6 (purple) ES cells. Numbers at the top left of each graph represent the number of reads for each ChIP-seq in a given gene.

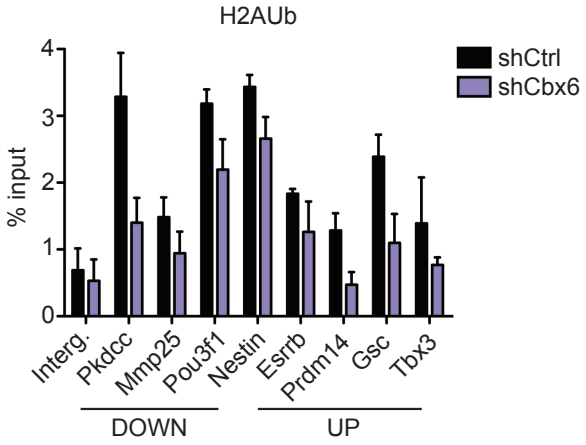
We next overlapped the 116 gene list with the RNAseq data and found that 27 genes that gained RING1B in their promoters were downregulated following CBX6 depletion (**Table R.5**).

A930011O12Rik, Ahnak, Camk1D, Cdx1, Col13a1, Gbx2, Gm53, H2afy2, Hopx, Inhbb, Itpka, Lrrn2, Mras, Nfib, Pou2f3, Prdm14, Sfrp1, Shisa2, Shisa7, Smox, Spsb1, Stac2, Tbx3, Tet2, Tmem108, ZC3hav1, Zfyve28

Table R.5. RING1B increased its binding at CBX6 target genes. List of genes that are CBX6 target genes and that were downregulated with increased RING1B binding in CBX6-depleted cells.

Overall, we observed that CBX6 depletion did not majorly affect the genome-wide distribution of cPRC1. A small subset of genes significantly gained RING1B in their promoters, but this RING1B gain only resulted in a decrease of transcription for 27 genes. The transcription reduction of the other downregulated genes (756) can thus be expected to be due to other repression mechanisms.

As RING1B distribution was not majorly affected, we hypothesized that changes in transcription levels could be related to an impairment in the catalytic efficiency of RING1B for histone H2A ubiquitination. We therefore checked the H2Aub levels by ChIP-qPCR experiments of promoters of genes that were up- or downregulated (**Figure R.40**).



FigureR.40. X CBX6 depletion does not affect RING1B catalytic activity. ChIP-qPCR of H2AUb in shCTL or shCbx6 ES cells. An intergenic region was used as a negative control gene. DOWN, downregulated genes. UP, upregulated genes. Results are shown relative to % of input. Error bars represent SD of three biological replicates.

We saw a moderate loss of H2AK119ub levels in some of the analysed promoters. However, H2AK119ub loss was observed in both up- and downregulated genes, making it challenging to explain the subsequent changes in gene transcription.

We also tested whether PRC2 occupancy and H3K27me3 deposition were affected by CBX6 depletion. ChIP-qPCR experiments revealed that neither SUZ12 nor H3K27me3 levels were affected following knockdown of *Cbx6* (**Figure R.41**).

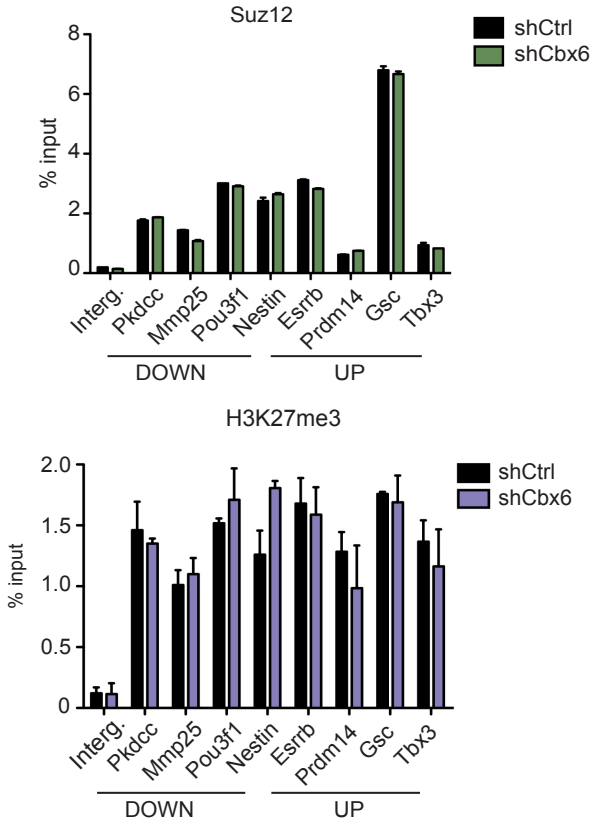


Figure R. 41. CBX6 depletion does not affect PRC2 occupancy. ChIP-qPCR of Suz12 and H3K27me3 in shCTL or shCbx6 ES cells. An intergenic region was used as a negative control gene. DOWN, downregulated genes. UP, upregulated genes. Results are shown relative to % of input. Error bars represent SD of two biological replicates.

We next studied the effect of ablating PRC2 function on the recruitment of Cbx6 to chromatin. We stably knocked down *Suz12* using an shRNA construct in the control and CBX6-HA expressing cell line. We successfully depleted both SUZ12 at the protein level and its resulting histone mark, H3K27me3 (**Figure R.42**).

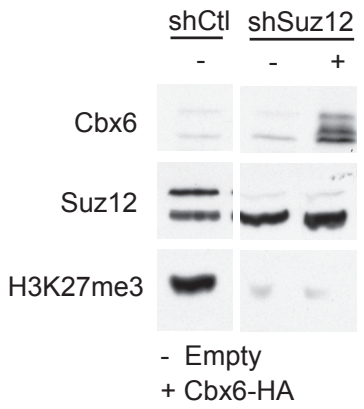


Figure R. 42. Suz12 and H3K27me3 lost upon Suz12knockdown. WB analysis of Cbx6, Suz12 and H3K27me3 in shCTL and shSuz12 ES cells.

ChIP-qPCR experiments (using the ChIP-IT kit) for RING1B, CBX7, and CBX6 further revealed that the level of occupancy of RING1B and CBX7 was significantly compromised in SUZ12-depleted ES cells. Unexpectedly, in an apparent contrast to previous results from our lab[112], we did not observe a large reduction of either RING1B or CBX7 signal at the promoters (**Figure R.43**). However, two aspects should be taken into account in our experiment: (i) residual SUZ12 protein and H3K27me3 marks could still recruit PRC1 to chromatin; and (ii) we used a high sensitivity protocol to compare conditions, which is able to ChIP low abundant proteins, such that if RING1B and CBX7 were not totally displaced from chromatin, any remaining would give a positive signal. Although CBX7 recruitment is fully dependent on H3K27me3, this is not the case for RING1B, which can be recruited as part of another complexes (for example, RYBP-PRC1 complex) in a PRC2-independent fashion.

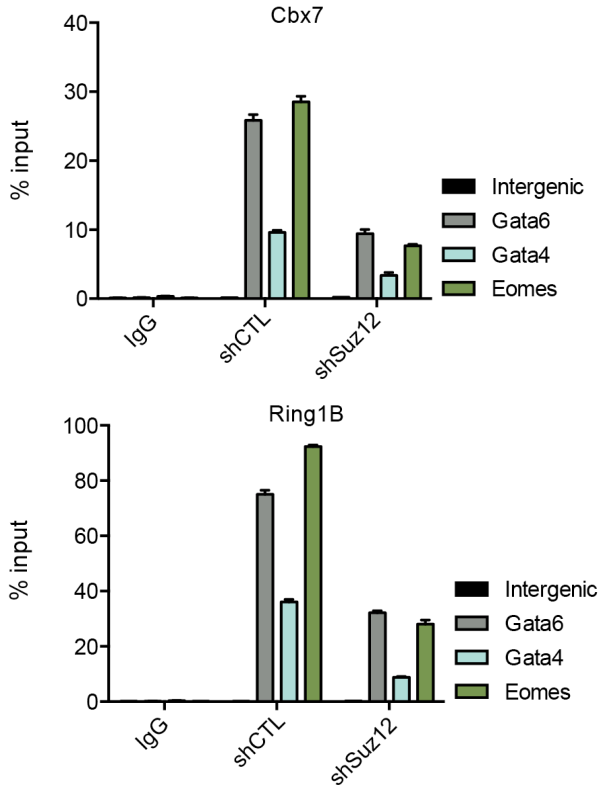


Figure R. 43. CBX7 and RING1B occupancy is partially lost upon PRC2 function ablation. ChIP-qPCR of CBX7 and RING1B in shCTL and shSuz12 ES cells. An intergenic region was used as a negative control gene. Results are shown relative to % of input. Error bars represent SD of two biological replicates.

Based on this information, we analysed what happened to CBX6 in a PRC2-depleted context. Interestingly we observed that CBX6 levels decreased upon Suz12 depletion, suggesting that CBX6 recruitment to chromatin depended on a functional PRC2 complex (**Figure R.44**).

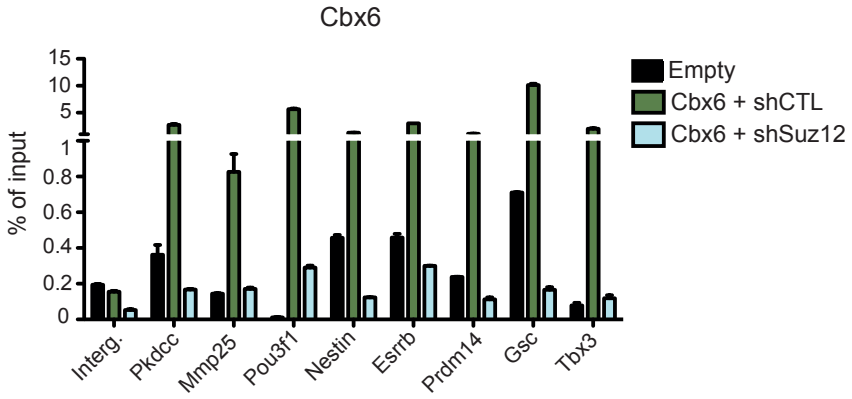


Figure R. 44. Cbx6 occupancy depends on a functional PRC2 complex. ChIP-qPCR of Cbx6 in shCTL and shCbx6 ES cells. An intergenic region was used as a negative control gene. Results are shown relative to % of input. Error bars represent SD of two biological replicates.

DISCUSSION

In recent, years comprehensive research in the ES cell biology field has provided essential cues for understanding the mechanisms underlying cell fate decisions, and for posing ES cells as promising tools for disease therapy. Epigenetic factors have emerged as indispensable regulators of ES cell biology with countless functions, such as modulation of chromatin structure and covalent post-translational modifications of histones. In this thesis, we have studied the function of CBX6, one of the CBX orthologs associated to Polycomb complexes, in ES cells. We show that CBX6 is a key epigenetic factor required for balancing pluripotency and differentiation of ES cells. We unequivocally show that CBX6 depletion strongly favours ES cell differentiation. At the molecular level, we show that CBX6 interacts with canonical and non-canonical PRC1 subunits, while its genome-wide distribution utterly overlaps with the cPRC1 complex.

1. CBX6 is essential for maintaining the pluripotent state of ES cells

Our data provide support for CBX6 being an indispensable factor for ES cell biology, as it is involved in controlling pluripotency and self-renewal. Specifically, our results indicate that CBX6 regulates the expression of gene networks that control these fundamental biological functions. Several observations support this view. In CBX6-depleted ES cells, expression of central pluripotency network regulating genes (*i.e. Klf4*, *Nanog*, and *Rex1*) was strongly downregulated. Concomitantly, there was a premature expression of mesoderm (*Brachyury*) and ectoderm (*Pax6* and *Otx2*) markers. Moreover, CBX6-depleted ES cells are poorly stained by AP, a specific stem-cell staining assay. CBX6 is not

only required for pluripotency but also for self-renewal, as depletion of CBX6 resulted in less proliferation as compared to control cells (differentiated cells do not re-attach to the plate after trypsinization). Taken together, these observations strongly indicate that the CBX6 functions in ES cells are critical for suppressing differentiation.

Notably, this is the first time that a differentiation phenotype has been observed following depletion of a canonical PRC1 complex subunit. CBX6 and CBX7 are the most highly-expressed CBX proteins in ES cells. In contrast to CBX6, CBX7 is not necessary to preserve self-renewal of ES cells. Previous work from our lab showed that the pluripotent markers OCT4 and NANOG are not affected after CBX7 depletion[112]. Furthermore, AP staining indicated that CBX7-depleted ES cells did not spontaneously differentiate. These results are similar to those observed for RING1B-mutant ES cells. However, complete impairment of PRC1 function (following a *Ring1B* and *Ring1A* double depletion) strongly affects self-renewal[111]. The finding that ES cells with a CBX6 depletion exhibited a phenotype that is not seen upon single depletion of other core PRC1 components suggests that CBX6 possesses a PRC1-independent, or at least a canonical PRC1–independent role, and that other family members cannot compensate for its deletion.

2. CBX6- depleted ES cells spontaneously differentiate into ectoderm.

In line with the effect of CBX6 knockdown in cell differentiation, our genome-wide expression study shows that CBX6-regulated genes have a significant overrepresentation of cell fate

commitment regulators. Among the top one-hundred genes that are up- or downregulated after CBX6 depletion are a significant group of targets involved in cell-cell adhesion, regulation of the FGFR signaling pathway, neuron differentiation, epithelial cell differentiation, and extracellular matrix organization. Claudin3, 6 and 7 are overexpressed in CBX6-knockdown ES cells. Claudins are integral membrane proteins found in tight junctions of all epithelia and endothelia. In addition to acting as a selective barrier between cells, they have also been involved in determining cell morphology. Moreover, CBX6-depleted cells overexpress receptors that are essential to determine epithelial commitment, such as the WNT receptor Frizzled2 (FZD2)[136] and LEF1, which also participates in the WNT signaling pathway for ectoderm specification[137]. Additionally, CBX6-depleted ES cells are enriched in neuronal genes that have been described to be indispensable for accurate neuroectoderm specification, such as *Otx2*, *Nestin*, and *Pax6*. Changes in the expression of these genes clearly match changes in cell morphology due to the spontaneous differentiation observed in CBX6-depleted ES cells, and suggest that these cells are biased towards the ectoderm fate. Precise differentiation experiments should be applied to CBX6 depleted ES cells to better characterize CBX6 contribution on lineage specification.

Downregulated genes are related to extracellular matrix homeostasis, with collagens (*Col4a1*, *Col4a2*, and *Col13a1*) and *Laminin1* the most downregulated genes following CBX6 depletion. Because collagen-related networks helps the basement membranes to interact with nearby cells, and thus are associated with cell proliferation and differentiation, we postulate that these results partially reflect the observed phenotype of CBX6-depleted

ES cells.

CBX6-depleted ES cells are biased towards an ectoderm lineage, which might be explained due to a particular combinatorial downregulation of pluripotency factors. In addition to their role in maintaining ES cell pluripotency and self-renewal, pluripotency factors are also involved in early cell fate decisions. Loh and Lim pursued the idea that pluripotency is defined by transcriptional competition between the lineage-specifying actions of pluripotency factors[3]. In fact, it has been shown that they are indispensible for proper differentiation, as loss of individual pluripotency transcription factors frequently prompts ES cell differentiation to specific lineages, and their overexpression often induces differentiation.

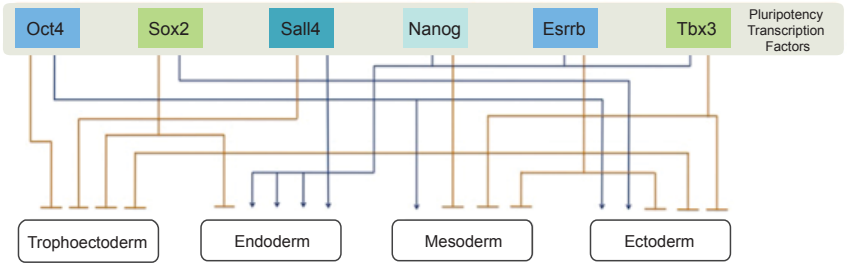


Figure D.1. Pluripotency factors as lineage specifiers. Pluripotency factors function as classical lineage specifier transcription factors and confer ES cells with the ability to differentiate to any of the germ layers. Perturbations in the expression of these transcription factors will trigger differentiation. Adapted from [3].

In CBX6-depleted ES cells, pluripotency factors like *Klf4*, *Nanog*, *Esrrb*, and *Tbx3* are downregulated, whereas *Oct4* and *Sox2*

transcript levels are not affected. Perturbations in the expression of these transcription factors would result in collapse of this fragile transcriptional equilibrium, with the consequence that lineage commitment initiated by OCT4 and SOX2 towards the ectoderm fate could take place (**Figure D.1**)

3. CBX6 regulates the pluripotency network trough the MAPK and WNT signalling pathways

In addition to elucidating the connection between the transcriptomic changes resulting from the loss of CBX6 in ES cells and the subsequent differentiation phenotype, which could be described based on the top one-hundred up- and downregulated genes, we also aimed to understand what triggers cell differentiation upon CBX6 depletion. We thus thoroughly studied CBX6 genome-wide distribution, as well as changes in expression of CBX6-target genes, which we assumed are the leading changes that are occur in the cell following CBX6 depletion. GO analysis derived from the KEGG pathway database revealed that CBX6 target genes are enriched in genes involved in signalling pathways regulating pluripotency of stem cells, in particular, the WNT and MAPK signalling pathways.

Soluble factors, including WNT, signal stem cells to continue self-renewal. In standard ES cell culture conditions, low levels of autocrine WNT molecules lead to activation of the pathway, positively influencing the expression of a subset of pluripotency-associated genes. CBX6 localizes to the promoters of numerous genes involved in the WNT signaling pathway where it regulates

their expression. The transcriptional outcome after CBX6 depletion was not directional, with transcription levels of genes either not changed or up- or downregulated. Based on the CBX6-depletion–derived phenotype, one could expect to find an overexpression of genes that negatively influence the pathway, and a global repression of genes that enable WNT signal transduction. Nevertheless, we found transcriptional changes in different directions, the combination of which will ultimately determine the degree of activation of the pathway. For example, CBX6 targets many Frizzled (*Fzd*) family receptors (*Wnt* receptors): after Cbx6 depletion, *Fzd2* is upregulated and *Fzd4*, downregulated. Other Cbx6 target genes are downregulated in CBX6- depleted ES cells, like such as *Sfrp1* (soluble frizzled-related protein 1), *Smad3*, (smad family member 3) and *Frat1* (frequently rearranged in advanced T-cell lymphomas 1), among others. SFRP1 modulates the WNT signaling pathway through its direct interaction with WNTs, decreasing intracellular levels of β-Catenin[138]. In contrast, SMAD3 has been shown to favor β-Catenin translocation to the nucleus[139]. FRAT1 also positively regulates the WNT signaling pathway through GSK3 inhibition[140].

The MAPK signaling pathway is also altered in CBX6-depleted ES cells. FGF and its receptor tyrosine kinases play an important role in regulating pluripotency and lineage segregation in ES cells. Autocrine FGF-induced ERK1/2 signaling (also known as MAPK) occurs in ES cells and is needed to exit from self-renewal and initiate differentiation[50]. CBX6 also targets several genes implicated in the MAPK cascade signaling pathway, such as FGF factors, FGFR, MAPK, and DUSP (dual-specificity phosphatase). As described above for CBX6 target genes involved in the WNT

signaling pathway, CBX6 also influences MAPK-related genes at the transcriptional level. In CBX6-depleted ES cells, genes such as *Fgf5*, *Mapk12*, *Fgfr2*, and *Dusp4/5* are upregulated. DUSP phosphatases generate a negative feedback mechanism to reduce FGF signaling; in contrast, FGF5, MAPK12 and FGFR2 enhance it. Overstimulation of the FGF pathway would irreversibly direct ES cells towards differentiation, justifying why FGF signaling is so tightly regulated and modulated at multiple layers. Thus, both Wnt and MAPK signaling pathways are perturbed in CBX6-depleted ES cells. Fine-tuning the balance between multiple and opposing signals downstream of WNT and FGF receptors generate contrasting functional outcomes, either maintaining pluripotency or instructing lineage differentiation. In CBX6-depleted ES cells, alteration of these extrinsic pathways directly affects the expression of pluripotency genes, which in turn triggers differentiation.

The phenotype of CBX6-depleted ES cells is not only determined by the misbalance between WNT and FGF signaling pathways; rather, the fine-tuning might be further influenced by additional signals that contribute to the final outcome. In the “2i” culture regimen, wild-type ES cells can be kept permanently undifferentiated as pluripotency is not threatened by autocrine FGF stimulation, and WNT signaling is constantly active[52]. Under “2i” culture conditions, ES cells treated with chemical inhibitors of MAPK (PD0325901) and of GSK3 (CHIR99021) revealed that the *Cbx6*-depleted ES cells did not spontaneously differentiate, as they did when not treated with the inhibitors. In the “2i” culture regimen, ES cells can be kept permanently undifferentiated as pluripotency is not threatened by autocrine

FGF stimulation, and Wnt signaling is constantly active[52]. CBX6 depleted ES cells cultured in these conditions do not spontaneously differentiate. We hypothesize that this is due to perturbations of the WNT and MAPK signaling pathways following CBX6 depletion, which are then alleviated by using the PD0325901 and CHIR99021 inhibitors.

We have seen that ES cells depleted of either CBX6 or CBX7, which encode two similar proteins of the same family with genomic distribution that strongly overlaps, display extremely different phenotypes. CBX6 and CBX7 depletion in ES cells generate different consequences also at the level of gene expression. While there are a considerable number of genes whose expression is affected in both CBX6- and CBX7- depleted ES cells, there are also genes that are specifically up- or down-regulated in each of the two cell lines. Nonetheless, analyzing these gene lists did not highlight any biological category that would allow us to explain the differences observed for the ES cell phenotypes. CBX7-depleted ES cells additionally show a perturbation of factors involved in the WNT and MAPK signaling pathways[112], although these changes have no apparent affects on cell differentiation.

4. CBX6 displays binding preferences for histone H4 peptides and does not bind to H3K27me3 *in vitro*

The CBX family of proteins contain a conserved chromodomain, which has been extensively studied. It has been previously shown that most of these chromodomains recognize methylated histone

residues[120]; however, none of the histone modifications tested to date resulted positive for CBX6 chromodomain binding.

Here, we have shown that CBX6 does not bind to a specific single modification; rather, it displays affinity for weakly-charged H4 histone peptides—specifically, for unmodified and methylated H4 peptides. Our data suggest that binding is determined both by a slight sequence specificity and by charge density: an H4 scrambled peptide partially loses its interaction with CBX6, and positively-charged acetylated H4 peptides cannot bind to CBX6. Importantly, this is the first demonstration that the full-length CBX6 protein is unable to bind H3K27me3 peptides *in vitro.* To date, all CBX6 *in vitro* binding experiments had been performed with the isolated chromodomain[120].

Based on results from the peptide pull-down experiments, we cannot conclude if CBX6 has a any preference for mono-, di-, or tri-methylated H4K20, or tri-methylated H4K16, as compared to the respective unmethylated counterparts. A more quantitative approach, such as fluorescent polarization or isothermal titration calorimetry (ITC), should be implemented to calculate the relative affinity of CBX6 for these histone peptides. These modifications comprise diverse biological functions: H4K20me1/2 is involved in replication and DNA damage repair[141,142], and H4K20me3 has a role in heterochromatin compaction[143]. However, genome-wide CBX6 data showed that it is majorly associated with cPRC1 complex binding sites, and thereby with H3K27me3-enriched regions that are poorly decorated with H4K16me3 and H4K20me1/2/3. Although CBX6 *in vitro* binding with H3K27me3 has never been observed, it could still be that CBX6 interacts with unmodified H4 histone tails from nucleosomes containing H3K27me3, or with other (yet-unidentified) H4 histone

modifications, enriched in these regions. Our findings indicating the CBX6 occupies H3K27me3 regions, and that it is displaced from chromatin upon PRC2 depletion, is in apparent conflict with our *in vitro* peptide pulldown data; this issue remains to be clarified with further experiments.

Comparative analyses between CBX6 and CBX7 clearly show that their histone binding preferences are very different, although both share a very conserved chromodomain. Dissimilar binding capabilities between the two proteins could rely on subtle differences within the protein sequence of the aromatic cage pocket and the adjacent hydrophobic pocket. Milosevich and coworkers[122] showed that CBX6 hydrophobic pocket cannot bind the conserved histone alanine at the –2 position of the trimethyllysine site H3K27me3 (AR**K**S); rather, a hydrophobic residue in the position –2 of the methyllysine is key for CBX6 chromodomain binding[122]. Based on our peptide pulldowns of histone H4(11-39), we propose either H4R23 (RKVL**R**DNIQ) or H4K31 (QGIT**K**PAIR) as possible CBX6-binding sites. Using the H4R23 binding site would require an arginine residue to occupy the aromatic cage of CBX6. This may sound unusual for a chromodomain, but is not unprecedented for other reader-protein families. Using the H4K31 site would require an isoleucine residue to be tolerated at the –2 position. Methylation at both the H4R23 and H4K31 sites have been identified in proteomic PTM surveys (PhosphositePlus), but these have not been further studied to date. It would be very interesting to determine the CBX6 binding affinity to these modifications.

Our experiments performed with the mutated chromodomain are difficult to interpret. While in the *in vitro* experiments, the chromodomain seems to be indispensable for CBX6 binding, expression of CBX6^{AA} in CBX6-depleted ES cells did not rescue the CBX6 phenotype, suggesting that the chromodomain is indispensable for CBX6 function. It is possible that the results from the H4 pulldown are artifacts, or that binding occurred via a different part of CBX6 than its chromodomain. However, it could also be that the presence of the –2 hydrophobic pocket in the CBX6^{AA} mutant was enough to allow binding.

Several limitations of this study should be considered. The CBX6-GST recombinant protein production is extremely challenging, and the resulting purified protein is easily degraded, which could interfere in its binding to the histone peptides. Moreover, CBX6-GST protein precipitates if frozen and must be freshly generated for every experiment. We also need to consider that we are producing our proteins of interest in a bacterial system, in which protein folding as well as any post-transcriptional modifications could be affected. It would be interesting to use a baculovirus system for expressing fully-functional recombinant proteins, as post-transcriptional processing and folding of recombinant proteins produced in insect cells resembles more closely to mammalian processes. Finally, it should be noted that we used an *in vitro* recombinant assay devoid of any other factor (i.e., other proteins, RNA) that may be required for methyl-histone binding. For a more biological approach, we would like to perform an endogoneous CBX6 immunoprecipitation, to identify modified histone-binding substrates by mass spectrometry.

5. CBX6: a canonical or non-canonical complex subunit?

One of the objectives of this thesis was to characterize CBX6-associated proteins in ES cells, in order to better characterize the CBX6 function. Although CBX6 is generally accepted to be part of the cPRC1 complex, there was no experimental evidence demonstrating that this also holds true in ES cells. Thus, we demonstrated for the first time that CBX6 interacts with the cPRC1 complex in ES cells. Importantly, the endogenous CBX6 genome-wide distribution shows a high overlap with the cPRC1, supporting the validity of the MS data and demonstrating that this result is not an overexpression artefact. We believe that, based on the very low CBX6 protein levels in vivo, overexpression is indispensable to capture CBX6 partners. This would explain why CBX6 was never observed in previous PRC1 purifications[144,145].

It is also worth noting that MS analysis has several detection limitations, and that some proteins are easier to detect because they are more abundant or their biochemical properties are more suitable, influencing the final results. This could explain why PCGF2, although it is more abundant in the cell than CBX6[145], was not detected in our MS analysis but was detected in our coimmunoprecipitation experiments. Moreover, PCGF2 *in silico* trypsin digestion gives unreadable peptides (data not shown).

Unexpectedly, among the more confident candidates in our mass spectrometry analysis, we also found non-canonical interactors involved in the ncPRC1 complex, PRC1.6. It should be noted that we only detected these interactions under CBX6-overexpression conditions. Nonetheless, based on our results, we can conclude

that CBX6 interacts with members of both canonical and non-canonical complexes; however, we do not know if our results reflect the composition of a single unique complex, or whether we are dealing with several distinct CBX6-containing complexes. According to our results, our preliminary hypothesis is that at least two different complexes exist, as it has been demonstrated that PCGF subunits are mutually exclusive[93]. On the one hand, we speculate that CBX6, RING1B, PCGF2 and PHC1 could compose one of the CBX6-containing complexes. This postulation is based on the high degree of overlap between the genome-wide distributions of these subunits in ES cells. Thus, it will be worth the effort to understand why the cell would have a very similar but distinct complex to the cPRC1, composed by RING1B, CBX7, PCGF2 and PHC1; this question is further addressed in the following section. On the other hand, a second, distinct complex could comprise CBX6, RING1B, L3MBTL2, PCGF6, and MGA. As mentioned, L3MBTL2, RING1B, PCGF6, and MGA have been implicated as components of the repressive E2F6 complex in somatic cells, which contains Polycomb group proteins also linked to PRC1. However, despite containing three of the main subunits of the PRC1.6 complex, this should be considered as a new complex, given the lack of other proteins of the established E2F6 complex, such as RYBP.

The lack of overlap between our CBX6 ChIPseq data and the published PCGF6 genome-wide distribution, of only three shared genes, suggests that CBX6 and PCGF6 do not occupy similar regions. PCGF6 is a positive regulator of transcription and colocalizes with active chromatin marks, such as H3K4me3[135]. Similar to CBX6, it has recently been shown that PCGF6 is required to maintain ES cell self-renewal[135]. According to Chao-

Sun and colleagues, PCGF6-depleted ES cells spontaneously differentiate and downregulate pluripotency genes like *Oct4*, *Sox2*, and *Nanog*[135]. In contrast, Zdzieblo et al. show that PCGF6-depleted ES cells also downregulate pluripotency genes but without spontaneous differentiation[146]. Despite obtaining a similar phenotype to that observed from Chao-Sun's research, changes in gene expression upon depletion of CBX6 or PCGF6 are unrelated, as PCGF6 is involved in the repression of mesoderm-specific genes[135]. Using MS analysis, they did not detect CBX6 as co-immunoprecipitating with PCGF6 in ES cells[135]; incomprehensibly, also no E2F6 complex members were found. It is worth noting that the aforementioned experiments for PCGF6 were performed in CCE ES cells, which come from a different genetic background than the E14Tg2A mice lines used to obtained the ES cells we used. To clarify whether this is the basis for the inconsistencies between our and the Chao-Sun results, we plan to repeat their experiments in our cellular system.

Intriguingly, 29% of PRC1/2-bound genes are also bound by L3MBTL2, while around 20% of CBX6 target genes co-occupy with L3MBTL2. L3MBTL2 has been implicated in transcriptional repression and chromatin compaction[134]. L3MBTL2-depleted ES cells do not spontaneously differentiate, and there is no impairment in the expression of the pluripotency genes. Moreover, L3MBTL2-depleted ES cells have a severe proliferation defect. Genome-wide, L3MBTL2 target genes co-localize with H3K9me2, low acetylated histones, and H2AK119ub (although ubiquitination is not dependent on L3MBTL2). In general, L3MBTL2 predominantly binds and regulates genes in ES cells that are not bound by PRC1/PRC2. However, genes co-bound by Cbx6 and

L3MBTL2 are also co-occupied by PRC1/PRC2 members, making it difficult to assess if Cbx6 is binding to these target genes through the hypothetical complex with L3MBTL2, or within a cPRC1 complex.

Very few data exist about MGA in ES cells. MGA, another potential candidate for a CBX6-containing complex, is the least studied of the MAX-network transcription factors but is know to bind to and regulate transcriptional targets through heterodimerization with MAX. In ES cells, *Mga* knockdown leads to ES cell differentiation, suggesting that MGA plays a role in the maintenance of pluripotency through its interaction with OCT4[147].

It is critical to next determine the exact composition of the CBX6 complex/es. To do that, we plan to perform a gradient centrifugation experiment and to verify which components co-sediment with CBX6 by Western blot analysis. So far, because of the strong overlap of CBX6 with cPRC1 subunits, it seems more plausible that the CBX6 main complex is cPRC1; however, determining whether CBX6 exists within a non-canonical complex could help us to understand the different effects between deleting CBX6 and CBX7 from ES cells. Nonetheless, CBX6 interactions with non-canonical subunits could be due to an overexpression artifact, because of the low degree of overlap between the genome-wide distributions of L3MBTL2 and PCGF6 subunits.

6. CBX6: repressor or activator?

A comparative gene expression profile between control and CBX6-depleted ES cells reveals an altered expression of 1,486

genes, of which 703 are up-regulated and 783, down-regulated. Importantly, a high proportion of these genes (specifically, 189 and 206 genes, respectively) are CBX6 direct target genes. Because Polycomb complexes are widely accepted to be transcriptional repressors, one would expect to find a higher percentage of up-regulated target genes than down-regulated genes following depletion of any Polycomb subunit that impairs Polycomb function. However, this was not the case for CBX6-depleted ES cells. That depletion of Polycomb members activates gene transcription is not exclusive for CBX6 depletion. In fact, several examples show that binding of Polycomb proteins is not always associated with gene repression. For instance, in a previous report from our laboratory, we observed that some CBX7 direct target genes displayed a decreased expression in the absence of CBX7[112]. More recently, we have also reported that in PCGF2-depleted ES cells, a small number of PCGF2 target genes are downregulated[112]. However, it should be noted that in both cases, most of the target genes were up-regulated, as expected. Activation of gene expression by CBX6 could be explained through its interaction with a non-canonical subunit that enhances transcription, such as PCGF6, which is a positive regulator of transcription.

Curiously, analysis of representative CBX6 target genes in CBX6-depleted ES cells showed that changes in transcription do not strictly correlate with the epigenetic status of the analyzed promoters. Concretely, neither RING1B nor SUZ12 recruitment was affected in any of the promoter cohorts (up- and down-regulated genes), whereas CBX7 recruitment was slightly increased, regardless of the transcriptional output of CBX6 target

genes. In contrast, H2AK119ub was slightly decreased in both sets of genes. From these experiments, we observed that PRC1 and PRC2 stability was not highly compromised upon CBX6 depletion. This is completely different to what we observed for CBX7-depleted ES cells, in which RING1B recruitment is dramatically reduced, H2AK119ub diminished, and PRC2 occupancy compromised[112]. Moreover, if CBX6 forms part of a canonical PRC1 complex, we suspect that the CBX6-containing cPRC1 is less abundant than cPRC1-CBX7, as CBX6 depletion did not cause displacement of RING1B from chromatin. Moreover, we have never detected the interaction between RING1B and CBX6 at endogenous levels, supporting the idea that cPRC1-CBX6 complex is in low abundance in the cell.

Additionally, we have evidence that soluble CBX7 in the nucleus tries to compensate for CBX6 loss. We hypothesize that decreases in H2AK119ub could be due to the low efficiency of cPRC1-CBX7 for ubiquitinating histone H2A.

Overall, we have elucidated that transcription of CBX6 target genes is differently affected upon CBX6 depletion. Several questions remain to be resolved; in particular, we need to better understand the observed dual functionality of CBX6 in gene repression and activation. We believe that the answer to this requires a comprehensive determination of the different CBX6-containing complexes. Also, as we have seen that both active and repressed CBX6 target genes display a similar epigenetic landscape following CBX6 depletion, we are further analyzing additional marks to help us understand differences in gene expression (*i.e.* H3K4me3, H3K36me3, and RNAPolII).

CONCLUSIONS

We can draw the following main conclusions from the resulted presented in this thesis:

1. CBX6 is required to maintain ES cell pluripotency. CBX6-depleted ES cells lose their pluripotency capacity and spontaneously differentiate into an ectoderm fate-

2. CBX6 depletion impairs ES cell self-renewing capability.

3. CBX6 is indispensable for precise ES cell differentiation.

4. CBX6 effects are specific. Its chromodomain and PcR box are essential for mediating the Cbx6-specific function.

5. CBX6 controls the pluripotency network through the transcriptional regulation of key components of the MAPK and WNT signalling pathways.

6. CBX6 depletion causes both repression and activation of transcription CBX6 target genes, suggesting that CBX6 has a dual function in transcription.

7. CBX6 displays *in vitro* binding preferences for histone H4 unmodified and methylated peptides but not for H3K27me3.

8. CBX6 interacts with canonical and non-canonical PRC1 proteins.

9. CBX6 genome-wide distribution completely overlaps with that of canonical PRC1.

10. CBX6 is not required for either canonical PRC1 or PRC2 distribution.

MATERIALS & METHODS

1. Cell culture and differentiation

1.1 ES cell culture and embryoid body differentiation

Wild-type (E14Tg2A) ES cells were cultured feeder-free in plates coated with 0.1% gelatin. Coating was achieved by covering the plates with gelatin 15 min at 37°C. After removing the remaining gelatin, ES cells were cultured with Glasgow minimum essential medium (Sigma) supplemented with β-mercaptoethanol, sodium pyruvate, penicillin-streptomycin, non-essential amino acids, GlutaMAX, 20% fetal bovine serum (Hyclone), and leukemia inhibitory factor (LIF).

For 2i/LIF cultured cells, medium was composed of DMEM/F12:Neurobasal 1:1, supplemented with N2, B27, β-mercaptoethanol, glutaMax, non-essential amino acids , penicillin-streptomycin, PD0325901 and CHIR99021 (Sigma), and LIF.

ES cells were differentiated into EBs using the hanging-drop method (with 1,000 cells/drop) in Glasgow minimum essential medium (Sigma) with 10% serum and no LIF.

1.2 Alkaline phosphatase staining

ES cells (1×10^3) were plated in a 6-well plate for 5 days. Alkaline phosphatase assay was performed with the alkaline phosphatase detection kit (Millipore) following the manufacturer's instructions.

2. Cell transfection and infection

2. 1. Calcium phosphate transfection

HEK-293T cells (2×10^6) were plated in a p10 plate. The following day, the calcium phosphate-DNA precipitates were prepared by pooling together the plasmid (different amounts of plasmid depending on the plasmid, see Lentiviral infection method) in 0.25 m $CaCl_2$. While vortexing, 1× volume of the calcium phosphate-DNA solution was added dropwise to an equal volume of HBS 2× (HEPES-buffered saline solution pH 7.05, with 0.28 NaCl, 0.05 M HEPES, 1.5 mM Na_2HPO_4) at room temperature. The solution was incubated for 15 min at room temperature and then added to the HEK-293T cells for lentivirus production.

2. 2. Lentivirus production and infection

Lentivirus were produced by transfecting HEK-293T packaging cells with 5 µg of pCMV-VSV-G, 6 µg of pCMVDR-8.91, and 7 µg of the pLKO-shRNA (Sigma) plasmid (either pLKO-shCTL or pLKO-shCbx6), using the calcium phosphate transfection method. Cells were incubated with the transfection mix for not more than 16 h, after which the medium was replaced by ES cell (LIF-free) fresh medium. Lentiviral particles were collected 48 h later and filtered using a 0.45 µm filter. The day of infection, 2×10^5 target cells were plated in a 6-well plate. Filtered medium (2 ml) containing the lentiviral particles was used to culture the cells in the presence of LIF (1:500), and Polybrene (1:1000) was added

to the culture medium. The following morning, medium was replaced with fresh medium for selection. The infected cells were selected using the appropriate antibiotic (2 μg/ml puromycin, 50 μg/ml hygromycin, or 50 μg/ml for G418) for 3 days.

3. Protein analysis

3.1. Protein extracts preparation

Whole cell extracts for Western blot analysis were prepared in lysis buffer IP300 (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 10% glycerol, 0.2% NP-40). Lysates were incubated for 5 min on ice and then sonicated 5 cycles (30 sec ON/30 sec OFF) in a Bioruptor (Diagenode). Cell extracts were centrifuged for 25 min at maximum speed at 4°C. Protein concentration was quantified by Bradford assay (Bio-Rad) according to manufacturer instructions. 40 μg of protein were loaded into SDS-PAGE gel for further analysis.

3.2. Quantification of protein concentration by Bradford assay

The Bradford solution (Bio-Rad) was diluted to 1x, and a standard curve was prepared with bovine serum albumin (BSA). 1 μl of the cell extract lysate was mixed with 1 ml of the freshly-prepared Bradford solution and incubated for 5 min at room temperature. Absorbance was measured in a UV/VIS photo-spectrometer at a wavelength of 595 nm. The protein concentration was calculated based on the standard curve.

3.3. Western blot

Whole cell extracts for Western blot analysis were prepared in lysis buffer IP300 (buffer). Lysates were incubated for 5 min on ice and then sonicated for 5 cycles (30 sec ON/30 sec OFF) in a Bioruptor (Diagenode). Cell extracts were centrifuged for 25 min at maximum speed at 4°C. Protein extracts (supernatant fractions) were quantified by Bradford assay (Bio-Rad). Protein (usually around 40 µg) was diluted with 5x Laemli buffer and heated for 7 min at 100°C. Samples were analysed by SDS-PAGE using acrylamide gels in running buffer (25 mM Tris-base, 200 mM glycine, 0.1% w/v SDS) at 100 V. Proteins were transferred onto nitrocellulose membranes at 300 mA for 70 min at 4°C in Transfer buffer (25 mM Tris-HCl pH 8.3, 200 mM glycine, 20% v/v methanol). Protein transfer was checked by Ponceau S (Sigma) staining. Transferred membranes were blocked with 5% w/v milk in TBS-Tween (10 mM Tris-HCl pH 7.5, 100 mM NaCl, and 0.1% Tween-20) for 30 min with rotation at room temperature. Blocked membranes were incubated overnight with the primary antibody (in with 5% w/v milk in TBS-Tween) at 4°C with rotation (see antibody table). The next day, membranes were washed twice for 5 min with TBS-Tween followed by incubation with the secondary antibody conjugated to the horseradish peroxidase (1:5000. Dako), diluted in TBS-Tween, for 1 h at room temperature. After two washes of 15 min with TBS-Tween (at room temperature), proteins were detected by a enhanced chemilluminescence reagent (Pierce ECL Western Blotting Substrate, Thermo Scientific).

3.4. Cellular fractionation of ES cells

Trypsinized cells (from a p150mm plate with approximately 30 million cells) were pelleted, resuspended in 10× v/v Buffer A, and incubated for 10 min on ice. One volume of SDS 2× was added, and samples were boiled at 98°C for 10 min (to give the Total extract fraction). Cells were centrifuged for 5 min at 1,300 g (4°C). Supernatant was clarified by centrifugation at 15,000 g × 15 min at 4°C. One volume of SDS 2× was added, and samples were boiled at 98°C for 10 min (to give the cytoplasmic fraction). Pellet were washed with 5 vol. Buffer A, centrifuged 1,300g × 5 min at 4°C, and resuspended in 1 volume buffer A. One volume of SDS 2× was added, and samples were boiled 98°C 10 min (to give the nuclear fraction). 10 volumes of Buffer B were added, and samples were vortexed, incubated 30 min on ice, and centrifuged at 1700 g × 5 min at 4°C. Supernatants were clarified by centrifugation at 15,000g × 15 min at 4°C. One volume of SDS 2× was added, and samples were boiled at 98°C for 10 min (to give the nucleoplasmic fraction). Pellets were washed with 5 vol. Buffer B and centrifuged at 1,700 g × 5 min at 4°C. Pellets were resuspended in 10 vol. B SDS 1×, boiled at 98°C for 10 min (to give the chromatin fraction), and sonicated until clarification.

BUFFER A	**BUFFER B**
HEPES [pH 7.9] 10 mM	EDTA 3 mM
KCl 10 mM	EGTA 0.2 mM
$MgCl_2$ 1.5 mM	DTT 1 mM
Sucrose 0.34 M,	bGlycerolph. 10 mM

Glycerol 10%	SOV 1 mM
DTT 1 mM	Protease Inhibitor
bGlycerolph.	
SOV 1mM	
Protease Inhibitor	
TritonX100. 0.1%	

4. Histone extraction protocol

Approximately 5×10^6 cells were used to extract histones for Western blot analysis. Cells were collected, washed once with PBS 1×, and resuspended by vortexing 3× with 500 µl of lysis buffer (10 mM Tris pH 6.5, 50 mM sodium bisulfite, 1% Triton X-100, 10 mM $MgCl_2$, 8.6% sucrose, and 10 mM sodium butyrate). Cell pellets were then washed once with 500 µl of washing buffer (10 mM Tris pH 7.4, 13 mM EDTA) and centrifuged at full speed for 15 seconds. The cell pellet was mixed with 100 µl of 0.4 M H_2SO_4 and incubated on ice for 1 h. After five min of centrifugation at full speed, the supernatant (containing histones) was precipitated with ice-cold acetone (80%) overnight at –20°C. After 10 min of centrifugation at full speed, pellets were air dried and resuspended in 30 µl of water. Around 2–3 µg of histones were loaded into a SDS-PAGE gel for further analysis.

5. Pull-down Assay of Biotin-labeled Histone peptides

5.1. Production of recombinant proteins

BL21 bacteria were transformed with the plasmid of interest (GST, Cbx6-GST, or Cbx7-GST). A starter culture was made with 1 ml

and left overnight at 37°C. The following morning, the starter culture was diluted into 100 ml of LB until the optical density was around 0.5–0.7, at which point it was induced with 0.2 mM IPTG overnight at 18°C.

Bacteria were centrifuged at 6,000 rpm for 15 min at 4°C, resuspended in 10 ml of lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.05% NP40) with protein inhibitors and 100 μl of lysozyme, and incubated on ice for 30 min. Lysates were sonicated (with a Branson sonicator) at 18% of intensity, using 3 rounds of 30 min with cooling down on ice in between. After sonication, samples were centrifuged at 12,000 rpm for 25 min at 4°C. Supernatants (containing the protein extract) were incubated with 125 μl of glutathione sepharose beads (pre-washed with lysis buffer 3×) overnight at 4°C.

The Samples were centrifuged the next day at 800 g for 3 min, and the beads were first washed with 10 ml lysis buffer for 5 min with rotation at 4°C, washed with 10 ml elution buffer (50 mM Tris pH 8.0) for 5 min with rotation at 4°C, and finally washed with 1 ml elution buffer for 5 min with rotation at 4°C. GST-tagged proteins were eluted with 300 μl elution buffer + glutathione 10 mg/ml for 4 h at 4°C on rotation.

5.2. Peptide pulldown

About 1 μg of the different biotinylated histone peptides (produced in the Laboratory of Proteomics & Protein Chemistry at Pompeu Fabra University) were added to 200 μl of binding buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.05% NP-40) containing 1 μg of each GST-tagged recombinant protein; recombinant protein

without histone peptide was used as a negative control (for each pulldown, 10% of input was taken). Tubes were rotated at 4°C overnight.

The following morning, streptavidin sepharose beads (Amersham) were washed 3× with 5–10 ml of cold binding buffer for 5 min with rotation at 4°C, with each wash followed by centrifugation at 2,000 rpm for 2 min. Washed streptavidin beads (15–20 μl) were added to each tube and rotated at 4°C for 1 h. Beads were washed 3× with 1 ml of binding buffer at 5 min on rotation at 4°C, with each wash followed by centrifugation at 2,000 rpm for 2 min. Beads were resuspended in 60 μl of Laemli buffer 2× (Roth) and boiled for 15 min at 100°C for elution, and 30 μl was used for SDS/PAGE analysis.

6. RNA extraction, cDNA synthesis, and gene expression analysis

RNA was extracted using the RNeasy mini kit (Qiagen) following manufacturer's instructions. cDNA was synthesized by reverse transcription using a cDNA synthesis kit (Quanta-Bioscience).

RNA from samples to be sequenced was extracted using the miRNeasy Mini kit.

Real-time PCR reactions were performed using SYBR Green I PCR Master Mix (Roche) and the Roche LightCycler 480. Expression was normalized to the housekeeping gene *Rpo*.

Paired-end RNA-sequencing was performed using 1 μg RNA and using two samples per lane to achieve maximum sequencing depth. The CRG Genomics Unit performed the quality control and

library preparation. Libraries were sequenced using Illumina HiSeq2000 sequencer. Genes with a fold change of at least 1.5 were considered to be differentially expressed.

7. Chromatin immunoprecipitation (ChIP)

Two 15-cm plates for each cell line to be tested were prepared at 70–80% confluency. Cells were trypsinized and crosslinked in 1% formaldehyde for 10 min at room temperature in a shaker. To stop the fixation reaction, 0.125 m glycine was added to the existing culture media and incubated for 5 min. Sample pellets were then washed twice with PBS 1× at room temperature. After aspirating PBS, crosslinked pellets were resuspended in 1.3 ml ice-cold IP buffer (1× volume SDS Buffer, 0.5 volume Triton dilution buffer) with proteinase inhibitors in a polystyrene conical tube. Samples were sonicated for 12 min (30 sec ON/30 sec OFF) with a Bioruptor (Diagenode) at maximum output. After sonication, samples were centrifuged at 4°C at maximum speed for 20 min. To check chromatin size, 20 μl supernatant was mixed with 80 μl of PBS 1× and de-crosslinked for 3 h at 65°C in a shaker (1000 rpm), followed by PCR purification kit (Qiagen). DNA was eluted in 30 μl of water and quantified by Nanodrop. Around 800 ng were analyzed in a 1% agarose gel to assure that chromatin was between 200–500 bp. ChIP reactions with 40 μg of chromatin and 5 μg of antibody, in a final volume of 500 μl, were incubated overnight at 4°C on rotation. The following day, 30 μl washed agarose beads were added to the ChIP reactions and incubated for 2 h at 4°C. After incubation, beads were washed (with 5 min

with rotation at 4°C) 3× with 1 ml of low salt buffer and 1× with 1 ml of high salt buffer.

After washes, all supernatant was removed with a syringe, and 110 μl of freshly-prepared elution buffer (1% SDS, 100 mM NaHCO3) was added per ChIP reaction (including 1% of every input). De-crosslinking was performed at 65°C for 3 h in a shaker (1000 rpm). DNA was purified using a PCR purification kit and eluted in 100 μl of water (two consecutive elutions of 50 μl), and 2 μl were used for ChIP-qPCR analysis.

SDS buffer

Reagent	[Final]
NaCl	100 mM
Tris-HCl pH 8.1	50 mM
EDTA pH 8	5 mM
SDS	2 %

Add fresh: protease inhibitors

Triton Dilution Buffer

Reagent	[Final]
NaCl	100 mM
Tris-HCl pH 8.6	100 mM
EDTA pH 8	5 mM
Triton X-100	5 %

Add fresh: protease inhibitors

ChIP Wash Low Salt

Reagent	[Final]
NaCl	140 mM
HEPES pH 7.5	50 mM
Triton X-100	1 %

Add fresh: protease inhibitors

ChIP Wash High Salt

Reagent	[Final]
NaCl	500 mM
HEPES pH 7.5	50 mM
Triton X-100	1 %

Add fresh: Protease inhibitors

For histone modifications, minor modifications were introduced. Cells were crosslinked and chromatin was prepared as described above. Crosslinked cells were resuspended in 1 ml of swelling buffer and incubated on ice for 10 min. Swollen cells were homogenized with a Douncer with an A pestle 20 times and centrifuged for 7 min at 2000 rpm at 4°C. Nuclei were resuspended in 1 ml sonication buffer followed by sonication in a Bioruptor for 10 min (30 sec ON/10 sec OFF), and supernatant was collected after 15 min of centrifugation at maximum speed at 4°C. 200 μg of protein was used per ChIP reaction in rotation overnight at 4°C. Thereafter, 30 μl of agarose beads was added, and samples were incubated for 2 h at 4°C. Immunoprecipitated material was washed twice in 1 ml of sonication buffer, once with 1 ml washing buffer, and once in TE buffer. Reverse crosslinking was done for 3 h at 65°C in a shaker at 1000 rpm in 100 μl of

elution buffer (1% SDS, 40 mM NaHCO3). DNA was purified following DNA purification kit. DNA was eluted with 100 μl of water (two consecutive elutions of 50 μl). 2 μl were used for ChIP-qPCR analysis.

Swelling Buffer

Reagent	[Final]
MgCl2	1.5 mM
HEPES pH 7.5	25 mM
KCl	10 mM
NP-40	0.1%
DTT	1 mM

Add fresh: protease inhibitors

Sonication Buffer

Reagent	[Final]
NaCl	140 mM
HEPES pH 7.5	50 mM
EDTA	1 mM
Triton X-100	1%
Sodium deoxycholate	0.1%
SDS	0.1%

Add fresh: protease inhibitors

Washing Buffer

Reagent	[Final]

Tris pH 8	2 mM
EDTA	0.02 mM
NP-40	0.1%
LiCl	50 mM
Sodium deoxycholate	0.1%

Add fresh: protease inhibitors

ChIP experiments on the Cbx6-3×HA cell line were performed using the ChIP-IT High Sensitivity Kit from Active Motif (53040) according to the manufacturer's instructions.

ChIP-seq for all samples were performed using 2–10 ng (depending on the experiment) of precipitated ChIP DNA followed by sequencing library preparation, quality control, and quantification (by Qubit). Libraries were sequenced in an Illumina Hiseq2000 sequencer.

8. Cbx6 cloning and rescue experiments

Cbx6 cDNA was amplified using specific primers from mES cell total cDNA. PCR product was cloned into the pCR8-TOPO-GW vector and sequenced to verify (TOPO-Cbx6). This plasmid was used as a backbone to amplify Cbx6 for further cloning.
Cbx6-3×FLAG construct was generated by restriction enzyme digestion. Cbx6 was amplified by PCR from the TOPO-Cbx6 plasmid using specific primers containing BglII (forward primer)

and HindIII (reverse primer) and cloned into the pCBA-FLAG×3 vector previously digested with the same enzymes (Aloia 2010). Cbx6-3×HA construct was generated by substituting the 3×FLAG construct by a 3×HA construct, obtained by PCR amplification from the pKH3 plasmid (Addgene).

Cbx6 cDNA resistant to shRNA#52 (Cbx6R) was produced by the insertion of a silent mutation in the 10th nucleotide of the shRNA#52 recognition site using the QuickChange Site-Directed Mutagenesis kit (Stratagene) following the manufacturer´s instructions. Cbx6-FLAG mutant for the chromodomain (Cbx6^{AA}) was generated by mutating Cbx6R in two residues: amino acids W33 and W36 (tryptophan) were switched to glycine. Cbx6-FLAG mutant for the PcR box (Cbx6$^{\Delta PcR}$) was generated by depleting the PcR box domain as described[148].

The rescue experiment was performed by overexpressing the Cbx6-FLAG versions first (Cbx6R, Cbx6^{AA}, and Cbx6$^{\Delta PcR}$) and infecting with pLKO-shCTL or pLHO-shCbx6 once the cell lines were established.

9. Proliferation curves

100,000 cells for each condition were plated in a 6-well plate. Every two days, cells were trypsinized and counted, and 100,000 cells were re-plated, for 5 passages.

10. Cbx6 gene editing

10.1. Crispr/Cas9 vector construction

Three sgRNAs targeting the Cbx6 locus were designed using the online software: http://crispr.mit.edu.
Primers coding for the sgRNAs were annealed and assembled with a pX459 (Puromycin selection) and a pX458 (GFP selection) vectors (Addgene) using the method described by Zhang at the Broad Institute of MIT. Targeting efficiencies were calculated through T7 endonuclease surveyor assay.

10.2. Donor vector construction

Left and right Cbx6 homology arms were generated by PCR using specific primers and cloned by Gibson technology into the HDR donor vector. To create the 3×HA tag construct, two oligonucleotides to amplify the tag were generated that contained an overlapping sequence specific to the vector, in order to insert the fragment into the the HDR vector using a Gibson reaction.

10.3. Stable cell line generation

sgRNAs were cloned into the diferent vectors (pX458 and pX459, 6ug) and the HDR linearized vector (3 μg) were co-transfected in 3×10^6 mES cells by nucleofection (Nucleofection Amaxa kit) and incubated for 24 h.
Cells transfected with the pX459 vector were selected for 48 h with puromycin (1 μg/ml), whereas cells transfected with pX458 were checked for GFP efficiency using a fluorescence

microscope.

Single-cell sorted by FACS using GFP fluorescence or size were plated into two 96-well plates (96 clones per condition). Once clones were let to grow, genomic DNA was extracted from each clone, and a PCR was performed to check if the tag had been successfully inserted. Positive clones were sequenced.

11. Protein immunoprecipitation (IP)

11.1. IP followed by Western blot analysis

ES cells expressing FLAG tagged constructs were lysed in IP300 buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 10% glycerol, 0.2% NP-40) supplemented with protease and phosphatase inhibitors. Cells were sonicated 5 cycles (30 sec ON/30 sec OFF) with a Bioruptor sonicator (Diagenode) followed by full speed centrifugation. 1 mg of protein was incubated with 30 μl of prewashed FLAG M2 (SIGMA) beads (with IP300) and incubated 1 h (depending on the IP) on a rotating wheel at 4°C. Samples were washed three times with IP300 buffer. Elution was performed by incubating the dried beads with 60 μl of Laemli buffer 2× (Roth Karlsruhe) at 100°C for 15 min. Alternatively, elution was performed by incubating the beads with FLAG peptide at 0.2 μg/ml in PBS. For endogenous IPs, 1 mg of protein was incubated with the antibody 1 h, followed by incubation of A or G sepharose beads for 2 h at 4°C. Elution was performed with Laemli buffer 2× (Roth Karlsruhe) at 100°C for 15 min.

11.2. IP followed by mass spectrometry analysis

11.2.1. Protein extract isolation and FLAG-affinity purification

Cryomilled material from empty control and Cbx6(3×)FLAG cells was produced as described Domanski et al[149]. 900 mg of cell powder were used for protein extraction. Experiments were performed in triplicates for both conditions (Empty control cell line and Cbx6-(3×)FLAG cell line). Extraction was performed with 20 mM HEPES pH 7.4, 400 mM NaCl, 0.5% Triton X-100 (v/v). Short sonication was performed to improve protein extraction from chromatin (25 amplitude × 2 seconds). After sonication, samples were centrifuged 20 k rcf, 4°C for 10 min. Supernatants were combined with 27 μl of anti-FLAG beads for 30 min at 4°C. Beads were eluted with 20 μl of 1× LDS at 70°C for 5 min with mixing.

11.2.2 Sample preparation for mass spectrometry

Gel plugs were cut into small cubes for processing. Samples were distained with ACN buffer. After dehydration, samples were trypsinized in-gel (12.5 μg/ml in 50 mM ammonium bicarbonate), and trifluoroacetic acid (0.5% w/v, final concentration) was then added to stop the digestion. 50 μl of 0.1% trifluoroacetic acid was added to each tube to extract the gel pieces. Samples were desalted using RPC tips (Variant A57003100), and peptides were eluted twice in 100 μl of 40% (v/v) ACN, 0.5% HAc (first eluation, E1) and in 100 μl of 80% (v/v) ACN, 0.5% HAc (second elution, E2). E1 and E2 were combined and snap-frozen in liquid nitrogen.

Dried peptide samples were resuspended in 10 µL 5% methanol, 0.2% formic acid. Mass spectra were recorded on a Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific).

11.2.3. Data analysis

Database searching and label-free quantitation were performed by MaxQuant 1.5.2.8 using the UP000000589 mouse database. The match between runs feature was disabled and intensities were based on maximum peak height. The 'proteingroups.txt' file was uploaded to Perseus 1.5.3.0 and protein identifications from the decoy database were removed. LFQ intensities were logarithmized. Control experiments were grouped together, as were Cbx6 experiments. Proteins were filtered according to the constraint that at least one group (Cbx6 or control) contain at least 3 valid values. Missing values were imputed from a normal distribution. A two sample Student's t-test was performed with an arbitrary minimum fold change of 2 required for significance and a permutation-based FDR = 0.05 used for truncation.

12. Bioinformatic analysis

The ChIP-seq samples were mapped against the mm9 mouse genome assembly using bowtie with the option -m 1 to discard those reads that could not be uniquely mapped in just one region. MACS (Zhang et al., 2008) was run with the default parameters but adjusting the shiftsize to 75 bp to perform the peak calling and each set of target genes was retrieved by matching those ChIPseq peaks in the region from 2.5 kb upstream of the TSS until the end of the transcripts as annotated in RefSeq.

The plots showing the distribution of ChIP-seq reads 5 kb around the TSS of around each target gene set were generated by counting the number of reads on this regions for each gene (according to RefSeq) and then by averaging this value for the total number of mapped reads of each sample and the number of targets of the gene set.

For the metagene plots, graphical distribution of normalized count of reads between the TSS and TES of target genes for each ChIP-seq was generated by calculating the weighted number of reads on each position from 5 kb upstream of the TSS to 5 kb downstream of the TES of these genes according to RefSeq, normalizing these values within each gene by its length to build the plot of uniform size representing an idealized gene.

The heatmaps displaying the density of ChIP-seq reads 5 kb around the TSS of each target gene set were generated by counting the number of reads on this region for each individual gene and normalizing this value with the total number of mapped reads of the sample. Genes on each ChIP heat map are ranked by the logarithm of the averaged number of reads on the same genomic region.

Distribution of ChIP-seq peaks across different regions of the genome was calculated by counting the number of peaks fitting on each class of region: DISTAL region is the region within 2.5 Kbp and 0.5 Kbp upstream of the TSS; PROXIMAL region is the region within 0.5 Kbp and the TSS. UTR is UnTRanslated sequence; CDS is the protein CoDing sequence; INTRONS are intronic regions; INTERGENIC is the rest of the genome;

and TSS is the transcription start site.

Gene ontology (GO) and other term enrichment analyses were done using DAVID and Enrichr web-base tools (Chen et al., 2013; Huang da et al., 2009).

RNAseq samples were mapped against the mm9 mouse genome assembly using TopHat (Langmead B et al GBiology) with the option -g 1 to discard those reads that could not be uniquely mapped in just one region. CuffLinks was run to quantify the expression in FPKMs of each annotated transcript in RefSeq. Genes showing one or more FPKMs are considered to be expressed. Up- and downregulated gene lists in control versus knockdown samples were generated applying a fold-change threshold of two.

13. Primers Used

ChIP-qPCR

Gene	Primer 5'-3'	Gene	Primer 5'-3'
EOMES	GGCGCAGGGAATCTTAACTG	Nestin	CTCGGGAGAGTCGCTTAGAG
	AAGACCCAACATGAGCCTGA		GCTTGGTTTTACCAGGGACA
Esrrb	TTCTCCTCCAACTGGGAATG	Pax6	ACCAAGGACAGGCAGAGAGA
	CTAGGTCCCTGCCACTTCAG		AAAGGGATGAGAGCCAGGAT
Fgf5	TGTACTGCAGAGTGGGCATC	Pkdcc	GCGCTTAAAGCAGTGGACTT
	ACGAAACCCTACCGGACTCT		TGGACGCCCTGAGACTTATT
FoxA2	CCCAAAGAGAGCCGGAAAAG	Pou3f1	CCTGGGGTCCTTCTAACTCC
	CCTGGGATAGCAACGGAGAT		GGAGGAGAGGGGAAGAGAAA
Gata4	TCTCCAGCACCCATCAGTTT	Prdm14	GTGTCACCCGACTGAAAGGT
	TTCTGGGAAACTGGAGCTGG		GTTCGTTTTGTTGGGCTGAT
Gata6	TCTTCCTGCTCTCCCCTTTG	Smo	CTTGCACCCGGAATAGTTGG
	CACAAACCGCCTTACAACCA		CCCATGGACTGTTGCTTCAC
Gsc	GCCAGGTGAGTAAAGCAAGC	Sox3	GTCCTTCTTGAGCAGCGTCT
	GCCTGGGCTACAACAGCTAC		CACAACTCCGAGATCAGCAA
JAK2	GTGGCTCTTTGGGACCCTAT	Tbx3	GGCTGACTGTTGACGTTTGA
	GAGACTTTCCCCTTCCTGCT		CGGTCTCCTTCAATGGAAAA
Mmp2	CGATGTCGCCCCTAAAACAG	VEGFa	CTTCCTCCCCTCTCTTCTCG
	CTGGCCAAGAAAGAGACCCT		GGGATTGCACGGAAACTTTT
Mmp25	GCAAAGAGGTCCCCTACTCC		
	GACGTTGGAAGATTGGGCTA		

Table MM.1. Mouse primers used for ChIP analysis. The sequences of the different mouse primers used in this study are listed. The first primer is forward and the second reverse. All of them are shown in 5' to 3' direction.

mRNA-qPCR

Gene	Primer 5'-3'	Gene	Primer 5'-3'
Cbx6	GCCGAATCCATCATTAAACG	Nr0b1	TCCAGGCCATCAAGAGTTTC
	TTGGGTTTAGGTCCCCTCTT		ATCTGCTGGGTTCTCCACTG
Cbx7	AGCCTCGGGGTATAGGAAGA	Oct4	GAGGAGTCCCAGGACATGAA
	CGGTGATGTCAGTCACGGTA		AGATGGTGGTCTGGCTGAAC
Cdx2	TCTCCGAGAGGCAGGTTAAA	Pax6	ACCGCCCTTGGTTAAAGTCT
	GCAAGGAGGTCACAGGACTC		GTTGGTGTGTTCCCTGTCCT
Eomes	GGCAAAGCGGACAATAACAT	Pcgf6	GGAGGAGATGGAGGAGGAG
	AGCCTCGGTTGGTATTTGTG		TTTGCAAATCGAACACAGGA
Esrrb	GGCGTTCTTCAAGAGAACCA	Rest	GTGCGAACTCACACAGGAGA
	TCCGTTTGGTGATCTCACAT		AAGAGGTTTAGGCCCGTTGT
Fgf5	GCTGTGTCTCAGGGGATTGT	Rpo	TTCATTGTGGGAGCAGAC
	TCTCGGCCTGTCTTTTCAGT		CAGCAGTTTCTCCAGAGC
Gata4	CACAAGATGAACGGCATCAACC	Sox17	CCGAGATGGGTCTTCCCTAC
	CAGCGTGGTGGTGGTAGTCTG		CGTCAAATGTCGGGGTAGTT
Gsc	GAAGCCCTGGAGAACCTCTT	Sox2	CTGCAGTACAACTCCATGACCAG
	GGCTTTTGAGGACGTCTTGT		GGACTTGACCACAGAGCCCAT
Hand1	ACGCACATCATCACCATCAT	T	GAACCTCGGATTCACATCGT
	CTACTGCGGTGGTAGGTGGT		TTCTTTGGCATCAAGGAAGG
Klf4	CAGCCATGTCAGACTCGCC	Tcfcp2l1	AGCATCATCCGTGTCGTTTT
	GTTTTTAATCTTCGTTGACTTTGGG		CAGTGCACCTGAATGAATGC
LifR	CTTGCAATGTGCCACTCACT	Tcl1	GATCTGGGAGAAGCACGTGTA
	CGAGCACCACTTTGTCTTGA		CCACATTAAAGGCAGCTCGT
Nanog	AGGCTGATTTGGTTGGTGTC	Tet2	GCCAGAAGCAAGAAACCAAG
	CCAGGAAGACCCACACTCAT		CCTTCCTTCAGACCCAAACA
Nestin	TCCCCTGAGGACCAGGAGT	Wnt3a	CCATGAACCGTCACAACAAT
	GTCTCAGGACAGTGCTGAGCCTT		CTTGAGGTGCATGTGACTGG

Table MM.2. Mouse primers used for mRNA analysis. The sequences of the different mouse primers used in this study are listed. The first primer is forward and the second reverse. All of them are shown in 5' to 3' direction.

14. Antibodies used

Antibody	Specie	Provider	Reference	ChIP	IP	WB
Cbx6	Rabbit pAb	In-house				1:1000
Cbx7	Rabbit pAb	Abcam	ab21873	5 µg		1:500
Ezh2	Mouse mAb	In-house				1:2000
Flag	Mouse mAb	Sigma	A2220		5 µg	1:1000
Gapdh	Mouse mAb	Santa Cruz	SC-32233			1:5000
HA	Rabbit pAb	BioLegend	902301	5 µl		1:2000
	Mouse mAb	In-house				1:100
H2A	Rabbit pAb	Abcam	ab18255	3 µl		1:5000
H2AK119Ub	Rabbit pAb	Cell Signaling	8240s	5 µl		1:5000
H3	Rabbit pAb	Abcam	ab1791	2 µg		1:5000
IgG	Rabbit pAb	Abcam	ab46540	5 µg	5 µg	
Mel18	Rabbit pAb	Santa Cruz	sc 10744	5 µg	5 µg	1:2000
Pcgf6	Rabbit pAb	Dr. Pasini lab		5 µg	2 µg	1:1500
Phc1	Mouse mAb	Active motif	39723	5 µg	5 µg	1:2000
Ring1B	Rabbit pAb	In-house		8 µl		1:2000
	Mouse mAb	MBL	D139-3		5 µg	
Rybp	Rabbit pAb	Millipore	3637	5 µg	5 µg	1:2000
Suz12	Rabbit pAb	Abcam	ab12073-100	5 µg		1:2000
L3mbtl2	Rabbit pAb	Sigma	HPA000815			1:100

Table MM.3. Antibodies and their applications. The antibodies used in this study, their commercial information, and dilution for use are given.

15. Histone peptides used

Peptide	Histone	Sequence
H4	H4 (11-27)	Ac-GKGGAKRHRKVLRDNIQ-Peg-Biot
H4scramble	H4 (11-27)	Ac-RKVLAKRHGKGGIQRDN-Biot
H4K16(Me3)	H4 (11-27)	Ac-GKGGAK(Me3)RHRKVLRDNIQ-Biot
H4K16Ac	H4 (11-27)	Ac-GKGGAK(Ac)RHRKVLRDNIQ-Peg-Biot
H4K20 me1	H4 (11-27)	Ac-GKGGAKRHRK(Me)VLRDNIQ-Biot
H4K20 me2	H4 (11-27)	Ac-GKGGAKRHRK(Me2)VLRDNIQ-Biot
H4K20 me3	H4 (11-27)	Ac-GKGGAKRHRK(Me3)VLRDNIQ-Peg-Biot
H4K20Ac	H4 (11-27)	Ac-GKGGAKRHRK(Ac)VLRDNIQ-Peg-Biot
H4K16AcK20Ac	H4 (11-27)	Ac-GKGGAK(Ac)RHRK(Ac)VLRDNIQ-Peg-Biot
H4K12AcK16AcK20Ac	H4 (11-27)	Ac-GK(Ac)GGAK(Ac)RHRK(Ac)VLRDNIQ-Peg-Biot
H2A	H2A (4-20)	Ac-GKQGGKARAKAKTRSSR-Biot
H3	H3 (19-35)	Ac-QLATKAARKSAPATGGVGG-K(Biotin)-CONH2
H3K27me3	H3 (19-35)	Ac-QLATKAAR[K(me3)]SAPATGGVGG-K(Biotin)-CONH2

Table MM.4. Histone peptide sequences. The histone peptides used in this study and their amino acid sequences are given. All peptides were produced in the Laboratory of Proteomics & Protein Chemistry at Pompeu Fabra University (Barcelona).

REFERENCES

1 Martello, G. & Smith, A. The nature of embryonic stem cells. *Annu Rev Cell Dev Biol* **30**, 647-675, doi:10.1146/annurev-cellbio-100913-013116 (2014).
2 Efroni, S. *et al.* Global transcription in pluripotent embryonic stem cells. *Cell stem cell* **2**, 437-447, doi:10.1016/j.stem.2008.03.021 (2008).
3 Loh, K. M. & Lim, B. A precarious balance: pluripotency factors as lineage specifiers. *Cell stem cell* **8**, 363-369, doi:10.1016/j.stem.2011.03.013 (2011).
4 Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. *Sci Adv* **1**, e1500737, doi:10.1126/sciadv.1500737 (2015).
5 Bhaumik, S. R., Smith, E. & Shilatifard, A. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol* **14**, 1008-1016, doi:10.1038/nsmb1337 (2007).
6 Selwood, L. & Johnson, M. H. Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: evolution and origins. *Bioessays* **28**, 128-145, doi:10.1002/bies.20360 (2006).
7 Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Developmental cell* **10**, 615-624, doi:10.1016/j.devcel.2006.02.020 (2006).
8 Brook, F. A. & Gardner, R. L. The origin and efficient derivation of embryonic stem cells in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5709-5712 (1997).
9 Batlle-Morera, L., Smith, A. & Nichols, J. Parameters influencing derivation of embryonic stem cells from murine embryos. *Genesis* **46**, 758-767, doi:10.1002/dvg.20442 (2008).
10 Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156 (1981).
11 Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 7634-7638 (1981).
12 Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917-929, doi:10.1016/j.cell.2005.08.040 (2005).
13 Kidder, B. L., Palmer, S. & Knott, J. G. SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem cells* **27**, 317-328, doi:10.1634/stemcells.2008-0710 (2009).
14 Yuan, P. *et al.* Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes & development* **23**, 2507-2520, doi:10.1101/gad.1831909 (2009).

15 Singh, A. M., Hamazaki, T., Hankowski, K. E. & Terada, N. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem cells* **25**, 2534-2542, doi:10.1634/stemcells.2007-0126 (2007).
16 Frankenberg, S. *et al.* Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Developmental cell* **21**, 1005-1013, doi:10.1016/j.devcel.2011.10.019 (2011).
17 Bessonnard, S. *et al.* Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* **141**, 3637-3648, doi:10.1242/dev.109678 (2014).
18 Rosner, M. H. *et al.* A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**, 686-692, doi:10.1038/345686a0 (1990).
19 Okamoto, K. *et al.* A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60**, 461-472 (1990).
20 Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-391 (1998).
21 Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell stem cell* **2**, 151-159, doi:10.1016/j.stem.2008.01.004 (2008).
22 Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature genetics* **24**, 372-376, doi:10.1038/74199 (2000).
23 Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development* **17**, 126-140, doi:10.1101/gad.224503 (2003).
24 Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature cell biology* **9**, 625-635, doi:10.1038/ncb1589 (2007).
25 Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem cells* **26**, 903-911, doi:10.1634/stemcells.2007-0951 (2008).
26 Arnold, K. *et al.* Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell stem cell* **9**, 317-329, doi:10.1016/j.stem.2011.09.001 (2011).
27 Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-655 (2003).
28 Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).

29 Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737, doi:10.1016/j.cell.2009.07.039 (2009).
30 Jeon, H. *et al.* Comprehensive Identification of Kruppel-Like Factor Family Members Contributing to the Self-Renewal of Mouse Embryonic Stem Cells and Cellular Reprogramming. *PloS one* **11**, e0150715, doi:10.1371/journal.pone.0150715 (2016).
31 Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature genetics* **38**, 431-440, doi:10.1038/ng1760 (2006).
32 Rodda, D. J. *et al.* Transcriptional regulation of nanog by OCT4 and SOX2. *The Journal of biological chemistry* **280**, 24731-24737, doi:10.1074/jbc.M502573200 (2005).
33 Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-1117, doi:10.1016/j.cell.2008.04.043 (2008).
34 Rizzino, A. Concise review: The Sox2-Oct4 connection: critical players in a much larger interdependent network integrated at multiple levels. *Stem cells* **31**, 1033-1039, doi:10.1002/stem.1352 (2013).
35 Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364-368, doi:10.1038/nature05284 (2006).
36 van den Berg, D. L. *et al.* An Oct4-centered protein interaction network in embryonic stem cells. *Cell stem cell* **6**, 369-381, doi:10.1016/j.stem.2010.02.014 (2010).
37 Liang, J. *et al.* Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nature cell biology* **10**, 731-739, doi:10.1038/ncb1736 (2008).
38 Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435, doi:10.1038/nature09380 (2010).
39 Gao, Z. *et al.* Determination of protein interactome of transcription factor Sox2 in embryonic stem cells engineered for inducible expression of four reprogramming factors. *The Journal of biological chemistry* **287**, 11384-11397, doi:10.1074/jbc.M111.320143 (2012).
40 Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-690, doi:10.1038/336688a0 (1988).
41 Williams, R. L. *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-687, doi:10.1038/336684a0 (1988).
42 Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & development* **12**, 2048-2060 (1998).
43 Bourillot, P. Y. *et al.* Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation

in cooperation with Nanog. *Stem cells* **27**, 1760-1771, doi:10.1002/stem.110 (2009).
44 Do, D. V. *et al.* A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. *Genes & development* **27**, 1378-1390, doi:10.1101/gad.221176.113 (2013).
45 Martello, G., Bertone, P. & Smith, A. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *The EMBO journal* **32**, 2561-2574, doi:10.1038/emboj.2013.177 (2013).
46 Stewart, C. L. *et al.* Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature* **359**, 76-79, doi:10.1038/359076a0 (1992).
47 Thomson, M. *et al.* Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875-889, doi:10.1016/j.cell.2011.05.017 (2011).
48 Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-292 (2003).
49 Yuan, H., Corbi, N., Basilico, C. & Dailey, L. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes & development* **9**, 2635-2645 (1995).
50 Kunath, T. *et al.* FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895-2902, doi:10.1242/dev.02880 (2007).
51 Burdon, T., Stracey, C., Chambers, I., Nichols, J. & Smith, A. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Developmental biology* **210**, 30-43, doi:10.1006/dbio.1999.9265 (1999).
52 Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523, doi:10.1038/nature06968 (2008).
53 Zeineddine, D. *et al.* Oct-3/4 dose dependently regulates specification of embryonic stem cells toward a cardiac lineage and early heart development. *Developmental cell* **11**, 535-546, doi:10.1016/j.devcel.2006.07.013 (2006).
54 Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871 (1974).
55 Fan, Y. *et al.* Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123**, 1199-1212, doi:10.1016/j.cell.2005.10.028 (2005).
56 Eissenberg, J. C. & Reuter, G. Cellular mechanism for targeting heterochromatin formation in Drosophila. *Int Rev Cell Mol Biol* **273**, 1-47, doi:10.1016/S1937-6448(08)01801-7 (2009).
57 Huisinga, K. L., Brower-Toland, B. & Elgin, S. C. The contradictory definitions of heterochromatin: transcription and

silencing. *Chromosoma* **115**, 110-122, doi:10.1007/s00412-006-0052-x (2006).
58 Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**, 846-856, doi:10.1038/nrc1991 (2006).
59 Campos, E. I. & Reinberg, D. Histones: annotating chromatin. *Annu Rev Genet* **43**, 559-599, doi:10.1146/annurev.genet.032608.103928 (2009).
60 Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science* **330**, 612-616, doi:10.1126/science.1191078 (2010).
61 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705, doi:10.1016/j.cell.2007.02.005 (2007).
62 Brownell, J. E. & Allis, C. D. Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr Opin Genet Dev* **6**, 176-184 (1996).
63 Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental cell* **10**, 105-116, doi:10.1016/j.devcel.2005.10.017 (2006).
64 Park, S. H. *et al.* Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells. *Ultrastruct Pathol* **28**, 229-238 (2004).
65 Ahmed, K. *et al.* Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PloS one* **5**, e10531, doi:10.1371/journal.pone.0010531 (2010).
66 Gutierrez, J. L., Chandy, M., Carrozza, M. J. & Workman, J. L. Activation domains drive nucleosome eviction by SWI/SNF. *The EMBO journal* **26**, 730-740, doi:10.1038/sj.emboj.7601524 (2007).
67 Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707-719, doi:10.1016/j.cell.2007.01.015 (2007).
68 Sif, S. ATP-dependent nucleosome remodeling complexes: enzymes tailored to deal with chromatin. *J Cell Biochem* **91**, 1087-1098, doi:10.1002/jcb.20005 (2004).
69 Bultman, S. *et al.* A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol Cell* **6**, 1287-1295 (2000).
70 Faast, R. *et al.* Histone variant H2A.Z is required for early mammalian development. *Current biology : CB* **11**, 1183-1187 (2001).
71 Stopka, T. & Skoultchi, A. I. The ISWI ATPase Snf2h is required for early mouse development. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14097-14102, doi:10.1073/pnas.2336105100 (2003).
72 Bosman, E. A. *et al.* Multiple mutations in mouse Chd7 provide models for CHARGE syndrome. *Hum Mol Genet* **14**, 3463-3476, doi:10.1093/hmg/ddi375 (2005).

73 Keenen, B. & de la Serna, I. L. Chromatin remodeling in embryonic stem cells: regulating the balance between pluripotency and differentiation. *Journal of cellular physiology* **219**, 1-7, doi:10.1002/jcp.21654 (2009).
74 Lee, J. H., Hart, S. R. & Skalnik, D. G. Histone deacetylase activity is required for embryonic stem cell differentiation. *Genesis* **38**, 32-38, doi:10.1002/gene.10250 (2004).
75 Meshorer, E. & Misteli, T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nature reviews. Molecular cell biology* **7**, 540-546, doi:10.1038/nrm1938 (2006).
76 Breiling, A., Bonte, E., Ferrari, S., Becker, P. B. & Paro, R. The Drosophila polycomb protein interacts with nucleosomal core particles In vitro via its repression domain. *Molecular and cellular biology* **19**, 8451-8460 (1999).
77 Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326, doi:10.1016/j.cell.2006.02.041 (2006).
78 Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-570 (1978).
79 Whitcomb, S. J., Basu, A., Allis, C. D. & Bernstein, E. Polycomb Group proteins: an evolutionary perspective. *Trends Genet* **23**, 494-502, doi:10.1016/j.tig.2007.08.006 (2007).
80 Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147-1155, doi:10.1038/nsmb.2669 (2013).
81 Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature reviews. Molecular cell biology* **10**, 697-708, doi:10.1038/nrm2763 (2009).
82 Cao, R. & Zhang, Y. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* **14**, 155-164, doi:10.1016/j.gde.2004.02.001 (2004).
83 Morey, L. & Helin, K. Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci* **35**, 323-332, doi:10.1016/j.tibs.2010.02.009 (2010).
84 Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *The EMBO journal* **25**, 3110-3122, doi:10.1038/sj.emboj.7601187 (2006).
85 Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-678, doi:10.1016/j.cell.2011.12.029 (2012).
86 Comet, I. & Helin, K. Revolution in the Polycomb hierarchy. *Nat Struct Mol Biol* **21**, 573-575, doi:10.1038/nsmb.2848 (2014).
87 Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445-1459, doi:10.1016/j.cell.2014.05.004 (2014).

88 Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a polycomb group protein complex. *Science* **306**, 1574-1577, doi:10.1126/science.1100576 (2004).
89 Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature cell biology* **9**, 1428-1435, doi:10.1038/ncb1663 (2007).
90 Richly, H. *et al.* Transcriptional activation of polycomb-repressed genes by ZRF1. *Nature* **468**, 1124-1128, doi:10.1038/nature09574 (2010).
91 Aloia, L. *et al.* Zrf1 is required to establish and maintain neural progenitor identity. *Genes & development* **28**, 182-197, doi:10.1101/gad.228510.113 (2014).
92 Nakagawa, T. *et al.* Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes & development* **22**, 37-49, doi:10.1101/gad.1609708 (2008).
93 Gao, Z. *et al.* PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* **45**, 344-356, doi:10.1016/j.molcel.2012.01.002 (2012).
94 Huynh, K. D., Fischle, W., Verdin, E. & Bardwell, V. J. BCoR, a novel corepressor involved in BCL-6 repression. *Genes & development* **14**, 1810-1823 (2000).
95 Wu, X., Johansen, J. V. & Helin, K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol Cell* **49**, 1134-1146, doi:10.1016/j.molcel.2013.01.016 (2013).
96 Farcas, A. M. *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* **1**, e00205, doi:10.7554/eLife.00205 (2012).
97 Trojer, P. *et al.* L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Mol Cell* **42**, 438-450, doi:10.1016/j.molcel.2011.04.004 (2011).
98 Vizan, P., Beringer, M., Ballare, C. & Di Croce, L. Role of PRC2-associated factors in stem cells and disease. *FEBS J* **282**, 1723-1735, doi:10.1111/febs.13083 (2015).
99 Nowak, A. J. *et al.* Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding sites. *The Journal of biological chemistry* **286**, 23388-23396, doi:10.1074/jbc.M110.207407 (2011).
100 Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* **15**, 57-67, doi:10.1016/j.molcel.2004.06.020 (2004).
101 Peng, J. C. *et al.* Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**, 1290-1302, doi:10.1016/j.cell.2009.12.002 (2009).

102 Savla, U., Benes, J., Zhang, J. & Jones, R. S. Recruitment of Drosophila Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. *Development* **135**, 813-817, doi:10.1242/dev.016006 (2008).
103 Qin, S. *et al.* Tudor domains of the PRC2 components PHF1 and PHF19 selectively bind to histone H3K36me3. *Biochem Biophys Res Commun* **430**, 547-553, doi:10.1016/j.bbrc.2012.11.116 (2013).
104 Ballare, C. *et al.* Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat Struct Mol Biol* **19**, 1257-1265, doi:10.1038/nsmb.2434 (2012).
105 Walker, E. *et al.* Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell stem cell* **6**, 153-166, doi:10.1016/j.stem.2009.12.014 (2010).
106 Zhang, Z. *et al.* PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. *Stem cells* **29**, 229-240, doi:10.1002/stem.578 (2011).
107 Forzati, F. *et al.* CBX7 is a tumor suppressor in mice and humans. *J Clin Invest* **122**, 612-623, doi:10.1172/JCI58620 (2012).
108 de Napoles, M. *et al.* Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Developmental cell* **7**, 663-676, doi:10.1016/j.devcel.2004.10.005 (2004).
109 Riising, E. M. *et al.* Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* **55**, 347-360, doi:10.1016/j.molcel.2014.06.005 (2014).
110 Leeb, M. & Wutz, A. Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J Cell Biol* **178**, 219-229, doi:10.1083/jcb.200612127 (2007).
111 Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**, 1513-1524, doi:10.1242/dev.014340 (2008).
112 Morey, L. *et al.* Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell stem cell* **10**, 47-62, doi:10.1016/j.stem.2011.12.006 (2012).
113 Hisada, K. *et al.* RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. *Molecular and cellular biology* **32**, 1139-1149, doi:10.1128/MCB.06441-11 (2012).
114 Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *The EMBO journal* **23**, 4061-4071, doi:10.1038/sj.emboj.7600402 (2004).

115 Faust, C., Lawson, K. A., Schork, N. J., Thiel, B. & Magnuson, T. The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development* **125**, 4495-4506 (1998).
116 Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & development* **24**, 265-276, doi:10.1101/gad.544410 (2010).
117 Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* **32**, 491-502, doi:10.1016/j.molcel.2008.10.016 (2008).
118 Fisher, C. L. & Fisher, A. G. Chromatin states in pluripotent, differentiated, and reprogrammed cells. *Curr Opin Genet Dev* **21**, 140-146, doi:10.1016/j.gde.2011.01.015 (2011).
119 Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306-310, doi:10.1038/nature08788 (2010).
120 Bernstein, E. *et al.* Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Molecular and cellular biology* **26**, 2560-2569, doi:10.1128/MCB.26.7.2560-2569.2006 (2006).
121 Senthilkumar, R. & Mishra, R. K. Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. *BMC Genomics* **10**, 549, doi:10.1186/1471-2164-10-549 (2009).
122 Milosevich, N. *et al.* Selective Inhibition of CBX6: A Methyllysine Reader Protein in the Polycomb Family. *ACS Med Chem Lett* **7**, 139-144, doi:10.1021/acsmedchemlett.5b00378 (2016).
123 Pemberton, H. *et al.* Genome-wide co-localization of Polycomb orthologs and their effects on gene expression in human fibroblasts. *Genome Biol* **15**, R23, doi:10.1186/gb-2014-15-2-r23 (2014).
124 Vandamme, J., Volkel, P., Rosnoblet, C., Le Faou, P. & Angrand, P. O. Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Mol Cell Proteomics* **10**, M110 002642, doi:10.1074/mcp.M110.002642 (2011).
125 Li, G. *et al.* Altered expression of polycomb group genes in glioblastoma multiforme. *PloS one* **8**, e80970, doi:10.1371/journal.pone.0080970 (2013).
126 Yap, K. L. *et al.* Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* **38**, 662-674, doi:10.1016/j.molcel.2010.03.021 (2010).
127 Simhadri, C. *et al.* Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). *J Med Chem* **57**, 2874-2883, doi:10.1021/jm401487x (2014).

128 Munoz, J. *et al.* The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *The EMBO journal* **31**, 3079-3091, doi:10.1038/emboj.2012.166 (2012).
129 Rinaldi, L. *et al.* Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis. *Cell stem cell*, doi:10.1016/j.stem.2016.06.020 (2016).
130 Ren, C. *et al.* Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. *Chem Biol* **22**, 161-168, doi:10.1016/j.chembiol.2014.11.021 (2015).
131 Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**, 1025-1040, doi:10.1038/nsmb1338 (2007).
132 Morey, L., Santanach, A. & Di Croce, L. Pluripotency and Epigenetic Factors in Mouse Embryonic Stem Cell Fate Regulation. *Molecular and cellular biology* **35**, 2716-2728, doi:10.1128/MCB.00266-15 (2015).
133 Pasini, D. & Di Croce, L. Emerging roles for Polycomb proteins in cancer. *Curr Opin Genet Dev* **36**, 50-58, doi:10.1016/j.gde.2016.03.013 (2016).
134 Qin, J. *et al.* The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. *Cell stem cell* **11**, 319-332, doi:10.1016/j.stem.2012.06.002 (2012).
135 Yang, C. S., Chang, K. Y., Dang, J. & Rana, T. M. Polycomb Group Protein Pcgf6 Acts as a Master Regulator to Maintain Embryonic Stem Cell Identity. *Sci Rep* **6**, 26899, doi:10.1038/srep26899 (2016).
136 Kadzik, R. S. *et al.* Wnt ligand/Frizzled 2 receptor signaling regulates tube shape and branch-point formation in the lung through control of epithelial cell shape. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 12444-12449, doi:10.1073/pnas.1406639111 (2014).
137 Zhang, Y. *et al.* Lef1 contributes to the differentiation of bulge stem cells by nuclear translocation and cross-talk with the Notch signaling pathway. *Int J Med Sci* **10**, 738-746, doi:10.7150/ijms.5693 (2013).
138 Lopez-Rios, J., Esteve, P., Ruiz, J. M. & Bovolenta, P. The Netrin-related domain of Sfrp1 interacts with Wnt ligands and antagonizes their activity in the anterior neural plate. *Neural Dev* **3**, 19, doi:10.1186/1749-8104-3-19 (2008).
139 Jian, H. *et al.* Smad3-dependent nuclear translocation of beta-catenin is required for TGF-beta1-induced proliferation of bone marrow-derived adult human mesenchymal stem cells. *Genes & development* **20**, 666-674, doi:10.1101/gad.1388806 (2006).
140 Thomas, G. M. *et al.* A GSK3-binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. *FEBS Lett* **458**, 247-251 (1999).

141 Beck, D. B., Oda, H., Shen, S. S. & Reinberg, D. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes & development* **26**, 325-337, doi:10.1101/gad.177444.111 (2012).
142 Pei, H. *et al.* MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature* **470**, 124-128, doi:10.1038/nature09658 (2011).
143 Wongtawan, T., Taylor, J. E., Lawson, K. A., Wilmut, I. & Pennings, S. Histone H4K20me3 and HP1alpha are late heterochromatin markers in development, but present in undifferentiated embryonic stem cells. *J Cell Sci* **124**, 1878-1890, doi:10.1242/jcs.080721 (2011).
144 O'Loghlen, A. *et al.* MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell stem cell* **10**, 33-46, doi:10.1016/j.stem.2011.12.004 (2012).
145 Kloet, S. L. *et al.* The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol* **23**, 682-690, doi:10.1038/nsmb.3248 (2016).
146 Zdzieblo, D. *et al.* Pcgf6, a polycomb group protein, regulates mesodermal lineage differentiation in murine ESCs and functions in iPS reprogramming. *Stem cells* **32**, 3112-3125, doi:10.1002/stem.1826 (2014).
147 Hammachi, F. *et al.* Transcriptional activation by Oct4 is sufficient for the maintenance and induction of pluripotency. *Cell Rep* **1**, 99-109, doi:10.1016/j.celrep.2011.12.002 (2012).
148 Kaustov, L. *et al.* Recognition and specificity determinants of the human cbx chromodomains. *The Journal of biological chemistry* **286**, 521-529, doi:10.1074/jbc.M110.191411 (2011).
149 Domanski, M. *et al.* Improved methodology for the affinity isolation of human protein complexes expressed at near endogenous levels. *Biotechniques* **0**, 1-6, doi:10.2144/000113864 (2012).

RESEARCH ARTICLES

Research article resulting from this thesis:

> **Santanach A**, Jiang H, Blanco E, Sansó M, Lacava J, Morey L, Di Croce L. The Polycomb group protein Cbx6 is an essential regulator of embryonic stem cell identity. Manuscript under preparation.

Collaboration to determine the role of the Polycomb group protein Mel18 in cardiac differentiation:

> Morey L, **Santanach A**, Blanco E, Aloia L, Nora EP, Bruneau BG, Di Croce L. Polycomb Regulates Mesoderm Cell Fate-Specification in Embryonic Stem Cells through Activation and Repression Mechanisms. Cell Stem Cell. 2015 Sep 3;17(3):300-15.

Collaboration to elucidate the changes in PRC1 and PRC2 composition and genomic distribution during neural differentiation:

> Kloet SL, Makowski MM, Baymaz HI, van Voorthuijsen L, Karemaker ID, **Santanach A**, Jansen PW, Di Croce L, Vermeulen M. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. Nat Struct Mol Biol. 2016 Jul;23(7):682-90.

Collaboration to understand the role of Zrf1 in neural progenitors:

> Aloia L, Di Stefano B, Sessa A, Morey L, **Santanach A**, Gutierrez A, Cozzuto L, Benitah SA, Graf T, Broccoli V, Di Croce L. Zrf1 is required to establish and maintain neural

progenitor identity. Genes Dev. 2014 Jan 15;28(2):182-97.

Review article:

Morey L, **Santanach A**, Di Croce L. Pluripotency and Epigenetic Factors in Mouse Embryonic Stem Cell Fate Regulation. Mol Cell Biol. 2015 Aug;35(16):2716-28.

ACKNOWLEDGEMENTS

Encara recordo el dia que vaig entrar per primera vegada al laboratori. En Luciano encara no havia arribat, i jo, estava tan nerviosa que no vaig ni atrevir-me a esperar-lo a dins, em vaig quedar tímidament esperant al passadís. Qui m'ho diria que després de cinc anys, aquell laboratori al que no gosava entrar, s'acabaria convertint en la meva segona casa (o la primera, en funció dels experiments!). Sou molts els que m'heu acompanyat durant tot aquest trajecte i és per això, que no hi ha res que em faci més feliç que deixar gravat en aquestes pàgines una petita mostra del meu agraïment.

En primer lugar, gracias a ti Luciano, por dejarme formar parte de esta pequeña gran "famiglia". Gracias por tu positividad, y por haber confiado siempre en Cbx6. Por ayudarme a abrir ventanas cuando parecía que todas las puertas estaban herméticamente cerradas. Hacer el PhD ha sido una experiencia de un valor incalculable donde he aprendido mucho más que ser una buena científica. Grazie.

Lluís, gràcies per haver estat sempre al meu costat. Per contagiar-me la teva passió per la ciència, per la teva implicació des del minut 1. Per tenir sempre temps per poder-me ajudar, inclús ara, a milers de kilòmetres de distància. Perquè el meu iPod té cançons com 'Take a walk' o 'Midnight city' que em fan recordar els mil i un bons moments viscuts.

A mis compañeros de laboratorio, siempre he pensado que soy muy afortunada de formar parte de un grupo como el nuestro. El clan Di Croce, siempre en pié de guerra, dispuesto a ayudar científica- y personalmente. A mis veteranas: Arantxa, gràcies per

les infinites converses, per poder parlar de tot, i pels consells de 'mami' quan els he necessitat. Cecilia, simplemente gracias por ser 'Cecita', por estar siempre dispuesta a ayudarme aunque eso significara retrasar tus experimentos, por compartir aventuras de palomas y gatitos.

A mis chicos de oro. Pedro, se'm dibuixa un somriure quan penso què dir-te, i no, no és perquè m'enrecordi de les tonteries que feies aquesta tarda abans que marxés. Gràcies pels consells, per les nostres reflexions sobre la vida, perquè no ha passat ni un dia que no m'hagis fet riure. He après molt de tu. Malte, empezamos y acabaremos esto juntos. Gracias por entenderme siempre, muchas veces no han hecho falta palabras para saber qué pensábamos.

Gloria, per la bondat que et caracteritza, per tenir la porta sempre oberta, per veure i creure gràcies a tu, que les dones, en ciència, poden arribar molt lluny (et trobarem a faltar!). Sergi, per tot el que ens estàs aportant a mi i al grup, creu-me que és incalculable. Miriam, per ser una alenada d'aire fresc mallorquí al laboratori (i al meu projecte!). Gràcies per implicar-te i ajudar-me, i no només a nivell de 'bench'. Santi, per ser el meu PhD sènior referent, company d'MHs i d'indignacions. Gràcies per ajudar-me també en aquest últim tram. Enrique, per aportar-me serenitat quan les coses no sortien com pensàvem. A mi MVP. Fichaje estrella. No me sorprende nada que en poco tiempo todo el mundo te quiera. A mis queridos juniors. Paul, because I see in you the enthusiasm that I had when I started the PhD. You will go far. Marc i Anna, perquè en poc temps heu passat a ser imprescindibles. No us imagineu com m'hagués agradat poder

començar el PhD amb vosaltres.

A mis tortugas ninjas, Guio, Jordi, Lore. Por esos dos primeros años en el 622.04. Gracias por hacerme sentir una 'Salvatora' más. Porqué nos quitaron los cafés y los lunchs en la terraza, pero no han podido separar ese vínculo inicial que nació después de unas cuantas horas en un jacuzzi, e intento de asesinato de un radiocasete. Gracias por dejarme compartir tantos momentos con vosotros.

Livia, la mia amica italiana. Gracias por tu amistad y estar siempre a mi lado. Por todas y cada una de nuestras aventuras juntas donde parecía que se acababa el mundo. Porqué ha pasado el tiempo, y te has ido a vivir con las llamas, pero continuamos unidas.

També m'agradaria agraïr a aquelles persones que han estat al meu costat, des de l'inici de la meva carrera com a científica. Núria, Moncu, Sara, Marta, Albert... Va ser genial fer la carrera al vostre costat, i encara més veure que passen els anys i que tot continua com sempre. Estem aconseguint els nostres objectius!
I qui ho hagués dit, que de la meva primera experiència en un laboratori m'emportaria tan gran amistat, Esther, gràcies.

Però sense dubte, gràcies a la meva família. Sou vosaltres qui realment heu estat al meu costat incondicionalment. Al papa, per ser el meu pilar de força durant tot aquest camí. A la mama supermami, per cuidar-me tant i pensar sempre en com ajudar-me. A la meva animadora número 1, ànima bessona i millor germana. Al meu germà des de fa 11 anys. A la iaia, per les

abraçades estrujadores que m'han fet marxar els nervis de cop. A l'avi, per entendre'm a la seva manera.

I a tu, la meva meitat. Perquè al teu costat ha estat més fàcil, per escoltar-me, per completar-me i per fer-me feliç.