Variations of Polycomb assembly in mouse embryonic stem cells and early differentiation

Ivano Mocavini

TESI DOCTORAL UPF / 2021

Thesis Supervisor
Dr. Luciano Di Croce

Centre for Genomic Regulation (CRG)
Gene Regulation, Stem Cells and Cancer Program

Universitat Pompeu Fabra
Department of Experimental and Health Sciences

Abstract

Polycomb repressive complex (PRC)1 and 2 are major players in gene regulation, devoted to maintenance of the epigenetic memory of gene silencing. Mouse embryonic stem cells (mESCs) have been extensively exploited as a model system to study Polycomb complexes composition and function, leading to the identification of many accessory factors. However, the mechanisms regulating PRC1/2 composition in mESCs are still poorly understood. Moreover, several reports have shown that changes in composition occur upon cell differentiation, although little is known about the functional relevance of these changes. In this doctoral thesis I aim to address these questions by focusing on two examples of these aspects: first, I characterize the PRC1 interactome upon mESCs differentiation to primitive endoderm (PrE) and try to assess its role in this cell fate transition. Secondly, I report the identification of a novel splicing isoform of PRC2 core component *Suz12*, with implications in PRC2 composition and activity on chromatin

Resumen

Los complejos de represión Polycomb (PRC)1 y 2 juegan un papel central en la regulación de la expresión génica, preservando la memoria epigenética del estado de silenciamiento. El uso de células madre embrionarias de ratón (mESCs) como modelo para el estudio de los complejos Polycomb, ha permitido la identificación de muchos factores accesorios. Sin embargo, no se conocen los mecanismos que regulan la composición de PRC1/2 en las mESCs. Además, varios informes han demostrado que, tras la diferenciación celular, ocurren ciertos cambios en la composición, aunque se sabe poco sobre la relevancia funcional de estos. En esta tesis doctoral trato de abordar estas cuestiones centrándome en dos ejemplos concretos: primero, caracterizo la red de proteínas asociadas a PRC1 tras la diferenciación de mESCs al endodermo primitivo (PrE), e intento evaluar su rol durante esta transición de identidad celular. En segundo lugar, informo del hallazgo de una nueva isoforma de splicing de *Suz12*, un componente central de PRC2, con implicaciones para la composición y la actividad de PRC2 en la cromatina.

Preface

Differential gene expression allows for the diversification of cell types starting from the same identical genetic material. This process is driven by both intrinsic and extrinsic cues, that ultimately direct the development of one single cell into a whole body. However, these signals are transitory, while changes in the cells can last for longer periods. This phenomenon is called epigenetic memory and is mediated by several nuclear factors that catalyze the deposition of epigenetic modifications. Among these factors, are the Polycomb group (PcG) of proteins, which catalyze the deposition of marks related with the silent state and contribute directly to gene repression. Epigenetic modifications are specific chemical moieties deposited on the proteins associated with the genome. These molecules provide a chemical baton that, upon DNA replication, is passed from the parental to the newly synthesized double helices. In physiological conditions, this mechanism allows the different cells of a growing embryo, to resume their task after every cell division, without forgetting which genes were being expressed, which ones where not, or which ones were ready to be activated. Importantly, this is fundamental for the symmetry breaking, for the correct segregation of cell lineages during development, and failure to do so, results in embryonic lethality or strong developmental defects. Consistently, mutations in Polycomb genes are found in patients suffering from overgrowth syndromes, including Sotos, Weaver, and Cohen-Gibson syndromes.

However, the role of Polycomb factors is also fundamental for cell differentiation in adult tissues. This is the case of the hematopoietic compartment, where a reservoir of stem cells must constantly replenish the whole hematopoietic system, throughout life, to ultimately produce all blood cell types. Owing to their importance in this process, Polycomb factors can be found mutated in several hematopoietic malignancies, including leukemias and lymphomas.

PcG proteins organize into separate Polycomb repressive complexes (PRC) namely, the PRC1 and PRC2. The activity of these complexes heavily relies on a growing list of accessory factors, many of which are probably still unknown. Most of the studies focusing on this, make use of mouse embryonic stem cells as a model system. However, several reports have shown that many of these genes are expressed in a dynamic fashion, depending on several

factors, including cell cycle phase, cell type, and developmental stage. Therefore, effort is being put into trying to characterize the composition of PRC1/2 in different contexts, to try to better understand their mode of function.
Importantly, PRC1/2 accessory factors mediate targeting of the complex to chromatin, thereby shaping the local epigenetic landscape, which in turn affect the cellular transcription program. Study of their structure and function could therefore provide novel strategies for targeting of the Polycomb complexes in diseases.
In this framework, I have tried to contribute to the understanding of the Polycomb machinery by means of complementary approaches. I have addressed how different forms of PRC2 coexist and cooperate in mESCs. Moreover, I have tried to characterize the changes that occur in PRC1 composition during the early stages of embryonic development, when the activity of PRC1 and PRC2 is essential. The data presented provides novel clues, laying the bases for the investigation of fundamental aspects in PRC1 and PRC2 activity.

Table of contents

Page

Introduction

Gene regulation and epigenetics

All cells of the body originate from the zygote. The information contained in this cell is replicated and inherited at every cell duplication happening along the way, such that virtually all the cell types of the organism share the same information, encoded in the hereditary material. By and large, the goal of many biological sciences is to understand how the same information can be interpreted in such a diverse variety of ways to perform the different tasks that an organism requires.

Study of the physical and functional nature of this information, has allowed researchers to define its operational unit, called the gene – a term coined by the danish botanist Wilhelm Johannsen to refer to Mendel's units of heredity (Johannsen, 1909), who is also quoted for having named the difference between the genotype (set of hereditary information) and the phenotype (its physical manifestation). Hence, while genetics is the study of the genotype, epigenetics is, in its very first definition, "the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being" (Waddington, 1968, 2012). In this broad definition, epigenetics very much overlaps with the study of gene expression.

More than half a century, and dozens of definitions of epigenetics later, we have learned quite a lot about genes, products, and their causal interactions. Genes are well defined regions of a double helix of deoxyribonucleic acid (DNA), whose informational content is given by the series of nucleobases embedded in it, namely adenine (A), cytosine (C), guanine (G) and thymine (T). Instead, products are, for most of the genes, represented by polymer chains of amino acids called proteins. Production of proteins from DNA requires the transcription of the gene into a temporary intermediate nucleic acid called messenger ribonucleic acid (mRNA, in which T is substituted with uracil, U), whose information is then translated three nucleobases at a time by transfer RNAs (tRNAs) to corresponding amino

acids. This process happens within the ribosome, a specialized ribonucleoprotein particle that catalyses the formation of covalent binds between corresponding amino acids, forming proteins.
In eukaryotes, the informational content of a gene is partitioned into smaller bits (exons) separated by stretches of nucleobases without coding potential (introns). The process of excluding the introns from the initial RNA transcript is called splicing. Importantly, splicing can be regulated to adjust the portions included and excluded from the mRNA. The differential splicing of the transcripts in certain situations (e.g., cell types, developmental stages, species, environmental stimuli etc.) is called alternative splicing (AS), and, together with RNA transcription (see below) represents a source of phenotypic variability (reviewed in (Ule and Blencowe, 2019)).
In its ample meaning, the study of gene expression refers to the study of all the mechanisms acting along this process, from DNA transcription to protein post-translation regulation. However, RNA content (also referred to as the transcriptome) by itself, represents the most efficient predictor when it comes to telling the identity of a cell. In other words, each different cell type interprets the genome in a different way, by transcribing certain genes and not others. Therefore, most of the efforts in the study of epigenetics, are focused on the study of the mechanisms that regulate RNA transcription from DNA.
Stripped down to the essential, RNA transcription of protein coding genes is carried out by an enzyme known as RNA polymerase II (RNA Pol II). Engagement of this enzyme with the transcription start site (TSS) of the gene results in RNA Pol II run across the gene body and concomitant synthesis of an RNA strand that is complementary to the DNA template (reviewed in (Cramer, 2019)). However, the DNA in the nucleus is not naked but rather associated to a group of positively charged relatively small proteins known as histones. There are 5 families of histone paralogs, namely H1, H2A and H2B, H3 and H4. All histones except H1 associate in two dimers of heterodimers (H2A with H2B, H3 with H4) to compose a histone octamer, around which 147 bases of the DNA double helix are

wrapped, forming a nucleosome. Instead, histone H1 contacts the DNA that connects two neighbouring nucleosomes, namely the linker DNA. This complex of proteins and DNA is referred to as the chromatin fibre. Importantly, this organization has two main roles: one is that of compacting very long molecules into the relatively small space of the nucleus, and the other one is to regulate access to the genetic information contained in the DNA.

Whether a gene must be transcribed or not in a certain cell or at a certain time, depends on an extra layer of information that is deposited on chromatin, in the form of chemical modifications.

One of these modifications is the deposition of a methyl group (CH_3) on the fifth atom of the cytosine ring (5mC) which typically occurs on cytosines that are followed by guanines (CpG dinucleotide). This chemical modification does not alter the information contained in the DNA, and mainly appears near the promoter regions, which are regions of the DNA immediately adjacent to the TSS (often referred as CpG islands, CGIs), whose ultimate role is to regulate the engagement of RNA Pol II for transcription. Methylation of the CpG dinucleotides at CGIs or within the gene promoter is associated with stable silencing (reviewed in (Jones, 2012)).

However, most of the epigenetic modifications involve histone residues. Particularly, the residues of the N-terminal tails of the histones, protruding from the core of the nucleosome, are heavily modified from a plethora of nuclear enzymes. Importantly, these modifications are dynamic in the sense that they can be added and removed, by epigenetic “writers” and “erasers” respectively. A large and growing list of epigenetic modifications has been associated with these residues (reviewed in (Lawrence et al., 2016)), the best studied of which are the addition of methyl or acetyl ($COCH_3$) moieties to the lysine residues. Some of them are associated with actively transcribed genes, like tri-methylation of the lysine 4 of histone H3 (H3K4me3), or lysine 36 (H3K36me3), or acetylation at most of the lysines; others are associated with either low or no transcription of the corresponding gene, such as H3K9me3, H3K27me3, or H4K20me3. Importantly, while acetylation directly affects the ability

of nucleosomes to interact with each other, therefore resulting in chromatin decompaction, methylation does not have a direct impact on nucleosomes. This means that its information has to be read and interpreted, a role that is accomplished by specific proteins defined chromatin readers (reviewed in (Musselman et al., 2012)). Therefore, the coordinated action of epigenetic writers, readers and erasers, results in the timely regulation of gene expression.

The Polycomb group of genes

One of the key questions underlying developmental biology is how cells acquire a different fate depending on their localisation in the developing embryo. In other words, how do the different parts of the body acquire distinct features in the embryo that will result in specific body characteristics in the adult. The body patterning of the fruit fly *drosophila melanogaster* has been a useful system to address such question, leading to the discovery of a set of genes dedicated to instructing pieces of the body on their identity: at the molecular level, axial body patterning in drosophila is determined by the expression of homeobox-containing transcription factors, known as homeotic (or Hox) genes.

In *drosophila*, most of these genes are found in two clusters (or complexes) on chromosome 3 (Fig. I1A): the Antennapedia Complex (ANT-C), composed of *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*) and *Antennapedia* (*Antp*); and the Bithorax Complex (BX-C), composed of *Ultrabithorax* (*Ubx*), *abdominal-A* (*abdA*), and *Abdominal-B* (*AbdB*) (Wellik, 2007). These two complexes are collectively referred to as the Homeotic complex (HOM-C). These genes are coregulated to achieve the final body plan, in a way that is known as ‘collinearity’ (or spatial collinearity, see below), meaning that Hox genes are expressed along the antero-posterior axis according to their linear position in the HOM-C cluster (Fig. I1A,B) (Lewis, 1978). Mutations in individual homeotic genes usually result in the loss of identity in those segments where the homeotic gene is normally expressed. The affected segments acquire the features of the segment immediately

anterior (Lewis, 1963), a phenomenon known as anterior transformation.

The adult fly presents 8 abdominal, 3 thoracic and 4 to 6 head segments, each displaying distinct morphological properties that distinguish it from the rest (Lewis, 1963). Those include for example the presence of antennae and eyes on the head, legs on the three thoracic segments or the ovary or testis in the abdomen.

Among those features is the presence of sex combs in male flies, that is restricted to the first pair of legs, corresponding to the prothoracic segment (T1). In 1942, Eleanor Slifer from Iowa State University reported the emergence of a naturally occurring mutant strain that presented extra pairs of sex combs on the second and third pairs of legs (hence the name *extra sex comb*, *esc*, Fig. I1D) (Slifer, 1942). Interestingly, Slifer acknowledged that a similar phenotype – appearance of sex combs on the second pair of legs – had been described two years earlier from A. Buzzati-Traverso in a strain carrying a mutation in the *spineless* (*ss*) locus (Buzzati-Traverso, 1940), also known as aristapedia, a gene involved in the definition of the identity of the distal region of both the antennae and legs (Duncan et al., 1998; Struhl, 1982). Five years after Slifer's report, Pamela Harrah Lewis who worked at Caltech together with her husband and future Nobel recipient Edward Lewis, would report a similar mutant displaying multiple sex combs naming it *Polycomb* (*Pc*) (Lewis, 1947).

Study of this phenotype suggested that the defects did not involve one specific segment (as it is the case for Hox individual mutants) but encompassed the entire body axis. The most evident defects regarded transformation of meso and metathoracic legs into prothoracic ones (appearance of sex combs), transformation of wings into halteres, and ventral into dorsal wings (Denell and Frederick, 1983; Lewis, 1963, 1978; Lewis, 1947), all defects that can be classified as posterior transformations. Successive screenings would lead to isolation of many other genes whose mutations gave similar phenotypes (Fig. I2A) and that are collectively referred to as the Polycomb group (PcG) of genes (Jürgens, 1985). Those included

Pleiohomeotic (*Pho*) (Gehring, 1970), *Suppressor of zeste 2* (*Su(z)2*) (Kalisch and Rasmuson, 1974), *Polycomblike* (*Pcl*) (Duncan, 1982), *Enhancer of* Polycomb (*E(Pc)*), *super sex combs* (*sxc*) (Ingham, 1984), *Additional sex combs* (*Asx*), *Posterior sex combs* (*Psc*) (Nusslein-Volhard et al., 1984), *Sex comb on midleg* (*Scm*) (Jürgens, 1985), *polyhomeotic* (*ph*) (Dura et al., 1985; Dura et al., 1987), *Sex combs extra* (*Sce*) (Breen and Duncan, 1986), *Enhancer of zeste* (*E(z)*) also known as *Polycombeotic* (*pco*) (Jones and Gelbart, 1990; Phillips and Shearn, 1990), *Suppressor of zeste 12* (*Su(z)12*) (Birve and Rasmuson-Lestander, 1994; Birve et al., 2001), and *calypso* (Gaytan de Ayala Alonso et al., 2007). Note: in this work I will use the words Polycomb and PcG with an ample meaning, referring to the whole class of genes/proteins/complexes depending on the context, while the *Polycomb* gene *stricto sensu* will be referred to as *Pc*.

Polycomb and the Hox cluster

Despite the clear transformations displayed by PcG mutants, more than 30 years would pass before a direct role in regulating the homeotic gene complex was formally proven for *Pc* and *Esc* (Lewis, 1978; Struhl, 1981), mutations of which, generated altered RNA expression pattern of many Hox genes (Struhl and Akam, 1985; Wedeen et al., 1986). The way in which *Pc* would exert such a regulation was – and still is – very debated. One fundamental finding in the understanding of Polycomb mode of function is that mutations on PcG factors are additive, such that embryos carrying more than one mutant allele have higher penetrance and more severe transformations (Jürgens, 1985), suggesting that these genes may act together in the regulation of Hox genes expression. Another key step in understanding the Polycomb network is represented by a subtle observation made by G. Struhl and M. Akam in 1985: while studying the expression of *Ubx*, they notice that, around 4-4.5 h after fertilization, wt and esc^{-} embryos display the same pattern of expression, suggesting that *Ubx* is correctly turned on in response to developmental stimulus. However, around 6-7 h post fertilization,

in esc⁻ embryos, *Ubx* expression starts to extend to more anterior parts of the body, reaching the prothoracic segments. Thus, they conclude that the product of esc is dispensable for correct initial *Ubx* pattern establishment, while it is necessary for maintaining its off state in those cells that did not turn it on in the first place (Struhl and Akam, 1985). Importantly, this model will turn out to be true for many other PcG mutants, including *Pc*, *E(Pc)*, *Asx*, *Pcl*, *Psc*, and *Su(z)2* (Soto et al., 1995).

However, the idea that PcG factors were involved in maintenance of a certain state rather than enforcing gene silencing had been around for a while: homeotic defects in PcG mutants show a certain level of variability from embryo to embryo, suggesting some degree of stochasticity in the development of the phenotype (Denell, 1982). Moreover, this variability is present down to the cellular level, with cells in close proximity generating distinct degrees of transformation (Denell and Frederick, 1983; Struhl, 1981), thus suggesting that Polycomb must be involved in the clonal transmission of a given developmental path, hence in its absence, cells become unstable and develop into something else. This theory was further reinforced by the large degree of similarity between the *Pc* gene product and the one of heterochromatin protein 1 (*Hp1*) (Paro and Hogness, 1991). The latter had been already proven to act in the spreading of heterochromatin into ectopically placed euchromatic regions, a phenomenon called position effect variegation (James and Elgin, 1986; Locke et al., 1988). This similarity was instrumental in understanding that PcG-mediated gene regulation had to do with heritable features of the chromatin.

PcG genes and complexes

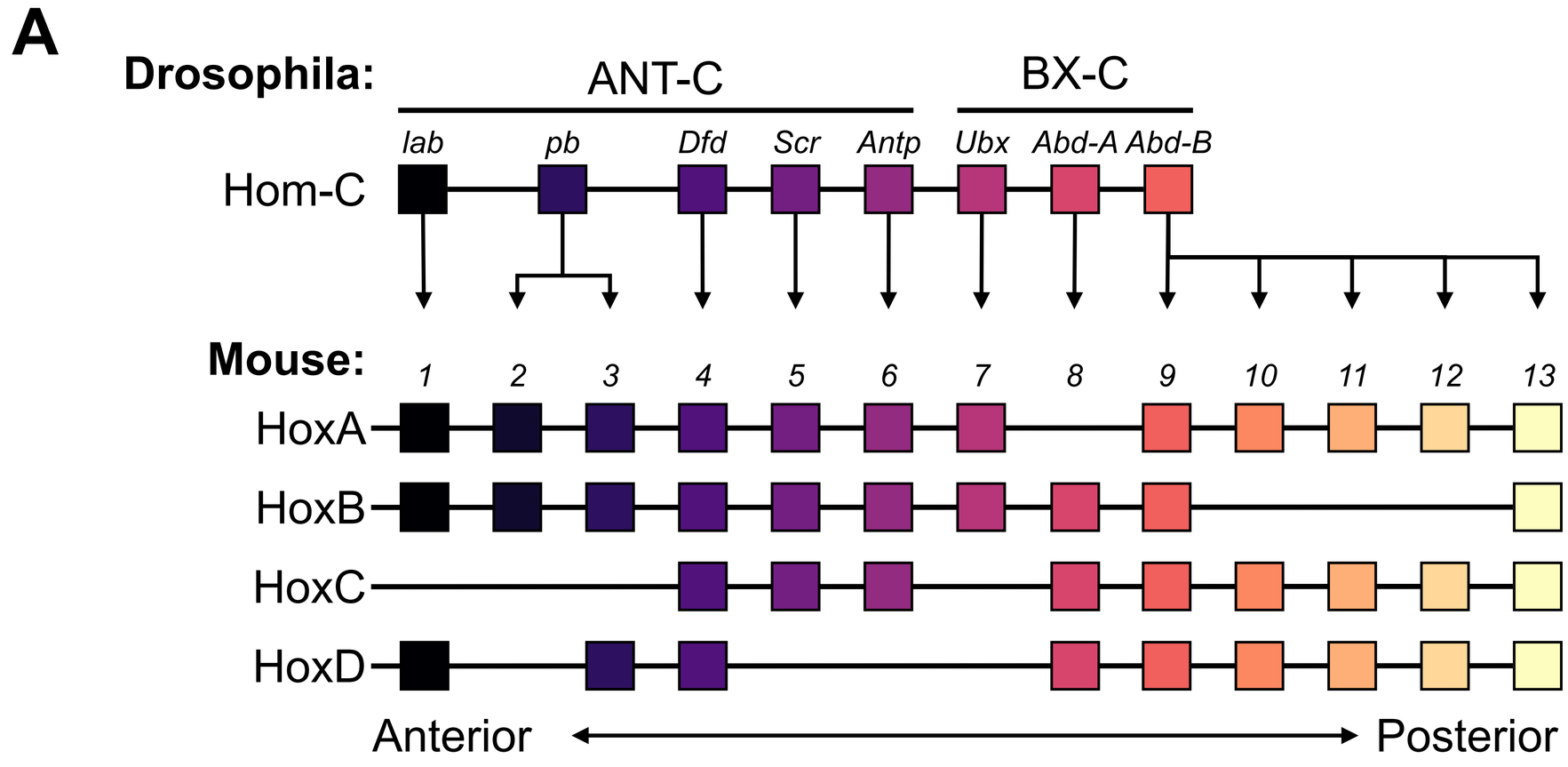
Insights on the molecular features of PcG mechanism of action, came from the evidence that Pc gene product was physically associated with the ANT-C and the BX-C chromosomal regions (Zink and Paro, 1989), and that this association was dependent on its chromodomain (Messmer et al., 1992). Moreover, subsequent studies on the drosophila salivary glands polytene chromosomes,

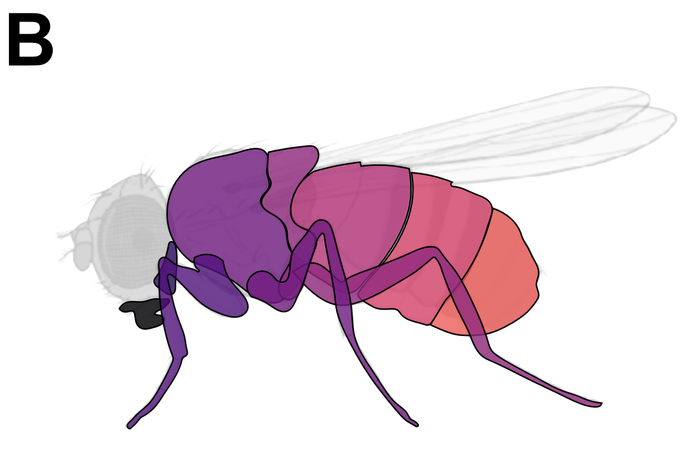
showed that the same binding sites were shared with Ph, Psc and Su(z)2 (DeCamillis et al., 1992; Martin and Adler, 1993; Rastelli et al., 1993). Even more interestingly, their presence on chromatin was dependent on E(z), suggesting that these proteins might be physically or functionally interdependent. The first evidence of a direct protein-protein interaction came for Pc and Ph (Franke et al., 1992) and was later extended, to define two groups of interactors, one composed of *Pc*, *Ph*, *Psc*, *Scm* and *Sce*, and termed Polycomb Repressive Complex (PRC)1 (Kyba and Brock, 1998a; Kyba and Brock, 1998b; Peterson et al., 1997; Saurin et al., 2001; Shao et al., 1999; Strutt and Paro, 1997), and another one containing those of *esc*, *E(z)*, *Caf1-55* (see below), *Su(z)12*, and *Pcl* (Jones et al., 1998; Muller et al., 2002; Ng et al., 2000; O'Connell et al., 2001; Tie et al., 1998; Tie et al., 2003), termed PRC2 (Kuzmichev et al., 2002).
However, this data did not provide an explanation as to why PRC1 binding on the HOM-C cluster and, therefore, its silencing, was dependent on E(z). The answer to this question would come from comparative studies: in 1993, Jones and Gelbart reported that *E(z)* has homology with *Trithorax* (*Trx*), a gene whose mutation in drosophila generated a phenotype opposite to the one of *Pc* (anterior transformation) (Ingham and Whittle, 1980). Like *Pc*, genes that would give a *Trx*-like phenotype where collectively labelled as the Trx group (TrxG) of genes. Similarity between E(z) and Trx proteins regarded an unknown domain that, according to the authors, would be necessary “to interact with a common target” although “it is not obvious what this target might be” (Jones and Gelbart, 1993). One year after, homology was extended to the product of the drosophila gene *Su(var)3-9*, and the homologous domain was named after these three proteins (SET, Suppressor of variegation, Enhancer of zeste, Trithorax) (Tschiersch et al., 1994). The SET domain was found to have homology with the Rubisco LSMT of *pisum sativum*, the first protein described to have methyltransferase activity (Klein and Houtz, 1995; Rea et al., 2000), and from then, many chromatin-related proteins were found to have histone methyltransferase (HMT) activity. Interestingly, most of them showed specificity towards one

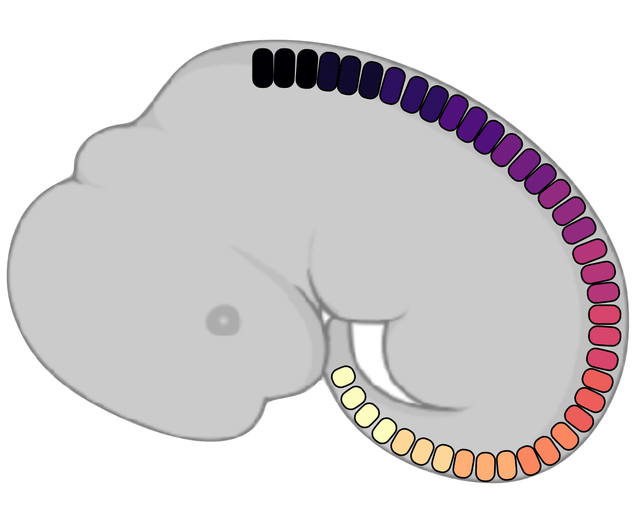
single Lysine residue on the histone N-terminal tails, as for example the yeast Clr4 with specificity for histone H3 lysine 9 (H3K9) (Nakayama et al., 2001), human SET7 with specificity for H3K4 (Wang *et al*, 2001), or drosophila Pr-Set7 modifying H4K20 (Strahl *et al*, 2002). Initial studies failed to detect HMT activity in E(z) (Rea et al., 2000) (probably due to the post-Set auto-inhibitory mechanism, see below), while others proposed that the silencing was mediated by interaction with deacetylases (Tie et al., 2001). However, later it was shown that the esc/E(z) complex is able to mono- di- and try-methylate H3K27 (H3K27me1/2/3), and that this mark is necessary for silencing of the HOM-C complex (Czermin et al., 2002; Muller et al., 2002). The last piece of this puzzle came from the finding that Pc can specifically recognize the H3K27me3 mark, providing a mechanism for PRC1 recruitment, similar to how the chromodomain of Hp1 can recognize H3K9me3 deposited by Su(var)3-9 (Fischle et al., 2003; Min et al., 2003; Rea et al., 2000) – although this separation of function is blurrier in mammals (see below).

Thus, the interdependency of PRC1 and PRC2 for Hox silencing (and the rest of the targets) depends on the presence of a molecular mark, the H3K27me3, that can be passed on through cell divisions and maintain a memory of the chromatin state, via recruitment of PRC1.

However, H3K27me3 is not the only modification marking the chromatin silent state: it had long been known that around 5 to 15% of the histone H2A molecules in higher eukaryotes are ubiquitinated, although a clear link with transcription regulation was missing (Zhang, 2003). In an attempt to define the enzyme responsible for this modification, Wang and colleagues found that *drosophila* dRING (product of *Sce*) and its mouse homologs RNF1/2 (RING1A/B, see below) catalyze H2A mono-ubiquitination at Lysine 118 (H2AK118ub, H2AK119ub in mouse), and that this modification is necessary for the silencing of the homeotic gene *Ubx* (Wang et al., 2004).







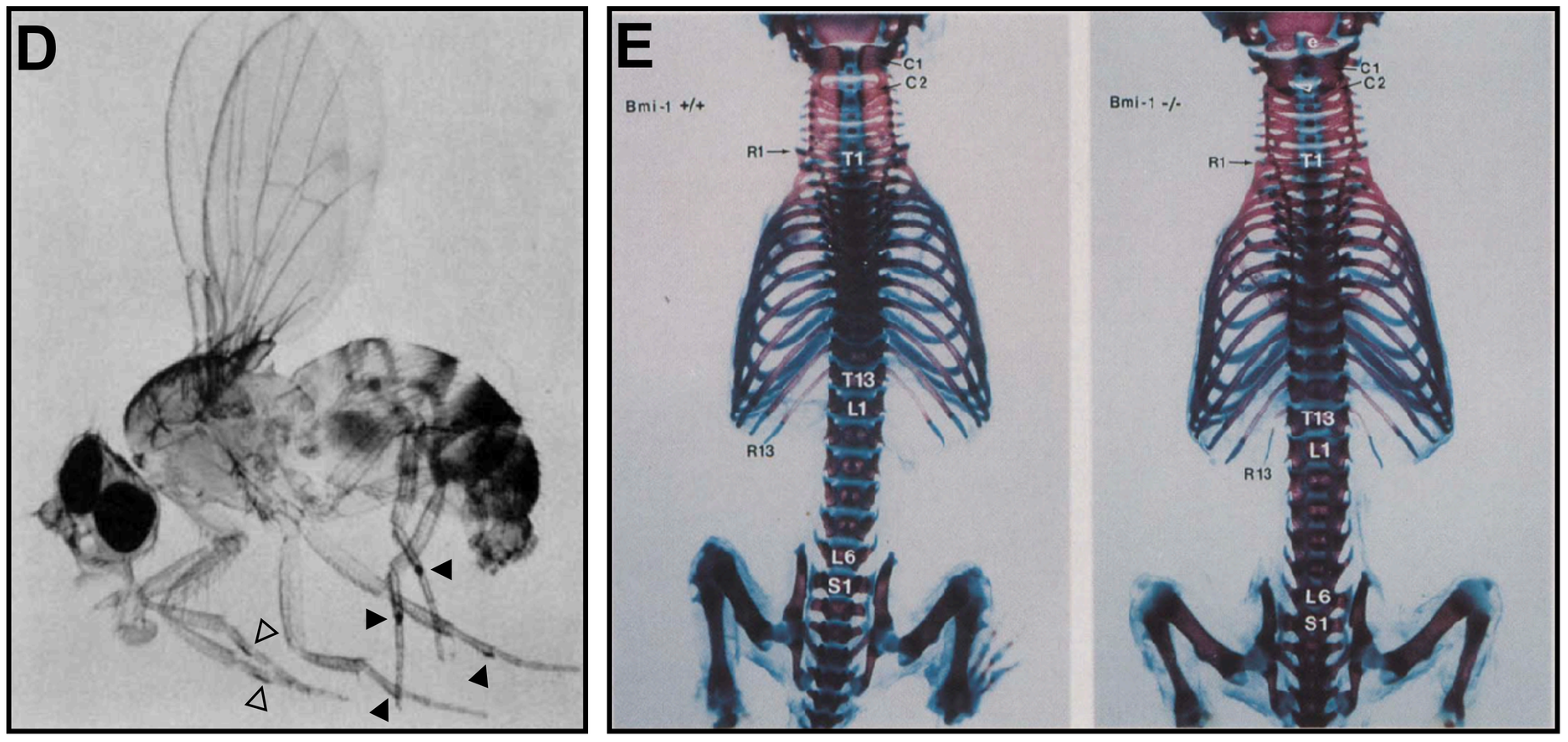


Figure I1 Hox genes collinearity and the homeotic phenotype A scheme of the Hox gene clusters in drosophila (HOM-C) and mouse (HoxA/B/C/D); **B–C** schematic representation of the expression boundaries of the Hox genes in drosophila (**B**) and mouse embryo (**C**, approx. E11.5–12.5)**; D–E** Polycomb-like phenotypes; **D** picture of an *esc* mutant *drosophila*. Open arrowheads indicate sex combs on pro-thoracic legs, while solid arrowheads indicate the appearance of extra sex combs on meso- and meta-thoracic legs, typical of the mutant flies (adapted from Slifer, 1942); **E** axial skeleton analysis of WT (left) and Bmi1 –/– mice (right). Indicated are the various phenotypes associated with posterior transformation in this mutant: appearance of an extra piece of bone rostral to C1 (e); a broadened atlas (open arrowhead, C1 to C2 transformation); fusion of the C7 rib with the sternum (R1, C7 to T1 transformation); degeneration of the 13th rib (R13, partial T13 to L1 transformation); lateral fusion of the last lumbar vertebra (L6) with the first sacral (S1, partial L6 to S1 transformation).

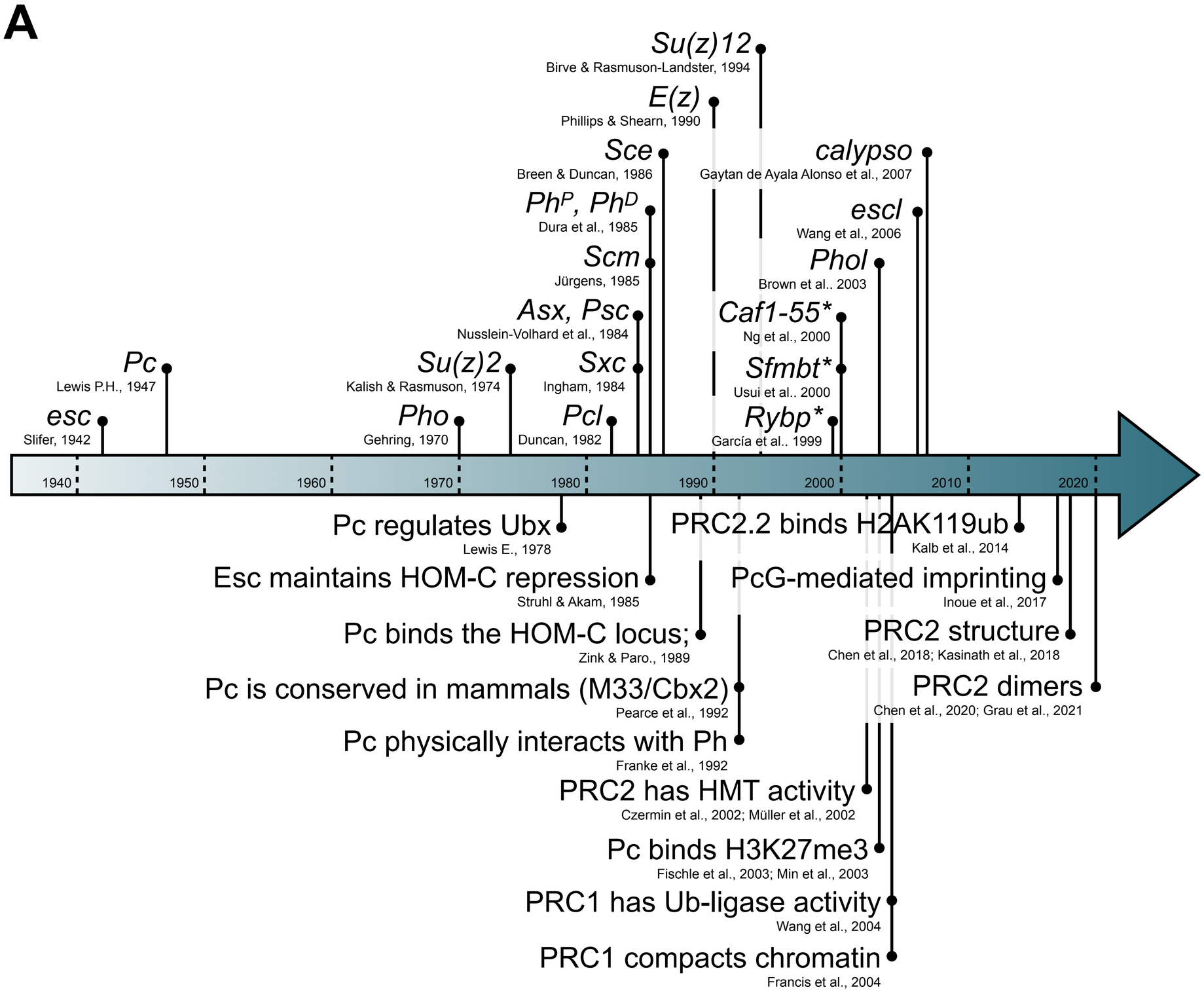


Figure I2 A timeline of the discovery of PcG genes and milestone observations (* these PcG genes were isolated as interactors of known PcG genes); **B–D** Graphical representation of the linear structure of the CBX (**B**) and PHC (**C**), and PCGF (**D**) groups of parlogs in mammals (the length refers to the human protein), including annotated domains. (AT: AT-hook; Chromo: Chromodomain; CaPS: Chromatin compaction and Phase Separation domain; Pc: Pc-box, SAM: Sterile Alpha Motif).

Of note, an additional Polycomb complex has been described, which includes Asx, OGT (the product of *Sxc*) and BAP1 (the product of *calypso*). This complex exerts H2A-specific deubiquitinase activity, endowed in the ubiquitin carboxy-terminal hydrolase domain contained in BAP1 (Scheuermann et al., 2010). For this reason, the complex has been termed Polycomb repressive deubiquitinase (PR-DUB).

PcG conservation and evolution

While most of the research on Polycomb role in the regulation of the Hox genes has been historically linked to the *drosophila* experimental model, early '90s reports showed that many of the PcG genes had counterparts in mouse and human. For example, two genes isolated in screenings involving mouse melanoma and B-cell lymphoma tumorigenicity, namely *Mel-18* and *Bmi1* were found to be homologous to *drosophila Psc* and *Su(z)2* (Brunk et al., 1991; Ishida et al., 1993; van Lohuizen et al., 1991). Soon, homologs of *Pc*, *Ph*, and *Sce* were also found (see below), confirming the functional and structural conservation of the PRC1 complex (Levine et al., 2002). Conservation of *esc* and *E(z)* (*Eed* and *Enx1*/*Ezh2*, respectively) confirmed the conservation of PRC2 too (Hobert et al., 1996; Kuzmichev et al., 2002; Shumacher et al., 1996).

While being highly conserved, PcG genes have evolved through at least a couple of gene duplications happening in vertebrates and in mammals (Whitcomb et al., 2007), that provided the latter with more than one homolog for each drosophila gene, suggesting the evolution of redundant or specialized functions.

The best example of this is probably represented by *Pc* and the abovementioned *Hp1*: these two genes in drosophila have homology to eight mouse and human genes, all named after their chromobox (Cbx) domain. Three of them are thought to be closely related to *Hp1* (*Cbx1*, *Cbx3*, and *Cbx5*, also known as *Hp1*β, *Hp1*γ, and *Hp1*α respectively), while the other five evolved from *Pc* (*Cbx2*, *Cbx4*, *Cbx6*, *Cbx7*, and *Cbx8*, Fig. I2B) (Alkema et al., 1997b; Bardos et al., 2000; Hemenway et al., 2000; Pearce et al., 1992).

Importantly, while mammalian HP1 homologs maintain a strong preference for H3K9me3 over H3K27me3 binding, Pc homologs seem to be more promiscuous, although their occupancy on chromatin generally correlates with that of the H3K27me3 mark (see below). Similar to *Pc*, *Sce*, also known as *Really interesting new gene* (*Ring*) has two homologs in mammals, *Rnf1* (or *Ring1a*) and *Rnf2* (or *Ring1b*) (Satijn et al., 1997a; Schoorlemmer et al., 1997). Moreover, the *drosophila* Ph locus, that harbours two closely related genes, named proximal (Ph^P) and distal (Ph^D) (Dura et al., 1985; Dura et al., 1987), has homology with three mammalian genes, named Phc1–3 (Fig. I2C) (Hemenway et al., 1998; Nomura et al., 1994). Finally, Apart from *Mel-18* and *Bmi1*, mammals have 4 more homologs for the drosophila *Psc*/*Su(z)2*, those are the *Nervous system Polycomb-1* (*NSPc1*), the *Mel-18 and Bmi1-like RING finger protein* (*MBLR*), and the *Ring finger protein 3* and *159* (*RNF3*/*Pcgf3*, *RNF159*/*Pcgf5*, Fig. I2D) (Akasaka et al., 2002; Nunes et al., 2001). A general trend towards evolution of more than one homolog per each drosophila PcG gene is also observed for genes belonging to PRC2.

Subsequent comparative studies revealed that PcG genes show a high degree of conservation across all metazoans (Schuettengruber et al., 2017; Whitcomb et al., 2007), with core PRC2 components being present in nearly all eukaryotes, with the notable exception of *S. cerevisiae* and *S. pombe*, likely due to a secondary loss (Shaver et al., 2010; Zhao et al., 2021). Not only PcG genes are conserved, but also their target genes are, highlighting the fundamental role played by this network in metazoans (Schuettengruber et al., 2007).

Homeotic transformations in PcG mutant mice

The increased complexity in mammalian PcG genes is mirrored by that of Hox genes: while drosophila has 8 Hox genes controlling the segmentation of about 16 axial segments, mouse displays 39 Hox genes distributed over 4 gene clusters that specify the identity of about 60 somites (Fig. I1A,C) (Theiler, 1989) following exactly the same co-linearity principle (Burke et al., 1995; Wellik, 2007). In

mouse PcG mutants, homeotic transformations manifest in the form of rostral or caudal shift of identity in the vertebrae, a phenomenon firstly observed by van der Lugt and co-workers in the *Bmi1* KO mouse model (Fig. I1E) (van der Lugt et al., 1994). This phenotype is mostly observable between those adjacent vertebrae that normally show morphological differences, for example between a cervical and a thoracic vertebra, with the appearance of an ectopic rib on the last cervical vertebra or vice versa. Similar to *drosophila*, distinct PcG mutants in mouse show variable degrees of penetrance in the transformation, and they display dosage effect, as observed in *Bmi1*/*Mel-18* (Akasaka et al., 2001), *Mel-18*/*Rnf2* (Suzuki et al., 2002), *Phc1*/*2* (Isono et al., 2005), and Phf19/Epop double mutants (Chammas et al., unpublished).

RNA in situ hybridization experiments in PcG mutant mice showed that the anterior boundary of expression of many hox genes is shifted rostrally or caudally by one or two somites, thus determining the onset of the features that are typical of the adjacent one. Importantly, hox genes expression in mouse starts as early as embryonic day 7.0 (E7.0) in the anterior area of the primitive streak (Deschamps and Wijgerde, 1993; Forlani et al., 2003), and misexpression of hox genes in PcG mutants is already visible by E9.0-9.5 (Akasaka et al., 2001; Bel et al., 1998) suggesting that during this window of time, Polycomb-mediated regulation is instrumental to ensure the correct expression pattern. Indeed, this time corresponds to the onset of somitogenesis, the process through which the caudal part of the embryo, harboring the presomitic mesoderm (PSM), starts to extend caudally, while differentiating rostrally, periodically generating consecutive pairs of somites. Those represent the segmented structure of the mesoderm that will generate the skeleton of the mouse. Importantly, somitogenesis is a timely regulated process, that involve the use of molecular clocks that dictate the time of differentiation of the PSM (Maroto et al., 2012). As time passes, in the PSM, the Hox cluster gets progressively activated in a 3' to 5' linear fashion, such that at each time point this region will acquire expression of new Hox paralogs, a phenomenon called

'temporal collinearity'. Therefore, each consecutive pair of somites that is generated from the PSM during somitogenesis will inherit the expression pattern of the PSM at that time, representing the anterior boundary of expression of that Hox paralog – fixing the temporal collinearity into the spatial collinearity. Molecularly, this process is mediated by the shifting of epigenetic marks along the hox clusters: at each new somite generated from the presomitic mesoderm, the broad H3K27me3 decorating the Hox clusters gets progressively erased, starting from the 3' end, where the first paralogs are located. This is accompanied by newly deposition of H3K4 methylation and H3 acetylation, together with increased occupancy of the RNA Pol II, leading to activation of new Hox paralogs (Soshnikova and Duboule, 2009). Therefore, the most likely hypothesis is that hox misexpression in PcG mutants arise from a defect in this timely process of H3K27me3 erasure, with faster erasure giving rise to anticipated upregulation of Hox genes and, therefore, rostral shift of their expression boundary, giving rise to posterior transformation (this is the case for most of the PcG mutants). Vice versa, lagging erasure, would result in a delay in Hox activation, shifting the boundary of expression caudally, and generating anterior transformations (the case for *Aebp2*, *Rnf1*, and *Pcgf6* mutants, (del Mar Lorente et al., 2000; Endoh et al., 2017; Grijzenhout et al., 2016)). However appealing, this hypothesis has not been formally proven for any PcG KO mouse model.

Polycomb complexity in the present days

Since the discovery of two functionally and structurally distinct Polycomb complexes, protein interaction studies have helped expand the number of PcG genes, and nowadays we tend to include not only those genes originally isolated as polycomb-like mutants, but also homologous genes and factors identified as physical interactors of those genes. An example is given by the *chromatin assembly factor 1, p55 subunit* (*Caf1-55*, also known as *p55* or *Nurf55*) whose product is a WD-repeat containing protein that is able to bind histone tails through its acidic pocket (Murzina et al., 2008; Schmitges

et al., 2011). Caf1-55 was identified as a direct interactor of E(z)/esc (Muller et al., 2002): in Caf1-55 mutant flies (and mice), homeotic transformations are not observed, since more pronounced defects arise. This is likely due to the fact that, apart from PRC2, Caf1-55 gene product (and its mammalian counterparts Rbbp4/7) takes part in many other multi-protein complexes such as the nucleosome remodelers SIN3, NURD, and NURF but also the Linker of Nucleoskeleton and Cytoskeleton (LINC) complex (Hauri et al., 2016), suggesting that its nucleosome-binding activity can be applied to several different nuclear tasks.

In the early 2000s, proteomic studies started to identify a large number of PRC1 and -2 protein interactors in mouse and human models, and began to classify them into distinct complex subtypes (Beringer et al., 2016; Conway et al., 2018; Dietrich et al., 2007; Dou et al., 2005; Elderkin et al., 2007; Gearhart et al., 2006; Hauri et al., 2016; Levine et al., 2002; Ogawa et al., 2002; Sanchez et al., 2007; Tahiliani et al., 2007; Trojer et al., 2011; Vandamme et al., 2011; Wiederschain et al., 2007). Both PRC1 and -2 share a common organization, whereby a core of fundamental factors endowing the complex with their respective enzymatic activity, is supported by auxiliary factors responsible for the recruitment of the complex to chromatin.

The current classification of PRC1 into six subtypes was established by Gao and colleagues (Gao et al., 2012) (Fig. I3A): according to them, the mammalian PRC1 complex is built around a core composed of one of the two *Sce* paralogs (RNF1/RING1A or RNF2/RING1B) in association with one of the six *Psc* paralogs, named *Polycomb group RING finger* (*Pcgf*) proteins (for they possess an inactive RING domain), namely NSPC2/PCGF1, MEL-18/PCGF2, RNF3/PCGF3, BMI1/PCGF4, RNF159/PCGF5, and MBLR/PCGF6. The RING and PCGF groups of paralogs share a certain degree of homology between them, both displaying an N-terminal RING (hence their names) and a C-terminal Ring-finger and WD40 associated Ubiquitin-Like (RAWUL) domain, which is-similar to the RING domain but lacks the characteristic diglycine

motif that is necessary to catalyse ubiquitin-ligase to the lysine substrate (Sanchez-Pulido et al., 2008). The interaction between RNF1/2 and PCGF1–6 is mediated by heterodimerization of their RING fingers (Fig. I2A) (Buchwald et al., 2006), providing a mechanism for mutual exclusivity of the 6 PCGF paralogs. Instead, the RAWUL domains are necessary for the interaction with non-core components (Fig. I2A): the RING1 and YY1-binding protein (RYBP), its homolog YY1-associated factor 2 (YAF2), and the five CBX paralogs (CBX2,4,6–8) compete in a dose-dependent manner for the binding to RNF1/2 RAWUL domain (Gao et al., 2012; Garcia et al., 1999; Wang et al., 2010). On the other hand, the RAWUL domain of each PCGF shows binding specificity towards one gene or a group of paralogs: MEL-18/PCGF2 and BMI1/PCGF4 interact with PHC1–3; NSPC1/PCGF1 interacts with *BCL-6 corepressor* (BCOR) or its homolog *BCOR-like* (BCORL) (Junco et al., 2013; Sanchez et al., 2007); the RAWUL domains of PCGF3 and -5 both bind the *DDB1- and CUL4-associated factor 7* (DCAF7) also known as WDR68 due to its WD repeat (Hauri et al., 2016), whereas the one of PCGF6 binds another WD repeat containing protein, WDR5 (Sanchez-Pulido et al., 2008). Notably, WDR5 is usually classified as a TrxG protein as it takes part in the formation of the Mixed Lineage Leukemia (MLL) complex (also known as COMPASS) responsible for the deposition of the H3K4me3 mark, associated with active transcription.

Importantly, the inclusion of PCGFs and CBX/RYBP/YAF2 is not independent, reducing the field of possible complex compositions: interaction with MEL-18 and BMI1 favors that of CBXs, whereas that with PCGF1/3/5/6 stabilizes RYBP/YAF2 interaction. Importantly, the complex containing RNF1/2, MEL-18/BMI1, CBXs and PHCs (PRC1.2/4) is usually referred to as the canonical PRC1, for this is the complex originally purified in mammals (Levine et al., 2002), and it is evolutionarily homologous to the drosophila PRC1 (*Sce*, *Psc*, *Pc*, *Ph*). This complex only contains four factors and a relatively simple composition, with paralogs competing for their inclusion in the complex. Instead, the complex formed by the inclusion of RYBP

and PCGF1/3/5/6 is referred to as non-canonical PRC1. Composition of this complex depends on the primary interactor of the corresponding PCGF: the PCGF1-BCOR module mediates the interaction with the *Lysine Demethylase 2B* (KDM2B), the *Ubiquitin Specific Peptidase 7* (USP7), and the *S-phase Kinase Associated Protein 1* (SKP1) to compose the PRC1.1 complex (Junco et al., 2013; Rose et al., 2016; Wong et al., 2016). Importantly, a similar complex has been found in *drosophila*, containing dRING (the product of *Sce*), PSC and dKDM (the drosophila homolog of KDM2B), known as the dRING associated factors (dRAF) complex (Lagarou et al., 2008); the PCGF3/5-DCAF7 module interacts with the *Casein Kinase 2* group (CSK21/2/B), one of the three *Fibrosin* paralogs (FBRS/FBSL/AUTS2), and the *Upstream Transcription Factor* (USF)1/2, forming two functionally and compositionally homologous complexes termed PRC1.3/5 (Gao et al., 2012; Scelfo et al., 2019); finally, the PCGF6-WDR5 module serves as an interaction platform for the many PRC1.6 interactors. Those include two pairs of transcription factor heterodimers, namely the DP-1/E2F6, and the MAX/MGA, the H3K4-specific *Lysine Demethylase 5C* (KDM5C/JARID1C), the *Histone Deacetylase* (HDAC)1/2, an H3K9 methyl transferase G9A and its reader CBX3, and the *Lethal(3)Malignant Brain Tumor-Like Protein 2* (L3MBTL2), a protein that is able to interact with both H3K9me1/2 and H4K20me1/2 (Guo et al., 2009; Ogawa et al., 2002; Tahiliani et al., 2007; Trojer et al., 2011).

On the other hand, PRC2 subtypes are defined by accessory proteins, as proposed by our lab (Beringer et al., 2016): the core of the complex is composed of four stoichiometric components, that are homologous to the drosophila PRC2, namely, EZH1/2, SUZ12, EED, RBBP4/7 (Fig. I3B). In mESCs, association of these factors with one Polycomb-like (PCL)1–3 (also known as PHF1, MTF2 and PHF19, respectively) and the Elongin BC and Polycomb repressive complex 2-associated protein (EPOP) defines the PRC2.1 subtype (Fig. I2B); instead, inclusion of the adipocyte enhancer-binding protein 2 (AEBP2) and the Jumonji and AT-rich interaction domain 2

(JARID2) defines the PRC2.2 (Beringer et al., 2016; Conway et al., 2018; Liefke et al., 2016). In addition to these factors, other proteins have been described to interact with PRC2 in certain situations: this is the case of the Elongin B/C (ELOB/C) complex, whose interaction with PRC2 is mediated by EPOP via an N-terminal BC box. Although the functional relevance of this interaction for mESCs is still not clear, in cancer cell lines, mutations disrupting EPOP binding to ELOB/C impair cell proliferation (Liefke et al., 2016). Another factor termed the PRC2-associated LCOR/L isoform (PALI1/2) was found to interact with PRC2 in the place of EPOP, especially in differentiating mESCs (Conway et al., 2018; Hauri et al., 2016; Kloet et al., 2016). This protein is an internal substrate for methylation at residue K1241 by PRC2 itself, which in turn enhances EZH2 catalytic activity (Zhang et al., 2021). Recent reports have described another PRC2 associated factor whose role is to restrict PRC2 catalytic activity, hence its name of Ezh2 inhibitory protein (EZHIP, also known as catalytic antagonist of Polycomb, CATACOMB) (Jain et al., 2019; Piunti et al., 2019; Ragazzini et al., 2019). This protein has been associated with lower H3K27me3 deposition during sperm and oocyte maturation, as well as in posterior fossa ependymoma group A (PFA). It does so by binding to the catalytic pocket, mimicking the histone tail substrate, a mechanism that is remindful of the H3K27M-mediated EZH2 inhibition in diffuse intrinsic pontine glioma (DIPG) (Aranda and Di Croce, 2019).

Polycomb complexes recruitment to chromatin

Although core components of both PRC1 and PRC2 alone retain their respective enzymatic activities, accessory factors are collectively required for PRC1 and PRC2 binding to chromatin (Fig. I3,4A) (Hojfeldt et al., 2019; Hojfeldt et al., 2018; Huseyin and Klose, 2021). Similar to *drosophila*, chromatin binding of mammalian canonical PRC1 depends on CBX-mediated H3K27me3 recognition (Morey et al., 2012; Tavares et al., 2012). However, recognition of the epigenetic mark is supported by concomitant binding to DNA by

the AT-hook motif adjacent to the Chromo domain (Fig. I3A,4A) (Connelly et al., 2019; Tardat et al., 2015; Zhen et al., 2016).
Instead, non-canonical PRC1 complexes exclusively rely on DNA-binding factors for their recruitment to target genes: KDM2B, through its CxxC domain, is responsible for targeting of PRC1.1 at hypomethylated CpG islands (CGIs, Fig. I4A) (Farcas et al., 2012; He et al., 2013; Koyama-Nasu et al., 2007; Wu et al., 2013b). PRC1.6 recruitment depends on at least two independent mechanisms: one is mediated by the MGA/MAX heterodimer that can recognize the Myc E-box (CACGTG) and T-box (AGGYGYGAGG) (Hurlin et al., 1999), and accounts for most of the PRC1.6 binding sites on chromatin in mESCs (Stielow et al., 2018). The second mechanism involves the recognition of the E2F sequence (GCGGGAA) by the DP-1/E2F6 heterodimer (Fig. I4A) (Gaubatz et al., 1998) that recruits PRC1.6 to a subset of loci mostly related with cell cycle (Stielow et al., 2018). Notably, the other two epigenetic readers in the complex, CBX3 and L3MBTL2, were proven to be not necessary for targeting, although L3MBTL2 is necessary for silencing of PRC1.6 target genes (Huang et al., 2018; Qin et al., 2012). Instead, chromatin binding of PRC1.3/5 has been recently shown to be dependent on the recognition of a variant E-box DNA motif (TCACGTG) by USF1/2 (Scelfo et al., 2019).
Similar to PRC1, distinct PRC2 subtypes have different ways of engaging chromatin: of note, the core component EED has the ability to interact with chromatin, at the sites modified by PRC2 (Fig. I3B) (Margueron et al., 2009). However, this interaction does not mediate PRC2 recruitment, rather it promotes positive feedback via allosteric activation of the EZH2 catalytic pocket (Justin et al., 2016; Lee et al., 2018), a mechanism that has been proposed to be responsible for the propagation of the H3K27me3 mark along the chromatin fiber (see below). Other core members also display a certain affinity to nucleosomes (e.g., the CXC domain of EZH2, the ZnnF of SUZ12, the acidic pocket of RBBP4, see below), however, in the context of the PRC2 complex this is either masked or needed for local

stabilization of the interaction with the nucleosome substrate (Chammas et al., 2020).

PRC2.1 binding to chromatin fully depends on the presence of PCL1–3 (Healy et al., 2019). The extended homology (EH) domain contained in all three paralogs recognizes and binds unmethylated CpG dinucleotides (Fig. I3B,4A) (Choi et al., 2017; Li et al., 2017), a feature that is strongly associated with PRC2 chromatin binding sites genome-wide (Tanay et al., 2007). In addition to CpG binding, PCL proteins can bind specifically the H3K36me3 mark via their PHD domain (Ballare et al., 2012; Brien et al., 2012). However, given the association of this feature to actively transcribed genes, it is not clear how this contributes to PRC2 recruitment.

On the other hand, PRC2.2 is recruited to chromatin via recognition of the H2AK119ub mark by AEBP2 and JARID2 (Fig. I3B,4A) (Blackledge et al., 2020; Healy et al., 2019; Kalb et al., 2014; Kasinath et al., 2021). Importantly, presence of this mark is crucial, as KO of *Rnf2* or its catalytically dead mutation completely abolish JARID2 and AEBP2 binding to chromatin (Blackledge et al., 2020). This interaction happens through a specialized N-terminal ubiquitin interacting motif (UIM) of JARID2 and two of the three Zn fingers (ZNF) of AEBP2 (Kasinath et al., 2021). Importantly, this interaction does not only target the complex to chromatin but also induces EZH2 catalytic activity, resulting in more H3K27me3 deposition (Kasinath et al., 2021). JARID2 can also be methylated by EZH2 at K116 (Sanulli et al., 2015). This modification is recognized by EED, similar to the H3K27me3 peptide, stimulating the allosteric activation of EZH2 HMT activity (Kasinath et al., 2018; Sanulli et al., 2015).

Historically the H3K27me3-mediated PRC1 targeting has been given the name of “classical” model of PcG recruitment, given that this is the first mechanism for PRC1 targeting that was discovered and it is conserved from *drosophila* (Cao et al., 2002; Rastelli et al., 1993). However, the growing body of evidence pointing towards a PRC2-independent PRC1 recruitment eventually conveyed into an “alternative” model (Schoeftner et al., 2006; Tavares et al., 2012).

Nowadays, these two models have been revisited, in the light of the mechanisms described before, and can be considered as two complementary ways that serve for the reciprocal reinforcement of PRC1 and PRC2 targeting to most of its chromatin targets (Fig. 4A). As a result of this, canonical and non-canonical PRC1 subtypes share most of their targets with PRC2.1 and PRC2.2 (Blackledge et al., 2020; Fursova et al., 2019; Healy et al., 2019; Hojfeldt et al., 2019).

To add to this complexity, RNA has been shown to play a role in Polycomb complexes targeting to chromatin: despite no PRC2 factors displaying any canonical RNA binding motif, PRC2 can interact with RNA in a non-sequence-specific manner, although preferring GC-rich sequences and G-quadruplexes (Cifuentes-Rojas et al., 2014; Davidovich et al., 2013; Wang et al., 2017). Molecularly, RNA binding does not involve any specific protein domain, rather it happens on dispersed amino acid patches on the surface of PRC2, including both core and accessory factors (Long et al., 2017; Zhang et al., 2019). While most of the reports point towards an antagonistic effect of RNA in PRC2 binding to chromatin (Beltran et al., 2019; Beltran et al., 2016), recent data provided by the labs of Thom Cech and Roberto Bonasio, suggest that actually RNA might be necessary for bridging PRC2 interaction with chromatin (Long et al., 2020). We commented these findings, in the general context of the role of RNA in PRC2 recruitment, in a *News & Views* article in Nature Genetics (Mocavini and Di Croce, 2020).

Gene silencing by PRC1 and PRC2

Notably, canonical and non-canonical PRC1 complexes are not only compositionally but also catalytically different, and this depends on the influence of PCGFs: interaction of RNF2 with MEL-18 or BMI1 results in attenuated E3 ubiquitin ligase activity, a mechanism that is conserved from drosophila, and involves the formation of an auto-inhibited complex (Taherbhoy et al., 2015). Instead, PCGF1/3/5/6, all stimulate H2A ubiquitination (Rose et al., 2016; Taherbhoy et al., 2015). Moreover, RYBP is able to bind the H2AK119ub mark

through its Npl4 Zinc Finger (NZF, Fig. I3A) domain, which results in further catalytic activation of RNF1/2 (Arrigoni et al., 2006; Rose et al., 2016; Zhao et al., 2020). These results suggest that non-canonical PRC1 subtypes are more catalytically active on chromatin. This data is in line with recent reports showing that most of the H2A ubiquitination in mESCs is dependent on non-canonical PRC1 (Blackledge et al., 2014; Fursova et al., 2019). Importantly, H2AK119ub deposition by PRC1.1 and PRC1.6 is fundamental for Polycomb-mediated gene silencing at non-Hox targets (Blackledge et al., 2020; Endoh et al., 2012). During the years, several hypotheses have been put forward to explain the mechanism through which H2AK119ub prevents active transcription at target genes, those include preventing H3K4 methylation at TSS, resulting in impaired RNA Pol II recruitment (Nakagawa et al., 2008), or preventing RNA Pol II release (Stock et al., 2007), possibly due to impaired recruitment of the FACT histone chaperone (Zhou et al., 2008). Regardless of the mechanism involved, recent reports have demonstrated that the presence of RNF2 and H2AK119ub at promoters, constrains transcriptional burst frequency at the TSS of target genes (Dobrinic et al., 2021), demonstrating that this mark has a direct role in gene silencing. On the other hand, H3K27me3 deposition by PRC2 is not a cause but rather a consequence of transcription shut down (Dobrinic et al., 2021; Hosogane et al., 2013; Riising et al., 2014; Yuan et al., 2012), although prolonged EZH2 inhibition can lead to irreversible transcriptional changes (Holoch et al., 2021), demonstrating that PRC2 is necessary for maintenance of epigenetic memory.

Canonical PRC1 mediates gene silencing memory through an indirect mechanism. although not completely clear, this mechanism likely involves the physical compaction of the chromatin fiber, a property that has long been associated with PRC1: early reports on reconstituted (canonical) PRC1 complex from *drosophila*, showed that this is able to compact nucleosomal arrays *in vitro*, independently of its H2A-ubiquitination activity (Eskeland et al., 2010; Francis et al., 2004). This phenomenon has been proposed to be

mediated by PHC1–3: these proteins contain a Sterile Alpha Motif (SAM, Fig. I2B) that is able to oligomerize forming a helical shape and bringing together distinct PRC1 complexes (Isono et al., 2013; Kim et al., 2002). Upon immunofluorescence visualization, Polycomb factors display a typical pattern of nuclear speckles (Alkema et al., 1997a; Buchenau et al., 1998; Messmer et al., 1992), that was named 'Polycomb bodies' (PBs) (Saurin et al., 1998). Recent reports performing single particle tracking have revealed that the nucleus of mESCs contains around 100 PBs, accounting for approx. 2% of the total RNF2. These spots contain preferentially canonical PRC1, and PRC2, while the rest of (non-canonical) PRC1 is dispersed into smaller/milder spots on chromatin (Blackledge et al., 2020; Huseyin and Klose, 2021). Notably, mutation of the SAM domain of PHC2 results loss of PBs (Isono et al., 2013), suggesting that PHC-mediated clustering is a driver of subnuclear localization of PRC1. However, other evidence suggest that this phenomenon might be also mediated by CBX proteins: a disordered positively charged region contained in CBX2 may be involved in determining chromatin compaction and phase separation (CaPS, Fig. I2A), resulting in the formation of such droplets *in vitro* and in the nucleus (Grau et al., 2011; Kundu et al., 2017; Plys et al., 2019). At the genome-wide level, chromatin compaction by PHCs and CBXs results in long range topological interactions (Boyle et al., 2020). Importantly, mouse models bearing loss of function mutations of either the PHC2 SAM domain (L307R) or the CBX2 CaPS domain (mutation of positively charged residues into alanines), both result in posterior transformation (Isono et al., 2013; Lau et al., 2017) indicating that both mechanisms are fundamental for correct maintenance of the inactive chromatin state, especially at the Hox clusters.

This chromatin conformation-driven regulation preferentially drives gene silencing at regions characterized by high Polycomb occupancy and vast deposition of the H3K27me3 mark, known as "Broad domains" (Boyle et al., 2020; Healy et al., 2019; Hojfeldt et al., 2019). These regions are characterized by a high degree of interplay between PRC1 and PRC2, through both the classical and the

alternative recruitment models, and are therefore, more stably silenced (Healy et al., 2019).
The scheme emerging from all this data seems to be one with clustered targets stably bound and compacted by the mass-action of several PRC1 and PRC2 components via the formation of PBs (comprising very well-known PcG targets e.g., Hox clusters, *Cbx2/4/8* cluster, *Cdkn2a/b*), against dispersed targets individually regulated via H2AK119ub-mediated gene silencing. However, silencing of broad domains by canonical PRC1 depends on H3K27me3 deposition by PRC2, which in turn needs H2AK119ub deposition by PRC1.1/6 for its recruitment (Blackledge et al., 2020; Hernandez-Munoz et al., 2005). In line with this dependency, PCGF1, PCGF2, and PCGF6 are found together at the majority of RNF2 chromatin targets (Scelfo et al., 2019), including the Hox cluster. Moreover, mouse models lacking PCGF6, E2F6, or the CxxC domain of KDM2B all display homeotic transformations (Blackledge et al., 2014; Endoh et al., 2017; Storre et al., 2002), demonstrating that both canonical and PRC1.1/6 complexes are necessary for correct regulation of the Hox cluster.
A separate mention must be made for PRC1.3/5. These complexes do not contribute to silencing of individual genes, rather they are responsible for pervasive deposition of H2AK119ub over large regions of the chromatin (Fursova et al., 2019; Huseyin and Klose, 2021). This mechanism is particularly relevant for the silencing of inactive X chromosome in female mammals: PRC1.3/5 is firstly recruited on the X chromosome via hnRNPK-mediated recognition of the *X-inactive specific transcript* (*Xist*), initiating the cascade of H2A ubiquitination that leads to subsequent gathering of PRC1/2 complexes (Almeida et al., 2017; Pintacuda et al., 2017). The importance of PRC1.3/5 for dosage compensation is highlighted by the fact that double KO of *Pcgf3* and *Pcgf5* results in female-specific embryo lethality around E9.5 (Almeida et al., 2017).

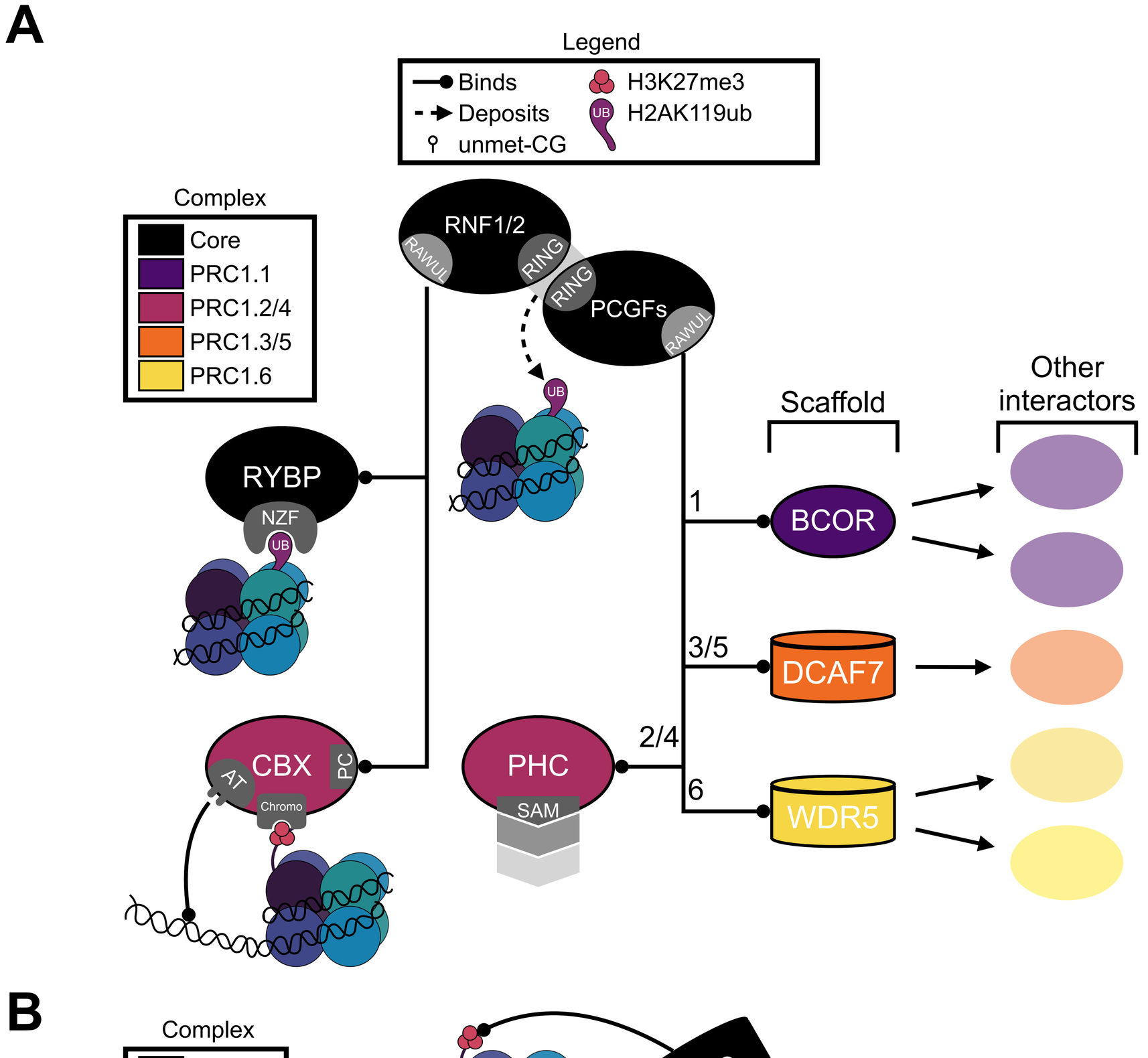
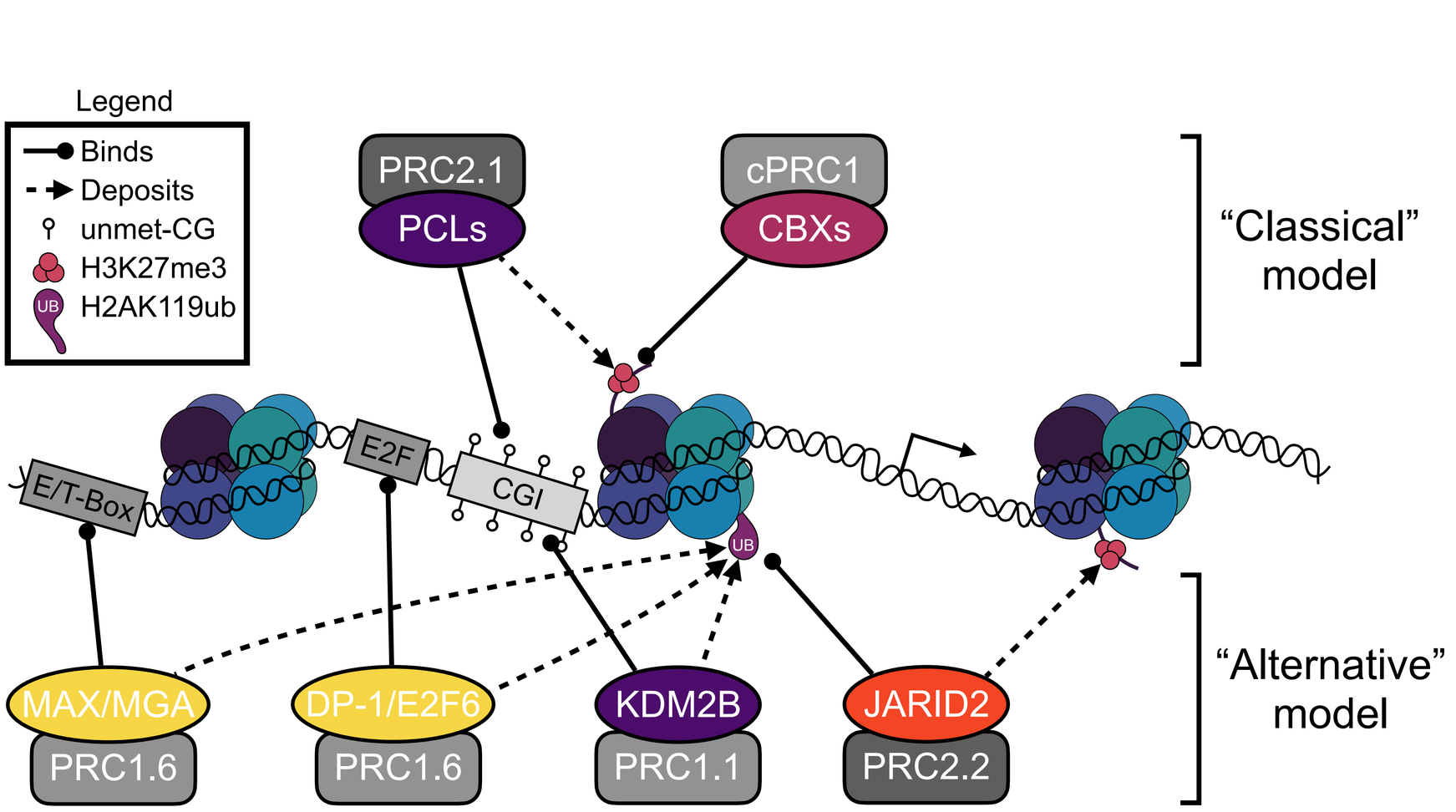


Figure I3 Polycomb factors assemble separate complexes with distinct characteristics
A–B Graphical representation of the main interactions among PRC1 (**A**) and PRC2 (**B**) factors, as well as their main chromatin interacting/modifying domains. Cylinders represent WD propeller proteins. SUZ12 structure is extended to allow visualisation of its protein-interaction domains. Please refer to the main text for domain acronyms.



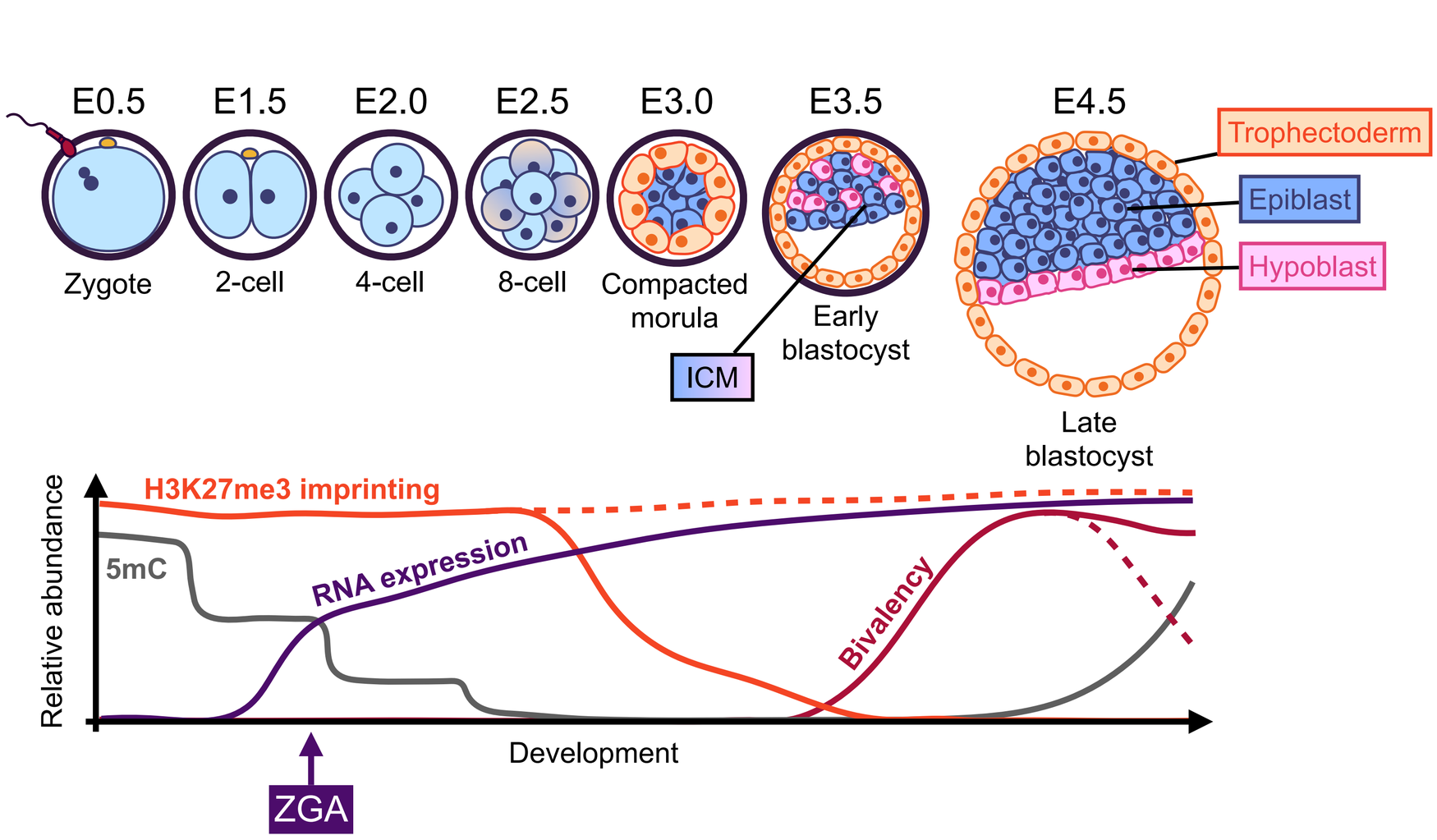


Figure I4 PRC1 and PRC2 cooperate to achieve silencing at target genes A graphical representation of the revisited "classical" and "alternative" models of Polycomb complexes recruitment to chromatin; **B** Representation of the various phases of mouse early development (upper panel) and the chromatin changes that occur during this period (ICM: inner cell mass; ZGA: zygotic genome activation). Note: solid lines refer to the chromatin of embryonic lineages while dashed lines refer to extra embryonic lineages.

The role of the PR-DUB complex in Polycomb-mediated gene regulation has been largely understudied. However, recent reports have shown that this complex is necessary to constrain pervasive H2A ubiquitination by PRC1 genome-wide (Fursova et al., 2021). This activity has the double effect of preventing the silencing of active genes, while maintaining proper targeting of Polycomb complexes at their canonical sites (Fursova et al., 2021). In addition to this genome-wide activity, the PR-DUB has also been shown to maintain transcriptional activity at specific genes by FOXK1/2 mediated targeting (Kolovos et al., 2020). Importantly, PR-DUB also contains the O-linked N-acetylglucosamine (O-GlcNAc) transferase (OGT), the product of the *Sxc* gene (Scheuermann et al., 2010; Sinclair et al., 2009). Glycosylation by OGT has been shown to be necessary for Polycomb-mediated silencing, in a mechanism that might involve direct GlcNAcylation of canonical PRC1 component Ph (Gambetta et al., 2009). Altogether, this data supports an essential role for PR-DUB in aiding Polycomb-mediated gene repression. Of note, recent reports have drawn the attention on possible roles of Polycomb complexes in activation of gene transcription. These non-canonical roles seem to be restricted to specific physiological or malignant contexts. Examples include activation of mammary oncogenic enhancers by canonical PRC1 (Chan et al., 2018), or the targeting of PRC1.1 to active metabolic genes that are important for acute myeloid leukemia (van den Boom et al., 2016); but also activation of neurogenesis-related genes in the mouse brain by an AUTS2-containing PRC1.3 complex (Gao et al., 2014; Liu et al., 2021).

Polycomb in mouse early development

Following fertilization, the newly formed mouse zygote undergoes three rounds of cell division, without increasing its size, which by E2.5 will have given rise to 8 equivalent blastomeres composing a morula. Compaction of the morula (E2.75) results in the polarization of the blastomeres, breaking the symmetry of the embryo into an inner cell mass (ICM) and an outer monocellular layer, the latter

committed to become the extra-embryonic trophectoderm (TE) (Leung and Zernicka-Goetz, 2015). By E3.5, cavitation of the morula will have resulted in the formation of the blastocyst. During maturation of the blastocyst, (E3.5–4.5) another break in the symmetry of the ICM will generate two subpopulations, a monolayer of cells facing the cavity, called hypoblast (also known as primitive endoderm), and the epiblast, on top of it (Fig. I4B).

The whole process of early preimplantation development is accompanied by a profound rewiring of the epigenetic landscape (Fig. I4B, reviewed in (Eckersley-Maslin et al., 2018)). Epigenetic determinants that have been best characterized in this process include DNA methylation, H3K4 methylation, and H3K27 methylation and acetylation. These marks are already present in mature oocytes, although they collectively suffer a progressive erasure and re-establishment: in particular, DNA methylation is globally lost from the genome of embryos following progressive dilution at each round of cell division. Exception to this rule is provided by imprinted genes, of which one of the two alleles will retain DNA methylation throughout the whole process, determining a parent-of-origin specific expression. Of note, recent reports have found that a group of maternal genes undergoes a DNA methylation-independent imprinting mechanism based on Polycomb-mediated silencing (Inoue et al., 2018; Inoue et al., 2017a; Inoue et al., 2017b). Maternal alleles of these genes are marked by H3K27me3 already in the oocyte and maintain the methylation unaltered until the morula stage. Upon trophectoderm specification, these genes will maintain imprinted expression in the extraembryonic tissues, while gradually losing it in the ICM (Inoue et al., 2017a).

Notably, among the genes regulated through this mechanism is *Xist* (Inoue et al., 2017b): indeed, imprinted expression of *Xist* is responsible for paternal-specific inactivation of the X chromosome in the female pre-implantation embryo and in extraembryonic tissues (Huynh and Lee, 2003; Okamoto et al., 2004; Takagi and Sasaki, 1975). Consistent with this, maternal KO of *Eed* results in random X inactivation in these cells (Inoue et al., 2018).

In the zygote and until early 2-cell stage, broad regions of the genome are covered in H3K4me3 and H3K27me3, although these two marks do not overlap with each other and do not correlate with the presence of genes (Dahl et al., 2016; Zheng et al., 2016). At late two-cell stage, around the time when the genome starts to be transcribed, major H3K4 demethylation by JARID1A/B results in the establishment of the canonical H3K4me3 pattern, with peaks in correspondence of the TSS, overlapping with H3K27ac in the case of active genes (Zhang et al., 2016). Instead, H3K27me3 is maintained in a broad gene-distal pattern until the formation of the ICM (Zheng et al., 2016).

In mouse, a molecular description of how the ICM gives rise to the epiblast and hypoblast is still lacking. Recent scRNA-seq studies have highlighted that prior to implantation (E3.5), cells of the ICM co-express markers for both lineages (e.g., *Nanog*/*Sox2* together with *Gata6*/*Sox17*, (Mohammed et al., 2017)). During blastocyst maturation, co-expression of these markers slowly gives way to transcriptional heterogeneity, resulting in sorting of two subpopulations, expressing either the pluripotency or the primitive endoderm master regulators. By the late blastocyst stage, these subpopulations will have adopted the typical spatial distribution of the epiblast and hypoblast, respectively (Rossant and Tam, 2009). Interestingly, PRC1 and PRC2 are necessary to ensure transcriptional heterogeneity in the ICM, to allow for expression of both epiblast- and PrE-priming genes (Illingworth et al., 2016). Depletion of *Eed* or transient catalytic inhibition of EZH2, results in priming towards a PrE fate (Illingworth et al., 2016), suggesting that loss of PRC2 is necessary for PrE commitment. In line with this, extraembryonic endoderm (XEN) cells derived from the mouse blastocyst, show very low levels of H3K27me3 (Rugg-Gunn et al., 2010).

Importantly, the ICM of E3.5 mouse embryos can be extracted from the blastocyst, disaggregated, and grown in vitro indefinitely, constituting the so-called mouse embryonic stem cells (mESCs) model. These cells are pluripotent, in that they can give rise to virtually all cell types of the embryo (Martello and Smith, 2014). For this reason,

they have been intensively exploited as a model system to characterise the molecular changes accompanying cell differentiation. Notably, from a transcriptomic and epigenomic point of view, these cells resemble more the E4.5 epiblast, than the E3.5 ICM, including the pattern of H3K27 methylation: in mESCs, around 3600 promoters are marked by H3K27me3 (Zheng et al., 2016). Importantly, the majority of these are also marked by the H3K4me3 modification, an observation that was first made by Bernstein and colleagues, who termed regions with concomitant presence of these two marks, one normally associated with silent genes, and the other associated with active transcription, as 'bivalent domains' (Bernstein et al., 2006a). Notably, these domains mark the promoters of genes mostly enriched in developmental transcription factors (Ku et al., 2008; Mantsoki et al., 2015; Mas et al., 2018; Mikkelsen et al., 2007). Moreover, upon mESCs differentiation, bivalency is 'resolved' into either activation, marked by H3K4me3 alone, or repression, marked by either H3K27me3 or DNA methylation (Bernstein et al., 2006a; Mikkelsen et al., 2007; Mohn et al., 2008). These observations led researchers to hypothesize that bivalency is necessary to restrict the expression of developmental genes in ESCs while allowing for timely activation or full repression upon exit from pluripotency (Bernstein et al., 2007). This hypothesis was backed up by data showing that bivalent genes are not completely silent, rather they produce low levels of transcripts and are occupied by an unproductive (or 'poised') RNA Pol II form, bearing a phosphorylation mark on the Serine 5 of the C-terminal domain (CTD S5P) (Bernstein et al., 2006a; Brookes et al., 2012; Mikkelsen et al., 2007; Stock et al., 2007). Importantly, the status of the RNA Pol II was found to be dependent on H2AK119ub deposition by PRC1 at bivalent genes (Stock et al., 2007), suggesting that the presence of Polycomb proteins is instrumental to directly constrain transcriptional activity of bivalent promoters. In line with this observation, depletion of *Eed* results in transcriptional activation of bivalent genes in mESCs (Azuara et al., 2006). However, bivalency is not a unique characteristic of mESCs: indeed, during neural differentiation, the number of

bivalent genes is maintained relatively stable, but the H3K27me3 mark is relocated to those genes that will be crucial for further lineage specification during each stage (Mohn et al., 2008). Therefore, this model applies to virtually all stages of potency. Our lab has recently provided an overview on the different aspects of bivalency, including its establishment during development and maintenance during cell cycle (Blanco et al., 2020).

Importantly, the long oversighted dilemma of whether bivalency is really the co-occupancy of the two histone marks or it is just an artifact of cell heterogeneity in ChIP-seq experiments (Voigt et al., 2013), has been resolved by sequential ChIP-seq experiments and in vitro studies, demonstrating that the H3K4me3 and H3K27me3 marks are present at the same time on the same nucleosome, but on opposite H3 tails (Kinkley et al., 2016; Mas et al., 2018; Voigt et al., 2012; Weiner et al., 2016). However, recent reports suggest that co-occurrence of these marks in population ChIP-seq experiments actually represents their alternate presence in each nucleosome, suggesting a bistable, rather than a bivalent state (Sneppen and Ringrose, 2019).

Regardless of the exact mechanism, establishment and enforcement of bivalency by Polycomb complexes is fundamental for mouse early development. In fact, KO of *Eed*, *Ezh2*, *Suz12*, or *Rnf2* results in embryonic lethality around the gastrulation stage (E7.5–8.5) (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004; Voncken et al., 2003). This is the process through which the post-implantation epiblast separates into the three major embryonic layers, namely the ectoderm, mesoderm and endoderm, that will give rise to all the tissues of the body. Importantly, KO of the same factors in mESCs does not alter self-renewal, suggesting that bivalency is not necessary for maintenance of pluripotency. Instead, when challenged to differentiate, these cells fail to correctly activate cell fate-specific transcription factors (Leeb et al., 2010; Morey et al., 2012; Pasini et al., 2007), suggesting that bivalency is indeed important to prime these genes for transcription upon differentiation stimuli. Therefore, loss of bivalency upon PRC1 and PRC2 full

depletion, impairs activation of those developmental genes that specify the three germ layers, ectoderm, mesoderm, and endoderm, ultimately resulting in the arrest of the embryonic development.

CBX proteins in the spotlight

The CBX family of proteins represents an intriguing system to study Polycomb complexity in mammals and has led many researchers to try to understand what the commonalities and specificities of the members of this family are. Major systematic comparative studies of this family include (Bernstein et al., 2006b; Kaustov et al., 2011; Klauke et al., 2013; Morey et al., 2012; van den Boom et al., 2013; Vandamme et al., 2011; Vincenz and Kerppola, 2008). Regarding their amino acid sequence, CBX proteins share a common structure, with an N-terminal part dedicated to chromatin binding, composed of the Chromodomain (>80% conservation), responsible for methyl-lysine binding, and the adjacent AT-hook motif (>80% conservation), a motif that is responsible for binding to the minor groove of the DNA double helix; on the other terminus of the protein, a PC-box (>45% conservation) mediates the interaction with RNF1/2 (Fig. I2B) (Ma et al., 2014; Schoorlemmer et al., 1997; Senthilkumar and Mishra, 2009). Importantly, all three domains are highly conserved from *drosophila* PC, suggesting their importance for CBX proteins function (Whitcomb et al., 2007). However, despite their conservation, these domains display some degree of variability in their binding affinity: for example, while *drosophila* PC shows strong specificity for H3K27me3, the chromodomain of CBX7 can bind both the H3K9 and H3K27 trimethylated marks with virtually no preference *in vitro*, while CBX4 displays preferential binding towards H3K9me3. Instead, CBX6 and CBX8 have very little binding affinity towards H3K27me3 and display no interaction with H3K9me3. Moreover, while CBX2 has high affinity for H3K27me3, its capacity to interact with H3K9me3 is not clear (Bernstein et al., 2006b; Kaustov et al., 2011). However, regardless of their affinity for H3K9me3, CBX proteins interaction with this mark has not been observed *in vivo*, suggesting that this promiscuous binding might be an artefact of the

experimental conditions, or that the chromatin environment, or the rest of the interactors, are needed to achieve specificity towards the H3K27me3 mark. Furthermore, recent studies reported that CBX7 can bind to non-histone trimethylated substrates, including the SET domain bifurcated 1 (SETDB1), and the euchromatic histone lysine methyltransferase (EHMT)1/2, expanding the potential PRC1 targeting mechanisms (Jung et al., 2019).

Differences among CBX protein paralogs also arise in their affinity for the PRC1 core: the RAWUL domain of RNF2 having high affinity for the PC-box of CBX4 and CBX7, and lower affinity for that of CBX2, CBX6, and CBX8 (Wang et al., 2008), suggesting preferential incorporation of the former isoforms with respect to the latter.

Furthermore, reports have described the existence of paralog-specific domains: for example, the abovementioned CaPS domain in CBX2, that mediates chromatin compaction through a PRC1-mediated phase separation (Grau et al., 2011; Plys et al., 2019), in a way that is remindful of that of HP1 in heterochromatin (Larson et al., 2017). This domain is found in the central part of the protein and is composed of a low complexity stretch of positively charged (K, R) residues that are predicted to acquire a disordered conformation. Importantly, these characteristics can also be found in the central part of CBX4 and -8, suggesting that they also share this ability (Kim and Kingston, 2020), although no reports have tested this hypothesis yet. On the other hand, CBX7 lacks most of this region, as it is the shortest paralog, and has a negative net charge, indicating that this protein does not have any phase separation-mediated chromatin compaction ability. Importantly, ectopic introduction of the CaPS domain into the Cbx7 gene in mESCs (where Cbx7 is the only paralog expressed, see below), results in impaired exit from pluripotency due to retained occupancy of PRC1 on developmental genes (Jaensch et al., 2021). This suggests that the CaPS domain is necessary for a stable silencing of target genes, and the lack of this domain in mESCs suggests that these cells need to have a flexible epigenetic state, to allow for effective lineage specification.

Another example of divergent evolution is given by the emergence of an E3 domain in CBX4 (Fig. I2B) (Kagey et al., 2003). This domain can catalyse the addition of a small ubiquitin-related modification (SUMO) to protein substrates. However, only a few specific targets have been described for CBX4, including the C-terminal Binding Protein (CtBP), which binds an amino acid motif contained in CBX4 (PIDLR) but not in other paralogs, the Smad-interacting protein 1 (SIP1), the homeodomain interacting protein kinase 2 (HIPK2), and fundamental chromatin regulators like the CCCTC-binding factor (CTCF), DNMT3A (Kagey et al., 2003; Li et al., 2007; Long et al., 2005; MacPherson et al., 2009; Roscic et al., 2006; Wotton and Merrill, 2007). Interestingly, in some of these cases, SUMOylation happens at Polycomb bodies and enhances the transcription repressive function of the target proteins (e.g., CTCF, CtBP), although this function is independent of canonical PRC1. Moreover, CBX4 E3 ligase activity appears to be particularly important for the DNA damage response pathway: CBX4 is able to SUMOylate its PRC1 binding partner BMI1 at the residue K88, mediating its recruitment to the sites of damage (Ismail et al., 2012). Indeed, PRC1 has a role in ubiquitinating the sites of damage to elicit further recruitment of proteins implicated in the DNA repair pathway (e.g., BRCA1, 53BP1) (Gieni et al., 2011). Another protein essential for DNA resection and homologous recombination, the CTBP-interacting protein (CTIP), is recruited to the sites of damage upon SUMOylation by CBX4 at residue K896 (Soria-Bretones et al., 2017).

CBX8 was also found to have a PRC1-independent role: this protein interacts with a conserved C-terminal domain of the super elongation complex subunit AF9, via a unique motif that maps around amino acids 202 and 333 of CBX8 (Fig. I2B) (Garcia-Cuellar et al., 2001; Hemenway et al., 2001). However, the importance of this interaction in physiological conditions is not fully understood. Nevertheless, it has been observed that this interaction is conserved in the MLL-AF9 fusion oncoprotein. In MLL-AF9-driven leukemogenesis CBX8 is necessary to activate *Hoxa9*, a target gene that is

critical for leukemic transformation by the fusion protein (Tan et al., 2011). A similar function was observed in breast cancer, where CBX8 interacts with the TrxG/PcG protein WDR5, cooperating in the maintenance of H3K4me3 levels at the *Notch* gene and its network of targets, ultimately contributing to mammary tumorigenesis (Chung et al., 2016). Notably, CBX8 acts as an oncogene in several other cancer types, including oesophageal and hepatocellular carcinomas, glioblastoma, and muscle invasive bladder cancer (Li et al., 2013; Tang et al., 2019; Xiao et al., 2014; Yuan et al., 2017).

The other CBX proteins are altered in several cancer types and are associated with a role as tumor suppressors or oncogenes depending on the context. For example, CBX4 can act as a tumor suppressor in cMyc-mediated transformation and in colorectal cancer (Satijn et al., 1997b; Wang et al., 2016), while behaves as an oncogene in hepatocellular carcinoma and osteosarcoma (Li et al., 2014; Wang et al., 2013; Yang et al., 2016). Similarly, while CBX7 is overexpressed in acute myeloid leukemia and luminal A and -B breast cancer subtypes (Jung et al., 2019; Liang et al., 2017), its expression correlates with good prognosis in hepatocellular carcinoma patients (Ning et al., 2018).

Importantly, in physiological conditions, the expression of CBX proteins is highly cell type-specific and regulated throughout the development: Cbx paralogs are organized in the mouse genome in two clusters, one in Chr 11 containing *Cbx2, Cbx4*, and *Cbx8*, and another one in Chr 15 containing *Cbx6* and *Cbx7*. As previously mentioned, the *Cbx2/4/8* cluster is among the so-called broad Polycomb domains in mESCs, owing to the high PcG factors occupancy and the large region covered in PcG-associated histone PTMs. Therefore, this region is repressed, whereas both *Cbx6* and *Cbx7* are expressed in mESCs. In particular, Cbx7 promoter is directly regulated by the pluripotency factor Oct4 (Kagey et al., 2010). CBX6 is rarely found associated with the PRC1 complex, unless exogenously overexpressed (Morey et al., 2012; Santanach et al., 2017), therefore CBX7 is the main paralog included in the PRC1 complex in mESCs (Kloet et al., 2016; Morey et al., 2012). CBX7-PCR1 is mainly found

on developmental genes that are silent in mESCs, including the *Cbx2/4/8* cluster, while CBX6 occupies the promoters of several metabolic genes devoid of PRC1 or -2 (Morey et al., 2012). Upon mESCs exit from pluripotency, Cbx7 is readily downregulated via targeting by the miR-125 and miR-181 family of micro RNAs. Importantly, downregulation of *Cbx7* results in de-repression of *Cbx2*, *Cbx4*, and *Cbx8*, that substitute CBX7 in the complex in differentiated cells (Kloet et al., 2016; Morey et al., 2012; Morey et al., 2015; O'Loghlen et al., 2012). This substitution is accompanied by relocalization of the complex to new targets that do not need to be expressed in that specific lineage: for example, upon differentiation to the mesodermal lineage, CBX2 occupies the promoters of genes related with neurogenesis, together with RNF2, MEL18, and RYBP (Morey et al., 2015). Instead, during embryoid body formation, which give rise to all three embryonic layers, CBX2 and CBX4 repress the expression of germ cell related genes (Morey et al., 2012). Similarly, upon differentiation to the neuroectodermal lineage, a canonical PRC1 comprising either CBX2, CBX4 or CBX8 is located at the promoters of developmental regulators and pluripotency factors *Sox2* and *Nanog* (Kloet et al., 2016; Ning et al., 2017), while in terminally differentiated cortical neurons, Cbx6 becomes the most expressed paralog (Bonev et al., 2017). Importantly, overexpression of CBX4 in mESCs results in increased binding to differentiation-specific loci (Kloet et al., 2016), suggesting that the switch in composition might be enough to determine PRC1 relocalization on chromatin. However, no molecular insights have been provided to explain how CBX proteins could potentially discriminate between target genes. Instead, comparative analysis of chromatin occupancy, suggest highly comparable binding patterns among CBX protein homologs (Pemberton et al., 2014), indicating that the context, but not the paralog, might actually drive differential binding.

Cbx paralogs are also highly regulated in adult stem cells. Hematopoietic stem cells (HSCs) are, in many ways, remindful of the ESCs system: CBX7 is the main paralog incorporated in the PRC1 complex in these cells, and it is necessary to maintain self-renewal and

proliferation of this niche (Klauke et al., 2013; Scott et al., 2007), suggesting that CBX7 plays a more general role in maintenance of pluripotency. Importantly, exogenous overexpression of CBX7 can lead to T-cell leukaemia or lymphoma in mice. Similar to ESCs, upon HSCs exit from pluripotency, CBX7 is downregulated and substituted with differentiation-specific paralogs CBX2/4/8. A CBX8-containing PRC1 is retargeted to myeloid-specific genes in multipotent progenitors and in the lymphoid lineage (Klauke et al., 2013). Moreover, CBX8 is necessary for B-cells germ centre formation (Beguelin et al., 2013; Beguelin et al., 2017; Beguelin et al., 2016; Caganova et al., 2013). CBX2 and CBX4 also play a role in the differentiation of hematopoietic stem cells, by preventing HSCs proliferation and regulating commitment of the hematopoietic progenitors towards the B/T-cells lymphoid lineages (Core et al., 1997; Liu et al., 2013; van den Boom et al., 2013). Our study of the role of PcG proteins in normal haematopoiesis and in leukemic conditions, led to the publication of a review article in the Journal of Cell Biology (Di Carlo et al., 2019).

Outside of the hematopoietic compartment, CBX4 has been implicated in the maintenance of various types of adult stem cells. For example, CBX4 is necessary for maintaining quiescence of human epidermal stem cells as well as to preserve cell identity of dermal and thymic epithelia (Liu et al., 2013; Luis et al., 2011; Mardaryev et al., 2016). Moreover, this protein prevents senescence of mesenchymal stem cells through repression of the rDNA loci (Ren et al., 2019).

Polycomb structural studies

As highlighted by the intricate network of interplay between complexes and even within each complex, the composition and structure of PRC1 and PRC2 complexes is fundamental to understand its function on chromatin. In this direction, structural studies have partially helped understand many of the features described above. Notable examples are the first crystal structures of Pc bound to H3K27me3 (Fischle et al., 2003; Min et al., 2003), the mutual

exclusivity between RYBP and CBX7 for RNF2 binding (Wang et al., 2010), or the minimal RNF2-BMI1 module ubiquitinating a nucleosome (Bentley et al., 2011; McGinty et al., 2014). However, we are still lacking much information about the overall structure of multiprotein PRC1 subtypes. The only exception to this is provided by Wong and colleagues, who have resolved the minimum PRC1.1 complex, comprising the PCGF1 RAWUL domain, the BCORL1 PUFD domain, the KDM2B LRR and FBOX domains, and SKP1 (Wong et al., 2016; Wong et al., 2020). The structure obtained has helped to provide an explanation for how PCGF1 and BCORL1 cooperate to the selective binding of the rest of the factors, specifying the PRC1.1 complex.

On the other hand, more information is available on PRC2 subunits, their association into complex, and their interaction with chromatin: as for PRC1, early reports on individual proteins have highlighted important aspects of PRC2 factors activity, as, for example, the recognition of H3K27me3 by EED (Margueron et al., 2009), the interaction of PHF19 with H3K36me3 (Ballare et al., 2012), or the auto-inhibited conformation of individualized EZH2 (Antonysamy et al., 2013; Wu et al., 2013a). However, resolution of the structure of the whole PRC2 complex has provided an outlook at how all these proteins interact with each other (Chen et al., 2018; Jiao and Liu, 2015; Kasinath et al., 2018): the PRC2 complex is structured in two main lobes that are encompassed by SUZ12, which serves as a scaffold to seal together core and accessory factors. The upper catalytic lobe is composed of the WD propeller EED, EZH2, and the VRN2-EMF2-FIS2-Su(z)12 (VEFS) domain of SUZ12. These three parts compose the minimal PRC2 complex retaining HMT activity (Schmitges et al., 2011). Importantly, as discussed above, in cells expressing a truncated SUZ12 version containing only the VEFS domain, bulk H3K27me3 levels are maintained but deposition does not follow the canonical pattern (Hojfeldt et al., 2018), owing to the lack of targeting by accessory factors. The lower lobe is composed by the WD propeller RBBP4/7 and the N-terminal part of SUZ12. This latter is responsible for the binding of accessory factors to the

complex (see below); therefore, this lobe is termed the scaffolding lobe.

Resolution of the entire PRC2 structure has also helped understand how subunits influence each other's activity, highlighting emerging properties of the PRC2 complex (Fig. 5). One such example is the regulatory axis that exists between EED and EZH2: recognition of the H3K27me3 mark by the aromatic cage of EED results in the structuring of the EZH2 stimulation responsive motif (SRM) that in turn induces a 20º rotation of the adjacent part of the SET domain, resulting in the opening of the substrate binding cleft, and therefore, stimulating EZH2 HMT catalytic activity (Jiao and Liu, 2015; Justin et al., 2016; Lee et al., 2018). This mechanism was even visualized using Cryo-electron microscopy of PRC2 together with nucleosomal arrays, where it has been observed how EED contacts the modified tail from one nucleosome, while EZH2 engages the unmodified tail of the adjacent nucleosome (Poepsel et al., 2018).

Structural studies have also helped define the molecular bases for PRC2 subtypes formation: resolution of the scaffolding lobe of PRC2 (including the N-terminal part of SUZ12 and RBBP4) in association with the JARID2 transrepression (TR) domain, and the AEBP2 C2-binding (C2B) domain has highlighted that JARID2 and EPOP, as well as AEBP2 and PCL proteins, compete for the binding of the same residues on the surface of SUZ12 Zn/Zn binding (ZnB) domain and C2 domains, respectively (Fig. 5A,B) (Chen et al., 2018). Another interesting aspect emerging from this structure is that RBBP4 nucleosome binding activity is impaired upon inclusion in the PRC2.2 complex: the acidic patch on the central pocket of the WD propeller of RBBP4 that would normally mediate the binding to nucleosomes (Schmitges et al., 2011), is now occupied by two positively charged residues of AEBP2 (K502, R503) and one of SUZ12 (R196, contained in the C2 domain) collectively referred to as the KR finger (Fig. I5C,D), which mimic residues R3 and K4 of the histone H3 (Chen et al., 2018). We performed a systematic review of these and more aspects of PRC2 structural-functional relationship, including an analysis of possible loss and gain of function mutations

in cancer. This review article was published in the British Journal of Cancer (Chammas et al., 2020).
More recent structural works have reported that PRC2 can establish higher order interactions: while in the monomeric state the SUZ12 R196 (part of the KR finger) contacts RBBP4 acidic patch within the same PRC2 complex (closed conformation), a change in the orientation of the SUZ12 C2 domain can project this critical residue outwards (open conformation), allowing to reach out for the same binding site on a neighboring PRC2 (Chen et al., 2020; Grau et al., 2021). This interaction can be reciprocal, generating a PRC2 dimer. In other words, SUZ12 C2 domain can interact either within the same PRC2 complex or across two different complexes, swapping from one RBBP4 to the other (Fig. 5E). Importantly, domain swapping is a widespread mechanism for higher order interactions (Liu and Eisenberg, 2002). Interestingly, due to the relatively high flexibility of the C2 domain, at least two dimer orientations have been reported, one with antiparallel symmetry, with the two scaffolding lobes facing each other and orienting catalytic lobes in opposite directions (Fig. I5F) (Chen et al., 2020), and a mirrored symmetry, with both complexes wrapping around the same nucleosome with their catalytic lobes (Grau et al., 2021). PRC2 ability to dimerize is intrinsic to the core components. Interestingly, though, inclusion of MTF2 or PHF19 to the complex stabilizes the dimeric conformation (Fig. I5F), while AEBP2 prevents it, suggesting that PRC2.1 might preferentially rely on this mechanism (Chen et al., 2020). Moreover, capability to dimerize has been shown to be important for H3K27me3 maintenance at Polycomb target genes in serum/LIF grown mESCs (Chen et al., 2020).
Notably, the recently resolved structure of the PR-DUB complex containing Calypso and Asx suggests that this complex might also act as a dimer (Foglizzo et al., 2018). Dimerization was found to increase nucleosome binding and, therefore, deubiquitinase activity by BAP1. These results suggest that both PRC2 and PR-DUB may adopt dimerization as a way to maximize their activity on chromatin

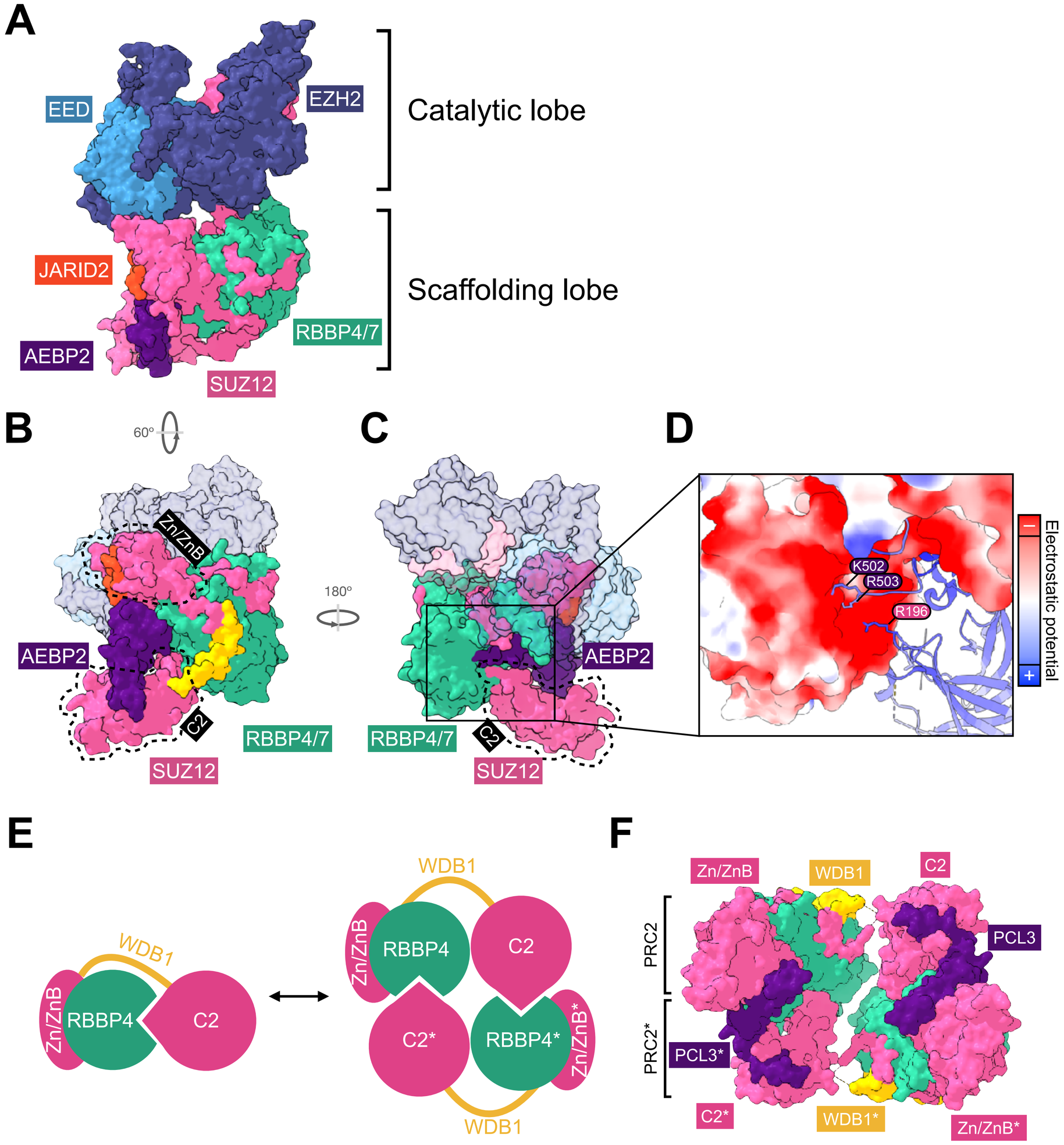


Figure I5 Structural characterisation of PRC2 unveils emerging properties of this complex A 3D rendering of the PRC2 structure including JARID2 TR domain and AEBP2 C2B domain (adapted from Kasinath et al., 2018 and Chen et al., 2018, PDB:6C23+5WAI) showing the two lobes of the complex; **B** focus on the scaffolding lobe, highlighting SUZ12 domains implicated in the binding of accessory factors (Zn/ZnB, C2). SUZ12 WDB1 domain is highlighted in yellow; **C–D** opposite side of the structure in **B**, showing the back of the WD propeller of RBBP4, with a close-up of the residues of AEBP2 and SUZ12 forming the KR finger that interacts with RBBP4 acidic pocket (**D**). Note that AEBP2 R502 and K503, and SUZ12 R196 have positive electrostatic potential while RBBP4 pocket is decorated with negatively charged residues; **E** graphical representation of the C2 swapping mechanism (same colors as in **A–C**); **F** Structure of the SUZ12-RBBP4-PCL3 dimer (adapted from Chen et al., 2020, PDB: 6NQ3). Note that PCL3 RC domain is sitting between the C2 domain of one complex and the Zn/ZnB domain of the other complex, stabilising the dimeric conformation.

Aim of the project

The goal of this project is to investigate the role of PRC1 and PRC2 accessory factors in directing the activity of these complexes during early developmental decisions. To achieve this, the project was subdivided into two parts, which in turn enclose different objectives:

Part I: characterization of the variations in PRC1 assembly upon differentiation of mESCs to PrE, and study of their functional relevance. This part includes the following objectives:

1) To establish an *in vitro* model system that allows for high-yield and reproducible immunoprecipitation of the PRC1 complex, with a special focus on CBX proteins.
2) To establish a differentiation protocol that allows for *in vitro* derivation of PrE from mESCs.
3) To characterize the interactome of canonical and non-canonical PRC1 complexes in mESCs and PrE.
4) Upon selection of interesting candidates, to investigate their impact on PRC1 functioning in this process
5) To evaluate the role of PRC1 in the mESCs-to-PrE transition

Part II: investigating the function of a novel *Suz12* splicing isoform with potential impact on PRC2 composition and structure. This part includes the following objectives:

1) To characterize the impact of this exon in shaping PRC2 formation and its interaction with accessory factors.
2) To analyze the role of this isoform in Polycomb activity, using mESCs as a model system

Results Part I

Generation of a mESCs system to study CBX proteins during differentiation

In mESCs, CBX7 is the main paralog incorporated into PRC1 and it contributes to the transcriptional silencing of the other members of the Cbx family (Morey et al., 2012). Upon exit from pluripotency, *Cbx7* gets rapidly downregulated, determining derepression of the other paralogs, that can now be incorporated into the PRC1 complex. This generates a duality, between a pluripotency-specific paralog, *Cbx7*, and differentiation-specific ones, *Cbx2*, *Cbx4*, and *Cbx8*. While the role of the former has been well documented in mESCs, little is known about the significance of the role swap that happens upon differentiation in intact cells, and how this regulates mouse early development *in vivo*. However, the lack of reliable specific antibodies, and generally the low levels of expression, have made it difficult to study these proteins in mESCs and upon differentiation. For example, to study protein-protein interactions, most of the works have tried to overcome these limitations by overexpressing exogenous tagged copies of PRC1 factors (Creppe et al., 2014; Klauke et al., 2013; Kloet et al., 2016; Morey et al., 2015; Sanchez et al., 2007; Santanach et al., 2017). Since PRC1 complexes composition results from the competition between accessory factors to bind core components RNF1/2, overexpressing one component is destined to compromise the overall network of interactions, ultimately impacting the biology of the system. For this reason, and to overcome the barrier of antibody sensitivity, we decided to build a system that would allow the study of PRC1 composition and activity, with a special focus on Cbx members, in differentiation, without interfering with the PRC1 interactome. To this end, we decided to generate mESCs cell lines containing a tag in the endogenous copy of the *Cbx2*, *Cbx4*, *Cbx6*, and *Cbx8* genes, to allow for efficient and comparable pull-down of the bait, without disrupting the physiological conditions. Moreover, we decided to tag *Rnf2* and *Rybp*,

representative of total and non-canonical PRC1, respectively, to have a complete picture of PRC1 subtypes.

The initial idea was to apply this system to mESCs differentiation towards the neural lineage. This was motivated by several reasons: i) PRC1 and PRC2 play a critical role in mouse neurogenesis, although the role of the different CBX proteins was unexplored (Corley and Kroll, 2015; Testa, 2011); ii) all differentiation-specific Cbx paralogs are expressed during neural differentiation (Bonev et al., 2017); iii) in our laboratory, we have established a reliable protocol for differentiation of mESCs to neural precursor cells (NPCs), using the Sox1:GFP reporter to monitor the efficiency of differentiation (Ying et al., 2003). For these reasons we decided to introduce the tag into the mES Sox1:GFP cell line. However, halfway through this project, some reports took advantage of the same differentiation system to study PRC1 composition and chromatin occupancy, as well as its role in chromatin conformation and gene regulation (Kloet et al., 2016; Kundu et al., 2017; Zhao et al., 2020). Therefore, we maintained the same cell line but applied it to the investigation of another cell fate decision that relies on Polycomb-mediated regulation (see below), in which the role of CBX family members has been so far poorly investigated.

Briefly, a 3xHA-3xFLAG tag peptide sequence was inserted into the endogenous loci of target genes via CRISPR-mediated DSB induction around the stop codon, combined with HDR template-mediated insertion (Fig R1A). Importantly, the homology arms used for HDR-mediated insertion contained a silent mutation of the PAM sequence, preventing the Cas9-gRNA complex from cutting again the same site once the insertion has happened. This allowed us to carry out multiple rounds of CRISPR, when necessary, without interfering with successfully edited alleles.

Practically, Sox1:GFP cells were subject to an initial round of CRISPR editing and screened for the insertion of the tag by PCR on the targeted genomic locus. One round was sufficient to isolate clones carrying homozygous insertion of the tag for *Rybp* (two clones), *Cbx6* (two clones), and *Cbx8* (one clone). For the rest of

the lines, clones with heterozygous insertions from the first round were pooled together and used for a second round of CRISPR editing. This way, we obtained two homozygous clones for *Rnf2*, one for *Cbx2*, and one for *Cbx4*.

Overall, considering every insertion as an independent event, editing in PX458-based transfection was 5- to 15-fold more frequent than in Cas9/RNP transfection (Fig. R1B). The differences observed are likely due to the fact that PX458 allows for enrichment of transfected cells (GFP positive), while RNP-based method doesn't.

Insertion of the tag was confirmed by Sanger sequencing (data not shown): insertions were placed in the correct position and editing did not result in any mutations within the open reading frame of the target gene. The presence of the protein tag was confirmed by Western Blot analysis (Fig. R1C–G). We were able to detect FLAG-RNF2 and FLAG-RYBP from whole cell extracts of the corresponding tagged lines. However, none of the other tagged proteins were visible, so we decided to enrich the tagged protein using immunoprecipitation (IP) of the FLAG epitope: while RNF2, RYBP and CBX6 were readily visible upon FLAG-IP in ESCs, CBX2, CBX4, and CBX8 were not (Fig. R1C), this was likely due to the low abondance of the proteins in undifferentiated ES cells. Thus, to allow for effective visualization of these proteins, cells were differentiated to NPCs (Fig. R1D), where expression levels are higher (Fig. R1E). Tagged CBX4 was visible upon differentiation, while no band for CBX2 or CBX8 was observed. However, FLAG-IP in CBX2-tagged NPCs led to Co-IP of RNF2, providing indirect evidence of the presence of the tag. Regarding *Cbx8*, while Western Blot failed to detect any band in ESC or NPC, IP-MS in ESC-derived Primitive Endoderm confirmed the presence of the FLAG tag in this protein (Table R2).

Tagged cell lines displayed normal ESC morphology with dome-shaped colonies (Fig. R2) and had a similar proliferation rate with respect to the parental Sox1:GFP cell line (Fig. R1H). Importantly, cells retained the potential to differentiate to NPCs, as visualized by cell morphology and Sox1:GFP expression (Fig. R2).

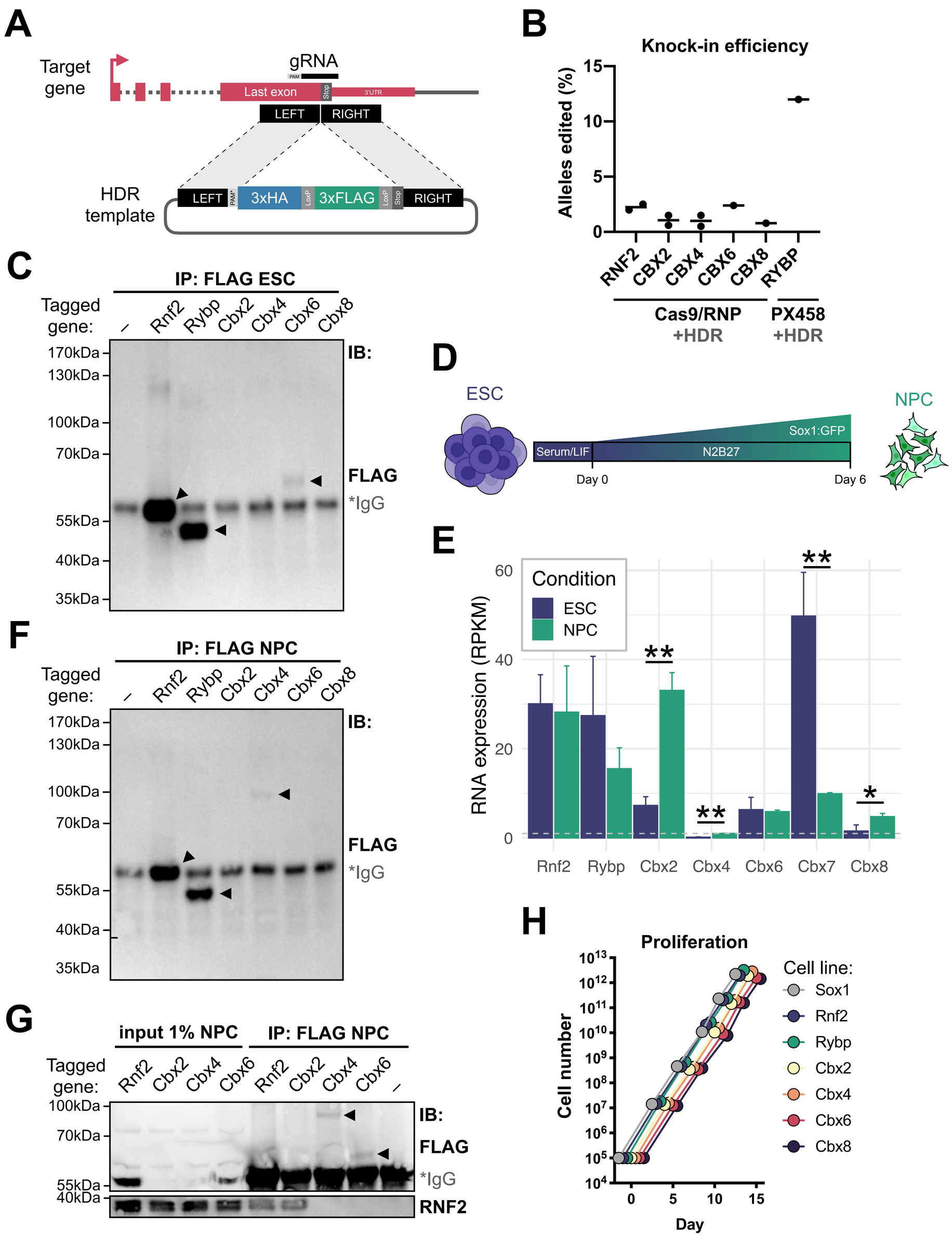


Figure R1 Generation of endogenously tagged PRC1 cell lines A Scheme of the CRISPR knock-in tagging strategy, targeting the C-terminal site of *Rnf2*, *Rybp*, *Cbx2*/*4*/*6*/*8*; **B** CRISPR knock-in efficiency in the different tagging experiments; **C** Western Blot of FLAG IP performed in the different ESC cell lines. Black arrowheads indicate bands corresponding to the tagged protein; **D** Schematic protocol used to differentiate ESCs to NPCs (adapted from Ying et al., 2003); **E** Expression of the genes of interest in ESC and NPC cells (*padj<0.05; **padj<0.01); **F–G** Western Blot of FLAG IP performed in the different lines upon NPC differentiation. Note: *Rnf2* clone used in **G** is heterozygous for the insertion, therefore it has a wt allele, visible at ~38 kDa both in the input and in Co-IP with FLAG-RNF2 and -CBX2; **H** Growth curve of parental cells (Sox1) and tagged lines grown in Serum/LIF conditions.

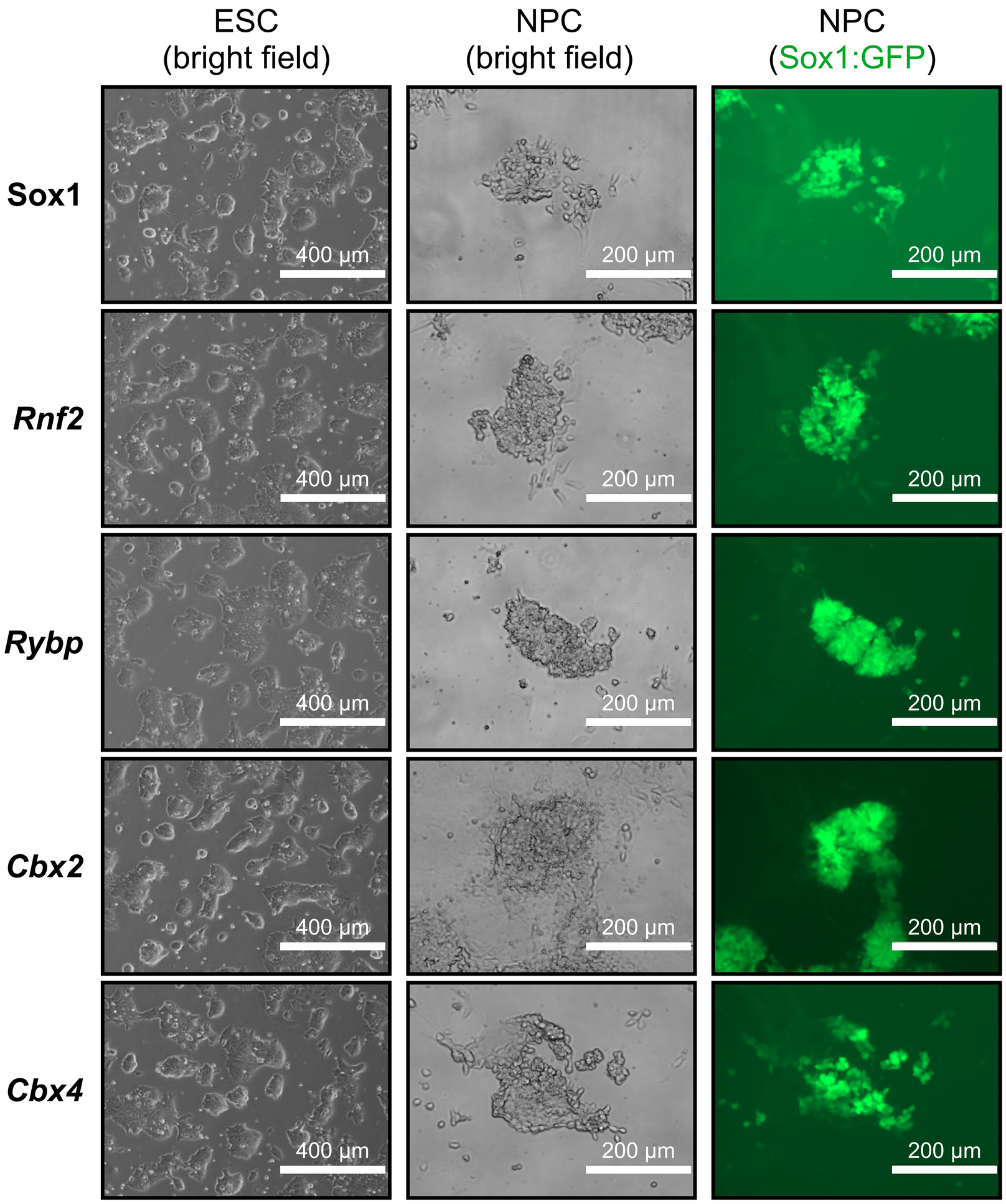


Figure R2 PRC1 tagged mESCs potential to generate NPCs Bright field picture of ESCs and day 6 NPCs of parental (Sox1) and tagged cell lines.

Derivation of primitive endoderm from mouse embryonic stem cells

During the early blastocyst stage, the ICM will form two functionally distinct layers, the epiblast, responsible for the formation of all three germ layers of the embryo, and the hypoblast, forming the primitive endoderm, necessary to mediate the exchange of gas and nutrients with the mother (Rossant and Tam, 2009). The molecular determinants of this decision are not yet fully understood, although it was demonstrated that both PRC1 and PRC2 take part in ensuring the correct balance between the two lineages, by allowing transcriptional heterogeneity in the ICM (Illingworth et al., 2016). Reports have shown that RNF2 is necessary for early embryonic development, as the *Rnf2* KO mouse model doesn't reach past the gastrulation stage (E6.5-8.5) (Posfai et al., 2012; Voncken et al., 2003). However, mice bearing a mutation that kills *Rnf2* catalytic activity (I53A) undergo normal development up until E15.5 (Illingworth et al., 2015). This data suggests that RNF2, but not its E3 ligase activity, is necessary for fate decisions prior to implantation. Given that canonical PRC1 regulates gene expression via an H2Aub-independent manner (Boyle et al., 2020; Francis et al., 2004; Fursova et al., 2019), we speculated that this complex might be involved in the epiblast-to-hypoblast fate choice. Therefore, we sought to investigate the role of canonical PRC1 in this fate decision.

Recently, the lab of Prof. Joshua M. Brickman has reported that *in vitro* grown mESCs have the potential to generate primitive endoderm (PrE) cells, upon activation of the Wnt and Tgfβ pathway in the absence of insulin (Anderson et al., 2017). It is worth noting that the transition of naïve mESCs to PrE does not represent a faithful imitation of the events happening *in vivo*: indeed, while in vivo PrE derives from the inner cell mass (E3.5), lab-grown mESCs resemble more the E4.5 epiblast (Martello and Smith, 2014). Therefore, it seems that *in vitro* grown mESCs still retain the capacity to generate PrE cells, in a process that blends the boundary between differentiation and trans-differentiation. For the sake of simplicity, I will refer to the practice of PrE derivation from mESCs as *differentiation*.

We adapted the protocol to derive PrE from our serum/LIF grown mESCs using Activin A and Chiron for nine days to differentiate the cells, adding LIF to maintain PrE cells proliferation by day 7 of differentiation (Fig. R3A). During differentiation, cells acquired an epithelial morphology, gradually expanding around the original dome-shaped ESC colony (Fig. R3B). We monitored PrE differentiation via staining with an APC-conjugated antibody specific for the Platelet-Derived Growth Factor Receptor Alpha (PDGFRA), a cell surface marker of endodermal lineage, while the internal Sox1:GFP reporter was used as a control (Fig. R3C). After nine days in the differentiation medium, around 70-80% of the cells stained positive for this marker, while no upregulation of the internal Sox1:GFP marker was detected (Fig. R3C–D). Importantly, differentiation speed was dependent on Activin A concentration, although at later time points cells reached a similar percentage of PDGFRA positive cells (Fig. R3D). For subsequent PrE differentiation assays, we decided to use the lowest Activin A concentration that would allow cells to efficiently differentiate (4 ng/ml). Differentiation was accompanied by downregulation of pluripotency markers *Nanog* and *Oct4* (Fig. R3E), and upregulation of endodermal transcription factors *Foxa2*, *Gata4*, *Gata6* and *Sox7* (Fig. R3F).

RNA-seq analysis of PrE revealed that cells undergo major transcriptomic changes, with approximately 6500 differentially expressed genes (padj < 0.05) upon differentiation, of which 3098 were downregulated and 3404 upregulated (Fig. R4A). Importantly, while endoderm marker genes were upregulated, pluripotency and germline specific genes were downregulated, and little to no changes in expression were observed in ectoderm and mesoderm genes (Fig. R4B), confirming the specificity of the differentiation. Downregulated genes were mostly enriched in cellular processes related to response to LIF and cell proliferation (e.g., DNA replication and repair, chromosome segregation, Fig. R4C). In accordance with this, we observed a dramatic drop in the growth rates of PrE with respect to ESCs (Fig. R3G–H).

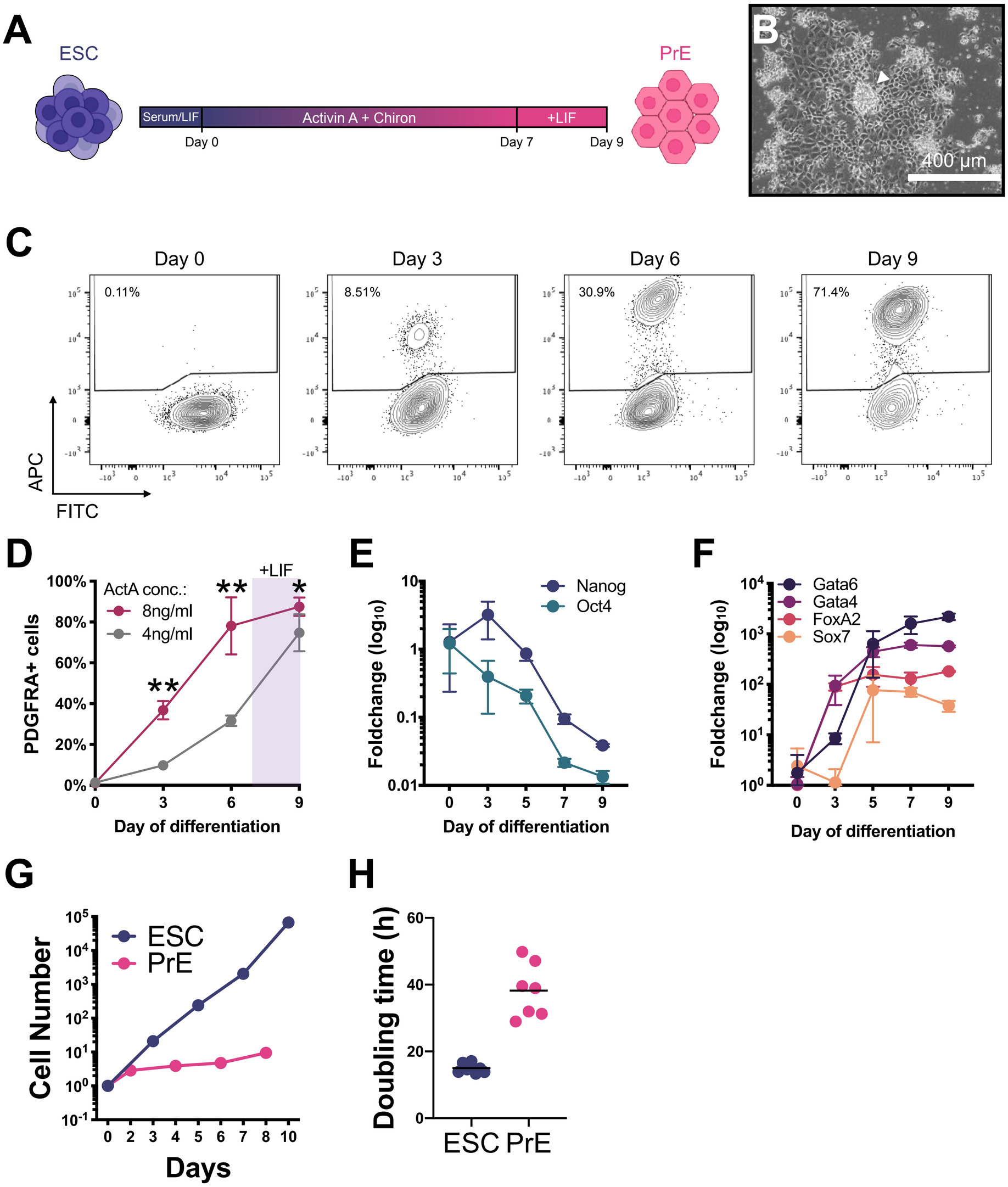


Figure R3 Setup of the PrE differentiation protocol A Schematic protocol used to differentiate ESCs to PrE (adapted from Anderson et al., 2017). **B** a PrE colony as seen through the microscope. The white arrowhead indicates the original ESC colony that generated it. **C** FACS plot of a representative PrE differentiation experiment, showing the gradual increase in APC-PDGFRA^{+} staining (y-axis) during the differentiation protocol. Internal Sox1:GFP (FITC) reporter is used as a negative control (x-axis); **D** Quantification of PrE differentiation efficiency, as reported by PDGFRA staining, using different concentrations of Activin A (* $p < 0.05$, ** $p < 0.01$, t-test); **E-F** RT-qPCR of pluripotent (**E**) and endodermal (**F**) markers during PrE differentiation protocol (normalized on the average expression of *Rplp0* and *Tbp*) **G-H** Growth curve of ESCs and PrE cells (**G**) and the inferred doubling time (**H**).

Figure R4 Characterization of *in vitro* derived PrE A MA plot displaying transcriptomic changes upon PrE differentiation of mESCs. Significantly downregulated genes are shown in blue while upregulated ones are in magenta. Selected pluripotency and endodermal markers are highlighted; **B** boxplot showing the expression levels of selected signature genes for each lineage (see methods section for the complete list of genes); **C–D** Cleveland dot plot of GO Biological Processes enriched among the downregulated (**C**) and upregulated (**D**) genes upon PrE differentiation. D–E Bar plot showing the changes in expression of genes associated to PRC1 (**D**) or PRC2 (**E**).

Conversely, upregulated genes were involved in response to Tgfβ, and processes related with cell adhesion and migration (Fig. R4D), consistent with the switch from a dome-shaped colony growth to a more epithelial one. Next, we focused on the expression of PcG genes: regarding PRC1, we noticed that *Phc1*, *Cbx6* and *Cbx7* were downregulated, while concomitant upregulation of *Bmi1*, *Phc2* and *Phc3*, and *Cbx4* and *Cbx8* was observed, indicating the possibility that canonical PRC1 might undergo a complete switch of composition in PrE (Fig. R4E). *Rybp* was slightly downregulated, albeit not significantly, while its homolog *Yaf2* was significantly upregulated, suggesting another switch in the non-canonical PRC1. Moreover, both *Pcgf3*, and *Pcgf5*, as well as *Fbrs* were significantly upregulated, suggesting that PRC1.3/5 play a role in PrE (Fig. R4E). On the other hand, all of the PRC2 factors, except for *Rbbp7*, were either downregulated or unchanged in expression (Fig. R4F), somehow suggesting that PRC2 might be less fundamental for PrE biology.

Characterization of the PRC1 interactome in mESCs

To assess whether changes in expression reflected switches in PRC1 composition, we went on to characterize PRC1 interactome upon PrE differentiation. The protein interaction network of PRC1 in mESCs was recently described by Kloet and colleagues, using a transgene with GFP-tagged *Rnf2* (Kloet et al., 2016). Importantly, the authors try to mimic the levels of expression of the endogenous protein by introducing the endogenous promoter in the transgene. However, the levels of expression of RNF2-GFP that they achieve are much lower than the endogenous protein. Therefore, we decided to test our endogenously tagged system and build the mESCs PRC1 interactome with endogenous levels of proteins. FLAG IP was performed for three technical replicates on benzonase-treated nuclear extracts, in order to focus on the stable interactions happening between PRC1 factors in the nucleus, while discarding transient, indirect (DNA/RNA-mediated), or spurious ones. As a control, FLAG mock IP was performed in parental Sox1:GFP line (not expressing

any FLAG-tagged protein). Probabilistic interaction based on SAINT analysis on RNF2 and RYBP FLAG IP-MS, identified respectively 24 and 15 highly confident interactors (Fold-change ≥ 10; BFDR ≤ 0.05), the majority of which are known Polycomb factors (23/24 for RNF2, 12/15 for RYBP, Fig. R6C). All PCGF proteins were detected in RNF2 IP, confirming the presence of all six PRC1 subtypes in mESCs (Fig. R5A, R6A). Instead, none of the canonical PRC1 components were significantly enriched in RYBP IP (Fig. R5B, R6B), consistent with the fact that RYBP is only included in non-canonical PRC1 subtypes. We confirmed by Western blot that canonical PRC1 factors PHC1 and MEL-18 do not Co-IP with RYBP, while PRC1.6 factor L3MBTL2 does (Fig. R6D).

PCGF6 was the most enriched PCGF paralog in both RNF2 and RYBP IP (Fig. R5A–B, Fig. R9A–B), suggesting that PRC1.6 is the prevalent subtype in mESCs, in line with previous proteomics studies (Kloet et al., 2016; Scelfo et al., 2019). This observation was corroborated by fractionation of nuclear protein extracts (Fig. R6E): while still being present at fractions 8–12 containing canonical PRC1 factors PHC1, CBX7, MEL-18 (PCGF2), RNF2 was found to mainly co-sediment with L3MBTL2, a PRC1.6 factor (Fig. R6E, fractions 14–18). Notably, while RNF2 and RYBP interacted with the same set of PRC1.6 factors (Fig. R5A–B, R6A–B), RYBP was more enriched in fractions of lower weight (2–4), in line with previous reports (Gao et al., 2012), suggesting that a large fraction of RYBP proteins in the nucleus does not take part in any complex formation and/or that RYBP might interact with PRC1 factors separately or, eventually, be involved in the formation of intermediate PRC1.6 subcomplexes. Regarding canonical PRC1, the most likely configuration in mESCs seems to be the one containing PHC1, MEL-18, and CBX7, along with RNF2.

Notably, RNF2 (RING1B) was found to interact with its paralog RNF1 (RING1A, Fig. R5A). Co-IP of these two proteins had been already observed in previous PRC1 proteomic studies (Gao et al., 2012; Kloet et al., 2016) but not in others (Maat et al., 2021; van den Boom et al., 2013). While in the classical view the presence of these

two proteins in the core of PRC1 is mutually exclusive, no structural evidence was provided in support of this view. However, besides the distinct stringencies of IP conditions used in this and other studies, another interesting explanation is that Co-IP occurs due to higher-order interactions between distinct cPRC1 complexes due to PHC-mediated oligomerization (Isono et al., 2013).

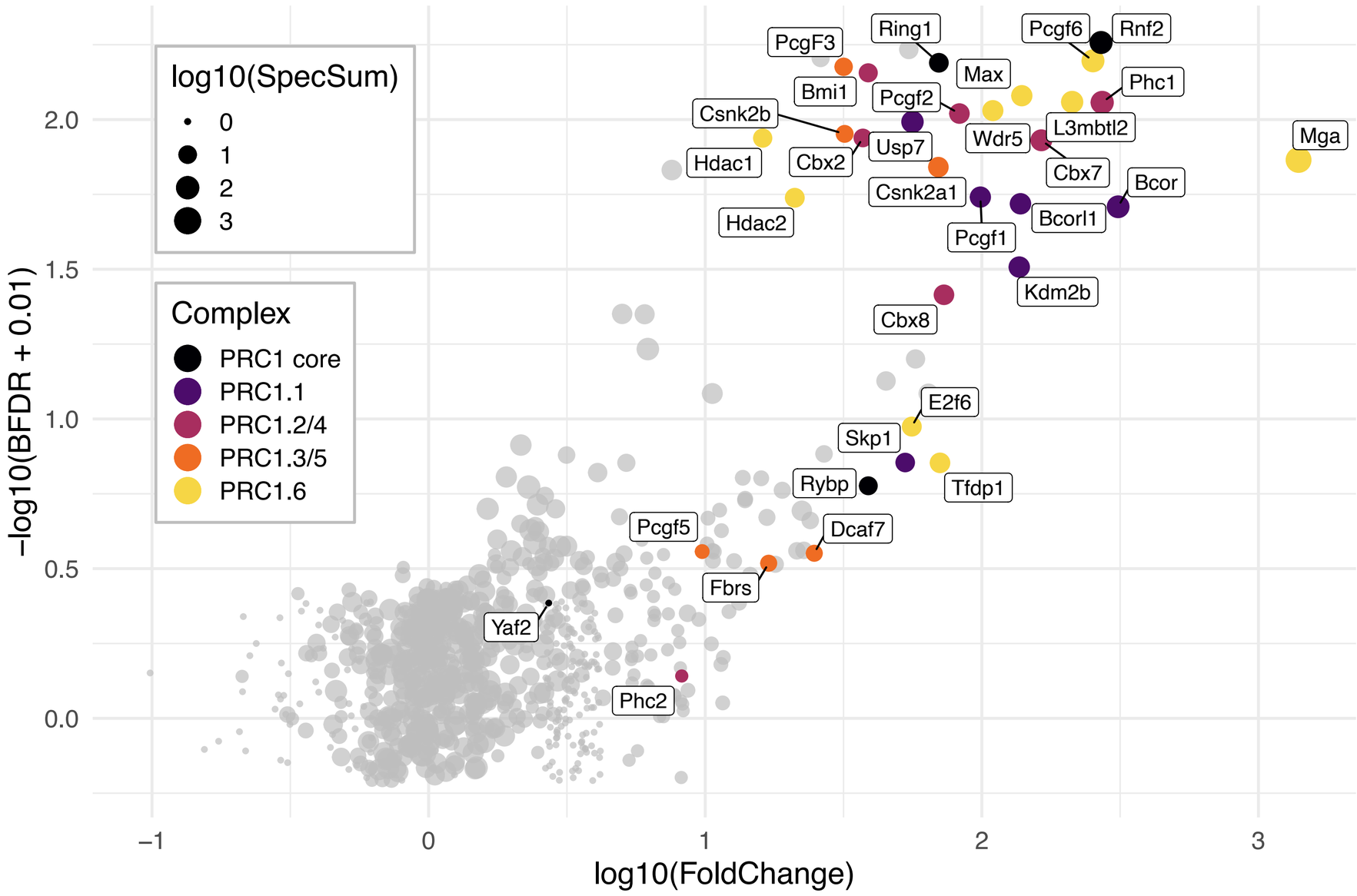
FLAG IP-MS in *Cbx2*-, *Cbx4*-, and *Cbx8*-tagged cell lines failed to retrieve any peptide belonging to the bait in any of the three replicates of each experiment (Table R1), therefore, these samples were discarded for further analysis. However, CBX8-FLAG IP-MS resulted in the detection of at least one peptide of PHC1 and RNF2, confirming RNF2 interaction with CBX8 in the reciprocal IP (Table R1). Notably, RNF2 was the only protein detected in all three replicates of both RNF2- and CBX8-IP (and never in the FLAG mock IP in Sox1:GFP cells).

Conversely, FLAG IP-MS in *Cbx6*-tagged cell line yielded at least one CBX6 peptide in each replicate (Table R1), while no other PRC1 protein was detected, suggesting CBX6 is indeed present in ESC although not incorporated in the PRC1 complex unless overexpressed (Santanach et al., 2017), in line with previous PRC1 proteomic studies (Kloet et al., 2016; Morey et al., 2012).

Table R1 *Peptides of canonical PRC1 proteins detected upon FLAG IP-MS in mESCs*

Prey \ Bait	**– (Sox1)**	**RNF2**	**CBX2**	**CBX4**	**CBX6**	**CBX8**
RNF2	0	**88**	0	0	0	3
CBX2	0	11	**0**	0	0	0
CBX4	0	0	0	**0**	0	0
CBX6	0	0	0	0	**3**	0
CBX8	0	25	0	0	0	**0**
CBX7	0	48	0	0	0	0
MEL-18/PCGF2	0	25	0	0	0	0
BMI1/PCGF4	0	13	0	0	0	0
PHC1	0	98	0	0	0	1
PHC2	0	2	0	0	0	0

A



B

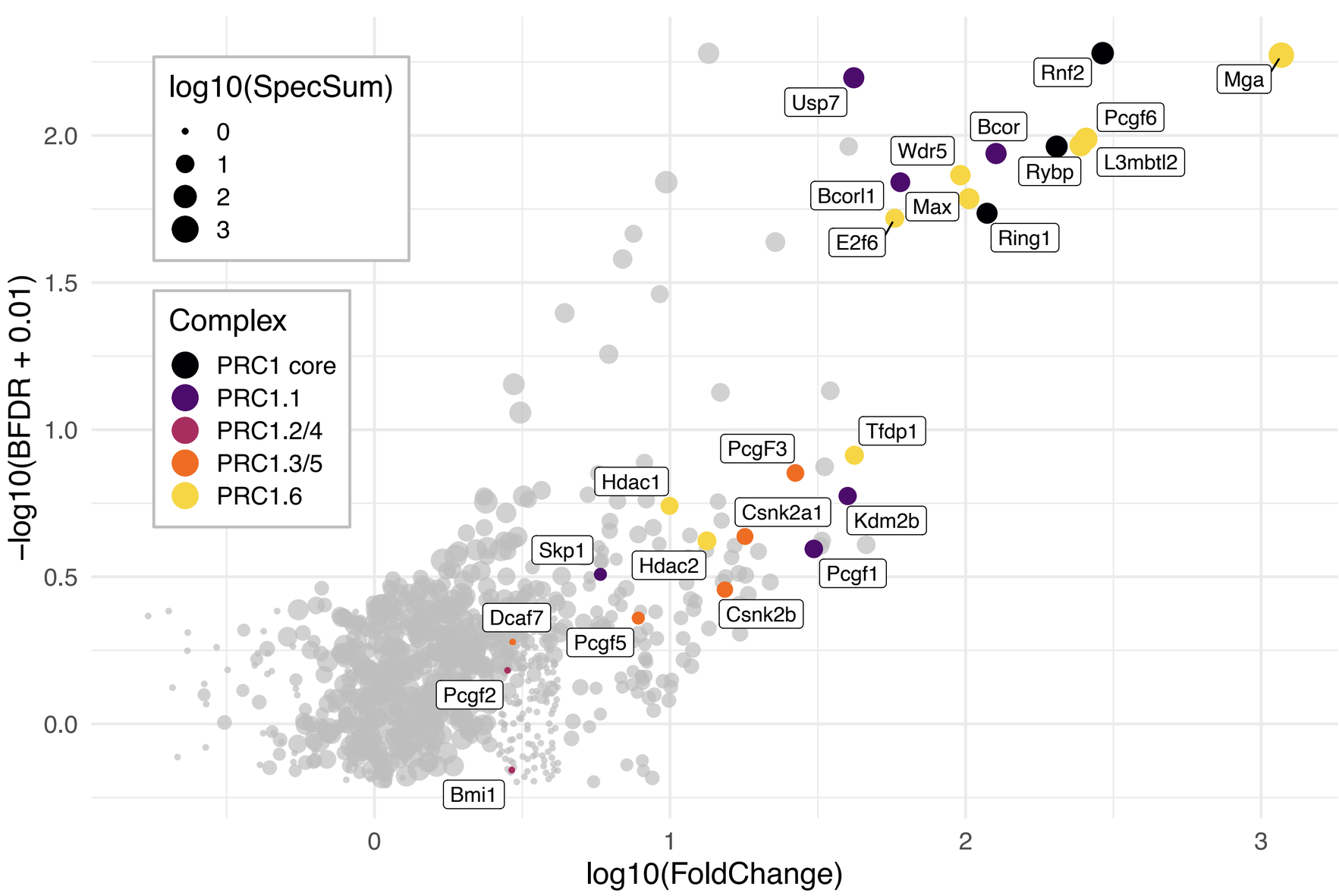


Figure R5 RNF2 and RYBP interactome in mESCs (1) A–B Volcano plot of the proteins identified in FLAG IP-MS in RNF2 (**A**) and RYBP (**B**) tagged cell lines in mESCs. Known PRC1 factors are coloured according to the complex they take part in. Note: to allow for visualisation, points were slightly jittered using the parameter *position_jitter(height = 0.2, width = 0.2, seed = 5)*

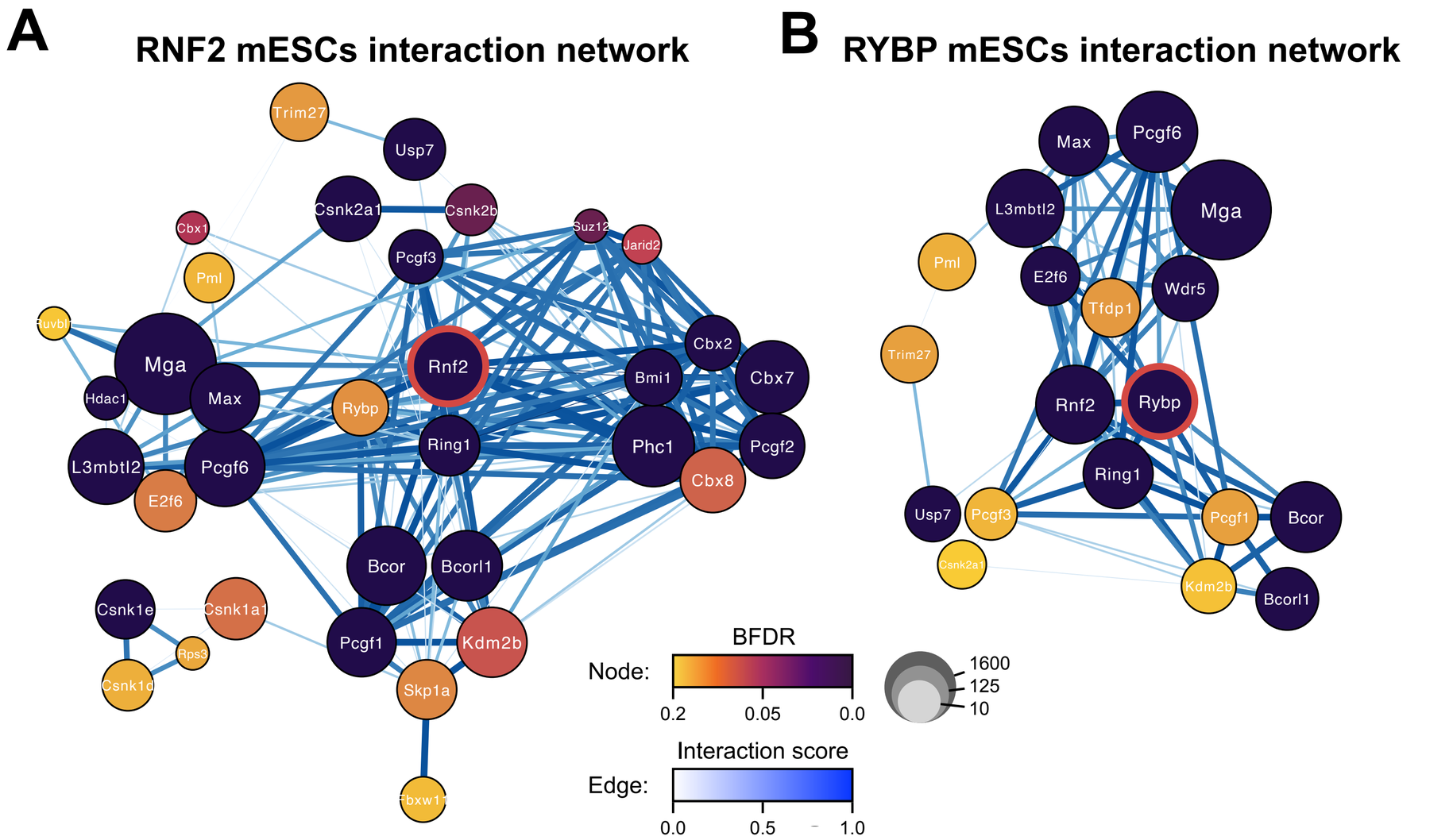


Figure R6 RNF2 and RYBP interactome in mESCs (2) A–B Protein-protein interaction network of RNF2 (**A**) and RYBP (**B**) interactors found in mESCs (FC ≥ 5, BFDR ≤ 0.2). The bait of each experiment is marked with a thick red circle. Individual peptides with no annotated interaction with any other protein of the network were excluded from the graph; **C** bar plot indicating the number PcG factors among the highly confident interactors (FC ≥ 5, BFDR ≤ 0.05) in RNF2 and RYBP FLAG IP-MS; **D** Western blot of FLAG IP-MS confirming some of the interactors detected from the IP-MS; **E** Western blot of a glycerol gradient fractionation of *Rnf2*-tagged mESCs nuclear extract. Note that some of the blots were performed over previous blots (indicated with a grey arrowhead). Molecular weight reported below indicates the approximate fraction at which the protein marker of the corresponding size run (see methods section).

Characterization of the PRC1 interactome in PrE

To define PRC1 interactome in PrE, we followed the same protocol as for mESCs FLAG IP-MS, harvesting PrE cells after nine days of differentiation.

We started from *Rnf2*- and *Rybp*-tagged cell lines, to have an exploratory view of the interactome in PrE. The first thing we noticed is that the average amount of the bait was, in both cases, reduced with respect to ESC (Fig. R8B–C), in line with a slight decrease in expression observed at the RNA level (Fig. R4B). Focusing on the interactors, we found 24 (Foldchange ≥ 10; BFDR ≤ 0.05) highly confident interactors for RNF2, of which, only 6 were already annotated PcG factors (Fig. R8D) Conversely, only 4 proteins were strongly enriched in RYBP FLAG IP-MS, namely RYBP, RNF2, PCGF6, and MGA (Fig. R7B, see below). Protein interaction network analysis (Fig. R8A) revealed that the majority of non-PcG RNF2 interactors in PrE was composed of proteins involved in DNA methylation (DNMT3A/B, and DNMT3L), mRNA splicing (e.g., SF3A1, SF3B1, CDC5L), ribosome constitutive components (RPLs and RPSs) and factors related to ribosome biogenesis (e.g., FBL, NOP56, and NOP58, GNL3). Among all the non-PcG proteins in the RNF2 interactome, Fibrillarin (FBL) was the most enriched one. This protein is an S-adenosyl-L-methyonine (SAM)-dependent RNA 2'-O-methyltransferase, involved in the maturation of rRNA within the C/D box snoRNP complex (Monaco et al., 2018). However, FBL does not only modify RNA substrates, but it is also capable of methylating histone H2A at glutamine 104 (H2AQ104me) (Iyer-Bierhoff et al., 2018; Tessarz et al., 2014). FBL methylation of H2A happens in the nucleolus, and this modification is thought to promote high rRNA transcription rates by counteracting FACT complex binding at the rDNA locus (Tessarz et al., 2014). Notably, FBL substrate potentially sits only 15 amino acids away from the one of RNF2 (H2AK119), raising the question of a functional interplay between these two proteins. Therefore, we sought to investigate the interaction between RNF2 and FBL. FLAG IP followed by Western blot using an antibody specific for FBL confirmed that RNF2 interacts with

FBL in PrE but not in ESC (Fig. R8E), although probably to a lower extent. Moreover, this experiment confirmed that RNF2 binding to PHC1 and L3MBTL2 drops in PrE, albeit still present. Interestingly, we observed RNF2 Co-IP with the H3K9me3 mark uniquely in PrE, suggesting a possible interplay between Polycomb repression and heterochromatinization, upon differentiation. This interaction is even more relevant considering that H3K9me3 bulk levels are reduced upon PrE differentiation (Fig. R8E).

Since FBL is located at rDNA in the nucleolus, we tested whether RNF2-FBL interaction happens at these chromatin loci. To do so we performed RNF2 and FBL ChIP: as expected FBL was enriched at rDNA loci both in ESC and PrE. Surprisingly, though, we found RNF2 association with coding regions of the rDNA in ESC, but not in PrE (Fig. R8F). This finding is somewhat unexpected, as, to our knowledge, there is no report of RNF2 (or PRC1) binding at rDNA genomic loci in mESCs (see discussion).

To confirm these observations, we performed double staining for FLAG-RNF2 and FBL and revealed their subnuclear localization in ESC and PrE by immunofluorescence (Fig. R8G). As expected, in ESCs RNF2 is found in discrete spots that are diffused around the nuclear volume but specifically excluded from the nucleolar space. However, in PrE cells, RNF2 signal was much rarer, making it impossible to identify a clear localization in the nucleus. While this is probably due to the lower amount of RNF2 in PrE, we could not define whether RNF2 is present in the nucleolus. Although difficult to interpret, this data did not support a functional interaction between RNF2 and FBL and was therefore not investigated further (see discussion).

Concerning PRC1 composition dynamics, upon PrE differentiation we observed a generalized loss of interaction with most of the factors we had identified in ESC, in both RNF2- and RYBP-IPs (Fig. R9A–C): many of the non-canonical PRC1 factors were not detected at all in either of the two PrE interactomes (E2F6, DP-1, PCGF3, KDM2B, DCAF7, BCORL1, FBRS, KDM5C), suggesting that PRC1.1 and PRC1.3/5 complexes are not formed in PrE. This

is particularly unexpected considering that, for example, *Pcgf3* and *Pcgf5* were found to be upregulated at the mRNA level upon PrE differentiation (Fig. R4E), although we did not test whether this up-regulation results in higher protein levels in PrE cells.

Among the factors we could detect, only MGA and PCGF6 were highly enriched in both RNF2 and RYBP IP-MS, indicating that non-canonical PRC1 complexity shrinks down to a simplified PRC1.6 subcomplex, containing only RNF2, PCGF6, and MGA (and possibly RYBP, although this is barely detectable in RNF2 IP, see Fig. R7A, R9A). Canonical PRC1 complexity was also greatly reduced: major PRC1.2/4 components like CBX2, CBX7 and MEL-18 (PCGF2) could not be detected in PrE RNF2 interactome (Fig. R9A) in line with their downregulation at the RNA level (Fig. R4D), while higher gene expression of *Bmi1* and *Cbx8* resulted in their inclusion in the PRC1 complex in PrE (Fig. R4D, R7A, R9A–C). To confirm this finding, we performed FLAG IP-MS in *Cbx8*-tagged PrE cells: as expected, we were able to identify peptides belonging to CBX8, RNF2, PHC1, and PHC2 (Table R2), providing further support to the abovementioned canonical PRC1 composition in PrE.

Table R2 *Peptides of canonical PRC1 proteins detected upon FLAG IP-MS in PrE*

Prey \ Bait	**– (Sox1)**	**RNF2**	**CBX8**
RNF2	0	**66**	3
CBX8	0	15	**3**
BMI1/PCGF4	0	6	0
PHC1	0	14	3
PHC2	0	3	1
PHC3	0	1	0

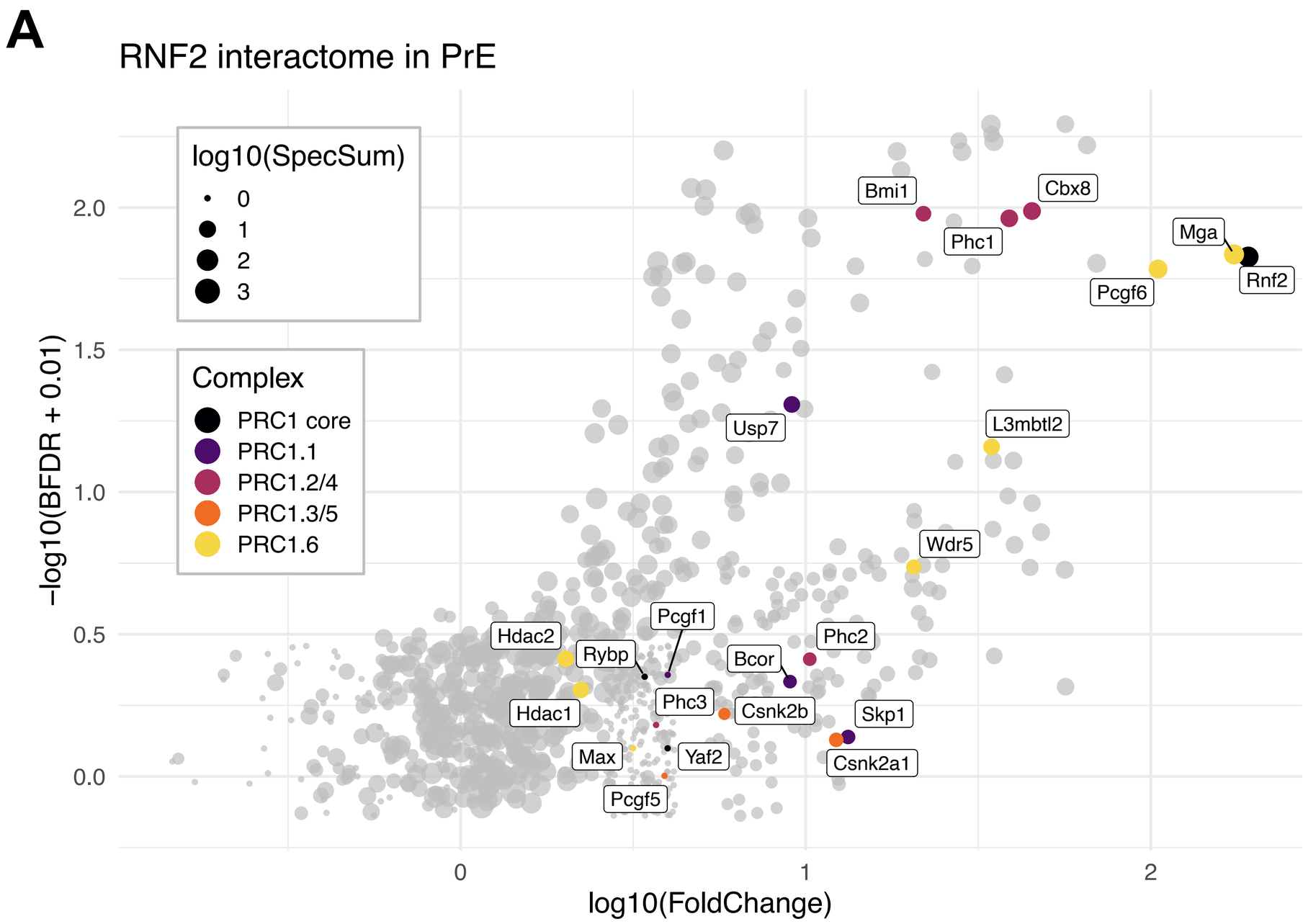


Figure R7 RNF2 and RYBP interactome in PrE (1) A–B Volcano plot of the proteins identified in FLAG IP-MS in RNF2 (**A**) and RYBP (**B**) tagged cell lines upon PrE differentiation. Same options as in Fig. R5

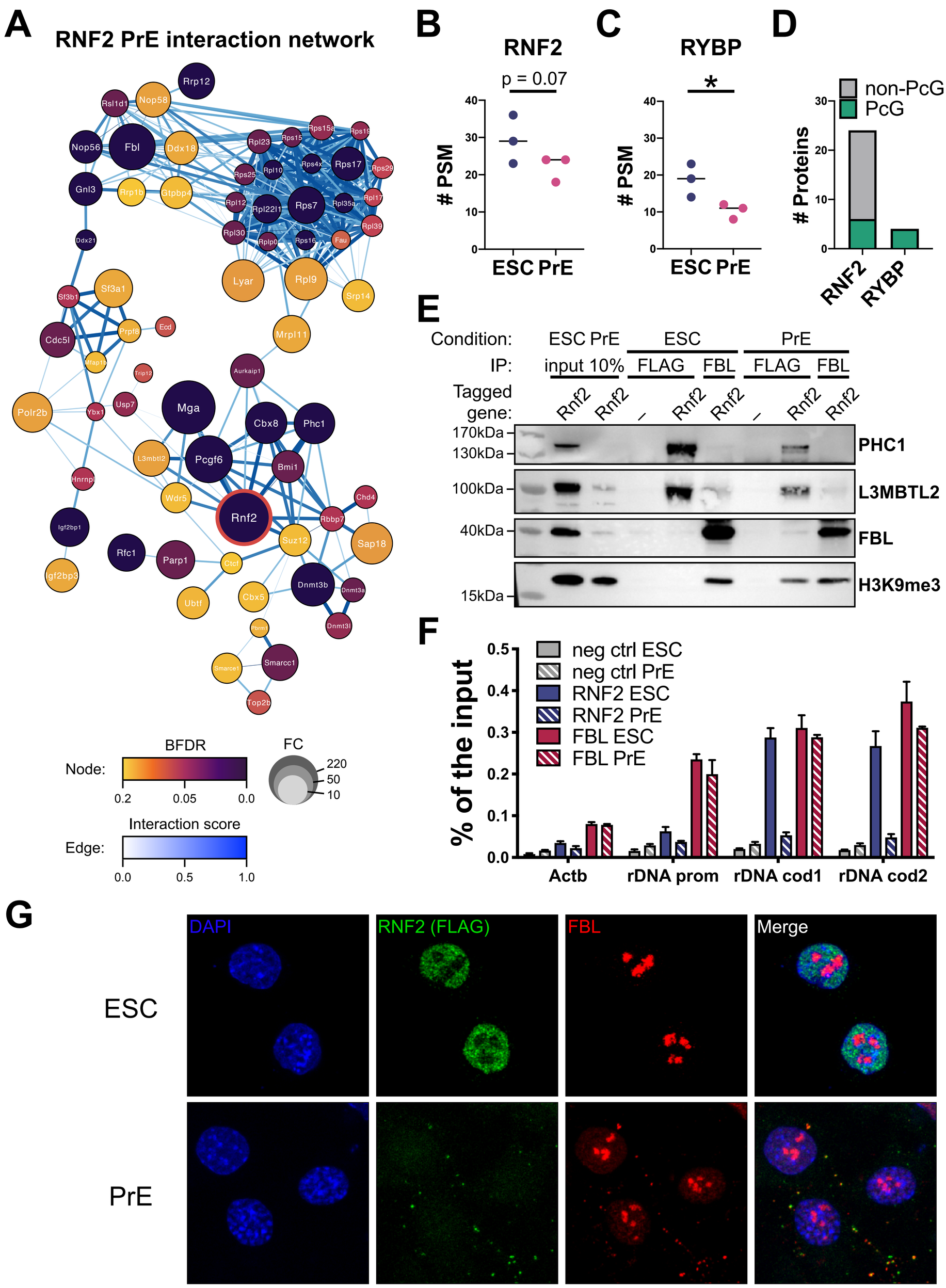


Figure R8 RNF2 and RYBP interactome in PrE (2) A Protein-protein interaction network of RNF2 in PrE. Same options as in R6A–B. **B–C** Retrieval of spectral counts of the bait in RNF2 (**B**) and RYBP (**C**) FLAG IP-MS (* $p < 0.05$, t-test); **D** Western blot of FLAG and FBL IP in ESC and PrE cells; **E** ChIP of mock (anti-FLAG in Sox1 parental cell line), RNF2 (anti-FLAG in *Rnf2*-tagged lines), and FBL (anti-FBL in *Rnf2*-tagged line), followed by qPCR at Actin B promoter (negative ChIP control), and at both rDNA promoter (prom) and transcribed region (cod1 and cod2); **F** immunofluorescence of RNF2 and FBL in *Rnf2*-tagged cells at ESC or PrE stage.

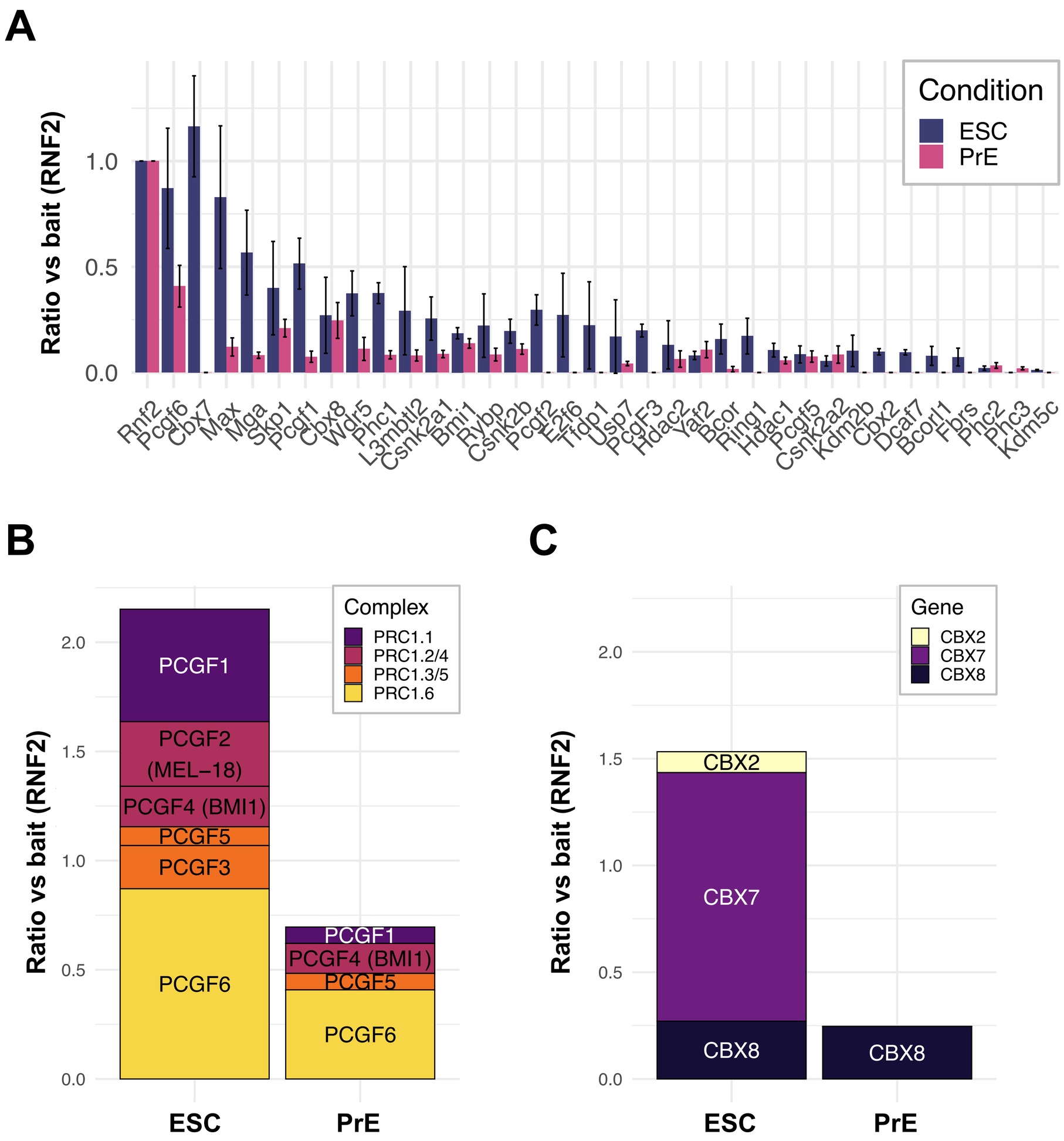


Figure R9 Changes in PRC1 interactome upon differentiation to PrE A Evaluation of the peptides ratio with respect to the bait (RNF2) for all the known PRC1 interactors. The value and the error bar indicate average and standard deviation respectively for three biological replicates; **B,C** same as in A but focusing on PCGFs and CBXs group of protein interactors respectively.

Generation of a Cbx8$^{-/-}$ cell line

As discussed in the introduction, Cbx paralogs have a cell state-specific expression, that reflects in compositional switches in canonical PRC1. Taken together, the transcriptomics and proteomics data suggest that upon PrE commitment, CBX7 is substituted by CBX8 as the main paralog included in the complex (Fig. R9C). Given that CBX proteins are essential for canonical PRC1 recruitment on chromatin we reasoned that CBX8 might play a role in shaping the Polycomb network in PrE. Therefore, we set out to investigate the role of this protein.

To gain insights into the role of *Cbx8* in PrE cells, we generated a *Cbx8* KO mESCs cell line. To do so, by making use of a CRISPR Cas9 technology, we induced two DNA DSBs, separated by a region of ~2 kb, spanning *Cbx8* exons 2 to 5, to completely remove most of the coding sequence (Fig. R10A). Importantly, given that the *Cbx8* gene clusters in the same genomic region of *Cbx2* and *Cbx4* (mm10, Chr 11:119,017,030-119,086,590), we decided to target the gene body without affecting the TSS and the region encoding the first exon, to maintain a normal transcription initiation of the locus, avoiding any possible indirect transcription-mediated. We induced the deletion in *Rnf2*-tagged cells, as we reasoned this would be the most suitable system to investigate any molecular perturbations of PRC1 upon *Cbx8* KO. Upon puromycin selection for successfully transfected cells, we manually picked 96 clones, of which only 66 led to successful genotyping. Upon PCR screening of the *Cbx8* locus, we observed that most of these clones displayed a deletion of the desired size on either one (8%) or both alleles (55%), indicating a high editing efficiency (Fig. R10B). We isolated two clones bearing a *bona fide* clean deletion on both alleles of *Cbx8*, confirmed the size by PCR (Fig. R10C), and sequenced the amplicon. While clone #1 presented a unique sequence, indicating that both alleles had undergone the same deletion (1776 bp), clone #2 presented distinct deletions on each allele, although similar in size (1803 and 1783 bp), along with a few base changes (Fig. R10E).

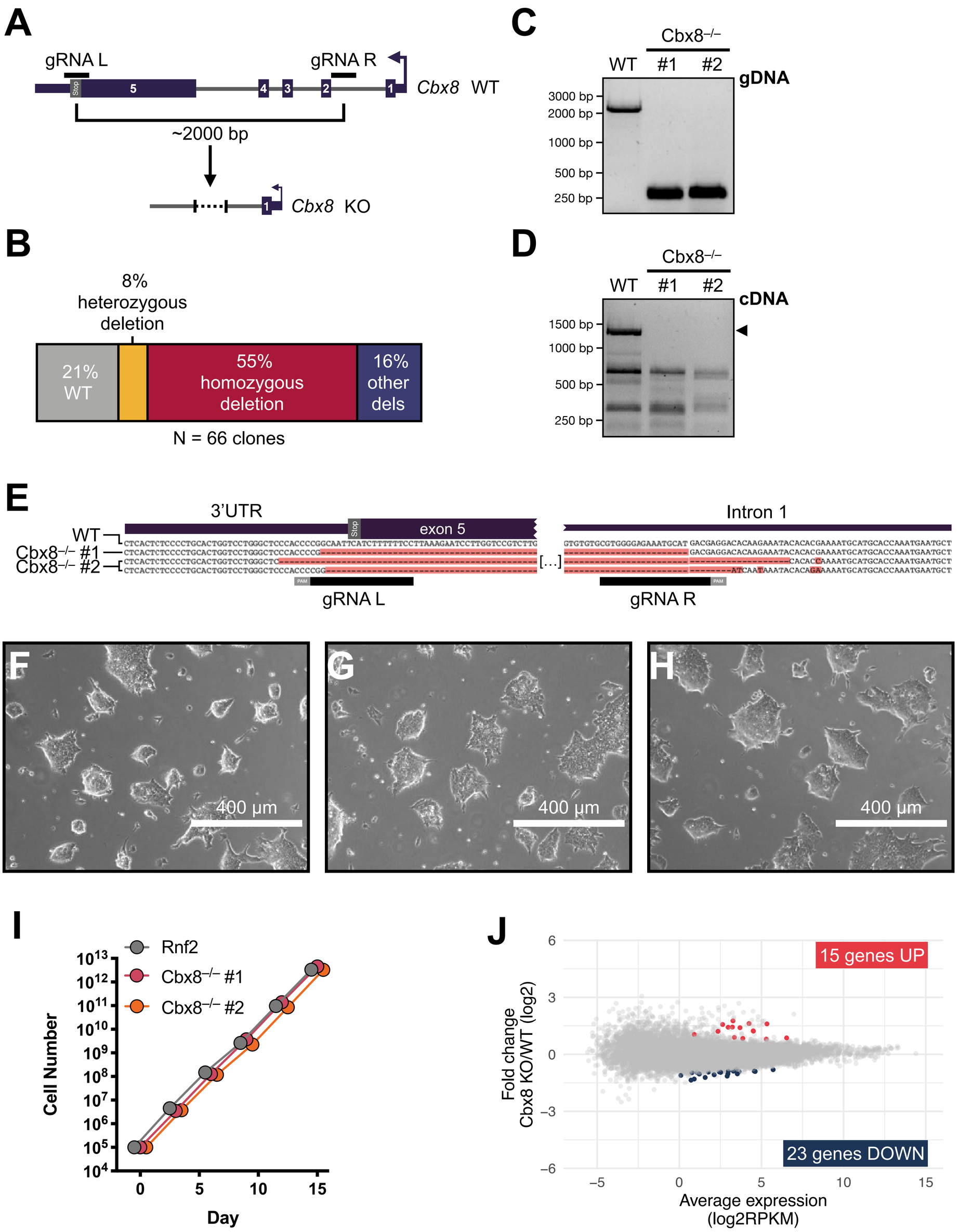


Figure R10 Generation of Cbx8$^{-/-}$ mESCs A Scheme of the CRISPR KO strategy adopted for Cbx8; **B** Allele editing results from the gDNA screening of PURO-selected clones. **C–D** Genotyping of Cbx8$^{-/-}$ clones. **C** PCR of the Cbx8 locus; **D** RT-PCR of the Cbx8 mRNA,. Note: in C–D primers external to the targeted region were used; **E** Sequence alignment of the alleles of WT and Cbx8$^{-/-}$ clones showing the exact deletion generated; **F–H** Picture of WT (**F**) and Cbx8$^{-/-}$ clone #1 (**G**) and #2 (**H**) mESCs grown in Serum/LIF conditions; **I** Growth curve of WT and Cbx8$^{-/-}$ cells; **J** MA plot showing differential gene expression upon Cbx8 KO in mESCs. Upregulated genes are marked in red, while downregulated ones in blu (padj < 0.05)

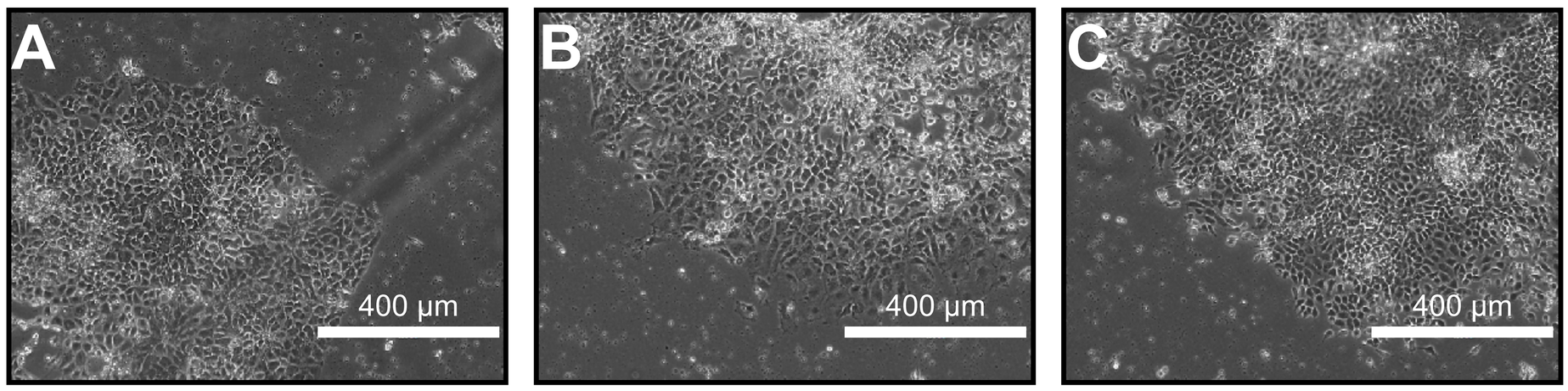
To confirm that the deletion results in a truncated mRNA, we performed RT-PCR on RNA from WT and *Cbx8*$^{-/-}$ mESCs (Fig. R10D): despite the unspecific amplification, none of the two KO clones showed the WT band size. Both *Cbx8*$^{-/-}$ mESC clones adopted the typical dome-shaped ESC colony morphology, when grown in Serum/LIF (Fig. R10F–H) and displayed proliferation rates similar to WT cells (Fig. R10I). To evaluate the impact of *Cbx8* depletion on transcription, we performed RNA-seq using WT and *Cbx8*$^{-/-}$ mESCs. Upon DeSeq2 analysis, we detected only 15 significantly upregulated genes and 23 downregulated ones upon *Cbx8* depletion (padj < 0.05, Fig. R10J).
Overall, this data suggests that *Cbx8* is not essential for ESC pluripotency, proliferation, and transcriptional regulation.

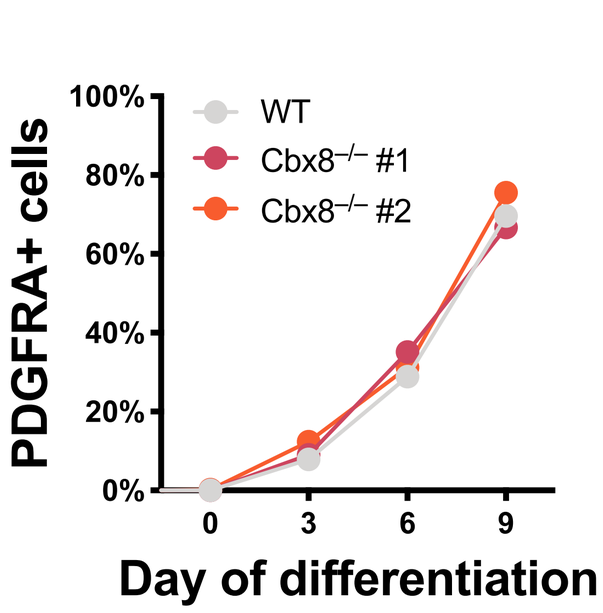
Cbx8 depletion does not affect PrE differentiation

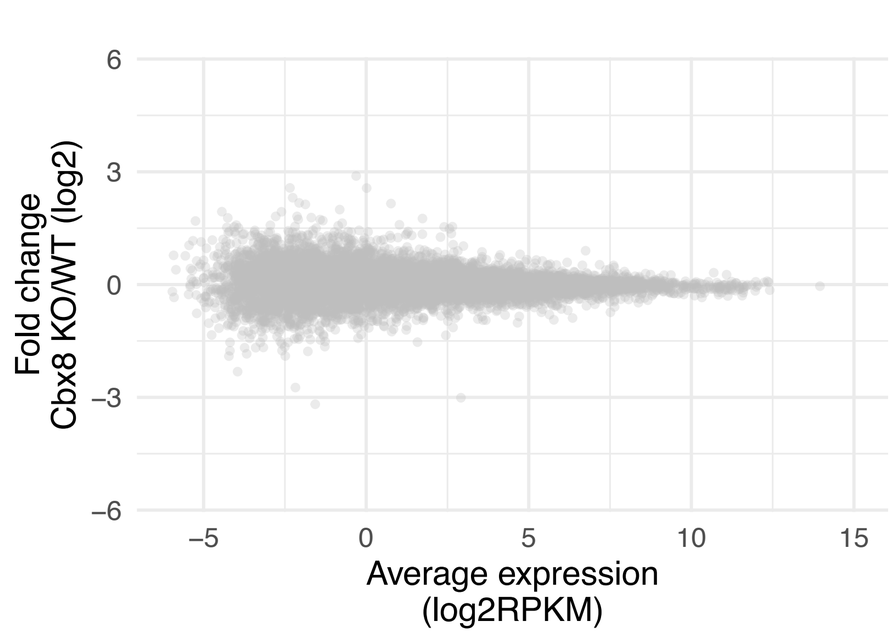
We then tested the ability of *Cbx8*$^{-/-}$ mESCs to generate PrE. We did not observe any major differences in cells morphology (Fig. R11A–C) or in the percentage of PDGFRA^{+} cells (Fig. R11D). Analysis of the transcriptome revealed 0 DEGs (*sic!*) (Fig. R11E), and no GSEA pathway enrichment relevant to PrE biology (data not shown), indicating that PrE produced by *Cbx8* depleted cells is largely similar to that of WT cells. Importantly, we also tested *Cbx8*$^{-/-}$ mESCs ability to derive NPCs and produce all three germ layers upon EBs differentiation, and observed no major defects (Fig. R11F–H)
We reasoned that maybe any defect upon *Cbx8* depletion could be masked by functional compensation of other Cbx family members. However, we did not observe any changes in transcription upon *Cbx8* depletion in mESCs (Fig. R12A) nor in PrE (Fig. R12B), indicating that Cbx paralogs expression is not dependent on CBX8 in these cells. Nonetheless, to rule out this possibility, we generated single CBX member knockdown in either WT or *Cbx8*$^{-/-}$ mESCs. To do so, we prepared lentiviral particles loaded with pLKO-Puro plasmids expressing shRNAs against *Cbx2*, *Cbx4*, and *Cbx7* (plus a nontargeting as control).

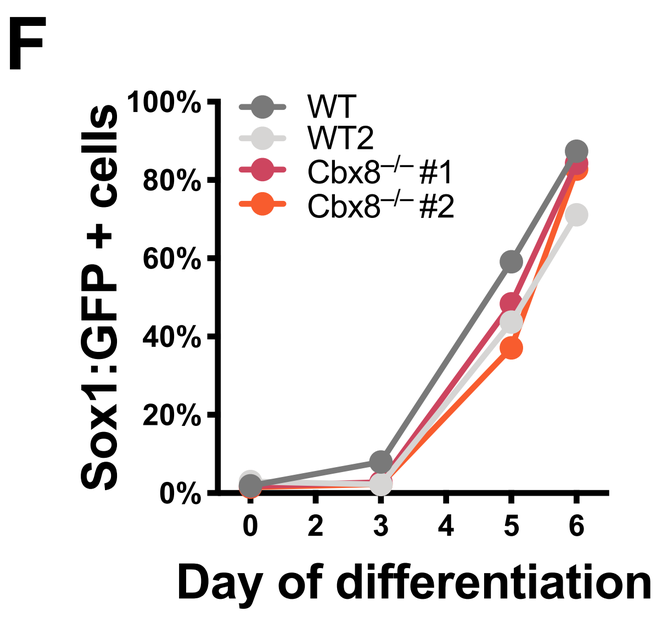
Sequences of the shRNAs were described in (Morey et al., 2012). We did not consider *Cbx6* since this was not incorporated in the complex in ESCs and became even less expressed in PrE (compare Fig. R12A and B). After the infection, a population of puromycin resistant clones was selected. RT-qPCR confirmed the efficiency of the knockdowns and unveiled a certain degree of inter-dependency in Cbx genes expression (Fig. R12C): *Cbx4* expression was reduced upon shCbx2 and shCbx7 and a very similar pattern was observed for *Cbx8*, where knockdown of any of the other Cbx family members resulted in some degree of downregulation. Interestingly, *Cbx7* followed the opposite trend, with higher expression in cells depleted for *Cbx2* or *Cbx4*, suggesting possible internal regulatory feedbacks between Cbx family members. Next, we assessed the capacity of these cells to undergo PrE differentiation. Interestingly, we observed that depletion of *Cbx2* resulted in higher differentiation rates with respect to the control, in both WT and *Cbx8*$^{-/-}$ cells, suggesting that this protein might be counteracting PrE fate choice in mESCs, and that this effect is independent of *Cbx8*. Furthermore, we observed a moderate delay in PrE differentiation in WT cells depleted for *Cbx4* and *Cbx7*, although these effects were partially rescued in *Cbx8*$^{-/-}$ cells. Overall, double depletions did not unmask any phenotype that was not already present in WT cells, indicating that the lack of phenotype in PrE cells upon *Cbx8* depletion is not due to functional compensation between Cbx proteins.

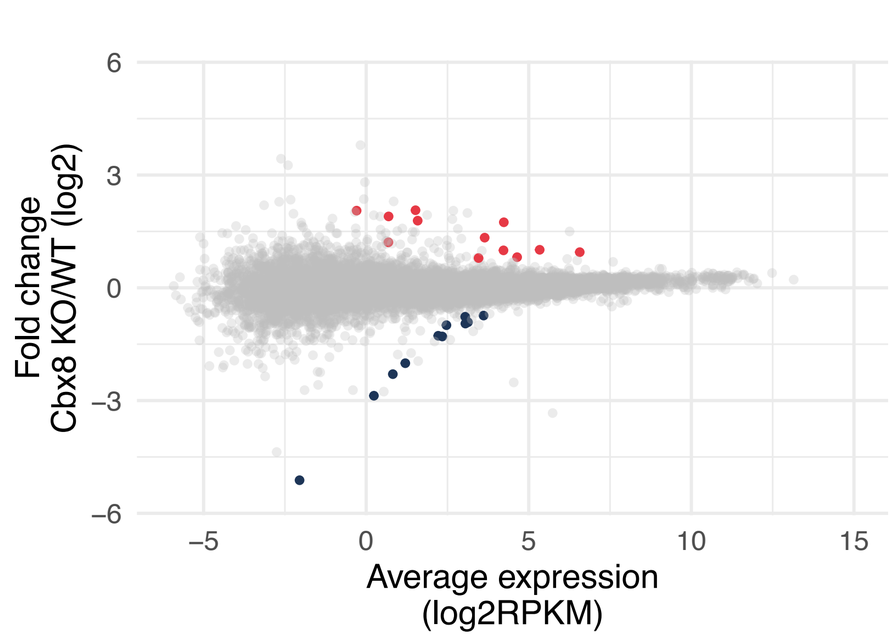
Upon differentiation to PrE, we observed a switch from *Cbx7* to *Cbx8* both in expression and in canonical PRC1 composition (Fig. R9C). Importantly, *Cbx7* is downregulated upon ESC exit from pluripotency (Fig. R1E, R4D) and independently of *Cbx8* (Fig. R12B), suggesting that the switch between the two proteins is regulated at the transcriptional level. Since we were not able to detect the functional relevance of CBX8 in PrE formation, we wondered whether the switch in composition *per se* has any impact in this process. To address this, we forced the expression of *Cbx7*, by introducing a transgene containing a constitutive promoter expressing the coding sequence of Cbx7 fused to a triple flag tag (Fig. R13A).











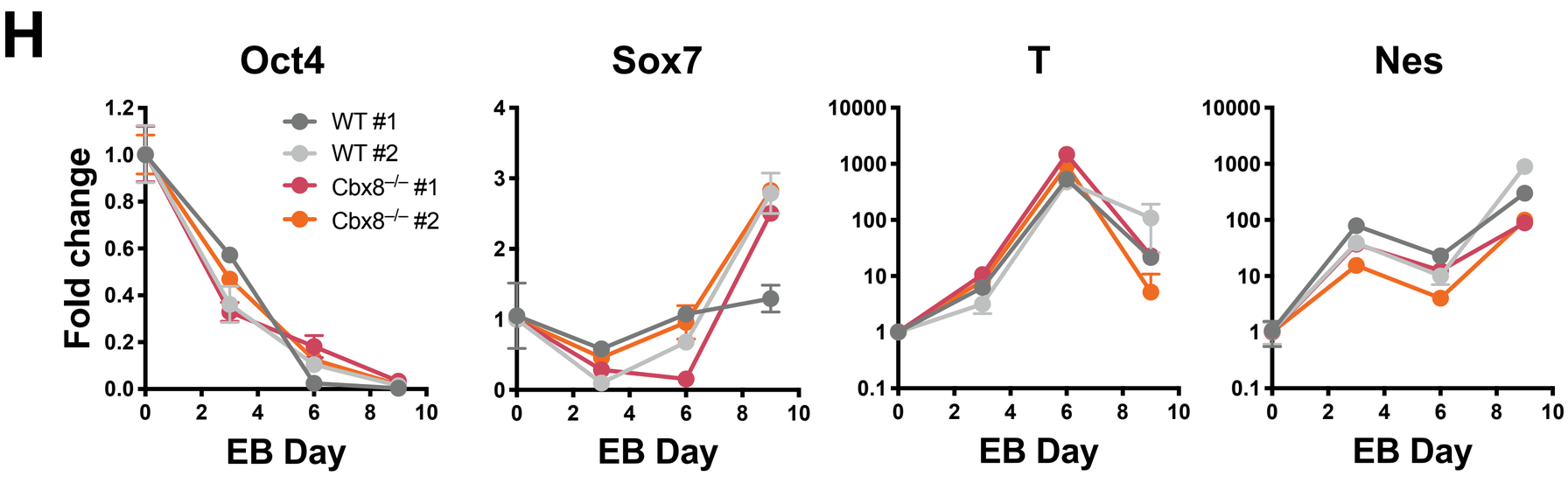


Figure R11 Cbx8 is not necessary for *in vitro* PrE derivation A–C Bright field images of WT (**A**), Cbx8$^{-/-}$ #1 (**B**), and Cbx8$^{-/-}$ #2 (**C**) ESC derived PrE; **D** percentage of PrE differentiated WT and Cbx8 KO cells as measured by FACS using PDGFRA staining; **E** MA plot showing differential gene expression between WT- and Cbx8 KO-derived PrE. Cbx8 gene was excluded from the graph for clearer visualisation (log2FC = -9.28; padj = 4.83 x 10^{-13}). **F** percentage of NPC differentiated cells from WT (Sox1 and Rnf2-tagged lines) and Cbx8$^{-/-}$ clones as measured by the expression of the Sox1:GFP internal reporter; G MA plot showing differential gene expression between WT- and Cbx8 KO-derived NPCs. Cbx8 excluded (log2FC = -7.99; padj = 0.014); **H** qRT-PCR at different time points of embryo body formation of genes representative of pluripotency (*Oct4*), endodermal (*Sox7*), mesodermal (*T*), and neuroectodermal (*Nes*) lineages.

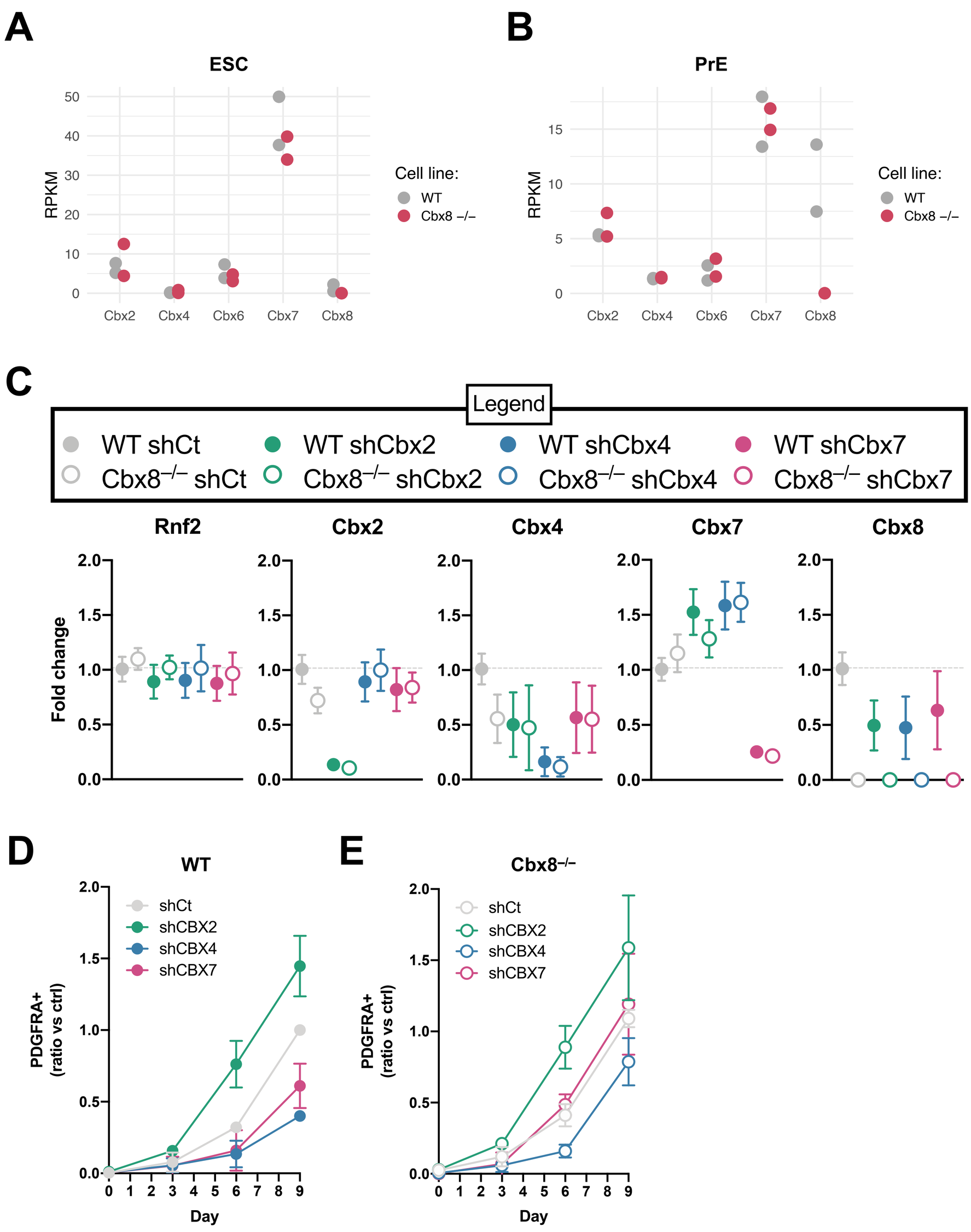


Figure R12 Cbx paralogs do not compensate for the loss of Cbx8 A–B Expression of Cbx paralogs in WT and *Cbx8*$^{-/-}$ mESCs (**A**) and PrE (**B**). Each dot corresponds to one RNA-seq replicate. **C** RT-qPCR of *Rnf2*, *Cbx2*, *-4*, *-7*, and *-8* in various mESCs conditions listed in the legend. The grey dashed line indicate the expression levels of the given gene in the WT shCt condition, which is normalised to 1; **D–E** PrE differentiation efficiency of WT (**D**) and *Cbx8*$^{-/-}$ (**E**) cells upon knockdown of various Cbx paralogs. The values in (**D**) and (**E**) are normalised on the percentage of PDGFRA+ cells of the WT shCt sample per each replicate.

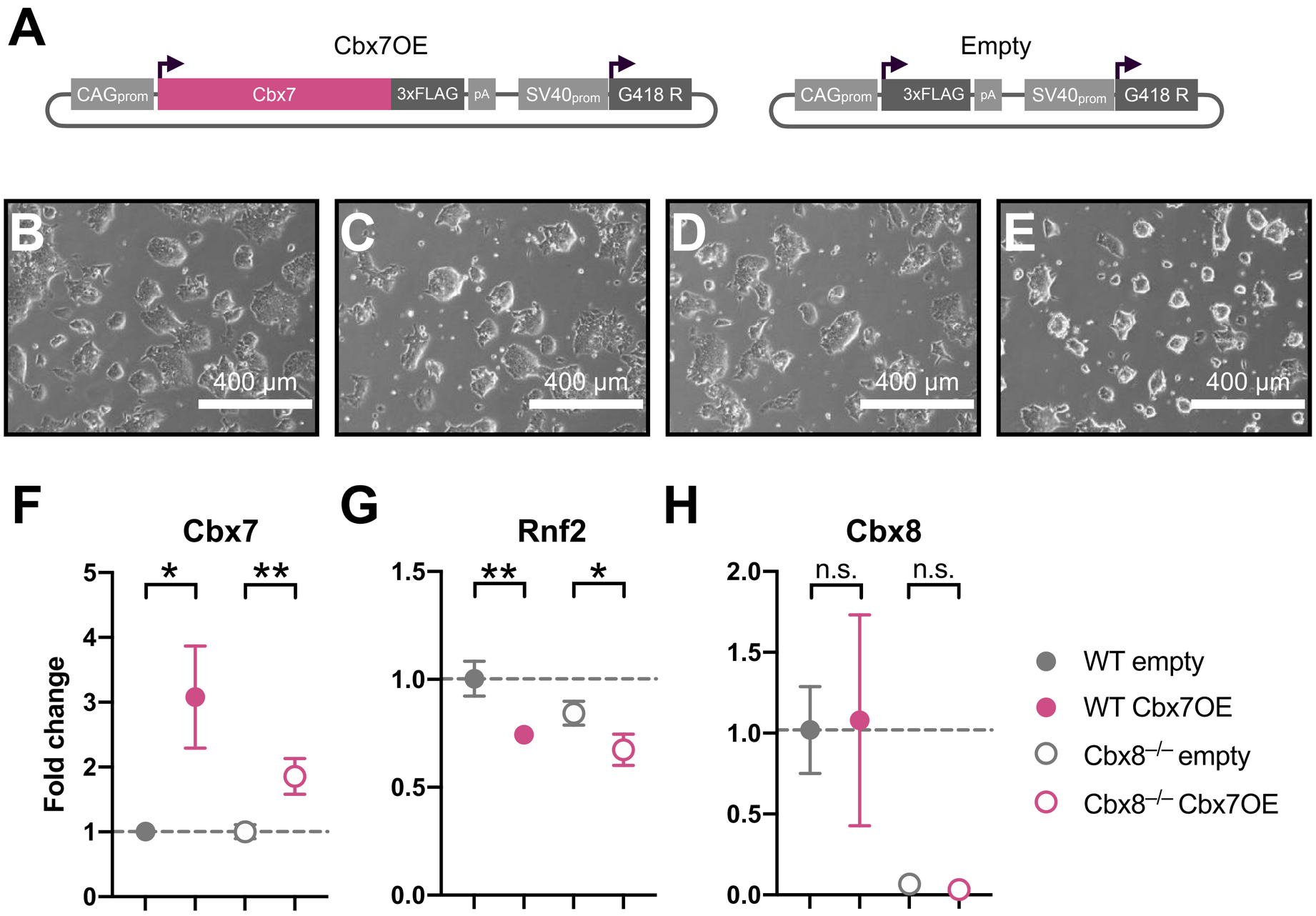


Figure R13 Cbx7 overexpression impairs PrE differentiation capacity A Scheme of the plasmid used for over expression of Cbx7 or of an empty vector; **B–E** Picture of Serum/LIF grown WT mESCs expressing an empty vector (**B**), or Flag-Cbx7 (**C**), and *Cbx8*$^{-/-}$ cells expressing an empty vector (**D**), or Flag-Cbx7 (**E**); **F–H** RT-qPCR of *Cbx7* (**F**), *Cbx8* (**G**) and *Rnf2* (**H**) in the cells in **B–E** (normalized on the average expression of *Rplp0* and *Tbp*). **I** Western blot of whole cell extract from untransfected cells or cells in **B–E**. **J** PrE differentiation efficiency of cells in **B–E** as measure by PDGFRA staining of the cells Values in **J** are normalised on the percentage of PDGFRA+ cells of the WT empty sample per each replicate.

Importantly, the plasmid also contained resistance to G418, therefore, by drug selection, we could isolate the population of cells that had undergone random integration of the transgene. Cells selected for integration of the Cbx7 transgene, or an empty vector showed similar morphology (Fig. R13B–E). Cbx7 overexpression was confirmed by RT-qPCR and Western blot (R13F, I). Importantly, Rnf2 showed a mild albeit reproducible downregulation (Fig. R13G) that resulted in lower RNF2 protein levels (Fig. R13I) independently of the presence or absence of *Cbx8*. Interestingly, we did not observe changes in *Cbx8* expression upon shCbx7 (Fig. R13H), suggesting that additional mechanisms might be involved in the regulation of the Cbx8 locus.

We then tested the impact of *Cbx7* on PrE differentiation and found that Cbx7 overexpression displayed a lower percentage of PDGFRA^{+} cells, regardless of whether this was performed with WT or *Cbx8*$^{-/-}$ cells. This indicates that physiological *Cbx7* downregulation is necessary for an efficient transition to PrE, and that this phenomenon is not dependent on *Cbx8*.

Cbx8 could be involved in further PrE differentiation

Around the time when the embryo is implanted, the PrE will have expanded all around the inner part of the trophectoderm, and along with the egg cylinder, constituting the parietal and the visceral extraembryonic endoderm respectively (Rossant and Tam, 2009) (PExE and VExE, Fig. R14A). These two epithelia will contribute to form the yolk sac, that together with the trophoblast provides support to the early developing embryo, prior to the formation of the chorioallantoic placenta, which will replace the yolk sac (Dickson, 1979; Hogan et al., 1980). Anderson and colleagues have reported that *in vitro* derived PrE is able to acquire a parietal-like phenotype when plated onto fibronectin in N2B27 medium (Anderson et al., 2017).

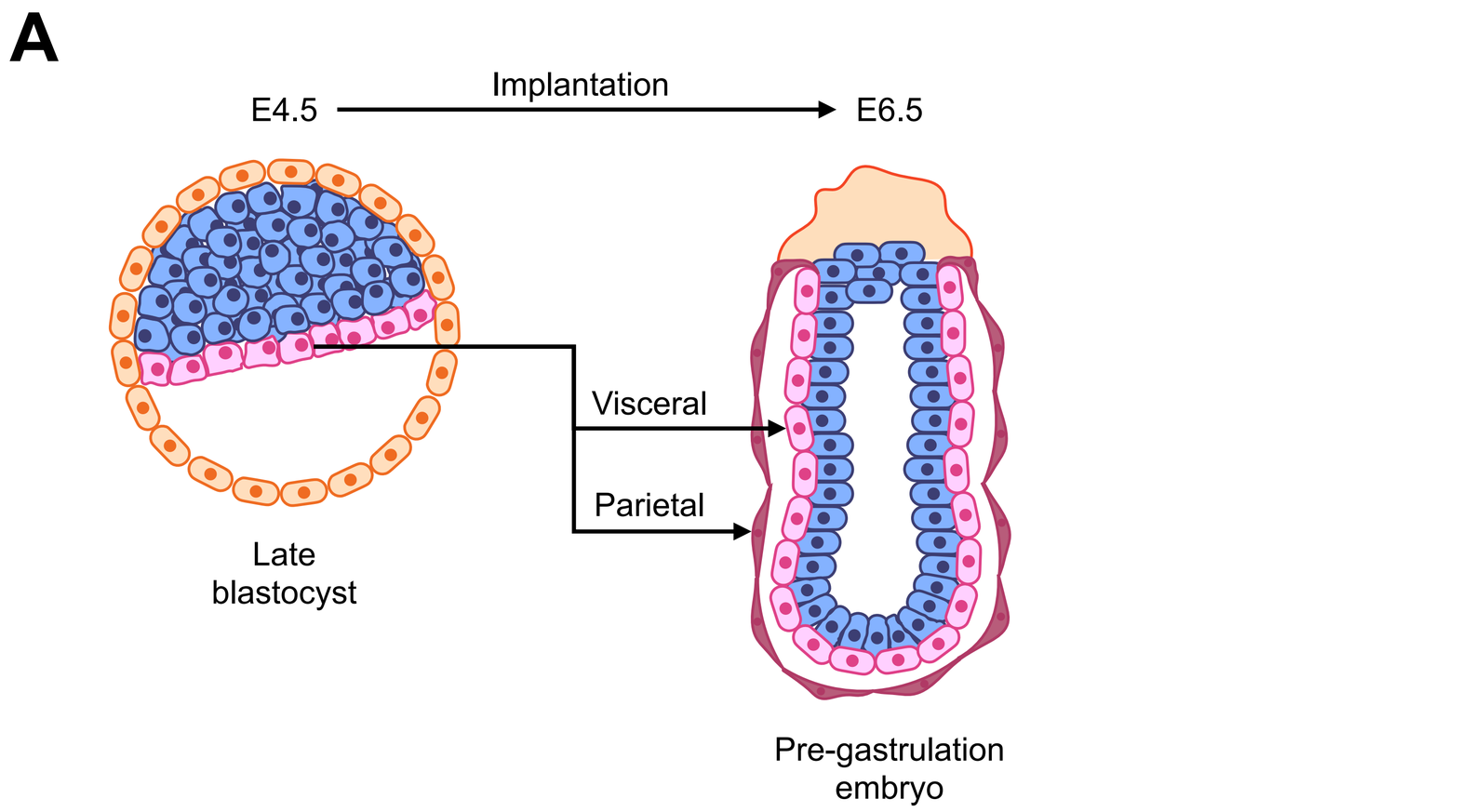


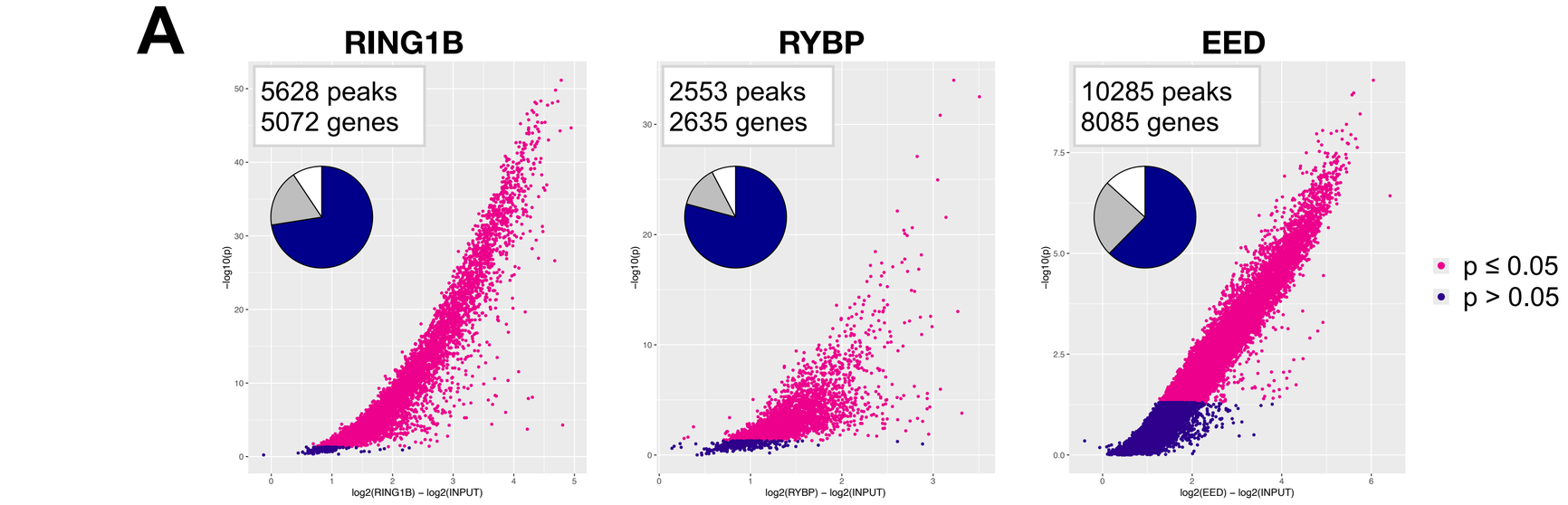
Figure R14 Cbx8 might be involved in PrE differentiation to PExE A Scheme depicting PrE derivation of Parietal and Visceral extra embryonic endoderm over the course of early implantation of the embryo; **B** Scheme of the PExE in vitro differentiation protocol, including subsequent passages (P1, P2, Pn); **C** panel of cell markers expression assessed by RT-qPCR at PrE, PExE, and subsequent PExE passages in WT and *Cbx8*$^{-/-}$ cells (normalized on the average expression of *Gapdh* and *Tbp*). Note: this data is the result of n=1 experiment, therefore no statistical conclusions can be drawn.

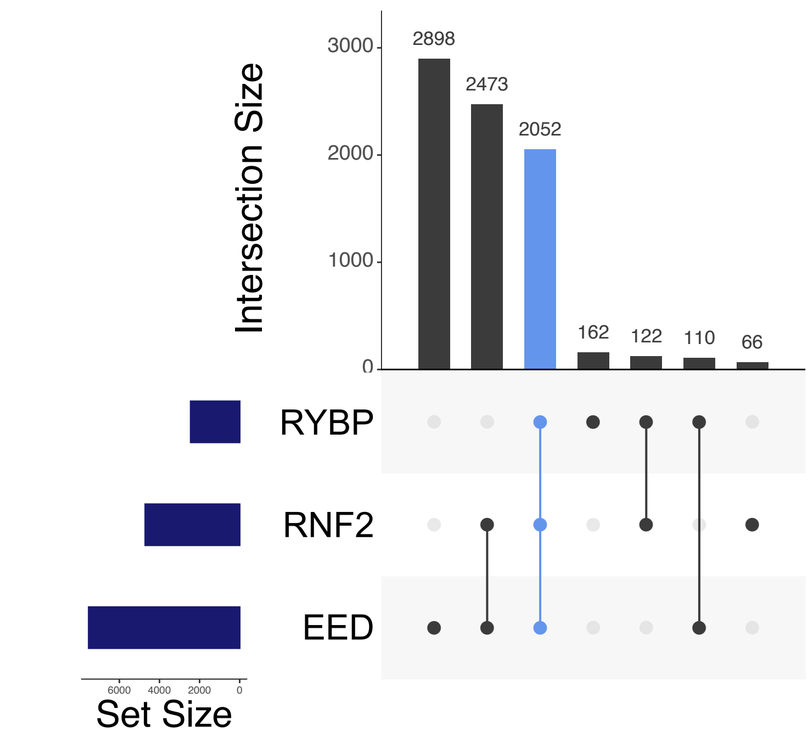
Since *Cbx8* does not have an impact on the determination of PrE, we hypothesized that its presence in a PrE-specific PRC1 complex might be necessary for those epigenetic changes that accompany further differentiation of the PrE. To test this possibility, we derived PrE from WT and *Cbx8*$^{-/-}$ mESCs and then followed differentiation as described, to get PExE cells (Fig. R14B). In addition, we reasoned that PRC1 might be involved in the maintenance of the differentiated state, so we cultured PExE for several passages, collecting RNA at different time points to monitor markers of cell identity. Upon early derivation of PExE (5 days), we observed an initial wave of activation of typical markers of the VExE (*Afp*, *Ttr*, *Hnf4*), although these genes became less expressed right after the first replating (Fig. R14C). Conversely, markers of the general endoderm (*FoxA2*, *Gata6*) and PExE (*Snai1*, *Thbd*) started accumulating at later passages, suggesting that in our case, cells take longer to acquire the PExE phenotype and shut down completely pluripotency factors expression (Fig. R14C). In most of the cases, the genes we tested showed very similar trajectories of expression in WT and *Cbx8*$^{-/-}$ cells, indicating that upon *Cbx8* depletion, PrE retains the capacity to further progress into differentiation. However, for some of the genes, we observed changes in the amplitude and/or kinetics of expression (e.g., *Afp* and *Ttr* at early PExE, *Thbd* at late PExE, Fig. R14C), suggesting that maintenance of the differentiated cell types originating from PrE could be impaired upon *Cbx8* KO.

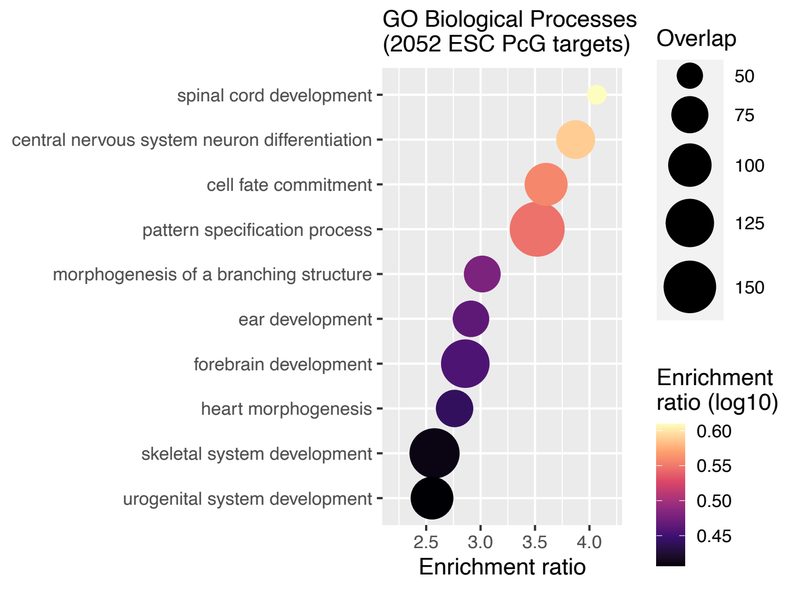
Assessing the role of PRC1 in PrE gene repression

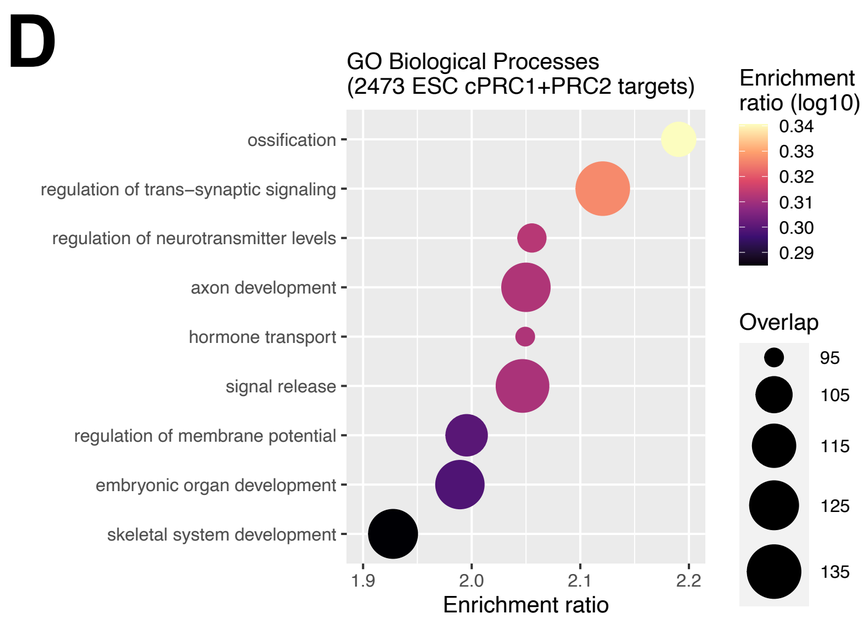
Since Cbx genes are essential for canonical PRC1 recruitment on chromatin (Morey et al., 2012; Tavares et al., 2012), the lack of phenotype upon depletion of the only paralog incorporated in PRC1 in PrE prompted us to ask whether PRC1 has any relevance in the regulation of PrE transcriptional program. To address this point, we characterized genome-wide PRC1 chromatin occupancy using FLAG ChIP-seq in *Rnf2*-, *Cbx8*, and *Rybp*-tagged cell lines at both ESC and PrE stages (differentiation day 9). Moreover, in order to

have a full perspective on the Polycomb network, we also performed EED ChIP-seq in Sox1:GFP cells at both stages.
Importantly, all ChIP-seq experiments were performed using a fixed amount of drosophila chromatin spike-in control. Target regions on chromatin were defined using MACS2 peak caller on individual replicates against an input sample and then evaluated pairwise using DiffBind. With this approach, we identified 5628 peaks for RNF2, 2553 for RYBP, and 10285 for EED. Most of these peaks mapped in the immediate vicinity of a TSS (62 to 79%) or within a gene body (13 to 24%) (Fig. R15A), corresponding to a total of 5072, 2635, and 8085 target genes respectively. No peaks were detected for CBX8 (data not shown), indicating that the ChIP was unsuccessful. This could be due to the fact that CBX8 does not bind chromatin in PrE, or that its chromatin occupancy cannot be detected due to the low abundance of the bait (see results from IP-MS of CBX8), or due to the specific conditions in which the ChIP was carried out. Overlapping of the lists of target genes (Fig. R15B) allowed us to define a set of 2052 common PcG targets (RNF2+RYBP+EED) were canonical and non-canonical PRC1 are present together with PRC2; a set of 2473 canonical PRC1 targets (RNF2+EED only) where only canonical PRC1 is present, likely due to H3K27me3-mediated recruitment; and finally, a set of 162 non-canonical PRC1 targets (RYBP only). As expected, all three sets are enriched for developmental genes, although showing slightly different terms: while common PcG targets belong to all sorts of morphogenetic processes (e.g., central nervous system, ear, heart, skeletal and urogenital systems, Fig. R15C), targets of canonical PRC1 seem to be more related with development and functioning of neurons (Fig. R15D). Instead, we found that non-canonical PRC1 targets are specifically enriched for genes related to gametes generation and meiosis (Fig. R15E), in line with studies showing that PCGF6 bound loci – of which approximately 40% are devoid of RNF2 (Scelfo et al., 2019) – are mainly associated with germ-cell related genes (Endoh et al., 2017; Scelfo et al., 2019; Zhao et al., 2017).









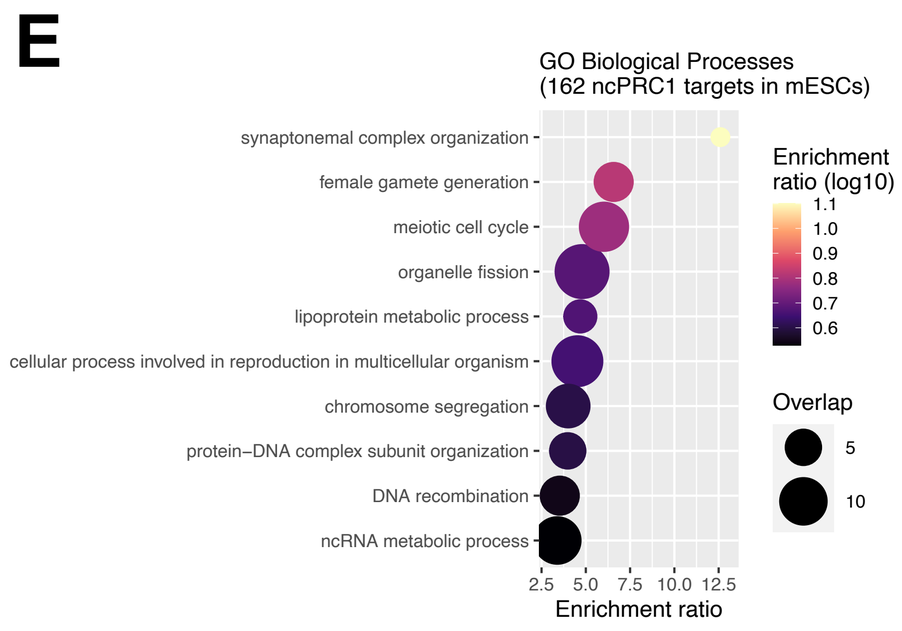


Figure R15 PRC1 chromatin localisation in mESCs A Volcano plot of the peaks detected for RNF2, RYBP and EED ChIP-seq in mESCs, and corresponding pie chart indicating the distribution of the peaks (blue for promoter-proximal, grey for gene body, and white for intergenic peaks); **B** upset plot of the list of genes targeted by RNF2, RYBP and EED in mESCs. The set of genes targeted by all three factors is highlighted in light blue; **C–E** Enrichment analysis of GO biological processes for the common set of PcG targets (**C**), the set of canonical PRC1 (**D**), and the set of non-canonical PRC1 (**E**). Same options as in R4C–D.

We went on to characterize the PcG chromatin localization in PrE. We were able to define 830, 239, and 970 peaks for RNF2, RYBP and EED respectively, the vast majority of which mapped to promoters or gene bodies (Fig. R16A). By analyzing the corresponding target genes (923 for RNF2, 283 for RYBP, and 991 for EED), we observed that the degree of overlap between the three proteins was much lower than that in mESCs (Fig. R16B), with only 90 genes marked by all three proteins, being only the 5th most abundant group (3rd in mESCs). To understand whether this group represents a certain specific subset of ESC PcG targets, we analysed GO biological processes enriched among these 90 genes, using the 2052 mESCs PcG common targets as the background (Fig. R16C): similar to ESC, among the enriched terms, were only embryonic developmental processes, suggesting that Polycomb has to maintain repression of these genes in the extraembryonic lineage.

To get more insights at the single protein level, we compared ESCs and PrE targets for each protein: in all three cases, target genes in PrE were merely a subset of the targets in mESCs, with very few targets being acquired upon PrE differentiation (Fig. R16D). When looking at all the RNF2 target loci in mESCs, we observed that upon PrE differentiation all three proteins suffered a general loss in chromatin occupancy (Fig. R17A), including fundamental targets of PcG repression, like the Hox gene clusters (Fig. R17B). Importantly, loss of PRC1 binding upon PrE differentiation was associated with transcription upregulation (Fig. R17D) suggesting that Polycomb does play a role in the repression of these loci in mESCs and needs to be removed for efficient transcription activation. This was particularly evident at the Hox gene clusters, where the loss of PRC1 and PRC2 was accompanied by a general derepression (Fig. R17B–C). Conversely, only a small fraction of downregulated genes was de novo targeted by PRC1 (43 out of 3098 downregulated genes), indicating that gene repression is not accompanied by acquired PRC1 binding in PrE. Taken together, this data suggests that, apart from a small subset of developmental promoters, gene repression is largely independent of PRC1 in PrE.

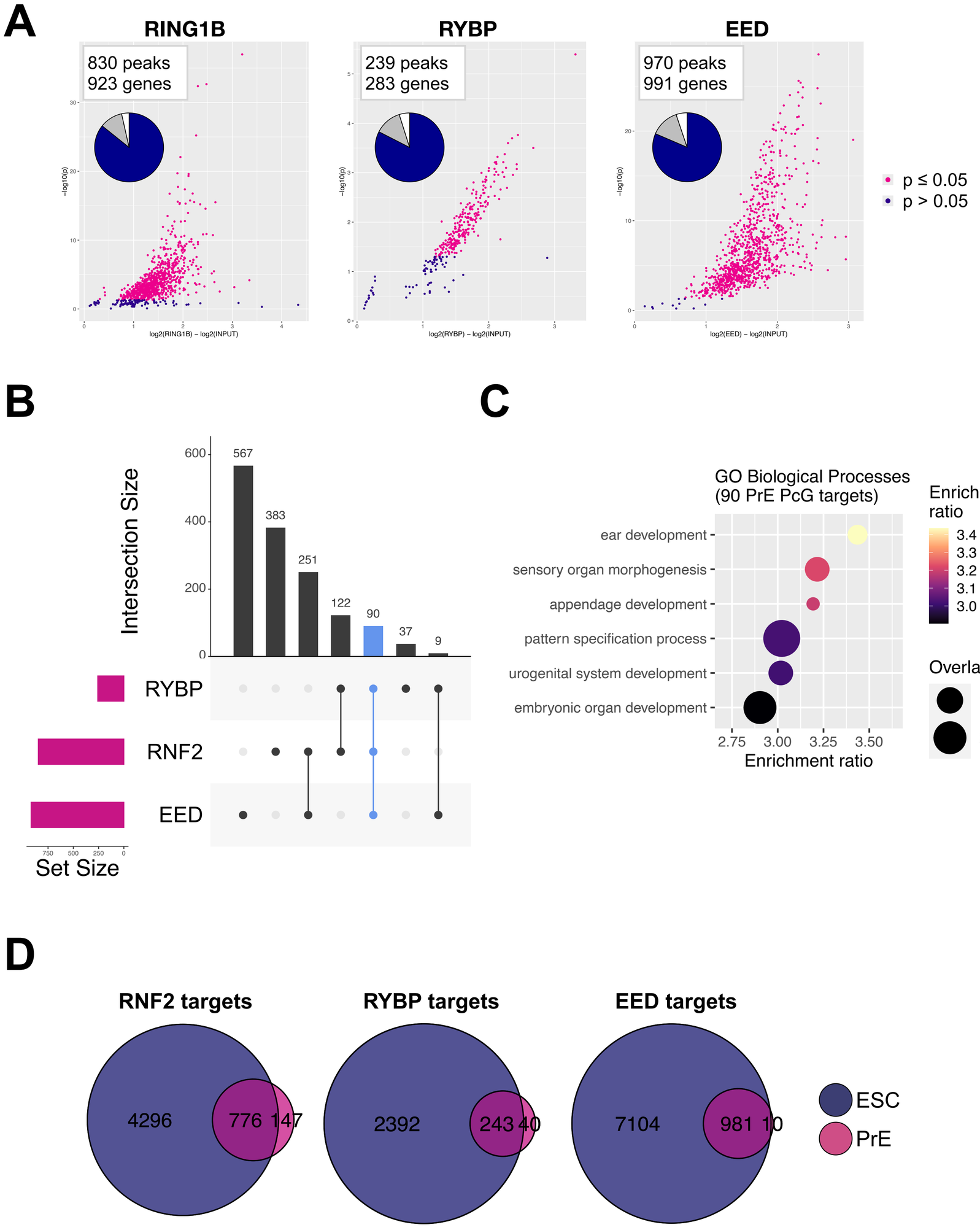


Figure R16 PRC1 chromatin localisation in PrE A Volcano plot of the peaks detected for RNF2, RYBP and EED ChIP-seq in PrE, and corresponding pie chart indicating the distribution of the peaks (blue for promoter-proximal, grey for gene body, and white for intergenic peaks); **B** upset plot of the list of genes targeted by RNF2, RYBP and EED in PrE. The set of genes targeted by all three factors is highlighted in light blue; **C** Enrichment analysis of GO biological processes for the common set of PcG targets in PrE using as background the set of common PcG targets in mESCs; **D** Venn diagram comparing ESC and PrE target genes for RNF2, RYBP and EED.

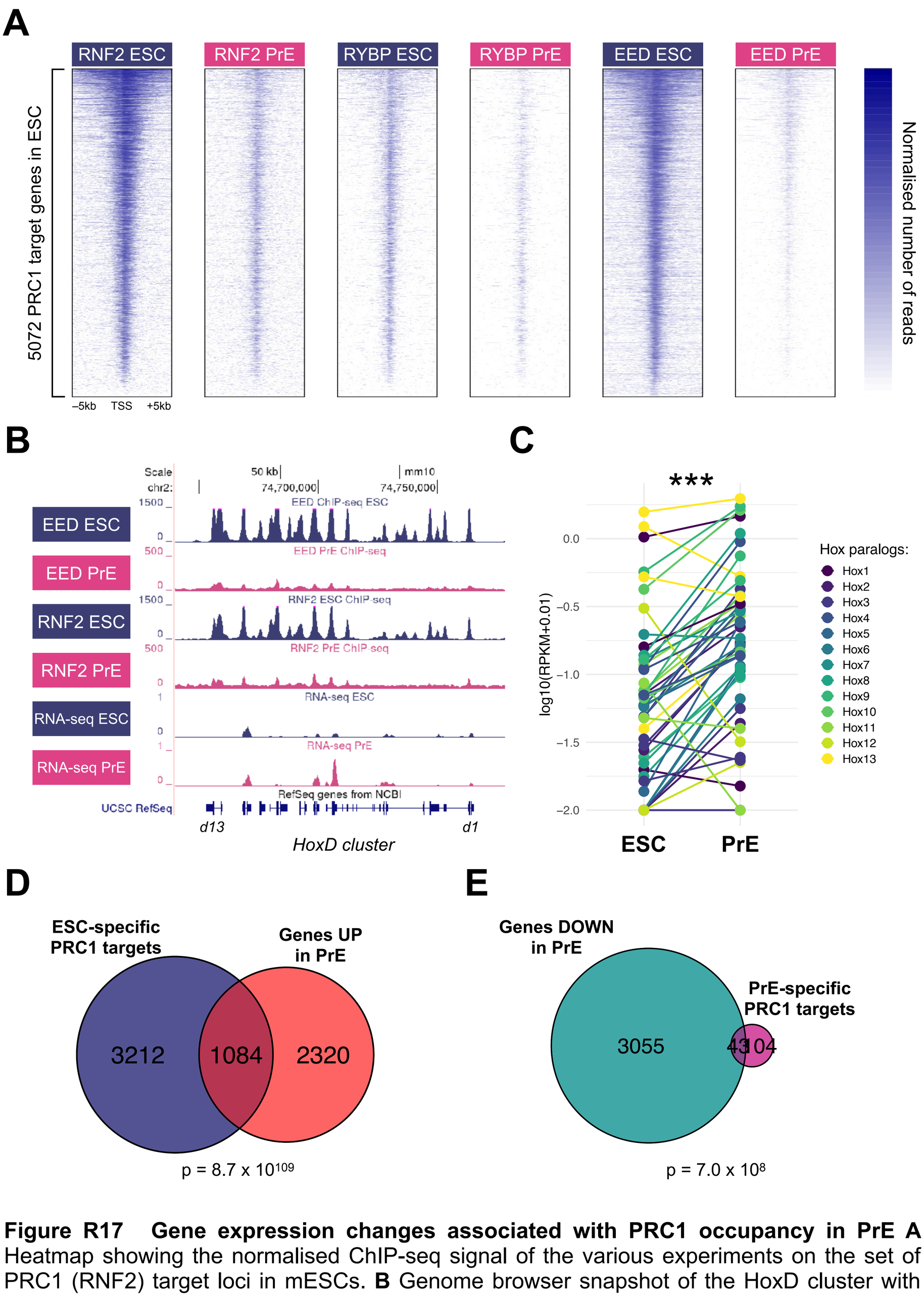


Figure R17 Gene expression changes associated with PRC1 occupancy in PrE A Heatmap showing the normalised ChIP-seq signal of the various experiments on the set of PRC1 (RNF2) target loci in mESCs. **B** Genome browser snapshot of the HoxD cluster with the indicated ChIP-seq and RNA-seq tracks; tracks showed represent the merge of both replicates together; **C** Expression of Hox genes in ESC and PrE: each dot represent one Hox gene, coloured according to the paralog group (*** $p < 0.0001$, 2-way ANOVA); **D** Venn diagram showing overlap between ESC-specific PRC1 (RNF2) targets and upregulated genes upon PrE differentiation (Fisher's exact test); **E** Venn diagram showing overlap between PrE-specific PRC1 (RNF2) targets ant downregulated genes upon PrE differentiation (same as in D)

Results Part II

Disclaimer: the results of the part II regard the findings of a joint project that was carried out together with Dr. Niccolò Arecco (N.A. hereafter), from the lab of Dr. Manuel Irimia (CRG), and that involves the characterization of a novel Suz12 splicing isoform. For the sake of this section, I will exclusively focus on the description of the experiments related with the PRC2 complex formation and activity, excluding those aspects that are out of the scope of this thesis (e.g., splicing regulation, evolutionary analysis), although they will be mentioned in the discussion.

Suz12 exon 4 is a panAS event conserved in mouse and human

In recent years, many structural studies have significantly advanced our understanding of the overall organization and protein-protein interactions that happen within the PRC2 complex. During my PhD thesis, I had the chance to perform a systematic review of the field (Chammas et al., 2020). The work would mainly stem from the finding that interaction of mutually exclusive sets of accessory factors, leading to the formation of PRC2 subtypes PRC2.1 (with EPOP and PCL proteins) and PRC2.2 (with AEBP2 and JARID2), is encoded in the structure of SUZ12 (Chen et al., 2018). SUZ12 C2 and Zn/ZnB domains provide the binding surface for both sets of interactors, in the form of a clamp that encloses their binding domains (Fig. R18A).
Inspection of the vast-db web server (Tapial et al., 2017) revealed that mice and humans can produce two *Suz12* alternative splicing isoforms, depending on the inclusion or exclusion of the exon 4 (hereafter referred to as the 'long' isoform and the 'short' isoform, respectively). These two isoforms were also annotated in the GenBank database (Clark et al., 2016) with the name NM_015355.4 (long) and NM_001321207.2 (short) for human, corresponding to peptides NP_056170.2 and NP_001308136.1, respectively.
Notably, this 69 nucleotides exon encodes for a portion of 23 amino acids of SUZ12 corresponding to the WD binding 1 (WDB1) domain

(Fig. R18B), also termed hinge, that sits on top of RBBP4 WD propeller, connecting the SUZ12 Zn and C2 domains (Chen et al., 2018; Kasinath et al., 2018) (Fig. R18A). We reasoned that because of its localization in the structure, the lack of this portion could have the potential to determine changes in SUZ12 conformation, with a clear impact on PRC2 holocomplex formation.
Regarding the splicing pattern of *Suz12* exon 4, we observed that this splicing event is alternative in most of the samples of vast-db, with a percentage spliced in (PSI) value in the range of 75 to 95% across mouse embryonic and adult tissues (Fig. R18C–D), and cancer cell lines (Fig. R18E). Moreover, by RT-PCR we confirmed that lab-grown mESCs retain alternativity in this event and maintain the same PSI upon neural differentiation (Fig. R18F). Notably, splicing events showing alternativity in most of the tissues, have been defined as panAS: interestingly, panAS are more conserved between humans and mice compared to splicing events that are not as alternative across tissues. Moreover, they are enriched in transcription and chromatin related factors (Tapial et al., 2017), therefore we argue that *Suz12* exon 4 is a panAS and might be important for SUZ12 activity as a chromatin modifier.

Generation of a Suz12 ∆exon4 mESC cell line

To address this point, we decided to generate mESCs depleted of *Suz12* exon 4 (∆exon4). We used a CRISPR KO approach, targeting two gRNAs around the exon, to induce a deletion of around 500 bp encompassing exon 4 and the surrounding intronic regions (Fig. R19A). Upon screening for the successful deletion, we selected three clones that had undergone deletion of the exon on both alleles of *Suz12* (∆exon4 #1–3), resulting in the sole expression of the short isoform (Fig. R19B). Additionally, as a control, we also isolated three isogenic clones that did not show deletion of the exon (WT-like #1–3). Importantly, while WT-like #1 and #2 clones expressed both the long and the short isoform of *Suz12*, WT-like #3 only expressed the long isoform (Fig. R19B) suggesting that, while deletion of the exon was unsuccessful, excision repair around the region of

the cut might have influenced the regulation of exon 4 splicing (see discussion).
We confirmed that deletion of exon 4 results in the production of a shorter protein, corresponding to the lower band of SUZ12 (Fig. R19C). Despite the lower expression of *Suz12* upon deletion of the exon 4(Fig. R19B), the relative expression of the short isoform becomes higher than in WT cells, resulting in a partial compensation of the total SUZ12 protein (Fig. R19C, see quantification below the SUZ12 blot). Moreover, upon exon 4 deletion, MTF2 was partially destabilized, and cells showed a dramatic loss in bulk H3K27me3 deposition (Fig. R19C). Cell fractionation analysis revealed that in WT cells the long isoform is the main chromatin isoform, although upon deletion of the exon 4 and forced expression of the short isoform, this latter could be found on chromatin as well (Fig. R19D). Interestingly, cytoplasmic fractions of both WT and ∆exon4 cells displayed a band the height of SUZ12 short isoform (Fig. R19D) raising the possibility that SUZ12 short isoform could be cytoplasm-specific. This hypothesis is currently being tested in our lab and won't be further discussed here.

Suz12 exon 4 is necessary for PRC2.2 holocomplex formation

To characterize the impact of SUZ12 exon 4 depletion on PRC2 holocomplex formation, we performed SUZ12 IP-MS in WT (parental WT and WT-like #1–2), and ∆exon4 cells (clones #1–3). As expected, we did not detect any peptide mapping on the exon 4 in ∆exon4 clones (Fig. R19E), and could confirm that the amount of bait retrieved – as evaluated by either the total number of SUZ12 peptide spectral match (# PSM) or the area of the three most abundant SUZ12 peptides, (TOP3 Area) – was tendentially lower in ∆exon4 cells then in WT (Fig. R19F), although the difference was not statistically significant (#PSM $p = 0.11$, TOP3 Area $p = 0.23$, unpaired t-test).

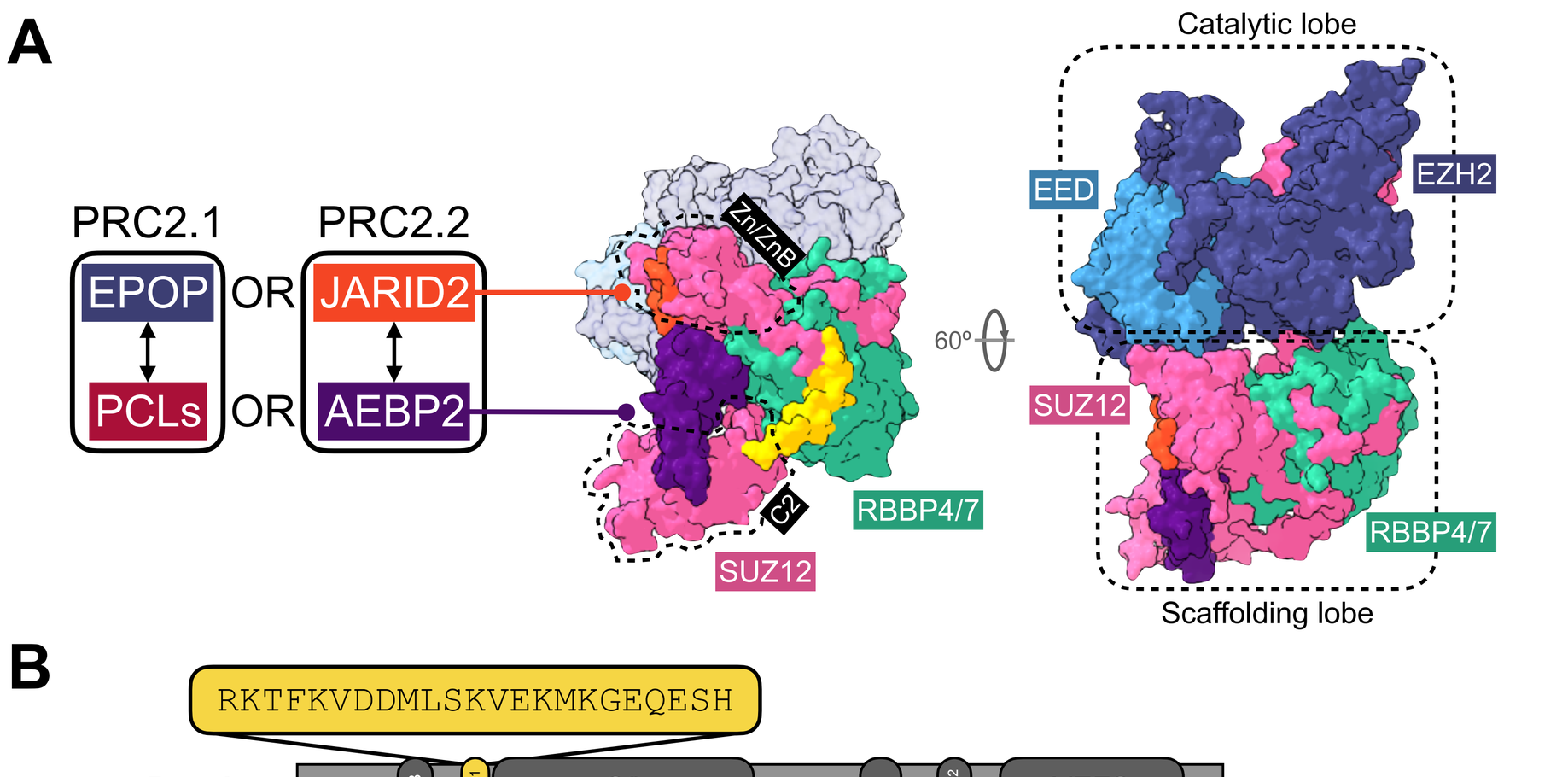


Figure R18 Suz12 exon 4 is a panAS A–B A 3D structure of PRC2 highlighting the scaffolding lobe composed of SUZ12 (pink), RBBP4 (green), JARID2 TR domain (orange, analogous to EPOP CTR), and AEBP2 C2B domain (indigo, analogous to PCLs RC domain). Dashed lines indicate SUZ12 Zn/ZnB (upper) and C2 (lower) domains. The region encoded by *Suz12* exon4 is highlighted in yellow, corresponding to WDB1 (**B**). **C–E** *Suz12* exon4 inclusion levels across embryonic (**C**) and adult cell types (**D**), and selected cancer cell lines (**E**); **F** RT-PCR to evaluate the PSI of *Suz12* exon 4 upon mESCs differentiation tu neurons

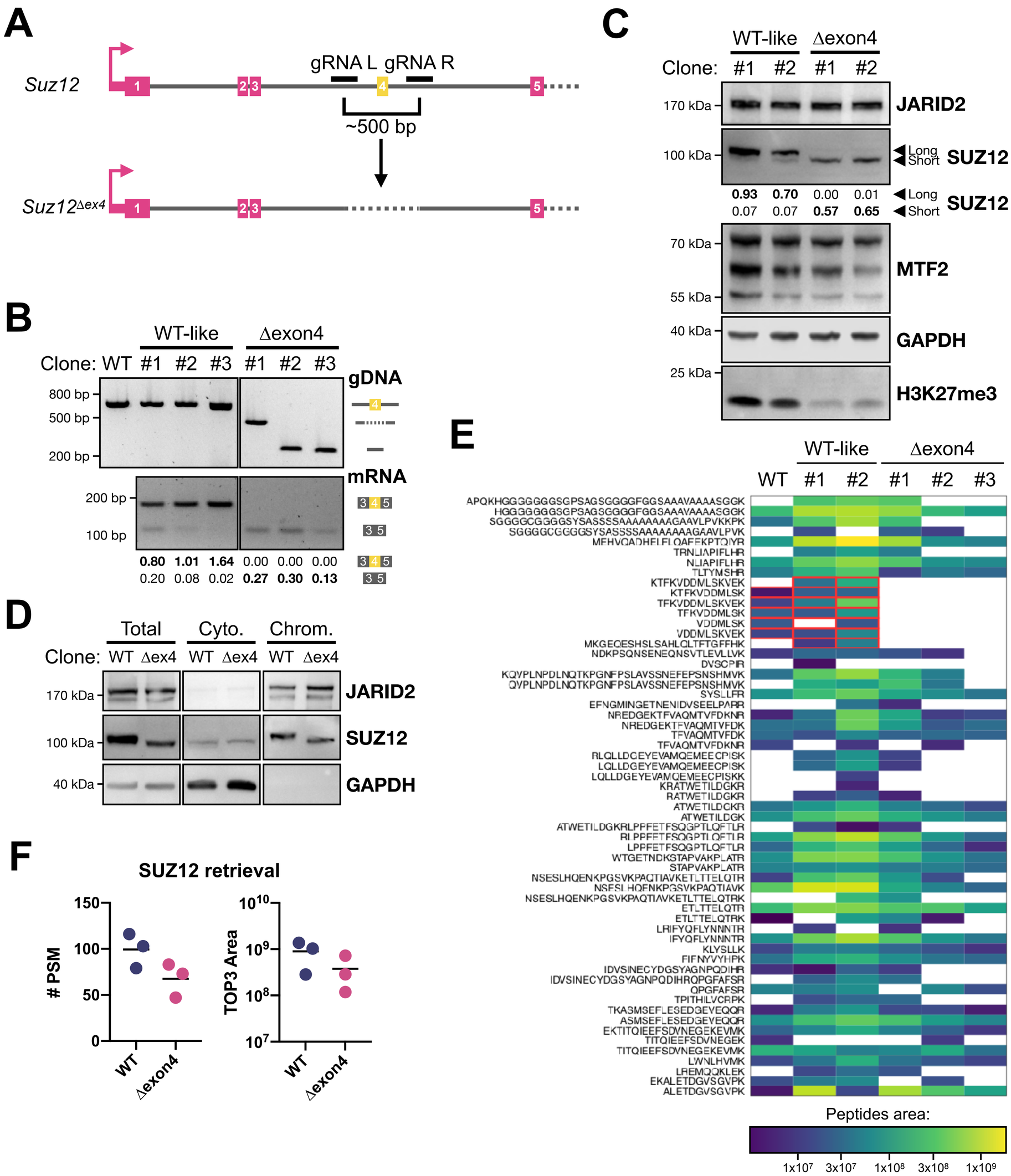


Figure R19 Generation of Suz12 Δexon4 cell line A Scheme of CRISPR-K.O. approach used to remove *Suz12* exon 4; **B** Genotyping of the targeted genomic locus and the resulting *Suz12* mRNA. WT indicates the parental mESCs line, while WT-like indicates isogenic clones that have undergone CRISPR editing unsuccessfully. Δexon4 indicates those CRISPR edited clones showing deletion of *Suz12* exon 4; **C** Western Blot of whole cell extract of WT-like (clones #1–2) and Δexon4 (clones #1–2) cells. The two arrowheads indicate the long and short isoforms of SUZ12, corresponding to the inclusion and exclusion of the exon 4 respectively (see also the quantification below SUZ12 blot). **D** Western blot of total, cytoplasmic and chromatin fractions of WT and Δexon4 (clone #1) cells; **E** Heatmap showing the intensity (area of the spectrum) of SUZ12 peptides identified in SUZ12 IP-MS in each sample, ordered according to their N- to C-terminal appearance in the sequence. Peptides encoded by the exon 4 are marked in red; **F** Quantification of SUZ12 total abundance in the IP-MS in WT (WT, WT-like #1, and #2) and Δexon4 samples, assessed by number of peptide spectral match (#PSM) or by the area of the top 3 spectra (TOP3 Area). Each dot represents one biological replicate.

We went on to evaluate the abundance of PRC2 components in the two sets of samples: no changes were observed in the interaction of SUZ12 with core factors EZH1/2, EED, and RBBP4/7. Moreover, PRC2.1-specific accessory factors EPOP, MTF2, PHF19, and TCEB1/2 (ELOB/C) were unchanged or slightly increased in the case of TCEB2. Strikingly, the interaction of PRC2.2 components AEBP2 and JARID2 dropped dramatically upon exon 4 depletion (Fig. R20A), suggesting that this exon is important for PRC2.2 complex formation. While interaction with AEBP2 was almost completely abolished (from a total of 72 PSMs in the three WT replicates, to 4 in ∆exon4), JARID2 binding was 4 to 5-fold reduced (243 PSM in WT down to 54 in ∆exon4). This data suggests that a SUZ12 isoform lacking exon 4 has a lower affinity for PRC2.2 accessory factors.

IP-WB of SUZ12 and JARID2, confirmed that a JARID2-containing PRC2 is still present in ∆exon4 cells (Fig. R20B–C), albeit to a lower extent (see EZH2 amounts in lane 4 and 5, Fig. R20C). Importantly, disruption of AEBP2 binding upon Suz12 exon 4 depletion did not lead to the formation of a hybrid complex containing JARID2 and MTF2 (Fig. R20C) as observed in *Aebp2* KO mESCs (Grijzenhout et al., 2016). Taken together, this data suggests that SUZ12 short isoform is not able to form a holo-PRC2.2, since binding with JARID2 is reduced and interaction with AEBP2 is completely abolished.

Generation of a Suz12$^{-/-}$ mESCs cell line

To further characterize the functionality of SUZ12 short isoform, we decided to produce a *Suz12* full KO mESC line, as a reference for the complete loss-of-function of PRC2. To do so, we adopted a similar approach as for exon 4 deletion, targeting a region of approximately 400 bp encompassing exons 2 and 3 of Suz12, as we observed that a 1-to-4 or a 1-to-5 spliced transcript would eventually lead to frameshift early on (Fig. R21A).

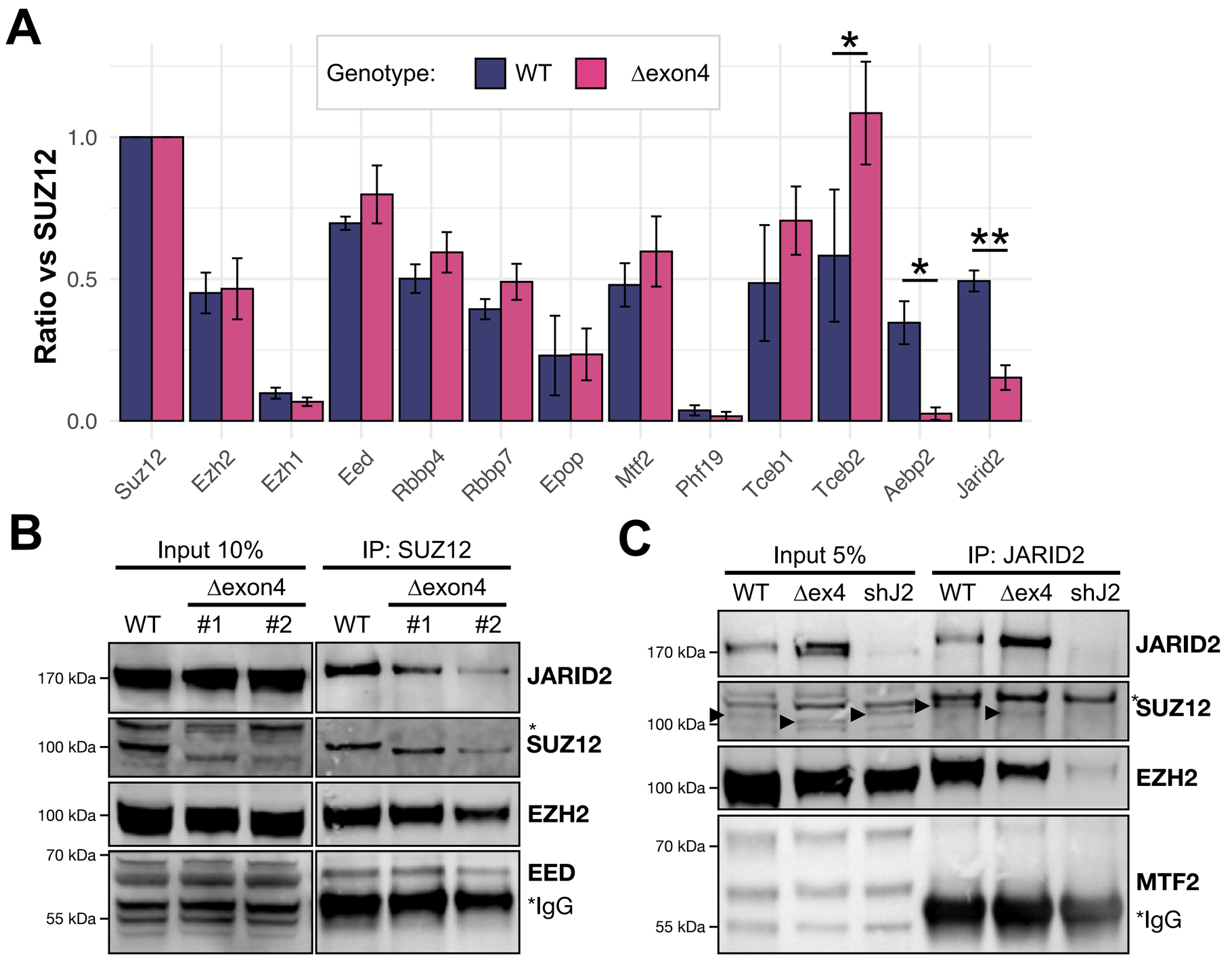


Figure R20 Suz12 exon4 regulates PRC2 interaction with accessory factors A Bar plot showing the ratio of the peptides belonging to PRC2 components with respect to the bait of the IP-MS (SUZ12) (*p < 0.05; ** p <0.01); **B** Western blot of SUZ12 IP in WT and Δexon4 (clones #1–2) mESCs (*unspecific band); **C** Western blot of JARID2 IP in WT, Δexon4 (clones #1) and mESCs expressing a constitutive shRNA against JARID2 (shJ2, Beringer et al., 2016).

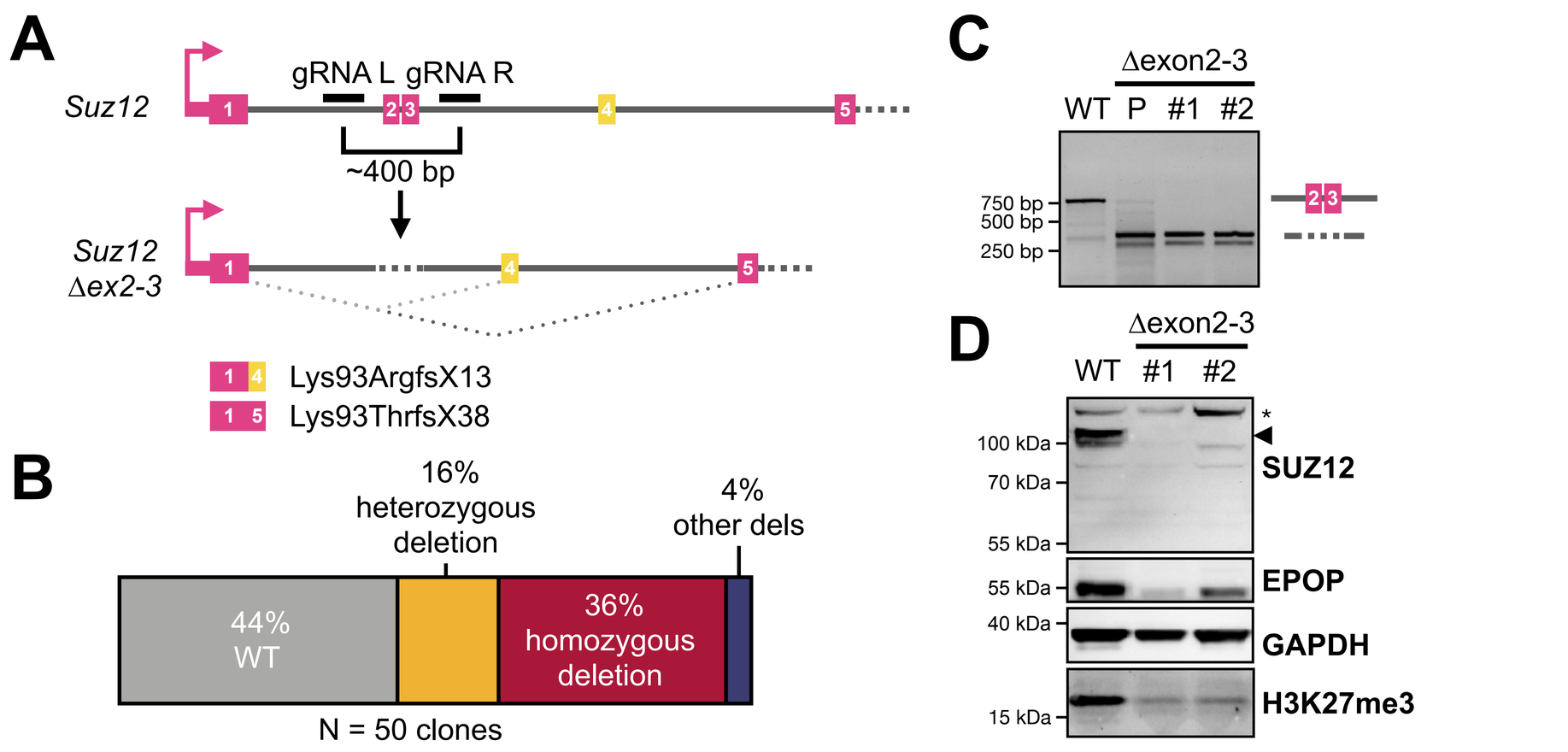


Figure R21 Generation of Suz12$^{-/-}$ mESCs A Scheme of the CRISPR KO approach adopted for *Suz12*, indicating the possible isoforms produced upon deletion of exons 2–3; **B** Allele editing results from the gDNA screening of manually picked CRISPR edited clones; **C** Genotyping of *Suz12* Δexon2-3 cells, including the pool of edited cells and selected clones #1 and #2; **D** Western blot of WT and Δexon2-3 clones #1 and #2.

Deletion of the locus was very efficient, and we isolated several clones displaying homozygous deletion of both exons (∆exon2-3, Fig. R21B–C, note Fig. R21C lane 2 showing that the band corresponding to the deletion is predominant in the pool of transfected cells). We confirmed that the *Suz12* ∆exon2-3 allele is effectively a *Suz12* full KO (*Suz12*$^{-/-}$*),* which results in a dramatic drop in H3K27me3 bulk levels (Fig. R21D). Moreover, we observed that EPOP protein levels drop upon *Suz12* depletion, in accordance with its protein stability depending on its inclusion in the PRC2 complex (Hojfeldt et al., 2019).

Lack of Suz12 exon 4 results in hypermethylation of Polycomb targets

To try to explore changes in H3K27me3 deposition upon *Suz12* exon 4 manipulation we performed H3K27me3 ChIP-seq in WT-like, ∆exon4, and Suz12 KO cells. We are still in the process of analyzing the data, therefore, all the observations are still very preliminary.

To our surprise, when looking at H3K27me3 levels at Polycomb target promoters, we detected an increase in cells lacking exon 4, compared to WT (Fig. R22A), in stark contrast with the reduction of H3K27me3 bulk levels that we observed by Western blot (Fig. R19C). Conversely, WT-like #3 clone that expressed only the long isoform (Fig. R19B), had lower levels of H3K27me3 on the same sites. We confirmed these observations by performing H3K27me3 ChIP-qPCR and observed that the increase in H3K27 methylation upon depletion of exon 4 was specific of PcG targets (Fig. R22B).

Taken together this data suggests that the complex formed around the short SUZ12 isoform is more proficient in catalyzing H3K27 methylation at Polycomb marked loci while depositing fewer methylation marks elsewhere.

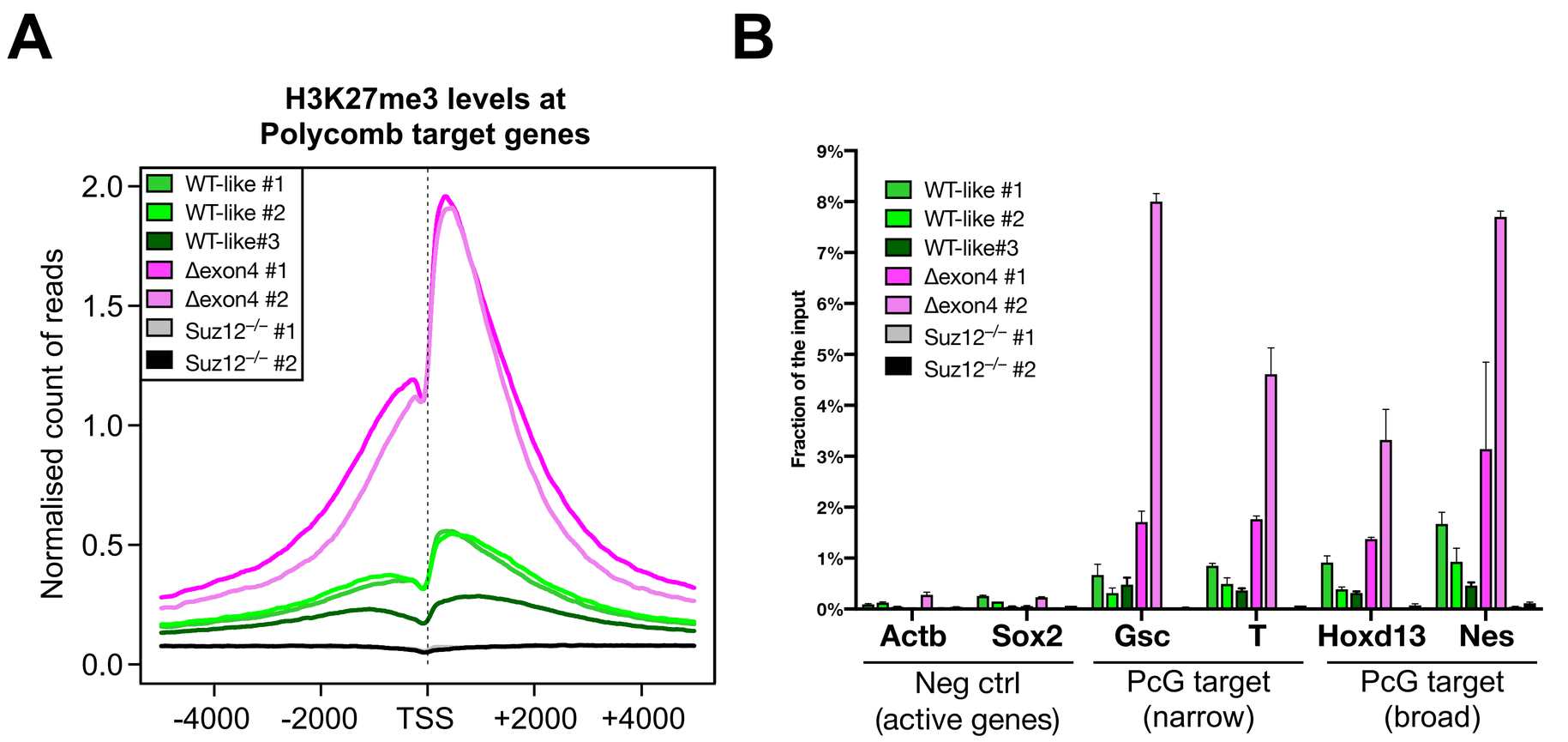


Figure R22 Suz12 exon 4 is regulates PRC2 activity on chromatin A Metaplot of H3K27me3 ChIP-seq signal over the TSS (+/– 5 kb) of Polycomb target genes (n = 5186 genes as defined using DiffBind on H3K27me3 in WT cells); **B** H3K27me3 ChIP followed by qPCR to check levels at Polycomb target promoters (narrow and broad) or untargeted (active) genes.

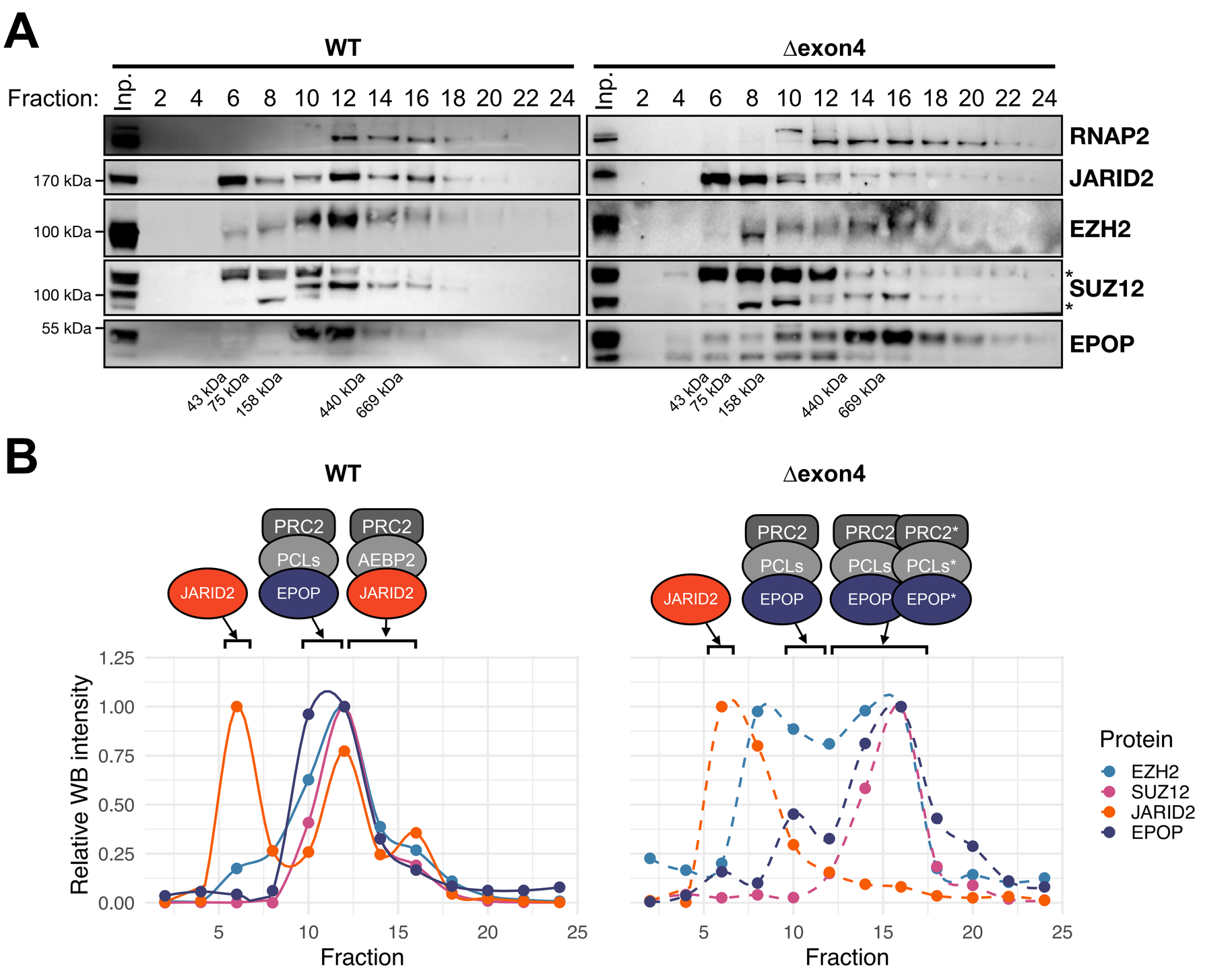


Figure R23 Suz12 exon4 regulates PRC2 dimerisation A Western blot of glycerol gradient fractionated nuclear extracts of WT and Δexon4 mESCs. **B** quantification of the blot intensity of selected proteins from A, together with a cartoon depicting the interpretation of the gradient.

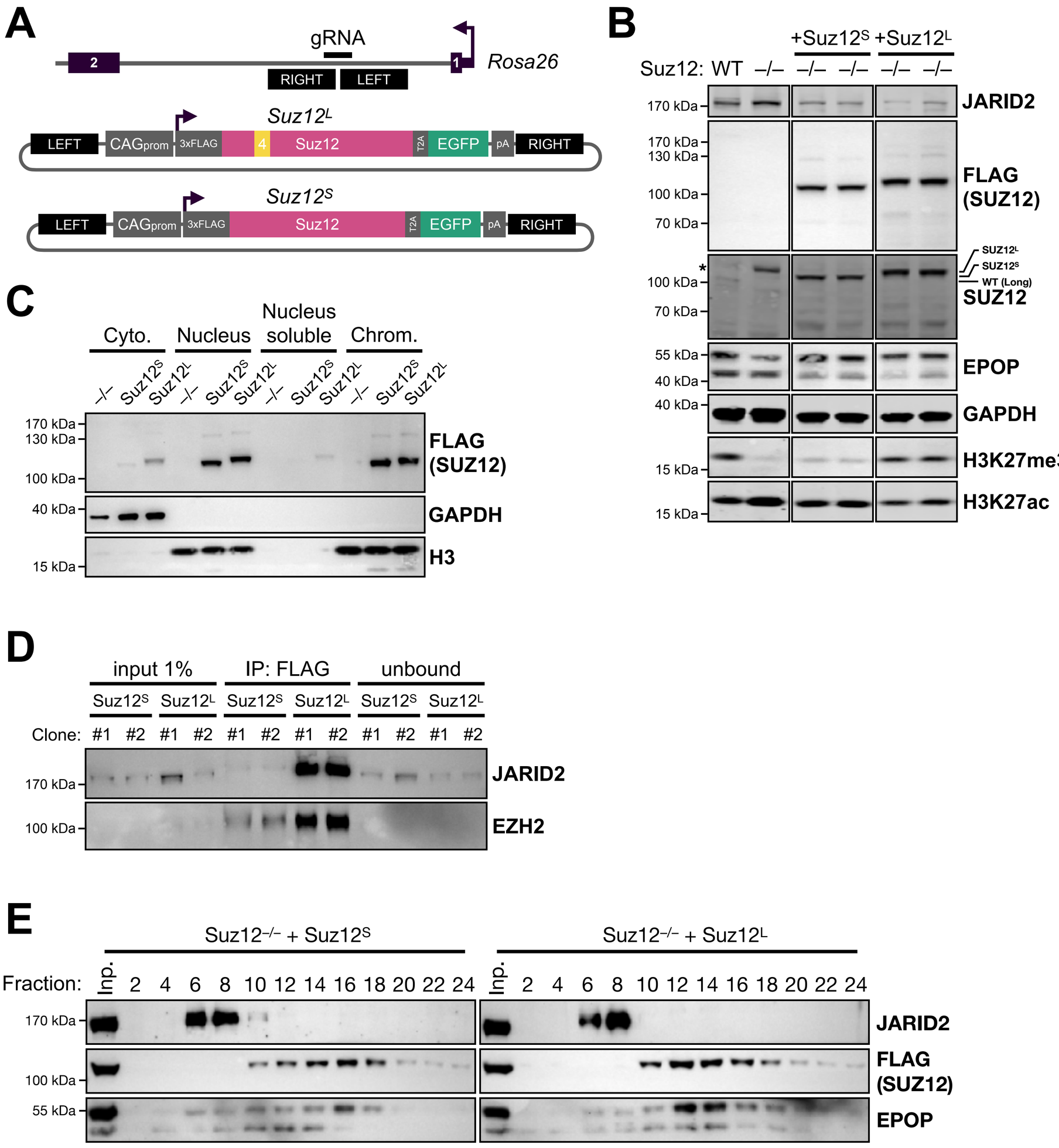


Figure R24 Generation of isoform-specific Suz12 rescue mESC cel lines A Scheme of the CRISPR KI strategy. **B** Western blot whole cell extract of WT, *Suz12*$^{-/-}$, and Suz12^{S} or Suz12^{L} rescue cell lines (*unspecific band); **C** Western blot of cytoplasmic, nuclear, nucleoplasmic and chromatin fractions of *Suz12*$^{-/-}$ and Suz12^{S} or Suz12^{L} rescue cell lines; **D** Western blot of FLAG IP in Suz12^{S} or Suz12^{L} rescue cell lines (2 clones each); **E** Western blot of glycerol gradient fractionated nuclear extracts of Suz12^{S} or Suz12^{L} rescue cell lines.

Suz12 exon 4 influences PRC2 propensity to dimerize

PRC2 ability to dimerize relies on the capacity of the SUZ12 C2 domain to adopt different orientations (Chen et al., 2020). Notably, in the structure of the antiparallel PRC2 dimer resolved by Chen and colleagues, the chain connecting the C2 domain, and the Zn domain is not resolved (Chen et al., 2020). Moreover, the authors infer that the length of this tract should be shorter than the normal hinge length that would allow C2 folding on its own protomer (Chen et al., 2018). This suggests that the distance between SUZ12 Zn and C2 domains, which is mostly covered by the region encoded by exon 4, might be critical for PRC2 propensity to dimerize. Therefore, we set out to test PRC2 capability to dimerize upon exon 4 manipulation. To do this, we performed glycerol gradient fractionation of nuclear extracts from WT and ∆exon4 mESCs. We confirmed that in WT cells EZH2 and SUZ12 co-sediment around fractions 10–14, with EPOP being present mainly at fractions 10–12, and JARID2 at 12–16 (Fig. R23A), compatible with a theoretical weight of approximately 360 kDa for PRC2.1 and 460 kDa for PRC2.2. As expected, upon exon 4 depletion, JARID2 was lost from fractions 12–16, and enriched in fractions 6–8, in line with the disruption of the PRC2.2 complex. However, we observed that SUZ12, EZH2 and EPOP, whilst still being present at fractions 10–12, were now mostly enriched at fractions 14–16 (quantified in Fig. R23B), compatible with the theoretical weight of a PRC2.1 dimer (~720 kDa).

To confirm the observations upon exon 4 genetic manipulation, we decided to test a different strategy: starting from *Suz12*$^{-/-}$ cells, we rescued expression of either the short or the long *Suz12* isoform (Suz12^{S} and Suz12^{L}, respectively). The transgene used contained *Suz12* coding sequence fused to a triple FLAG tag for membrane detection and immunoprecipitation, and an eGFP to facilitate selection of successfully recovered clones, connected via a self-cleavable T2A linker (Fig. R24A). The transgene was inserted in *Suz12*$^{-/-}$ mESCs via CRISPR Cas9-mediated knock-in in the *Gt(ROSA)26Sor* locus, to allow for stable expression (Fig. R24A). We isolated various clones and selected a couple for each isoform

to perform experiments. We confirmed by Western Blot the presence of the tagged SUZ12 protein (Fig. R24B). Notably, while rescue of the long isoform restored nearly WT levels of H3K27me3, the short isoform failed to do so, showing only a mild increase. Importantly, this recapitulates observations made in the ∆exon4 cells, indicating that the short isoform alone is not sufficient to establish WT bulk levels of H3K27me3 (Fig. R19C, R24B). Conversely, H3K27ac levels (increased upon *Suz12* KO), were lowered upon rescue with the long, but not the short, isoform (Fig. R24B), mirroring the changes observed in the methylation levels. We confirmed that both the long and short isoforms are mostly present on chromatin (Fig. R24C) and that JARID2 co-immunoprecipitated with the long but not the short SUZ12 isoform (Fig. R24D). Notably, the amount of EZH2 retrieved was lower in short than long SUZ12 isoform IP, suggesting that there might be less PRC2 formation upon expression of the short isoform alone (Fig. R24D). We then tested protein complex fractionation in these cells: in $Suz12^{S}$ rescue cells, SUZ12 and EPOP co-sediment around fraction 16, while in $Suz12^{L}$ rescue cells they are found in fractions 12–14 (Fig. R24E), confirming the hypothesis that the short isoform assembles a PRC2 complex with higher molecular weight. Strangely enough, we could not detect JARID2 at PRC2-compatible molecular weights, in none of the rescues. This is particularly unexpected, at least for $SUZ12^{L}$ cells, since we observed Co-IP of JARID2 with the long isoform (Fig. R24D) and deserves further investigation.

Overall, this data indicates that expression of specific Suz12 isoforms recapitulates most of the observations made upon exon 4 manipulation (Table R3), thus reinforcing the hypothesis that Suz12 alternative splicing of exon 4 generates two functionally distinct proteins that have specific characteristics in terms of PRC2 composition, multimerization, and, eventually, activity on chromatin.

Table R3 *Summary of the characteristics observed in the various cell lines generated (* variable between clones; n.t. : not tested)*

Cell line	L:S ratio	PRC2.2 formation	Dimer formation	Bulk H3K27me3	Target H3K27me3
WT	90:10	+++++	+	+++	+++
WT-like #1,2	90:10*	+++++	+	+++	+++
WT-like #3	98:2	n.t.	n.t.	n.t.	+
∆exon4 #1–3	0:100	+	++++	+	+++++
$Suz12^{-/-}$ #1,2	0:0	–	n.t.	–	–
$Suz12^{-/-}$ + $Suz12^{S}$ #1,2	0:100	+	++++	+	n.t.
$Suz12^{-/-}$ + $Suz12^{L}$ #1,2	100:0	+++++	+	+++	n.t.

Discussion

Approaching Polycomb complexity

Around the time when this thesis will be discussed, 80 years will have passed since the identification of the very first *ante litteram* Polycomb drosophila mutant (*esc*, *Eed* in mouse) (Slifer, 1942). Since then, a plethora of genes have been added to the growing list of PcG factors, a group that nowadays, in mouse, counts more than 30 groups of paralogs, for a total of more than 60 factors, if we include all the interactors found associated to PcG complexes. In an attempt to contribute to making sense of this complex system, in this thesis I have tried to explore certain compositional switches among the polycomb repressive complexes, be it within the same cell type, or during cell fate transitions. In the first chapter, I have focused on the changes occurring at PRC1 compositions upon commitment of mESCs towards the extraembryonic endoderm, and how these could be relevant for its biology. I have found that, mESCs display a huge variety of PRC1 factors, whereas only a small subset of them are present in PrE cells. Despite a reduced complexity, I was able to identify switches in PRC1 composition, especially within the canonical part: for instance, whilst canonical PRC1 in mESCs incorporates both BMI1 and MEL-18 (probably preferring the latter), in PrE, BMI1 is the only PCGF paralog incorporated in this complex. Interestingly, this switch is also observed upon mESCs differentiation to NPC (Kloet et al., 2016), where BMI1 is necessary for their maintenance and correct differentiation (Bruggeman et al., 2005; Zencak et al., 2005). Conversely, MEL-18 is maintained during the transition towards the mesodermal lineage and, eventually, cardiac differentiation (Morey et al., 2015), while being substituted for BMI1 in hematopoietic stem cells, where it is important for self-renewal (Oguro et al., 2010). However, these two proteins display a high degree of similarity (65% sequence identity and 93% structural similarity), have similar transcription repressor activity (Scelfo et al., 2019), and they both inhibit RNF2 E3 ubiquitin ligase activity (Taherbhoy et al., 2015). Therefore, further studies will be needed

to address the significance of a PCGF exchange within canonical PRC1.

Another interesting switch is the one represented by Cbx paralogs: upon exit from pluripotency, *Cbx7* is usually downregulated, leading to derepression of the genomic cluster containing *Cbx2/4/8* (Morey et al., 2012). In most of the cases, loss of CBX7 in more committed cells is supplied by either CBX2 or CBX4: this is the case for embryoid body formation (Morey et al., 2012), post-implantation epiblast cells (A. Lackner, unpublished), differentiation to neuroectodermal (Kloet et al., 2016; Morey et al., 2012), and mesodermal lineages (Morey et al., 2015). The data presented here, indicates that CBX8 is the PrE-specific paralog that gets incorporated into PRC1 (Fig. R7–9, Table 2). It is worth noting that despite being present in the PRC1 interactome of pluripotent cells, CBX8 is a not among the main interactors, and is not necessary for the regulation of the steady-state levels of the transcriptome in mESCs (Fig. R10). It is therefore telling that it is maintained in PrE, in spite of a major interactor like CBX7, raising the question of whether CBX8 is necessary for PrE gene regulation. However, RNA-seq data of *Cbx8* KO PrE cells reveal that this is not the case. Double depletion of *Cbx8* in combination with other Cbx paralogs resulted in no unmasking of the phenotype (Fig. R12), reinforcing the observation that *Cbx8* is unnecessary for PrE differentiation. Nevertheless, this preliminary data suggest that *Cbx2* might eventually play part in this process: *Cbx2* depletion results in a higher differentiation efficiency, suggesting that the lack of *Cbx2* might be involved in preventing premature PrE differentiation of mESCs. Interestingly, this is similar to what is observed upon enzymatic inhibition or depletion of EZH2 (Illingworth et al., 2016), raising the possibility that a CBX2-containing PRC1 might cooperate with PRC2 at specific target loci. Further investigation will be necessary to prove this hypothesis.

Upon Retinoic acid-induced exit from pluripotency, CBX8 was shown to outcompete CBX7 from chromatin, supporting the activation of Polycomb targets (Creppe et al., 2014). We tested whether this might be the case in the transition to PrE. However, results

suggest that the presence of CBX8 is not sufficient to counteract differentiation defects arising from Cbx7 overexpression (Fig. R13), although, we cannot exclude that the extent of our CBX7 overexpression might be saturating any potential competition by CBX8.
Notably, *Cbx8*$^{-/-}$ cells were able to undergo differentiation to NPCs without any impediment and could correctly form EBs, giving rise to all three germ layers (Fig. R11F–H), suggesting that *Cbx8* might not be necessary at all for early embryonic development. To our knowledge this has not been directly addressed yet. However, a report from Tan and colleagues studying the role of *Cbx8* in leukemogenesis, adopted a Cre-inducible Cbx8 KO mouse model generated by the lab of Dr. H. Koseki (Tan et al., 2011). Strangely enough, no general characterization of this mouse model has been published so far.

Gene repression in Primitive Endoderm

Upon differentiation to PrE, PRC1 suffers a dramatic drop in chromatin occupancy, while maintaining the targeting to a small subset of gene promoters, mainly related with development of embryonic tissues (Fig. R16). One possible explanation to this is that PRC2 factors are largely downregulated upon PrE differentiation (Fig. R4F). In line with this, XEN cells present barely detectable levels of H3K27me3 (Rugg-Gunn et al., 2010). Lack of this mark could, in turn, explain the loss of canonical PRC1 from chromatin (including the fact that we were not able to detect any ChIP signa for CBX8). However, despite the loss in Polycomb binding, we did not observe a disproportioned transcription activation (roughly the same number of up- and downregulated genes upon PrE), suggesting that a different mechanism might be responsible for the silencing of newly repressed genes. Needless to say, given the general loss of PRC1 from chromatin, it is not surprising that Cbx8 depletion has no effect on the transcriptome of PrE cells (Fig. R11).
Regarding the possible alternatives to gene repression, one possibility is given by DNA-methylation. Indeed, upon exit from pluripotency, those bivalent genes that were not activated, are likely to gain

DNA methylation at their promoter (Mohn et al., 2008). This suggests a chain of events in which bivalent genes that are not activated, become stably silenced through DNA methylation. Interestingly, we detected DNMT3A, -B and -L in the interactome of RNF2 in PrE. However, interaction between RNF2 and DNMT3A was not confirmed by IP Western blot (data not shown) and therefore not investigated any further.

Another possible mechanism could be provided by H3K9me3 deposition. Of note, this mechanism was already proposed to explain gene silencing upon lack of H3K27me3 in XEN cells (Rugg-Gunn et al., 2010). Interestingly, although overall H3K9me3 levels were slightly reduced upon differentiation to PrE, we observed Co-IP of this modification with RNF2 in PrE but not in mESCs, raising the possibility that the chromodomain of CBX proteins might be involved in this targeting. Of note, while CBX4 and CBX7 have high affinity for H3K9me3 *in vitro*, CBX8 is not able to bind this histone mark (Bernstein et al., 2006b; Kaustov et al., 2011), complicating even more the interpretation of this piece of data. Early reports describing the composition of PRC1.6 identified G9A (EHMT2) and HP1γ (CBX3), namely a writer and reader of the H3K9 methylation mark (Ogawa et al., 2002; Sanchez et al., 2007). None of these two proteins were identified in RNF2 or RYBP interactome, neither in mESCs nor in PrE, providing no support to a potential interplay between PRC1 and H3K9me3 via these proteins. Instead, another paralog of the HP1 family, namely HP1α (CBX5) was identified in the PrE RNF2 interactome (Fig. R8A), raising the possibility that this protein could be involved in RNF2 targeting at constitutive heterochromatin regions marked by H3K9me3. However, the short-reads sequencing method with which our ChIP-seq data was generated, does not allow for proper quantification of the occupancy at repetitive regions (that constitute a large fraction of H3K9me3 marked regions), given that short reads will suffer from multilocus alignment. To address this limitation and extend the potential of the ChIP-seq technique to repetitive regions (including ribosomal DNA, see

below), E. Blanco is now working on developing new pipelines of analysis, and hopefully this will help to shed some light on this aspect.

Polycomb meets ribosomal DNA

Among all cell types in vertebrates, naïve and primed ESCs display one of the highest proliferation rates and fastest cell cycle (doubling time ~13-15 h and ~4-8 h, respectively), as they must expand the epiblast in a very narrow window of time during early embryonic development. To achieve this task, they enter a state of so-called "hypertranscription", expressing high levels of rRNAs and ribosomal proteins, resulting in the presence of extraordinarily large nucleoli (Efroni et al., 2008; Percharde et al., 2017). Upon exit from pluripotency, the rate of cell division is lowered, and cells usually start to silence part of the ribosomal genes and rDNA loci. Notably, PRC1 and PRC2 have recently been reported to take part in this process: while in ESCs the DEAD-box helicase 18 (DDX18) RNA-binding protein prevents PRC2 from entering the nucleolus, upon differentiation, downregulation of DDX18 allows PRC2 to bind and silence rDNA loci (Zhang et al., 2020). Moreover, CBX4 was shown to limit rRNA expression in human mesenchymal stem cells (hMSC) via recruitment of FBL, maintaining nucleolar homeostasis and preventing senescence (Ren et al., 2019). In this work, we observed that while mESCs transition towards the PrE, their proliferation rate drops dramatically. In parallel, RNF2 acquires interaction with FBL, along with other proteins involved in ribosome biogenesis (Fig. R8), which led us to hypothesize that RNF2 might be targeting rDNA loci in PrE. However, IF and ChIP experiments revealed that RNF2 does not localize at nucleoli upon differentiation, suggesting that this interaction might be happening somewhere else in the nucleus, or be an artifact of the IP conditions (eg. upon nuclear lysis). Importantly, FBL does not localize at Polycomb target genes, as assessed by ChIP-qPCR in mESCs and PrE (data not shown). Interestingly, though, we observed that RNF2 might be located at rDNA coding sequence in mESCs, together with FBL. However, given that mouse

displays an average of approximately 200 copies of the rDNA in the genome (Parks et al., 2018), these two factors may target distinct sets of rDNA genes (eg. silent ones and active ones, respectively). This would suggest a scenario where the ribosomal production is eventually regulated by the balance between FBL-activated rDNA copies and PRC1-repressed ones. Although this may seem too farfetched, silent rDNA genes were shown to be marked by repressive chromatin marks like H3K9me3 and H3K27me3 (Murayama et al., 2008; Salifou et al., 2016; Xie et al., 2012), which have the potential to recruit PRC1 via CBX chromodomain recognition. This is especially interesting considering that the epigenetic and transcriptional state of rDNA is inherited through cell division (Conconi et al., 1989), similar to the Polycomb-mediated epigenetic memory of the silent state. Moreover, a bivalent state bearing H3K4me3 and H3K27me3 was also described for rDNA promoters that undergo transcriptional activation upon environmental stimuli (Salifou et al., 2016; Xie et al., 2012), much like many developmental genes that are targeted by Polycomb in mESCs. Therefore, in the future it would be interesting to have a closer look at the role of Polycomb in regulating rDNA expression, especially in a high-proliferative system like pluripotent stem cells.

PRC2 and splicing

Exception made for the mono cistronic *Epop*, virtually all PRC2 genes contain at least one alternative splicing event annotated in vast-db. Whilst potentially very impactful, this aspect has been long overlooked by researchers. As for many other aspects regarding the study of PRC2, the only reports addressing splicing have focused on the catalytic subunit EZH2: an isoform lacking exon 4 and a small part of exon 8 (named Ezh2β) was described to have a similar expression pattern and catalytic activity with respect to the canonical one (Grzenda et al., 2013). Instead, another isoform generated by skipping of the exon 14 generates a less catalytically active protein that acts as a dominant negative in cancer cell lines (Chen et al., 2017; Li et al., 2019; Mu et al., 2018). Exon 14 encodes for part of

the CXC domain that is used by EZH2 to contact DNA at its nucleosome targets, therefore, the lack of this exon may impair PRC2 engagement of its substrate by destabilizing its presence on chromatin. Interestingly, inclusion of this exon is regulated during cell cycle in testis, mouse embryonic fibroblasts (Mu et al., 2018), and in Serum/LIF grown mESCs (N.A. personal communication). We reasoned that since cell cycle is a general mechanism that is present in virtually all cell types, this might provide an explanation for Suz12 exon 4 high alternativity. However, alternative splicing analysis conducted by N.A. on cell cycle specific RNA-seq data from mESCs (kindly provided by Dr. Sergi Aranda) revealed that this exon has similar inclusion levels in the different phases of the cell cycle (data not shown).

Another example of a PRC2 gene undergoing extensive splicing regulation is *Aebp2*: this gene shows at least 4 distinct splicing isoforms in human and mouse, all of which stem from alternative splicing events involving exons 8 to 10, leading to the generation of different C-terminal regions (N. A., personal communication). Interestingly, this region is the one involved in the formation of the KR finger, together with SUZ12 (Fig. I5), suggesting that alternative splicing of *Aebp2* could modulate its interaction with the RBBP4/7 acidic patch, resulting in a dramatic impact in PRC2 conformation. Given the importance of the KR finger in intra- and inter-PRC2 interactions, it would be interesting to study the effect of alternative splicing of *Aebp2* on PRC2 dimerization capacity, in the future.

Overall, alternative splicing has the potential to diversify the structure and function of many chromatin related factors, of which the PRC2 factors, given their high associability and inter-dependency, represent a very peculiar example. In the next years, effort will have to be put in trying to understand the real contribution of splicing-mediated transcription variability in the regulation of epigenetic players, of which, this thesis was just a very modest attempt.

Who regulates the regulator?

In an attempt to induce the deletion of *Suz12* exon 4 in mESCs, we came across a clone in which exon deletion was not successful but instead displayed constitutive exon inclusion, named WT-like clone #3. Analysis of the sequence around the targeting site of the gRNA in the intron 3, revealed that, while most of the WT-like clones had undergone small indels around the position of the cut (ranging from 1 to 5 nt), WT-like #3 presented a 16 bp deletion in one of the two alleles (and a 1 nt insertion in the other, data not shown). To get more insights, we conducted bioinformatic prediction of the possible regulatory elements in this area. We found that the sequence deleted in WT-like #3 is a potential purine-rich binding site for the serine-arginine splicing factor (SRSF) family of proteins. Indeed, overexpression of SRSF3, (but not of the other members of the family), resulted in a drop in Suz12 exon 4 inclusion levels, as evidenced by analysis with a mini gene-splicing model conducted by N.A. (data not shown). This data supports the hypothesis that SRSF3 is responsible for the regulation of exon 4 alternative splicing. Importantly, evolutionary analysis conducted by N.A. suggests that the purine-rich motif recruiting SRSF3 appeared in Eutheria, which would roughly correspond to the appearance of alternativity in this event, as in non-mammalian vertebrates, exon 4 is constitutively included in the transcript.

Notably, another member of the SRSF family, namely SRSF2, was shown to regulate *Ezh2* splicing, its mutation in myelodysplastic syndrome altering sequence specificity and leading to the inclusion of a cryptic exon bearing a premature stop codon (Kim et al., 2015).

Suz12 short isoform has impaired AEBP2 binding

Previous structural studies identified SUZ12 residues that are critical for the interaction with accessory factors (Chen et al., 2018; Kasinath et al., 2018). A recent report from the lab of Prof. Thom Cech has taken advantage of this information to generate human iPSCs expressing SUZ12 mutants that would assemble either one or the other PRC2 subtype (Youmans et al., 2021). Notably, the

presence of PRC2.2 alone determined lowered PRC2 occupancy and H3K27me3 deposition at target genes, leading to a certain degree of derepression. Instead, forced PRC2.1 assembly led to the opposite phenotype, with aberrantly high H3K27me3. Importantly, these results are in line with our observations in Δexon4 cells, where preferential PRC2.1 formation (due to impaired interaction with AEBP2 and JARID2, Fig. R20), results in higher H3K27me3 deposition at PcG target promoters. While both PRC2 subtypes have been reported to bind the same genomic regions and have largely redundant function (Healy et al., 2019; Hojfeldt et al., 2019), they are recruited through distinct mechanisms: indeed, while PRC2.2 presence on chromatin is completely dependent on H2AK119ub deposition by PRC1 (recognized by JARID2), PRC2.1 recruitment relies on PCLs recognition of CpG islands (Healy et al., 2019). Therefore, one way to interpret this data is that PRC2.1 is a more efficient H3K27 methyltransferase owing to its independent mechanism of recruitment that puts it on top of the methylation chain of events, whereas PRC2.2 may not be sufficient to do the whole job by itself (eg. lack of PRC2.1 → less H3K27me3 deposition → less PRC1 recruitment → less H2Aub → less PRC2.2 recruitment).
However, other observations suggest that this interpretation might not fully recapitulate the data: upon deletion of exon 4, SUZ12 interaction with JARID2 is not completely abolished, rather reduced, likely due to the lack of mutual stabilization with AEBP2. Therefore, the local increase in H3K27me3 happens in the presence of a JARID2-contaning PRC2. Notably, AEBP2 full KO in mESCs does not impair JARID2 inclusion in the complex and leads to higher H3K27me3 deposition at PcG targets (Grijzenhout et al., 2016), suggesting that lack of AEBP2 does not impair the formation of a functional JARID2-containing complex. Moreover, recent reports showed that H2AK119ub-mediated PRC2 recruitment mostly depends on JARID2 alone (Kasinath et al., 2021), indicating that a functional 'PRC2.2' is defined by the presence of JARID2. Therefore, changes in H3K27me3 deposition at PRC2 targets upon deletion of Suz12 exon 4, may happen in the presence of (nearly)

functionally competent PRC2 subtypes – although further experiments will be needed to prove that this is the case. As explained in the next paragraph, higher order PRC2 interactions might provide an alternative explanation (see below).

Suz12 exon 4 and dimers

Recent reports addressing PRC2 higher order interaction established that whilst AEBP2 prevents the natural tendency of PRC2 to form dimers, MTF2 and PHF19 can favor it (Chen et al., 2020). In line with this data, we observe that SUZ12 impaired interaction with AEBP2 upon exon 4 deletion is associated with a higher proportion of a high molecular weight complex, whose size is compatible with a PRC2 dimer (Fig. R23). While these observations will have to be confirmed by means of orthogonal approaches (e.g., complementation assays, Co-IP, etc.), we speculate that this can be due to structural differences between the long and short isoform of SUZ12: lack of the exon generates a shorter SUZ12 peptide lacking the WDB1 or hinge domain (23 amino acids). Importantly, given that this peptide connects the Zn and the C2 domains of SUZ12, the isoform lacking exon 4 will necessarily shorten the distance between these two domains, bringing them closer in space and potentially restricting the flexibility of the C2. This is likely to prevent the C2 domain from reaching the back of the WD propeller, disrupting both the KR finger and the clamp around accessory factors. The only possible conformation left is that of a C2 protruding from the complex. Notably, this model has the potential to explain both proteomics and protein fractionation data: (I) protrusion of the C2 domain from the complex is remindful of the structure of the antiparallel dimer (Chen et al., 2020), therefore, structural constrains of the SUZ12 short isoform may prime the protein for dimerization; (II) rotation of the C2 domain would disrupt the clamp that seals together the JARID2 TR and AEBP2 C2B domains, with obvious implications for their binding. Moreover, on the other side of the lobe, the interactions between AEBP2 and SUZ12 at the level of the KR finger would be lost as well, further destabilizing its binding to the complex; (III) while

disruption of the clamp might impair AEBP2 binding, it does not affect PCL binding, as it was shown that PHF19 can form interactions across the two protomers of a PRC2 dimer, sitting between the protruding C2 of one complex and the ZnB of the other complex, gluing the two protomers together. We are currently in contact with collaborators to try to solve the structure of the SUZ12 short isoform-containing PRC2 complex, possibly in its mono- and dimeric form.

Suz12 splicing may define two distinct PRC2 functions

Upon deletion of the genomic sequence encoding *Suz12* exon 4, the expression of *Suz12* is overall reduced (Fig. R19B and data not shown), pointing towards the fact that exon 4 inclusion is necessary to preserve the stability of the transcript. One possible explanation for this is that the lack of the exon may lead to the occurrence of other splicing events (eg. intron retention, or inclusion of a cryptic exon with a premature termination codon) that eventually result in non-sense mediated decay of *Suz12* mRNA. We were not able to provide an answer to this yet. However, the relative abundance of the short isoform was higher in ∆exon4 cells than in WT cells, suggesting that the 3' splice site (ss) of exons 4 and 5 compete for the same 5'ss of exon 3 in WT cells, such that, upon removal of exon 4, the short isoform is more likely to be produced.

Regarding the subcellular distribution of the two isoforms, cell fractionation experiments suggest that the long is the main (if not the only) isoform present on chromatin in WT cells (Fig. R19D), although the short isoform retains the ability to bind chromatin as observed in ∆exon4 cells. However, this might be due to the low resolution of the gel or to a disparity of signal between the long and the short band. We have tried to address this point by quantifying skipping-specific peptides (eg. peptides uniquely generated by an exon3-5 junction, therefore from a SUZ12 short isoform) in SUZ12 IP from nuclear extracts by targeted mass-spectrometry. We were able to identify only one of these peptides (namely, SLSAHLQLTFTGFFHK, in one replicate per each condition (data not shown)

While this result does not provide significant evidence for the presence of the short isoform on chromatin, it is a proof of concept that this might be present in the nucleus of WT cells, besides the long isoform.
Moreover, we tested the hypothesis that exon 4 may contain a nuclear localization signal (NLS). However, despite its high composition in positively charged amino acids, none of the six potential NLSs found in the sequence of SUZ12 mapped within the region encoded by exon 4 (NLS-db, data not shown) (Nair et al., 2003).
Preliminary ChIP-seq data analysis suggest that ∆exon4 cells display surprisingly higher levels of H3K27me3 at Polycomb target promoters (Fig. R22), while showing lower levels of this mark on bulk chromatin (Fig. R19C). When faced with this paradox, we must not forget that H3K27 methylation at Polycomb target promoters only represents a small fraction of the total methylation happening at this residue. Indeed, various estimates have established that around 80-90% of the histone H3 molecules in the chromatin are methylated at lysine 27 in various human cell lines (Leroy et al., 2013; Young et al., 2009; Yuan et al., 2011), while Polycomb target regions only represent a fraction of the coding genome. Therefore, an increase at these sites, is compatible with the overall loss of the mark genome wide. For this reason, we speculate that in WT cells, PRC2 including the short isoform is responsible for very little deposition of H3K27me3 at non-target sites, while catalyzing a considerable part of the H3K27 methylation at Polycomb targets (Fig. R 19D). This hypothesis is further corroborated by the fact that the WT-like clone #3, displaying a depletion of the short isoform, has less H3K27me3 at these sites. Importantly, this would mean that even small variations in the PSI of *Suz12* exon 4 could determine big changes in H3K27me3 deposition.
Importantly, the short isoform is not detectable in the chromatin fraction of WT cells (Fig. R19D), while it becomes clearly enriched in this fraction upon stabilization of the short isoform in ∆exon4 cells. While we cannot rule out that this is simply due to masking of the signal by the upper band (long isoform) that is prevalent in WT cells,

one intriguing possibility is that the two isoforms compete for binding to chromatin – probably at canonical Polycomb recruitment sites (eg. CGIs, H2AK119ub). We are currently conducting experiments to prove this hypothesis, by tracing SUZ12 isoforms binding to chromatin upon exogenous expression.
Regarding their function, one possible speculation is that long SUZ12 isoform-containing PRC2 is dedicated to depositing H3K27me1/2 marks on large fractions of the genome, while short SUZ12-containing one is more focused on the H3K27me3 at target promoters. This hypothesis is currently being addressed by means of Histone PTMs MS and ChIP-seq of H3K27 intermediate methylation states. Another possibility is that the long and short isoforms may instruct the complex to either extend the mark along the chromatin fiber (spreading) or reinforce its presence on the recruiting spot (nucleation). This latter hypothesis is corroborated by the observation that a SUZ12 mutant that is unable to form PRC2 dimers, has lower affinity to CpG islands (Chen et al., 2020). Nonetheless, these two scenarios are not mutually exclusive, meaning that, for instance, spreading may be associated with deposition of the H3K27me1 and -me2 marks whilst focusing may result into local positive feedback leading to more proficient H3K27me3 deposition.
When faced with the question of why cells would express two functionally distinct *Suz12* isoforms, we must not forget that *Suz12* exon 4 acquires alternativity in placental mammals. One fundamental trait arising in this group is dosage compensation mediated by random X-chromosome inactivation in female individuals. This process heavily relies on H3K27 methylation by PRC2 to silence large portions of the X chromosome through *Xist*-coating. Notably, imprinted gene expression also arises in this group. While being historically linked with DNA methylation, recent reports described the existence of a set of H3K27me3-dependent imprinted loci, including *Xist* itself (Inoue et al., 2018; Inoue et al., 2017a; Inoue et al., 2017b; Santini et al., 2021). However, further experiments will be needed to establish whether *Suz12* isoform specialization has anything to do with these processes.

Conclusions

1) We generated a system that enables molecular studies of the PRC1 complex both in mouse embryonic stem cells and upon differentiation.

2) We successfully adapted a protocol to obtain primitive endoderm cells *in vitro* from mouse embryonic stem cells.

3) In primitive endoderm, PRC1 is mainly associated with a small set of Polycomb proteins, including CBX8 and BMI1.

4) CBX8 is not necessary for maintenance of the transcriptome in mouse embryonic stem cells.

5) CBX8 is not necessary for generation of the primitive endoderm *in vitro*.

6) Gene silencing in primitive endoderm is largely independent of PRC1.

7) Exon 4 of the mammalian *Suz12* gene is alternatively spliced in most of the tissues and cell types, generating two isoforms.

8) Inclusion of the exon is necessary for the formation of a PRC2.2 holocomplex and for genome wide H3K27 methylation.

9) H3K27 methylation at Polycomb targets strongly depends on the presence of the short SUZ12 isoform.

10) The short isoform of SUZ12 assembles a complex with higher molecular weight, compatible with that of a dimeric PRC2.

Materials and Methods

Cell culture conditions

Mouse embryonic stem cells (mESCs) from either E14Tg2a (Sox1:GFP (Ying et al., 2003), chapter I) or 129 B13 (WT, chapter II) background were cultured in Serum/LIF conditions: Glasgow minimum essential medium (GMEM, Sigma, G5154) supplemented with 20% inactivated fetal bovine serum (Cytiva HyClone SV30160.03), Glutamax (Gibco, 35050-038), Pen/Strep (Gibco, 15140-122), non-essential amino acids (NEAA, Gibco, 11140-050), β-Mercaptoethanol (Gibco, 31350-010), and Leukemia inhibitory factor (LIF, produced in-house). Cells were grown on cell culture treated plates coated with 0.1% gelatin (Millipore, ES-006-B) at a maximum confluency of 2 x 10^5 cells cm^{-2} and passaged every 2-3 days for a maximum of 2 weeks. Cells were frozen in FBS 10% DMSO using Mr. Frosty (Nalgene, C1562), and thawed in a 37 ºC water bath.

mESCs differentiation

Neural precursor cells (NPC) were obtained from Sox1:GFP mESCs as described in (Ying et al., 2003). Briefly, 1.8 x 10^3 cells cm^{-2} were plated on gelatin-coated plate in Serum/LIF conditions. After 24 h the medium was changed to N2B27 differentiation medium (1:1 mixture of Neurobasal (Gibco, 21203-049) supplemented with N2 (17502-048) and DMEM-F12 (11320-074) supplemented with B27 (17504-044), to which Glutamax, Pen/Strep, NEAA, and 0.33% BSA fraction V (15260-037) were added. Differentiation medium was changed every other day. Cell differentiation was monitored via cytometre using the Sox1:GFP internal reporter, that marks early neuroectoderm committed cells (Wood and Episkopou, 1999; Ying et al., 2003) and harvested after 6 days of differentiation. Primitive Endoderm (PrE) was derived from mESCs adapting the protocol described in (Anderson et al., 2017). mESCs were plated on gelatin-coated plates in Serum/LIF conditions at a confluency of 3-5 x 10^4 cells cm^{-2}. After 24 h the medium was changed to PrE differentiation medium, composed of RPMI-1640 (Lonza, BE12-167F) supplemented with B27 minus insulin (Gibco, A18956-01), Glutamax, and Pen/Strep, to which 4-8 ng/ml Activin A (PeproTech 120-14E), and 3 μM CHIR-99021 (Selleckchem, S1263) were added. The differentiation medium was changed every other day.

After 7 days, LIF was added to the medium to obtain a PrE maintenance medium. Cells were gathered after 9 days for further applications or passaged to a confluency of 5-10 x 10^4 cells cm^{-2} in human fibronectin (HFN, Millipore, FC010)-coated plates in PrE maintenance medium. For HFN coating, a 16 µg/ml solution of HFN in PBS was applied to the well and incubated at 37 ºC for 30 min.
Induction of parietal extraembryonic endoderm (PExE) was performed according to (Anderson et al., 2017). Briefly, PrE cells were plated in PrE maintenance medium to a confluency of approximately 5-10 x 10^4 cells cm^{-2} in HFN-coated plates. After 24 h, medium was changed to N2B27. Cells were harvested after 5 days for further use or passaged similarly to PrE cells every 2-3 days approximately.

Cytometry
Cells were detached with Trypsin and washed twice with PBS. For cell surface marker staining, cells were resuspended in a 1:300 PBS dilution of anti-PDGFRA-APC conjugated antibody (eBioscience 17-1401-81), supplemented with 0.1% BSA, and incubated in the dark on ice for 20 min. After the incubation, cells were washed with PBS and resuspended in PBS supplemented with DAPI. Data on viability and positive staining was acquired using a BD LSR II cytometre (BD Biosciences) using the APC-A detector. Cytometry data was analyzed using the FlowJo software (BD Biosciences).

Gene expression analysis
RNA extraction was performed using the RNeasy kit (Qiagen, 74134) according to manufacturer's instructions. For RT-qPCR, up to 1 µg of RNA was retrotranscribed using the qScript cDNA Synthesis Kit (Quanta bio, 95047) and diluted 1:2 to 1:6 prior to use for amplification. qPCR was performed using the SYBR Green I Master mix (Roche, 04887352001), in a Lightcycler 480 qPCR machine (Roche, 28005). Oligos for qPCR amplification are listed in TABLE M2.
For RNA-seq application, RNA samples were processed by the CRG Genomics Unit as follows: samples were quantified using the Nanodrop spectrophotometer (Thermo Fisher Scientific) and libraries were were prepared using the TruSeq stranded mRNA Library Prep (Illumina, 20020595) according to the manufacturer's protocol. Briefly, 1000 ng of total RNA was used for poly(A)-mRNA selection using poly-T oligo attached magnetic beads using two rounds of

purification. During the second elution of the poly-A RNA, the RNA was fragmented under elevated temperature and primed with random hexamers for cDNA synthesis. Then, the cleaved RNA fragments were copied into first strand cDNA using the SuperScript II reverse transcriptase (Invitrogen, 18064-014) and random primers. After that, second strand cDNA was synthesized, removing the RNA template and synthesizing a replacement strand, incorporating dUTP in place of dTTP to generate ds cDNA using DNA Polymerase I and RNase H. These cDNA fragments, then had the addition of a single 'A' base to the 3' ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation. A corresponding single T nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragments. Subsequent ligation of the multiple indexing adapter to the ends of the ds cDNA was done. Finally, PCR selectively enriched those DNA fragments that had adapter molecules on both ends. The PCR was performed with a PCR Primer Cocktail that anneals to the ends of the adapters. Final libraries were analyzed using Bioanalyzer DNA 1000 (Agilent, 5067-1504) to estimate the quantity and validate the size distribution, and were then quantified by qPCR using the KAPA Library Quantification Kit KK4835 (Roche, 07960204001) prior to the amplification with Illumina's cBot.

Libraries were sequenced on a single end for 50+8bp on Illumina's HiSeq2500. A minimum of 40 x 10^6 reads per sample was generated.

Raw sequencing data was analyzed by Dr. Enrique Blanco (E. Blanco hereafter) in our laboratory as follows: The RNA-seq samples were mapped against the mm10 mouse genome assembly using TopHat (Trapnell et al., 2009) with the option –g 1 to discard those reads that could not be uniquely mapped in just one region. DESeq2 (Love et al., 2014) was run to quantify the expression of every annotated transcript using the RefSeq catalogue of exons (O'Leary et al., 2016) and to identify each set of differentially expressed genes between the two conditions.

MA-, box-, dot- and barplots were generated using the ggplot2 R package. Signature genes for boxplot generation was manually curated according to various sources: pluripotency: *Crebbp*, *Dppa2*, *Dppa4*, *Eras*, *Esrrb*, *Fgf4*, *Fgfr1*, *Id1*, *Id2*, *Id3*, *Klf2*, *Klf4*, *Klf5*, *Lin28a*, *Lin28b*, *Myc*, *Mycn*, *Nanog*, *Nog*, *Nr0b1*, *Pou5f1*, *Sall4*,

Sox2, *Sox4*, *Stat3*, *Tbx3*, *Tcf3*, *Utf1*, *Wnt7a*, *Zfp42*, *Zhx2*; ectoderm: *Col5a1*, *Crabp1*, *Evpl, Fgf5*, *Foxj3*, *Gbx2*, *Krt18*, *Lhx5*, *Meis1*, *Meis2*, *Nes*, *Neurod1*, *Pax6*, *Runx3*, *Sox1*, *Sox18*, *Tcf15*, *Tgm3*, *Zic1*; mesoderm: *Bmp6*, *Eomes*, *Gas1*, *Lef1*, *Lefty2*, *Mesp1*, *Mesp2*, *Msx2*, *Myocd*, *Sox6*, *T*, *Tal1*, *Tbx6*, *Wnt3a*, *Wnt5b*, *Wnt8a*; endoderm: *Amn*, *Cfc1*, *Foxa2*, *Foxa3*, *Foxc1*, *Foxg1*, *Foxh1*, *Foxq1*, *Gata4*, *Gata6*, *Gsc*, *Hhex*, *Hnf1b*, *Nodal*, *Pdgfra*, *Sox17, Sox7*, *Thbd*, *Ttr*; germline: *Blimp1*, *Ctcfl*, *Cxcr4*, *Ddx4*, *Iftm3*, *Kit*, *Nanos3*, *Prdm1*, *Prdm14*, *Sycp3*, *Stella*.

Enrichment analysis methods

Over-representation analysis (ORA) and gene set enrichment analysis (GSEA) were performed using the Web-based Gene Set Analysis Toolkit (www.webgestalt.org) (Liao et al., 2019). For GSEA, genes were ranked according to the 'stat' provided by DeSeq2 analysis. Unless specified, the whole genome was used as a reference gene set. Results were graphed in a Cleveland dot plot using the ggplot2 R package. The 'Overlap' measure represents the number of genes in common between the queried gene set and the one represented, while the 'Enrichment ratio' is the ratio between the observed and the expected overlap. This measure was used to rank entries in the plot. Only entries with an FDR < 0.01 were reported.

CRISPR Cas9 general strategies for genetic manipulation

Cbx8 full knock out (KO) cells were generated via induction of two DNA double strand breaks (DSBs) to induce a deletion encompassing Cbx8 exon 2 to the stop codon (approx. 2kb). This was achieved by double transfection of two sgRNA-Cas9-PuroR expressing vectors (PX459, Addgene #62988). Cells were selected for 5 days with 2µg/ml puromycin, then let to grow for 5 more days without puromycin. Colonies were manually picked and grown into multi-96 well plate and PCR screened for the genomic deletion.

Suz12 exon 4 specific KO cells were generated by N.A. using a similar strategy.

Suz12 full KO cells were generated with the same approach, although deletion involved only Suz12 exons 2 and 3, the remaining transcript, formed by the splicing of exon 1 with either exon 4 or 5, resulting in an early frameshift either case. In this case, gRNAs and plasmids were generated by N.A. while I performed transfection, selection of clones and further validation.

N-term endogenous tagging was carried out via induction of DNA DSB around the TSS of target genes, providing in parallel a template for homology directed repair (HDR), containing the in-frame tag sequence flanked by homology arms. Regarding the tag, to obtain high-affinity reliable immunoprecipitations, we decided to tag endogenous genes with a triple HA (YPYDVPDYAGYTYDVPDYAGSYPYDVPDYA) and triple FLAG (DYKDHDGDYKDHDIDYKDDDD) tags. This latter was flanked by two LoxP sites so that it could be easily removed upon Cre treatment. Given that these two sites are in frame with the rest of the protein, the canonical 34bp sequence was complemented to 36 bp to maintain the correct frame (ATA ACT TCG TAT AAT GTA TGC TAT ACG AAG TTA Tcc, which translates to ITSYNVCYTKLS).
DSB was induced via introduction of either Cas9/RNP (this is the case for *Rnf2*, *Cbx2*, *Cbx4*, *Cbx6*, and *Cbx8* tagging) or combined expression of sgRNA-Cas9-GFP using PX458 plasmid (Addgene #48138) (*Rybp*, see below). Clones' selection was carried out as follows: 48h post transfection, cells were single-cell sorted into a multi-96 well plate previously coated with gelatin and inactivated mouse embryonic fibroblasts (iMEFs, seeding density of 1.3 x 10^5 cells cm^{-2}). For PX458 transfected cells, only GFP positive cells were sorted. 10-14 days after sorting, cells were expanded and PCR-screened for the successful insertion of the tag.
gRNA selection was carried out using T7 assay starting from the top 3 candidate gRNAs (see below). gRNAs and HDR plasmids for endogenous tagging experiments were designed and cloned by the CRG Biomolecular Screening & Protein Technology Unit.

Transfection for CRISPR-Cas9 editing
24h prior to transfection cells were seeded to a 1.0-2.0 x 10^4 cells cm^{-2} density. Transfection was carried using Lipofectamine 2000 (Invitrogen, 11668-019) according to manufacturer's instructions. Briefly, for 1 well of a multi-6 well plate a "Lipo mix" containing 125 µl Optimem and 10 µl Lipofectamine 2000 was prepared. In parallel, a "Cas9 mix" was prepared as follows: in the case of Cas9/RNP complex, 5 µg of Cas9 protein (Sigma, C120010), 75 pmol sgRNA, and 75 pmol of SpCas9 tracrRNA (Sigma) were added, while in the case of plasmid expression, 500ng of plasmid (PX458 or PX459) were used for transfection. For knock-in, 2-3 µg HDR plasmid were added to the Cas9 mix. Once ready, Lipo mix and Cas9 mix were

joined and let to incubate for 30 min, then added directly to the cell medium.

T7 assay for gRNA selection

Three guides per each target region were designed using the CRISPR tool in the Benchling web platform (www.benchling.com). 24h prior to transfection, cells were seeded at a density of 1.0–1.5 x 10^4 cells cm^{-2}. Transfection was carried out as described previously (without HDR template plasmid). 48h after transfection, gDNA was extracted and the targeted locus was amplified by PCR. Oligos were designed so that the cut site does not sit exactly in the middle of the amplicon, but rather skewed towards one of the two oligos. Once amplified, the PCR product was purified from gel and denatured at 95ºC for 5min in 1X NEB buffer 2 (B7002S). This was then reannealed by ramping down the temperature to 25ºC with a speed of 0.1 ºC sec^{-1}. T7 digestion was carried out by addition of 1µl T7 endonuclease I (M0302) and incubation at 37ºC for 15min. The result of the digestion was run on agarose gel for product visualization (Fig. M1B–C). The intensity of the full-size band (unedited alleles) was compared with that of the digested products (edited alleles). Those gRNAs yielding a higher proportion of edited alleles were selected for further application.
Selection of gRNAs for *Rnf2*, *Rybp*, *Cbx2*, *Cbx4*, *Cbx6*, and *Cbx8* tagging was carried out by Dr. Valerio Di Carlo. Selection of gRNAs for targeting of *Suz12* exons 2-3 and exon 4 was carried out by Dr. Niccolò Arecco. I carried out selection of gRNAs for *Cbx8* KO (Fig. M1A–C)

Genomic DNA extraction

For large scale genotyping, gDNA exctaction was carried out directly on multi 96 well cell culture plates. After removing the medium, cells were washed twice with PBS and lysated by adding 50 µl/well of Bradley buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 10 mM NaCl, 0.5% SDS) supplemented with 0.1 mg/ml Proteinase K (Thermo, EO0491) and incubated at 65 ºC in a humified chamber for 2h to o.n. To precipitate DNA, 100 µl/well of a 75 mM NaCl ice-cold ethanol solution was added and let to incubate at r.t. for 30 min. The plate was then spinned at 3000 rpm in a 96-well plate holder centrifuge for 20 min. the supernatant was decanted, and the pellet was rinsed twice with 70% ethanol, each time spinning at 3000 rpm

for 10 min. Finally, the pellet was let to dry for ~20 min and then resuspended in 30–50 µl/well of pre-warmed TE buffer pH 8.0. For genotyping PCR, 2-5 µl of extract were used.
For small scale use, genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, 69504) according to manufacturer's instructions.

Protein extraction
To obtain a whole-cell protein extract (WCE), cells were resuspended in WCE lysis buffer (25 mM TrisHCl pH7.4, 1 mM EDTA, 1 mM EGTA, 1% SDS) supplemented with protease/phosphatase inhibitors (1 mM PMSF, 1 mM Na_3VO_4, and Protease inhibitor cocktail, (Roche, 45148300). To resuspend the chromatin insoluble fraction, the lysate was boiled for 10min at 98ºC and finally sonicated for 4–5 cycles – or until no pellet is visible –in a Bioruptor (Diagenode, UCD-200) set to 30s ON, 30s OFF, low intensity. Protein concentration of the extract was quantified using the Pierce BCA Protein Assay Kit (Thermo Scientific, 23227) with incubation of 30 min at 37 ºC and a Tecan Infinite M-200 plate reader to measure the absorbance.

Immunoprecipitation
To isolate the nuclei, cells were resuspended in hypotonic buffer (10 mM TrisHCl pH 7.4, 10 mM KCl, 15 mM $MgCl_2$) and incubated for 10min on a rotating wheel at 4ºC. Nuclei were then pelleted via centrifugation at 700 g for 5 min at 4 ºC. Nuclei pellet was then resuspended in nuclear lysis buffer (150mM NaCl, 50 mM Hepes pH 7.5, 2.5 mM $MgCl_2$, 0.5% NP-40) supplemented with 100 U/ml Benzonase (Novagen, #71205) and protease and phosphatase inhibitors, and incubated for 1 h at 4 ºC on a rotating wheel. The extract was then clarified by centrifugation at max speed for 10min in a tabletop centrifuge at 4ºC. Protein concentration was quantified with BCA assay. Per each IP, 0.8 to 1mg of protein were diluted at least 1:3 in IP buffer (150 mM NaCl, 50 mM Hepes pH 7.5, 0.5% NP-40, 5 mM EDTA) supplemented with protease/phosphatase inhibitors. To this, 4 ug of antibody and 20 µl of Protein A Dynabeads (Invitrogen, 10002D) were added and incubated overnight at 4ºC on a rotating wheel. The day after, beads were washed three times with IP buffer. To elute the proteins, beads were resuspended in 2X LDS sample

buffer (Invitrogen, NP0007) supplemented with 100 mM dithiothreitol and incubated at 70ºC for 15min.

Western blot

LDS sample buffer was added to the protein samples. Before loading on the gel, samples were denatured by boiling at 98 ºC for 5 min, then cooled down quickly on ice. NuPAGE 4-15% Bis-Tris gradient gels of 10/15 wells (Invitrogen, NP0335/6) were used to separate the proteins in a XCell SureLock Mini-Cell (Invitrogen, EI001), using NuPAGE MOPS SDS Running Buffer (Invitrogen NP0001). Membrane transfer was carried out in NuPAGE Transfer Buffer (Invitrogen, NP0006) supplemented with 15% methanol in a mini-trans cuvette (Bio-rad) at 4 ºC. After the transfer, the membrane was blocked using 5% milk in TBS-T for 1 h at r.t., washed three times with TBS-T, and incubated overnight with the primary antibody (see table M1) diluted in 2.5% milk in TBS-T. The day after, the membrane was washed three times with TBS-T and incubated with the secondary antibody solution in 2.5% milk in TBS-T (see table M1). For HRP-conjugated secondary antibodies, the signal was revealed using ECL Prime Western Blotting Detection System (Cytiva, RPN2232) in an iBright 1500 machine (Invitrogen) using the chemiluminescence mode, while for IRDye-conjugated secondary antibodies, signal was detected using an Odissey machine (LI-COR).

Proteomic analysis

For mass-spectrometry analysis, IP was carried out starting from 2.5 to 5 mg of protein extract and using 20 µg of antibody and 100 µl of Dynabeads. After immunoprecipitation, beads were washed three times with ABC (200 mM ammonium bicarbonate) and resuspended in 60 µl of 6 M Urea in ABC. Samples were then reduced with dithiothreitol (30 nmol, 37 ºC, 60 min), alkylated in the dark with iodoacetamide (60 nmol, 25 ºC, 30 min) and diluted to 1 M urea with ABC for trypsin digestion (1 µg, 37 ºC, 8h, Promega, V5113).
After digestion, peptide mix was acidified with formic acid and desalted with a MicroSpin C18 column (The Nest Group, Inc) prior to mass-spectrometry analysis. Raw data acquisition and analysis was performed by the CRG Proteomics Unit using an LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific) coupled to an EASY-nLC 1000 (Thermo Fisher Scientific).

Acquired spectra were analyzed using the Proteome Discoverer software suite (v1.4, Thermo Fisher Scientific) and the Mascot search engine (v2.6, Matrix Science, (Perkins et al., 1999). The data were searches against a Swiss-Prot mouse database (as in June 2020, 17056 entries) plus E9PW15 (SUZ12 short isoform) and a list of common contaminants and all the corresponding decoy entries (Beer et al., 2017). For peptide identification a precursor ion mass tolerance of 7 ppm was used for MS1 level, trypsin was chosen as enzyme and up to three missed cleavages were allowed. The fragment ion mass tolerance was set to 0.5 Da for MS2 spectra. Oxidation of methionine and N-terminal protein acetylation were used as variable modifications whereas carbamidomethylation on cysteines was set as a fixed modification. FDR in peptide identification was set to a maximum of 0.05.

Protein-protein interactions were scored using the SAINTexpress algorithm (Teo et al., 2014). The ratio with respect to the bait for each protein was calculated as following: the number of PSM for each prey in each replicate (+1) was divided by the number of PSM of that prey in the corresponding mock IP (+1) and normalized by the length of the prey (in amino acids). This number was then divided by the one calculated for the bait (such that the bait itself would have a value of 1) and averaged for the three replicates, to obtain a mean ratio. This measure approximates the stoichiometry of each prey with respect to the bait, but cannot be considered as such, for it does not consider the differences in 'detectability' that preys can have in a mass-spectrometry experiment.

Volcano plots and barplots for protein interaction data visualization were generated using the ggplot2 R package. Visual rendering of protein interactions network was obtained using the string-db.com web server (version 11.0) (Szklarczyk et al., 2019). Briefly, the list of interactors (FC ≥ 5 and BFDR ≤ 0.2) was fed to the string-db algorithm. Only reported physical interactions with a degree of confidence greater than 0.5 were used to trace edges between the nodes, and nodes with no connecting edges towards the rest of the network were hidden. Graphical rendering of the network was performed using Cytoscape software (version 3.9.0) (Shannon et al., 2003), integrating the data from string-db with proteomics data (FC and BFDR).

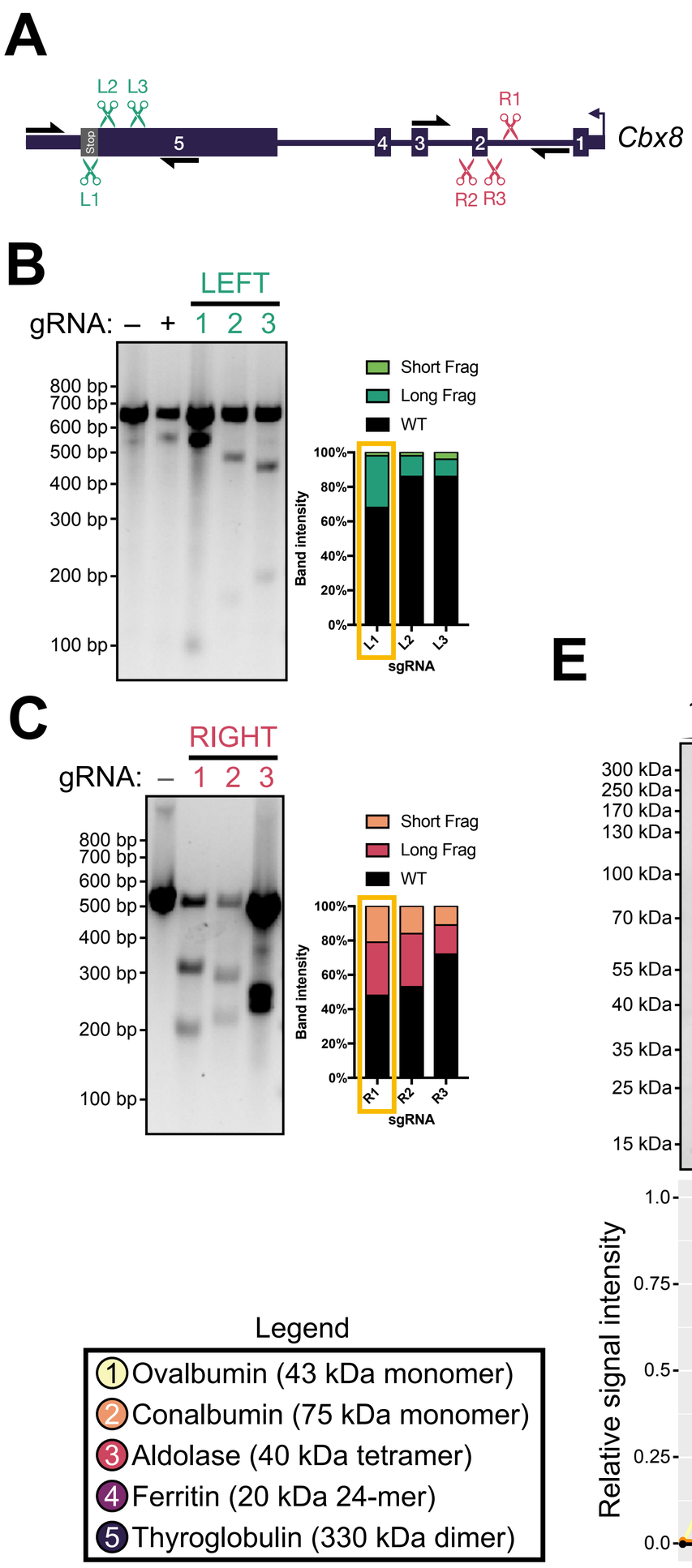


Figure M1 A–C T7 assay for Cbx8 CRISPR KO gRNA selection. **A** Scheme depicting the location of gRNAs and oligos used for T7 assay. **B–C** Agarose gel separation and quantification of the intensity of the bands corresponding to wt and digested fragments in cells transfected with L1–3 (B) or R1–3 (C). '–' = untransfected cells; '+' = cells transfected with a gRNA of known function (the one used for Cbx8 tagging). The yellow square indicates the best performing gRNA that was selected for the CRISPR KO experiment; D Gel separation of chromatin at different cycles of sonication; E coomassie staining of a glycerol gradient fractionation of the protein weight ladder (upper panel) and corresponding quantification of the intensity of the bands (lower panel).

Chromatin immunoprecipitation

Cells were grown to the desired confluency, washed twice with PBS and then crosslinked by incubation with 1% formaldehyde (Sigma, F8775) in PBS for 10 min at r.t. To stop the crosslinking, Glycine was added to a final concentration of 125 mM and incubated for an additional 5 min. To remove the formaldehyde, the solution was discarded, and cells were washed twice with PBS and then collected using a cell lifter (Corning, 3008) in PBS supplemented with protease inhibitors in a 15 ml conical tube. Cells were pelleted by centrifugation at 300 g for 5 min at 4 ºC. After discarding the supernatant, cell pellets were either used directly for chromatin immunoprecipitation (ChIP), or snap-frozen in dry ice and stored at -80 ºC until use.
ChIP of FLAG and EED in mESCs and PrE was performed as follows: lysis was induced by resuspending the cells in lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 0.3% SDS) supplemented with protease and phosphatase inhibitors. Chromatin was sonicated using a Bioruptor Pico (Diagenode) in 15 ml sonication tubes (Diagenode, C01020031) using 350 mg/ml of sonication beads (Diagenode, C03070001) for 40 cycles of 30 s ON, 30 s OFF (see optimization of this step in Fig. M1D). To avoid interference with the anti-FLAG antibody, SDS was removed by incubation of the chromatin extract on ice for 1 h followed by max-speed centrifugation at 4 ºC for 20 min. The precipitated SDS pellet was discarded.
Chromatin size and concentration was assessed on a 2.5% aliquot of the extract: this was diluted 1:8 in PBS and decrosslinked with 0.2 mg/ml Protease K and incubation at 65 ºC for 5h to o.n. DNA was extracted using the PCR Purification Kit (Qiagen, 28106) according to manufacturer's instructions. Concentration was measured by nanodrop, and size was assessed by agarose gel electrophoresis.
Per each ChIP sample, a chromatin amount equivalent to 30-50 µg of DNA was diluted 1:10 or more in Dilution buffer (20 mM Tris-HCl pH8.0, 1.1 mM EDTA pH 8.0, 167 mM NaCl, 1.1% Triton) supplemented with protease and phosphatase inhibitors. To each sample, a 0.1% equivalent of sonicated drosophila chromatin (prepared in-house) was added for spike-in normalization. IP was performed by addition of 4 µg of antibody (see table X), 1 µg of Spike-in Antibody (Active Motif, 104597), and 20 µl of Protein A Dynabeads and incubation overnight at 4ºC on a rotating wheel. The day after, beads were washed three times with Dilution buffer. Chromatin was then

eluted by resuspending the beads in Elution buffer (0.1 M $NaHCO_3$, 1% SDS) and incubating at 65 ºC for 30 min in a Thermomixer shaking at 1000 rpm. Eluted chromatin was then decrosslinked by addition of NaCl to a final concentration of 200 mM and incubation at 65 ºC overnight. Protein digestion was carried out by addition of Proteinase K to a final concentration of 0.1 mg/ml, EDTA to a final conc. of 10 mM, and Tris-HCl pH 6.5 to a final conc. of 40 mM and incubated at 45 ºC for 1 h. DNA was purified using the PCR Purification Kit (Qiagen, 28106) and eluted in ultrapure H_2O. ChIP was performed in biological duplicates.

ChIP of SUZ12 and H3K27me3 in mESCs was performed in the same way, except for the following: (1) Lysis was induced by resuspending the cells in Lysis/IP buffer composed of 1 volume of SDS buffer (100 mM NaCl, 50 mM Tris-HCl pH 8.0, 5 mM EDTA pH 8.0, 0.5% SDS) and 0.5 volumes of Triton buffer (100 mM NaCl, 100 mM Tris-HCl pH 8.6, 5 mM EDTA pH 8.0, 5% Triton) supplemented with protease and phosphatase inhibitors. (2) No SDS removal was carried out after sonication. (3) IP was carried out by diluting the same Lysis/IP buffer. (4) in the case of H3K27me3, 2.5% of equivalent drosophila chromatin was used, and no Spike-in antibody was added, given that the anti-H3K27me3 antibody would recognize the histone mark from both mouse and drosophila chromatin. (5) Post-IP washes were carried out with Triton buffer.

For ChIP-qPCR, 1-2% of the eluted DNA was used per well.

For ChIP-seq, DNA was quantified using Qubit (Thermo Fisher Scientific) and libraries were prepared using the NEBNext Ultra DNA Library Prep (Illumina, E7370) according to the manufacturer's protocol. Briefly, starting amount of 4 ng of input and ChIP enriched DNA were subjected to end repair and addition of ‘A’ bases to 3′ ends, ligation of adapters and USER excision. All purification steps were performed using AgenCourt AMPure XP beads (Beckman Coulter, A63882). Library amplification was performed by PCR using NEBNext Multiplex Oligos for Illumina. Final libraries were analyzed using Agilent Bioanalyzer to estimate the concentration and size and were then quantified by qPCR using the KAPA Library Quantification Kit prior to amplification with Illumina’s cBot. Libraries were sequenced on a single end for 50+8bp on Illumina’s HiSeq2500. A minimum of 30 x 10^6 reads per sample was generated.

Raw data was analyzed by Dr. E. Blanco as follows: ChIP-seq samples including spike-in were mapped against a synthetic genome constituted by the mouse and the drosophila chromosomes (mm10 + dm3) using Bowtie with the option -m 1 to discard reads that did not map uniquely to one region (Langmead et al., 2009).
Peak calling was carried out by running MACS on each replicate individually, with the default parameters but with the shift-size adjusted to 100 bp to perform the peak calling against the corresponding control sample (Zhang et al., 2008). DiffBind (Ross-Innes et al., 2012) was run next over the union of peaks from each pair of replicates of the same experiment to find those peaks that were significantly enriched in both replicates in comparison to the corresponding controls (DiffBind 2.0 arguments: categories = DBA_CONDITION, block = DBA_REPLICATE and method = DBA_DESEQ2_BLOCK). In all cases, DiffBind peaks with P value < 0.05 were selected for further analysis.
The genome distribution of each set of peaks was calculated with SeqCode (Blanco et al., 2021) by counting the number of peaks fitted on each class of region according to RefSeq annotations (O'Leary et al., 2016). Promoter is defined as the region between 2.5 kbp upstream and 2.5 kbp downstream of the transcription start site (TSS). Genic regions correspond to the rest of the gene (the part that is not classified as promoter) and the rest of the genome is considered as intergenic. Peaks that overlapped with more than one genomic feature were proportionally counted the same number of times.
Target genes for each sample were retrieved using SeqCode by matching the ChIP-seq peaks in the region 2.5 Kbp upstream of the TSS until the end of the transcripts as annotated in RefSeq.
The heatmaps displaying the density of ChIP-seq reads around the TSS of each target gene set were generated with SeqCode by counting the number of reads in this region for each individual gene and normalizing this value with the total number of reads mapped on the spike-in genome (Orlando et al., 2014). Genes on each ChIP heatmap were ranked by the average number of reads in the same genomic region.
Screenshots of ChIP-seq tracks were generated using the UCSC genome browser (Kent et al., 2002). Briefly, tracks of replicates were grouped and merged using the ‘add’ option, using a windowing

function with the 'maximum' option and a smoothing window of 5 pixels.

Immunofluorescence

Cells were plated onto a gelatin-coated µ-slide 8-well chambers (Ibidi, 80827) to a confluency of 4-5 x 10^4 cells cm^{-2}. The day after, cells were washed twice with PBS and fixed using 4% Paraformaldehyde (Electron Microscopy Sciences, 15710) in PBS for 10 min at r.t. After fixation, cells were washed twice with PBS and stored with PBS at 4ºC until use. Cells were permeabilized by incubation with PB buffer (10% Goat serum (Merck, G9023), 2.5% Bovine Serum Albumin (Sigma, A7906), and 0.4% Triton) for 30 min at r.t. After washing twice with PBS, staining was carried by incubating the cells with the primary antibody (see table M1) in a 1:2 PB buffer dilution at 37 ºC for 3 h. After the incubation, cells were washed with PBS and stained with the secondary antibody (see table M1) in 1:2 PB buffer dilution supplemented with DAPI (Biotium, 40043) for 1 h at r.t. in the dark. After staining with the secondary antibody, cells were washed twice and kept in PBS until imaging.
Confocal imaging of the staining was performed using a Leica SP5 inverted microscope in the CRG Lightsheet Microscopy Unit. Data was then analyzed and rendered using ImageJ.

Cbxs knockdown/overexpression

Stable knockdown of *Cbx2*, *Cbx4* and *Cbx7* were obtained according to (Morey et al., 2012). Briefly, to produce lentivirus containing shRNA, 2 x 10^6 293T cells were transfected with 10 µg of pLKO-shRNA (Sigma). Supernatant was harvested after 48 h, filtered and used to replace the medium of mESCs overnight. After the medium was washed and Puromycin was added for selection to a concentration of 2 µg/ml. Cells were kept in puromycin and experiments were carried out within two weeks from the transfection.
Cbx7 overexpression was achieved using transient transfection of the pCBA-FlagX3 plasmid (Aloia et al., 2010) containing the cDNA of Cbx7 from (Morey et al., 2012) or not (empty), under the constitutive CAG promoter (Fig. R13A). This plasmid also expresses resistance to G418 (geneticin) under the SV40 constitutive promoter. Briefly, cells were transfected using Lipofectamine 2000 according to manufacturer's protocol with 5 µg of plasmid. After 24 h the

medium was changed, and stable clones were selected using G418 for 6 days.

Cell fractionation

Cell fractionation protocol was adapted from (Mendez and Stillman, 2000). Briefly: cells were detached, washed with PBS, and resuspended in 10 volumes of Buffer A (10 mM Hepes pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M Sucrose, 10% Glycerol, 1 mM dithiothreitol, and 0.1% Triton) supplemented with protease and phosphatase inhibitors, to induce hypotonic lysis of the plasma membrane. 1/10 of the sample was collected and diluted 1:2 in 2X SDS Lysis buffer (50 mM Tris-HCl pH 7.4, 4 mM EDTA pH 8.0, 2% SDS) to obtain the total fraction. After 5 min of incubation on ice, nuclei were pelleted by centrifugation at 1300 g at 4 ºC for 5 min. The supernatant was then clarified by centrifugation at max-speed at 4 ºC for 15 min and diluted 1:2 in 2X SDS Lysis buffer to obtain the cytoplasmic fraction. The pelleted nuclei were washed once with Buffer A and then resuspended in 10 volumes of Buffer A. 1/10 of the sample was collected and diluted 1:2 in 2X SDS Lysis buffer. Nuclei were pelleted again and resuspended in Buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM dithiothreitol) supplemented with protease and phosphatase inhibitors, to induce breakdown of the nuclear membrane by incubating on ice for 30 min. Samples were then quickly vortexed and centrifuged at 1700 g at 4 ºC for 5 min to pellet the chromatin insoluble fraction. Supernatant was clarified by max speed centrifugation at 4 ºC for 15 min and diluted 1:2 in 2X SDS buffer to obtain the nucleoplasmic fraction. The pelleted chromatin was washed with Buffer B and finally resuspended in 10 volumes of 1X SDS buffer to obtain the chromatin fraction.

All fractions were boiled at 98 ºC for 10 min. Total, nuclear, and chromatin fractions were sonicated on a Bioruptor until no insoluble fraction was visible.

Protein concentration of the total fraction was quantified using BCA as described above. For Western blot analysis, equal volumes of each fraction were loaded.

Protein fractionation

Nuclear extracts were obtained as described for the IP protocol and then concentrated by centrifugation in a 5 kDa Amicon Ultra filter unit (Millipore, UFC900508) at 4000 rpm at 4 ºC for 20 min or until

the extract reached the desired volume (see below). The concentrated extract was then diluted at least 1:4 in Dilution buffer (150 mM KCl, 0.2 mM EDTA, 0.1% NP-40, 20 mM Hepes pH 7.9, 0.5 mM dithiothreitol) in a total volume of 2 ml.
The protein weight ladder was obtained using the Gel Filtration HMW Calibration Kit (Cytiva, 28403842) according to manufacturer's instructions to obtain a solution of 500 µl containing 0.8 mg Ovalbumin (43 kDa, monomer), 0.5 mg Conalbumin (75 kDa, monomer), 0.8 mg Aldolase (158 kDa, tetramere), 4.5 ng Ferritin (440 kDa, 24 subunits), and 1.25 mg Thyroglobulin (669 kDa, dimer). This was diluted 1:4 in Dilution buffer for a total of 2ml.
The glycerol gradient was prepared as following: 5 ml of a 5-10% Glycerol solution (150 mM KCl, 0.2 mM EDTA pH 8.0, 0.1% NP-40, 20 mM Hepes pH 7.9, 5-10% Glycerol, see below) were overlayed on the same volume of a 45-50% Glycerol solution (150 mM KCl, 0.2 mM EDTA pH 8.0, 0.1% NP-40, 20 mM Hepes pH 7.9, 45-50% Glycerol) in an Ultra Clear 13.2 ml tube (Beckman Coulter, 344059), and mixed using a Gradient Master 107 IP (Biocomp). For the fractionation of PRC1 components, a 5-45% Glycerol gradient was used, while for PRC2, a 10-50% gradient was used instead (Fig. M1E).
Gradients were incubated at 4 ºC for 2h prior to centrifugation. To separate the proteins, the extract and the protein ladder were overlayed onto separate gradients and centrifuged with an SW41Ti rotor (Beckman Coulter, S/N 20U113807) in an Optima XPN-100 Ultracentrifuge (Beckman Coulter) at 33000 rpm at 4 ºC for 22 h.
After centrifugation, 24 fractions of approximately 500 µl were collected, starting from the top of the tube, and stored at -20 ºC until use. For Western blot analysis, 3-5% of each fraction were used. To obtain a reference of the weight of protein complexes in the gradient fractions, the protein ladder gradient was run on a NuPAGE gel, stained using BlueSafe (Nzytech, MB15201) for 1 h at r.t. and finally washed with water overnight to remove the excess staining. The day after it was imaged with iBright using the 'protein gel' mode (Fig. M1E). Intensity of the bands was assessed using the iBright Analysis Software and visually rendered with the ggplot2 R package (Fig. M1E).

Table M4 *List of antibodies used and their dilution*

Epitope	Reference	WB	IP/ ChIP	IF
CBX7	Abcam ab21873	1:500		
EED	Abcam ab240650	1:1000	1:250	
EPOP	In house made	1:1000		
EZH2	Cell Signalling 3147S	1:1000		
FBL	Abcam ab166630	1:2000	1:250	1:1000
FLAG M2	Sigma F1804	1:500	1:250	1:1000
GAPDH	Santa Cruz sc-32233	1:5000		
H3	Abcam ab1791	1:5000		
H3K9me3	Abcam ab8898	1:2000		
H3K27ac	Active Motif 39685	1:2000		
H3K27me3	Millipore 07-449	1:2000	1:250	
JARID2	Novus NB100-2214	1:500	1:250	
L3MBTL2	Sigma HPA000815	1:200		
MEL-18	Santa Cruz sc-10744	1:1000		
MTF2	Protein Tech 16208-1-AP	1:1000		
PHC1	Active Motif 39723	1:500		
RNF2	In house made	1:500		
RYBP	Millipore AB3637	1:500		
RNA Pol II	Covance MMS-128P	1:2000		
SUZ12	In house made	1:1000	1:250	
VINC	Sigma V9131	1:2000		
Rb HRP	Dako P0448	1:5000		
Ms HRP	Dako P0260	1:5000		
Rb IRDye 680CW	Abcam ab216779	1:5000		
Ms IRDye 800CW	Abcam ab216774	1:5000		

Table M5 *gRNAs for CRISPR Knock-In/Knock-Out*

Target	Sequence
Cbx8 intron 1	TGGGGAGAAATGCATGACG
Cbx8 TTS	GAAAAAAGATGAATTGCCG
Gt(Rosa)26 intron 1	GTCTTTCTAGAAGATGGGCG
Suz12 intron 1	TATCTAGAAAACTAGTAAGC
Suz12 intron 3	GAAAGTTTTGTCTTAGAGTG

Table M6 *oligos for CRISPR Knock-In genotyping*

Gene	Forward	Reverse
Cbx2	GCAGTCTCATCGAACACGTC	CCACCCCTCTCCTCAACGG
Cbx4	CCCTGCCTCAGAGAAGCC	AGGTGGACAACTCAATCGGG
Cbx6	GGACCAGGCGCTACATACAG	GCCAGGTACACTAGGGAGCA
Cbx8	CTCGAAGGCAGGACTCAGAC	CCAGAACTGAGGACGGAGGA
Rnf2	ATGCCTGTTGAGCTCCTTGG	GGAACCCCGTTTTCTGTAAGC

Rybp	AAGACTCTCCAGGGTGCAAG	CAGAGAGCACAGACAAAGGC

Table M7 *oligos for CRISPR Knock-Out genotyping*

Gene	Forward	Reverse
Cbx8	GGGAAGAGGGGTGTGAAGAA	CTCCTTAGACTTGCAGCCCT
Suz12	TGCTTGGCGTAAATGTGTGTGT	AGAAACTAGTGACAAACTA-GAATTACAA

Table M8 *oligos for qRT-PCR*

Gene	Forward	Reverse
Afp	CATGCTGCAAAGCTGACAA	CTTTGCAATGGATGCTCTCTT
Cbx2	CTGGAGTACCTGGTCAAGTGG	CGGGTCCAAAATGTTCTCTTC
Cbx4	AGGCTGGTCCCCCAAATA	GCTCCTGCCTTTCCCTGT
Cbx7	AGCCTCGGGGTATAGGAAGA	CGGTGATGTCAGTCACGGTA
Cbx8	GAGCAAGCTGGATCACACTG	GGCCTTGGAGAGTACCTGAA
Foxa2	CCCTTCTCTATCAACAAC-CTCATGT	GGGTAGTGCATGAC-CTGTTCGT
Gata4	TTCGCTGTTTCTCCCTCAAG	CAATGTTAACGGGTTGTGGA
Gata6	GGTCTCTACAGCAAGATGAATGG	TGGCACAGGACAGTCCAAG
Hnf4a	CAGCAATGGACAGATGTGTGA	TGGTGATGGCTGTGGAGTC
Nanog	TTCTTGCTTACAAGGGTCTGC	AGAGGAAGGGCGAGGAGA
Nes	CCTTCTCTAGTGCTCCACGTCC	CTGCTCCTCCAGCGTCTT-GACC
Oct4	GAGGAGTCCCAGGACATGAA	AGATGGTGGTCTGGCTGAAC
Rnf2	TAAAACGGTGAACTGGCCAC	CCAAGTATCTGGCTGTGAGG
Rplp0	TTCATTGTGGGAGCAGAC	CAGCAGTTTCTCCAGAGC
Snai1	TGTGTGGAGTTCACCTTCCAG	AGAGAGTCCCAGATGAGGGT
Sox2	TCCAAAAACTAATCACAACAATCG	GAAGTGCAATTGGGATGAAAA
Sox7	GCGGAGCTCAGCAAGATG	GGGTCTCTTCTGGGACAGTG
T	CGAGATGATTGTGACCAAGAAC	GGCCTGACACATTTACCTTCA
Tbp	GGACCAGAACAACAGCCTTC	CCGTAAGGCATCATTGGACT
Thbd	ATGCGTGGAGCATGAGTG	CTGGCATCGAGGAAGGTC
Ttr	ATTTCCCCGTTCCATGAAT	GATGGTGTAGTGGCGATGG

Table M9 *oligos for ChIP-qPCR*

Gene	Forward	Reverse
Actb	GCCTAGTAACCGAGACATTGA	AGAAAGCGAGATTGAGGAAG
Gsc	GTTGGTGCCAGGTGAGTAAA	CCCCAGGTAGGGTCGC
Hoxd13	GTGGAACAGCCAGGTGTACT	TAACCACTCCCAAATAGGGGC
Nes	AGGGTTAAGGCCTAGGGACC	GCACTAGAGAAGGGAGTGGC
Sox2	TCCAAAAACTAATCACAACAATCG	GAAGTGCAATTGGGATGAAAA
rDNA prom	GACCAGTTGTTCCTTTGAGG	ACCTATCTCCAGGTCCAATAG
rDNA cod1	GCATCGGTGTGTCGGCATCG	CTGAGCAGTCCCACCACACC
rDNA cod2	GCGACCTCAGATCAGACGTGG	CTGTTCACTCGCCGTTACTGAG
T	GCAGCATGCGTTCCAACAAT	AGCGACCCTAGTAGGAGGTG

Bibliography

Akasaka, T., Takahashi, N., Suzuki, M., Koseki, H., Bodmer, R., and Koga, H. (2002). MBLR, a new RING finger protein resembling mammalian Polycomb gene products, is regulated by cell cycle-dependent phosphorylation. Genes Cells *7*, 835-850.

Akasaka, T., van Lohuizen, M., van der Lugt, N., Mizutani-Koseki, Y., Kanno, M., Taniguchi, M., Vidal, M., Alkema, M., Berns, A., and Koseki, H. (2001). Mice doubly deficient for the polycomb group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. Development *128*, 1587-1597.

Alkema, M.J., Bronk, M., Verhoeven, E., Otte, A., van 't Veer, L.J., Berns, A., and van Lohuizen, M. (1997a). Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. Genes Dev *11*, 226-240.

Alkema, M.J., Jacobs, J., Voncken, J.W., Jenkins, N.A., Copeland, N.G., Satijn, D.P., Otte, A.P., Berns, A., and van Lohuizen, M. (1997b). MPc2, a new murine homolog of the Drosophila polycomb protein is a member of the mouse polycomb transcriptional repressor complex. J Mol Biol *273*, 993-1003.

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M.*, et al.* (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. Science *356*, 1081-1084.

Aloia, L., Parisi, S., Fusco, L., Pastore, L., and Russo, T. (2010). Differentiation of embryonic stem cells 1 (Dies1) is a component of bone morphogenetic protein 4 (BMP4) signaling pathway required for proper differentiation of mouse embryonic stem cells. J Biol Chem *285*, 7776-7783.

Anderson, K.G.V., Hamilton, W.B., Roske, F.V., Azad, A., Knudsen, T.E., Canham, M.A., Forrester, L.M., and Brickman, J.M. (2017). Insulin fine-tunes self-renewal pathways governing naive pluripotency and extra-embryonic endoderm. Nat Cell Biol *19*, 1164-1177.

Antonysamy, S., Condon, B., Druzina, Z., Bonanno, J.B., Gheyi, T., Zhang, F., MacEwan, I., Zhang, A., Ashok, S., Rodgers, L.*, et al.* (2013). Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain. PLoS One *8*, e84147.

Aranda, S., and Di Croce, L. (2019). Inhibitory protein puts a lid on an epigenetic marker. Nature *573*, 38-39.

Arrigoni, R., Alam, S.L., Wamstad, J.A., Bardwell, V.J., Sundquist, W.I., and Schreiber-Agus, N. (2006). The Polycomb-associated protein Rybp is a ubiquitin binding protein. FEBS Lett *580*, 6233-6241.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M.*, et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Ballare, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O.*, et al.* (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nat Struct Mol Biol *19*, 1257-1265.

Bardos, J.I., Saurin, A.J., Tissot, C., Duprez, E., and Freemont, P.S. (2000). HPC3 is a new human polycomb orthologue that interacts and associates with RING1 and Bmi1 and has transcriptional repression properties. J Biol Chem *275*, 28785-28792.

Beer, L.A., Liu, P., Ky, B., Barnhart, K.T., and Speicher, D.W. (2017). Efficient Quantitative Comparisons of Plasma Proteomes Using Label-Free Analysis with MaxQuant. Methods Mol Biol *1619*, 339-352.

Beguelin, W., Popovic, R., Teater, M., Jiang, Y., Bunting, K.L., Rosen, M., Shen, H., Yang, S.N., Wang, L., Ezponda, T.*, et al.* (2013). EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. Cancer Cell *23*, 677-692.

Beguelin, W., Rivas, M.A., Calvo Fernandez, M.T., Teater, M., Purwada, A., Redmond, D., Shen, H., Challman, M.F., Elemento, O., Singh, A.*, et al.* (2017). EZH2 enables germinal centre formation through epigenetic silencing of CDKN1A and an Rb-E2F1 feedback loop. Nat Commun *8*, 877.

Beguelin, W., Teater, M., Gearhart, M.D., Calvo Fernandez, M.T., Goldstein, R.L., Cardenas, M.G., Hatzi, K., Rosen, M., Shen, H., Corcoran, C.M.*, et al.* (2016). EZH2 and BCL6 Cooperate to Assemble CBX8-BCOR Complex to Repress Bivalent Promoters, Mediate Germinal Center Formation and Lymphomagenesis. Cancer Cell *30*, 197-213.

Bel, S., Core, N., Djabali, M., Kieboom, K., Van der Lugt, N., Alkema, M.J., and Van Lohuizen, M. (1998). Genetic interactions and dosage effects of Polycomb group genes in mice. Development *125*, 3543-3551.

Beltran, M., Tavares, M., Justin, N., Khandelwal, G., Ambrose, J., Foster, B.M., Worlock, K.B., Tvardovskiy, A., Kunzelmann, S., Herrero, J.*, et al.* (2019). G-tract RNA removes Polycomb repressive complex 2 from genes. Nat Struct Mol Biol *26*, 899-909.

Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T.*, et al.* (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res *26*, 896-907.

Bentley, M.L., Corn, J.E., Dong, K.C., Phung, Q., Cheung, T.K., and Cochran, A.G. (2011). Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. EMBO J *30*, 3285-3297.

Beringer, M., Pisano, P., Di Carlo, V., Blanco, E., Chammas, P., Vizan, P., Gutierrez, A., Aranda, S., Payer, B., Wierer, M.*, et al.* (2016). EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. Mol Cell *64*, 645-658.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669-681.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006a). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006b). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol *26*, 2560-2569.

Birve, A., and Rasmuson-Lestander, A. (1994). Genetic analysis of the Su(z)12 gene in Drosophila melanogaster. Hereditas *121*, 209.

Birve, A., Sengupta, A.K., Beuchle, D., Larsson, J., Kennison, J.A., Rasmuson-Lestander, A., and Muller, J. (2001). Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. Development *128*, 3371-3379.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y.*, et al.* (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445-1459.

Blackledge, N.P., Fursova, N.A., Kelley, J.R., Huseyin, M.K., Feldmann, A., and Klose, R.J. (2020). PRC1 Catalytic Activity Is Central to Polycomb System Function. Mol Cell *77*, 857-874 e859.

Blanco, E., Gonzalez-Ramirez, M., Alcaine-Colet, A., Aranda, S., and Di Croce, L. (2020). The Bivalent Genome: Characterization, Structure, and Regulation. Trends Genet *36*, 118-131.

Blanco, E., Gonzalez-Ramirez, M., and Di Croce, L. (2021). Productive visualization of high-throughput sequencing data using the SeqCode open portable platform. Sci Rep *11*, 19545.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A.*, et al.* (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557-572 e524.

Boyle, S., Flyamer, I.M., Williamson, I., Sengupta, D., Bickmore, W.A., and Illingworth, R.S. (2020). A central role for canonical PRC1 in shaping the 3D nuclear landscape. Genes Dev *34*, 931-949.

Breen, T.R., and Duncan, I.M. (1986). Maternal expression of genes that regulate the bithorax complex of Drosophila melanogaster. Developmental Biology *118*, 442-456.

Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L.*, et al.* (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat Struct Mol Biol *19*, 1273-1281.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N.*, et al.* (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157-170.

Bruggeman, S.W., Valk-Lingbeek, M.E., van der Stoop, P.P., Jacobs, J.J., Kieboom, K., Tanger, E., Hulsman, D., Leung, C., Arsenijevic, Y., Marino, S.*, et al.* (2005). Ink4a and Arf differentially affect cell proliferation and neural stem cell self-renewal in Bmi1-deficient mice. Genes Dev *19*, 1438-1443.

Brunk, B.P., Martin, E.C., and Adler, P.N. (1991). Drosophila genes Posterior Sex Combs and Suppressor two of zeste encode proteins with homology to the murine bmi-1 oncogene. Nature *353*, 351-353.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D.J. (1998). The distribution of polycomb-group proteins during cell division and development in Drosophila embryos: impact on models for silencing. J Cell Biol *141*, 469-481.

Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., and Sixma, T.K. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. EMBO J *25*, 2465-2474.

Burke, A.C., Nelson, C.E., Morgan, B.A., and Tabin, C. (1995). Hox genes and the evolution of vertebrate axial morphology. Development *121*, 333-346.

Buzzati-Traverso, A.A. (1940). [New mutants report]. Drosophila Information Service *13*, 49.

Caganova, M., Carrisi, C., Varano, G., Mainoldi, F., Zanardi, F., Germain, P.L., George, L., Alberghini, F., Ferrarini, L., Talukder, A.K.*, et al.* (2013). Germinal

center dysregulation by histone methyltransferase EZH2 promotes lymphomagenesis. J Clin Invest *123*, 5009-5022.

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

Chammas, P., Mocavini, I., and Di Croce, L. (2020). Engaging chromatin: PRC2 structure meets function. Br J Cancer *122*, 315-328.

Chan, H.L., Beckedorff, F., Zhang, Y., Garcia-Huidobro, J., Jiang, H., Colaprico, A., Bilbao, D., Figueroa, M.E., LaCava, J., Shiekhattar, R.*, et al.* (2018). Polycomb complexes associate with enhancers and promote oncogenic transcriptional programs in cancer through multiple mechanisms. Nat Commun *9*, 3377.

Chen, K., Xiao, H., Zeng, J., Yu, G., Zhou, H., Huang, C., Yao, W., Xiao, W., Hu, J., Guan, W.*, et al.* (2017). Alternative Splicing of EZH2 pre-mRNA by SF3B3 Contributes to the Tumorigenic Potential of Renal Cancer. Clin Cancer Res *23*, 3428-3441.

Chen, S., Jiao, L., Liu, X., Yang, X., and Liu, X. (2020). A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin. Mol Cell *77*, 1265-1278 e1267.

Chen, S., Jiao, L., Shubbar, M., Yang, X., and Liu, X. (2018). Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. Mol Cell *69*, 840-852 e845.

Choi, J., Bachmann, A.L., Tauscher, K., Benda, C., Fierz, B., and Muller, J. (2017). DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. Nat Struct Mol Biol *24*, 1039-1047.

Chung, C.Y., Sun, Z., Mullokandov, G., Bosch, A., Qadeer, Z.A., Cihan, E., Rapp, Z., Parsons, R., Aguirre-Ghiso, J.A., Farias, E.F.*, et al.* (2016). Cbx8 Acts Non-canonically with Wdr5 to Promote Mammary Tumorigenesis. Cell Rep *16*, 472-486.

Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. Mol Cell *55*, 171-185.

Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2016). GenBank. Nucleic Acids Res *44*, D67-72.

Conconi, A., Widmer, R.M., Koller, T., and Sogo, J. (1989). Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. Cell *57*, 753-761.

Connelly, K.E., Weaver, T.M., Alpsoy, A., Gu, B.X., Musselman, C.A., and Dykhuizen, E.C. (2019). Engagement of DNA and H3K27me3 by the CBX8 chromodomain drives chromatin association. Nucleic Acids Res *47*, 2289-2305.

Conway, E., Jerman, E., Healy, E., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D.J., Mucha, M.*, et al.* (2018). A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. Mol Cell *70*, 408-421 e408.

Core, N., Bel, S., Gaunt, S.J., Aurrand-Lions, M., Pearce, J., Fisher, A., and Djabali, M. (1997). Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. Development *124*, 721-729.

Corley, M., and Kroll, K.L. (2015). The roles and regulation of Polycomb complexes in neural development. Cell Tissue Res *359*, 65-85.

Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45-54.

Creppe, C., Palau, A., Malinverni, R., Valero, V., and Buschbeck, M. (2014). A Cbx8-containing polycomb complex facilitates the transition to gene activation during ES cell differentiation. PLoS Genet *10*, e1004851.

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

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y.*, et al.* (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537*, 548-552.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. Nat Struct Mol Biol *20*, 1250-1257.

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

del Mar Lorente, M., Marcos-Gutierrez, C., Perez, C., Schoorlemmer, J., Ramirez, A., Magin, T., and Vidal, M. (2000). Loss- and gain-of-function mutations show a Polycomb group function for Ring1A in mice. Development *127*, 5093-5100.

Denell, R.E. (1982). Homoeosis in Drosophila: Evidence for a maternal effect of the polycomb locus. Developmental Genetics *3*, 103-113.

Denell, R.E., and Frederick, R.D. (1983). Homoeosis in Drosophila: a description of the Polycomb lethal syndrome. Dev Biol *97*, 34-47.

Deschamps, J., and Wijgerde, M. (1993). Two phases in the establishment of HOX expression domains. Dev Biol *156*, 473-480.

Di Carlo, V., Mocavini, I., and Di Croce, L. (2019). Polycomb complexes in normal and malignant hematopoiesis (Rockefeller University Press), pp. 55-69.

Dickson, A.D. (1979). The disappearance of the decidua capsularis and Reichert's membrane in the mouse. J Anat *129*, 571-577.

Dietrich, N., Bracken, A.P., Trinh, E., Schjerling, C.K., Koseki, H., Rappsilber, J., Helin, K., and Hansen, K.H. (2007). Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus. EMBO J *26*, 1637-1648.

Dobrinic, P., Szczurek, A.T., and Klose, R.J. (2021). PRC1 drives Polycomb-mediated gene repression by controlling transcription initiation and burst frequency. Nat Struct Mol Biol *28*, 811-824.

Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L., and Roeder, R.G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell *121*, 873-885.

Duncan, D.M., Burgess, E.A., and Duncan, I. (1998). Control of distal antennal identity and tarsal development in Drosophila by spineless-aristapedia, a homolog of the mammalian dioxin receptor. Genes Dev *12*, 1290-1303.

Duncan, I.M. (1982). Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of {Drosophila} melanogaster. Genetics *102*, 49-70.

Dura, J.M., Brock, H.W., and Santamaria, P. (1985). Polyhomeotic: a gene of Drosophila melanogaster required for correct expression of segmental identity. Mol Gen Genet *198*, 213-220.

Dura, J.M., Randsholt, N.B., Deatrick, J., Erk, I., Santamaria, P., Freeman, J.D., Freeman, S.J., Weddell, D., and Brock, H.W. (1987). A complex genetic locus, polyhomeotic, is required for segmental specification and epidermal development in {D}. melanogaster. Cell *51*, 829-839.

Eckersley-Maslin, M.A., Alda-Catalinas, C., and Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. Nat Rev Mol Cell Biol *19*, 436-450.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., *et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437-447.

Elderkin, S., Maertens, G.N., Endoh, M., Mallery, D.L., Morrice, N., Koseki, H., Peters, G., Brockdorff, N., and Hiom, K. (2007). A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. Mol Cell *28*, 107-120.

Endoh, M., Endo, T.A., Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W.A., *et al.* (2012). Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS Genet *8*, e1002774.

Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T., Onaga, N., *et al.* (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. Elife *6*.

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

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., *et al.* (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. Elife *1*, e00205.

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

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev *17*, 1870-1881.

Foglizzo, M., Middleton, A.J., Burgess, A.E., Crowther, J.M., Dobson, R.C.J., Murphy, J.M., Day, C.L., and Mace, P.D. (2018). A bidentate Polycomb Repressive-Deubiquitinase complex is required for efficient activity on nucleosomes. Nat Commun *9*, 3932.

Forlani, S., Lawson, K.A., and Deschamps, J. (2003). Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. Development *130*, 3807-3819.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574-1577.

Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H.W., and Paro, R. (1992). Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. EMBO J *11*, 2941-2950.

Fursova, N.A., Blackledge, N.P., Nakayama, M., Ito, S., Koseki, Y., Farcas, A.M., King, H.W., Koseki, H., and Klose, R.J. (2019). Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. Mol Cell *74*, 1020-1036 e1028.

Fursova, N.A., Turberfield, A.H., Blackledge, N.P., Findlater, E.L., Lastuvkova, A., Huseyin, M.K., Dobrinic, P., and Klose, R.J. (2021). BAP1 constrains pervasive H2AK119ub1 to control the transcriptional potential of the genome. Genes Dev *35*, 749-770.

Gambetta, M.C., Oktaba, K., and Muller, J. (2009). Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. Science *325*, 93-96.

Gao, Z., Lee, P., Stafford, J.M., von Schimmelmann, M., Schaefer, A., and Reinberg, D. (2014). An AUTS2-Polycomb complex activates gene expression in the CNS. Nature *516*, 349-354.

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

Garcia, E., Marcos-Gutierrez, C., del Mar Lorente, M., Moreno, J.C., and Vidal, M. (1999). RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. EMBO J *18*, 3404-3418.

Garcia-Cuellar, M.P., Zilles, O., Schreiner, S.A., Birke, M., Winkler, T.H., and Slany, R.K. (2001). The ENL moiety of the childhood leukemia-associated MLL-ENL oncoprotein recruits human Polycomb 3. Oncogene *20*, 411-419.

Gaubatz, S., Wood, J.G., and Livingston, D.M. (1998). Unusual proliferation arrest and transcriptional control properties of a newly discovered E2F family member, E2F-6. Proc Natl Acad Sci U S A *95*, 9190-9195.

Gaytan de Ayala Alonso, A., Gutierrez, L., Fritsch, C., Papp, B., Beuchle, D., and Muller, J. (2007). A genetic screen identifies novel polycomb group genes in Drosophila. Genetics *176*, 2099-2108.

Gearhart, M.D., Corcoran, C.M., Wamstad, J.A., and Bardwell, V.J. (2006). Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. Mol Cell Biol *26*, 6880-6889.

Gehring, W. (1970). A recessive lethal (l(4)29) with a homoeotic effect in Drosophila melanogaster. Drosophila Information Service *45*, 103-103.

Gieni, R.S., Ismail, I.H., Campbell, S., and Hendzel, M.J. (2011). Polycomb group proteins in the DNA damage response: a link between radiation resistance and "stemness". Cell Cycle *10*, 883-894.

Grau, D., Zhang, Y., Lee, C.H., Valencia-Sanchez, M., Zhang, J., Wang, M., Holder, M., Svetlov, V., Tan, D., Nudler, E.*, et al.* (2021). Structures of monomeric and dimeric PRC2:EZH1 reveal flexible modules involved in chromatin compaction. Nat Commun *12*, 714.

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

Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M.R., Szumska, D., McGouran, J.F., Bhattacharya, S., Kessler, B.M., Brockdorff, N., and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. Development *143*, 2716-2723.

Grzenda, A., Lomberk, G., Svingen, P., Mathison, A., Calvo, E., Iovanna, J., Xiong, Y., Faubion, W., and Urrutia, R. (2013). Functional characterization of EZH2beta reveals the increased complexity of EZH2 isoforms involved in the regulation of mammalian gene expression. Epigenetics Chromatin *6*, 3.

Guo, Y., Nady, N., Qi, C., Allali-Hassani, A., Zhu, H., Pan, P., Adams-Cioaba, M.A., Amaya, M.F., Dong, A., Vedadi, M.*, et al.* (2009). Methylation-state-specific recognition of histones by the MBT repeat protein L3MBTL2. Nucleic Acids Res *37*, 2204-2210.

Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A High-Density Map for Navigating the Human Polycomb Complexome. Cell Rep *17*, 583-595.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat Cell Biol *15*, 373-384.

Healy, E., Mucha, M., Glancy, E., Fitzpatrick, D.J., Conway, E., Neikes, H.K., Monger, C., Van Mierlo, G., Baltissen, M.P., Koseki, Y.*, et al.* (2019). PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. Mol Cell *76*, 437-452 e436.

Hemenway, C.S., de Erkenez, A.C., and Gould, G.C. (2001). The polycomb protein MPc3 interacts with AF9, an MLL fusion partner in t(9;11)(p22;q23) acute leukemias. Oncogene *20*, 3798-3805.

Hemenway, C.S., Halligan, B.W., Gould, G.C., and Levy, L.S. (2000). Identification and analysis of a third mouse Polycomb gene, MPc3. Gene *242*, 31-40.

Hemenway, C.S., Halligan, B.W., and Levy, L.S. (1998). The Bmi-1 oncoprotein interacts with dinG and MPh2: the role of RING finger domains. Oncogene *16*, 2541-2547.

Hernandez-Munoz, I., Taghavi, P., Kuijl, C., Neefjes, J., and van Lohuizen, M. (2005). Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. Mol Cell Biol *25*, 11047-11058.

Hobert, O., Sures, I., Ciossek, T., Fuchs, M., and Ullrich, A. (1996). Isolation and developmental expression analysis of Enx-1, a novel mouse Polycomb group gene. Mech Dev *55*, 171-184.

Hogan, B.L.M., Cooper, A.R., and Kurkinen, M. (1980). Incorporation into Reichert's membrane of laminin-like extracellular proteins synthesized by parietal endoderm cells of the mouse embryo. Developmental Biology *80*, 289-300.

Hojfeldt, J.W., Hedehus, L., Laugesen, A., Tatar, T., Wiehle, L., and Helin, K. (2019). Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. Mol Cell *76*, 423-436 e423.

Hojfeldt, J.W., Laugesen, A., Willumsen, B.M., Damhofer, H., Hedehus, L., Tvardovskiy, A., Mohammad, F., Jensen, O.N., and Helin, K. (2018). Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. Nat Struct Mol Biol *25*, 225-232.

Holoch, D., Wassef, M., Lovkvist, C., Zielinski, D., Aflaki, S., Lombard, B., Hery, T., Loew, D., Howard, M., and Margueron, R. (2021). A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals. Nat Genet.

Hosogane, M., Funayama, R., Nishida, Y., Nagashima, T., and Nakayama, K. (2013). Ras-induced changes in H3K27me3 occur after those in transcriptional activity. PLoS Genet *9*, e1003698.

Huang, Y., Zhao, W., Wang, C., Zhu, Y., Liu, M., Tong, H., Xia, Y., Jiang, Q., and Qin, J. (2018). Combinatorial Control of Recruitment of a Variant PRC1.6 Complex in Embryonic Stem Cells. Cell Rep *22*, 3032-3043.

Hurlin, P.J., Steingrimsson, E., Copeland, N.G., Jenkins, N.A., and Eisenman, R.N. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. EMBO J *18*, 7019-7028.

Huseyin, M.K., and Klose, R.J. (2021). Live-cell single particle tracking of PRC1 reveals a highly dynamic system with low target site occupancy. Nat Commun *12*, 887.

Huynh, K.D., and Lee, J.T. (2003). Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. Nature *426*, 857-862.

Illingworth, R.S., Holzenspies, J.J., Roske, F.V., Bickmore, W.A., and Brickman, J.M. (2016). Polycomb enables primitive endoderm lineage priming in embryonic stem cells. Elife *5*.

Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes Dev *29*, 1897-1902.

Ingham, P., and Whittle, R. (1980). Trithorax: A new homoeotic mutation of Drosophila melanogaster causing transformations of abdominal and thoracic imaginal segments. Molecular and General Genetics MGG *179*, 607-614.

Ingham, P.W. (1984). A gene that regulates the bithorax complex differentially in larval and adult cells of Drosophila. Cell *37*, 815-823.

Inoue, A., Chen, Z., Yin, Q., and Zhang, Y. (2018). Maternal Eed knockout causes loss of H3K27me3 imprinting and random X inactivation in the extraembryonic cells. Genes Dev *32*, 1525-1536.

Inoue, A., Jiang, L., Lu, F., Suzuki, T., and Zhang, Y. (2017a). Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature *547*, 419-424.

Inoue, A., Jiang, L., Lu, F., and Zhang, Y. (2017b). Genomic imprinting of Xist by maternal H3K27me3. Genes Dev *31*, 1927-1932.

Ishida, A., Asano, H., Hasegawa, M., Koseki, H., Ono, T., Yoshida, M.C., Taniguchi, M., and Kanno, M. (1993). Cloning and chromosome mapping of the human Mel-18 gene which encodes a DNA-binding protein with a new 'RING-finger' motif. Gene *129*, 249-255.

Ismail, I.H., Gagne, J.P., Caron, M.C., McDonald, D., Xu, Z., Masson, J.Y., Poirier, G.G., and Hendzel, M.J. (2012). CBX4-mediated SUMO modification regulates BMI1 recruitment at sites of DNA damage. Nucleic Acids Res *40*, 5497-5510.

Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. Dev Cell *26*, 565-577.

Isono, K., Fujimura, Y., Shinga, J., Yamaki, M., J, O.W., Takihara, Y., Murahashi, Y., Takada, Y., Mizutani-Koseki, Y., and Koseki, H. (2005). Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of Hox genes. Mol Cell Biol *25*, 6694-6706.

Iyer-Bierhoff, A., Krogh, N., Tessarz, P., Ruppert, T., Nielsen, H., and Grummt, I. (2018). SIRT7-Dependent Deacetylation of Fibrillarin Controls Histone H2A Methylation and rRNA Synthesis during the Cell Cycle. Cell Rep *25*, 2946-2954 e2945.

Jaensch, E.S., Zhu, J., Cochrane, J.C., Marr, S.K., Oei, T.A., Damle, M., McCaslin, E.Z., and Kingston, R.E. (2021). A Polycomb domain found in committed cells impairs differentiation when introduced into PRC1 in pluripotent cells. Mol Cell.

Jain, S.U., Do, T.J., Lund, P.J., Rashoff, A.Q., Diehl, K.L., Cieslik, M., Bajic, A., Juretic, N., Deshmukh, S., Venneti, S.*, et al.* (2019). PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism. Nat Commun *10*, 2146.

James, T.C., and Elgin, S.C. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. Mol Cell Biol *6*, 3862-3872.

Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science *350*, aac4383.

Johannsen, W. (1909). Elemente der exakten Erblichkeitslehre. [Elements of an Exact Theory of Heredity.] (Jena, Germany: Gustav Fischer).

Jones, C.A., Ng, J., Peterson, A.J., Morgan, K., Simon, J., and Jones, R.S. (1998). The Drosophila esc and E(z) proteins are direct partners in polycomb group-mediated repression. Mol Cell Biol *18*, 2825-2834.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet *13*, 484-492.

Jones, R.S., and Gelbart, W.M. (1990). Genetic analysis of the enhancer of zeste locus and its role in gene regulation in Drosophila melanogaster. Genetics *126*, 185-199.

Jones, R.S., and Gelbart, W.M. (1993). The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. Mol Cell Biol *13*, 6357-6366.

Junco, S.E., Wang, R., Gaipa, J.C., Taylor, A.B., Schirf, V., Gearhart, M.D., Bardwell, V.J., Demeler, B., Hart, P.J., and Kim, C.A. (2013). Structure of the

polycomb group protein PCGF1 in complex with BCOR reveals basis for binding selectivity of PCGF homologs. Structure *21*, 665-671.

Jung, J., Buisman, S.C., Weersing, E., Dethmers-Ausema, A., Zwart, E., Schepers, H., Dekker, M.R., Lazare, S.S., Hammerl, F., Skokova, Y.*, et al.* (2019). CBX7 Induces Self-Renewal of Human Normal and Malignant Hematopoietic Stem and Progenitor Cells by Canonical and Non-canonical Interactions. Cell Rep *26*, 1906-1918 e1908.

Jürgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in Drosophila. Nature *316*, 153-155.

Justin, N., Zhang, Y., Tarricone, C., Martin, S.R., Chen, S., Underwood, E., De Marco, V., Haire, L.F., Walker, P.A., Reinberg, D.*, et al.* (2016). Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. Nat Commun *7*, 11316.

Kagey, M.H., Melhuish, T.A., and Wotton, D. (2003). The Polycomb Protein Pc2 Is a SUMO E3. Cell *113*, 127-137.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W., Muller, C.W., Vermeulen, M., and Muller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat Struct Mol Biol *21*, 569-571.

Kalisch, W.E., and Rasmuson, B. (1974). Changes of zeste phenotype induced by autosomal mutations in Drosophila melanogaster. Hereditas *78*, 97-104.

Kasinath, V., Beck, C., Sauer, P., Poepsel, S., Kosmatka, J., Faini, M., Toso, D., Aebersold, R., and Nogales, E. (2021). JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. Science *371*.

Kasinath, V., Faini, M., Poepsel, S., Reif, D., Feng, X.A., Stjepanovic, G., Aebersold, R., and Nogales, E. (2018). Structures of human PRC2 with its cofactors AEBP2 and JARID2. Science *359*, 940-944.

Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G.A., Li, Z., Vedadi, M., Schapira, M.*, et al.* (2011). Recognition and specificity determinants of the human cbx chromodomains. J Biol Chem *286*, 521-529.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.

Kim, C.A., Gingery, M., Pilpa, R.M., and Bowie, J.U. (2002). The SAM domain of polyhomeotic forms a helical polymer. Nat Struct Biol *9*, 453-457.

Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.B., Murphy, M.E.*, et al.* (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell *27*, 617-630.

Kim, J., and Kingston, R.E. (2020). The CBX family of proteins in transcriptional repression and memory. J Biosci *45*.

Kinkley, S., Helmuth, J., Polansky, J.K., Dunkel, I., Gasparoni, G., Frohler, S., Chen, W., Walter, J., Hamann, A., and Chung, H.R. (2016). reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells. Nat Commun *7*, 12514.

Klauke, K., Radulovic, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K.*, et al.* (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. Nat Cell Biol *15*, 353-362.

Klein, R.R., and Houtz, R.L. (1995). Cloning and developmental expression of pea ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase. Plant Mol Biol *27*, 249-261.

Kloet, S.L., Makowski, M.M., Baymaz, H.I., van Voorthuijsen, L., Karemaker, I.D., Santanach, A., Jansen, P., Di Croce, L., and Vermeulen, M. (2016). The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. Nat Struct Mol Biol *23*, 682-690.

Kolovos, P., Nishimura, K., Sankar, A., Sidoli, S., Cloos, P.A., Helin, K., and Christensen, J. (2020). PR-DUB maintains the expression of critical genes through FOXK1/2- and ASXL1/2/3-dependent recruitment to chromatin and H2AK119ub1 deubiquitination. Genome Res *30*, 1119-1130.

Koyama-Nasu, R., David, G., and Tanese, N. (2007). The F-box protein Fbl10 is a novel transcriptional repressor of c-Jun. Nat Cell Biol *9*, 1074-1080.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S.*, et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet *4*, e1000242.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Mol Cell *65*, 432-446 e435.

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

Kyba, M., and Brock, H.W. (1998a). The Drosophila polycomb group protein Psc contacts ph and Pc through specific conserved domains. Mol Cell Biol *18*, 2712-2720.

Kyba, M., and Brock, H.W. (1998b). The SAM domain of polyhomeotic, RAE28, and Scm mediates specific interactions through conserved residues. Developmental Genetics *22*, 74-84.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev *22*, 2799-2810.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature *547*, 236-240.

Lau, M.S., Schwartz, M.G., Kundu, S., Savol, A.J., Wang, P.I., Marr, S.K., Grau, D.J., Schorderet, P., Sadreyev, R.I., Tabin, C.J.*, et al.* (2017). Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. Science *355*, 1081-1084.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet *32*, 42-56.

Lee, C.H., Yu, J.R., Kumar, S., Jin, Y., LeRoy, G., Bhanu, N., Kaneko, S., Garcia, B.A., Hamilton, A.D., and Reinberg, D. (2018). Allosteric Activation Dictates PRC2 Activity Independent of Its Recruitment to Chromatin. Mol Cell *70*, 422-434 e426.

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

Leroy, G., Dimaggio, P.A., Chan, E.Y., Zee, B.M., Blanco, M.A., Bryant, B., Flaniken, I.Z., Liu, S., Kang, Y., Trojer, P.*, et al.* (2013). A quantitative atlas of histone modification signatures from human cancer cells. Epigenetics Chromatin *6*, 20.

Leung, C.Y., and Zernicka-Goetz, M. (2015). Mapping the journey from totipotency to lineage specification in the mouse embryo. Curr Opin Genet Dev *34*, 71-76.

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

Lewis, E.B. (1963). Genes and developmental pathways. Am Zool *3*, 33-56.

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

Lewis, P.H. (1947). [New mutants report]. Drosophila Information Service *21*, 69-69.

Li, B., Zhou, J., Liu, P., Hu, J., Jin, H., Shimono, Y., Takahashi, M., and Xu, G. (2007). Polycomb protein Cbx4 promotes SUMO modification of de novo DNA methyltransferase Dnmt3a. Biochem J *405*, 369-378.

Li, D., Wang, H.L., Huang, X., Gu, X., Xue, W., and Xu, Y. (2019). Identification and Functional Characterization of a New Splicing Variant of EZH2 in the Central Nervous System. Int J Biol Sci *15*, 69-80.

Li, G., Warden, C., Zou, Z., Neman, J., Krueger, J.S., Jain, A., Jandial, R., and Chen, M. (2013). Altered expression of polycomb group genes in glioblastoma multiforme. PLoS One *8*, e80970.

Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L.*, et al.* (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. Nature *549*, 287-291.

Li, J., Xu, Y., Long, X.D., Wang, W., Jiao, H.K., Mei, Z., Yin, Q.Q., Ma, L.N., Zhou, A.W., Wang, L.S.*, et al.* (2014). Cbx4 governs HIF-1alpha to potentiate angiogenesis of hepatocellular carcinoma by its SUMO E3 ligase activity. Cancer Cell *25*, 118-131.

Liang, Y.K., Lin, H.Y., Chen, C.F., and Zeng (2017). Prognostic values of distinct CBX family members in breast cancer. Oncotarget *8*, 92375-92387.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res *47*, W199-W205.

Liefke, R., Karwacki-Neisius, V., and Shi, Y. (2016). EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape. Mol Cell *64*, 659-672.

Liu, B., Liu, Y.F., Du, Y.R., Mardaryev, A.N., Yang, W., Chen, H., Xu, Z.M., Xu, C.Q., Zhang, X.R., Botchkarev, V.A., *et al.* (2013). Cbx4 regulates the proliferation of thymic epithelial cells and thymus function. Development *140*, 780-788.

Liu, S., Aldinger, K.A., Cheng, C.V., Kiyama, T., Dave, M., McNamara, H.K., Zhao, W., Stafford, J.M., Descostes, N., Lee, P., *et al.* (2021). NRF1 association with AUTS2-Polycomb mediates specific gene activation in the brain. Mol Cell *81*, 4663-4676 e4668.

Liu, Y., and Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. Protein Sci *11*, 1285-1299.

Locke, J., Kotarski, M.A., and Tartof, K.D. (1988). Dosage-dependent modifiers of position effect variegation in Drosophila and a mass action model that explains their effect. Genetics *120*, 181-198.

Long, J., Zuo, D., and Park, M. (2005). Pc2-mediated sumoylation of Smad-interacting protein 1 attenuates transcriptional repression of E-cadherin. J Biol Chem *280*, 35477-35489.

Long, Y., Bolanos, B., Gong, L., Liu, W., Goodrich, K.J., Yang, X., Chen, S., Gooding, A.R., Maegley, K.A., Gajiwala, K.S., *et al.* (2017). Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. Elife *6*.

Long, Y., Hwang, T., Gooding, A.R., Goodrich, K.J., Rinn, J.L., and Cech, T.R. (2020). RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. Nat Genet *52*, 931-938.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Luis, N.M., Morey, L., Mejetta, S., Pascual, G., Janich, P., Kuebler, B., Cozutto, L., Roma, G., Nascimento, E., Frye, M., *et al.* (2011). Regulation of human epidermal stem cell proliferation and senescence requires polycomb- dependent and -independent functions of Cbx4. Cell Stem Cell *9*, 233-246.

Ma, R.G., Zhang, Y., Sun, T.T., and Cheng, B. (2014). Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. J Zhejiang Univ Sci B *15*, 412-428.

Maat, H., Atsma, T.J., Hogeling, S.M., Rodriguez Lopez, A., Jaques, J., Olthuis, M., de Vries, M.P., Gravesteijn, C., Brouwers-Vos, A.Z., van der Meer, N., *et al.* (2021). The USP7-TRIM27 axis mediates non-canonical PRC1.1 function and is a druggable target in leukemia. iScience *24*, 102435.

MacPherson, M.J., Beatty, L.G., Zhou, W., Du, M., and Sadowski, P.D. (2009). The CTCF insulator protein is posttranslationally modified by SUMO. Mol Cell Biol *29*, 714-725.

Mantsoki, A., Devailly, G., and Joshi, A. (2015). CpG island erosion, polycomb occupancy and sequence motif enrichment at bivalent promoters in mammalian embryonic stem cells. Sci Rep *5*, 16791.

Mardaryev, A.N., Liu, B., Rapisarda, V., Poterlowicz, K., Malashchuk, I., Rudolf, J., Sharov, A.A., Jahoda, C.A., Fessing, M.Y., Benitah, S.A.*, et al.* (2016). Cbx4 maintains the epithelial lineage identity and cell proliferation in the developing stratified epithelium. J Cell Biol *212*, 77-89.

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

Maroto, M., Bone, R.A., and Dale, J.K. (2012). Somitogenesis. Development *139*, 2453-2456.

Martello, G., and Smith, A. (2014). The nature of embryonic stem cells. Annu Rev Cell Dev Biol *30*, 647-675.

Martin, E.C., and Adler, P.N. (1993). The Polycomb group gene Posterior Sex Combs encodes a chromosomal protein. Development *117*, 641-655.

Mas, G., Blanco, E., Ballare, C., Sanso, M., Spill, Y.G., Hu, D., Aoi, Y., Le Dily, F., Shilatifard, A., Marti-Renom, M.A.*, et al.* (2018). Promoter bivalency favors an open chromatin architecture in embryonic stem cells. Nat Genet *50*, 1452-1462.

McGinty, R.K., Henrici, R.C., and Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. Nature *514*, 591-596.

Mendez, J., and Stillman, B. (2000). Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. Mol Cell Biol *20*, 8602-8612.

Messmer, S., Franke, A., and Paro, R. (1992). Analysis of the functional role of the Polycomb chromo domain in Drosophila melanogaster. Genes Dev *6*, 1241-1254.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

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

Mocavini, I., and Di Croce, L. (2020). RNA closing the Polycomb circle. Nat Genet *52*, 866-867.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J.*, et al.* (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep *20*, 1215-1228.

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

Monaco, P.L., Marcel, V., Diaz, J.J., and Catez, F. (2018). 2'-O-Methylation of Ribosomal RNA: Towards an Epitranscriptomic Control of Translation? Biomolecules *8*.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di Croce, L. (2012). Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell *10*, 47-62.

Morey, L., Santanach, A., Blanco, E., Aloia, L., Nora, E.P., Bruneau, B.G., and Di Croce, L. (2015). Polycomb Regulates Mesoderm Cell Fate-Specification in Embryonic Stem Cells through Activation and Repression Mechanisms. Cell Stem Cell *17*, 300-315.

Mu, W., Starmer, J., Yee, D., and Magnuson, T. (2018). EZH2 variants differentially regulate polycomb repressive complex 2 in histone methylation and cell differentiation. Epigenetics Chromatin *11*, 71.

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

Murayama, A., Ohmori, K., Fujimura, A., Minami, H., Yasuzawa-Tanaka, K., Kuroda, T., Oie, S., Daitoku, H., Okuwaki, M., Nagata, K.*, et al.* (2008). Epigenetic control of rDNA loci in response to intracellular energy status. Cell *133*, 627-639.

Murzina, N.V., Pei, X.Y., Zhang, W., Sparkes, M., Vicente-Garcia, J., Pratap, J.V., McLaughlin, S.H., Ben-Shahar, T.R., Verreault, A., Luisi, B.F.*, et al.* (2008). Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. Structure *16*, 1077-1085.

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

Nair, R., Carter, P., and Rost, B. (2003). NLSdb: database of nuclear localization signals. Nucleic Acids Res *31*, 397-399.

Nakagawa, T., Kajitani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., Koji, T., Matsuyama, T., Ikura, T., Muramatsu, M.*, et al.* (2008). Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes Dev *22*, 37-49.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Ng, J., Hart, C.M., Morgan, K., and Simon, J.A. (2000). A Drosophila ESC-E(Z) protein complex is distinct from other polycomb group complexes and contains covalently modified ESC. Mol Cell Biol *20*, 3069-3078.

Ning, B., Zhao, W., Qian, C., Liu, P., Li, Q., Li, W., and Wang, R.F. (2017). USP26 functions as a negative regulator of cellular reprogramming by stabilising PRC1 complex components. Nat Commun *8*, 349.

Ning, G., Huang, Y.L., Zhen, L.M., Xu, W.X., Jiao, Q., Yang, F.J., Wu, L.N., Zheng, Y.Y., Song, J., Wang, Y.S.*, et al.* (2018). Transcriptional expressions of Chromobox 1/2/3/6/8 as independent indicators for survivals in hepatocellular carcinoma patients. Aging (Albany NY) *10*, 3450-3473.

Nomura, M., Takihara, Y., and Shimada, K. (1994). Isolation and characterization of retinoic acid-inducible cDNA clones in F9 cells: one of the early inducible clones encodes a novel protein sharing several highly homologous regions with a Drosophila polyhomeotic protein. Differentiation *57*, 39-50.

Nunes, M., Blanc, I., Maes, J., Fellous, M., Robert, B., and McElreavey, K. (2001). NSPc1 , a novel mammalian Polycomb gene, is expressed in neural crest-derived structures of the peripheral nervous system. Mechanisms of Development *102*, 219-222.

Nusslein-Volhard, C., Wieschaus, E., and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster : I. Zygotic loci on the second chromosome. Wilehm Roux Arch Dev Biol *193*, 267-282.

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

O'Connell, S., Wang, L., Robert, S., Jones, C.A., Saint, R., and Jones, R.S. (2001). Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein. J Biol Chem *276*, 43065-43073.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D*., et al.* (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res *44*, D733-745.

O'Loghlen, A., Munoz-Cabello, A.M., Gaspar-Maia, A., Wu, H.A., Banito, A., Kunowska, N., Racek, T., Pemberton, H.N., Beolchi, P., Lavial, F*., et al.* (2012). MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. Cell Stem Cell *10*, 33-46.

Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D.M., and Nakatani, Y. (2002). A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. Science *296*, 1132-1136.

Oguro, H., Yuan, J., Ichikawa, H., Ikawa, T., Yamazaki, S., Kawamoto, H., Nakauchi, H., and Iwama, A. (2010). Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1. Cell Stem Cell *6*, 279-286.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*, 644-649.

Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep *9*, 1163-1170.

Parks, M.M., Kurylo, C.M., Dass, R.A., Bojmar, L., Lyden, D., Vincent, C.T., and Blanchard, S.C. (2018). Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. Sci Adv *4*, eaao0665.

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

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol *27*, 3769-3779.

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

Pearce, J.J.H., Singh, P.B., and Gaunt, S.J. (1992). The mouse has a Polycomb-like chromobox gene. Development.

Pemberton, H., Anderton, E., Patel, H., Brookes, S., Chandler, H., Palermo, R., Stock, J., Rodriguez-Niedenfuhr, M., Racek, T., de Breed, L.*, et al.* (2014). Genome-wide co-localization of Polycomb orthologs and their effects on gene expression in human fibroblasts. Genome Biol *15*, R23.

Percharde, M., Bulut-Karslioglu, A., and Ramalho-Santos, M. (2017). Hypertranscription in Development, Stem Cells, and Regeneration. Dev Cell *40*, 9-21.

Perkins, D.N., Pappin, D.J.C., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551-3567.

Peterson, A.J., Kyba, M., Bornemann, D., Morgan, K., Brock, H.W., and Simon, J. (1997). A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. Mol Cell Biol *17*, 6683-6692.

Phillips, M.D., and Shearn, A. (1990). Mutations in polycombeotic, a Drosophila polycomb-group gene, cause a wide range of maternal and zygotic phenotypes. Genetics *125*, 91-101.

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B.*, et al.* (2017). hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol Cell *68*, 955-969 e910.

Piunti, A., Smith, E.R., Morgan, M.A.J., Ugarenko, M., Khaltyan, N., Helmin, K.A., Ryan, C.A., Murray, D.C., Rickels, R.A., Yilmaz, B.D.*, et al.* (2019). CATACOMB: An endogenous inducible gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. Sci Adv *5*, eaax2887.

Plys, A.J., Davis, C.P., Kim, J., Rizki, G., Keenen, M.M., Marr, S.K., and Kingston, R.E. (2019). Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. Genes Dev *33*, 799-813.

Poepsel, S., Kasinath, V., and Nogales, E. (2018). Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. Nat Struct Mol Biol *25*, 154-162.

Posfai, E., Kunzmann, R., Brochard, V., Salvaing, J., Cabuy, E., Roloff, T.C., Liu, Z., Tardat, M., van Lohuizen, M., Vidal, M.*, et al.* (2012). Polycomb function during oogenesis is required for mouse embryonic development. Genes Dev *26*, 920-932.

Qin, J., Whyte, W.A., Anderssen, E., Apostolou, E., Chen, H.H., Akbarian, S., Bronson, R.T., Hochedlinger, K., Ramaswamy, S., Young, R.A.*, et al.* (2012). 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.

Ragazzini, R., Perez-Palacios, R., Baymaz, I.H., Diop, S., Ancelin, K., Zielinski, D., Michaud, A., Givelet, M., Borsos, M., Aflaki, S.*, et al.* (2019). EZHIP constrains Polycomb Repressive Complex 2 activity in germ cells. Nat Commun *10*, 3858.

Rastelli, L., Chan, C.S., and Pirrotta, V. (1993). Related chromosome binding sites for zeste, suppressors of zeste and Polycomb group proteins in Drosophila and their dependence on Enhancer of zeste function. The EMBO Journal *12*, 1513-1522.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D.*, et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Ren, X., Hu, B., Song, M., Ding, Z., Dang, Y., Liu, Z., Zhang, W., Ji, Q., Ren, R., Ding, J.*, et al.* (2019). Maintenance of Nucleolar Homeostasis by CBX4 Alleviates Senescence and Osteoarthritis. Cell Rep *26*, 3643-3656 e3647.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol Cell *55*, 347-360.

Roscic, A., Moller, A., Calzado, M.A., Renner, F., Wimmer, V.C., Gresko, E., Ludi, K.S., and Schmitz, M.L. (2006). Phosphorylation-dependent control of Pc2 SUMO E3 ligase activity by its substrate protein HIPK2. Mol Cell *24*, 77-89.

Rose, N.R., King, H.W., Blackledge, N.P., Fursova, N.A., Ember, K.J., Fischer, R., Kessler, B.M., and Klose, R.J. (2016). RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. Elife *5*.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R.*, et al.* (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389-393.

Rossant, J., and Tam, P.P. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development *136*, 701-713.

Rugg-Gunn, P.J., Cox, B.J., Ralston, A., and Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc Natl Acad Sci U S A *107*, 10783-10790.

Salifou, K., Ray, S., Verrier, L., Aguirrebengoa, M., Trouche, D., Panov, K.I., and Vandromme, M. (2016). The histone demethylase JMJD2A/KDM4A links ribosomal RNA transcription to nutrients and growth factors availability. Nat Commun *7*, 10174.

Sanchez, C., Sanchez, I., Demmers, J.A., Rodriguez, P., Strouboulis, J., and Vidal, M. (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. Mol Cell Proteomics *6*, 820-834.

Sanchez-Pulido, L., Devos, D., Sung, Z.R., and Calonje, M. (2008). RAWUL: A new ubiquitin-like domain in PRC1 Ring finger proteins that unveils putative plant and worm PRC1 orthologs, pp. 308-308.

Santanach, A., Blanco, E., Jiang, H., Molloy, K.R., Sanso, M., LaCava, J., Morey, L., and Di Croce, L. (2017). The Polycomb group protein CBX6 is an essential regulator of embryonic stem cell identity. Nat Commun *8*, 1235.

Santini, L., Halbritter, F., Titz-Teixeira, F., Suzuki, T., Asami, M., Ma, X., Ramesmayer, J., Lackner, A., Warr, N., Pauler, F.*, et al.* (2021). Genomic imprinting in mouse blastocysts is predominantly associated with H3K27me3. Nat Commun *12*, 3804.

Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S.T., Offer, J.*, et al.* (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. Mol Cell *57*, 769-783.

Satijn, D.P., Gunster, M.J., van der Vlag, J., Hamer, K.M., Schul, W., Alkema, M.J., Saurin, A.J., Freemont, P.S., van Driel, R., and Otte, A.P. (1997a). RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor. Mol Cell Biol *17*, 4105-4113.

Satijn, D.P., Olson, D.J., van der Vlag, J., Hamer, K.M., Lambrechts, C., Masselink, H., Gunster, M.J., Sewalt, R.G., van Driel, R., and Otte, A.P. (1997b). Interference with the expression of a novel human polycomb protein, hPc2, results in cellular transformation and apoptosis. Mol Cell Biol *17*, 6076-6086.

Saurin, A.J., Shao, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R.E. (2001). A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. Nature *412*, 655-660.

Saurin, A.J., Shiels, C., Williamson, J., Satijn, D.P., Otte, A.P., Sheer, D., and Freemont, P.S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. J Cell Biol *142*, 887-898.

Scelfo, A., Fernandez-Perez, D., Tamburri, S., Zanotti, M., Lavarone, E., Soldi, M., Bonaldi, T., Ferrari, K.J., and Pasini, D. (2019). Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities. Mol Cell *74*, 1037-1052 e1037.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Muller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature *465*, 243-247.

Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S.*, et al.* (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell *42*, 330-341.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J *25*, 3110-3122.

Schoorlemmer, J., Marcos-Gutierrez, C., Were, F., Martinez, R., Garcia, E., Satijn, D.P., Otte, A.P., and Vidal, M. (1997). Ring1A is a transcriptional repressor that interacts with the Polycomb-M33 protein and is expressed at rhombomere boundaries in the mouse hindbrain. EMBO J *16*, 5930-5942.

Schuettengruber, B., Bourbon, H.M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell *171*, 34-57.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome Regulation by Polycomb and Trithorax Proteins, pp. 735-745.

Scott, C.L., Gil, J., Hernando, E., Teruya-Feldstein, J., Narita, M., Martinez, D., Visakorpi, T., Mu, D., Cordon-Cardo, C., Peters, G.*, et al.* (2007). Role of the chromobox protein CBX7 in lymphomagenesis. Proc Natl Acad Sci U S A *104*, 5389-5394.

Senthilkumar, R., and Mishra, R.K. (2009). Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. BMC Genomics *10*, 549.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

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

Shaver, S., Casas-Mollano, J.A., Cerny, R.L., and Cerutti, H. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. Epigenetics *5*, 301-312.

Shumacher, A., Faust, C., and Magnuson, T. (1996). Positional cloning of a global regulator of anterior-posterior patterning in mice. Nature *383*, 250-253.

Sinclair, D.A., Syrzycka, M., Macauley, M.S., Rastgardani, T., Komljenovic, I., Vocadlo, D.J., Brock, H.W., and Honda, B.M. (2009). Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). Proc Natl Acad Sci U S A *106*, 13427-13432.

Slifer, E.H. (1942). A mutant stock of Drosophila with extra sex-combs. Journal of Experimental Zoology *90*, 31-40.

Sneppen, K., and Ringrose, L. (2019). Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent. Nat Commun *10*, 2133.

Soria-Bretones, I., Cepeda-Garcia, C., Checa-Rodriguez, C., Heyer, V., Reina-San-Martin, B., Soutoglou, E., and Huertas, P. (2017). DNA end resection requires constitutive sumoylation of CtIP by CBX4. Nat Commun *8*, 113.

Soshnikova, N., and Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes in vivo. Science *324*, 1320-1323.

Soto, M.C., Chou, T.B., and Bender, W. (1995). Comparison of germline mosaics of genes in the Polycomb group of Drosophila melanogaster. Genetics *140*, 231-243.

Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. PLoS Genet *14*, e1007193.

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

Storre, J., Elsässer, H.P., Fuchs, M., Ullmann, D., Livingston, D.M., and Gaubatz, S. (2002). Homeotic transformations of the axial skeleton that accompany a targeted deletion of E2f6. EMBO Reports *3*, 695-700.

Struhl, G. (1981). A gene product required for correct initiation of segmental determination in Drosophila. Nature *293*, 36-41.

Struhl, G. (1982). Spineless-Aristapedia: A Homeotic Gene That Does Not Control the Development of Specific Compartments in Drosophila. Genetics *102*, 737-749.

Struhl, G., and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. EMBO J *4*, 3259-3264.

Strutt, H., and Paro, R. (1997). The polycomb group protein complex of Drosophila melanogaster has different compositions at different target genes. Mol Cell Biol *17*, 6773-6783.

Suzuki, M., Mizutani-Koseki, Y., Fujimura, Y., Miyagishima, H., Kaneko, T., Takada, Y., Akasaka, T., Tanzawa, H., Takihara, Y., Nakano, M.*, et al.* (2002). Involvement of the Polycomb-group gene Ring1B in the specification of the anterior-posterior axis in mice. Development *129*, 4171-4183.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P.*, et al.* (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res *47*, D607-D613.

Taherbhoy, A.M., Huang, O.W., and Cochran, A.G. (2015). BMI1-RING1B is an autoinhibited RING E3 ubiquitin ligase. Nat Commun *6*, 7621.

Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y. (2007). The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. Nature *447*, 601-605.

Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature *256*, 640-642.

Tan, J., Jones, M., Koseki, H., Nakayama, M., Muntean, A.G., Maillard, I., and Hess, J.L. (2011). CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. Cancer Cell *20*, 563-575.

Tanay, A., O'Donnell, A.H., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proc Natl Acad Sci U S A *104*, 5521-5526.

Tang, B., Tian, Y., Liao, Y., Li, Z., Yu, S., Su, H., Zhong, F., Yuan, G., Wang, Y., Yu, H.*, et al.* (2019). CBX8 exhibits oncogenic properties and serves as a prognostic factor in hepatocellular carcinoma. Cell Death Dis *10*, 52.

Tapial, J., Ha, K.C.H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallieres, M., Permanyer, J., Sodaei, R., Marquez, Y.*, et al.* (2017). An atlas of alternative splicing profiles and functional associations reveals

new regulatory programs and genes that simultaneously express multiple major isoforms. Genome Res *27*, 1759-1768.

Tardat, M., Albert, M., Kunzmann, R., Liu, Z., Kaustov, L., Thierry, R., Duan, S., Brykczynska, U., Arrowsmith, C.H., and Peters, A.H. (2015). Cbx2 targets PRC1 to constitutive heterochromatin in mouse zygotes in a parent-of-origin-dependent manner. Mol Cell *58*, 157-171.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H.*, et al.* (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664-678.

Teo, G., Liu, G., Zhang, J., Nesvizhskii, A.I., Gingras, A.C., and Choi, H. (2014). SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. J Proteomics *100*, 37-43.

Tessarz, P., Santos-Rosa, H., Robson, S.C., Sylvestersen, K.B., Nelson, C.J., Nielsen, M.L., and Kouzarides, T. (2014). Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. Nature *505*, 564-568.

Testa, G. (2011). The time of timing: how Polycomb proteins regulate neurogenesis. Bioessays *33*, 519-528.

Theiler, K. (1989). The House Mouse: Atlas of Embryonic Development.

Tie, F., Furuyama, T., and Harte, P.J. (1998). The Drosophila Polycomb Group proteins ESC and E(Z) bind directly to each other and co-localize at multiple chromosomal sites. Development (Cambridge, England) *125*, 3483-3496.

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., and Harte, P.J. (2001). The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. Development *128*, 275-286.

Tie, F., Prasad-Sinha, J., Birve, A., Rasmuson-Lestander, A., and Harte, P.J. (2003). A 1-megadalton ESC/E(Z) complex from Drosophila that contains polycomblike and RPD3. Mol Cell Biol *23*, 3352-3362.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trojer, P., Cao, A.R., Gao, Z., Li, Y., Zhang, J., Xu, X., Li, G., Losson, R., Erdjument-Bromage, H., Tempst, P.*, et al.* (2011). L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. Mol Cell *42*, 438-450.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J *13*, 3822-3831.

Ule, J., and Blencowe, B.J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. Mol Cell *76*, 329-345.

van den Boom, V., Maat, H., Geugien, M., Rodriguez Lopez, A., Sotoca, A.M., Jaques, J., Brouwers-Vos, A.Z., Fusetti, F., Groen, R.W., Yuan, H*., et al.* (2016). Non-canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis. Cell Rep *14*, 332-346.

van den Boom, V., Rozenveld-Geugien, M., Bonardi, F., Malanga, D., van Gosliga, D., Heijink, A.M., Viglietto, G., Morrone, G., Fusetti, F., Vellenga, E*., et al.* (2013). Nonredundant and locus-specific gene repression functions of PRC1 paralog family members in human hematopoietic stem/progenitor cells. Blood *121*, 2452-2461.

van der Lugt, N.M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M*., et al.* (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. Genes Dev *8*, 757-769.

van Lohuizen, M., Frasch, M., Wientjens, E., and Berns, A. (1991). Sequence similarity between the mammalian bmi-1 proto-oncogene and the Drosophila regulatory genes Psc and Su(z)2. Nature *353*, 353-355.

Vandamme, J., Volkel, P., Rosnoblet, C., Le Faou, P., and Angrand, P.O. (2011). Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. Mol Cell Proteomics *10*, M110 002642.

Vincenz, C., and Kerppola, T.K. (2008). Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. Proc Natl Acad Sci U S A *105*, 16572-16577.

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

Voigt, P., Tee, W.W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes Dev *27*, 1318-1338.

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

Waddington, C.H. (1968). Towards a theoretical biology. Nature *218*, 525-527.

Waddington, C.H. (2012). The epigenotype. 1942. Int J Epidemiol *41*, 10-13.

Wang, B., Tang, J., Liao, D., Wang, G., Zhang, M., Sang, Y., Cao, J., Wu, Y., Zhang, R., Li, S.*, et al.* (2013). Chromobox homolog 4 is correlated with prognosis and tumor cell growth in hepatocellular carcinoma. Ann Surg Oncol *20 Suppl 3*, S684-692.

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

Wang, R., Ilangovan, U., Robinson, A.K., Schirf, V., Schwarz, P.M., Lafer, E.M., Demeler, B., Hinck, A.P., and Kim, C.A. (2008). Structural transitions of the RING1B C-terminal region upon binding the polycomb cbox domain. Biochemistry *47*, 8007-8015.

Wang, R., Taylor, A.B., Leal, B.Z., Chadwell, L.V., Ilangovan, U., Robinson, A.K., Schirf, V., Hart, P.J., Lafer, E.M., Demeler, B.*, et al.* (2010). Polycomb group targeting through different binding partners of RING1B C-terminal domain. Structure *18*, 966-975.

Wang, X., Goodrich, K.J., Gooding, A.R., Naeem, H., Archer, S., Paucek, R.D., Youmans, D.T., Cech, T.R., and Davidovich, C. (2017). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. Mol Cell *65*, 1056-1067 e1055.

Wang, X., Li, L., Wu, Y., Zhang, R., Zhang, M., Liao, D., Wang, G., Qin, G., Xu, R.H., and Kang, T. (2016). CBX4 Suppresses Metastasis via Recruitment of HDAC3 to the Runx2 Promoter in Colorectal Carcinoma. Cancer Res *76*, 7277-7289.

Wedeen, C., Harding, K., and Levine, M. (1986). Spatial regulation of Antennapedia and bithorax gene expression by the Polycomb locus in Drosophila. Cell *44*, 739-748.

Weiner, A., Lara-Astiaso, D., Krupalnik, V., Gafni, O., David, E., Winter, D.R., Hanna, J.H., and Amit, I. (2016). Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution. Nat Biotechnol *34*, 953-961.

Wellik, D.M. (2007). Hox patterning of the vertebrate axial skeleton. Dev Dyn *236*, 2454-2463.

Whitcomb, S.J., Basu, A., Allis, C.D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. Trends Genet *23*, 494-502.

Wiederschain, D., Chen, L., Johnson, B., Bettano, K., Jackson, D., Taraszka, J., Wang, Y.K., Jones, M.D., Morrissey, M., Deeds, J.*, et al.* (2007). Contribution of polycomb homologues Bmi-1 and Mel-18 to medulloblastoma pathogenesis. Mol Cell Biol *27*, 4968-4979.

Wong, S.J., Gearhart, M.D., Taylor, A.B., Nanyes, D.R., Ha, D.J., Robinson, A.K., Artigas, J.A., Lee, O.J., Demeler, B., Hart, P.J.*, et al.* (2016). KDM2B Recruitment of the Polycomb Group Complex, PRC1.1, Requires Cooperation between PCGF1 and BCORL1. Structure *24*, 1795-1801.

Wong, S.J., Senkovich, O., Artigas, J.A., Gearhart, M.D., Ilangovan, U., Graham, D.W., Abel, K.N., Yu, T., Hinck, A.P., Bardwell, V.J.*, et al.* (2020). Structure and Role of BCOR PUFD in Noncanonical PRC1 Assembly and Disease. Biochemistry *59*, 2718-2728.

Wood, H.B., and Episkopou, V. (1999). Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. Mech Dev *86*, 197-201.

Wotton, D., and Merrill, J.C. (2007). Pc2 and SUMOylation. Biochem Soc Trans *35*, 1401-1404.

Wu, H., Zeng, H., Dong, A., Li, F., He, H., Senisterra, G., Seitova, A., Duan, S., Brown, P.J., Vedadi, M.*, et al.* (2013a). Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations. PLoS One *8*, e83737.

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

Xiao, W., Ou, C., Qin, J., Xing, F., Sun, Y., Li, Z., and Qiu, J. (2014). CBX8, a novel DNA repair protein, promotes tumorigenesis in human esophageal carcinoma. Int J Clin Exp Pathol *7*, 4817-4826.

Xie, W., Ling, T., Zhou, Y., Feng, W., Zhu, Q., Stunnenberg, H.G., Grummt, I., and Tao, W. (2012). The chromatin remodeling complex NuRD establishes the poised state of rRNA genes characterized by bivalent histone modifications and altered nucleosome positions. Proc Natl Acad Sci U S A *109*, 8161-8166.

Yang, J., Cheng, D., Zhu, B., Zhou, S., Ying, T., and Yang, Q. (2016). Chromobox Homolog 4 is Positively Correlated to Tumor Growth, Survival and Activation of HIF-1alpha Signaling in Human Osteosarcoma under Normoxic Condition. J Cancer *7*, 427-435.

Ying, Q.L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol *21*, 183-186.

Youmans, D.T., Gooding, A.R., Dowell, R.D., and Cech, T.R. (2021). Competition between PRC2.1 and 2.2 subcomplexes regulates PRC2 chromatin occupancy in human stem cells. Mol Cell *81*, 488-501 e489.

Young, N.L., DiMaggio, P.A., Plazas-Mayorca, M.D., Baliban, R.C., Floudas, C.A., and Garcia, B.A. (2009). High throughput characterization of combinatorial histone codes. Mol Cell Proteomics *8*, 2266-2284.

Yuan, G.J., Chen, X., Lu, J., Feng, Z.H., Chen, S.L., Chen, R.X., Wei, W.S., Zhou, F.J., and Xie, D. (2017). Chromobox homolog 8 is a predictor of muscle invasive bladder cancer and promotes cell proliferation by repressing the p53 pathway. Cancer Sci *108*, 2166-2175.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T.*, et al.* (2012). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science *337*, 971-975.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem *286*, 7983-7989.

Zencak, D., Lingbeek, M., Kostic, C., Tekaya, M., Tanger, E., Hornfeld, D., Jaquet, M., Munier, F.L., Schorderet, D.F., van Lohuizen, M.*, et al.* (2005). Bmi1 loss produces an increase in astroglial cells and a decrease in neural stem cell population and proliferation. J Neurosci *25*, 5774-5783.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q.*, et al.* (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537*, 553-557.

Zhang, H., Wu, Z., Lu, J.Y., Huang, B., Zhou, H., Xie, W., Wang, J., and Shen, X. (2020). DEAD-Box Helicase 18 Counteracts PRC2 to Safeguard Ribosomal DNA in Pluripotency Regulation. Cell Rep *30*, 81-97 e87.

Zhang, Q., Agius, S.C., Flanigan, S.F., Uckelmann, M., Levina, V., Owen, B.M., and Davidovich, C. (2021). PALI1 facilitates DNA and nucleosome binding by PRC2 and triggers an allosteric activation of catalysis. Nat Commun *12*, 4592.

Zhang, Q., McKenzie, N.J., Warneford-Thomson, R., Gail, E.H., Flanigan, S.F., Owen, B.M., Lauman, R., Levina, V., Garcia, B.A., Schittenhelm, R.B.*, et al.* (2019). RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. Nat Struct Mol Biol *26*, 237-247.

Zhang, Y. (2003). Transcriptional regulation by histone ubiquitination and deubiquitination. Genes Dev *17*, 2733-2740.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhao, J., Wang, M., Chang, L., Yu, J., Song, A., Liu, C., Huang, W., Zhang, T., Wu, X., Shen, X.*, et al.* (2020). RYBP/YAF2-PRC1 complexes and histone H1-dependent chromatin compaction mediate propagation of H2AK119ub1 during cell division. Nat Cell Biol *22*, 439-452.

Zhao, W., Tong, H., Huang, Y., Yan, Y., Teng, H., Xia, Y., Jiang, Q., and Qin, J. (2017). Essential Role for Polycomb Group Protein Pcgf6 in Embryonic Stem Cell Maintenance and a Noncanonical Polycomb Repressive Complex 1 (PRC1) Integrity. J Biol Chem *292*, 2773-2784.

Zhao, X., Rastogi, A., Deton Cabanillas, A.F., Ait Mohamed, O., Cantrel, C., Lombard, B., Murik, O., Genovesio, A., Bowler, C., Bouyer, D.*, et al.* (2021). Genome wide natural variation of H3K27me3 selectively marks genes predicted to be important for cell differentiation in Phaeodactylum tricornutum. New Phytol *229*, 3208-3220.

Zhen, C.Y., Tatavosian, R., Huynh, T.N., Duc, H.N., Das, R., Kokotovic, M., Grimm, J.B., Lavis, L.D., Lee, J., Mejia, F.J.*, et al.* (2016). Live-cell single-molecule tracking reveals co-recognition of H3K27me3 and DNA targets polycomb Cbx7-PRC1 to chromatin. Elife *5*.

Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X.*, et al.* (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. Mol Cell *63*, 1066-1079.

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

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