PhD degree in Systems Medicine (curriculum in Computational Biology)

European School of Molecular Medicine (SEMM)

Univerdity of Milan and University of Napes "Federico II"

Settore disciplinare: BIO/11

# CONTROL OF TRANSCRIPTIONAL IDENTITY BY POLYCOMB GROUP PROTEINS

*Daniel Fernández Pérez*

IEO, Milan

*Tutor:* Prof. Diego Pasini

IEO, Milan

*PhD Coordinator:* Prof. Saverio Minucci

Anno accademico 2020-2021

# Contents

# List of Figures

## Abstract

Establishment and maintenance of transcriptional identity is the cornerstone of tissue organization and normal development in multicellular organisms. By establishing facultative heterochromatin, Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2) constitute the main dynamic and plastic mechanism to maintain gene repression in eukaryotic cells. PRC1 and PRC2 modify histones by depositing H2AK119Ub1 and H3K27me3, respectively, and both can be further subdivided in multiple sub-complexes based on their biochemical composition. In particular, PRC1 can be subdivided in 6 different sub-complexes depending on which PCGF interacts with its catalytic subunits RING1A/B (PCGF1-6). It is not clear if these sub-complexes act redundantly to maintain PRC1 function or if instead they have specific activities. It is also under debate which is the role of its catalytic product, H2AK119Ub1, in establishing PcG-mediated gene repression. Here, by combining the use of engineered mouse embryonic stem cell lines (mESC) and transgenic mice, we have been able to delete individual PRC1 sub-complexes *in vitro* and *in vivo* to assess their roles in orchestrating PRC1 function. Our data show that individual PRC1 sub-complexes regulate distinct biological processes and are able to maintain binding specificity both *in vitro* and *in vivo*. We also demonstrate that PRC1.1 and PRC1.2/4 are responsible for the maintenance of most of H2AK119Ub1 levels associated with promoter regions, however, the residual PRC1 activity from PRC1.3/5 and PRC1.6 is enough to maintain low H2AK119Ub1 levels and PcG-mediated gene repression. Finally, by using a fully catalytic inactive PRC1 in mESC, we demonstrate that H2AK119Ub1 constitutes the central hub of PcG-mediated gene repression. Complete loss of H2AK119Ub1 abolishes PcG-mediated gene repression and induces the disassembly of PcG chromatin domains by displacing PRC2.2 and cPRC1

from chromatin. Importantly, removal of PRC2 and H3K27me3 does not phenocopy H2AK119Ub1 loss, demonstrating that H2AK119Ub1 is central to PcG system function.

# 1 Introduction

## 1.1 Transcription initiation and control of gene expression

The development of multicellular organisms and establishment of cellular identity relies on the capacity of single cells to activate and maintain specific gene expression programs. The genetic information contained in the DNA, as if it was a how to book, contains the protein-coding and non-coding regulatory sequences that will determine when, where and to which level every region of the genome will be transcribed [1].

Transcription usually initiates at a defined position, which is called Transcription Start Site (TSS). At this position, the transcription machinery -RNA polymerase II (Pol II) and general transcription factors (GTFs)- is recruited to the core promoter, which consists of a short sequence around 50bp located upstream and downstream of the TSS. The core promoter is associated with a basal level of transcription, which can be enhanced by distally located regulatory elements called enhancers. Enhancers can boost transcription at core promoters by binding transcription factors (TFs), which then recruit transcription cofactors that will potentiate the recruitment of Pol II. Contrary to core promoters, enhancers are not position- and orientation-

dependent, which allows them to boost transcription independently of their location respect to their target genes [2, 3].

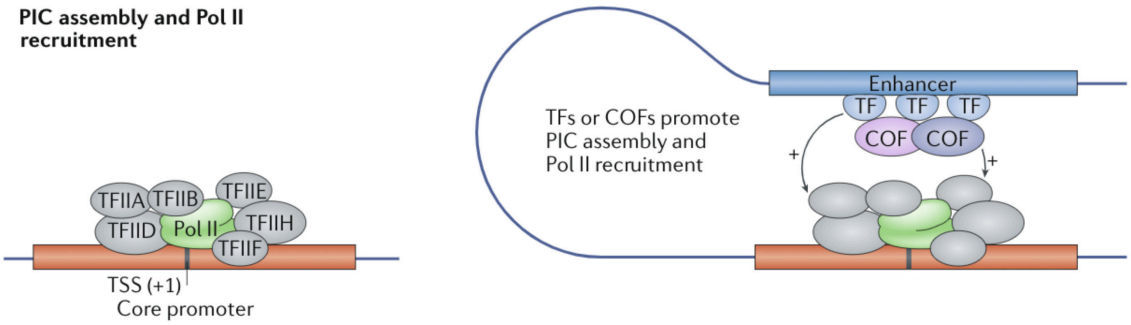


**Figure 1.1:** Recruitment of transcription machinery to core promoters. Adapted from [4].

### 1.1.1 CpG islands and their role in transcription initiation

In vertebrates, around 70% of all promoters are associated with CpG islands (CGIs) [5]. CGIs are genomic regions of around 1kb with a high frequency of CpG dinucleotides and high GC content, as opposite to most of the genome, which is CpG-deficient [6]. The methylation of vertebrate genomes is mainly accumulated at CpG dinucleotides, however, CGIs very frequently are depleted of DNA methylation [6]. Around 50% of CGIs are associated with TSS, while the other half, located within or between promoters, are named *orphan* CGIs and seem to be also related with transcription initiation [6].

CGIs do not contain sequence promoter elements such as TATA box, but are still correlated with promoter function. The high density of CpG dinucleotides and CG in CGIs seems to be related with higher affinity for TF binding, since many TF motifs are enriched in CpG dinucleotides, such as Sp1, Nrf-1, E2F and ETS [7]. In the same way, those motifs are enriched in the promoters of housekeeping genes [8].

CGIs have been shown to be able to stimulate the deposition of H3K4me3, H3K27me3 and H2AK119Ub1, which correlate with both active and repressed transcription [9]. Also, although just in specific contexts, CGIs can also be DNA methylated (for example, during development [10] or in cancer [11]), and thus, silenced. Overall it seems that these CGIs provide a tethering point with high affinity for different mechanisms of gene regulation, which can both activate and repress transcription. In this sense, CGIs would be like a battlefield where different complexes will compete with each other to ”conquer” CGIs and determine the transcriptional status of gene promoters. The competition of these complexes for CGIs will be also determined by the activity of other TFs, which will be activated in specific cellular contexts, and will determine the final transcriptional status and identity of a cell [9].

## 1.2 Chromatin: A template for transcriptional regulation

In order to regulate the transcriptional state of target promoters, TFs have to interact with regulatory elements. However, the complexity of eukaryotic genomes and their transcriptional programs require more than site-specific binding of TFs to be governed. In the recent years, it has been more and more clear that the physical and biochemical structure of chromatin, the template where transcription occurs, plays a key role in the regulation of gene expression. A question that arises now is how these DNA-binding proteins interact with other enzymes that chemically and physically modify chromatin, in order to determine, in specific cellular contexts, which areas of the genome will be transcribed [12]. The main mechanisms by which chromatin participates in the regulation of transcription can be generally divided in [12]:

1. **DNA accessibility**, which regulates the access of TFs to regulatory elements.
2. **Long-range interactions**, which determine the contacts between regulatory elements (enhancers) and target promoters.
3. **Specific histone modifications**, which can both stimulate and repress transcription.

During the following pages of this thesis I will discuss which are the characteristics of the chromatin template, and which are the resources that eukaryotic cells have in order be able to mold chromatin to their will to regulate gene expression.

### 1.2.1 Characteristics of chromatin

For eukaryotic cells, the organization of genomic information represents a dilemma. On one side, they need to be able to have easy and precise access to their DNA sequences for replicative and transcriptonal purposes, but at the same time, they need it to be tightly compact and structured due to space limitations and the need to finely regulate gene expression. In order to satisfy these needs, eukaryotic DNA is wrapped around an octameric protein complex termed nucleosome, which corresponds to the basic unit of the chromatin fiber [13]. 147bp of DNA hug each nucleosome, which is composed by 2 copies of each core histone (H2A, H2B, H3, and H4). The positive net charge of nucleosomes allows for a very stable interaction with DNA since the last one is negatively charged [14].

The structure of chromatin is very important for the regulation of the accessibility to DNA. Nucleosomes can force the DNA to bend and acquire 3D conformations that can reduce its accessibility to the transcription machinery and TFs, and also by themselves can physically block the interaction of those elements with DNA (Figure 1.2). By modifying biochemically and physically the structure of nucleosomes,

eukaryotic cells can regulate the interaction of nucleosomes with DNA and alter the chromatin structure to regulate gene expression [15].

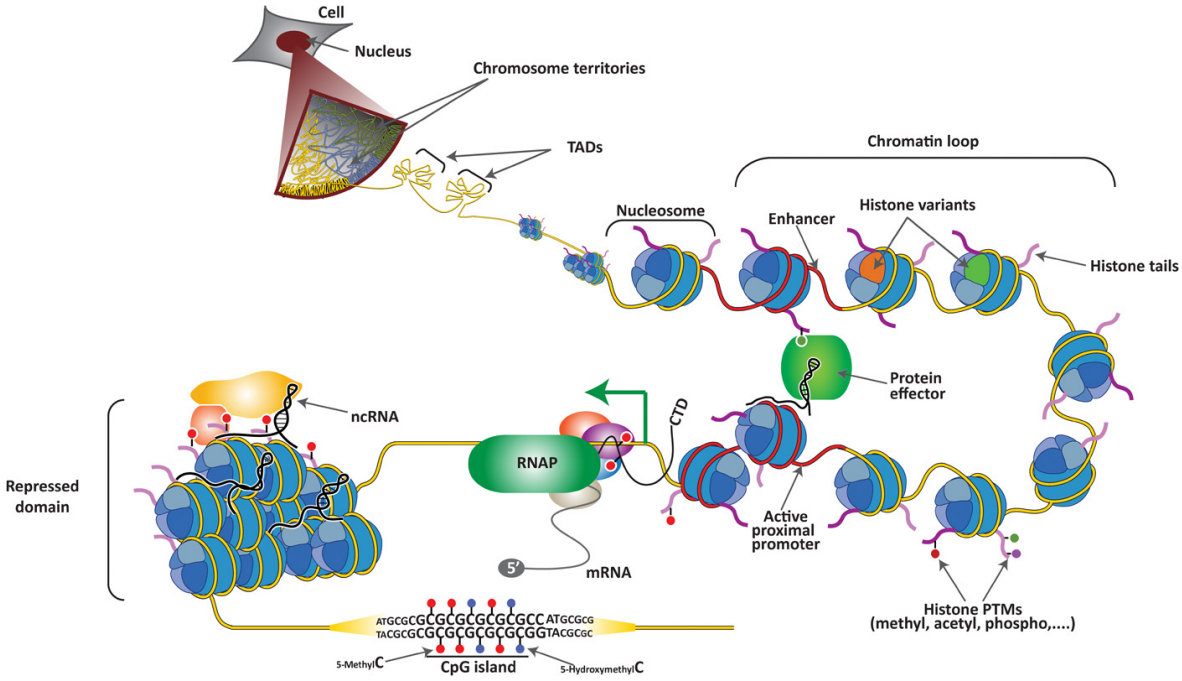


**Figure 1.2:** Chromatin structure and organization. Adapted from [16].

### 1.2.2 Chromatin accessibility

Chromatin is not structured in the same way across the whole genome. Different topological organizations of nucleosomes define regions of low and high nucleosomal density (Figure 1.3). Regions corresponding to facultative and constitutive heterocromatin are associated with very dense arrays of nucleosomes, meanwhile regions corresponding to euchromatin are more enriched in genes and are lightly packed. Then, regulatory regions such as enhancers and gene promoters are depleted of histones [17]. As a curiosity, just between 2 and 3 percent of the human genome correspond to accessible regions (depleted of nucleosomes) but it accounts for 90% of all the genomic loci bound by transcription factors [18]. This highly dynamic reorganization of nucleosomes, which allows to quickly change form acces-

sible to close chromatin, is at the basis of chromatin organization and function, and the regulation of this elements constitute the basic toolkit for any eukaryotic cell to regulate access to DNA.

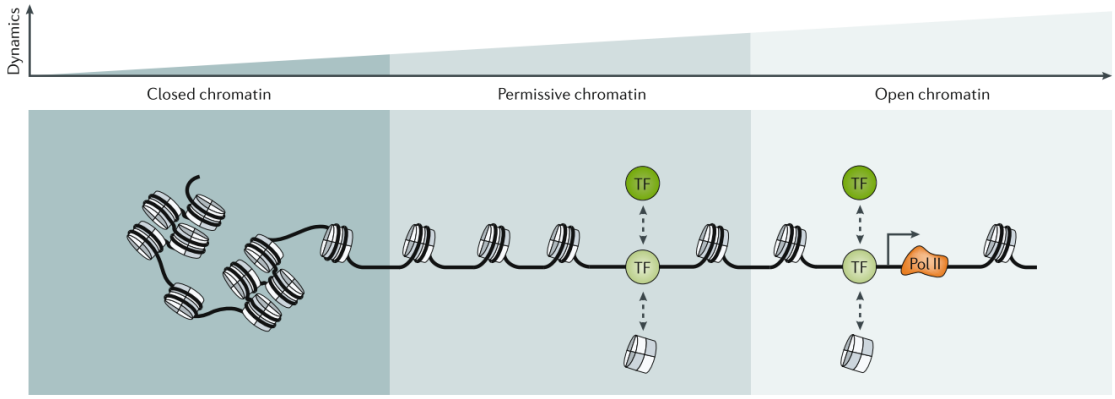


**Figure 1.3:** Different topological organization of nucleosomes define regions of low and high nucleosomal density. Adapted from [15].

The physical determinants of chromatin accessibility depend on multiple factors, in particular nucleosome occupancy and the activity of TFs and structural proteins. Nucleosome occupancy depends on its turnover rate, which depends on DNA sequences, TFs, histone variants, histone chaperones and local chromatin context [19, 20]. Nucleosomes are physically removed or displaced from chromatin by chromatin remodeling complexes, but those proteins usually lack DNA sequence recognition, thus, they need the action of TFs that first bind chromatin and then recruit the remodeling machinery to alter nucleosome organization [18]. TFs, which usually bind already open chromatin, can interact with DNA to destabilize histone-DNA interactions and create a local environment that can promote the binding of other TFs and cofactors to stabilize open chromatin. Passive competition with histones (which involves binding directly to DNA associated to nucleosomes, which seems to be transiently accessible [17]), displacement of linker histones (binding to linker -accessible- DNA sequences and then displacing H1 histones), distant interactions in trans or the

activity of pioneer TFs are thought to be the main mechanisms (reviewed in [15]). On top of that, the recruitment of TFs can be altered by the presence of non-canonical histone variants, which are enriched in regulatory regions such as promoters and enhancers and seem to be associated with different nucleosome turnover rates [21].

### 1.2.3 Chromatin long-range interactions

Packaging around 2 m of DNA into a 15 µm nucleus in eukaryotic cells is expected to produce non-specific chromatin interactions across distant genomic regions. However, thanks to functional and unbiased mapping studies [22], we know that in a wide range of organisms exist many functionally relevant high-frequency interactions that can't be explained stochastically [23]. Then, which are the biological advantages of having nonrandom spatial organization in 3D space?

Long-range interaction can be divided into intra and interchromosomal interactions. Intrachromosomal interactions, which occur between regions located in the same chromosome, usually correspond to interactions between promoters and regulatory elements (enhancers) or isulator-mediated contacts [24]. On the other side, interchromosmal interactions, which happen between different chromosomes, are less frequent, but still seem to be important for gene regulation and chromosomal translocations [23].

Long-range interactions play a key role in nuclear architecture, allowing the compartmentalization of eukaryotic genomes into different "areas" that play a key role in the regulation of gene expression [25]. For example, regions in the periphery and regions that are in contact with the nuclear lamina -called lamina-associated domains (LADs)- are associated with poor gene expression [26], meanwhile genes recruited to transcription factories, which are nuclear structures consisting of multiple genes

and copies of hyperphosphorylated Pol II, are associated with high gene expression [27]. Interactions between distant genomic elements are also important to bring enhancers in contact with promoters. One of the most famous cases is the gene *sonic hedgehog*, which is regulated by an enhancer located 1Mb away [28]. The *HoxD* locus is another example of the role of long-range interactions in the regulation of gene expression, where a single locus is regulated by many different enhancers, and those contacts occur in very controlled specific moments during development [29].

### 1.2.4 Histone modifications

Histone post-translational modifications (PTMs) are known to participate in the regulation of gene expression more than 50 years ago [30]. Since then, many histone PTMs have been discovered, such as acetylation, methylation, phosphorylation, sumoylation, ADP ribosylation, and deamination [31]. Until recently, research has been focused primarily on histone PTMs located at the N-terminal *tail* of histones, which are separated from the central structured histone domains and are potentially more accessible to be recognized by binder proteins than modifications in the globular part of the protein [32]. Novel globular modifications have recently demonstrated to play important roles in the definition of the chromatin state and regulation of gene expression [33]. In this introduction we will focus on the most studied PTMs located at the N-terminal region of histone, the tail.

Usually proteins that recognize histone PTMs are called *readers*, meanwhile proteins that modify the amino acids of histones are called *writers*. Following the guidelines from Henikoff and his team [34], in this thesis we will talk about *modifiers* (proteins that modify one amino acid at a time) and *binders* (proteins that bind one residue at a time) instead of writers and readers which are less accurate adjectives.

Another misleading way of referring to histone PTMs is to call them *activating* or *repressive*, which implies causality.Histone PTMs are not essential for transcription, for example, removing H3 or H4 terminal tails in budding yeast produces a viable phenotype [35]. The fact that histone PTMs are correlated with different transcriptional states does not mean that they are the direct *cause* of that state (although they can be important to potentiate or reduce it), and thus, the terminology that refers to histone PTMs should be more accurate.

#### 1.2.4.1 Histone acetylation

Histone acetylation was the first histone PTM to be described [36]. This modification reduces the overall positive charge of histone tails by neutralizing the lysines where it is incorporated, which decreases the affinity of histone tails for DNA. Principal sites of acetylation across the tails of histone H3 are K9, K14, K18, K23 and K27 [37]. Histone acetylation is usually associated with active transcription and concentrated at promoters and enhancers. The dynamics of histone acetylation depend on two families of proteins, histone acetyltransferases (HATs) and histone deacetylases (HDACs) [31]. Histone acetylation can be recognized by the bromodomain of chromatin remodelling complexes [38], which links acetylation to the regulation of nucleosome stability and gene expression. However, as discussed before, this does not mean that acetylation is *essential* for transcriptional activation [35].

#### 1.2.4.2 Histone methylation

As opposed to acetylation, histone methylation doesn't change the net charge of the histone tail. The methylation of lysine occurs at the primary amine of the amino acid by replacing its hydrogen/s, which allows the formation of monomethyl (me1),

dimethyl (me2) and trimethyl (me3) lysines. Arginines can also be mono, symmetrically and assymetricallt methylated [39].

Contrary to acetylation, methylation of lysines at the histone H3 can be correlated with both active and repressed transcription based on the aimo residue and the amount of methylation [37]. For example, monomethylation of H3K4 usually is associated with the presence of enhancers, trimethylation of H3K4 with active transcription and trimethylation of H3K9 or H3K27 with repressed transcription [31]. The addition of methyl groups is performed by a family of lysine methyltransferases (HKMT), in particular by their catalytic SET domain. Differently respect to HATs, HKMTs are very residue specific, for example, SET7/9 is just able to monomethylate H3K4 meanwhile DIM5 can mono, di and trymethylate H3K9 [31].

H3K4me3, which is usually accumulated at gene promoters, is associated with active transcription [40]. This histone PTM has a very interesting pattern deposition across gene bodies. The TSS of active genes is enriched in H3K4me3, then, if we move into the gene bodies, there is an enrichment of H3K4me2 and then of H3K4me1. So, there is a gradient from H3K4me3 to H3K4me1 from the TSS until the first half of the gene body. Then, opposite to H3K4, an increasing gradient of H3K36me3 starts from the beginning of the gene body (3') until the TES (5'). This intriguing gradient of methylation is explained by the affinity of HKMT for the CTD of Pol II. Set1, which catalyzes the methylation of H3K4, has high affinity for the initating form of Pol II (Ser-5 phosphorilation) [41] meanwhile Set2 (which methylates H3K36) could have higher affinity for the elongating form (Ser-2 phosphorilation). This is another example of how different histone modifications can be correlated with transcription without *activating* or *inhibiting* transcription.

Contrary to H3K4me3, H3K9me3 and H3K27me3 are correlated with repressed transcription. H3K9me3 is found mainly in constitutive heterochromatin, mean-

while H3K27me3 is enriched in facultative heterochromatin [40]. It is very interesting the fact that both histone PTMs seem to be mutually exclusive although both of them are associated with gene repression. One potential explanation to this phenomena is that H3K9me3 highly correlates with DNA methylation, which excludes Polycomb Repressive Complex 2 (PRC2) from chromatin (which is the protein complex that catalyzes the mono- di- and trimethylation of H3K27) and could explain the absence of H3K27me3 in H3K9me3+ regions [42].

#### 1.2.4.3 The bivalent genome

There are characteristic genomic domains that are particularly interesting for being decorated with histone PTMs associated with both active and repressed transcription, H3K4me3 and H3K27me3 [43]. These regions, which are enriched (but are not exclusive) in embryonic stem cells (ESCs), usually correspond to promoters and enhancers of key developmental genes. Bivalent domains are in a state termed *poised*, with low levels of basal transcription. It is thought that this poised state protects these genes from *de novo* DNA methylation and allows them to be quickly activated upon differentiating stimuli (Figure 1.4), moment in which one of the two histone marks will prevail over the other and will define a gene or enhancer as definitely activated or repressed [44].

Bivalent domains are defined by the activity of both Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2, both termed as PcG) and SWI/SNF and COMPASS complexes (which form part of the Thritorax gene family, TrxG). PRC2 catalyzes the deposition of H3K27me3 and COMPASS the deposition of H3K4me3 [45]. Many of these bivalent domains are located at promoter regions (termed bivalent promoters), which are rich in CpG islands. The high affinity of PcG and TrxG for CpG-rich regions is what allows to drive the activity of both complexes at bivalent promoters[44].

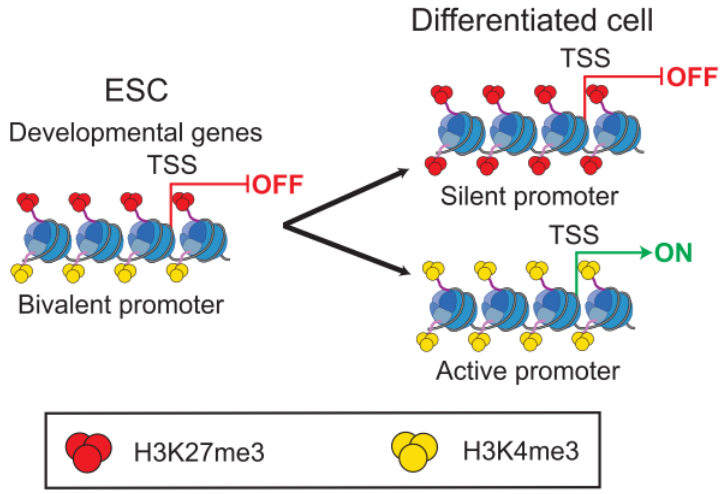


**Figure 1.4:** Development of bivalent genes upon differentiation. Adapted from [44].

It is important to note that the overlap of H3K27me3 and H3K4me3 at bivalent promoters is not an artifact of analyzing in bulk two different populations of cells, but that they are physically on the same nucleosome in opposite histone tails [46].

## 1.3 Polycomb Group Proteins

More than 70 years ago, Pam Lewis discovered the first PcG gene in *Drosophila melanogaster*, called *Polycomb* (*Pc*) [47], which caused partial sex combs on the second and third pairs of legs of adult males. 30 years later, using the mutant that Pam Lewis found, Ed Lewis proposed that *Pc* could be a repressor of *Homeotic* (*Hox*) genes, since mutations in *Pc* induced their ectopic expression [48], which set the path for the future investigations about PcG genes. As years passed, other groups started to find mutations in different genes that caused similar phenotypes to *Polycomb*, and finally they were defined as Polycomb Group (PcG) proteins [49].

Based on their catalytic activity, PcG proteins are divided in two main groups: Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2). With time we have learned that these two complexes are actually much more com-

plex than it was expected, and they can be further subdivided in many different sub-complexes, in particular PRC1. Both complexes will be further described in the following sections of this thesis. Since all the work that will be presented in this thesis has been performed using mouse models and mouse-derived cell lines, mammalian PRC1 and PRC2 complexes will be introduced.

### 1.3.1 Polycomb Repressive Complex 1

As stated previously, PRC1 is actually subdivided in different sub-complexes. All of them share 2 core subunits conserved in mammals and plants [50]. In mammals, they consist in the mutually exclusive catalytic subunits RING1A or RING1B and one of the the six Polycomb group ring-finger domain proteins (PCGF1–PCGF6) [51]. RING1A and RING1B catalyze the mono-ubiquitination of the lysine 119 of the histone H2A (H2AK119Ub1), which is associated with gene repression [50]. Based on which PCGF is forming the complex with RING1A/B a different PRC1 subcomplex will be assembled, since each PCGF will interact with different ancillary proteins. Thus, in total, there are 6 different PRC1 complexes (PRC1.1-6). PCGF2 and PCGF4 form what is called the canonical PRC1 complex (cPRC1) meanwhile PCGF1-3-5-6 form variant PRC1 complexes (vPRC1). PCGF2/4 and PCGF3/5 form biochemically identical complexes, so, biochemically speaking, we can actually distinguish 4 different types of PRC1 complexes (PRC1.1, PRC1.2/4, PRC1.3/5 and PRC1.6) [51].

#### 1.3.1.1 Canonical PRC1

Canonical PRC1 subcomplexes are characterized by the presence of PCGF2 or PCGF4 and one ancillary subunit from the Chromobox Homolog Protein family (CBX2, CBX4, CBX6, CBX7 or CBX8) and a Polyhomeotic Homolog (PHC1, PHC2, PHC3).

Although PCGF2 and PCGF4 have also been detected forming complexes with RYBP, which is present in all vPRC1 complexes, RYBP and CBX+PHC proteins seem to be mutually exclusive [51], making PCGF2 and PCGF4 the only PCGF proteins being able to form cPRC1 complexes.

CBX proteins, through their chromodomain, are able to bind H3K27me3, which is deposited by PRC2. In fact, this constitutes the main known mechanism by which cPRC1 complexes are able to be recruited to chromatin [52]. This fact led to the development of the canonical model of recruitment and establishment of polycomb domains, which will be discussed later in the following sections.

PHC proteins contain a specific domain called SAM, which allows the interaction with other SAM domains from different PHC proteins (so, the establishment of interactions between different cPRC1 complexes). This domain is key for the repressive activity of cPRC1 complexes since its removal causes the upregulation of a subset of PRC1 target genes [52].

Overall, thanks to PHC and CBX proteins, cPRC1 is able to regulate chromatin structure, in part independently of its catalytic activity, which is important to establish transcriptional repression in specific genomic loci [53, 54, 55].

#### 1.3.1.2 Variant PRC1

PRC1 complexes containing PCGF1, PCGF3/5 or PCGF6 form the variant PRC1. vPRC1, respect to cPRC1, is characterized by the presence of the ancillary subunits RYBP and YAF2 (which are mutually exclusive). vPRC1 complexes are more efficient at depositing H2AK119Ub1 than cPRC1 thanks to RYBP, which stimulates the E3 ligase activity of RING1A and RING1B [56] and is able to recognize H2AK119Ub1 [57, 58]. Contrary to PCGF2 and PCGF4, which have been found (in low amounts)

to interact with RYBP, PCGF1, PCGF3/5 and PCGF6 have not been found to interact with CBX and PHC proteins [51].

vPRC1 is more heterogeneous than cPRC1 since the complex is formed by different proteins based on which PCGF (1, 3/5, 6) is interacting with RING1A/B. In Figure 1.5 you can find the main proteins forming each vPRC1 and cPRC1 complex.

One particularity of vPRC1 complexes (except PRC1.3/5) is that they contain proteins with DNA binding capabilities, meanwhile cPRC1 can just be recruited by H3K27me3 (so, PRC2 activity). PRC1.1 has the H3K36 demethylase KDM2B, which recognizes unmethylated CpG islands [59]. In the case of PRC1.6, there are many TFs that recognize E-box and E2F motifs (MAX, MGA, L3MBTL2, E2F6, for example). It has been demonstrated that many of these PRC1.6 subunits are important for its recruitment to chromatin, and it seems that each of them confers different target specificity [60]. In the case of PRC1.3/5, how is it recruited to chromatin remains an open question because none of the proteins forming the complex has DNA or histone binding capabilities.

### 1.3.2 Polycomb Repressive Complex 2

PRC2, which catalyzes the mono-, di- and trimethylation of the lysine 27 of the hisotne H3 (H3K27me1, H3K27me2, and H3K27me3), although not as heterogeneus as PRC1, can also be subdivided in different subcomplexes. The core components of PRC2 consist in its mutually exclusive catalytic subunits enhancer of zeste homolog 1 1 and 2 (EZH1 and EZH2), embryonic ectoderm development (EED), suppressor of zeste (SUZ12), and the CAF1 histone-binding proteins RBBP4 and RBBP7.

Based on which accessory proteins will interact with the core subunits, we can distinguish 2 subtypes of PRC2 (Figure 1.6): PRC2.1 and PRC2.2. These 2 classes

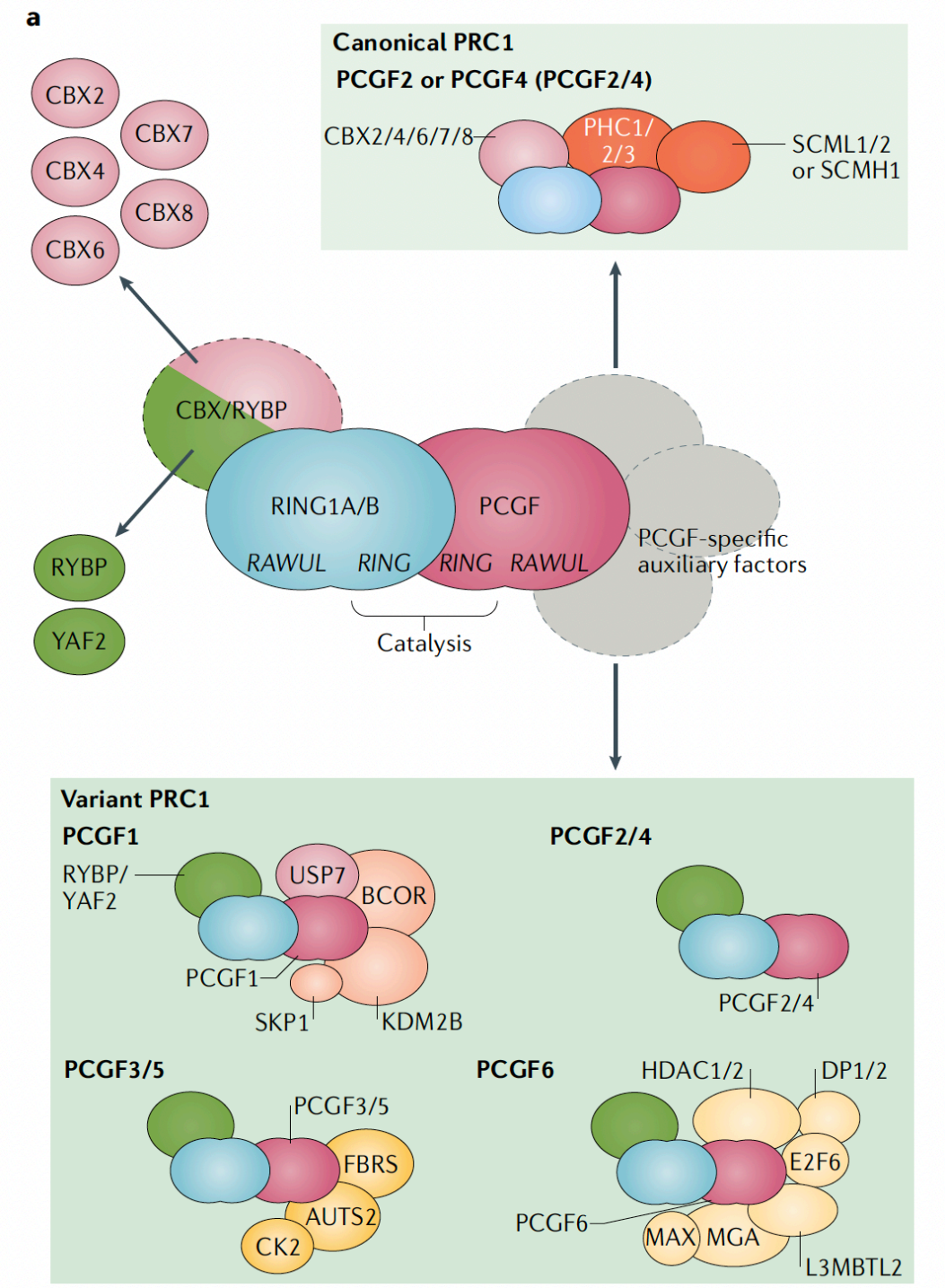


**Figure 1.5:** Biochemical composition of PRC1. Adapted from [61].

of PRC2, as it happens between cPRC1 and vPRC1, will be recruited to chromatin by different mechanisms determined by the accessory proteins. PRC2.1 is characterized by the presence of polycomb-like subunits (PHF1, MTF2, and PHF19) that stimulate EZH2 trimethylation activity and provide affinity for unmethylated CpG islands [62]. Also, one of the two mutually exclusive subunits, EPOP and PALI1, will form part of the complex. In the case of PRC2.2, it is composed by the accesory proteins JARID2 and AEBP2 and, in the case of JARID2 [63, 64], they provide affinity for the H2AK119Ub1 that is deposited by PRC1. So, as it happens for cPRC1 and vPRC1, we have two different groups of PRC2 complexes that are recruited to chromatin by directly binding DNA (PRC2.1) or by recognizing the catalytic product of PRC1, H2AK119Ub1 (PRC2.2).

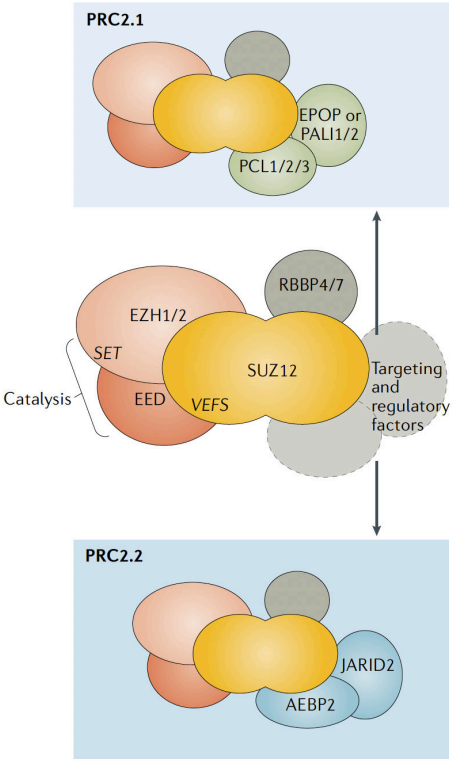


**Figure 1.6:** Biochemical composition of PRC2. Adapted from [61].

### 1.3.3 Recruitment of PcG proteins to chromatin

As we have seen during the previous sections, there are different groups of proteins that provide affinity for chromatin to PRC1 and PRC2. Some of them are able to directly recognize DNA sequences, meanwhile others are able to recognize nucleosomes. In *Drosophila melanogaster*, PcG proteins recognize specific DNA elements called PREs (Polycomb Responsive Elements) [65], however, this elements do not seem to have clear mammalian orthologues [66] and the recruitment of PcG proteins seem to have evolved and become more complex from invertebrates to vertebrates [66]. PcG proteins are recruited to chromatin, in mammals, by 3 different mechanisms: Recognition of specific DNA sequences by TFs, recruitment by RNA-dependent mechanisms and recognition of unmethylated CGIs.

#### 1.3.3.1 Recognition of specific DNA sequences by TFs

As shown in Figure 1.5, PRC1.6 contains many proteins that recognize specific DNA sequences. For example, MAX and MGA provide affinity for E- and T-box motifs, meanwhile E2F6 and DP1 for E2F6-DP1 motifs. This characteristic allows PRC1.6, based on the cellular context, to be recruited to specific targets and regulate specific pathways that do not overlap with classical PcG targets. This specific recruitment of PRC1.6 has an important role in the regulation of germline-related genes [60].

#### 1.3.3.2 Recruitment by chromatin-associated RNAs

Recently it has been postulated that in some cases PcG proteins could be tethered to chromatin by long non-coding RNAs (lncRNAs) [67]. The most studied case is the recruitment of PRC1 and PRC2 during X chromosome inactivation by the lncRNA *Xist*. This mechanism involves vPRC1.3/5, that if deleted, causes female-specific em-

bryo lethality (in mice) and also attenuates the X-chromosome silencing induced by *Xist* [68]. A 600nt region of *Xist*, called XR-PID, is recognized by the RNA-binding protein HNRNPK, which interacts with vPRC1.3/5 and mediates its recruitment to *Xist*, which then tethers vPRC1.3/5 to chromatin. This allows the establishment of megabase-sized H2AK119Ub1 domains, which then will be recognized by the PRC2.2 ancillary subunit JARID2 and will induce the deposition of H3K27me3 that, at the same time, will recruit cPRC1 through its CBX subunit. This mechanism allows the formation of large PcG domains that will be crucial to establish transcriptionally repressed genomic domains during the X chromosome inactivation [69].

Recently there have been also some evidences about the participation of R-loops in the recruitment of PcG proteins to chromatin. R-loops are DNA-RNA hybrids that are usually formed during transcription between nascent mRNA and DNA [70]. It has been demonstrated that a subset of PcG target genes form R-loop structures that contribute to PRC1 genome-wide recruitment and synergyze with PcG proteins to mediate transcriptional silencing at those specific subset of genes [71]. Still, the underlying biochemical mechanisms of PcG and R-loop gene regulation are unknown and need further investigation.

#### 1.3.3.3 PcG affinity for CGIs

TFs and RNAs participate in the recruitment of PcG to a subset of targets, however, in mammals, the main mechanism of PcG recruitment to chromatin is through the recognition of unmethylated CGIs [50]. PcG proteins bind primarily promoter regions, which usually (around 70% in mammals [5]) are rich in CGIs. This affinity of PcG complexes for CGIs is explained by the presence subunits that recognize unmethylated CpG dinucleotides. vPRC1.1 is able to bind CGIs through its subunit

KDM2B (which contains a zinc-finger-CXXC domain) [72] meanwhile PRC2.1 depends on its PCL subunits (which has a winged-helix domain) [62].

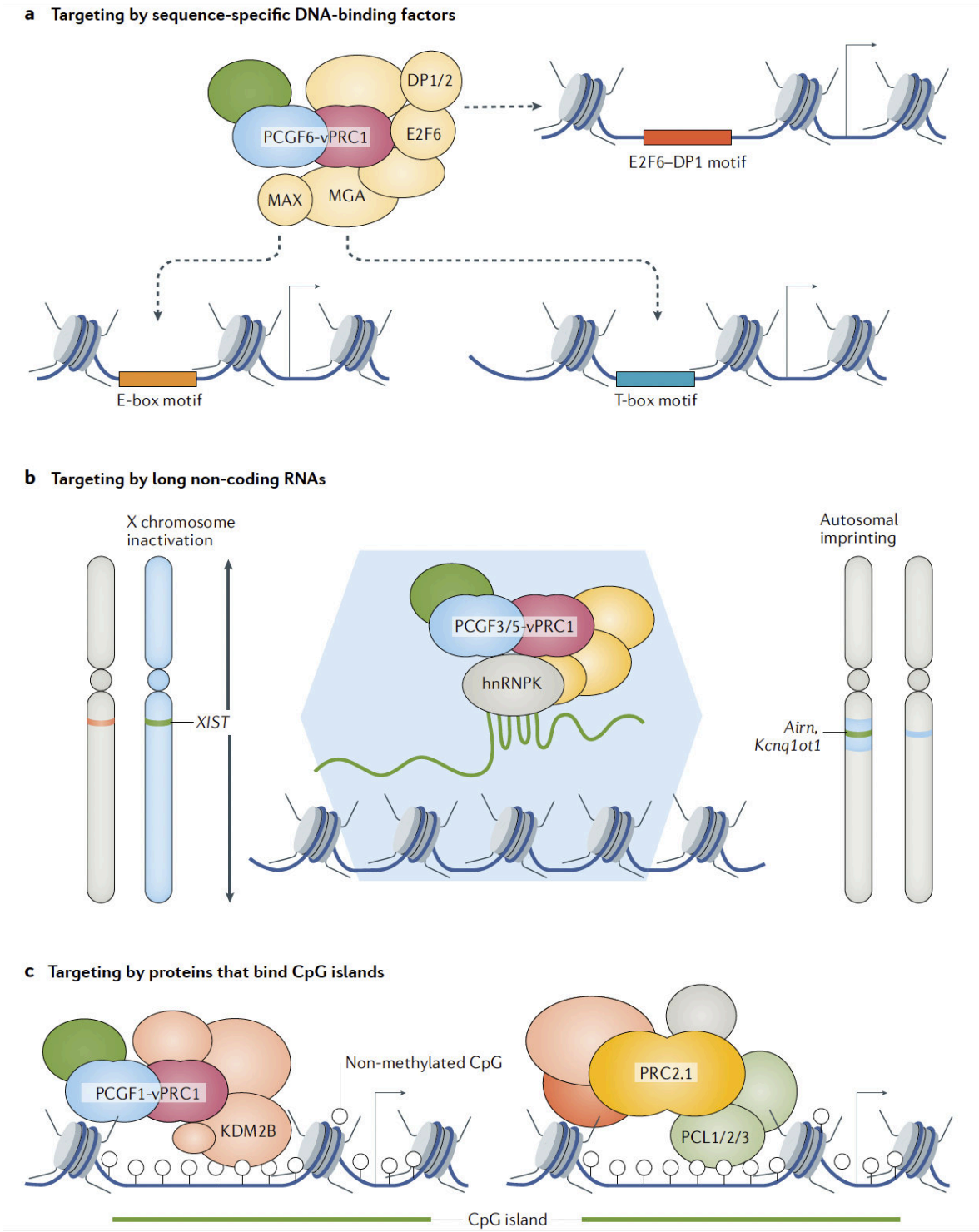


**Figure 1.7:** Recruitment of PcG proteins to chromatin. Adapted from [61].

### 1.3.4 Establishment and maintenance of PcG domains

PcG proteins are able to create very large PcG domains to regulate transcription. This is accomplished by the cooperation and establishment of positive feedback loops between PRC1 and PRC2, which allows the deposition of high levels of H3K27me3 and H2AK119Ub1 and also a high occupancy of PcG complexes on chromatin, which permits the establishment of repressive PcG domains [61].

#### 1.3.4.1 Canonical hierarchical model

The first model that explained the establishment of PcG domains is also known as the *Canonical* model, which is based on the activity of PRC2. As explained in sections 1.3.3.3 and 1.3.2, through its PCL subunits PRC2 can directly bind unmethylated CGIs and methylate H3K27. H3K27me3 is then recognized by the PRC2 subunit EED through a WD40-repeat domain, and this interaction also stimulates the catalytic activity of the complex. Also, when EED interacts with H3K27me3 the catalytic SET domain of EZH1/2 is placed exactly on the H3K27 of the next nucleosome, allowing an efficient spreading anf formation of H3K27me3 domains [73]. These H3K27me3 domains will then tether cPRC1 to chromatin, which interacts with them through the chromodomain of its CBX subunit. Interestingly, this recruitment of cPRC1 by PRC2 does not seem to play an important role in increasing H2AK119Ub1 levels, probably due to the low catalytic activity of cPRC1 [74]. However, the recruitment of cPRC1 seems to be important to establish high-order chromatin interactions and compaction [54]. cPRC1 complexes can polymerize even when located several megabases away through the SAM domains of its PHC subunits [54], allowing the formation of chromatin condensates with very high PcG occupancy called *Polycomb bodies* [75].

#### 1.3.4.2 Variant hierarchical model

Following a similar mechanism as in the canonical model, where PRC2.1 deposits H3K27me3 and then cPRC1 is recruited, vPRC1 can deposit H2AK119Ub1 (independently of PRC2) that can be recognized by PRC2.2 [74]. Very similarly to the role of EED in stimulating PRC2 binding to H3K27me3, the vPRC1 subunit RYBP/YAF2 is able to recognize H2AK119Ub1 through its zinc-finger domain allowing the propagation of H2AK119Ub1 to near nucleosomes [76]. Once chromatin is decorated with H2AK119Ub1, PRC2.2 can recognize it thanks to its subunits JARID2 and AEBP2 which contain a ubiquitin binding motif and tandem C2H2 zinc-fingers, respectively [63, 64]. Also, the interaction of JARID2 and AEBP2 with H2AK119Ub1 reduces the inhibitory effect of H3K4me3 and H3K36me3 on PRC2 [64]. Contrary to the canonical model, which was lacking high enzymatic PRC1 activity, this mechanism is more efficient at depositing high levels of H3K27me3 and H2AK119Ub1 and establishing PcG chromatin domains.

The variant and canonical models do not have to be interpreted as two independent mechanisms of PcG domain formation, but as a whole. The combination of both is what allows the formation of large PcG domains with very high occupancy of PRC2.1, PRC2.2, vPRC1 and cPRC1, which allows the establishment of very stable and reliable repressive chromatin domains.

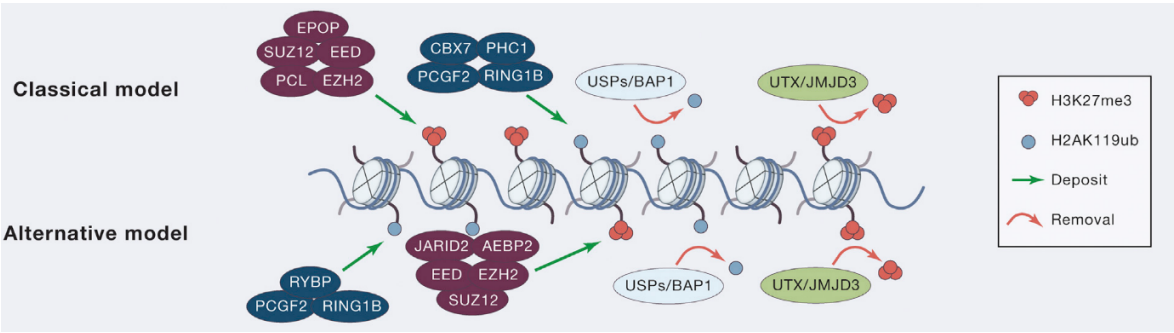


**Figure 1.8:** Canonical and Variant establishment of PcG domains. Adapted from [50].

#### 1.3.4.3 Fighting for chromatin control

PcG complexes bind most CGI-containing promoters, however, not all of them present high levels of PcG occupancy, H2AK119Ub1, H3K27me3 and transcriptional repression. This suggests that probably PcG proteins have to compete with other biochemical systems for the control of chromatin in order to establish stable and repressive PcG domains. This is also supported by the fact that many PcG domains seem to be cell-type specific and to change during development/differentiation, which means that they should depend on the activity of other nuclear proteins such as TFs and the transcription machinery [61].

A mechanism that explains how just a subset of potential PcG targets show very high occupancy of PcG proteins is the transcriptional activity of target genes. Genomic regions decorated by PcG domains (high levels of H2AK119Ub1 and H3K27me3) show very low transcriptional activity, which can be interpreted as a *consequence* of PcG binding, but could also represent the *cause* of PcG binding. Studies have demonstrated that a reduction in transcriptional activity with inhibitors can stabilize PRC2 and H3K27me3 [77], meanwhile tethering experiments have shown that PcG can establish *de novo* domains at CGIs just if they are depleted of TF activity [78, 79]. At the same time, PRC2 seems to be able to bind promiscuously to nascent RNAs, who would displace PRC2 from chromatin in very active transcriptional environments [80, 81]. And finally, H3K4me3 and H3K36me3, two histone PTMs that are deposited during transcription [41], seem to reduce the catalytic activity of PRC2. Summing up, there are many evidences that suggest that TF activity and active transcription counteract PcG binding, and would be their absence what would allow the establishment of PcG domains at CGI-containing loci.

Based on this idea, the role of PcG proteins would not be to induce *de novo* repressive domains to shut down transcription at target genes, but to keep ”safe” and ”defend” loci that do not have to be active in specific cellular contexts by fighting against aberrant activation of gene expression.

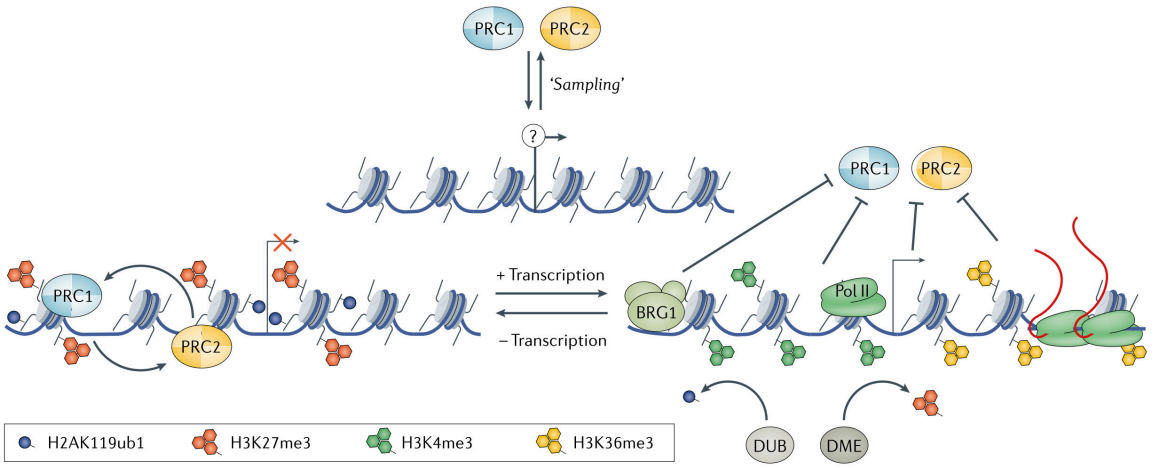


**Figure 1.9:** Maintenance of PcG domains on chromatin. Adapted from [61].

### 1.3.5 Role of PRC1 and PRC2 in gene repression

70 years after the discovery of the first PcG protein still is not very clear how these complexes are able to repress gene expression. Many advances have arrived in the last 20 years, and although we are closer than before in understanding how repression is established by PcG proteins, there is still a lot of research to do. There has been a lot of debate about who is responsible for the PcG-mediated gene repression. H2K119Ub1? H3K27me3? Chromatin compaction by PRC2-cPRC1? There is a lot of literature about this topic that could seem somewhat contradictory at first glance, which will be now discussed.

At first it was thought that PcG mediated repression was accomplished primarily by compacting chromatin via PRC2 and cPRC1 [53, 75, 82, 83]. In line with this, other groups claimed that PRC1-mediated repression could be independent of H2AK119Ub1

[54, 84] and that a catalytic null version of PRC1 delayed embryonic lethality of PRC1KO in mouse embryonic development [85]. This idea was challenged by the fact that removing PRC2 (and in consequence, also cPRC1) from mESC, which was in theory mediating gene repression through compaction, did not cause a strong gene derepression [86]. Also, other studies showed that there was not a clear relationship between PcG activity and increased chromatin compaction [87, 88] Also, removal of vPRC1 complexes in mice had a more severe phenotype than removing cPRC1 [89, 90].

This amount of "contradictory" data highlights the possibility that there is not a specific mechanism by which PcG proteins establish gene repression but, that depending on the cellular context, some mechanisms could play more important roles than others. In our lab we have been very interested during the recent years in understanding better the role of individual PRC1 subcomplexes in mediating PRC1 function, in particular the roles of vPRC1 and H2AK119Ub1 in establishing gene repression, that as discussed above, is still not very clear. Later on in this thesis our data regarding this topic will be shown and deeply discussed.

## 1.4 Intestinal homeostasis and PcG proteins

During my PhD I have invested most of my time studying the role of PRC1 in mESC, however, in the recent years we have also been very interested in the role of PcG proteins in adult stem cells, particularly in the intestine and hair follicles [91, 92, 93]. Since some of the data presented in this thesis will be about the role of PRC1 in mouse intestinal tissue, some pages of this introduction will be dedicated to introduce the intestinal tissue.

### 1.4.1 The intestinal epithelium

In humans, the intestinal epithelium is the second largest tissue by area after the lungs. Its main functions, absorption of metabolites and protection against external insults, have to be performed in a high-stress environment (pH variation, mechanical damage, presence of many different bacteria). In this context, the intestinal tissue is able to perform these functions thanks to a high proliferating rate and a spatial organization called crypt-villus structure [94].

The crypt-villus structure is a very smart way of compartmentalizing the intestinal tissue in two different domains that allows an efficient nutrient uptake (which takes place at the villus) and a maintenance of a high proliferating cell rate (which take place at the crypt). This division is very important due to the nature of the intestinal environment [95]. Cells in the intestine are exposed to high mechanical stress and environmental hazards, reason for which there is a very high cell turnover (lifetime of a cell averages between 3 and 5 days [96]). Proliferating stem cells are safely located at the crypts, which are invaginations of the intestinal tissue that keep them isolated from environmental hazards. On the other side, the villus is a protuberance of the intestinal wall that allows to maximize the amount of intestinal tissue in contact with nutrients. This allows for an efficient metabolite absorption at the expenses of being more exposed to hazards of the digestive tract [94].

The main populations of cells forming the intestine, which will be presented in the following lines, can be divided in two main groups based on their function: absorptive cells and secretory cells. In the absorptive category we find the enterocytes, meanwhile the more diverse secretory group consists in paneth, goblet, enteroendocrine and tuft cells. The production of new intestinal cells has to maintain correct balance between cell populations in order to ensure a correct intestinal homeostasis.

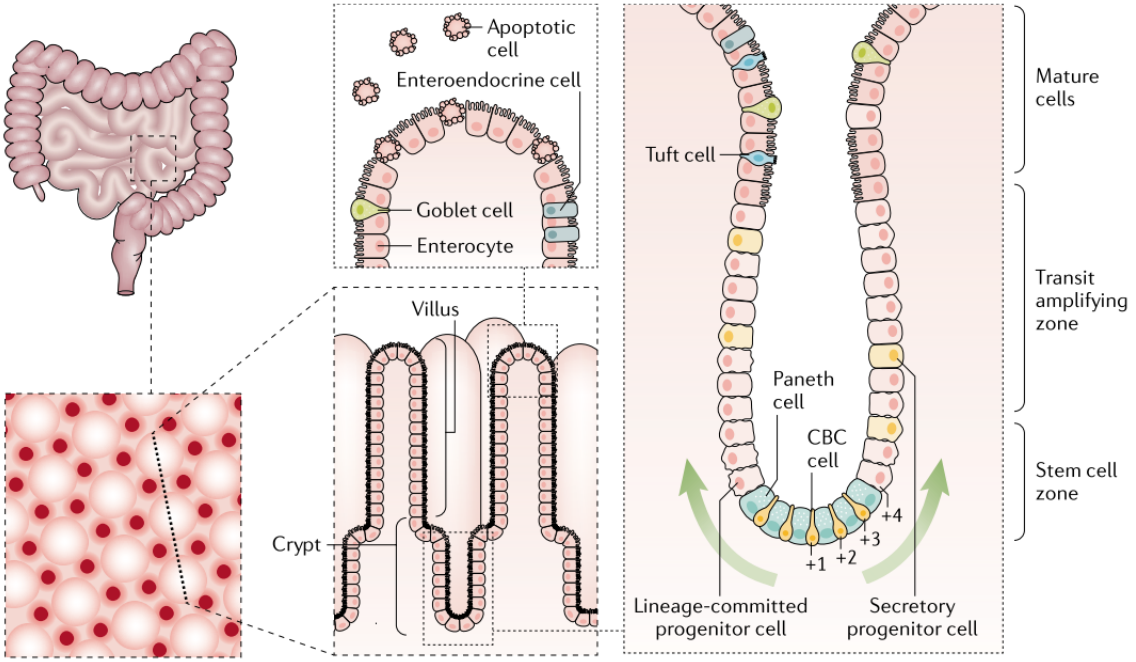


**Figure 1.10:** Structure of the intestinal epithelium. Adapted from [94].

## 1.4.2 Cell populations of the small intestine

### 1.4.2.1 Crypt base columnar cells

A pool of highly proliferating stem cells at the base of intestinal crypts, termed crypt base columnar cells (CBCs), keeps a constant flow of cells across the crypt-villus axis. These cells are characterized by the expression of the LGR5 protein, which is a target of the WNT signaling pathway [97]. Tracing experiments with reporter mice targeting CBCs ($Lgr5^{EGFP-IRES-CreERT2-R26R-lacZ}$) have demonstrated that after 5 days of tamoxifen injection, blue ribbons of $LacZ^{+}$ cells can be dectected from the bottom of the crypts to the villi, marking all cell types present in the intestine. This was the experiment that demonstrated that CBCs had the capacity to generate multiple cell lineages and to maintain a long-term self renewal potential. These features are what define CBCs as intestinal stem cells [97].

When CBCs asymmetrically divide they give rise to new progenitor cells that start to push their neighbour cells towards the tip of the villi. These cells, named

transit-amplifying cells, proliferate rapidly and move towards the villi in order to become a mature intestinal cell. The first fate decision they will have to make is between becoming secretory or absorptive cells, which will depend on their exposure to Notch signals.

#### 1.4.2.2 Absorptive cells

Absorptive cells consist in enterocytes and M cells. Enterocytes represent around 80% of the total intestinal cells, meanwhile M cells are very rare. M cells are located at lymphoid follicles in the intestine, where high concentrations of T an B cells can be found [98]. Enterocytes are responsible for the uptake of nutrients [99] and their fate is determined by the activity of Notch signaling. If transit-amplifying cells are in contact with DLL+ cells (Paneth cells) high Notch levels will be maintained and they will become enterocyes [100].

#### 1.4.2.3 Secretory cells

Contrary to absorptive cells, transit-amplifying cells need low levels of Notch to start their journey towards the secretory lineage. The lack of Notch signalling allows the expression of ATOH1, a transcription factor that drives transit-amplifying cells towards the secretory fate [100]. The main secretory lineages are:

- **Paneth cells**: Paneth cells are particular secretory cells because they stay at the bottom of the crypt instead of moving towards the villus as all the other cell populations of the intestine [101]. Their main role is antimicrobial (production of defensines and lysozyme) and to sustain the stem cell niche by providing Notch ligands [102].

- **Goblet cells**: They are the most abundant secretory cell population, and their role is to protect the intestinal tissue by producing a protective mucus layer [103].
- **Enteroendocrine cells**: These cells participate in the regulation of the metabolism by secreting hormones into the bloodstream upon activation [104]. The main driver of enterocyte specification is NEUROG3, which deletion abolishes all enterocyte cells from the intestine [105].
- **Tuft cells**: They are to the most rare secretory cell population of the intestine. They seem to be important against helimnth infections by regulating the immune response by secreting IL-25 and stimulaing ILC2s [106, 107]. A very particular characteristic of this cell population is that its specification is independent of ATOH1, or at least, there is an independent ATOH1 pathway (which depends on the TF POUF2F3) [108, 109].

### 1.4.3 Signaling pathways in the small intestine

Intestinal homeostasis and cell fate decisions depend on the equilibrium of different signaling pathways. The most relevant ones in the small intestine are WNT, NOTCH, EGF and BMP.

#### 1.4.3.1 WNT

The WNT pathway is primarily involved in the maintenance of the stem cell niche. There is a decreasing gradient of WNT from the bottom of the crypts towards the villi, which creates a perfect flux of differentiation signaling that allows the formation of the cypt-villus structure. The main downstream effector of the WNT pathway is CTNNB1, which becomes stabilized when WNT ligands bind to the Friz-

zled–LRP5–LRP6 receptor complex. This blocks the degradation of CTNNB1 by the APC destruction complex, and allows the accumulation of high CTNNB1 levels in the cytoplasm that will then move towards the nucleus and enhance the expression of its target genes [110]. Deletion of *Tcf4* (main effector of canonical WNT signaling) in adult mice induces the loss of the stem cell population [111]. WNT ligands are primarily produced by paneth cells in the form of WNT3, which is transmitted to the stem cells that are next to the paneth cells in the form of cell-to-cell contacts [102].

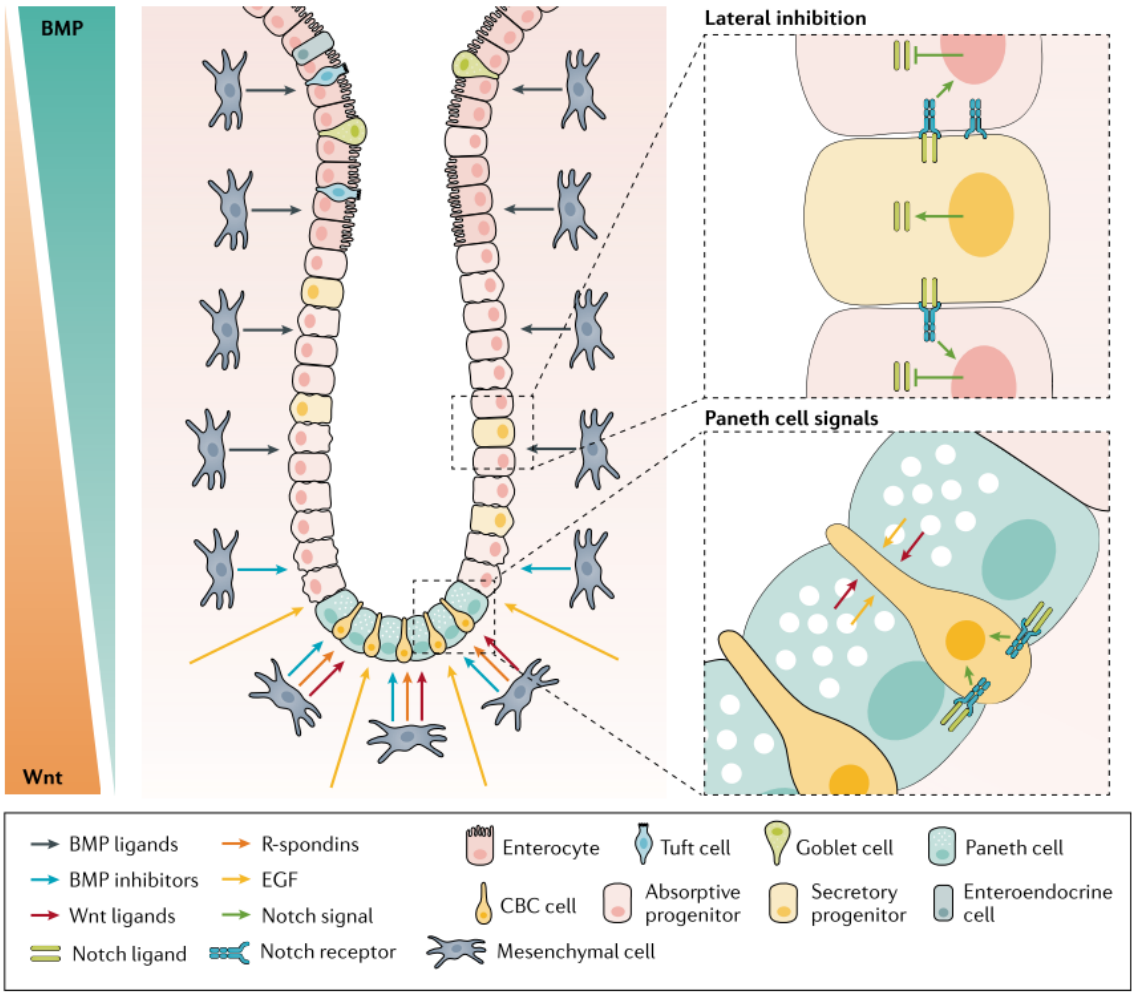


**Figure 1.11:** Regulatory signals of the small intestinal epithelium. Adapted from [94].

#### 1.4.3.2 Notch

In the intestine, Notch is the main signaling pathway that determines if a progenitor cell will become secretory or absorptive. The activation of this pathway requires direct contact between cells, one that expressed notch ligands (DLL1 and DLL4, such as paneth cells) and other that expressed the receptor (NOTCH1, as it happens in the stem cells) [112]. The activation of this pathway will activate HES1, which represses ATOH1, the main driver of secretory differentiation. Thus, meanwhile Notch signaling is active, ATOH1 will be repressed and progenitor cells will differentiate towards the absorptive lineage [112].

### 1.4.4 EGF

The EGF pathway is key for the proliferation of stem cells and turnover rate of the small intestine, however, in contrast to WNT and Notch, is no required to maintain stem cell identity [113]. CBCs express the EGF receptor ERBB1, meanwhile paneth cells secrete the ligands (TGF$\alpha$ and EGF) [102].

### 1.4.5 BMP

The decreasing WNT gradient and increasing BMP gradient from the bottom to the top of the crypts is what drives the differentiation path of intestinal cells across the crypt-villus axis. If WNT has a role in potentiating and maintaining the stem cell niche, BMP acts an antagonist promoting cell differentiation [114]. When BMP ligands contact their type II receptors SMAD1, SMAD5 or SMAD8 gets phosphorilated and translocated to the nucleus with SMAD4, activating the expression of their target genes [115]. Stem cells are protected from the BMP ligands by the mesencymal niche in contact with them, which secretes BMP antagonists [116].

### 1.4.6 Role of PcG proteins in the small intestine

PRC1 and PRC2 play important roles in the maintenance of intestinal homeostasis and intestinal regeneration upon injury [91, 117]. In particular, our group demonstrated that PRC1 is critical for CBCs to maintain their cellular identity. Without altering H3K27me3 levels (thus, independent of PRC2), PRC1 and H2AK119Ub1 loss produced a strong upregulation of non-intestinal related genes, which caused CBCs to lose their transcriptional identity and stemness. On top of that, PRC1 removal caused the upregulation of ZIC1 and ZIC2 transcription factors, which interacts with the CTNNB1/TCF complex inhibiting its function. This inhibition of WNT signaling plus the overexpression of non-intestinal TFs caused a loss of stem cells that totally disrupted intestinal homeostasis. Very importantly, these phenotype was independent of the activation of *Cdkn2a* [91]. Contrary to loss of PRC1, loss of PRC2 in the small intestine is compatible with intestinal homeostasis. The reduction in H3K27me3 levels didn't affect the deposition of H2AK119Ub1, meaning that in adult mouse intestinal stem cells PRC1 activity does not heavily depend on PRC2. However, loss of PRC2 negatively affected intestinal architecture and unbalanced the proportion of absorptive and secretory cells. Loss of H3K27me3 levels on *Atoh1* and *Spdef* promoters led to the uprergulation of these TFs, which unbalanced the transit-amplifying cell fate towards goblet cells, causing a goblet hyerplasia. PRC2 and H3K27me3 loss also induced a strong reduction in cell proliferation in the transit-amplifying compartment, caused by an upregulation of CDKN2A. The upregulation of CDKN2A was also the cause for the failure of PRC2-null small intestine to regenerate upon irradiation-induced damage [117].

# 2 Results

The data that will be presented in this thesis is the result of the collaboration with multiple members of my laboratory. In particular, Andrea Scelfo, Karin Ferrari, Simone Tamburri, Elisa Lavarone, Simona Amato and Annachiara del Vecchio have been the main contributors. All the figures and analyses regarding computational techniques -analysis of next-generation sequencing (NGS) data- have been developed and produced by me, meanwhile the figures regarding wet lab techniques (western blot, immunohistochemical analyses and qRT-PCR) as well as the preparation of all the sequencing libraries (RNA-seq and ChIP-seq) have been performed by these other members of our laboratory. I would like to thank all my colleagues for allowing me to participate in these amazing projects and involving me in all the scientific discussions that allowed us to move forward and generate and curate all the data that will be presented here, without them this thesis would have not been possible.

## 2.1 Landscape of PCGF proteins in mESC

In this first section of the results chapter, I will present our findings regarding the roles and functions of the different PRC1 sub-complexes in mESC. As discussed during the introduction, none of the PCGFs is able to reproduce the loss of RING1A/B

activity during development, which leaves an open question about which is the specific role of each PRC1 sub-complex, if they are able to compensate each other, and if they are functionally overlapped by regulating related biological pathways.

In order to answer these questions, my laboratory has developed a series of genetically engineered mESC lines that allowed us to specifically delete each of the different PCGFs in those cells (thus, removing specific PRC1 sub-complexes). With these cell lines, coupled with the development of specific custom antibodies to recognize each individual PCGF, we were able to perform a series of experiments that allowed us to better understand which are the specific roles of each PRC1 sub-complex, and how these roles may converge to orchestrate a PRC1 function that, as time goes on, is proving to be way more complex than initially expected.

### 2.1.1 Expression and genomic localization of PCGF proteins

To understand which PCGFs are expressed and which PRC1 sub-complexes are predominant in mESC, we measured their mRNA and protein levels using RNA-seq and mass spectrometry techniques. As shown in Figure 2.1 (left panel), PRC1.6 (PCGF6) and PRC1.2 (PCGF2) were the most abundant PRC1 sub-complexes, followed by PRC1.1 (PCGF1) and PRC1.3 (PCGF3). Instead, PRC1.4 (PCGF4) and PRC1.5 (PCGF5) were barely detected. Although the expression of PCGF1-3-4-5 was very similar in terms of mRNA (Figure 2.1, right panel), the amount of protein interacting with RING1B was significantly lower for PCGF3-4-5 respect to PCGF1, which may suggest that not all PCGFs bind to RING1B with the same affinity. Since PCGF2/4 and PCGF3/5 form biochemically identical complexes, we can confirm that, form the functional perspective, all PRC1 sub-complexesa are present in mESC.

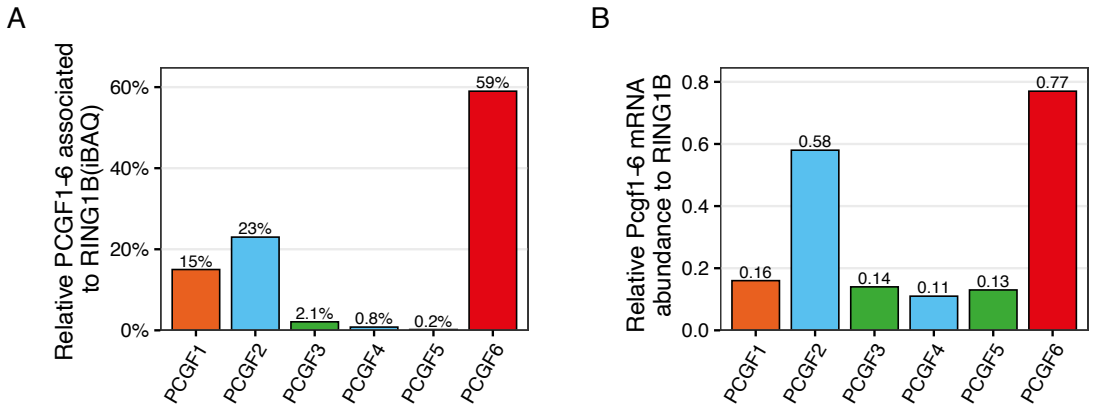


**Figure 2.1: Expression of PCGF1-6 relative to RING1B**. A) Relative abundance based on intensity Based Absolute Quantification (iBAQ) values of PCGF peptides identified in RING1B-Flag-bio tandem purifications coupled to mass-spectrometry experiment. B) Relative mRNA expression of each Pcgf relative to RING1B expression (RPKM).

By using custom antibodies developed in my laboratory, we were able to map the genome-wide localization of the most abundant PCGF proteins (PCGF1-2-3-6) in mESC. The use of specific stable mESC lines where each *Pcgf* was deleted (in the case of *Pcgf2/4* and *Pcgf3/5* both *Pcgfs* were removed to avoid any potential compensatory effects) allowed us to asses the specificity of each antibody, as they showed very little background signal when used on the KO cell lines, as shown in Figure 2.2. Also, any peaks called in the KO cell lines were used to filter any potential false positive found in the WT samples. With this setup, we have the right tools to study the specific activities of PCGF proteins in mESC and to understand how do they behave in the absence of other PRC1 sub-complexes.

Figures 2.2 (panel A) and 2.3 (panel A) show different binding profiles across PRC1 sub-complexes. PCGF1 and PCGF2 presented a more broad binding pattern (closer to the shape of H2AK119Ub1 and H3K27me3 deposition patterns), meanwhile PCGF3 and PCGF6 were narrower and more similar to a TF-like profile. Also, as expected, most PCGF binding sites are located in CpG rich regions, as is the case for RING1B.

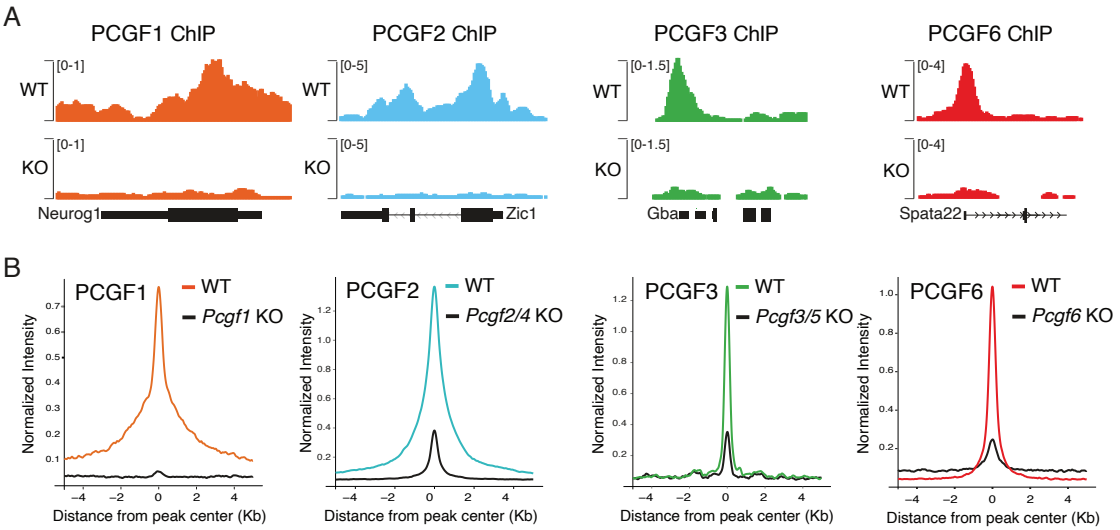


**Figure 2.2: Specificity of PCGF antibodies**. A) Genomic snapshot of the indicated ChIP-seq samples across different PCGF target genes in WT (colored) and KO (black) mESC lines. B) Intensity plots of the indicated ChIP-seq samples across each PCGF target loci in WT (colored) and KO (black) mESC lines.

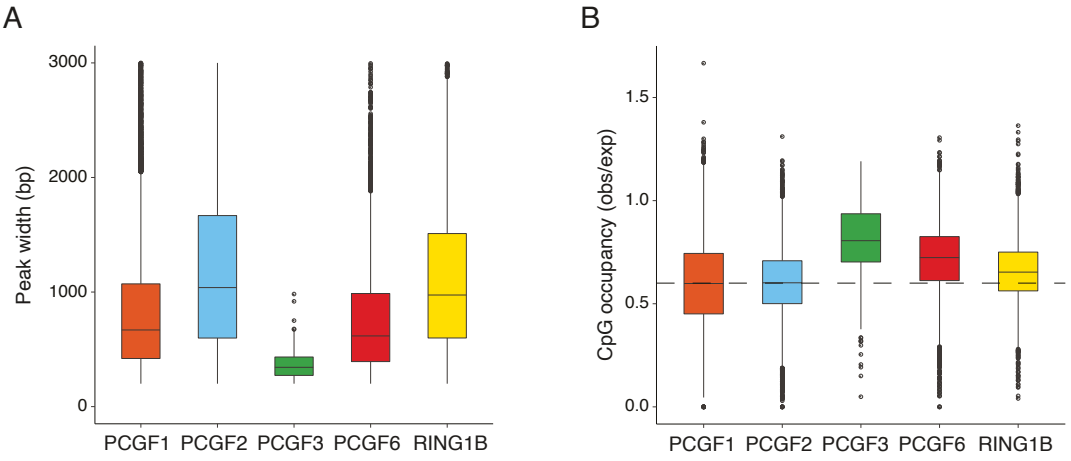


**Figure 2.3: Peak width and CpG occupancy of PCGF binding sites**. A) Peak width (bp) of WT PCGF binding sites. B) CpG occupany of WT PCGF binding sites. The dashed line indicates the threshold of to determine if the ratio of observed vs expected CpG occupancy that defines a CpG island (0.6).

Figure 2.4 shows that more than 75% of the binding sites of all PCGFs were located at promoter regions (+/- 2500bp from TSS), which is in accordance with the distribution of RING1B. From now on, all the analyses regarding the target sites of the PCGFs will be referred to these regions (promoter regions targeted by PCGFs, distal regions won't be included in the analyses). The motivation behind this decision is to perform more reliable analyses. It is already known that PcG proteins bind mainly unmethylatyed CpG-rich promoters [61], and mapping distal regions to target genes is not trivial. At the same time, it is possible to find more than 1 peak for a particular PCGF in the same promoter region. By working with genomic windows obtained with a peak caller, this situation would be counted as 2 binding sites, while by annotating those peaks as TSS instead, would be counted as 1, which to our view, is closer to reality from a functional point of view. Another reason for this decision is the definition of genomic loci bound by multiple PCGFs. By working with genomic windows, there are cases in which a peak from a TF could overlap other 2 TFs, but those 2 TFs not overlap between them. By having a common unit (TSS), which is the same for all the ChIP-seq samples, we also avoid this problem.

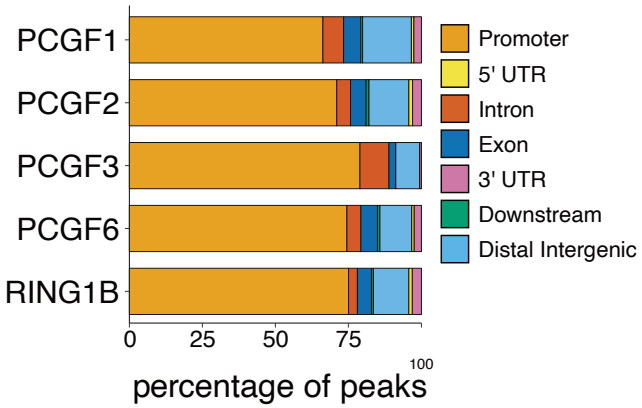


**Figure 2.4: Genome-wide distribution of PRC1 sub-complexes binding sites**. Regions annotated as *downstream* correspond to those that are <10kb form the TES, regions annotated as *distal intergenic* are those that are >10kb from TES and do not overlap with any promoter or gene body.

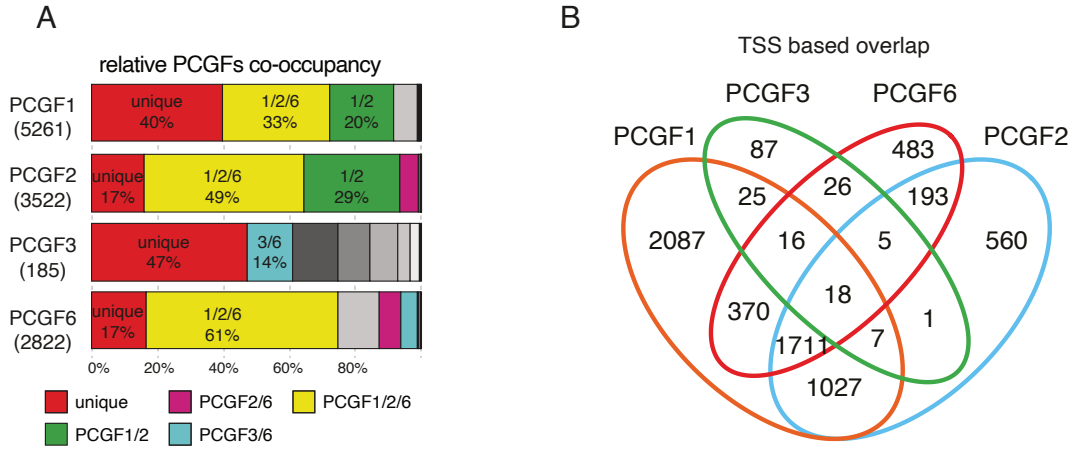


**Figure 2.5: Overlap of genome-wide PCGFs target genes**. A) Barchart showing the % of each PCGF target genes that overlap with the other PCGF proteins. B) VennDiagram's of PCGF1, PCGF2, PCGF3 and PCGF6 target genes.

By looking at the number of target genes in Figure 2.5, we can see that PCGF1, despite not being the most abundant PRC1 complex, showed the highest number of target genes (5261). This was followed by PCGF2 (3522) and PCGF6 (2882), while PCGF3 (185) showed the lowest number of binding sites by more than an order of magnitude. The predominance of PCGF1 regarding the number of binding sites could be explained by the fact that forms complex with KDM2B, which recognizes unmethylated CGIs [59], which would allow for a broader set of potential binding sites respect to PCGF2 and PCGF6, which are more limited to H3K27me3+ regions (PCGF2) [51] and E-BOX and E2F binding sites (PCGF6) [60]. Figure 2.5 shows that PCGFs, despite sharing an important amount of target genes, still retain binding specificity, in particular PCGF1 and PCGF6 (red color). We can also notice that the most frequent combination of PCGFs targeting the same genes are PCGF1-2-6 followed by PCGF1-2. Actually, around 80% of PCGF2 target genes overlap with PCGF1. Based on Figure 2.5, we defined for further analyses the following genomic loci based on the binding of different PCGFs: genes targeted by a single PCGF (red), genes targeted by PCGF1-2-6 (yellow), genes targeted by PCGF1-2 (green) and genes targeted

by PCGF3-6 (blue). The rest of combinations were not included in the analyses due to the reduced number of genes that they represented (less than 14% of the total targets).

### 2.1.2 Association of PCGF proteins with different chromatin domains

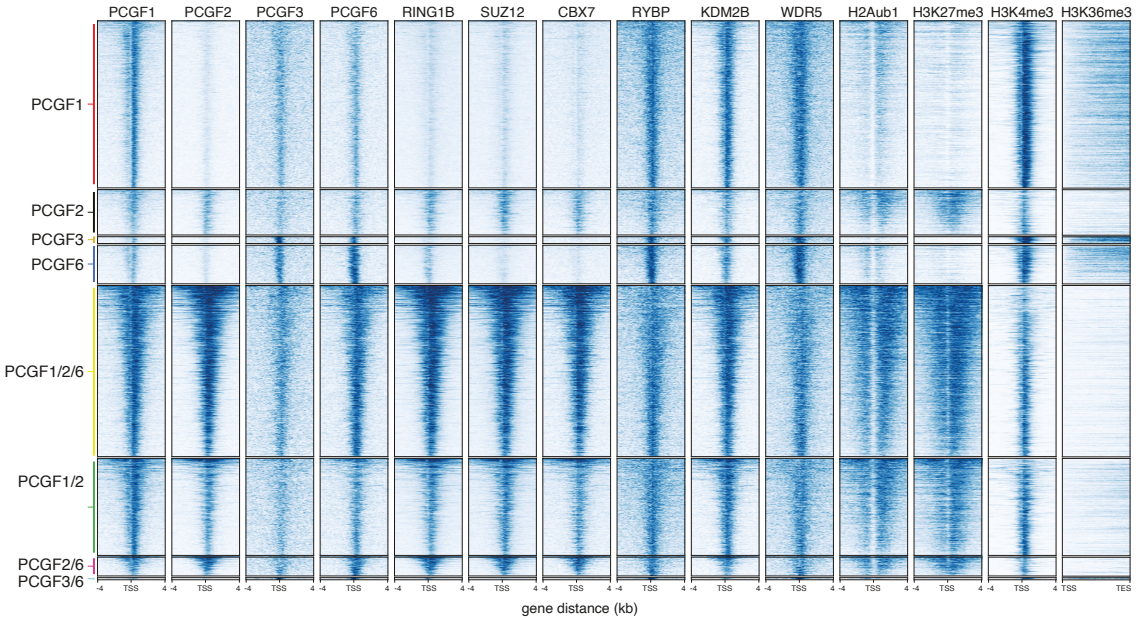


**Figure 2.6: Chromatin properties of PCGF target genes**. Heatmaps representing the normalized ChIP-seq intensities for the indicated PCGF proteins over ±4 kb around the TSS of the indicated loci stratified for PCGF co-occupancy in wild-type (WT) mESCs. H3K36me3 intensity was analyzed over the entire gene length (from TSS to TES). CBX7 and RYBP datasets from mESCs were obtained from Morey et al. (2013) and KDM2B from Farcas et al. (2012).

In order to better understand the functional properties of the different groups of PCGF targets that we defined, we represented in Figure 2.6 the signal of different general components of PcG (SUZ12, RING1B, H2AK119Ub1 and H3K27me3), components of vPRC1 (RYBP, KDM2B, WDR5), cPRC1 (CBX7), the catalytic product of TrxG (H3K4me3) and H3K36me3, which is associated with active transcription.

We noticed that the presence of PCGF2 was always correlated with very high accumulation of RING1B, SUZ12, H3K27me3 and H2AK119Ub1 (Figure 2.7 and 2.6) and low levels of H3K36me3 and mRNA expression (Figure 2.8), which are classic characteristics of PcG domains. In other words, the presence of cPRC1 was always associated to the presence of PcG domains and repressed transcription.

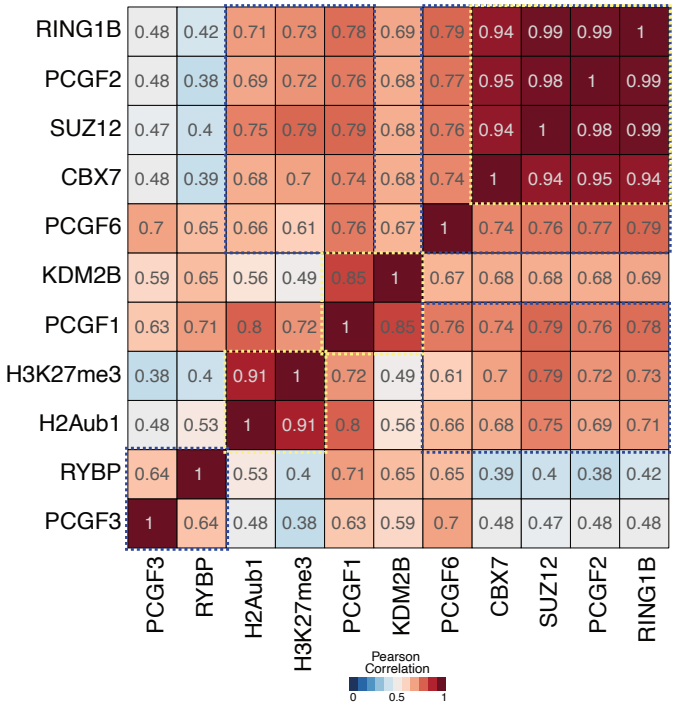


**Figure 2.7: Correlation of PcG proteins and transcriptional associated hPTMs at PCGF target genes.** Pearson correlation of ChIP-seq signal over the promoter regions (±4 kb from TSS) of annotated RefSeq coding genes.

On the other side, the presence of just PCGFs from vPRC1 complexes (PCGF1, PCGF3 and PCGF6) was associated with permissive transcription (Figure 2.8, PCGF1 and PCGF6 targets) or high levels of transcription (Figure 2.8, PCGF3 and PCGF3/6 targets). Also, interestingly, the strongest ChIP-seq signal for RYBP, which is able to recognize H2AK119Ub1 [76], was across PCGF6 unique targets, where H2AK119Ub1 levels were very low.

We also analyzed whether PCGF targets shared functional features or if each PCGF was targeting a specific type of genes. For that, we performed GO enrichment analyses of the different PCGF target categories, shown in Figure 2.9. We no-

ticed that targets from PCGF1 and PCGF2 were very often related to developmental processes and neuron development, which are classical pathways regulated by PcG proteins [50]. On the other side, PCGF3 and PCGF6 targets were enriched in distinct pathways, such as authophagy and germ cell specific processes, respectively. Based on these data, and regarding the functional role of target genes, we could group PCGFs in two modules: PCGF1/2, associated with developmental pathways, and PCGF3/6, associated with more diverse and "niche" biological processes.

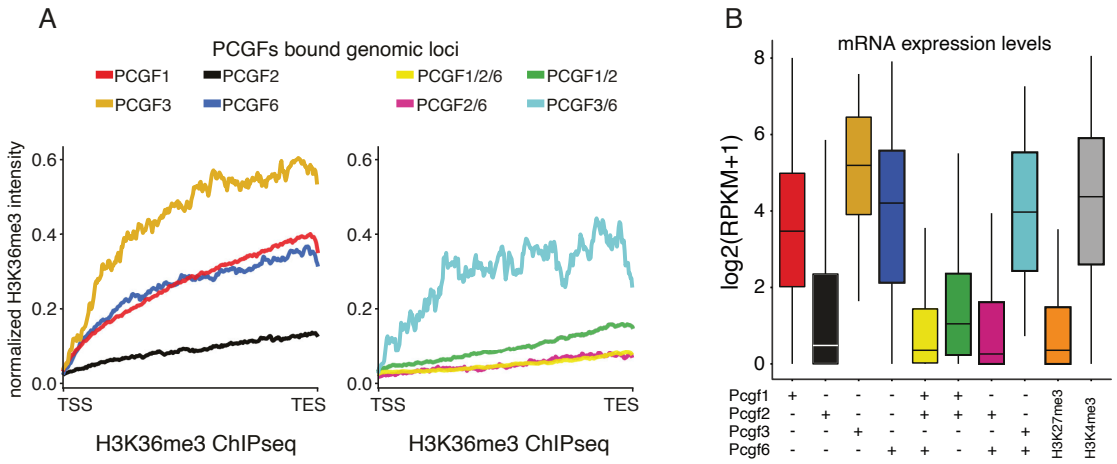


**Figure 2.8: H3K36me3 and mRNA expressoin of PCGF targets**. A) Average deposition profile of H3K36me3 in WT mESCs over the gene body (from TSS to TES) of PCGF unique bound promoters (left panel) or promoters co- occupied by at least two PCGF proteins (right panel), as indicated. B) Boxplots showing the expression levels obtained from RNA-seq analyses performed in WT mESC for the indicated PCGF target genes. H3K27me3- and H3K4me3-positive loci served as controls for repressed and active promoters, respectively.

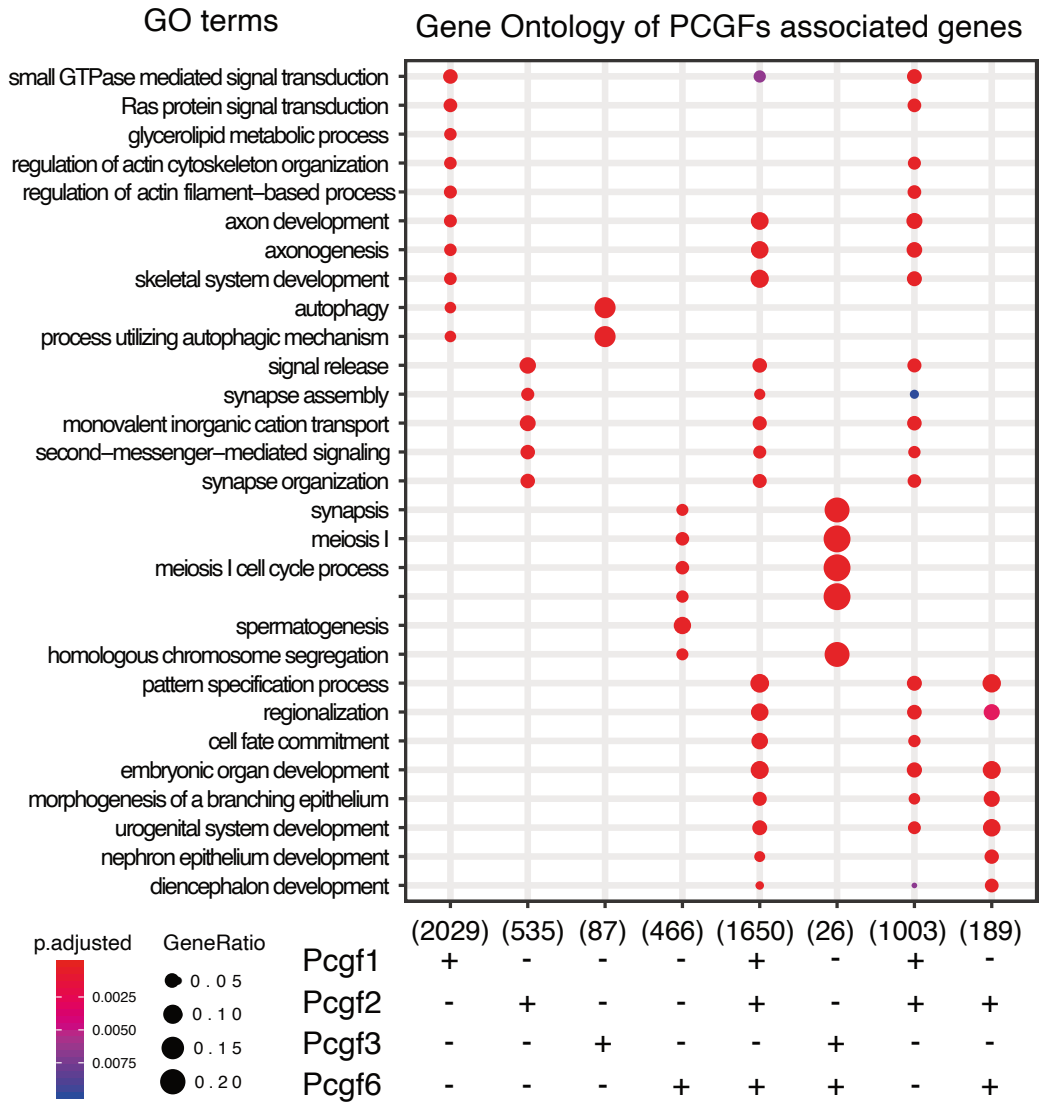


**Figure 2.9: GO enrichment analysis of PCGF target genes**. Gene ontology analysis for the indicated PCGF unique and co-occupied target genes. The most represented categories are highlighted. Dot size is proportional to the number of genes corresponding to that gene ontology category, color scale indicates statistical significance (adjusted p-value < 0.01 and q-value < 0.01).

### 2.1.3 Compensation between PRC1 sub-complexes

We have already seen that, although retaining unique sets of targets, PCGF proteins share a proportion of them, in particular PCGF1, 2 and 6. Based on this, another question that we wanted to answer was whether PRC1 sub-complexes were able to compensate each other activity and act redundantly, or instead, if the functions of RING1B were the sum of individual and non-overlapping PRC1 sub-complex activities. For that, we took advantage of different mESC cell lines where each different *Pcgf* was deleted (WT, *Pcgf1KO*, *Pcgf24KO*, *Pcgf35KO*, and *Pcgf6KO*). We quantified by ChIP-seq signal the binding of PCGF1, PCGF2, PCGF3 and PCGF6 in each of those cell lines, as shown in Figure 2.11.

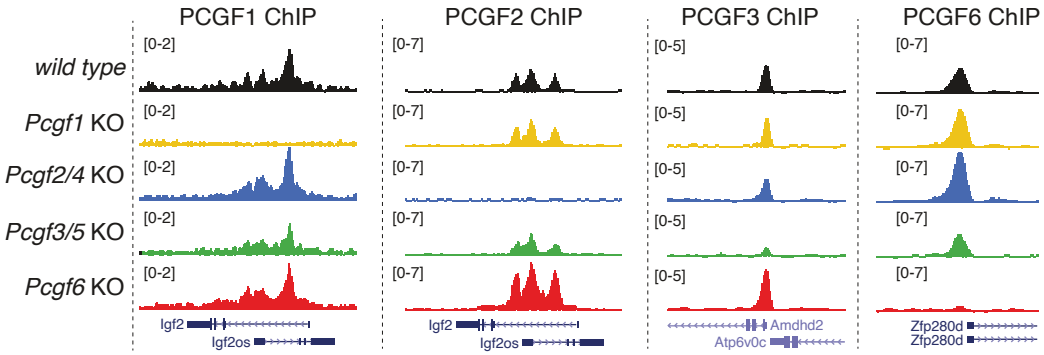


**Figure 2.10: Snapshots of ChIP-seq signal of PCGF proteins in different *Pcgf-null* backgrounds**. Genomic snapshots of the indicated ChIP-seq profiles at selected target gene loci, performed in WT and in the indicated Pcgf KO mESC clones.

Generally speaking, we didn't see a particular compensation across PCGFs in the different KO cell lines. There were some minor changes (although interesting), such as a minor global reduction in PCGF3 signal in *Pcgf1KO* cells and a very slight increase in genome-wide PCGF1 signal in *Pcgf35KO* cells. However, due to the small magnitude of the changes and the semi-quantitative nature of ChIP-seq, the readout that we interpret from these data is that PCGF proteins do not compensate each

other regarding chromatin occupancy (so, when one PCGF is deleted, there is not an increase in the occupancy across its target genes).

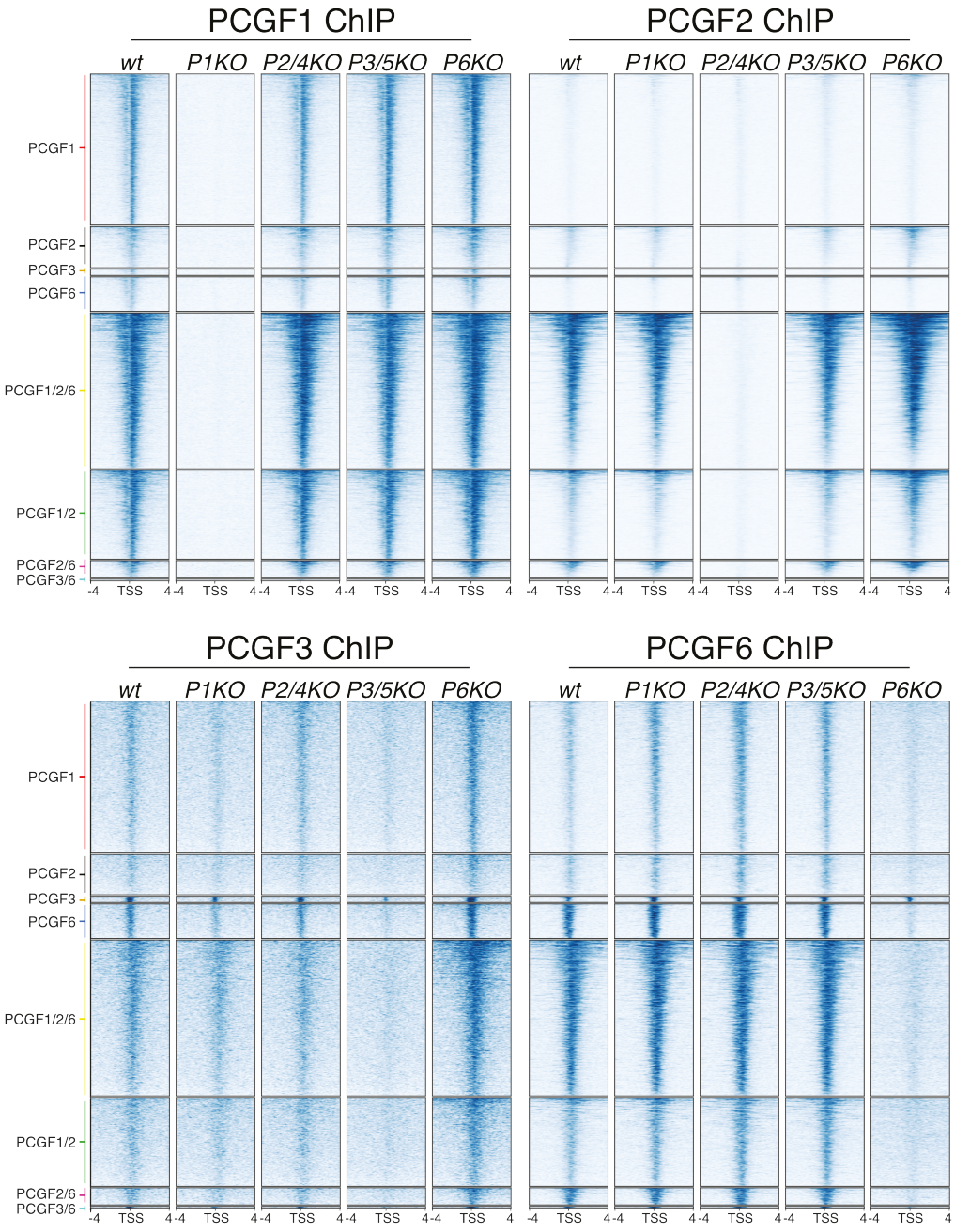


**Figure 2.11: Genome-wide ChIP-seq signal of PCGF proteins in different *Pcgf-null* backgrounds**. Heatmaps of the input subtracted normalized intensity profiles of ChIP-seq analyses for PCGF1, PCGF2, PCGF3, or PCGF6 performed in WT mESCs, Pcgf1, Pcgf2/4, Pcgf3/5, or Pcgf6 KO ESC clones. The regions plotted correspond to ±4 kb around TSS of unique and co-occupied target genes, as indicated.

Next, we proceeded to analyze the chromatin occupancy of core components of PRC1 and PRC2 (RING1B and SUZ12, respectively) and their catalytic products, H2AK119Ub1 and H3K27me3. Figure 2.12 shows that most RING1B signal was lost in *Pcgf24KO* mESC, meanwhile SUZ12, H3K27me3 and H2AK119Ub1 maintained their WT signal across all KOs. Interestingly, the strong reduction in RING1B signal did not translate into a reduction in H2AK119Ub1, meaning that PRC1 sub-complexes do not compensate each other regarding chromatin occupancy, but they do it regarding their catalytic activity. This is also in line with vPRC1 being more active than cPRC1 at depositing H2AK119Ub1 [74]. The strong reduction in RING1B signal probably is due to cPRC1 complexes being more stable than vPRC1 on chromatin, which usually account for 80% of RING1B signal [118]. In *Pcgf6KO* cells, RING1B signal and H2AK119Ub1 were completly lost, meaning that PRC1.6 is the only PRC1 complex that is able to maintain H2AK119Ub1 at those loci, in line with its specific role in regulating specific biological process, such as germ cell related genes [60].

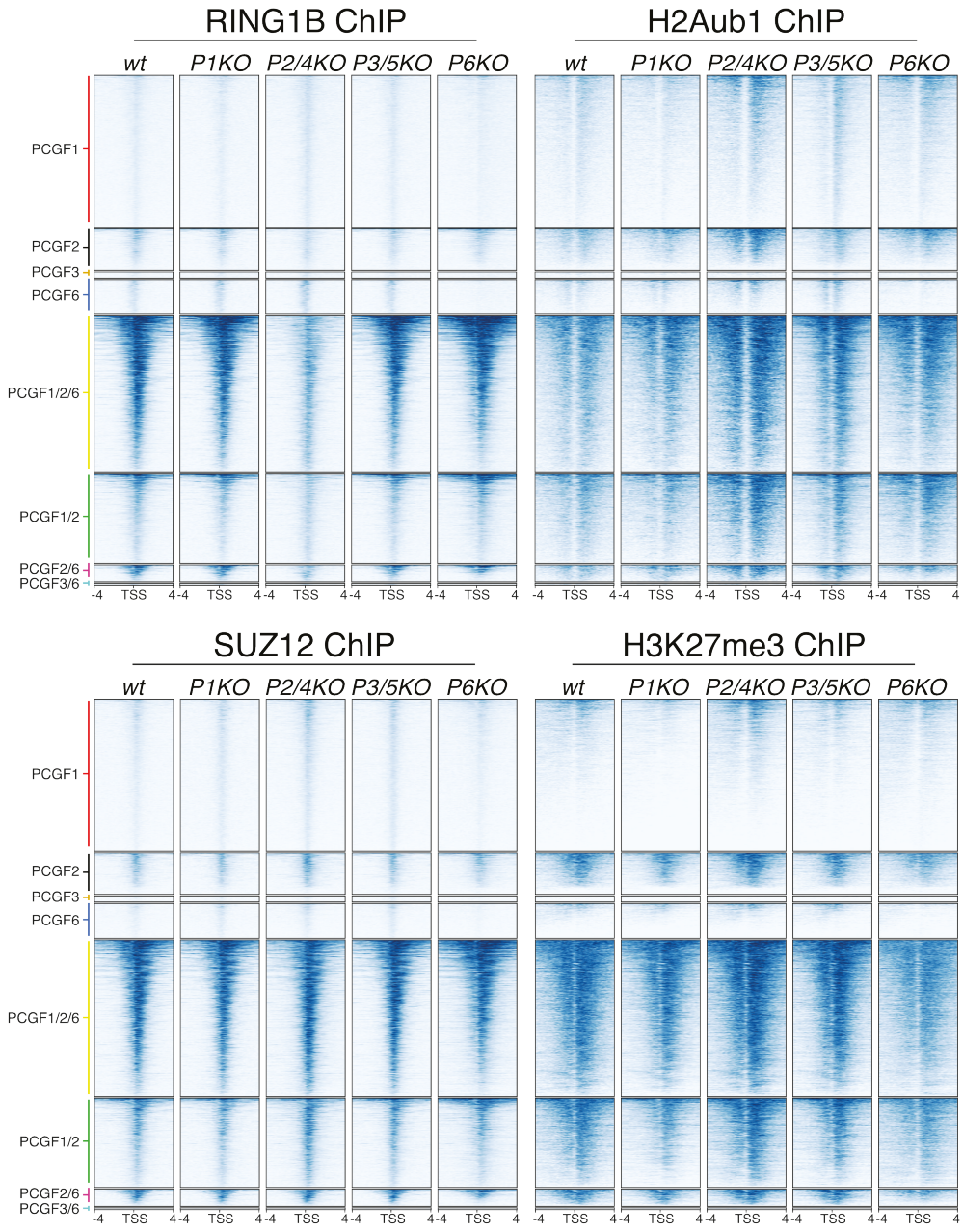


**Figure 2.12: Genome-wide ChIP-seq signal of PcG proteins in different *Pcgf-null* backgrounds**. Heatmaps of the input subtracted normalized intensity profiles of ChIP-seq analyses for RING1B, SUZ12, H3K27me3 or H2AK119Ub1 performed in WT mESCs, Pcgf1, Pcgf2/4, Pcgf3/5, or Pcgf6 KO ESC clones. The regions plotted correspond to ±4 kb around TSS of unique and co-occupied target genes, as indicated.

### 2.1.4 PCGF1/2/4 and PCGF3/5/6 modules

#### 2.1.4.1 Transcriptome of *Pcgf* KO mESCs

Based on their potential functional role (Figure 2.9) and their genome-wide co-occupancy (Figure 2.5), we decided to define two different PCGF modules: PCGF1/2/4 and PCGF3/5/6, for which we developed *Pcgf124*KO and *Pcgf356*KO mESC lines. *Pcgf124*KO cells were viable and didn't present any morphological defects when compared to WT cells (Figure 2.13), however, for *Pcgf356*KO cells we observed a strong effect on the morphology, since they acquired a fibroblast-like shape (Figure 2.13).

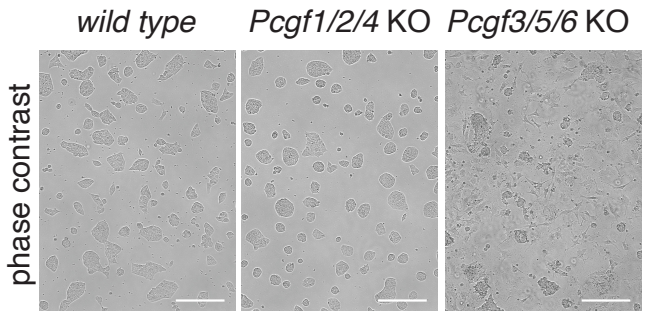


**Figure 2.13: Phase contrast of WT, *Pcgf124*KO and *Pcgf356*KO mESC**. Representative phase-contrast field of WT, *Pcgf124* and *Pcgf356* triple KO mESCs. Scale bars correspond to 200 µm.

We performed bulk RNA-seq for all *Pcgf* KOs (1, 2/4, 3/5, 6, 1/2/4 and 3/5/6) and then applied Principal Component Analysis (PCA) (Figure 2.14). In line with the phase contrast images (Figure 2.13), *Pcgf124*KO cells didn't show big transcriptional differences compared to WT cells, as happened for *Pcgf1*KO and *Pcgf24*KO cells. Instead, *Pcgf356*KO cells suffered stronger changes in their transcriptome (in particular a strong derepression, as seen in Figure 2.15), which were already evident in *Pcgf6*KO cells, in line with previous reports [60]. When looking at the results from GO enrichment (Figure 2.16), we found biological processes in *Pcgf356*KO cells related to extracellular matrix and muscle contraction, which could be associated with

the observed morphological phenotype. In line with this, there were also many up-regulated collagen and keratin related genes (Figure 2.17).

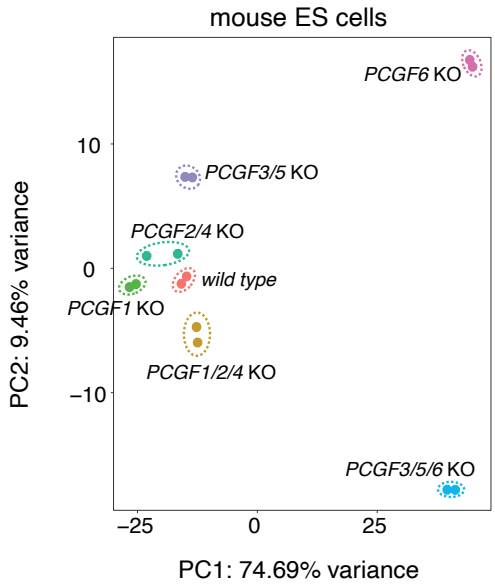


**Figure 2.14: Principal Component Analysis of *Pcgf* KOs transcriptomes**. Principal component analysis of gene expression levels from RNA-seq analysis performed in WT mESCs and in the indicated KOs. Dashed lines enclose experimental replicates.

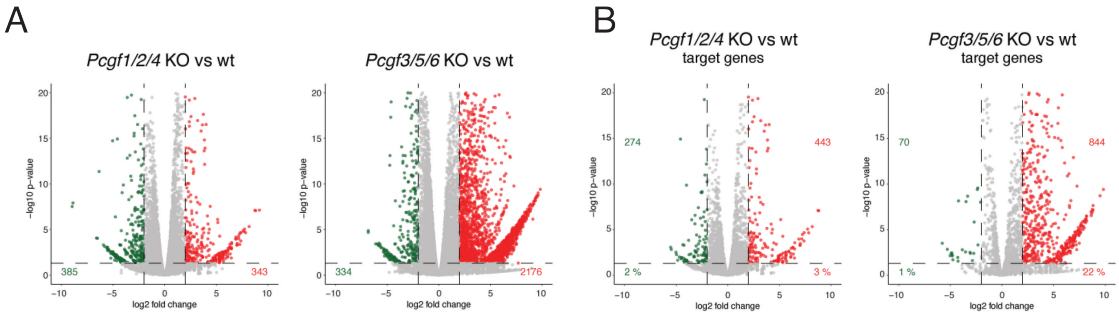


**Figure 2.15: Volcano plots of *Pcgf124* and *Pcgf356*KO mESC**. A) Volcano plots of $-\log_{10}(pvalue)$ against $\log_2(foldchange)$ representing the differences in gene expression between Pcgf1/2/4 and Pcgf3/5/6 KO mESC clones and WT for all protein coding genes. B) Same as A but for PCGF1 and PCGF2 targets or PCGF3 and PCGF6 targets, respectively.

In general there was a very weak correlation between differentially expressed genes and PCGF targets (Figure 2.18). The highest correlation was for PCGF6 targets (around 20% were upregulated) in *Pcgf6*KO cells. This low correlation between differentially expressed genes and PCGF targets could be expected since we have

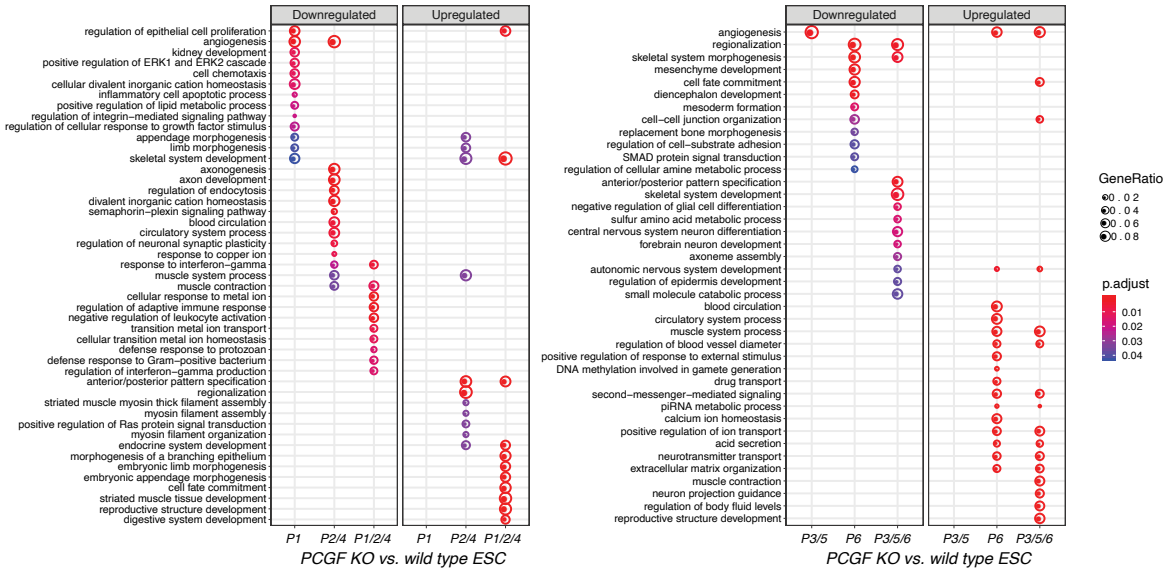


**Figure 2.16: GO enrichment analyses of *Pcgf* KOs**. Gene ontology analysis for the differentially expressed genes in the indicated Pcgf KOs. The most represented categories are highlighted. Dot size is proportional to the number of genes corresponding to that Gene ontology category, color scale indicates statistical significance (adjusted p-value < 0.01 and q-value < 0.01).

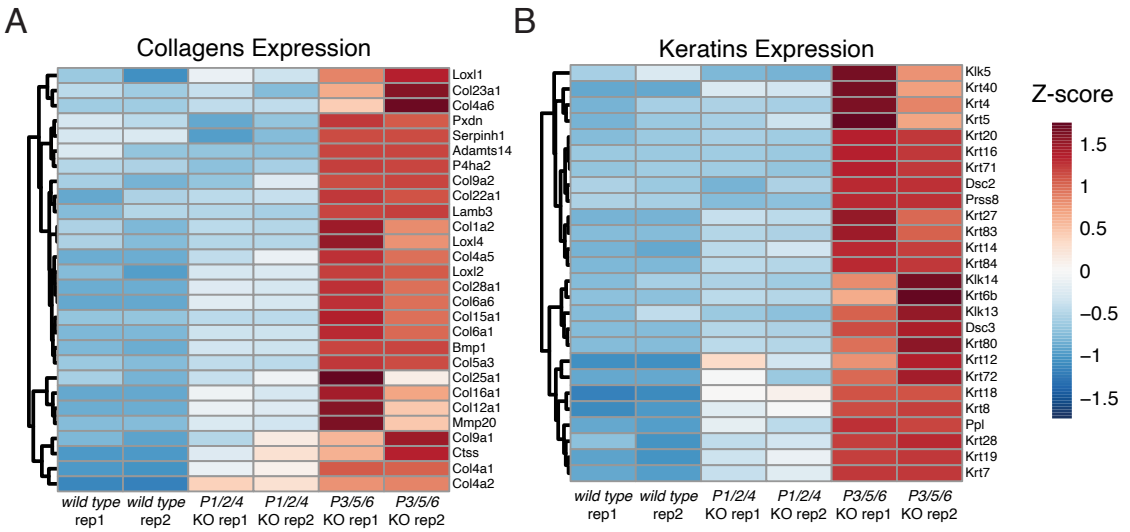


**Figure 2.17: Heatmap of collagen-related in *Pcgf124* and *Pcgf356*KO mESC**. Z-score expression heatmaps for collagen related genes from RNAseq analyses performed in WT, Pcgf1/2/4 or Pcgf3/5/6 triple KO mESCs.

shown that, enzymatically, PCGFs were able to compensate each other (except a for a subset of PCGF6 targets).

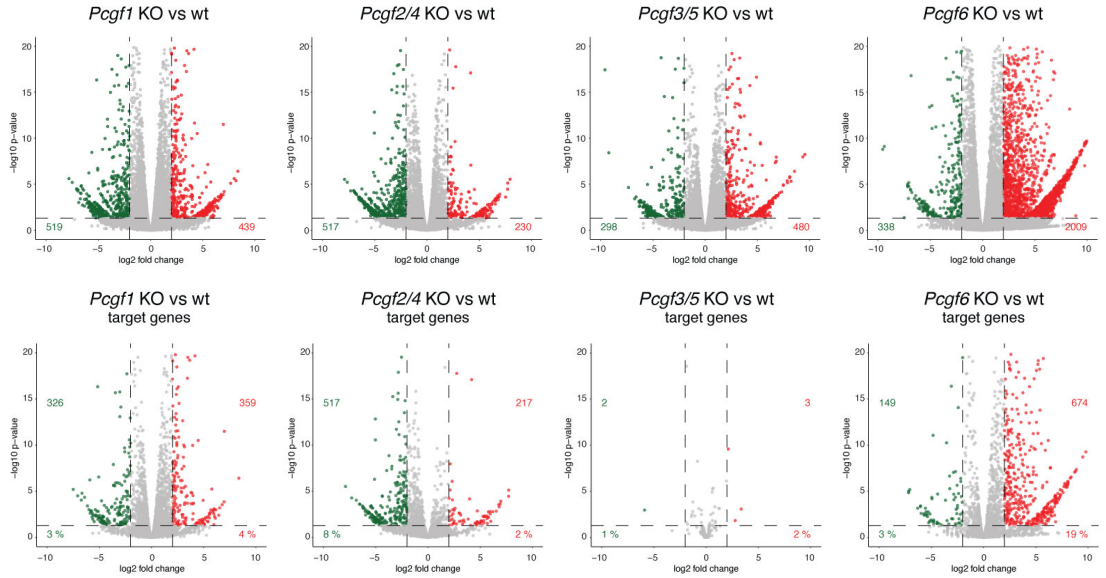


**Figure 2.18: Volcano plots of *Pcgf1, Pcgf24, Pcgf35* and *Pcgf6*KO mESC**. Volcano plots of $-\log_{10}(pvalue)$ against $\log_2(foldchange)$ representing the differences in gene expression between each *Pcgf* KO mESC clone and WT for all protein coding genes (upper panels) and for their corresponding target genes, respectively (bottom panels).

#### 2.1.4.2 PcG genome-wide occupancy in *Pcgf124* and *Pcgf356*KO mESCs

Chromatin occupancy of PRC1 (RING1B) in *Pcgf124*KO across PCGF1/2 and PCGF1/2/6 loci was lost (Figure 2.19). There was a residual RING1B signal at PCGF1/2/6 loci, probably due to the presence of PCGF6. Indeed, RING1B signal was unaltered at PCGF6 unique loci (which was lost upon *Pcgf6*KO, as shown in Figure 2.12). However, since H2AK119Ub1 levels were almost completely lost at those loci, we can conclude that although PCGF6 can have a role in recruiting PRC1 to chromatin at PCGF1/2/6 loci, most of the catalytic activity of PRC1 at those regions depends on PCGF1 and PCGF2, and not on PCGF6. Signal of PRC2 (SUZ12) and H3K27me3 was also reduced in *Pcgf124*KO cells, probably due to the loss of H2AK119Ub1, which is recognized by the ancillary protein JARID2 from PRC2.2 [63]. It is interesting to note that the strong reduction of H2AK119Ub1 did not correlate with gene expression upregulation (Figure 2.15).

RING1B, SUZ12, H2AK119Ub1 and H3K27me3 levels were maintained in *Pcgf356*KO cells, except across PCGF6 and PCGF3 loci. Paradoxically, these cells showed a strong gene upregulation (Figure 2.12), although just 22% of PCGF6 and PCGF3 targets were among upregulated genes. Probably most upregulated genes were indirect effects related to the acquisition of a fibroblast-like identity, however, these data are also in accordance with potential PRC1 roles that are indipendent of H2AK119Ub1 [84, 85].

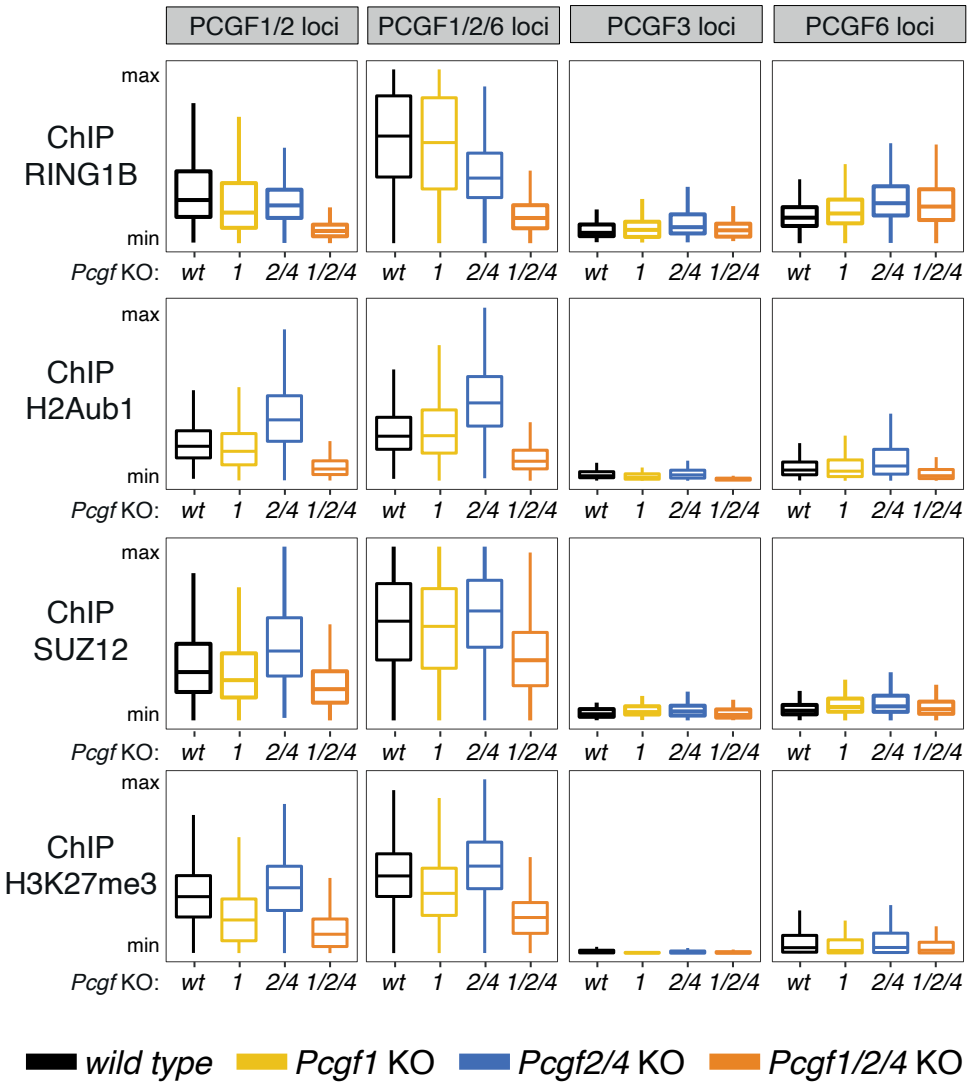


**Figure 2.19: ChIP-seq signal quantification of RING1B, H2AK119Ub1, SUZ12 and H3K27me3 in *Pcgf124*KO mESCs.**. Boxplots of normalized ChIP-seq intensity profiles of RING1B, H2AK119ub1 (H2Aub1), SUZ12, and H3K27me3 performed for WT or Pcgf1, Pcgf2/4, and Pcgf1/2/4 KO mESC clones, over ±500 bp (or ±4 kb for H2Aub1 and H3K27me3) around the TSS of target genes unique for PCGF3 or PCGF6, or common to PCGF1/2 or PCGF1/2/6.

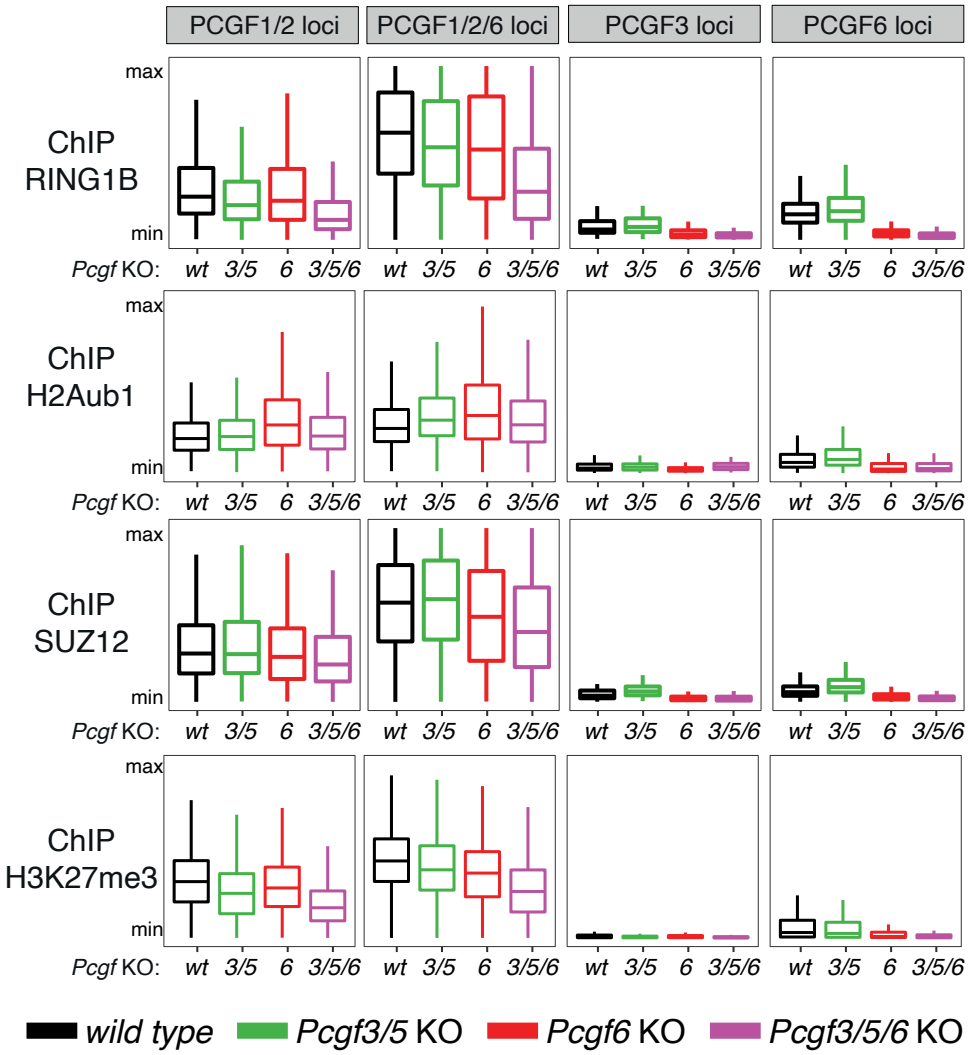


**Figure 2.20: ChIP-seq signal quantification of RING1B, H2AK119Ub1, SUZ12 and H3K27me3 in *Pcgf124*KO mESCs.**. Boxplots of normalized ChIP-seq intensity profiles of RING1B, H2AK119ub1 (H2Aub1), SUZ12, and H3K27me3 performed for WT or Pcgf3/5, Pcgf6, and Pcgf3/5/6 KO mESC clones, over ±500 bp (or ±4 kb for H2Aub1 and H3K27me3) around the TSS of target genes unique for PCGF3 or PCGF6, or common to PCGF1/2 or PCGF1/2/6.

### 2.1.5 PRC1 complex assembly in the absence of RING1A/B

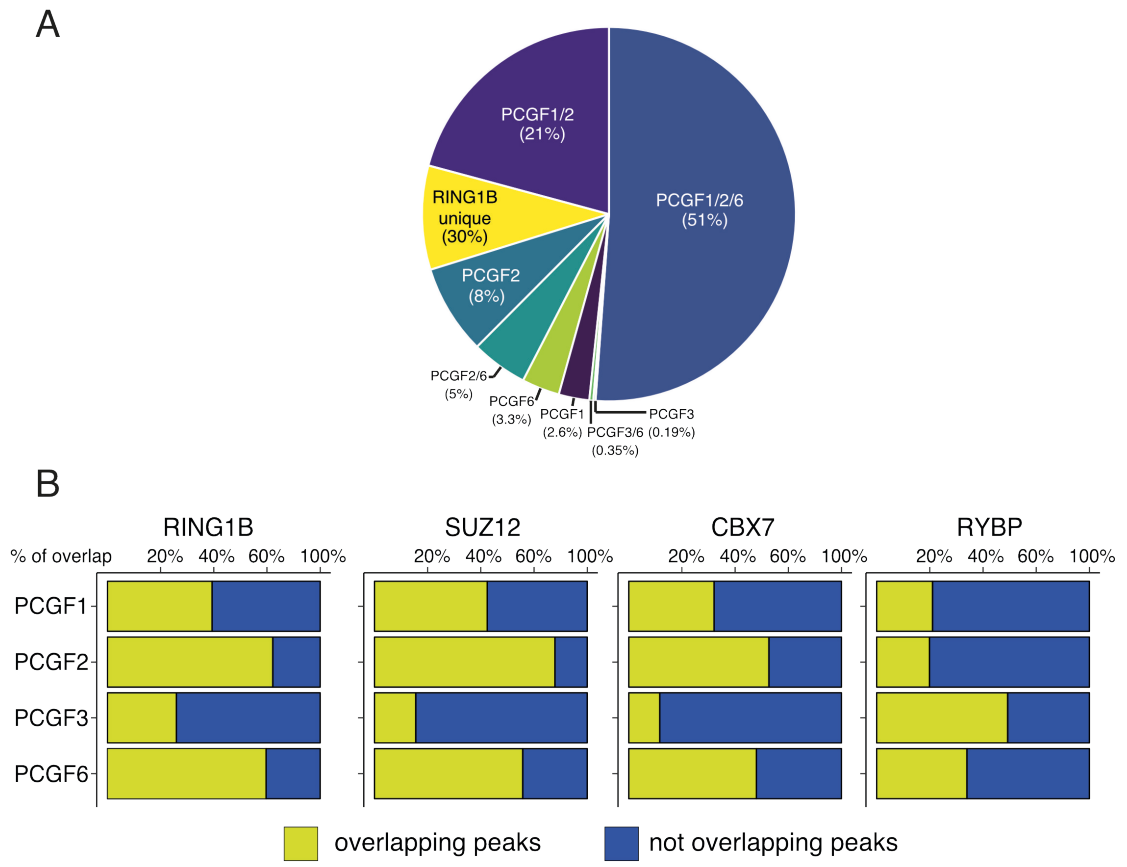


**Figure 2.21: vPRC1 complexes can bind chromatin in the absence of RING1B**. A) Percentage of occupancy of the different PCGF proteins at RING1B-bound promoters. B) Percentage of overlap of RING1B, SUZ12, CBX7, and RYBP at the indicated PCGF-bound promoters.

Overlapping RING1B and PCGF targets showed that 70% of RING1B targets overlap with at least 1 PCGF protein(Figure 2.21A). On the other side, some PCGFs, such as PCGF1 and PCGF3, had more than 50% of their targets not overlapping with RING1B (Figure 2.21B). These regions not overlapping with RING1B were always correlated with low H2AK119Ub1 and H3K27me3 and high levels of histone marks associated with active transcription (H3K4me3, H3K36me3) (Figure 2.23). The absence of RING1B binding and the high transcriptional activity at those sites were strong indicators showing that probably RING1B was indeed not present at those

regions, suggesting that PCGFs could be able to form RING1B-independent complexes that bind chromatin. To test that, we profiled the genome-wide occupancy of PCGF1, PCGF2, PCGF3 and PCGF6 in the absence of RING1A/B. This was achieved by using a Ring1A$^{-/-}$; Ring1B$^{fl/fl}$; Rosa26::CreERT2 conditional mouse ESC line [119], which will be referred as *R1A* KO-*R1B* FL. Our results showed that all PCGFs were able to retain binding at chromatin except PCGF2, which was degraded in *R1A* KO-*R1B* FL cells (Figure 2.22). PCGF1, although was able to still bind chromatin, was significantly displaced, as well as PCGF6 across RING1B+ targets. However, we think that this displacement is primarily due to the strong gene derepression caused by *Ringd*KO and due to the fact that mESC without PRC1 are not viable [119]. So, we conclude that PCGF1, PCGF3 and PCGF6 are able to bind chromatin and from complex in the absence of RING1B.

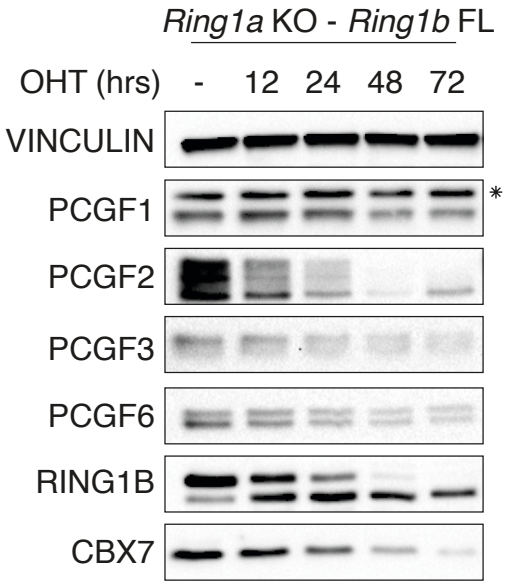


**Figure 2.22: Protein levels of PCGFs in R1A KO-R1B FL.** Western blot analysis with the indicated antibodies of protein lysates prepared from Ring1A–/–;Ring1Bfl/fl;Rosa26::CreERT2 (R1A KO-R1B FL) mESC treated with 4-OHT at the indicated time points. Vinculin served as loading control. *, unspecific signal.

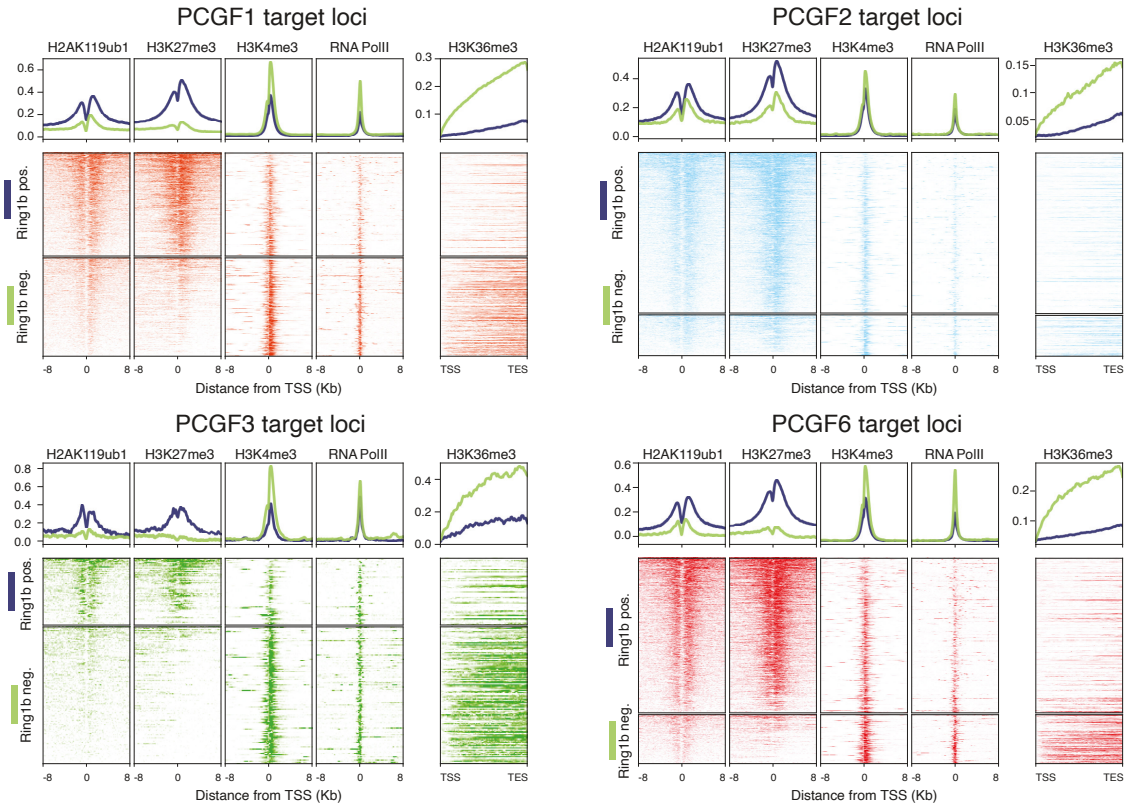


**Figure 2.23: PCGF-RING1B bound loci are associated with transcriptional repression**. Heatmap analysis representing the input subtracted normalized ChIP-seq intensities over ±8 kb centered at TSS of PCGF1 (A), PCGF2 (B), PCGF3 (C), PCGF6 (D) target loci stratified for RING1B co-occupancy in WT mESCs. H3K36me3 intensities are shown along the entire gene length (from TSS to TES).

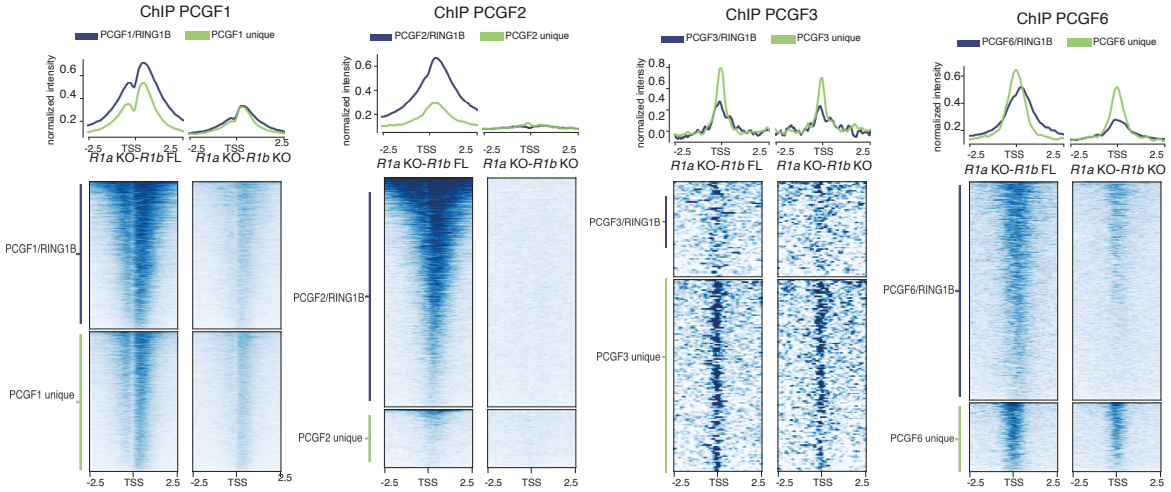


**Figure 2.24: Chromatin features of RIN1B positive and negative PCGF targets**. Heatmap showing the normalized signal of each PCGF ChIP-seq in R1A KO-R1B FL and R1A KO-R1B KO mESC over ±2.5 Kb across their target genes split into RING1B+ and RING1B-.

### 2.1.6 Different DNA-binding activities orchestrate PRC1.6 recruitment to chromatin

Motif enrichment analyses of PRC1.6 targets (PCGF6) revealed that E-box and E2F motifs were present in many PCGF6-bound loci (Figure 2.25A), which is in agreement with the fact that E2F6-DP1 (E2F) and MAX-MGA (E-box) proteins form part of the PRC1.6 complex [51]. Recently it was published that the MAX-MGA heterodimer was required for PRC1.6 recruitment to chromatin, since their downregulation using shRNAs let to a total displacement of PRC1.6 from chromatin [60], in accordance with the fact that almost all PCGF6 targets overlap with MGA (Figure 2.25B). However, the fact that other DNA-binding proteins form part of PRC1.6 and that we were able to find their motifs enriched across PCGF6 targets, made us to hypothesize that, maybe, MAX-MGA were not the only important subunits that mediate PRC1.6 chromatin recruitment. To test that, we quantified the genome-wide occupancy of PCGF6 across different mESC lines: WT, *Pcgf6*KO, sh*Mga*, *Mga*ΔHLH, sh*E2F6* and *Mga*ΔHLH-sh*E2F6* (Figure 2.27). We found, according to [60], that downregulation of MGA let do a displacement of PCGF6 binding similar to *Pcgf6*KO. However, if instead of knocking-down MGA, just the C-terminal HLH DNA-binding domain was removed (*Mga*ΔHLH), then PCGF6 was just slightly displaced from chromatin. This indicates that MGA is not dispensable for PRC1.6 recruitment, but probably, on top of its DNA-binding ability, it works as a scaffold protein that maintains PRC1.6 assembled. Indeed, western blot analyses of PCGF6 immunoprecipitations showed that in absence of RING1A/B (Figure 2.26A) or in *Mga*ΔHLH cells (Figure 2.26B) PCGF6 was able to associate with its ancillary proteins (except RYBP). However, donwregulation of *Mga* dissasembled the complex (Figure 2.26B). sh*E2F6* also produced a displacement of PCGF6 binding, and the combination of *Mga*ΔHLH-sh and sh*E2F6* further

displaced PRC1.6 from chromatin. These data strongly suggests that the recruitment of PRC1.6 is the result of the cooperation between different DNA-binding proteins that form PRC1.6, and not just the MAX-MGA dimer.

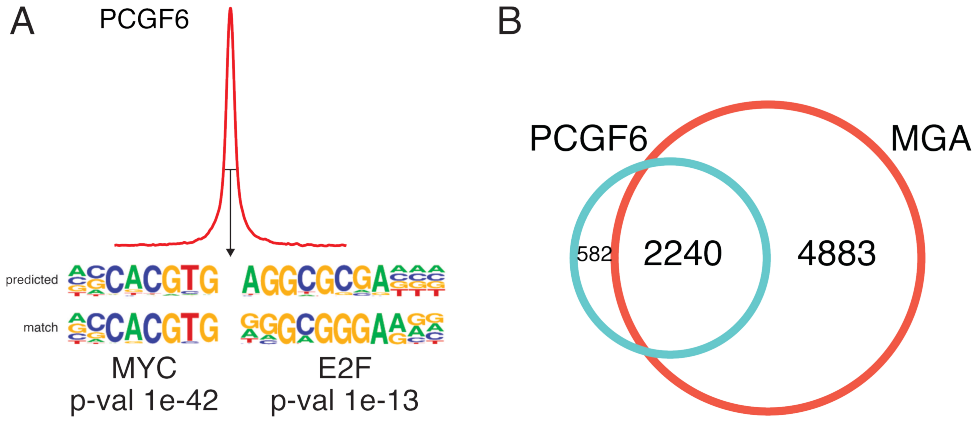


**Figure 2.25: PCGF6 interacts with E-box and E2F motifs**. A) *De novo* motif discovery analysis performed underneath the summit of PCGF6 peaks. Sequence weight matrices of predicted compared to match DNA binding motifs are shown together with relative p-values.B) Venn diagram of PCGF6 (blue) and MGA (red) target genes.

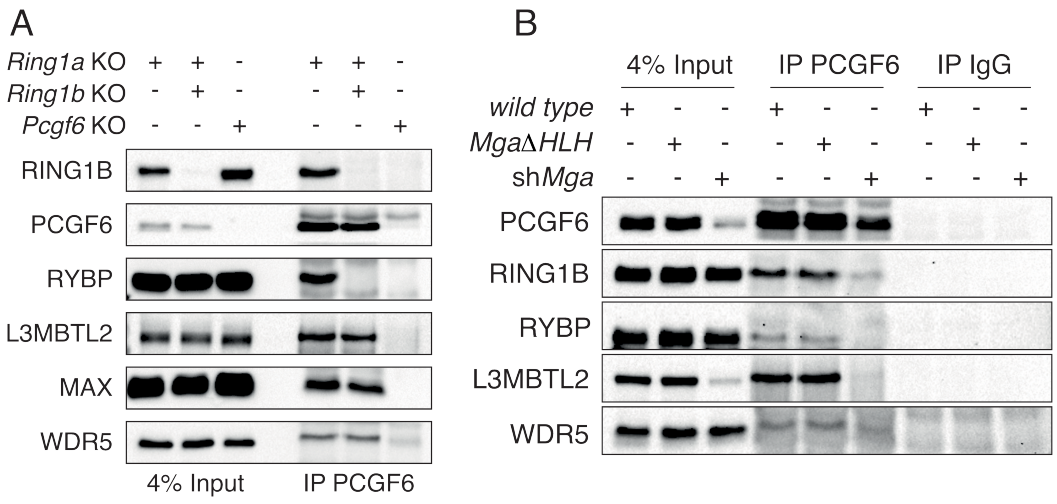


**Figure 2.26: MGA is required for PRC1.6 complex assembly**. A) Western blot analyses using the indicated antibodies in PCGF6 immunoprecipitations from nuclear extracts prepared from the indicated mESC lines. Input served as loading control. B) Same as A) but in *Mga*ΔHLH and sh*Mga* mESCs.

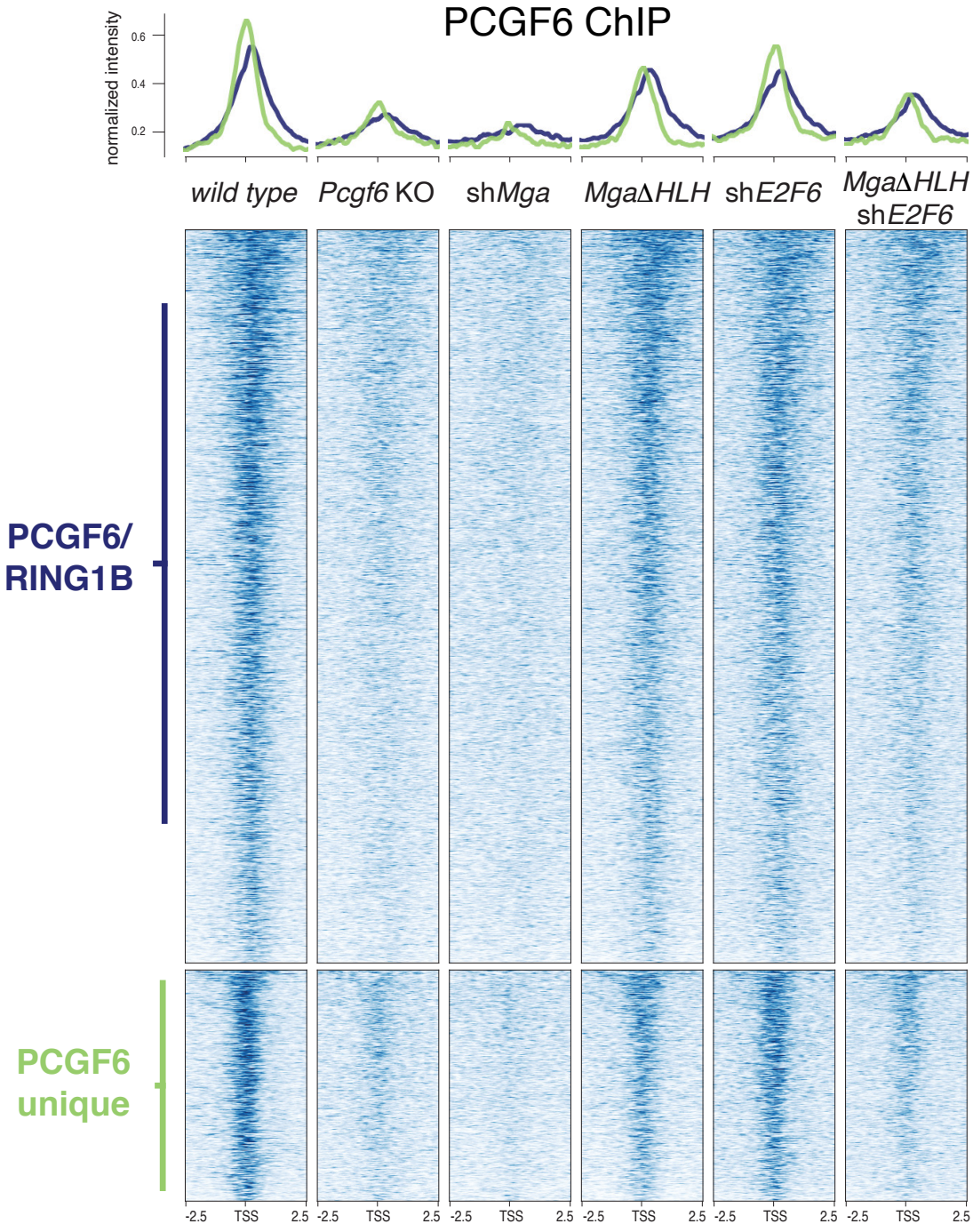


**Figure 2.27: PCGF6 Requires Cooperative E2F and E-Box Recognition for Target Recruitment**. Normalized intensity profiles and heatmap of PCGF6 binding in WT mESCs or Pcgf6 KO, a shMga, MgaΔHLH mutant, shE2f6, and a shE2f6 in MgaΔHLH mutant around ±2.5 kb of the TSS of common and unique target loci.

### 2.1.7 Recruitment to of PRC1.3 is mediated by USF1/2

PRC1.3/5 are the only PRC1 sub-complexes that do not contain a clear ancillary protein that mediates its recruitment to chromatin, contrary to PRC1.2, PRC1.2/4 and PRC1.6 [51]. By performing motif enrichment analysis on PCGF3-bound loci, we found just 1 very enriched motif, which was a type of E-box very similar to USF1/USF2 motifs (Figure 2.28A). USF1/USF2 are transcription factors with an HLH domain that form a dimer very similar to MAX-MGA [120].

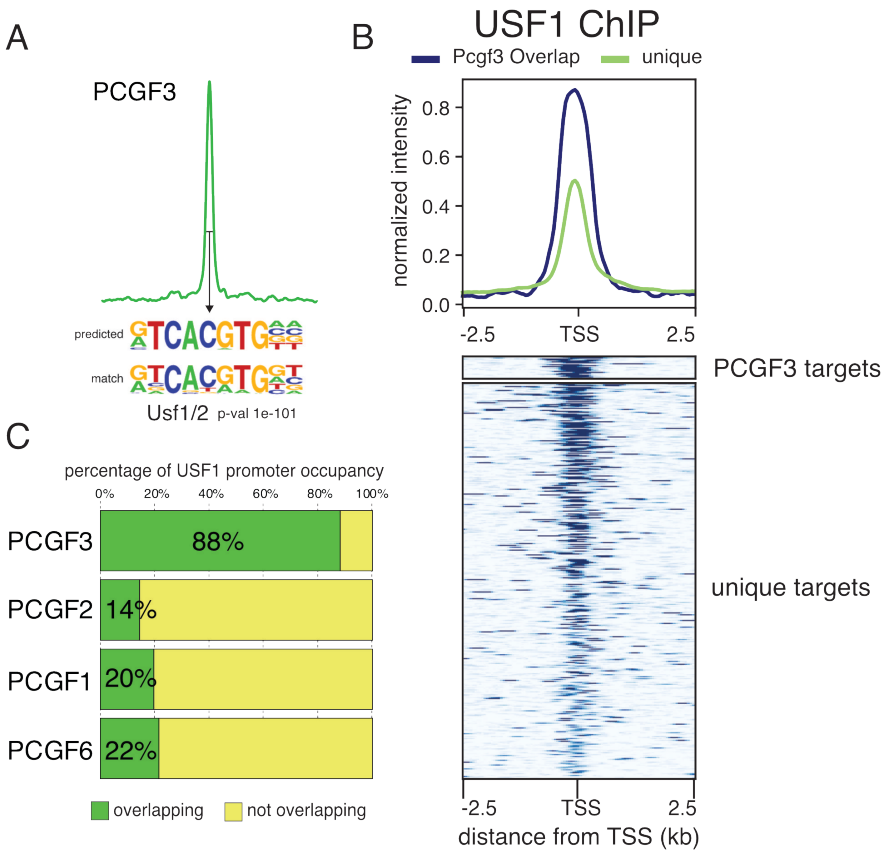


**Figure 2.28: USF1 motif is enriched in PCGF3 target loci.** A) *De novo* motif discovery analysis performed underneath the summit of PCGF3 peaks. Sequence weight matrices of predicted compared to match DNA binding motifs are shown together with relative p-values. B) Input subtracted normalized intensity profiles and heatmap of USF1 binding in WT mESCs around ±2.5 kb of the TSS of USF1-Pcgf3 common and USF1 unique target loci. C) Relative (percentage) of the different PCGF target genes co-occupied by USF1.

By performing a ChIP-seq for USF1 in WT mESC, we found that there was a strong signal enrichment across PCGF3 loci (Figure 2.28B). Also, 88% of PCGF3-bound loci overlapped with USF1, meanwhile other PCGFs did not overlap more than 20%, suggesting a strong specific link between USF1 and PRC1.3 (Figure 2.28C). To test this potential dependency of PCGF3 on USF1, we performed ChIP-seq analyses for PCGF3 in mESC after sh*Usf1* treatment. Two independent biological replicates showed a strong displacement of PCGF3 from chromatin upon *Usf1* knock-down (Figure 2.29), strongly suggesting that USF1 could participate in the recruitment of PRC1.3 to chromatin.

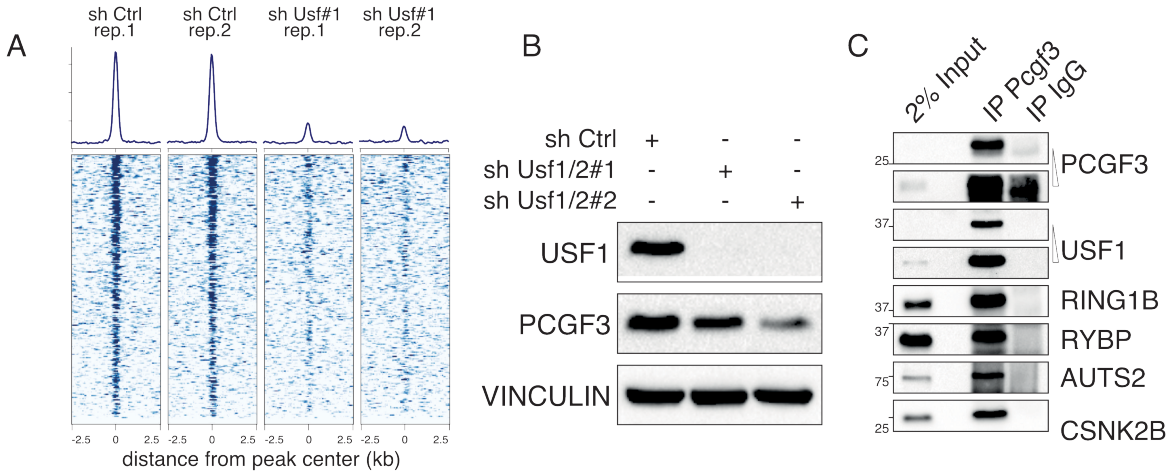


**Figure 2.29: A) USF1 is required for PRC1.3 recruitment or chromatin**. Relative normalized ChIP-seq intensity profiles for two biological replicates (right panels) over a ±2.5 Kb region at PCGF3 target loci. B) Western blot analysis with the indicated antibodies of protein lysates from cells described in (C). Vinculin served as loading control. C) Wstern blot analyses using the indicated antibodies in PCGF3 immunoprecipitations from nuclear extracts prepared from WTmESCs. Rabbit IgG served as an unrelated control antibody. Inputs are shown as loading control.

## 2.2 Role of H2AK119Ub1 in PRC1-mediated gene repression

For many years, the establishment of PcG chromatin domains was thought to start based on the canonical recruitment model. In this model, first PRC2 binds unmethylated CpG islands through its PCL proteins, decorates chromatin with H3K27me3, cPRC1 recognizes and binds to H3K27me3, and finally PRC1 starts to deposit H2AK119Ub1 [121, 122, 123]. However, we have shown that by removing cPRC1, gene repression of PcG targets is maintained (Figure 2.18), which is in line with the absence of strong gene upregulation upon PRC2KO [86]. We have also shown that, in the absence of PCGF6, PRC1 is not able to maintain H2AK119Ub1 levels at PCGF6-loci, which causes an upregulation of PCGF6 specific target genes. These data, at least, challenge the idea of PRC2 and cPRC1 being the central hub of PcG mediated gene repression.

Different studies tried to give some insights about the role of PRC1 and H2AK119Ub1 in gene repression, however, contradictory results from different laboratories still left this topic as an open question. On one side, studies *in vivo* and *in vitro*, in mice and mESC respectively, proposed that PRC1 activity was dispensable for gene repression in mESC and early development in mice [82, 85]. The problem with these studies is that just RINGB was removed, leaving PRC1-RING1A complexes, which can deposit H2AK119Ub1, intact and active. Indeed, upon full removal of PRC1 (RING1A/B KO) [124, 125], it was shown that PcG gene repression was severely compromised. However, by introducing inactive point mutations in PRC1, other laboratories showed in flies that H2AK118Ub1 activity was not required for PRC1-mediated repression [84] and that embryonic lethality of PRC1KO was extended from E10.5 to E15.5 in mice [85]. These data were then challenged by the fact that I53A RING1B mutation is not sufficient to fully block H2AK119Ub1 deposition *in vivo* [54].

Based on literature, PRC1 is required to maintain PcG-mediated gene repression, however, the role of H2AK119Ub1 in this process is not clear. In our laboratory we wanted to answer this question by developing a conditional mESC model in which PRC1 catalytic activity is truly achieved (Figure 2.30). The model, developed in my laboratory by Simone Tamburri and Elisa Lavarone, consists in a Ring1A$^{-/-}$; Ring1B$^{fl/fl}$; Rosa26::CreERT2 conditional mouse ESC line [119] -that will be called parental from now on- with an integrated vector that stably expresses a WT version of RING1B (called WT from now on) or RING1B with I53S point mutation, which renders PRC1 unable to deposit H2AK119ub1 [126, 127] (called I53S from now on). Both the WT and I53S RING1B are FLAG-HA-tagged, which allows us to specifically pull-down the PRC1 complexes using the exogenous RING1B version.

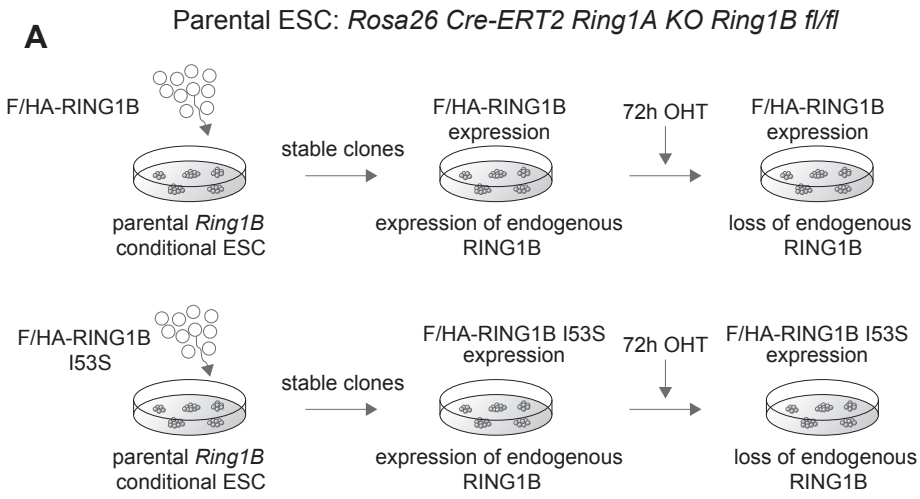


**Figure 2.30: Model to uncouple PRC1 activity from H2AK119Ub1**. Schematic representation of the strategy used for the generation of Ring1A$^{-/-}$; Ring1B$^{fl/fl}$; Rosa26::CreERT2 conditional mESCs stably expressing FLAG-HA (F/HA)-tagged RING1B WT or I53S.

### 2.2.1 Uncoupling PRC1 function from H2AK119Ub1

Upon 4-hydroxy tamoxifen (OHT) treatment on our 3 cell lines (parental, WT and I53S) the following happens: the parental cell lines remain without any RING1A/B

activity, and then the WT and I53S mESC lines express only the exogenous normal or inactive RING1B versions, respectively. Following previous work done in our lab [86], we chose 72h after OHT treatment to perform all the experiments since all endogenous RING1B and H2AK119Ub1 are lost at that time. Western blot analyses revealed that indeed all RING1B and H2AK119Ub1 levels were lost after 72h of OHT treatment in both parental and I53S samples, meanwhile WT samples were able to rescue the non-treated (-OHT) parental H2AK119Ub1 levels. We also checked by mass spectrometry (MS/MS) analyses that WT and I53S models were able to assemble all PRC1 subcomplexes (Figure 2.31), both variant and canonical. Confirming the western blot data, ChIP-seq analyses of H2AK119Ub1 showed no signal after 72h of OHT in the parental and I53S samples (Figure 2.32). WT cells were able to rescue H2AK119Ub1 levels compared to -OHT parental cells. With these tools, we are able to uncouple PRC1 from H2AK119Ub1 function, allowing us to analyze which is the role of this histone mark in establishing gene repression.

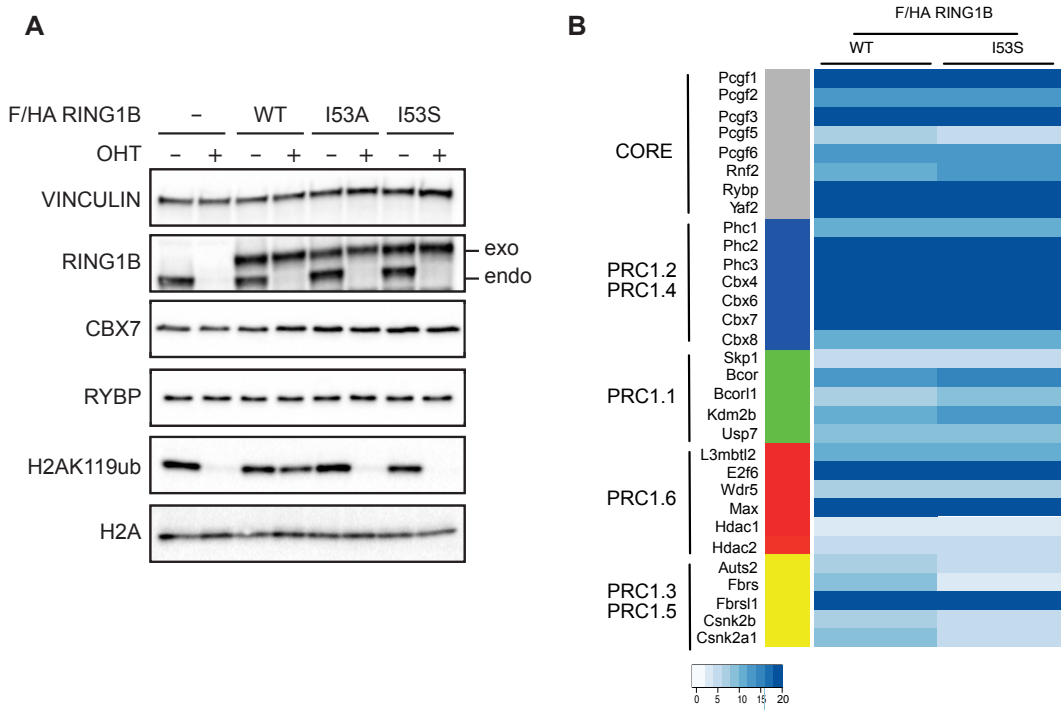


**Figure 2.31: WT and I53S mESC lines are able to assemble all PRC1 sub-complexes**. A) Western blot analysis with the indicated antibodies of total protein extracts obtained from the specified cell lines upon 72 h of treatment with OHT (+OHT) or EtOH (-OHT). Vinculin and histone H2A were used as loading controls. B) Values of the LFQ ratios of the RING1B WT and I53S obtained by tandem mass spectrometry (MS/MS) analyses in the RING1B immuno-purifications (anti- FLAG) from Ring1A$^{-/-}$; Ring1B$^{fl/fl}$; Rosa26::CreERT2 conditional mESCs stably expressing FLAG-HA (F/HA)-tagged RING1B WT or I53S upon 72 h of treatment with OHT (+OHT).

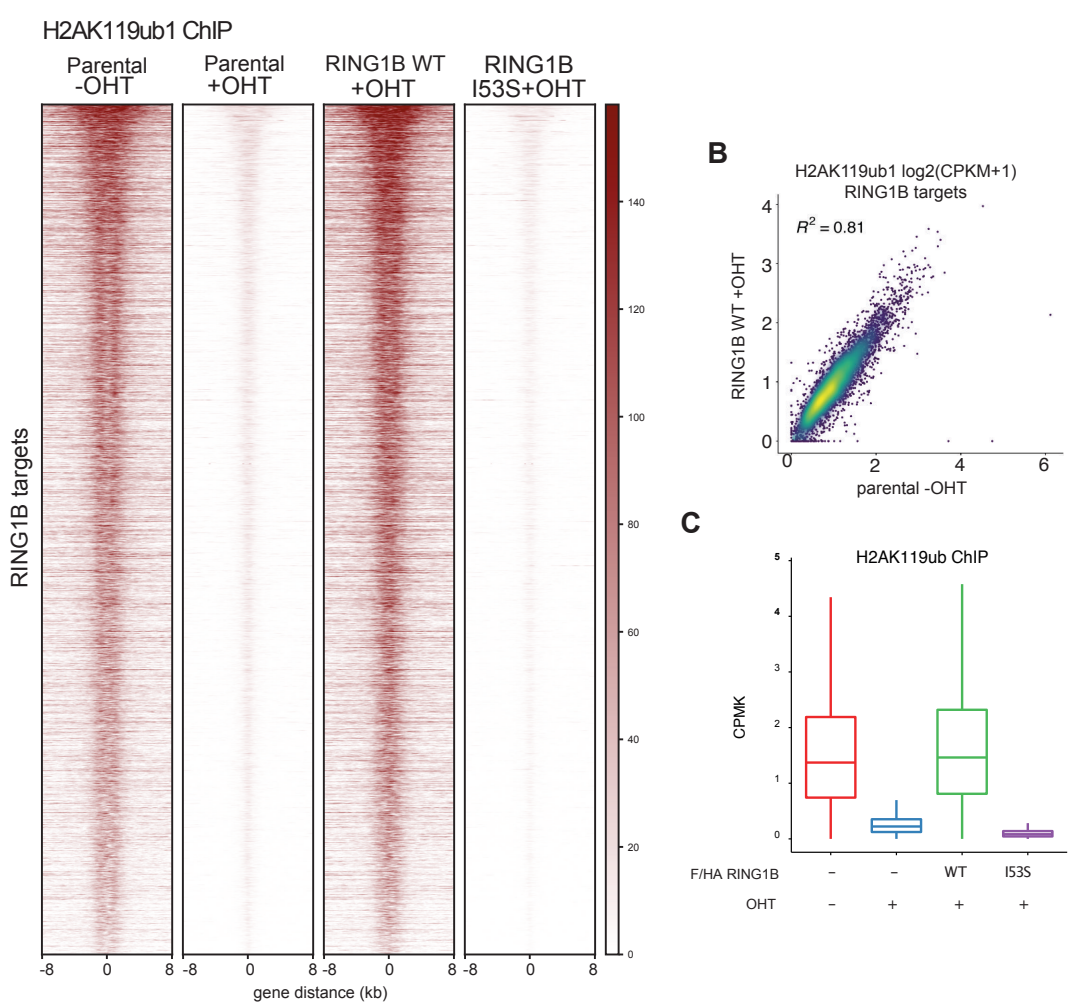


**Figure 2.32: H2AK119Ub1 is lost in I53S RING1B mESC**. A) Heatmaps representing normalized H2AK119ub1 ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines. B)Scatterplot showing the relationship between H2AK119ub1 CPMK levels (counts per million per kilobase) between parental EtOH treated (?OHT) and RING1B WT OHT-treated (+OHT) cells in RING1B target loci. R2 represents the coefficient of determination of linear regression. C)Boxplots representing H2AK119ub1 ChIP-seq CPMK levels in the indicated cell lines at RING1B target loci.

### 2.2.2 PRC1-mediated gene repression depends on H2AK119Ub1

To test whether H2AK119Ub1 is important to maintain PcG mediated repression, we performed RNA-seq in parental, WT and I53S samples. First, we checked that the co-existence of endogenous RING1B and the exogenous versions (WT and I53S) was not causing any changes in gene expressionn due a dominant-negative effect. As shown

in Figure 2.33 (panel A), there were no differentially expressed genes between the parental and WT/I53S lines when treated with the vehicle (EtOH), neither between I53S and WT cells. Upon 72h of OHT treatment, parental cells without RING1A/B showed a strong upregulation of gene expression, in line with previous reports [124] (Figure 2.33, panel B) and the role of PRC1 in maintaining gene repression. Interestingly, meanwhile WT samples were able to rescue the lose of endogenouse RINGA and RING1B, this gene de-repression was phenocopied by the I53S mutant, which forms PRC1 complexes that are not able to deposit H2AK119Ub1. The differentially expressed genes in parental and I53S were reproducible, as well as the extent of their up- and downregulation (Figure 2.34, panel A,B,C). Also, consistent with the repressive nature of PRC1, around 70% of upregulated genes were H2AK119Ub1 targets, meanwhile just 10% of downregulated genes were targets (Figure 2.34, panel D). Very importantly, this gene de-repression caused by the absence of H2AK119Ub1 was not phenocopied by the removal of PRC2, in line with previous data from our lab [86] (Figure 2.34, panel E).

### 2.2.3 Absence of H2AK119Ub1 alters H3K27me3 deposition

There has been a lot of discussion about which recruitment model, the canonical or the variant, is the starting point for the formation of PcG domains [50, 61]. Using this system, we are able to analyze what happens to PRC2 recruitment and H3K37me3 deposition in the absence of H2AK119Ub1. First, we checked that H2AK119Ub1-null mESC didn't alter the abundances of main PRC2 components, which include core components (EZH2, EED, SUZ12), PRC2.2 (JARID2) and PRC2.1 (MTF2) (Figure 2.35). Loss of H2AK119Ub1 produced a reduction in genome-wide H3K27me3 levels, in line with a displacement of SUZ12 from chromatin (Figure 2.36, A and B panels).

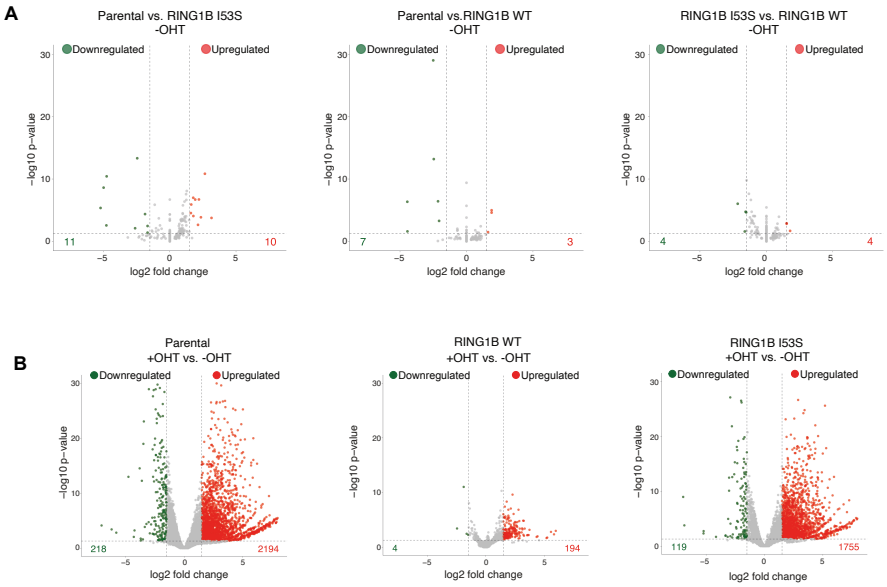


**Figure 2.33: H2AK119Ub1 is required for PRC1-mediated gene repression (I)**. A) Volcano plots of −log10 (p value) against log2 fold change representing the differences in gene expression, related to RNA-seq analysis, in the indicated cell lines upon EtOH treatment (-OHT). Upregulated (red) and downregulated (green) genes are highlighted. As in (A) upon OHT treatment (+OHT).

This demonstrates that H2AK119Ub1 is important to maintain high levels of PRC2 chromatin occupancy as well as H3K27me3.

#### 2.2.3.1 PRC2.1 and PRC2.2 are differentially affected by H2AK119Ub1

PRC2.1 and PRC2.2 are recruited to chromatin by different mechanisms. Meanwhile PRC2.1 recognizes unmethylated CpG islands through its PCL subunits [62], PRC2.2 can interact with H2AK119Ub1 through its JARID2 and AEBP2 subunits [63, 64]. Also, PRC2 can recognize its own catalytic product H3K27me3 [73]. Thus, we asked ourselves if the absence of H2AK119Ub1 would have different effects on PRC2.1 and PRC2.2 recruitment. To test this, we profiled by ChIP-seq the genome-wide occupancy of PRC2.1 and PRC2.2 specific ancillary proteins (MTF2 and EPOP for PRC2.1, JARID2 and AEBP2 for PRC2.2). We profiled JARID2 and MTF2 in a small time course, 48h and 72h after OHT treatment, to see the kinetics of PRC2 displace-

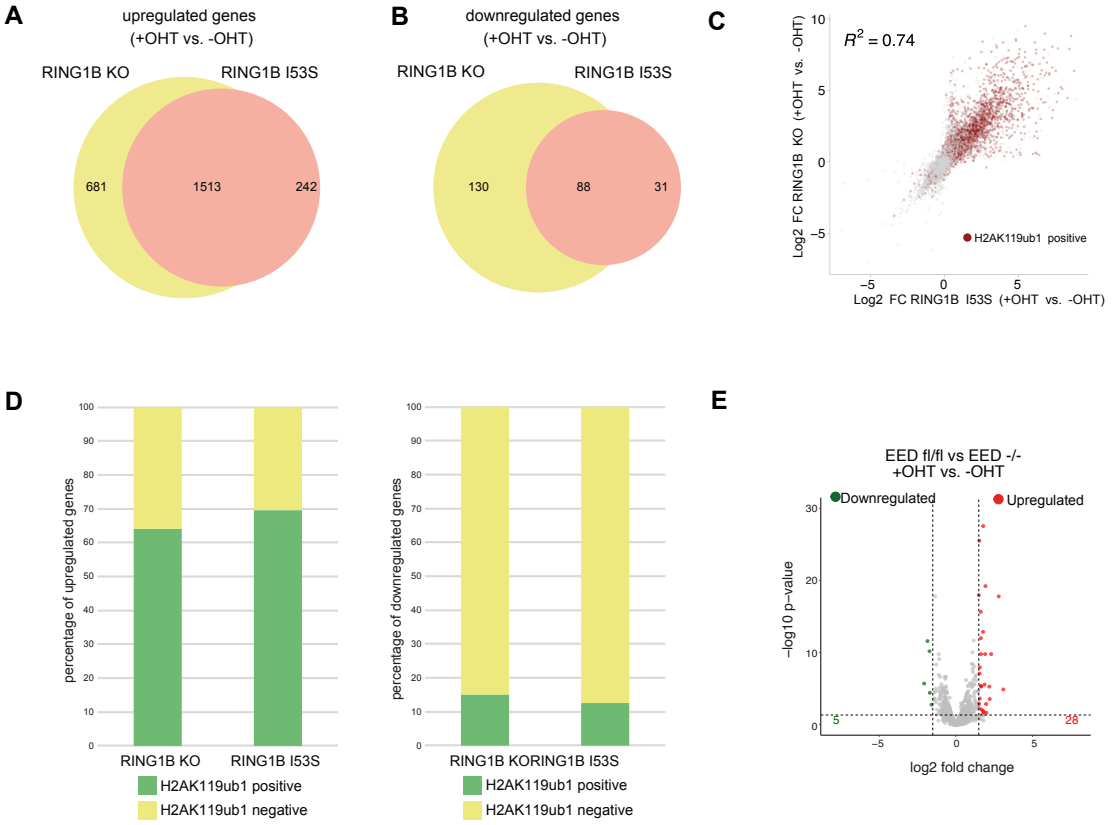


**Figure 2.34: H2AK119Ub1 is required for PRC1-mediated gene repression (II)**. A) Venn diagrams showing the overlap of upregulated genes between the indicated cell lines. B) As in (A) for downregulated genes. C) Scatterplot showing the relationship between log2 fold changes (FC) between the indicated cell lines at RING1B target loci. R2 represents the coefficient of determination of linear regression. Genes with promoters (±2.5 kb around transcription start site [TSS]) containing H2AK119ub1 peaks are highlighted in red. D) Barplots showing the percentage of upregulated (left) or downregulated (right) genes with promoters (±2.5 kb around TSS) containing H2AK119ub1 peaks in the indicated cell lines. E) Volcano plots of –log10 (p value) against log2 fold change representing the differences in gene expression, related to RNA-seq analysis, in $EED^{fl/fl}$ versus $EED^{-/-}$ ESCs. Upregulated (red) and downregulated (green) genes are highlighted.

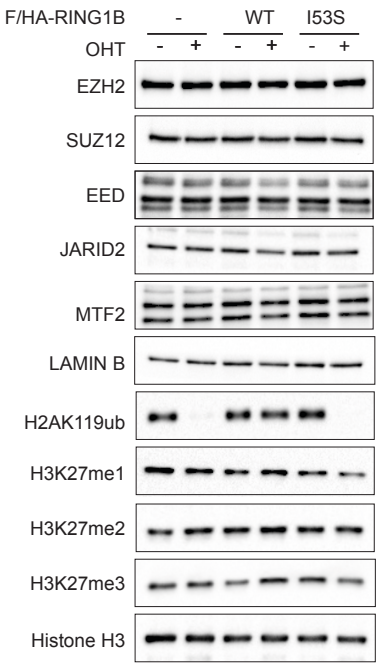


**Figure 2.35: PRC2 components expression is not altered in the absence of H2AK119Ub1.** Western blot analysis with the indicated antibodies of protein extracts obtained from the specified cell lines upon 72 h of treatment with OHT (+OHT) or EtOH (-OHT). LAMIN B and histone H3 were used as loading controls.

ment from chromatin. Displacement of both PRC2.1 (MTF2) and PRC2.2 (JARID2) was evident 48h after OHT treatment, and was stronger at 72h (Figure 2.37). The displacement of JARID2 was stronger at 48 and 72h compared to MTF2 (median of -4 log2FC and -2, respectively), according with its role in recognizing H2AK119Ub1 [63]. We observed the same trend when analyzing the genome-wide occupancy of other PRC2.1 and PRC2.2 subunits (EPOP and AEBP2, respectively) 72h after OHT treatment (Figure 2.38). In fact, AEBP2 ChIP-seq in I53S and parental cells showed no signal, making more evident the fact that PRC2.2 is differentially affected by loss of H2AK119Ub1. These data demonstrate that H2AK119Ub1 is required to recruit PRC2.2 to chromatin and also to stabilize PRC2.1 binding.

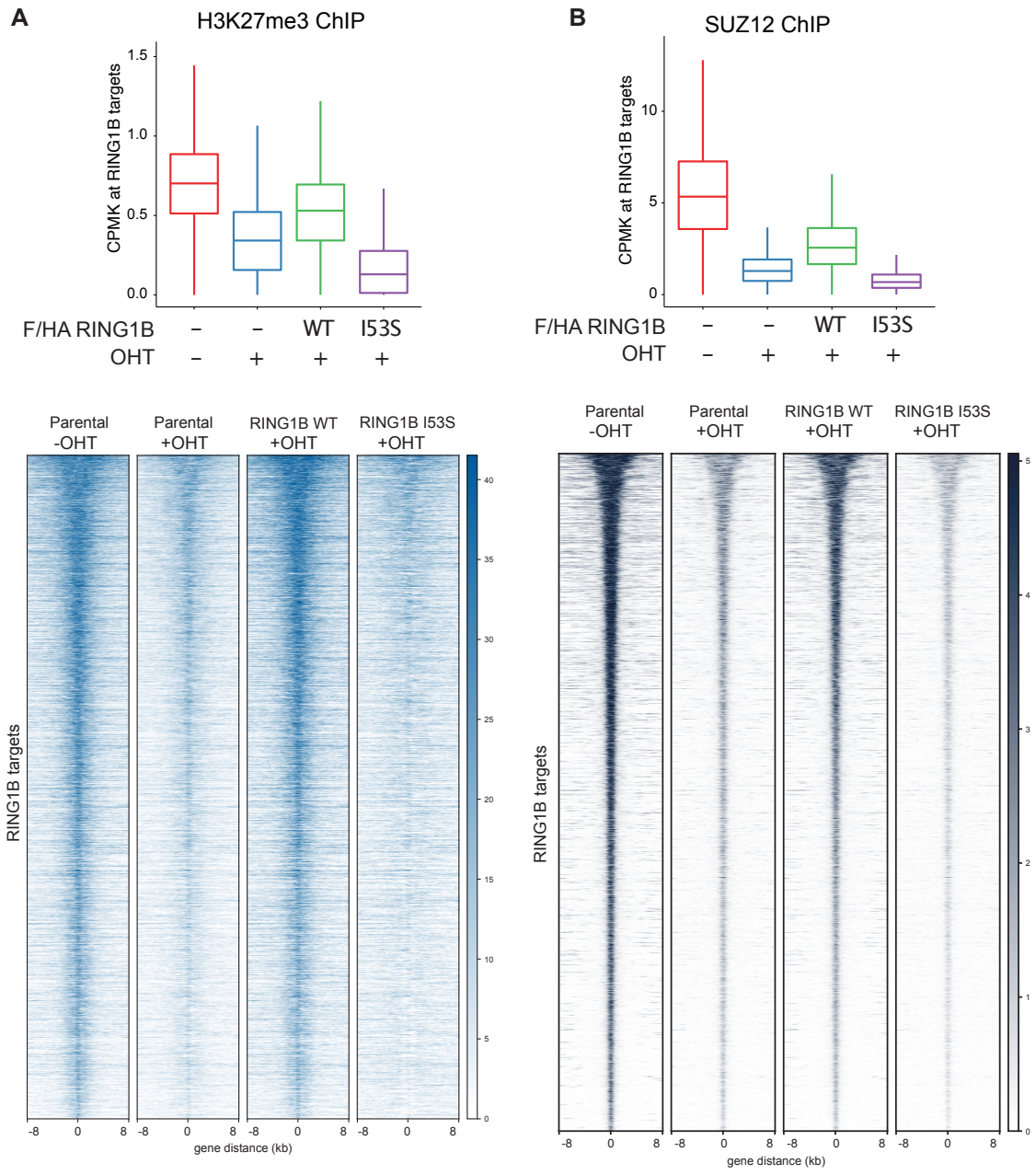


**Figure 2.36: H3K27me3 and SUZ12 are partially displaced from chromatin in the absence of H2AK119Ub1**. A) Heatmaps and their corresponding boxplots representing normalized H3K27me3 ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines. B) Same as A but for SU12 ChIP-seq.

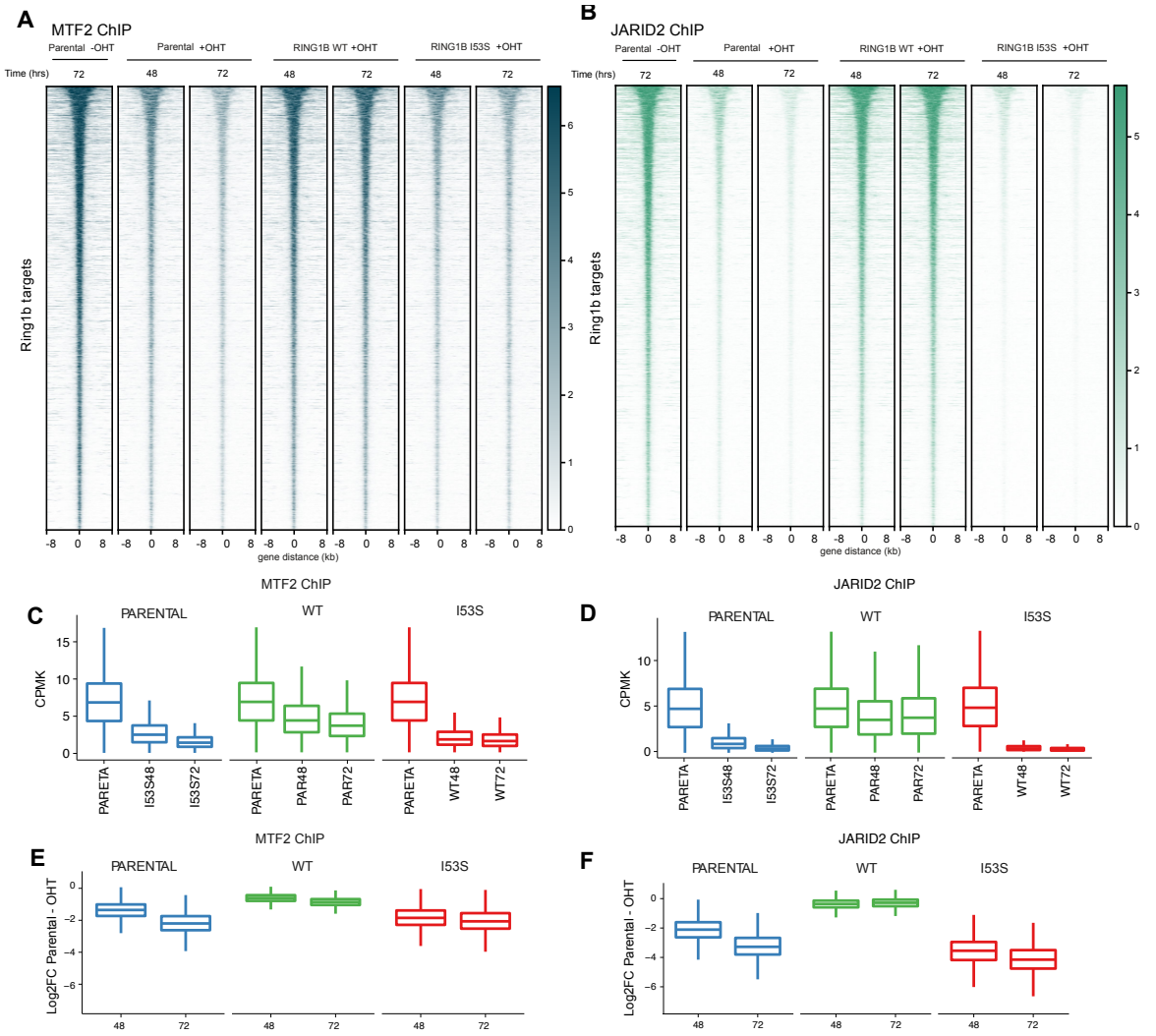


**Figure 2.37: PRC2.2 is more affected by H2AK119Ub1 loss than PRC2.1 (I).** (A) Heatmaps representing normalized MTF2 ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines at the indicated time point post OHT induction. (B) Heatmaps representing normalized JARID2 ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines at the indicated time point post OHT induction. (C) Boxplot representing MTF2 ChIP-seq CPMK levels in the indicated cell lines at RING1B target loci at the indicated time point post OHT induction. (D) Boxplot representing JARID2 ChIP-seq CPMK levels in the indicated cell lines at RING1B target loci at the indicated time point post OHT induction. (E) Boxplot representing the log2 ratio of MTF2 CPMK levels at RING1B target loci between RING1B WT- and I53S-expressing cells at the indicated time point post OHT induction. (F) Boxplot representing the log2 ratio of JARID2 CPMK levels at RING1B target loci between RING1B WT- and I53S-expressing cells at the indicated time point post OHT induction.

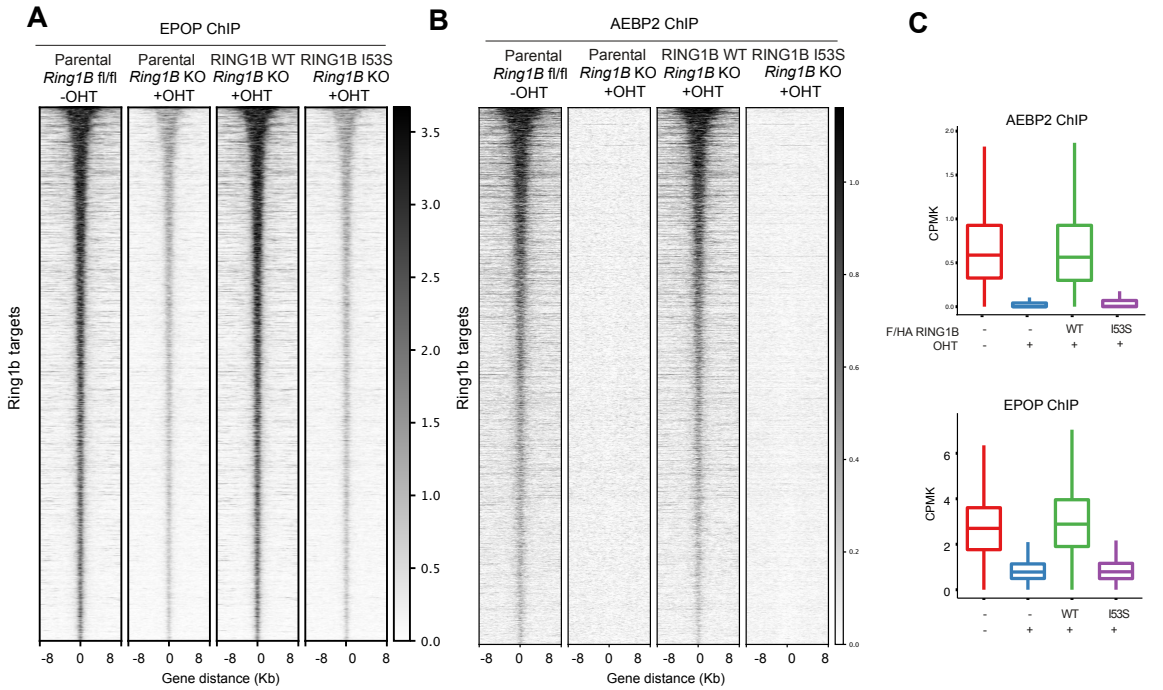


**Figure 2.38: PRC2.2 is more affected by H2AK119Ub1 loss than PRC2.1 (II).** (A) Heatmaps representing normalized EPOP ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines. (B) Heatmaps representing normalized AEBP2 ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines. (C) Boxplot representing AEBP2 (top) or EPOP (bottom) ChIP-seq CPMK levels in the indicated cell lines at RING1B target loci.

### 2.2.3.2 PRC2.1 maintains residual PRC2 binding in the absence of H2AK119Ub1

To further demonstrate that the residual PRC2 binding upon H2AK119Ub1 loss is due to PRC2.1, we decided to remove MTF2 (PRC2.1) from our cell conditional system. Removal of RING1A/RING1B in the parental line or expression of I53S RING1B in the absence of MTF2 reduced H3K27me3 levels almost to zero, as shown in Figure 2.39. This was confirmed by profiling the genome-wide binding of PRC2 (SUZ12), which was totally displaced from chromatin upon MTF2 removal and absence of H2AK119Ub1 (Figure 2.40).

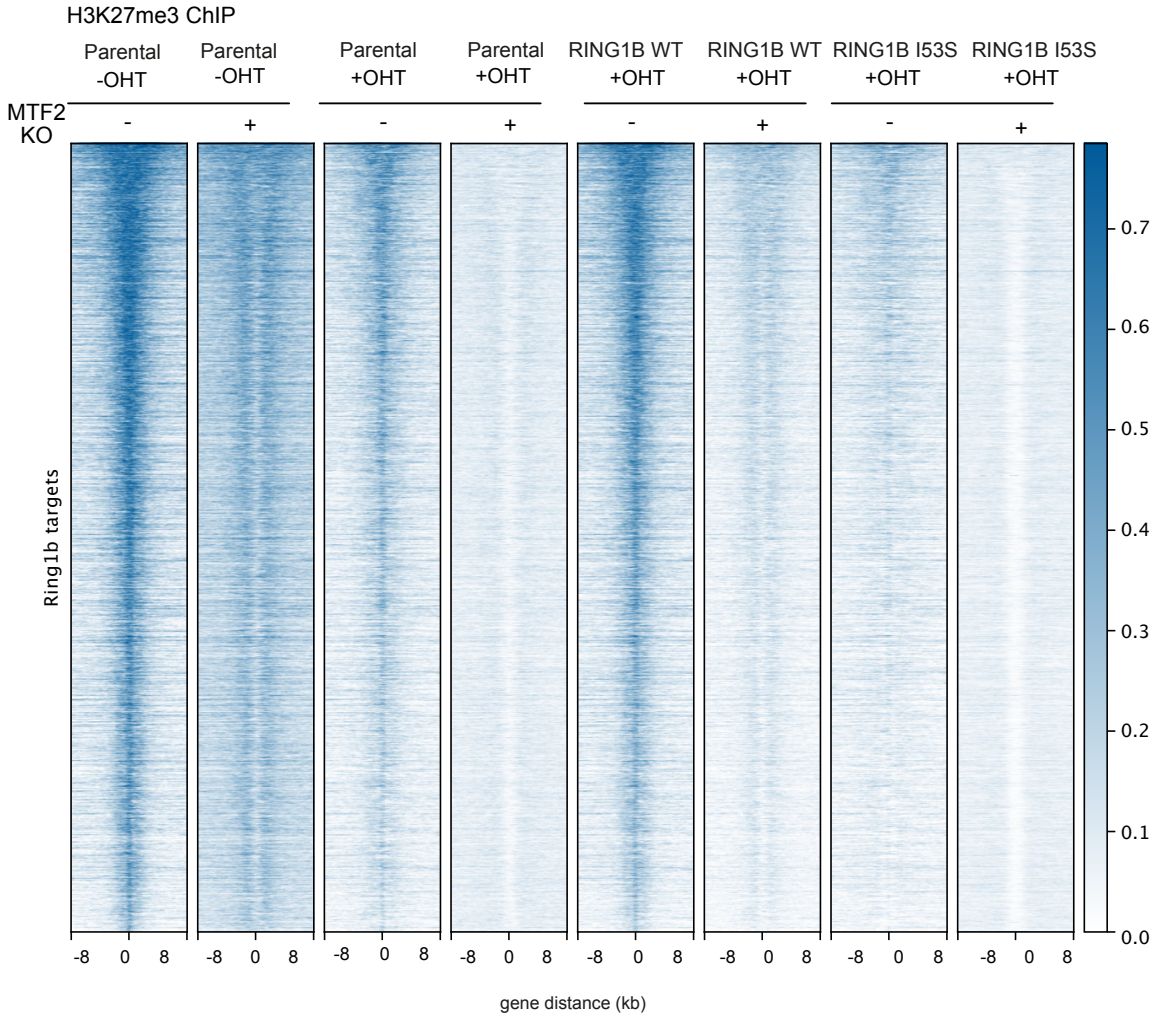


**Figure 2.39: MTF2 is required for H3K27me3 deposition in the absence of H2AK119Ub1**. Heatmaps representing normalized H3K27me3 ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines.

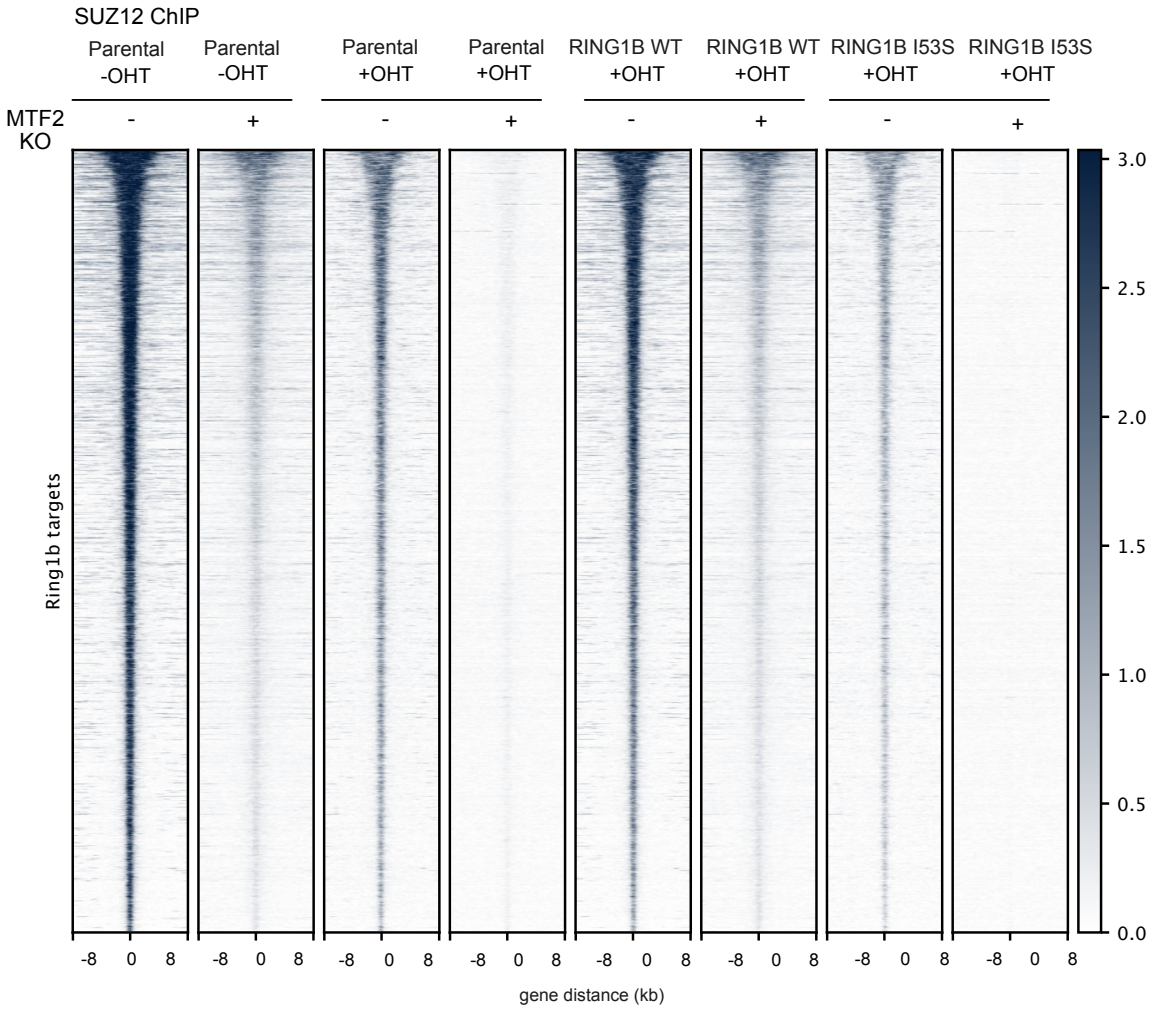


**Figure 2.40: MTF2 is required for SUZ12 recruitment in the absence of H2AK119Ub1.** Heatmaps representing normalized SUZ12 ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines.

### 2.2.4 Displacement from chromatin of PRC1 in the absence of H2AK119Ub1

We have shown that around 80% of RING1B signal in mESC is maintained by cPRC1 (Figure 2.11), which depends on H3K27me3 [43]. Based on that, we wanted to see if the reduction in H3K27me3 levels caused by the loss of H2AK119Ub1 was affecting also RING1B recruitment to chromatin. As expected, due to the absence of endogenous and exogenous RING1B, parental cells treated with OHT didn't show any

RING1B signal. WT cells were able to maintain physiological RING1B occupancy on chromatin, meanwhile in I53S it was severely displaced (Figure 2.41, panels A and B). It is important to note that the reduction in RING1B binding in I53S cells is not due to an intrinsic problem of the mutant, since in the absence of OHT treatment RING1B is able to bind chromatin in those cells (Figure 2.41, panel C). Also, supporting the thesis that RING1B displacement in I53S is due to reduced levels of H2AK119Ub1 and H3K27me3, PRC2KO (and thus, absence of H3K27me3) [86] was able to phenocopy the lose of RING1B binding in I53S cells.

#### 2.2.4.1 cPRC1 is preferentially displaced upon H2AK119Ub1 loss

To further demonstrate that PRC1 displacement from chromatin in the absence of H2AK119Ub1 is mainly due to H3K27me3 loss, we profiled cPRC1 and vPRC1 (PCGF2 and PCGF6) in our models. As we have shown previously (Figure 2.22), the absence of RING1A/B to interact with PCGF2 produces a destabilisation of the protein, which now we show that can be rescued by the WT and I53S versions of RING1B (Figure 2.42). Still, in I53S cells, PCGF2 and CBX7 chromatin localization was strongly compromised in the absence of H2K119Ub1 due to the strong reduction in H3K27me3 levels (Figure 2.43). Surprisingly, PCGF6 and RYBP (vPRC1) binding to chromatin was also reduced, but to a lesser extent respect to cPRC1, supporting the idea that PRC1 displacement from chromatin is mainly due to a reduction of H3K27me3 levels (Figure 2.44). In particular, RYBP levels were almost unaltered. This can be better observed when looking at the $\log_2(foldchange)$ of the signal between untreated parental and treated I53S cells (Figure 2.45). We think that these data suggest that the PRC1 displacement observed upon H2AK119Ub1 loss is a secondary effect produced by the previous destabilization of PRC2 and the consequent reduction in H3K27me3 levels.

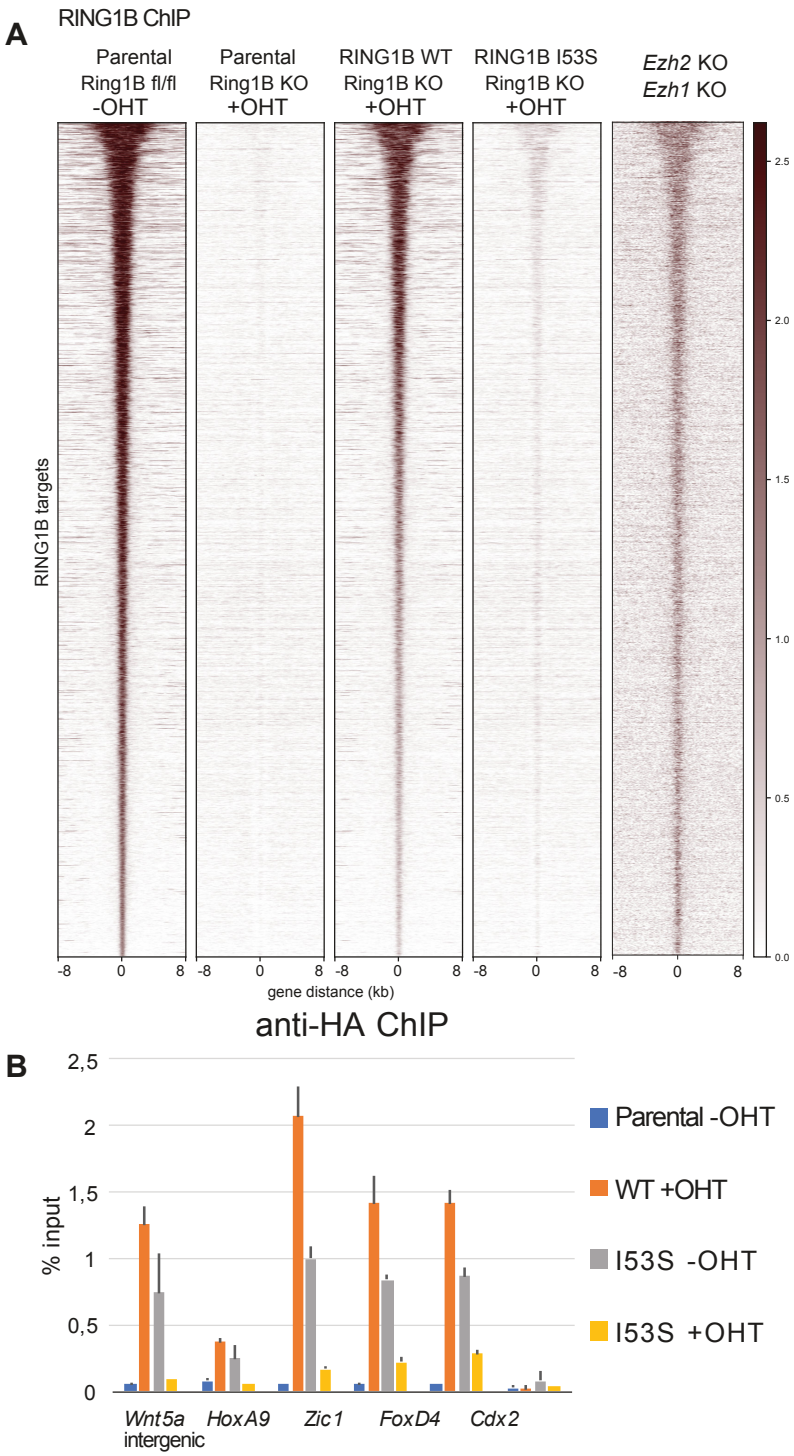


**Figure 2.41: H2AK119Ub1 is required to stabilize PRC1 on chromatin**. (A) Heatmaps representing normalized RING1B ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines. (B) ChIP-qPCR analysis of HA in the indicated cell lines at five specific polycomb targets and one intergenic region. Parental cells served as a negative control.

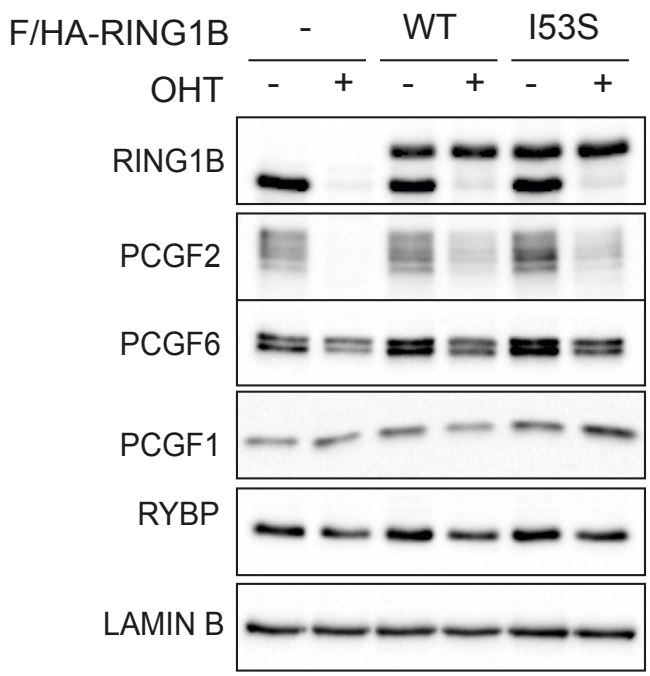


**Figure 2.42: Expression of cPRC1 and vPRC1 components in RING1B conditional models**. Western blot analysis with the indicated antibodies of nuclear protein extracts obtained from the specified cell lines upon 72 hours of treatment with OHT (+OHT) or EtOH (-OHT). LAMIN B served as a negative control.

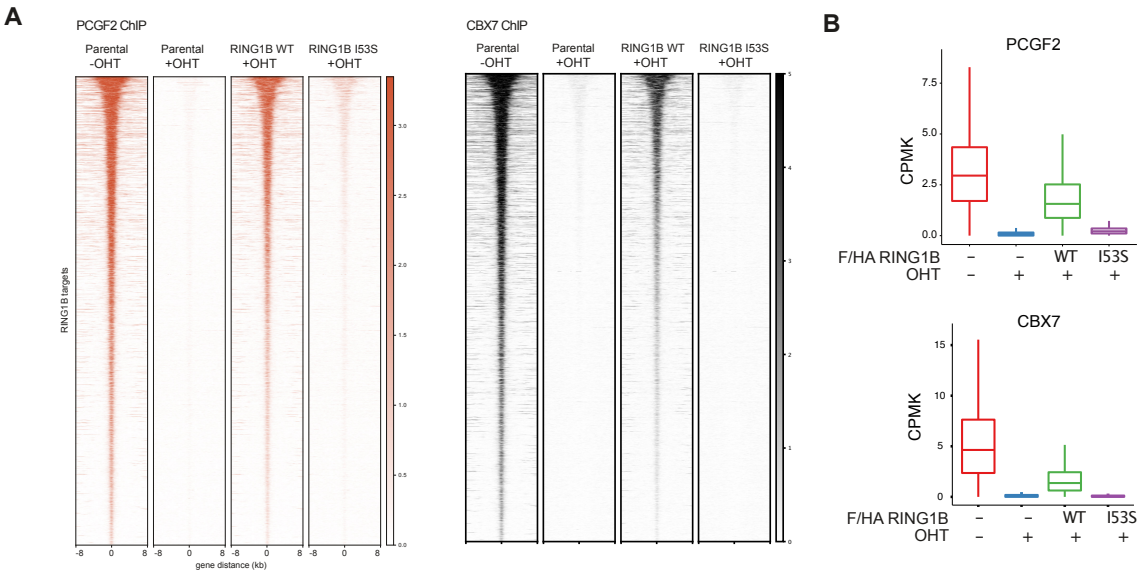


**Figure 2.43: cPRC1 binding is strongly compromised upon H2AK119Ub1 loss**. A) Heatmaps representing normalized PCGF2 and CBX7 ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines. (B) Boxplots representing PCGF2 ChIP-seq CPMK levels (top panel) and CBX7 (bottom panel) in the indicated cell lines at RING1B target loci.

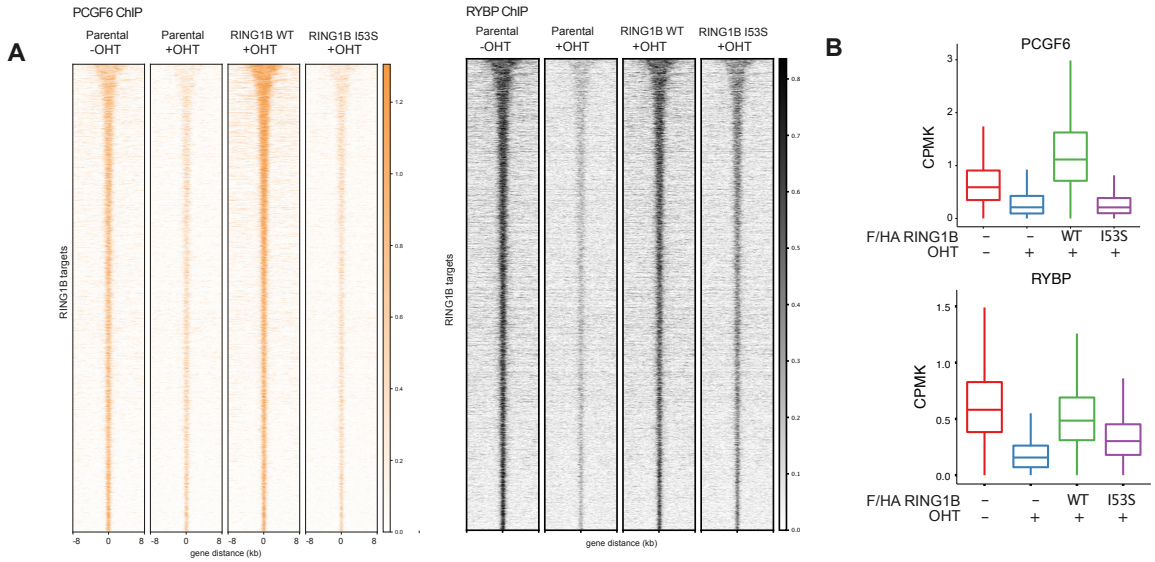


**Figure 2.44: vPRC1 binding is minimally compromised upon H2AK119Ub1 loss**. A) Heatmaps representing normalized PCGF2 and CBX7 ChIP-seq intensities ±8 kb around the center of RING1B target loci in the indicated cell lines. (B) Boxplots representing PCGF6 ChIP-seq CPMK levels (top panel) and RYBP (bottom panel) in the indicated cell lines at RING1B target loci.

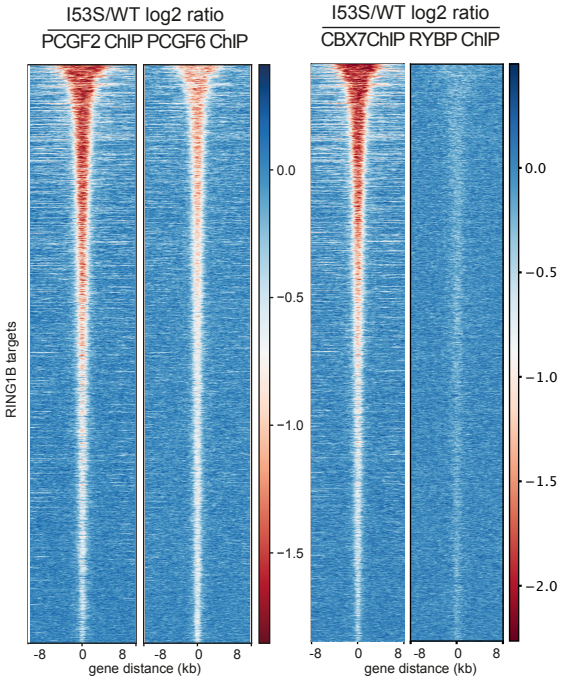


**Figure 2.45: cPRC1 and vPRC1 binding is differently affected by H2AK119Ub1 loss**. (F) Boxplots representing PCGF2 (upper panel) and PCGF6 (bottom panel) ChIP-seq CPMK levels ±250 bp around TSS of PCGF target genes in the indicated cell lines.

## 2.3 Landscape of PCGF proteins in mouse intestine

mESC are a very good *in vitro* system to study PcG proteins, however, they do not fully resemble the environment present in an *in vivo* tissue. Cells within a tissue are exposed to external stimuli, differentiation signals, stress conditions, and insults that constantly activate signaling cascades that trigger changes in gene expression. Thus, replicating *in vitro* observations using *in vivo* models would be very important to validate our data from mESC. Also, *in vivo* models would allow us to assess the role of different PRC1 sub-complexes in maintaining cell identity in challenging environments. Previous data from our lab showed that double deletion of *Ring1a* and *Ring1b* in mouse intestinal stem cells leads to intestinal stem cell exhaustion and eventually death [91], similar to what happens in mESC in the absence of PRC1 [119] -they are not viable-. Instead, PRC2 is dispensable for intestinal homeostasis [117] and does not phenocopy the loss of PRC1, as it occurs in mESC [86]. Instead, PRC2 seems to compromise the regenerative capacity of the intestinal epithelium upon irradiation damage and also produces a strong suppression of stem cell and transit amplifying cells' proliferation, which is driven by the activation of CDK2NA [117].

The aim of this section of the thesis is to try to understand if PCGFs' behaviour *in vivo* resembles our *in vitro* data, if the phenotype caused by the loss of PRC1 can be phenocopied by specific PRC1 sub-complexes and if the PRC2 phenotype depends on the activity of cPRC1. For that, our lab has developed different mouse models that allow the acute deletion of specific PRC1 sub-complexes specifically in mouse intestinal stem cells ($Lgr5^{EGFP-IRES-CreERT2-R26R-lacZ}$ models) or across the whole intestinal tract (AhCre models). These are still ongoing projects, so the data that will be presented from now on are preliminary results that still need to be further elab-

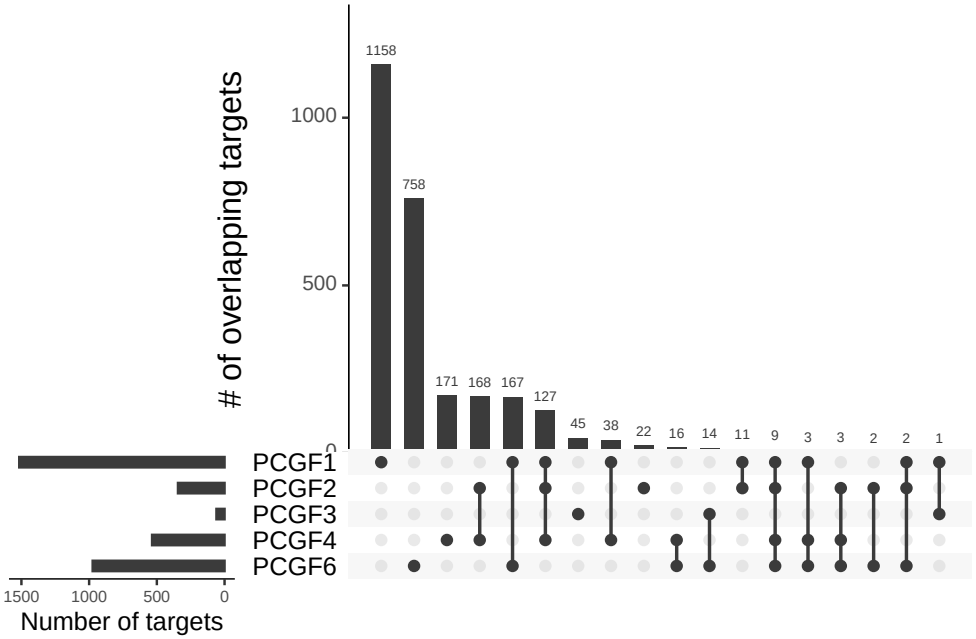


**Figure 2.46: PCGF proteins retain binding specificity in mouse intestinal tissue**. Upset plot showing the overlap between different PCGF targets genes. A gene is considered target when there is at least 1 peak associated with its promoter region (±2.5kb from TSS).

orated, validated and completed. Still, they offer a novel, general and preliminary overview of the PRC1 sub-complexes landscape in the mouse intestinal tissue.

### 2.3.1 Genome-wide occupancy of PCGF proteins correlates with different transcriptional states

To analyze the genome-wide occupancy of PRC1 sub-complexes in the WT mouse intestine, we performed ChIP-seq analyses across all PCGF proteins in mouse intestinal crypts. We obtained positive results for all PCGFs except PCGF5, from which we were not able to detect any enriched loci during the peak calling (Figure 2.46). Our results showed that PCGFs can also retain binding specificity in adult tissues, in particular PCGF1 and PCGF6, which resembles very well our previous data in mESC (Figure 2.5). PCGF1 was again the PCGF with more binding sites, followed by PCGF6 and PCGF2/4. Contrary to mESC, PCGF4 played a relevant role in the recruitment of

cPRC1 to chromatin, since exhibited a similar number of binding sites than PCGF2. Still, they shared most the their target genes, which is consistent with the biochemical redundancy of PRC1.2 and PRC1.4 complexes. Another interesting difference respect mESC was that the overlap between vPRC1 and cPRC1 in terms of shared loci was much lower. In mESC PCGF2 overlapped more than 80% with PCGF1 (Figure 2.5), and PCGF6 overlapped also around 60% with PCGF2. Instead, in intestinal crypts PCGF6 did not overlap with cPRC1, and less than 50% of cPRC1-bound loci contained PCGF1 (Figure 2.46).

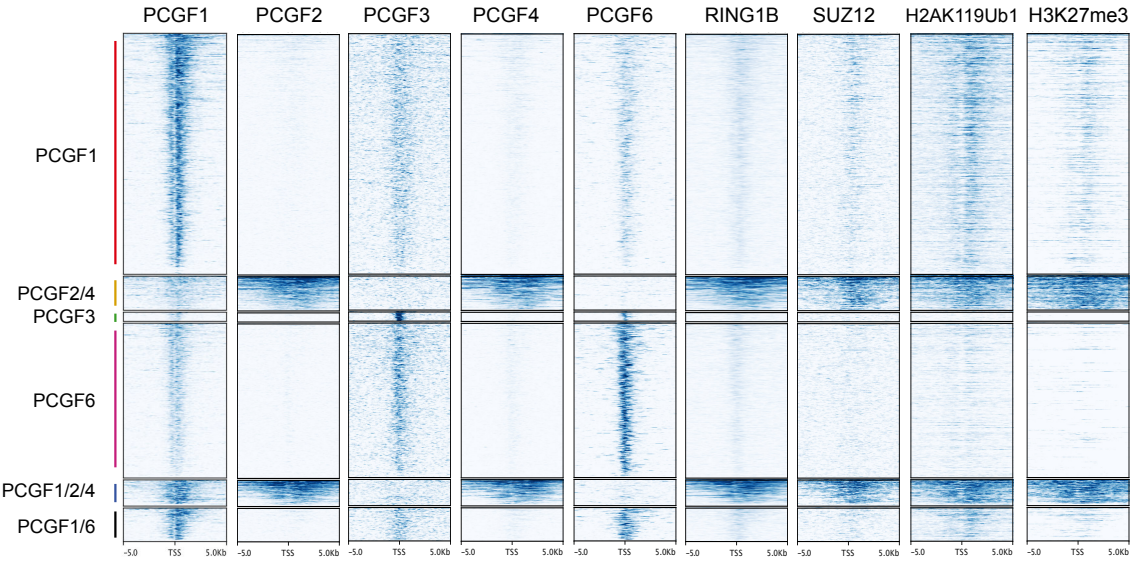


**Figure 2.47: H2AK119Ub1 decoration is correlated with cPRC1 occupancy**. Heatmaps representing the normalized ChIP-seq intensities for the indicated PCGF proteins over ±2.5 kb around the TSS of the indicated loci stratified for PCGF co-occupancy in wild-type (WT) mouse intestinal crypts.

By representing the signal of each PCGF and main PcG core units (RING1B, SUZ12, H2AK119Ub1 and H3K27me3) across the main loci defined in Figure 2.46 (PCGF1-2-3-4-6 unique loci and PCGF2/4, PCGF1/2/4 and PCGF1/6 loci), we observed that, as in mESC, cPRC1 occupancy was always correlated with high PRC1/PRC2 occupancy and high H3K27me3 and H2AK119Ub1 levels. We think that it was also interesting the fact that for vPRC1 complexes, it was possible to detect signal traces

across all their binding sites (so, weak PCGF1 signal across unique PCGF6 peaks, and the same for PCGF6), but not for cPRC1 across vPRC1-bound loci. This high levels of H2AK119Ub1 and H3K27me3 across cPRC1-bound loci were also correlated with a repressive transcriptional state, meanwhile vPRC1-bound loci (except if there was also cPRC1) were correlated with permissive transcription and low H2AK119Ub1 and H3K27me3 levels (Figure 2.48, panel A). Regarding the functional annotation of PCGF target genes, GO enrichment analyses revealed that, as in mESC, PCGF1-2-4 targeted mainly genes regarding to development and differentiation and PCGF3 bound authophagy-related genes (Figure 2.48, panel B). However, enrichment results for PCGF6 target genes were not related to germ cell specific processes, as it happened in mESC, but to cell cycle and regulation of metabolism, which are processes also regulated by MYC (which recognizes the same E-box motifs than PCGF6) [128].

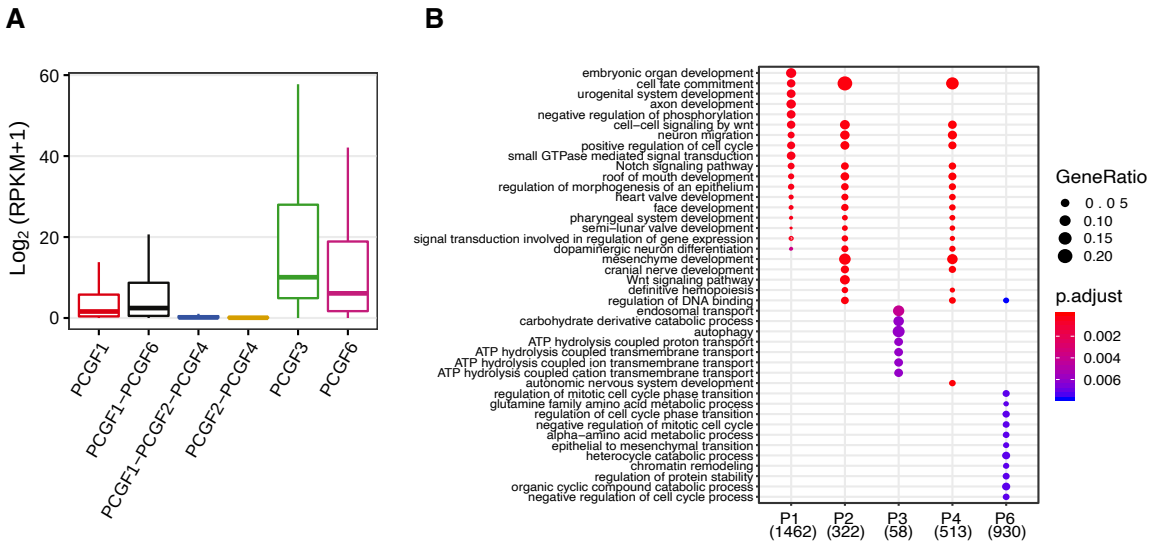


**Figure 2.48: cPRC1 is correlated with repressed transcription**. A) Boxplot showing the mRNA expression of different groups of PCGF target genes. B) Gene ontology analysis for the indicated PCGF target genes. The most represented categories are highlighted. Dot size is proportional to the number of genes corresponding to that gene ontology category, color scale indicates statistical significance (adjusted p-value < 0.01 and q-value < 0.01).

### 2.3.2 Individual PRC1 sub-complexes are dispensable for intestinal homeostasis

One of the main goals of the characterization of PRC1 sub-complexes in the small intestinal epithelium was to understand if the phenotype of PRC1-null small intestine could be recapitulated by specific PRC1 sub-complexes. Using our Lgr5 transgenic mice, we were able to acutely delete each of the different PRC1 sub-complexes in adult mice, specifically in the intestinal stem cells. This model allows a precise phenotypical analysis of the tissue because the expression of $CreER^{T2}$ is not uniform. This means that just some intestinal crypts will suffer the allele deletion, leaving intact the global integrity of the tissue, just producing local alterations that can be analyzed using imaging techniques, for example [97]. Lineage tracing analyses of PRC1-null intestinal stem cells showed that they were exhausted after 3 days of OHT injection, producing a collapse of the intestinal tissue integrity [91]. We performed the same experiments but removing specific PRC1 sub-complexes, as shown in Figure 2.49. We didn't observe any impairment in the ability of intestinal stem cells to maintain tissue homeostasis and give rise to differentiated cell populations in any of the different KOs. This indicates that the phenotype observed in PRC1-null intestinal stem cells has to be due to the loss of multiple PRC1 sub-complex activities. Indeed, by performing RNA-seq on sorted intestinal stem cells (LGR5+) after deletion of each specific PRC1 sub-complex, we observed that none of the KOs was able to resemble the transcriptional changes observed in PRC1-null cells. We were also able to perform RNA-seq on new generated mice models where multiple PRC1 sub-complexes were deleted (PRC1.1/2/4 and PRC1.3/5/6) and they neither were able to resemble the loss of PRC1 (Figure 2.50).

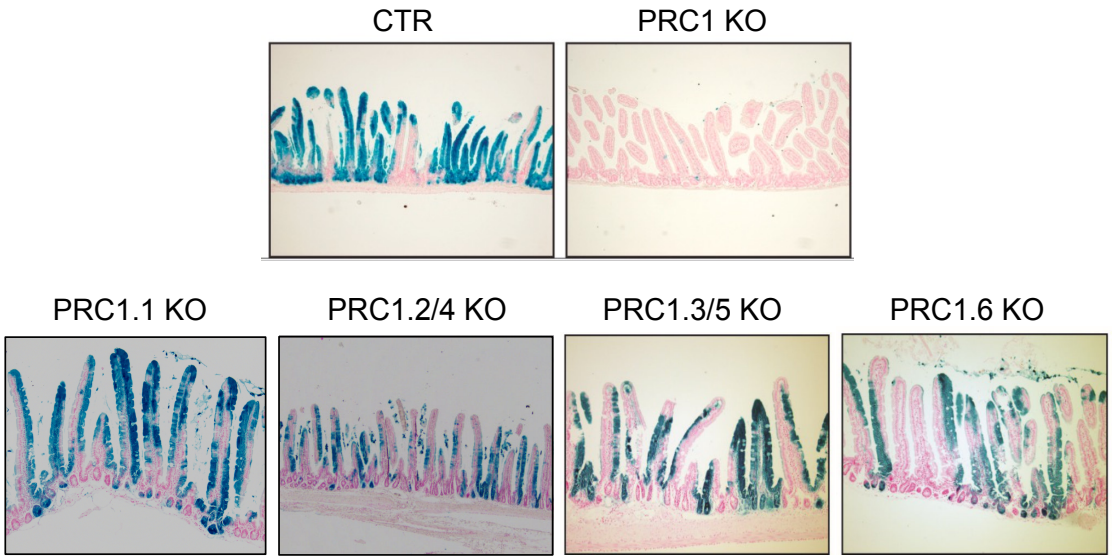


**Figure 2.49: Individual PRC1 sub-complexes are dispensable for intestinal homeostasis (I).** In vivo ISC lineage tracing. R26lSl-LACZ mice with the indicated genotypes were injected with tamoxifen and sacrificed after 15 days. For each sample, the small intestine was stained with X-GAL. Nuclei were counterstained using Nuclear Fast Red.

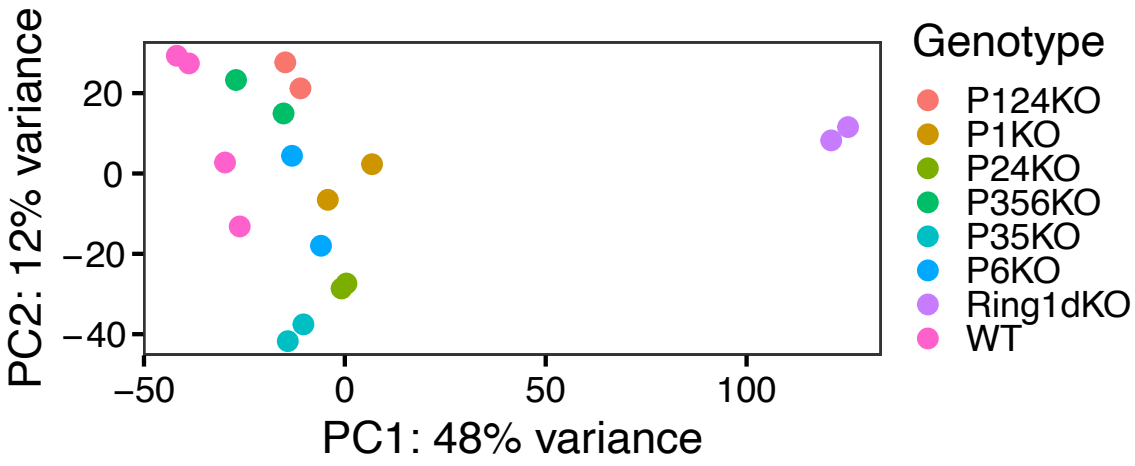


**Figure 2.50: PRC1 activity in LGR5+ cells is the result of multiple PRC1 sub-complex activities.** Principal Compontent Analysis (PCA) of the indicated LGR5+ sorted RNA-seq samples. *Ring1a/b* null LGR5+ cells were obtained from [91].

We also analyzed by ChIP-seq the genome-wide occupancy of RING1B, H2AK119Ub1 and H3K27me3 upon deletion of specific PRC1 sub-complexes. As explained above, since the Lgr5$^{EGFP-IRES-CreERT2-R26R-lacZ}$ models are not homogeneous, we used AhCre mice in order to delete PRC1 sub-complexes across all intestinal cells (except paneth cells) to avoid having contamination from non floxed cells. RING1B occupancy remained unaltered across all *Pcgf-null* backgrounds except in *Pcgf2/4*KO mice, where the signal was strongly reduced, as we already observed in mESC (Figure 2.51). Again, this displacement of RING1B from chromatin was not correlated with a reduction in H2AK119Ub1 or H3K27me3 levels, meaning that despite most RING1B ChIP-seq signal is cPRC1-dependent, the deposition of H2AK119Ub1 is mainly performed by vPRC1 complexes. We observed a strong reduction of H2AK119Ub1 in *Pcgf3/5*KO intestinal crypts. PRC1.3/5 has been recently described, by our lab and others, as complex with low affinity for specific chromatin regions, that instead interacts transiently with chromatin and deposits a genome-wide "stratum" of H2AK119Ub1 [118, 129, 130]. In particular it seems to be important to maintain intergenic H2AK119Ub1 [130]. To test this, we calculated the reduction in H2AK119Ub1 levels across promoter and distal H2AK119Ub1-enriched regions, as shown in Figure 2.52, panel A. This results showed that PRC1.3/5 mainly maintains intergenic H2AK119Ub1 levels, meanwhile PRC1.1 and PRC1.6 were able to maintain H2AK119Ub1 at promoters in the absence of PRC1.3/5. Indeed, H3K27me3 levels were also moderately reduced across intergenic regions in the absence of PRC1.3/5, probably due to a displacement of PRC2.2 from chromatin, which recognizes H2AK119Ub1 through its subunit JARID2 [63, 64].

Histological analyses of AhCre mice lacking the different vPRC1 sub-complexes showed completely unaltered homeostasis (Figure 2.53), since the structure of the epithelium (panel A) and cell proliferation (panel B) were comparable to WT mice.

These data suggest that intergenic levels of H2AK119Ub1 are not relevant for the maintenance of PcG-mediated gene repression in mouse intestinal crypts. Also, these histological results are in accordance with the unaltered H2AK119Ub1 levels in *Pcgf1, 2/4 and 6* KO samples (Figure 2.51). On the same line, absence of cPRC1 did not phenocopied PRC2 and H3K27me3 loss [117], neither regarding cell proliferation (Figure 2.54, A) or expansion of goblet cells (Figure 2.54, B). This means that, contrary to what we observed in mESC, H3K27me3 levels could be important to maintain gene repression at specific PcG target genes independently of cPRC1.

In summary, our data showed that PRC1 sub-complexes, with some exceptions, behaved very similarly in embryonic and adult tissues. The phenotype of PRC1-null intestine can't be explained by the lack of a particular PRC1 sub-complex, not even by the lack of multiple (PRC1.1/2/4 and PRC1.3/5/6), meaning that, as it happens in mESC, PRC1 function is the results of a combined and cooperative activity of multiple PRC1 sub-complexes.

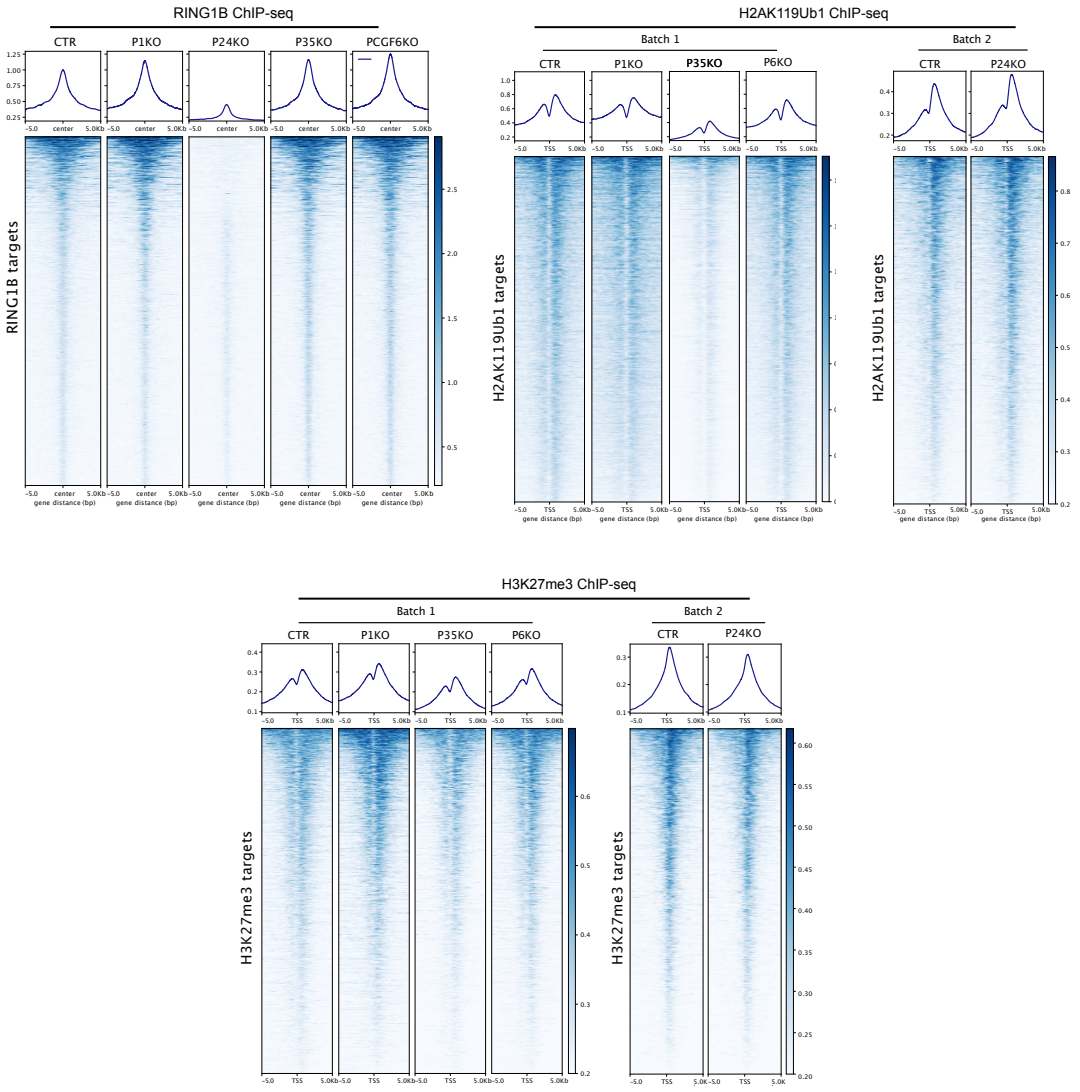


**Figure 2.51: H2AK119Ub1 levels are maintained by the cooperative effect of different PRC1 sub-complexes**. Heatmaps representing the normalized ChIP-seq intensities for the indicated samples (RING1B, H2AK119Ub1 and H3K27me3) over ±2.5 kb around the TSS of the indicated loci stratified for PCGF co-occupancy in WT, P1KO, P24KO, P35KO and P6KO mouse intestinal crypts. Batch 1 and Batch 2 indicate different chromatin and processing batches for cPRC1 (P24KO) and vPRC1 KOs (P1KO, P35KO, P6KO).

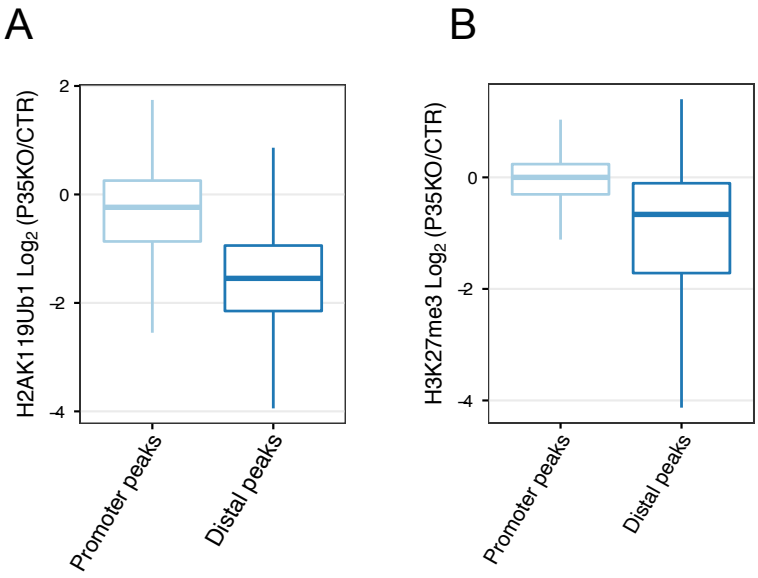


**Figure 2.52: PRC1.3/5 maintains intergenic H2AK119Ub1**. A) Boxplots showing the fold change of H2AK119Ub1 levels between control samples (WT mice) and *Pcgf3/5*KO intestinal crypts across intergenic and promoter H2AK119Ub1 peaks (called in WT samples). B) Same as A but for H3K27me3.

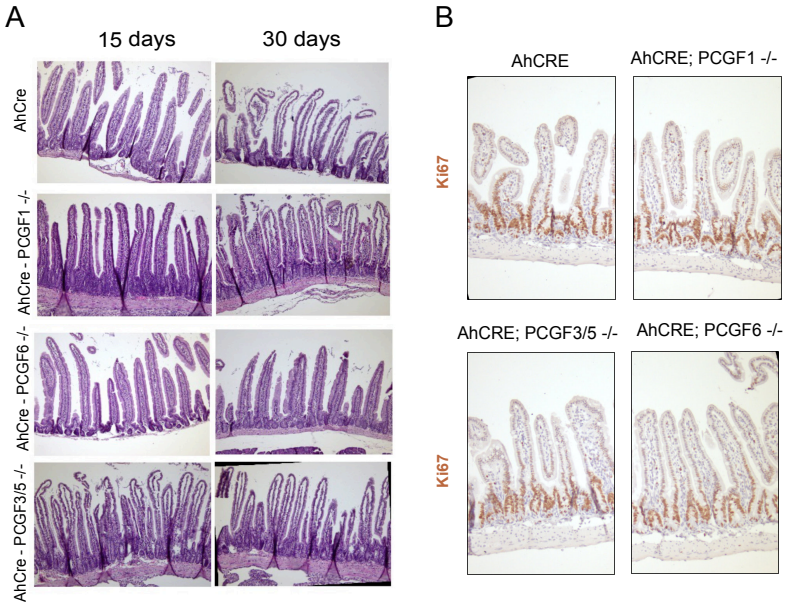


**Figure 2.53: Individual PRC1 sub-complexes are dispensable for intestinal homeostasis (II)**. A) HE staining on intestinal sections from the indicated AhCre mice prepared 15 and 30 days after induction. B) Immunostaining using anti-Ki67 on intestinal sections from the indicated AhCre mice prepared 15 days after induction.

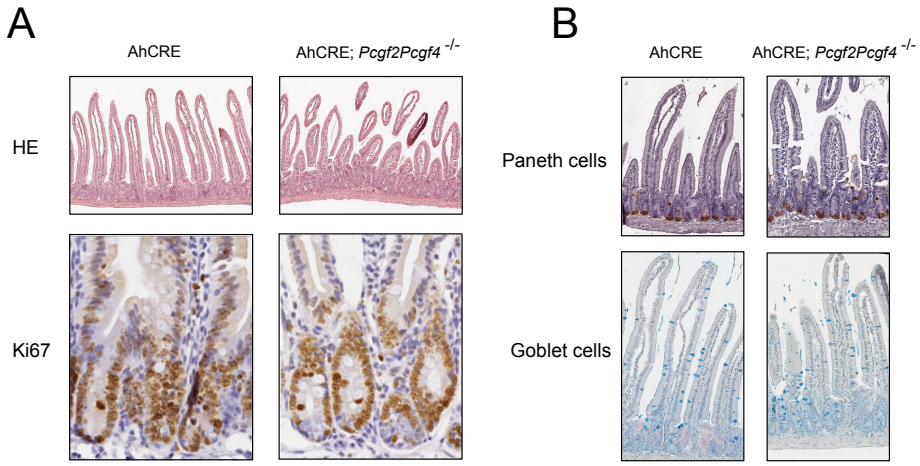


**Figure 2.54: cPRC1 does not phenocopy the loss of PRC2.** A) HE staining (top panels) and immunostaining using anti-Ki67 (bottom panels) of the indicated AhCre mice 15 days after induction. B) Lyz1 (top panels) and Alcian Blue (bottom panels) stainings prepared from small intestinal sections of the indicated AhCre mice 15 days after $\beta$-naphthoflavone injection.

### 2.3.3 PRC1.6 loss induces tuft hyperplasia

AhCre *Pcgf6*$^{-/-}$ mice did not show any alteration in proliferation or tissue integrity, consistent with the unmodified occupancy of RING1B and maintenance of H2AK119Ub1 levels. Also, RNA-seq analyses of these mice didn't show big changes in gene expression (around 200 differentially expressed genes) (Figure 2.55, A). However, across the 162 upregulated genes, we consistently found genes specifically related to tuft cells. By performing enrichment analyses on downloaded cell signatures from the intestinal epithelium [131], we found out that the tuft signature was extremely enriched across AhCre *Pcgf6*$^{-/-}$ upregulated genes (Figure 2.55, B). These data were confirmed also by GSEA analysis, which also showed a strong enrichment of this signature across upregulated genes (Figure 2.55, C). We performed western blot analyses to validate these data, which showed a strong upregulation of the Tuft marker DCLK1 (Figure 2.56, A). Then, to further validate that this increase of DCLK1 corresponded to an increase in the number of tuft cells, and not just an upregulation of the marker

gene, we validated by immunohistochemical analysis that indeed there was an increase in the number of tuft cells in (Figure 2.56, B) AhCre *Pcgf6*$^{-/-}$ mice compared to the controls. These data suggest that PRC1.6 may participate in the regulation of tuft cell differentiation, potentially by RING1B and H2AK119Ub1 independent mechanisms since removal of PRC1.6 didn't have any impact in PRC1 and H2AK119Ub1 occupancy (Figure 2.51).

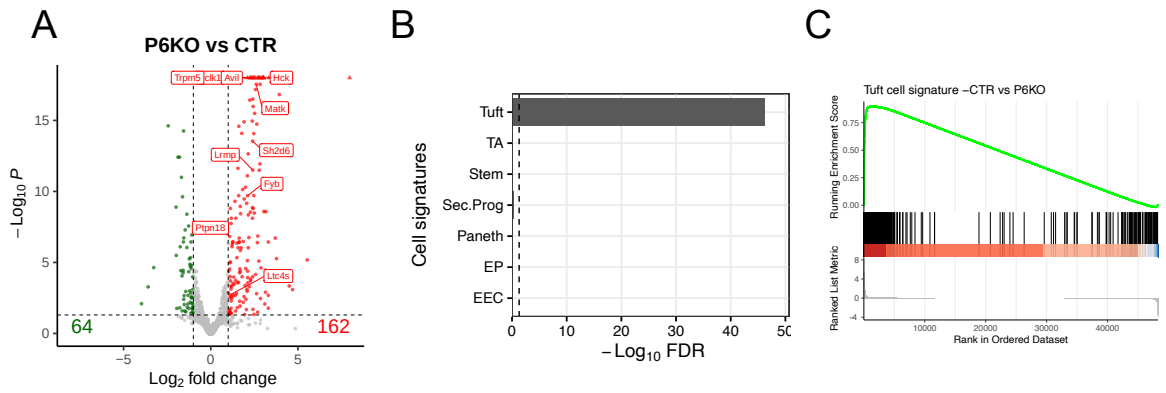


**Figure 2.55: PRC1.6 loss induces the upregulation of tuft cell-related genes**. A) Volcano plots of $-\log_{10}(pvalue)$ against $\log_2(foldchange)$ representing the differences in gene expression between *Pcgf6*KO and WT intestintal crypts for all protein coding genes. B) Enrichment analyses of different gene signatures corresponding to intestinal cell lineages. Dashed line indicates significance threshold (p-value < 0.05). C) GSEA analysis of the tuft cell signature between CTT (WT) and P6KO intestinal crypts.

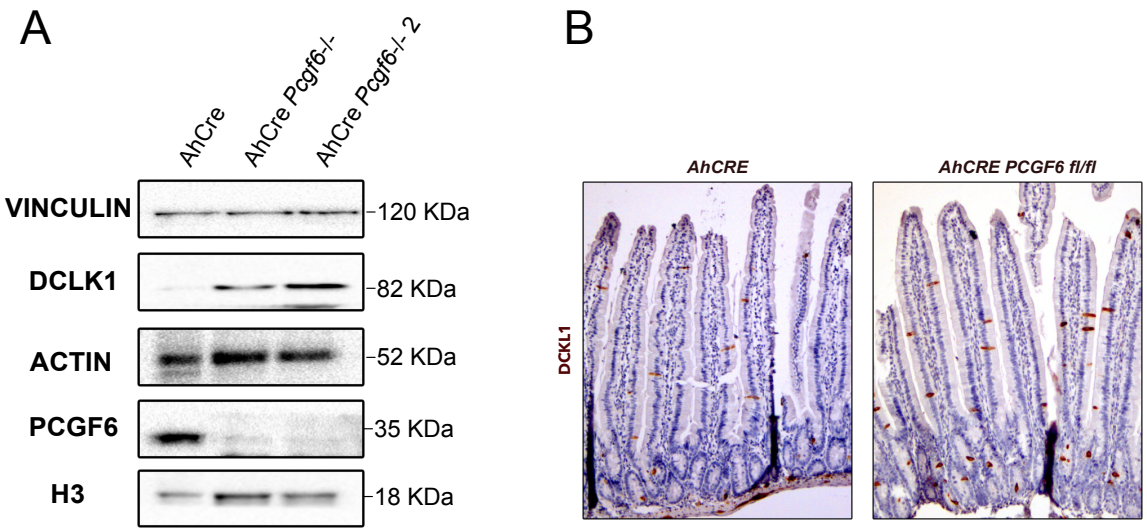


**Figure 2.56: PRC1.6 loss induces tuft hyperplasia**. A) Western blot analysis with the indicated antibodies of protein lysates prepared from AhCre and AhCre *Pcgf6*$^{-/-}$ mice, in duplicate. Vinculin served as loading control. B) Immunohistochemical analysis of AhCre and AhCre *Pcgf6*$^{-/-}$ mice using anti-DCLK1 specific antibody (tuft cells) on small intestinal sections.

# 3 Methods

## 3.1 Workflow managment with snakemake and docker

Organization and reproducibility are the basis of any large and successful bioinformatic project. Multiple command line applications and pipelines are chained together to create an analytical workflow. To be able to reproduce the analyses of any given bioinformatic project, it is very important to be able to re-execute all the code in any given machine, by any given person, and to be able to replicate the same results. To achieve this, in this thesis I have used two main toolkits:

- **Snakemake** [132]: A workflow management system that allows to automate pipelines and to re-execute all the code in the exact same order, with the exact same inputs and outputs by any person, in any machine, in any place.
- **Docker** (`https://www.docker.com/`): This open-source project allows the creation of *containers*, in which all the required software to run a specific workflow can be installed, saved, and shared with anyone. Docker is very useful in the management of bioinformatic workflows by facilitating reproducibility.

By combining a workflow managment system (snakemake) and the use of containers (docker) that contain all the software required to run the workflow, it is very

easy to create bioinformatic pipelines that can be shared to everybody and executed in any machine, allowing to reproduce exactly the same results.

All the snakeamke pipelines described and used for this thesis are publicly available in a github repository:

- ChIP-seq pipeline: https://github.com/dfernandezperez/ChIPseq-snakemake.
- RNA-seq pipeline: https://github.com/dfernandezperez/RNAseq-Snakemake.
- Docker images: https://github.com/dfernandezperez/Docker.

By downloading these snakemake pipelines with their corresponding docker images (and raw data), any person should be able to reproduce the ChIP-seq and RNA-seq data presented in this thesis (peak files, bigwig files and gene differential expression results).

## 3.2 R, tidyverse and bioconductor

All of the analyses presented in this thesis have been produced using the R programming language [133] with tidyverse [134] and bioconductor [135] packages. Tidyverse is a language to solve data science problems in R. It allows to manipulate, tidy, organize and visualize data in a very homogeneous and syntactically elegant way. Bioconductor, instead, offers an ecosystem of R packages to analyze high-throughput data, such as RNA-seq and ChIP-seq.

## 3.3 ChIP-seq analysis

### 3.3.1 Use of a calibrated internal control: Spike in

ChIP-seq is a semi-quantitative technique and, in order to precisely quantify ChIP-seq data, the use of a calibrated internal control is required. Due to the nature of a ChIP-seq, when the abundance of a given histone modification is globally reduced (let's say 60%), this difference can be masked by the normalization [136] if the data is normalized as Counts Per Million (CPM, calculated as $\frac{\text{N reads mapped to loci}}{\text{Total mapped reads}} * 10^6$). In our experience, we have seen that this is particularly relevant with very broad hPTMs, meanhwhile for "sharp" hPTMs (such as H3K27ac) or TFs, the use of a calibrated internal control does not offer extra benefits, in our experience.

The application of a calibrated internal control in ChIP-seq consists in mixing a fixed, previously known amount of external chromatin (from a different organism, for simplicity referred as *Drosophila* from now on) with the target chromatin (referred as *Mouse* from now) where the experiment is going to be performed. In our case, we add a 5% of the amount of *Mouse* chromatin as *Drosophila* chromatin (for example, if we measure 100$\mu$g of *Mouse* chromatin, we add 5$\mu$g of *Drosophila* chromatin). So, if there are no differences in the amount of a particular hPTM (H2AK119Ub1, as an example) between two samples (let's say A and B), we expect to pull down the same proportion of fragments from *Mouse* and *Drosophila* genomes in A and B samples, since we added the same proportion of spike-in respect to the target chromatin. So, the ratio $\frac{\text{reads } \textit{Drosophila}}{\text{reads } \textit{Mouse}}$ will be the same in A and B. If instead, the ratio in A is 10 and in B is 5, it means that B has a higher amount of H2AK119Ub1 (because for the same proportion of *Drosophila* and *Mouse* chromatins, we are pulling down more *Mouse* DNA fragments from sample B than A -because there is more H2AK119Ub1-). We have applied this normalization based on the work from Or-

lando and colleagues [136]. The normalization consists in the following: we want to obtain a normalization factor ($\alpha$) that when applied to each sample, the signal coming from the *Drosophila* genome ($\beta$) is the same across all samples. This can be achieved by applying the formula $\alpha * N_d = \beta$ ($N_d$ is the signal coming from the *Drosophila* genome, which is just the number of reads aligned to it). Since $\beta$ will be the same for all samples, it can be substituted by 1, and $\alpha$ can be calculated as:

$$\alpha = \frac{1}{N_d}$$

However, the previous formula is assuming that we are adding exactly a 5% of the *Mouse* chromatin as *Drosophila* chromatin. This assumption is not always satisfied due to technical variability introduced during chromatin quantification and pipetting processes. To avoid any potential variability, it is possible to sequence the input samples in order to calculate the *real* proportion of *Mouse* and *Drosophila* chromatins in the sample before performing the immunoprecipitation. So, for example, if by mistake we added 2 times more spike-in in sample A than B, if there are no differences in the amount of protein that we are immunoprecipitating, it will look like in sample B there is 50% less signal (because the ratio $\frac{\text{reads } \textit{Drosophila}}{\text{reads } \textit{Mouse}}$ for B will be 1/2 of the ratio in A) . However, if we sequence also the input samples, we will find out that in sample A the starting ratio $\frac{\text{reads } \textit{Drosophila}}{\text{reads } \textit{Mouse}}$ is 2 times bigger than in sample B. We can correct $\alpha$ based on this information and avoid the problem introduced by the error during sample preparation. So, by adding the sequencing of the input samples in the equation to correct $\alpha$ for technical variability, the final formula would be:

$$\alpha = \frac{1}{N_d} * \frac{N_{di}}{N_{mi}} * 10^6$$

where:

- $\alpha$ is the final normalization factor.
- $N_d$ is the number of reads mapped to *Drosophila* genome in the ChIP.
- $N_{di}$ is the number of reads mapped to *Drosophila* genome in the input.
- $N_{mi}$ is the number of reads mapped to *Mouse* genome in the input.
- $10^6$ is a constant multiplier added for convenience to produce calibrated CPM.

Now, multiplying by $\alpha$ the read counts will produce a calibrated and normalized ChIP-seq signal for every sample. We have applied this normalization strategy to all ChIP-seq samples of hPTMs from the results section 2.2, meanwhile for samples from section 2.1 we applied the formula $\alpha = \frac{1}{N_d}$ because the later normalization strategy still was not set up in the lab.

### 3.3.2 Main workflow

The main pipeline and scripts used to pre-process all ChIP-seq samples can be found in `https://github.com/dfernandezperez/ChIPseq-snakemake`. Raw fastq reads were aligned to mm10, or mm10 and dm6 genomes in case the samples contained spike-in, using the bowtie aligner v1.2.2 [137]. Settings for the aligment were left as by default, with the exception of the removal of multimapping reads (using the flag -m). Ambiguous reads that aligned against mm10 and dm6 genomes were discarded using with the script *remove_spikeDups.py*. Then, PCR duplicates were removed using the PICARD toolkit [138] with default parameters. PCR duplicate-free bam files were used as input for MACS2 v2.1.1 [139], which was run with parameters –format BAMPE -p 1e-10 -g mm. In order to avoid any kind of false positive peaks during the peak calling of PCGF1-2-3-6, we used the peaks found in their respective KOs to remove and filter non-specific peaks called by MACS2.

Association of PCGF peaks to CpG islands was done by overlapping PCGF-bound loci, using the ChIPeakAnno R package [140], with a public track of CpG mm10 regions downloaded from UCSC. Also, genomic annotation of PCGF-bound loci into Promoter, 5' UTR, Intron, Exon, 3' UTR, Downstream and Distal Interegenic was also done using ChIPeakAnno. Promoters were considered as windows of $\pm$ 2500bp around TSS.

Generation of quantification files (bigwig) of the ChIP-seq samples and representation of ChIP-seq signal with heatmaps was performed using deeptools [141]. Bigwig samples were created using the function *bamCoverage* with parameters -bs 50 -e. For spiked-in samples, the normalization factor $\alpha$, calculated as described above, was introduced in the function *bamCoverage* using the flag –scaleFactors. Heatmaps were produced with the functions *ComputeMatrix* and *plotHeatmap* with parameters -bs 50 –missingDataAsZero and using a list of blacklist regions from ENCODE to remove sticky regions from the analyses [142]. In order to equalize the scale of all boxplots, heatmaps, and intensity plots, signal intensity was scaled to 0–1 (represented by min-max in the boxplot figures) by applying the formula $\frac{1}{(P98-P5)}$ to all matrices generated by *computeMatrix*.

Enrichment analyses of PCGF peaks was performed using HOMER [143]. A bed file containing $\pm$ 25bp around the peak summits were used as input for HOMER, which was executed with default settings. A q-value of 0.01 was set as threshold to determine enriched motifs. Gene ontology analysis of PCGF target genes was performed using the R package clusterProfiler [144], setting as theshold a q-value and p-value of 0.01.

## 3.4 RNA-seq analysis

### 3.4.1 Removal of PCR duplicates

Removal of PCR duplicates in RNA-seq data is a topic that has been deeply discussed. When most sequencing machines were used to produce single-end reads, it seemed that not removing "PCR duplicates" in RNA-seq data could be more beneficial than doing so [145]. Basically the thesis behind this idea was that usually a very small amount of transcripts (usually housekeeping genes) represent >80% of the total number of reads. For single-end sequencing this can represent a problem, because reading just 50bp from a 500bp fragment means that we are just obtaining 10% of the real information. With the information of the first 50bp of a cDNA fragment, we can't know if two fragments that start at the same nucleotide are the same actual fragment (PCR duplicate) or actually are 2 different fragments (coming from 2 different RNA molecules) that were fragmented in the same 5' end but have different 3' ends (of which we don't have any information because single-end reads are just covering the 5' of the molecule). However, now most Illumina machines produce paired-end reads, which can sequence fragments longer than 1000bp. This means that now, in most cases, we can know the first and last 50bp of any given cDNA fragment. This extra piece of information is crucial in differentiating real PCR duplicates from false. Still there is no literature about the impact of paired-end sequencing in the detection of real and false PCR duplicates in RNA-seq. However, we think that with the information offered by paired-end sequencing the detection of real PCR duplicates is much more reliable. For this, I have decided to eliminate PCR duplicates from the RNA-seq samples analyzed for this thesis.

### 3.4.2 Main workflow

The main pipeline and scripts used to pre-process all RNA-seq samples can be found in https://github.com/dfernandezperez/RNAseq-Snakemake. Raw fastq reads were aligned to the mouse genome mm10 using STAR with default parameters, except for the removal of multimapping reads using the flag –outFilterMultimapNmax 1. PCR duplicates were removed using samblaster with default parameters [146]. PCR duplicate-free bam files were used to quantify gene expression with featureCounts [147] with parameters -s 0 -t exon -g gene_name with GRCm38 annotation (Gencode M21, https://www.gencodegenes.org/mouse/). Once gene counts were quantified, the R package DESeq2 [148] was used to perform library normalization and differential gene expression. DESeq2 normalizes count data by calculating a series of *sizeFactors* for each sample that will be used to normalize the expression of all genes to correct by library size and potential technical variability (by multiplying the gene counts by the sizeFactor). The size factor is calculated as following: First, a pseudo-reference count table is calculated by computing the geometric mean of every gene across the whole dataset, which consists in the following formula:

$$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \sqrt[n]{x_1 x_2 \cdots x_n}$$

Then, once we have generated a psueo-reference (that contains 1 value -geometric mean- per gene), we calculate a ratio for each gene of each sample by dividing the counts of the gene $a$ in the sample $x$ by the geometric mean $z$ of that specific gene from the pseudo-reference. This will give us, for every sample, a vector of ratios. By calculating the median of that vector for each sample, we obtain the *sizeFactor* that will be used to normalize the gene counts of each sample. $Log_2$ fold changes calculated by DESeq2 were corrected using the R package apeglm [149] in order to opti-

mize its calculation for genes with high variance. To increase the statistical power of differential expression testing, while controlling the false discovery rate (FDR), the package IHW [150] was used. This allows to correct the p-vales calculated by DESeq2, which improves the performance for lowly expressed genes. A threshold of $|Log_2$Fold Change$|$ of 1.5 and FDR of 0.05 was applied to annotate genes as differentially expressed. Gene ontology analysis of differentially expressed genes was performed using the R package clusterProfiler [144], setting as theshold a q-value and p-value of 0.01.

## 3.5 Wet lab methods

The data presented here have been produced in collaboration from multiple members from my laboratory. All the *in vitro* work using mESC models has already been published, and all the wet lab methods used to produce the figures presented in this thesis can be accessed in the following articles [151, 152]. Regarding the *in vivo* data, the methods are the following, based on the methods from [91].

### 3.5.1 Animal Procedures

Mice were housed accordingly to the guidelines set out in Commission Recommendation 2007/526/EC, June 18, 2007, on guidelines for the accommodation and care of animals used for experimental and other scientific purposes. All experiments were performed in accordance with the Italian Laws (D.L.vo 116/92 and following additions), which enforces EU 86/609 Directive (Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations, and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes).

### 3.5.2 Tissue morphology and immunohistochemistry

Intestines were dissected immediately after death. The proximal (duodenum), central (jejunum) and distal (ileum) parts of the small intestine were flushed with cold PBS, opened longitudinally and fixed overnight in 4% formaldehyde. The colon was opened longitudinally and processed following a "Swiss roll" protocol [153]. Samples were embedded in paraffin and sectioned at 5 µm thickness. Morphological analyses were performed using haematoxylin and eosin or Alcian blue staining. Alkaline phosphatase staining used the Alkaline Phosphatase Kit (Sigma) following manufacturer's instructions. Immunohistochemistry analyses were mediated by proteinase K or with heat-induced antigen unmasking using 10 mM sodium citrate. Anti-lysozyme (Dako, A0099), anti-DCLK1 (Cell Signalling 62257) and anti-KI67 (Abcam, ab16667) were used as primary antibodies. All images were acquired using an Olympus BX51 bright field microscopy and Leica TCS SP2 laser scanning confocal microscope.

# 4 Discussion

## 4.1 PRC1 sub-complexes display distinct functional features

Here we have profiled for the first time the genome-wide occupancy of all PRC1 sub-complexes in mESC. Our data showed that, despite sharing an important amount of target genes (Figure 2.5), PCGFs retained binding specificity. Not just their geome-wide occupancy was different, but also the functional role of their target genes was specific for different sub-complexes (Figure 2.9). The exception to this was PCGF2, which overlapped 80% with PCGF1. In part this can be explained by the fact that PCGF1 showed the highest number of binding sites (5261, which overlapped also 60% with PCGF6), probably due to the ability of its subunit KDM2B to recognize un-methylated CpG islands [51, 59]. Although in some cases, as for PCGF6, the number of specific binding sites was low (17%), I want to stress the specificity of PRC1.6 for those regions, since RING1B recruitment and H2AK119Ub1 levels at those loci was exclusively compromised in *Pcgf6*KO cells and unaltered in any other KO (Figures 2.19 and 2.20).

Interestingly, different PCGF-bound loci correlated with different transcriptional states. Genes bound by vPRC1 complexes showed a permissive transcriptional state

(Figure 2.8), meanwhile any gene targeted by cPRC1 (or a combination of cPRC1 + vPRC1) showed very low transcriptional levels. At first glance this could be read as cPRC1 or PRC2 establishing repression, however, removal of cPRC1 and PRC2 from mESC didn't alter H2AK119Ub1 and transcription at those genes. Actually, the only single PRC1 sub-complex that showed some kind of specific gene repression was PRC1.6, since in the absence of PCGF6, 20% of its targets were upregulated (Figure 2.18). We speculate that probably we did not observe gene upregulation upon *Pcgf* deletion for two reasons: 1. Because most of PCGF-specific targets (vPRC1 targets, mainly) were already transcriptional permissive, with very low PRC1 and PRC2 occupancy (RING1B, SUZ12, H2AK119Ub1 and H3K27me3, Figure 2.6). Thus, removing the small amount of PcG from those sites would be expected to have minimal effects on gene transcription. 2. We think that there is an evident enzymatic compensation across different PRC1 sub-complexes, in particular at PCGF1/2/6 loci. Indeed, by knocking out both vPRC1.1 and cPRC1, which shared 80% of cPRC1 targets, RING1B occupancy and H2AK119Ub1 levels decreased dramatically (Figure 2.19). This demonstrated that H2AK119Ub1 levels are maintained by the coordinated activity of different PRC1 sub-complexes, in accordance with the work of Fursova and colleagues [129]. In this specific case, is interesting to highlight the low participation of PRC1.6 (PCGF6) in the recruitment of RING1B and H2AK119Ub1 deposition at PCGF1/2/6 loci (Figure 2.19). However, the absence of gene upregulation in *Pcgf124*KO cells (Figure 2.15), despite the strong reduction in H2AK119Ub1, suggests that these low levels of H2AK119Ub1, maintained by PRC1.6 and PRC1.3/5, were sufficient to keep gene repression since later we showed that complete removal of H2AK119Ub1 let to PcG target gene upregulation (Figure 2.33, panel B).

## 4.2 PCGF1/2/4 and PCGF3/5/6 modules regulate different biological processes

Based on their genome-wide binding pattern and the functional role of their target genes, we grouped the 6 PCGFs in two different modules: the PCGF1/2/4 and PCGF3/5/6 modules. PCGF1/2/4 were characterized by a strong genome-wide overlap, and by the similarities in the function of their target genes -mainly developmental genes- (Figures 2.5 and 2.9). PCGF3/5/6 were characterized by regulating specific biological processes (such as authophagy and germ-cell related functions, Figure 2.9) and by not requiring RING1A/B to bind chromatin and recruit PRC1 ancillary proteins (Figure 2.24). Removal of the PCGF1/2/4 module didn't have any effect on cell viability, indeed, changes in gene expression were marginal when compared to single KOs (Figures 2.18 and 2.14). This was surprising taking into account the strong displacement of RING1B from chromatin and the reduction in H2AK119Ub1 levels. However, as discussed above, probably those low levels of RING1B occupancy and H2AK119Ub1 are enough to maintain gene repression, which is in line with recently published data [129] and the fact that just removal of RING1B (and not RING1A) or using the I53A RING1B instead of I53S does not have the same de-repressive effect as a full KO of PRC1 [85, 124].

Regarding the PCGF3/5/6 module, its absence in mESC induced the acquisition of a fibroblast-like phenotype. This was evident by the shape of the cells (Figure 2.13) and by the strong upregulation of keratins, collagens and cellular matrix related genes (Figure 2.16). Since global PRC1 and PRC2 activity was minimally compromised in these cells (Figure 2.20), and PCGF3 and PCGF6 could bind chromatin in the absence of RING1A/B, we speculate that part of this phenotype could be caused by RING1A/B-independent functions of both PCGFs. For example, PCGF6 binds to

E-box motifs through its MAX subunit, which is also able to form an heterodimer with MYC [128]. A competition bewteen PCGF6 and MYC for MAX and their specific target loci, which would not depend on RING1A/B activity, could determine the transcriptional status of those genes. Recent data from Tanaskovic et al. [154] (not peer-reviewed) showed how this competition between PCGF6 and MYC can have a key role in regulating MYC-induced lymphomagenesis. More experiments should be done in order to asses if PCGF3 and PCGF6 are carrying out RING1A/B-independent functions, but we think that our data, at least, suggest the possibility of the existence of such mechanisms in mESC.

## 4.3 PCGF6 recruitment to chromatin: Cooperation between E2F and E-box motifs

Recent published data [60, 155] showed that MGA was required for PRC1.6 recruitment to chromatin since in *Mga*KO mESCs PRC1.6 was not able to bind chromatin anymore. It was concluded that the DNA-binding properties of MGA were essential for PCGF6 binding and that E2F6 and L3MBTL2 subunits were required to enhance PRC1.6 recruitment. Indeed, we observed a strong overlap between MGA and PCG6 (Figure 2.25) and we were also able to reproduce the displacement of PRC1.6 from chromatin in the absence of MGA (Figure 2.27). In fact, when we checked the assembly of the complex in the absence of MGA, we saw that PCGF6 was not able to interact anymore with the rest of the complex (Figure 2.26). However, if instead of downregulating the whole protein we removed just a smal C-terminal region of the HLH domain, which allows MGA to bind chromatin, then PCGF6 signal was just slightly reduced (Figure 2.27) and was also able to assemble the whole complex. A similar displacement from chromatin was observed upon E2F6 knock-down, and,

when the knock-down was performed in the *Mga*ΔHLH cell line, PRC1.6 was further displaced (Figure 2.27). These data demonstrate that MGA, apart from its role in recruiting PRC1.6 to chromatin, is actually acting as a scaffold protein in the complex. Also, we showed that the recruitment of PRC1.6 to chromatin is the product of a cooperation of different DNA-binding activities from multiple subunits (MGA, MAX, E2F6, L3MBTL2) instead of just the activity of MGA enhanced by other subunits [60, 155].

## 4.4 Potential recruitment to chromatin of PRC1.3/5 by USF1/2

The mechanism of vPRC1.3/5 recruitment to chromatin is still missing. It is the only PRC sub-complex without a clear ancillary protein that mediates its tethering to chromatin. Many roles for vPRC1.3/5 have been proposed. Recently, Fursova et al. [129] has proposed a role for vPRC1.3/5 in establishing a low-level H2AK119Ub1 genome-wide "blanket", which would be dispensable for transcriptional repression. Other works in this line, where vPRC1.3/5 is important to establish megabase-sized H2AK119Ub1 domains, showed that this PRC1 sub-compelxes are key in the establishment of X chromosome silencing by *Xist*. On the other side, there are also very interesting data in which vPRC1.3/5 is associated with transcriptional activation [156, 157], in particular [156], where a subunit from vPRC1.3/5 complex, AUTS2, is able to recruit p300 and mediate the activation of transcription. As reported in [129], we observed that PCGF3/5 are not required for mESC homeostasis and to maintain PRC1-mediated gene repression. However, we didn't observe a genome-wide reduction in H2AK119Ub1 levels upon removal of PRC1.3/5 (Figure 2.12). One explanation for this striking difference would be the miss-calibration of our spike-in control in

some of those samples. As explained in the methods, when performing a calibrated (spike-in) ChIP-seq, it is very important to correct for any technical variability introduced in the protocol. This can be very easily fixed by sequencing the input samples of each IP, which allows to precisely quantify the relationship between the spike-in genome (*Drosophila*) and the target genome (mouse). While [129] applied this technique to their experiments, we did not. In fact, recently we replicated the same experiments with a proper calibration [130] and we observed a genome-wide H2AK119Ub1 reduction upon *Pcgf3/5*KO.

In line with the secondary role of PCGF3 in gene activation [156, 157], we found that PCGF3 targets were very actively transcribed (Figure 2.8). Also, we noticed that an E-box motif (CACGTG), corresponding to USF1/2, was extremely enriched in PCGF3-bound loci. This motif was very similar to motifs recognized by MYC (TCACGTG), which could explain the overlap between PCGF3 and PCGF6. Under the hypothesis of a potential connection between PCGF3 and USF1, genome-wide analysis of USF1-bound loci showed a strong overlap with PCGF3 (88%, Figure 2.28). Moving forward, we showed that, upon *shUSF1*, PCGF3 was strongly displaced from chromatin (Figure 2.29), in accordance with the hypothesis of USF1 driving PRC1.3 recruitment to chromatin. Since USF1 is a TF associated with transcriptional activation [158], we propose a model in which USF1 drives PRC1.3/5 to chromatin, who through its AUTS2 subunit then recruits p300 and promotes the activation of transcription.

## 4.5 H2AK119Ub1 is the driver of PcG-mediated gene repression

How PcG proteins establish and maintain repression is still a matter of debate [50]. Previous works have proposed that H2AK119Ub1 could play a minimal role in maintaining PcG-dependent repression [84, 85], meanwhile others proposed the opposite [124, 127]. To assess whether the catalytic activity of a protein complex is involved in a particular biological process is not a trivial task. *In vitro* assays are required to find mutations that render the complex inactive while not altering its 3D conformation and other non-catalytic properties, such as recruitment to chromatin. Also, the quantification of the activity has to be feasible and reliable *in vivo* in order to properly determine the role of the catalysis in regulating a specific biological process. My lab has developed a series of genetically engineered mESC lines that allowed us to specifically asses the role of H2AK119Ub1 in establishing PcG-mediated repression. Our model expresses, in a RING1A-null background, a fully catalytically inactive RING1B (I53S) that, coupled with genome-wide analyses of PcG chromatin occupancy and quantification of H2AK119Ub1 and H3K27me3 levels, allowed us determine that H2AK119Ub1 is the central hub of PcG-mediated repression in mESC. Very importantly, this acute model also avoids any potential adaptations that can take place in stable cell lines [129, 151] and is able to assemble all PRC1 sub-complexes and to be recruited to chromatin as a WT RING1B.

Our results demonstrate that, although I53S PRC1 was able to bind chromatin, in the absence of its catalytic activity is not able to maintain gene repression. This effect was not due to a displacement of cPRC1 from chromatin, since *Pcgf24*KO mESC did not show any sign of gene activation, and neither due to a displacement of PRC2, since *Eed*KO (and lose of H3K27me3) neither induced a gene de-repression, in line

with previous works from our lab and others [77, 86]. This places H2AK119Ub1 as the main mechanisms by which PcG proteins maintain gene repression in mESC. As discussed above, these data contrast with previous *in vivo* works recently published. We think that the delay in embryonic lethality presented in [85] doesn't properly answer to the question if H2AK119Ub1 is really involved in maintaining gene repression. The mice used in that study had perfectly functional RING1A-PRC1 complexes, and the RING1B mutation used to inactivate the complex (I53A) was later shown to be hypomorphic [127]. However, studies published in *Drosophila Melanogaster* models, where the role of H2AK118Ub1 (in mammals H2AK119Ub1) in regulating homeotic genes was assessed, showed that this system is more PRC2 oriented, since H3K27me3, opposite to H2AK118Ub1, seems to be crucial for this process [159]. Although H2AK118Ub1 is also recognized by PRC2 in flies, its removal does not have a strong effect in H3K27me3 levels, which suggests that PRC2.1 could play a more important role there. PRC1 has evolved by increasing its complexity from flies to mammals, going from 1 type of vPRC1 (equivalent of vPRC1.1) to 6 different sub-complexes. We speculate that, maybe, in mammals, through the development of more vPRC1 complexes that are more catalytically active than cPRC1, PcG proteins could have moved from a PRC2/H3K27me3-oriented system into a PRC1/H2AK119Ub1. A potential motivation for this change could be the more variety offered by vPRC1 regarding chromatin recruitment and function when compared to PRC2 and cPRC1.

Variant versus canonical recruitment of PcG proteins to chromatin, and the establishment of chromatin domains, have also been deeply discussed [50, 61, 160]. Our data point towards a leader role for PRC1 in the establishment of PcG domains in mESC, since abolishing H2AK119Ub1 caused a strong displacement of PRC2 from chromatin, in particular PRC2.2 (Figure 2.40). If we apply the same mechanics in the

opposite direction (thus, removing PRC2 or H3K27me3 from mESC) RING1B signal is reduced (due to loss of cPRC1 binding) but H2AK119Ub1 and transcriptional repression are maintained [86]. This means that H2AK119Ub1 is acting as a central PcG hub, which maintains the whole system tight at promoter regions of PcG target genes. Remainder PRC2 and H3K27me3 signal in H2AK119Ub1-null mESC is attributed to PRC2.1, which, contrary to what seems to happen in flies [84, 159], is not enough to maintain high H3K27me3 levels and gene repression. This strong displacement of PRC2.2 in the absence of H2AK119Ub1 fits well with the role of JARID2 and AEBP2 in recognizing PRC1-decorated chromatin [63, 64, 161].

Our data showed, interestingly, that also vPRC1 and PRC2.1 were displaced from chromatin in the absence of H2AK119Ub1 (Figures 2.44 and 2.36). These complexes are recruited to chromatin via H2AK119Ub1 and H3K27me3 independent mechanisms [50] and, at first, they could be expected to be minimally affected by the loss of PRC1 and PRC2 decorated chromatin. This reduction in chromatin occupancy could be partially explained by the fact that both vPRC1 and PRC2.1 catalytic products are able to stabilize them on chromatin, respectively [73, 76]. So, since H2AK119Ub1 is lost and H3K27me3 are reduced, this could led to a destabilization of both complexes from chromatin. Another possibility could be that the same activation of gene expression, in particular Pol II activity, could push PcG away from chromatin. We have shown that most of vPRC1 binding sites (not shared with cPRC1), show very low RING1B occupancy and very low H2AK119Ub1 levels (Figure 2.6). Those genes tended to be moderately transcribed, meaning that vPRC1 can bind, with medium-low affinity, active genes without being able to deposit high levels of H2AK119Ub1. We speculate that the reason why vPRC1 is displaced from chromatin in the absence of H2AK119Ub1 is the same, or very similar, to the reason why vPRC1 is not able to establish rich PcG domains, with high PRC2 and PRC1 occupancy, at those subsets

of active target genes (potentially due to competition with Pol II activity). These data support the idea of CpG-promiscuous PcG proteins that bind chromatin and establish PcG domains in the absence of transcription, which is something that was already proposed a few years ago [61, 77, 162]. Based on this idea, vPRC1 and PRC2.1 would be ”sampling” CpG-rich loci until they find a locus that is not being transcribed anymore. In that moment, through the canonical and variant recruitment mechanisms, a positive feedback loop would allow a fast establishment of a reliable, large and solid PcG domain that would help to maintain specific loci silenced and protected against spurious transcriptional activation. This system allocates PcG as ”responsive” and not ”proactive” protein complexes, that instead of dictating the transcriptional status of a gene, they help to maintain gene repression when the expression of those loci is not required anymore, in a given time, by the cell.

Although our work has clarified the roles of H2AK119Ub1 and H3K27me3 in establishing gene repression in mESC, still the biochemical mechanism by which this repression is obtained is not clear. There are many hypothesis: as discussed above, competition with Pol II for chromatin [77, 162], recruitment of binder proteins that promote gene repression [63, 163] or by not allowing the deposition of histone marks that enhance transcription [164, 165]. More work in this direction will be needed in order to understand the biochemical mechanisms involved in PcG-mediated gene repression.

## 4.6 Different PRC1 sub-complexes orchestrate PRC1 function in intestinal epithelium

Studying the functional role of individual PRC1 sub-complexes *in vivo* is a very important part of our research activity. We first characterized these complexes *in vitro*

using mESC models, which allowed us to better understand how PRC1 complexes cooperate in order to establish PcG-mediated gene repression. However, it is also important to extend this characterization to different cellular contexts, in particular to *in vivo* systems that represent a much more challenging environment. We think that this is particularly important to study PcG proteins due to their role in maintaining the transcriptional identity of cells. In these systems cells are constantly exposed to external stimuli and insults that trigger changes in gene expression, which does not happen in *in vitro* systems. PRC1 plays a key role in maintaining the cellular identity of mouse intestinal stem cells, and is required to maintain intestinal homeostasis independently of its role in regulating CDKN2A expression [91]. Here, by producing multiple mouse transgenic models that allowed us to acutely delete specific PRC1 sub-complexes across the whole intestinal epithelium (AhCre models) or just in intestinal stem cells ($Lgr5^{EGFP-IRES-CreERT2-R26R-lacZ}$ models), we showed that the phenotype associated to the loss of PRC1 and H2AK119Ub1 in the intestinal epithelium can't be phenocopied by the loss of individual PRC1 sub-complexes. This means that the role of PRC1 in maintaining intestinal homeostasis is the result of the cooperation of the activity of different PRC1 sub-complexes.

As we observed in mESC, in the intestinal epithelium PCGF proteins also retained binding specificity. In particular, vPRC1 was more associated with "unique loci" (so, genomic regions bound just by 1 PCGF), low H2AK119Ub1 and H3K27me3 and permissive/active transcription. Meanwhile, cPRC1 was associated with high H2AK119Ub1 and H3K27me3, gene repression, and the loci bound by PCGF2 or PCGF4 were also shared with vPRC1 complexes (Figures 2.46 and 2.47). We think that these data support the idea of vPRC1 complexes "sampling" chromatin, looking for transcriptionally inactive promoters that allow PRC1 and PRC2 to establish PcG repressive domains [61]. vPRC1 complexes, due to their biochemical nature, are

catalytically superior to cPRC1 complexes and also can potentially bind to a wider variety of genomic loci [51, 74, 151]. However, our data systematically showed that in mESC and mouse intestinal epithelium, vPRC1-bound loci were usually associated with low H2AK119Ub1 and H3K27me3 levels, except if also cPRC1 occupied those loci. We hypothesize that the explanation to this, at first counter-intuitive behaviour, is that vPRC1 complexes are just able to deposit high levels of H2AK119Ub1 if transcription is already silenced, which then stimulates the recruitment of PRC2.2, deposition of H3K27me3, and finally the recruitment of cPRC1. Then, the association of cPRC1 with repressed transcription would be a consequence of transcriptional silencing, rather than the cause.

In agreement with our data from mESC, single and acute deletion of PRC1 subcomplexes did not phenocopy the loss of PRC1 activity and H2AK119Ub1 at the transcriptional level (Figure 2.50), neither regarding H2AK119Ub1 deposition. Just vPRC1.3/5 demonstrated to be important to maintain intergenic H2AK119Ub1 levels, which were strongly decreased at distal sites. Importantly, this reduction was followed by a milder loss of H3K27me3 at intergenic regions, in accordance with H2AK119Ub1's role in recruiting PRC2.2 and in agreement with recent data from Fursova and collegues [129]. Still, this reduction in intergenic H2AK119Ub1 and H3K27me3 dit not have any effect at the phenotypic level, neither at transcriptomic level. This indicates that probably H2AK119Ub1 levels at promoter regions, and not intergenic, are important to maintain gene repression. These data are in agreement with recent work from our lab, where intergenic increase of H2AK119Ub1 caused gene de-repression of PcG target genes [130]. This gene de-repression was caused by a displacement of PcG proteins from promoter regions to intergenic loci, stimulated by the increase of H2AK119Ub1 and H3K27me3 at intergenic regions.

We showed in mESC that vPRC1.6 was important for the regulation of germ cell related genes. We did not find genes related to these biological processes among PRC1.6 targets in the intestinal tissue, instead, we found processes related to cell cycle and metabolism, which are also regulated by MYC (which binds to the same E-box motif as PRC1.6) [128]. Interestingly, PRC1.6-loss induced a strong expansion of tuft cells, which was confirmed by transcriptomic, WB and histological data (Figures 2.55 and 2.56). This role of vPRC1.6 in regulating tuft cell differentiation could potentially be H2AK119Ub1 independent, since we did not observe a reduction of H2AK119Ub1 levels in the absence of PCGF6. Speculating, a potential mechanism by which vPRC1.6 could affect gene expression in a H2AK119Ub1-independent manner could be by competing with other factors for E-box motifs, as we already discussed above. Some examples of this competition have already been described in MYC-induced lymphomagenesis [154].

Removal of PRC2 in the intestinal epithelium causes a strong reduction in cell proliferation, as well as an expansion of goblet cells [117]. Based on the strong phenotype of PRC1-loss in mouse intestinal stem cells [91], it was not clear if the effect of PRC2 loss could be a PRC2 independent mechanism or could be PRC1 dependent (through cPRC1). Based on our data, absence of cPRC1 in mouse intestine produced a strong reduction in RING1B signal, however, as observed in mESC, this was not corresponded with a reduction in H2AK119Ub1, again demonstrating that vPRC1 complexes play a more important role at decorating chromatin genome-wide with H2AK119Ub1. This, together with our histological data, demonstrates that the phenotype caused by the loss of PRC2 and H3K27me3 was not induced by an indirect displacement of cPRC1 from chromatin, meaning that the repression of the *Cdkn2a* locus by PRC2 is a PRC1 independent mechanism. These data, to some extent, contradict our mESC data regarding H2AK119Ub1. In that system, removal of PRC2 did

not cause any changes in gene expression, neither at the phenotypical level. Our data expose two different environments in which PRC1 and PRC2 roles in maintaining gene repression are slightly different. Gene repression in mESC seems to be PRC1 dependent and PRC2 independent, meanwhile in mouse intestinal cells is PRC1 dependent and PRC2 dependent (although the upregulation of gene expression and effect on cell homeostasis is much lower in the case of PRC2). Based on this, in some adult tissues, PRC1 and PRC2 could mediate gene repression by different and independent mechanisms, and the combination of both would allow the establishment of solid PcG repressive domains. This would be in agreement with recent data from epidermal tissue, where PRC1 and PRC2 act redundantly to maintain stem cell identity [166].

# Acronyms

| | |
|---|---|
| CBC | Crypt Columnar Cell |
| CGIs | CpG Islands |
| COMPASS | Complex Proteins Associated with Set1 |
| Counts Per Million | Counts per Million |
| cPRC1 | Canonical PRC1 |
| ESCs | Embryonic Stem Cells |
| False Discovery Rate | False Discovery Rate |
| GO | Gene Ontology |
| GTF | General Trancription Factor |
| HAT | Histone Acetyltransferase |
| HDAC | Histone Deacetylase |
| HE | Hematoxylin-eosin |
| HKMT | Lysine Methyltransferase |
| iBAQ | Intensity-based Absolute Quantification |
| IP | Immunoprecipitation |
| KO | Knock-out |
| LADs | Lamina-Associated Domains |
| lncRNA | Long non-coding RNA |
| mESC | Mouse Embryonic Stem Cells |

| | |
|---|---|
| NGS | Next-Generation Sequencing |
| OHT | 4-hydroxy tamoxifen |
| PCA | Principal Component Analysis |
| PcG | Polycomb Group |
| PCGF | Polycomb Group RING Finger Protein |
| Pol II | RNA polymerase II |
| PRC1 | Polycomb Repressive Complex 1 |
| PRC2 | Polycomb Repressive Complex 2 |
| PRE | Polycomb Responsive Element |
| PTM | Post-translational modification |
| qRT-PCR | Quantitative Real-Time PCR |
| SWI/SNF | SWItch/Sucrose Non Fermentable |
| TES | Transcription End Site |
| TF | Transcription Factor |
| TrxG | Thritorax Group |
| TSS | Transcription Start Site |
| vPRC1 | Variant PRC1 |
| WB | Western Blot |
| WT | Wild type |

# Bibliography


1. V. Haberle and A. Stark. "Eukaryotic core promoters and the functional basis of transcription initiation". *Nature Reviews Molecular Cell Biology* 19, 10 2018, pp. 621–637. ISSN: 14710080. DOI: 10.1038/s41580-018-0028-8.

2. F. Spitz and E. E. Furlong. "Transcription factors: from enhancer binding to developmental control". *Nature Reviews Genetics 2012 13:9* 13, 9 2012, pp. 613–626. ISSN: 1471-0064. DOI: 10.1038/nrg3207.

3. M. A. Zabidi and A. Stark. "Regulatory Enhancer–Core-Promoter Communication via Transcription Factors and Cofactors". *Trends in Genetics* 32, 12 2016, pp. 801–814. ISSN: 0168-9525. DOI: 10.1016/J.TIG.2016.10.003.

4. V. Haberle and A. Stark. "Eukaryotic core promoters and the functional basis of transcription initiation". *Nature reviews Molecular cell biology* 19:10, 2018, pp. 621–637.

5. S. Saxonov, P. Berg, and D. L. Brutlag. "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters". *Proceedings of the National Academy of Sciences* 103, 5 2006, pp. 1412–1417. ISSN: 0027-8424. DOI: 10.1073/PNAS.0510310103.

6. R. S. Illingworth, U. Gruenewald-Schneider, S. Webb, A. R. Kerr, K. D. James, D. J. Turner, C. Smith, D. J. Harrison, R. Andrews, and A. P. Bird. "Orphan CpG

Islands Identify Numerous Conserved Promoters in the Mammalian Genome". *PLOS Genetics* 6, 9 2010, e1001134. ISSN: 1553-7404. DOI: 10.1371/JOURNAL.PGEN.1001134.

7. J. M. Landolin, D. S. Johnson, N. D. Trinklein, S. F. Aldred, C. Medina, H. Shulha, Z. Weng, and R. M. Myers. "Sequence features that drive human promoter function and tissue specificity". *Genome Research* 20, 7 2010, p. 890. ISSN: 10889051. DOI: 10.1101/GR.100370.109.

8. J. M. Rozenberg, A. Shlyakhtenko, K. Glass, V. Rishi, M. V. Myakishev, P. C. FitzGerald, and C. Vinson. "All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues". *BMC Genomics* 9, 1 2008, pp. 1–13. ISSN: 14712164. DOI: 10.1186/1471-2164-9-67/FIGURES/8.

9. A. M. Deaton and A. Bird. "CpG islands and the regulation of transcription". *Genes & Development* 25, 10 2011, p. 1010. ISSN: 08909369. DOI: 10.1101/GAD.2037511.

10. F. Mohn, M. Weber, M. Rebhan, T. C. Roloff, J. Richter, M. B. Stadler, M. Bibel, and D. Schübeler. "Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors". *Molecular Cell* 30, 6 2008, pp. 755–766. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2008.05.007.

11. P. A. Jones and S. B. Baylin. "The Epigenomics of Cancer". *Cell* 128, 4 2007, pp. 683–692. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2007.01.029.

12. T. C. Voss and G. L. Hager. "Dynamic regulation of transcriptional states by chromatin and transcription factors". *Nature Reviews Genetics 2013 15:2* 15, 2 2013, pp. 69–81. ISSN: 1471-0064. DOI: 10.1038/nrg3623.

13. P. Oudet, M. Gross-Bellard, and P. Chambon. "Electron microscopic and biochemical evidence that chromatin structure is a repeating unit". *Cell* 4, 4 1975, pp. 281–300. ISSN: 0092-8674. DOI: 10.1016/0092-8674(75)90149-X.

14. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. "Crystal structure of the nucleosome core particle at 2.8 Å resolution". *Nature 1997 389:6648* 389, 6648 1997, pp. 251–260. ISSN: 1476-4687. DOI: 10.1038/38444.

15. S. L. Klemm, Z. Shipony, and W. J. Greenleaf. "Chromatin accessibility and the regulatory epigenome". *Nature Reviews Genetics 2018 20:4* 20, 4 2019, pp. 207–220. ISSN: 1471-0064. DOI: 10.1038/s41576-018-0089-8.

16. S. Aranda, G. Mas, and L. D. Croce. "Regulation of gene transcription by Polycomb proteins". *Science Advances* 1, 11 2015. ISSN: 23752548. DOI: 10.1126/SCIADV.1500737/ASSET/515C2088-6DF1-4C27-BBA8-2030A6DA417B/ASSETS/GRAPHIC/1500737-F5.JPEG.

17. M. G. Poirier, M. Bussiek, J. Langowski, and J. Widom. "Spontaneous Access to DNA Target Sites in Folded Chromatin Fibers". *Journal of Molecular Biology* 379, 4 2008, pp. 772–786. ISSN: 0022-2836. DOI: 10.1016/J.JMB.2008.04.025.

18. R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutyavin, B. Lajoie, B. K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J.

Dekker, G. E. Crawford, and J. A. Stamatoyannopoulos. "The accessible chromatin landscape of the human genome". *Nature 2012 489:7414* 489, 7414 2012, pp. 75–82. ISSN: 1476-4687. DOI: 10.1038/nature11232.

19. J. Riposo and J. Mozziconacci. "Nucleosome positioning and nucleosome stacking: two faces of the same coin". *Molecular BioSystems* 8, 4 2012, pp. 1172–1178. ISSN: 1742-2051. DOI: 10.1039/C2MB05407H.

20. A. Valouev, S. M. Johnson, S. D. Boyd, C. L. Smith, A. Z. Fire, and A. Sidow. "Determinants of nucleosome organization in primary human cells". *Nature 2011 474:7352* 474, 7352 2011, pp. 516–520. ISSN: 1476-4687. DOI: 10.1038/nature10002.

21. A. M. Deaton, M. Gómez-Rodríguez, J. Mieczkowski, M. Y. Tolstorukov, S. Kundu, R. I. Sadreyev, L. E. Jansen, and R. E. Kingston. "Enhancer regions show high histone H3.3 turnover that changes during differentiation". *eLife* 5, JUN2016 2016. ISSN: 2050084X. DOI: 10.7554/ELIFE.15316.

22. E. Splinter and W. D. Laat. "The complex transcription regulatory landscape of our genome: control in three dimensions". *The EMBO Journal* 30, 21 2011, pp. 4345–4355. ISSN: 1460-2075. DOI: 10.1038/EMBOJ.2011.344.

23. T. Misteli and E. Soutoglou. "The emerging role of nuclear architecture in DNA repair and genome maintenance". *Nature Reviews Molecular Cell Biology 2009 10:4* 10, 4 2009, pp. 243–254. ISSN: 1471-0080. DOI: 10.1038/nrm2651.

24. W. Deng and G. A. Blobel. "Do chromatin loops provide epigenetic gene expression states?" *Current Opinion in Genetics & Development* 20, 5 2010, pp. 548–554. ISSN: 0959-437X. DOI: 10.1016/J.GDE.2010.06.007.

25. J. Dekker and T. Misteli. "Long-Range Chromatin Interactions". *Cold Spring Harbor Perspectives in Biology* 7, 10 2015, a019356. ISSN: 19430264. DOI: 10.1101/CSHPERSPECT.A019356.

26. L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. D. Klein, L. Wessels, W. D. Laat, and B. V. Steensel. "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions". *Nature* 453, 7197 2008, pp. 948–951. ISSN: 1476-4687. DOI: 10.1038/NATURE06947.

27. C. H. Eskiw and P. Fraser. "Ultrastructural study of transcription factories in mouse erythroblasts". *Journal of Cell Science* 124, 21 2011, pp. 3676–3683. ISSN: 00219533. DOI: 10.1242/JCS.087981/-/DC1.

28. Y. Jeong, K. El-Jaick, E. Roessler, M. Muenke, and D. J. Epstein. "A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers". *Development* 133, 4 2006, pp. 761–772. ISSN: 0950-1991. DOI: 10.1242/DEV.02239.

29. T. Montavon, N. Soshnikova, B. Mascrez, E. Joye, L. Thevenet, E. Splinter, W. D. Laat, F. Spitz, and D. Duboule. "A regulatory archipelago controls hox genes transcription in digits". *Cell* 147, 5 2011, pp. 1132–1145. ISSN: 10974172. DOI: 10.1016/J.CELL.2011.10.023/ATTACHMENT/8C25BB94-A43B-4FA7-99FA-7AF3338B6FEC/MMC1.PDF.

30. V. G. ALLFREY, R. FAULKNER, and A. E. MIRSKY. "ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS". *Proceedings of the National Academy of Sciences of the United States of America* 51, 5 1964, p. 786. ISSN: 00278424. DOI: 10.1073/PNAS.51.5.786.

31. A. J. Bannister and T. Kouzarides. "Regulation of chromatin by histone modifications". *Cell Research 2011 21:3* 21, 3 2011, pp. 381–395. ISSN: 1748-7838. DOI: 10.1038/cr.2011.22.

32. M. Yun, J. Wu, J. L. Workman, and B. Li. "Readers of histone modifications". *Cell Research* 21, 4 2011, p. 564. ISSN: 10010602. DOI: 10.1038/CR.2011.42.

33. M. Lawrence, S. Daujat, and R. Schneider. "Lateral Thinking: How Histone Modifications Regulate Gene Expression". *Trends in Genetics* 32, 1 2016, pp. 42–56. ISSN: 0168-9525. DOI: 10.1016/J.TIG.2015.10.007.

34. S. Henikoff and A. Shilatifard. "Histone modification: cause or cog?" *Trends in Genetics* 27, 10 2011, pp. 389–396. ISSN: 0168-9525. DOI: 10.1016/J.TIG.2011.06.006.

35. X. Ling, T. A. Harkness, M. C. Schultz, G. Fisher-Adams, and M. Grunstein. "Yeast histone H3 and H4 amino termini are important for nucleosome assembly in vivo and in vitro: redundant and position-independent functions in assembly but not in gene regulation." *Genes & Development* 10, 6 1996, pp. 686–699. ISSN: 0890-9369. DOI: 10.1101/gad.10.6.686.

36. D. M. PHILLIPS. "The presence of acetyl groups in histones". *Biochemical Journal* 87, 2 1963, p. 258. ISSN: 02646021. DOI: 10.1042/BJ0870258.

37. H. Kimura. "Histone modifications for human epigenome analysis". *Journal of Human Genetics 2013 58:7* 58, 7 2013, pp. 439–445. ISSN: 1435-232X. DOI: 10.1038/jhg.2013.66.

38. R. Sanchez and M. M. Zhou. "The role of human bromodomains in chromatin biology and gene transcription". *Current opinion in drug discovery & development* 12, 5 2009, p. 659. ISSN: 13676733.

39. R. S. Blanc and S. Richard. "Arginine Methylation: The Coming of Age". *Molecular Cell* 65, 1 2017, pp. 8–24. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2016.11.003.

40. I. Dunham et al. "An integrated encyclopedia of DNA elements in the human genome". *Nature 2012 489:7414* 489, 7414 2012, pp. 57–74. ISSN: 1476-4687. DOI: 10.1038/nature11247.

41. N. J. Krogan, J. Dover, A. Wood, J. Schneider, J. Heidt, M. A. Boateng, K. Dean, O. W. Ryan, A. Golshani, M. Johnston, J. F. Greenblatt, and A. Shilatifard. "The Paf1 Complex Is Required for Histone H3 Methylation by COMPASS and Dot1p: Linking Transcriptional Elongation to Histone Methylation". *Molecular Cell* 11, 3 2003, pp. 721–729. ISSN: 1097-2765. DOI: 10.1016/S1097-2765(03)00091-1.

42. N. R. Rose and R. J. Klose. "Understanding the relationship between DNA methylation and histone lysine methylation". *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839, 12 2014, pp. 1362–1372. ISSN: 1874-9399. DOI: 10.1016/J.BBAGRM.2014.02.007.

43. B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander. "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells". *Cell* 125, 2 2006, pp. 315–326. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2006.02.041.

44. E. Blanco, M. González-Ramírez, A. Alcaine-Colet, S. Aranda, and L. D. Croce. "The Bivalent Genome: Characterization, Structure, and Regulation". *Trends in Genetics* 36, 2 2020, pp. 118–131. ISSN: 0168-9525. DOI: 10.1016/J.TIG.2019.11.004.

45. A. Piunti and A. Shilatifard. "The roles of Polycomb repressive complexes in mammalian development and cancer". *Nature Reviews Molecular Cell Biology 2021 22:5* 22, 5 2021, pp. 326–345. ISSN: 1471-0080. DOI: 10.1038/s41580-021-00341-1.

46. P. Voigt, G. LeRoy, W. J. Drury, B. M. Zee, J. Son, D. B. Beck, N. L. Young, B. A. Garcia, and D. Reinberg. "Asymmetrically Modified Nucleosomes". *Cell* 151, 1 2012, pp. 181–193. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2012.09.002.

47. P. H. Lewis. "New mutants report". *Drosoph. Inf. Serv* 21, 1947, p. 69.

48. E. B. Lewis. "A gene complex controlling segmentation in Drosophila". *Nature 1978 276:5688* 276, 5688 1978, pp. 565–570. ISSN: 1476-4687. DOI: 10.1038/276565a0.

49. G. Jürgens. "A group of genes controlling the spatial expression of the bithorax complex in Drosophila". *Nature 1985 316:6024* 316, 6024 1985, pp. 153–155. ISSN: 1476-4687. DOI: 10.1038/316153a0.

50. B. Schuettengruber, H. M. Bourbon, L. D. Croce, and G. Cavalli. "Genome Regulation by Polycomb and Trithorax: 70 Years and Counting". *Cell* 171, 1 2017, pp. 34–57. ISSN: 10974172. DOI: 10.1016/j.cell.2017.08.002.

51. Z. Gao. "PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes". *Mol. Cell* 45, 3 2012, pp. 344–356. DOI: 10.1016/j.molcel.2012.01.002.

52. A. K. Robinson, B. Z. Leal, L. V. Chadwell, R. Wang, U. Ilangovan, Y. Kaur, S. E. Junco, V. Schirf, P. A. Osmulski, M. Gaczynska, A. P. Hinck, B. Demeler, D. G. McEwen, and C. A. Kim. "The Growth-Suppressive Function of the Polycomb Group Protein Polyhomeotic Is Mediated by Polymerization of Its Ster-

ile Alpha Motif (SAM) Domain". *Journal of Biological Chemistry* 287, 12 2012, pp. 8702–8713. ISSN: 0021-9258. DOI: 10.1074/JBC.M111.336115.

53. M. S. Lau, M. G. Schwartz, S. Kundu, A. J. Savol, P. I. Wang, S. K. Marr, D. J. Grau, P. Schorderet, R. I. Sadreyev, C. J. Tabin, and R. E. Kingston. "Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning". *Science* 355, 6329 2017, pp. 1081–1084. ISSN: 10959203. DOI: 10.1126/SCIENCE.AAH5403/SUPPL_FILE/LAU-SM.PDF.

54. S. Kundu, F. Ji, H. Sunwoo, G. Jain, J. T. Lee, R. I. Sadreyev, J. Dekker, and R. E. Kingston. "Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation". *Molecular Cell* 65, 3 2017, 432–446.e5. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2017.01.009.

55. R. Tatavosian, S. Kent, K. Brown, T. Yao, H. N. Duc, T. N. Huynh, C. Y. Zhen, B. Ma, H. Wang, and X. Ren. "Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation". *Journal of Biological Chemistry* 294, 5 2019, pp. 1451–1463. ISSN: 0021-9258. DOI: 10.1074/JBC.RA118.006620.

56. N. R. Rose, H. W. King, N. P. Blackledge, N. A. Fursova, K. J. Ember, R. Fischer, B. M. Kessler, and R. J. Klose. "RYBP stimulates PRC1 to shape chromatin-based communication between polycomb repressive complexes". *eLife* 5, 2016. DOI: 10.7554/ELIFE.18591.

57. L. Tavares, E. Dimitrova, D. Oxley, J. Webster, R. Poot, J. Demmers, K. Bezstarosti, S. Taylor, H. Ura, H. Koide, A. Wutz, M. Vidal, S. Elderkin, and N. Brockdorff. "RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3". *Cell* 148, 4 2012, pp. 664–678. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2011.12.029.

58. R. Arrigoni, S. L. Alam, J. A. Wamstad, V. J. Bardwell, W. I. Sundquist, and N. Schreiber-Agus. "The Polycomb-associated protein Rybp is a ubiquitin binding protein". *FEBS Letters* 580, 26 2006, pp. 6233–6241. DOI: 10.1016/J.FEBSLET.2006.10.027.

59. A. M. Farcas, N. P. Blackledge, I. Sudbery, H. K. Long, J. F. McGouran, N. R. Rose, S. Lee, D. Sims, A. Cerase, T. W. Sheahan, H. Koseki, N. Brockdorff, C. P. Ponting, B. M. Kessler, and R. J. Klose. "KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands". *eLife* 2012, 1 2012. DOI: 10.7554/ELIFE.00205.

60. M. Endoh, T. A. Endo, J. Shinga, K. Hayashi, A. Farcas, K. W. Ma, S. Ito, J. Sharif, T. Endoh, N. Onaga, M. Nakayama, T. Ishikura, O. Masui, B. M. Kessler, T. Suda, O. Ohara, A. Okuda, R. Klose, and H. Koseki. "PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes". *eLife* 6, 2017. DOI: 10.7554/ELIFE.21064.

61. N. P. Blackledge and R. J. Klose. *The molecular principles of gene regulation by Polycomb repressive complexes*. 2021. DOI: 10.1038/s41580-021-00398-y.

62. H. Li, R. Liefke, J. Jiang, J. V. Kurland, W. Tian, P. Deng, W. Zhang, Q. He, D. J. Patel, M. L. Bulyk, Y. Shi, and Z. Wang. "Polycomb-like proteins link the PRC2 complex to CpG islands". *Nature 2017 549:7671* 549, 7671 2017, pp. 287–291. ISSN: 1476-4687. DOI: 10.1038/nature23881.

63. S. Cooper, A. Grijzenhout, E. Underwood, K. Ancelin, T. Zhang, T. B. Nesterova, B. Anil-Kirmizitas, A. Bassett, S. M. Kooistra, K. Agger, K. Helin, E. Heard, and N. Brockdorff. "Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2". *Nature*

*Communications 2016 7:1* 7, 1 2016, pp. 1–8. ISSN: 2041-1723. DOI: 10.1038/ncomms13661.

64. V. Kasinath, C. Beck, P. Sauer, S. Poepsel, J. Kosmatka, M. Faini, D. Toso, R. Aebersold, and E. Nogales. "JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications". *Science* 371, 6527 2021. ISSN: 10959203. DOI: 10.1126/SCIENCE.ABC3393/SUPPL_FILE/ABC3393S1.MP4.

65. M.-O. Fauvarque and J.-M. Dura. "polyhomeotic regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in Drosophila." *Genes & Development* 7:8, 1993, pp. 1508–1520.

66. M. Bauer, J. Trupke, and L. Ringrose. "The quest for mammalian Polycomb response elements: are we there yet?" *Chromosoma 2015 125:3* 125, 3 2015, pp. 471–496. ISSN: 1432-0886. DOI: 10.1007/S00412-015-0539-4.

67. M. Almeida, J. S. Bowness, and N. Brockdorff. "The many faces of Polycomb regulation by RNA". *Current Opinion in Genetics & Development* 61, 2020, pp. 53–61. ISSN: 0959-437X. DOI: 10.1016/J.GDE.2020.02.023.

68. M. Almeida, G. Pintacuda, O. Masui, Y. Koseki, M. Gdula, A. Cerase, D. Brown, A. Mould, C. Innocent, M. Nakayama, L. Schermelleh, T. B. Nesterova, H. Koseki, and N. Brockdorff. "PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation". *Science* 356, 6342 2017, pp. 1081–1084. ISSN: 10959203. DOI: 10.1126/SCIENCE.AAL2512/SUPPL_FILE/AAL2512S6.MP4.

69. G. Pintacuda, G. Wei, C. Roustan, B. A. Kirmizitas, N. Solcan, A. Cerase, A. Castello, S. Mohammed, B. Moindrot, T. B. Nesterova, and N. Brockdorff. "hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing". *Molecular Cell* 68, 5 2017, 955–969.e10. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2017.11.013.

70. A. Aguilera and A. G. Rondón. "R-Loops". *Brenners Encyclopedia of Genetics: Second Edition*, 2013, pp. 265–268. DOI: 10.1016/B978-0-12-374984-0.01343-7.

71. K. Skourti-Stathaki, E. T. Triglia, M. Warburton, P. Voigt, A. Bird, and A. Pombo. "R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes". *Molecular Cell* 73, 5 2019, 930–945.e4. ISSN: 10972765. DOI: 10.1016/j.molcel.2018.12.016.

72. M. D. Gearhart, C. M. Corcoran, J. A. Wamstad, and V. J. Bardwell. "Polycomb Group and SCF Ubiquitin Ligases Are Found in a Novel BCOR Complex That Is Recruited to BCL6 Targets". *Molecular and Cellular Biology* 26, 18 2006, pp. 6880–6889. ISSN: 0270-7306. DOI: 10.1128/MCB.00630-06/SUPPL_FILE/SUPPLEMENTARY_TABLE_1.ZIP.

73. S. Poepsel, V. Kasinath, and E. Nogales. "Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes". *Nature Structural & Molecular Biology 2018 25:2* 25, 2 2018, pp. 154–162. ISSN: 1545-9985. DOI: 10.1038/s41594-018-0023-y.

74. N. P. Blackledge, A. M. Farcas, T. Kondo, H. W. King, J. F. McGouran, L. L. Hanssen, S. Ito, S. Cooper, K. Kondo, Y. Koseki, T. Ishikura, H. K. Long, T. W. Sheahan, N. Brockdorff, B. M. Kessler, H. Koseki, and R. J. Klose. "Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation". *Cell* 157, 6 2014, pp. 1445–1459. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2014.05.004.

75. A. J. Saurin, C. Shiels, J. Williamson, D. P. Satijn, A. P. Otte, D. Sheer, and P. S. Freemont. "The Human Polycomb Group Complex Associates with Pericentromeric Heterochromatin to Form a Novel Nuclear Domain". *Journal of Cell*

*Biology* 142, 4 1998, pp. 887–898. ISSN: 0021-9525. DOI: 10.1083/JCB.142.4.887.

76. J. Zhao, M. Wang, L. Chang, J. Yu, A. Song, C. Liu, W. Huang, T. Zhang, X. Wu, X. Shen, B. Zhu, and G. Li. "RYBP/YAF2-PRC1 complexes and histone H1-dependent chromatin compaction mediate propagation of H2AK119ub1 during cell division". *Nature Cell Biology 2020 22:4* 22, 4 2020, pp. 439–452. ISSN: 1476-4679. DOI: 10.1038/s41556-020-0484-1.

77. E. M. Riising, I. Comet, B. Leblanc, X. Wu, J. V. Johansen, and K. Helin. "Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide". *Molecular Cell* 55, 3 2014, pp. 347–360. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2014.06.005.

78. P. Jermann, L. Hoerner, L. Burger, and D. Schübeler. "Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation". *Proceedings of the National Academy of Sciences of the United States of America* 111, 33 2014, E3415–E3421. ISSN: 10916490. DOI: 10.1073/PNAS.1400672111/-/DCSUPPLEMENTAL.

79. M. D. Lynch, A. J. Smith, M. D. Gobbi, M. Flenley, J. R. Hughes, D. Vernimmen, H. Ayyub, J. A. Sharpe, J. A. Sloane-Stanley, L. Sutherland, S. Meek, T. Burdon, R. J. Gibbons, D. Garrick, and D. R. Higgs. "An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment". *The EMBO Journal* 31, 2 2012, pp. 317–329. ISSN: 1460-2075. DOI: 10.1038/EMBOJ.2011.399.

80. S. Kaneko, R. Bonasio, R. Saldaña-Meyer, T. Yoshida, J. Son, K. Nishino, A. Umezawa, and D. Reinberg. "Interactions between JARID2 and noncoding RNAs

regulate PRC2 recruitment to chromatin". *Molecular cell* 53:2, 2014, pp. 290–300.

81. M. Beltran, M. Tavares, N. Justin, G. Khandelwal, J. Ambrose, B. M. Foster, K. B. Worlock, A. Tvardovskiy, S. Kunzelmann, J. Herrero, T. Bartke, S. J. Gamblin, J. R. Wilson, and R. G. Jenner. "G-tract RNA removes Polycomb repressive complex 2 from genes". *Nature Structural & Molecular Biology 2019 26:10* 26, 10 2019, pp. 899–909. ISSN: 1545-9985. DOI: 10.1038/s41594-019-0293-z.

82. R. Eskeland, M. Leeb, G. R. Grimes, C. Kress, S. Boyle, D. Sproul, N. Gilbert, Y. Fan, A. I. Skoultchi, A. Wutz, and W. A. Bickmore. "Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination". *Molecular Cell* 38, 3 2010, pp. 452–464. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2010.02.032.

83. N. J. Francis, R. E. Kingston, and C. L. Woodcock. "Chromatin compaction by a polycomb group protein complex". *Science* 306, 5701 2004, pp. 1574–1577. ISSN: 00368075. DOI: 10.1126/SCIENCE.1100576/SUPPL_FILE/FRANCIS.SOM.PDF.

84. A. R. Pengelly, R. Kalb, K. Finkl, and J. Müller. "Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation". *Genes & Development* 29, 14 2015, pp. 1487–1492. ISSN: 0890-9369. DOI: 10.1101/GAD.265439.115.

85. R. S. Illingworth, M. Moffat, A. R. Mann, D. Read, C. J. Hunter, M. M. Pradeepa, I. R. Adams, and W. A. Bickmore. "The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development". *Genes and Development* 29, 18 2015, pp. 1897–1902. ISSN: 15495477. DOI: 10.1101/gad.268151.115.

86. E. Lavarone, C. M. Barbieri, and D. Pasini. "Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity". *Nature*

*Communications* 10, 1 2019, pp. 1–16. ISSN: 20411723. DOI: 10.1038/s41467-019-09624-w.

87. H. C. Hodges, B. Z. Stanton, K. Cermakova, C. Y. Chang, E. L. Miller, J. G. Kirkland, W. L. Ku, V. Veverka, K. Zhao, and G. R. Crabtree. "Dominant-negative SMARCA4 mutants alter the accessibility landscape of tissue-unrestricted enhancers". *Nature Structural & Molecular Biology 2017 25:1* 25, 1 2017, pp. 61–72. ISSN: 1545-9985. DOI: 10.1038/s41594-017-0007-3.

88. H. W. King and R. J. Klose. "The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells". *eLife* 6, 2017. DOI: 10.7554/ELIFE.22631.

89. M. K. Pirity, J. Locker, and N. Schreiber-Agus. "Rybp/DEDAF is required for early postimplantation and for central nervous system development". *Molecular and cellular biology* 25, 16 2005, pp. 7193–7202. ISSN: 0270-7306. DOI: 10.1128/MCB.25.16.7193-7202.2005.

90. T. Akasaka, M. van Lohuizen, N. van der Lugt, Y. Mizutani-Koseki, M. Kanno, M. Taniguchi, M. Vidal, M. Alkema, A. Berns, and H. Koseki. "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, 9 2001, pp. 1587–1597. ISSN: 0950-1991. DOI: 10.1242/DEV.128.9.1587.

91. F. Chiacchiera, A. Rossi, S. Jammula, M. Zanotti, and D. Pasini. "PRC2 preserves intestinal progenitors and restricts secretory lineage commitment". *The EMBO Journal* 35, 21 2016, pp. 2301–2314. ISSN: 0261-4189. DOI: 10.15252/embj.201694550.

92. F. Chiacchiera and D. Pasini. "Control of adult intestinal identity by the Polycomb repressive machinery". *Cell Cycle* 16, 3 2017, pp. 243–244. ISSN: 15514005. DOI: 10.1080/15384101.2016.1252582.

93. S. Pivetti, D. Fernandez-Perez, A. D'Ambrosio, C. M. Barbieri, D. Manganaro, A. Rossi, L. Barnabei, M. Zanotti, A. Scelfo, F. Chiacchiera, and D. Pasini. "Loss of PRC1 activity in different stem cell compartments activates a common transcriptional program with cell type–dependent outcomes". *Science Advances* 5, 5 2019. ISSN: 23752548. DOI: 10.1126/SCIADV.AAV1594/SUPPL_FILE/AAV1594_SM.PDF.

94. H. Gehart and H. Clevers. *Tales from the crypt: new insights into intestinal stem cells*. 2019. DOI: 10.1038/s41575-018-0081-y.

95. E. T. Kimchi, N. J. Gusani, and J. T. Kaifi. "Anatomy and Physiology of the Small Intestine". *Greenfields Surgery: Scientific Principles and Practice: Fifth Edition*, 2019, pp. 817–841. DOI: 10.1016/B978-0-323-40232-3.00071-6.

96. A. S. Darwich, U. Aslam, D. M. Ashcroft, and A. Rostami-Hodjegan. "Meta-Analysis of the Turnover of Intestinal Epithelia in Preclinical Animal Species and Humans". *Drug Metabolism and Disposition* 42, 12 2014, pp. 2016–2022. ISSN: 0090-9556. DOI: 10.1124/DMD.114.058404.

97. N. Barker, J. H. V. Es, J. Kuipers, P. Kujala, M. V. D. Born, M. Cozijnsen, A. Haegebarth, J. Korving, H. Begthel, P. J. Peters, and H. Clevers. "Identification of stem cells in small intestine and colon by marker gene Lgr5". *Nature 2007 449:7165* 449, 7165 2007, pp. 1003–1007. ISSN: 1476-4687. DOI: 10.1038/nature06196.

98. N. A. Mabbott, D. S. Donaldson, H. Ohno, I. R. Williams, and A. Mahajan. "Microfold (M) cells: important immunosurveillance posts in the intestinal epithe-

lium". *Mucosal Immunology 2013 6:4* 6, 4 2013, pp. 666–677. ISSN: 1935-3456. DOI: 10.1038/mi.2013.30.

99. C. Chen, Y. Yin, Q. Tu, and H. Yang. "Glucose and amino acid in enterocyte: Absorption, metabolism and maturation". *Frontiers in Bioscience - Landmark* 23, 9 2018, pp. 1721–1739. ISSN: 27686698. DOI: 10.2741/4669.

100. Q. Yang, N. A. Bermingham, M. J. Finegold, and H. Y. Zoghbi. "Requirement of Math1 for secretory cell lineage commitment in the mouse intestine". *Science* 294, 5549 2001, pp. 2155–2158. ISSN: 00368075. DOI: 10.1126/SCIENCE.1065718/SUPPL_FILE/1065718S4_THUMB.GIF.

101. E. Batlle, J. T. Henderson, H. Beghtel, M. M. V. den Born, E. Sancho, G. Huls, J. Meeldijk, J. Robertson, M. V. de Wetering, T. Pawson, and H. Clevers. "β-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/EphrinB". *Cell* 111, 2 2002, pp. 251–263. ISSN: 00928674. DOI: 10.1016/S0092-8674(02)01015-2/ATTACHMENT/A1C2E976-5258-407D-8828-626DA9B1A915/MMC3.JPG.

102. T. Sato, J. H. V. Es, H. J. Snippert, D. E. Stange, R. G. Vries, M. V. D. Born, N. Barker, N. F. Shroyer, M. V. D. Wetering, and H. Clevers. "Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts". *Nature 2010 469:7330* 469, 7330 2010, pp. 415–418. ISSN: 1476-4687. DOI: 10.1038/nature09637.

103. G. M. Birchenough, M. E. Johansson, J. K. Gustafsson, J. H. Bergström, and G. C. Hansson. "New developments in goblet cell mucus secretion and function". *Mucosal Immunology 2015 8:4* 8, 4 2015, pp. 712–719. ISSN: 1935-3456. DOI: 10.1038/mi.2015.32.

104. J. J. Worthington, F. Reimann, and F. M. Gribble. "Enteroendocrine cells-sensory sentinels of the intestinal environment and orchestrators of mucosal immu-

nity". *Mucosal Immunology 2018 11:1* 11, 1 2017, pp. 3–20. ISSN: 1935-3456. DOI: 10.1038/mi.2017.73.

105. L. López-Díaz, R. N. Jain, T. M. Keeley, K. L. VanDussen, C. S. Brunkan, D. L. Gumucio, and L. C. Samuelson. "Intestinal Neurogenin 3 directs differentiation of a bipotential secretory progenitor to endocrine cell rather than goblet cell fate". *Developmental Biology* 309, 2 2007, pp. 298–305. ISSN: 0012-1606. DOI: 10.1016/J.YDBIO.2007.07.015.

106. J. V. Moltke, M. Ji, H. E. Liang, and R. M. Locksley. "Tuft-cell-derived IL-25 regulates an intestinal ILC2–epithelial response circuit". *Nature 2015 529:7585* 529, 7585 2015, pp. 221–225. ISSN: 1476-4687. DOI: 10.1038/nature16161.

107. F. Gerbe, E. Sidot, D. J. Smyth, M. Ohmoto, I. Matsumoto, V. Dardalhon, P. Cesses, L. Garnier, M. Pouzolles, B. Brulin, M. Bruschi, Y. Harcus, V. S. Zimmermann, N. Taylor, R. M. Maizels, and P. Jay. "Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites". *Nature 2016 529:7585* 529, 7585 2016, pp. 226–230. ISSN: 1476-4687. DOI: 10.1038/nature16527.

108. F. Gerbe, J. H. V. Es, L. Makrini, B. Brulin, G. Mellitzer, S. Robine, B. Romagnolo, N. F. Shroyer, J. F. Bourgaux, C. Pignodel, H. Clevers, and P. Jay. "Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secretory cell type in the intestinal epithelium". *Journal of Cell Biology* 192, 5 2011, pp. 767–780. ISSN: 0021-9525. DOI: 10.1083/JCB.201010127.

109. A. Banerjee, C. A. Herring, B. Chen, H. Kim, A. J. Simmons, A. N. Southard-Smith, M. M. Allaman, J. R. White, M. C. Macedonia, E. T. Mckinley, M. A. Ramirez-Solano, E. A. Scoville, Q. Liu, K. T. Wilson, R. J. Coffey, M. K. Washington, J. A. Goettel, and K. S. Lau. "Succinate Produced by Intestinal Microbes Promotes Specification of Tuft Cells to Suppress Ileal Inflammation". *Gastroenterology*

159, 6 2020, 2101–2115.e5. ISSN: 15280012. DOI: 10.1053/J.GASTRO.2020.08.029/ATTACHMENT/AF699495-918A-4AA9-83C5-F31DC499E36B/MMC1.PDF.

110. R. Nusse and H. Clevers. "Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities". *Cell* 169, 6 2017, pp. 985–999. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2017.05.016.

111. V. Korinek, N. Barker, P. Moerer, E. V. Donselaar, G. Huls, P. J. Peters, and H. Clevers. "Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4". *Nature Genetics 1998 19:4* 19, 4 1998, pp. 379–383. ISSN: 1546-1718. DOI: 10.1038/1270.

112. R. Sancho, C. A. Cremona, and A. Behrens. "Stem cell and progenitor fate in the mammalian intestine: Notch and lateral inhibition in homeostasis and disease". *EMBO reports* 16, 5 2015, pp. 571–581. ISSN: 1469-3178. DOI: 10.15252/EMBR.201540188.

113. O. Basak, J. Beumer, K. Wiebrands, H. Seno, A. van Oudenaarden, and H. Clevers. "Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells". *Cell Stem Cell* 20, 2 2017, 177–190.e4. ISSN: 18759777. DOI: 10.1016/J.STEM.2016.11.001/ATTACHMENT/BF23980E-A4F2-4651-895D-3A35A1E2FCD9/MMC3.XLSX.

114. X. C. He, J. Zhang, W. G. Tong, O. Tawfik, J. Ross, D. H. Scoville, Q. Tian, X. Zeng, X. He, L. M. Wiedemann, Y. Mishina, and L. Li. "BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt–β-catenin signaling". *Nature Genetics 2004 36:10* 36, 10 2004, pp. 1117–1121. ISSN: 1546-1718. DOI: 10.1038/ng1430.

115. J. Massagué. "TGFβ signalling in context". *Nature Reviews Molecular Cell Biology 2012 13:10* 13, 10 2012, pp. 616–630. ISSN: 1471-0080. DOI: 10.1038/nrm3434.

116. A. P. G. Haramis, H. Begthel, M. V. D. Born, J. V. Es, S. Jonkheer, G. J. A. Offerhaus, and H. Clevers. "De Novo Crypt Formation and Juvenile Polyposis on BMP Inhibition in Mouse Intestine". *Science* 303, 5664 2004, pp. 1684–1686. ISSN: 00368075. DOI: 10.1126/SCIENCE.1093587/SUPPL_FILE/HARAMIS.SOM.PDF.

117. F. Chiacchiera, A. Rossi, S. Jammula, A. Piunti, A. Scelfo, P. Ordóñez-Morán, J. Huelsken, H. Koseki, and D. Pasini. "Polycomb Complex PRC1 Preserves Intestinal Stem Cell Identity by Sustaining Wnt/β-Catenin Transcriptional Activity". *Cell Stem Cell* 18, 1 2016, pp. 91–103. ISSN: 18759777. DOI: 10.1016/j.stem.2015.09.019.

118. M. K. Huseyin and R. J. Klose. "Live-cell single particle tracking of PRC1 reveals a highly dynamic system with low target site occupancy". *Nature Communications 2021 12:1* 12, 1 2021, pp. 1–20. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21130-6.

119. M. Endoh, T. A. Endo, T. Endoh, Y. I. Fujimura, O. Ohara, T. Toyoda, A. P. Otte, M. Okano, N. Brockdorff, M. Vidal, and H. Koseki. "Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity". *Development* 135, 8 2008, pp. 1513–1524. ISSN: 0950-1991. DOI: 10.1242/DEV.014340.

120. A. Kiermaier, J. M. Gawn, L. Desbarats, R. Saffrich, W. Ansorge, P. J. Farrell, M. Eilers, and G. Packham. "DNA binding of USF is required for specific E-box dependent gene activation in vivo". *Oncogene* 18:51, 1999, pp. 7200–7211.

121. R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang. "Role of histone H3 lysine 27 methylation in polycomb-group silencing". *Science* 298, 5595 2002, pp. 1039–1043. ISSN: 00368075. DOI: 10.1126/SCIENCE.1076997/SUPPL_FILE/PAP.PDF.

122. L. Wang, J. L. Brown, R. Cao, Y. Zhang, J. A. Kassis, and R. S. Jones. "Hierarchical Recruitment of Polycomb Group Silencing Complexes". *Molecular Cell* 14, 5 2004, pp. 637–646. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2004.05.009.

123. J. Min, Y. Zhang, and R.-M. Xu. "Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27". *Genes & development* 17:15, 2003, pp. 1823–1828.

124. M. Endoh, T. A. Endo, T. Endoh, K. ichi Isono, J. Sharif, O. Ohara, T. Toyoda, T. Ito, R. Eskeland, W. A. Bickmore, M. Vidal, B. E. Bernstein, and H. Koseki. "Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity". *PLoS Genetics* 8, 7 2012. ISSN: 15537390. DOI: 10.1371/journal.pgen.1002774.

125. M. Leeb, D. Pasini, M. Novatchkova, M. Jaritz, K. Helin, and A. Wutz. "Polycomb complexes act redundantly to repress genomic repeats and genes". *Genes & Development* 24, 3 2010, pp. 265–276. ISSN: 0890-9369. DOI: 10.1101/GAD.544410.

126. G. Buchwald, P. V. D. Stoop, O. Weichenrieder, A. Perrakis, M. V. Lohuizen, and T. K. Sixma. "Structure and E3-ligase activity of the Ring–Ring complex of Polycomb proteins Bmi1 and Ring1b". *The EMBO Journal* 25, 11 2006, pp. 2465–2474. ISSN: 1460-2075. DOI: 10.1038/SJ.EMBOJ.7601144.

127. M. Tsuboi, Y. Kishi, W. Yokozeki, H. Koseki, Y. Hirabayashi, and Y. Gotoh. "Ubiquitination-Independent Repression of PRC1 Targets during Neuronal Fate

Restriction in the Developing Mouse Neocortex". *Developmental Cell* 47, 6 2018, 758–772.e5. ISSN: 18781551. DOI: 10.1016/j.devcel.2018.11.018.

128. B. Amati and H. Land. "Myc—Max—Mad: a transcription factor network controlling cell cycle progression, differentiation and death". *Current Opinion in Genetics & Development* 4, 1 1994, pp. 102–108. ISSN: 0959-437X. DOI: 10.1016/0959-437X(94)90098-1.

129. N. A. Fursova, N. P. Blackledge, M. Nakayama, S. Ito, Y. Koseki, A. M. Farcas, H. W. King, H. Koseki, and R. J. Klose. "Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression". *Molecular Cell* 74, 5 2019, 1020–1036.e8. ISSN: 10974164. DOI: 10.1016/j.molcel.2019.03.024.

130. E. Conway, F. Rossi, D. Fernandez-Perez, E. Ponzo, K. J. Ferrari, M. Zanotti, D. Manganaro, S. Rodighiero, S. Tamburri, and D. Pasini. "BAP1 enhances Polycomb repression by counteracting widespread H2AK119ub1 deposition and chromatin condensation". *Molecular Cell* 81, 17 2021, 3526–3541.e8. ISSN: 10974164. DOI: 10.1016/J.MOLCEL.2021.06.020/ATTACHMENT/C881B318-0C35-4FFF-9CF3-BA414434730D/MMC4.XLSX.

131. A. L. Haber, M. Biton, N. Rogel, R. H. Herbst, K. Shekhar, C. Smillie, G. Burgin, T. M. Delorey, M. R. Howitt, Y. Katz, et al. "A single-cell survey of the small intestinal epithelium". *Nature* 551:7680, 2017, pp. 333–339.

132. J. Köster and S. Rahmann. "Snakemake—a scalable bioinformatics workflow engine". *Bioinformatics* 28, 19 2012, pp. 2520–2522. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTS480.

133. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2020.

134. H. Wickham, M. Averick, J. Bryan, W. Chang, L. D, A. Mcgowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. "Welcome to the Tidyverse". *Journal of Open Source Software* 4, 43 2019, p. 1686. ISSN: 2475-9066. DOI: 10.21105/JOSS.01686.

135. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang, and J. Zhang. "Bioconductor: open software development for computational biology and bioinformatics." *Genome biology* 5, 10 2004, pp. 1–16. ISSN: 14656914. DOI: 10.1186/GB-2004-5-10-R80/FIGURES/3.

136. D. A. Orlando, M. W. Chen, V. E. Brown, S. Solanki, Y. J. Choi, E. R. Olson, C. C. Fritz, J. E. Bradner, and M. G. Guenther. "Quantitative ChIP-Seq normalization reveals global modulation of the epigenome". *Cell reports* 9:3, 2014, pp. 1163–1170.

137. B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome". *Genome Biology* 10, 3 2009, pp. 1–10. ISSN: 14747596. DOI: 10.1186/GB-2009-10-3-R25/TABLES/5.

138. Broad Institute. *Picard Tools*. http://broadinstitute.github.io/picard/. (Accessed: 2018/02/21; version 2.17.8).

139. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, and X. S. Shirley. "Model-based anal-

ysis of ChIP-Seq (MACS)". *Genome Biology* 9, 9 2008, pp. 1–9. ISSN: 14747596. DOI: 10.1186/GB-2008-9-9-R137/FIGURES/3.

140. L. J. Zhu, C. Gazin, N. D. Lawson, H. Pagès, S. M. Lin, D. S. Lapointe, and M. R. Green. "ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data". *BMC Bioinformatics* 11, 1 2010, pp. 1–10. ISSN: 14712105. DOI: 10.1186/1471-2105-11-237/TABLES/2.

141. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, and T. Manke. "deepTools2: a next generation web server for deep-sequencing data analysis". *Nucleic Acids Research* 44, W1 2016, W160–W165. ISSN: 0305-1048. DOI: 10.1093/NAR/GKW257.

142. H. M. Amemiya, A. Kundaje, and A. P. Boyle. "The ENCODE Blacklist: Identification of Problematic Regions of the Genome". *Scientific Reports 2019 9:1* 9, 1 2019, pp. 1–5. ISSN: 2045-2322. DOI: 10.1038/s41598-019-45839-z.

143. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. "Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities". *Molecular Cell* 38, 4 2010, pp. 576–589. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2010.05.004.

144. G. Yu, L. G. Wang, Y. Han, and Q. Y. He. "ClusterProfiler: An R package for comparing biological themes among gene clusters". *OMICS A Journal of Integrative Biology* 16, 5 2012, pp. 284–287. ISSN: 15362310. DOI: 10.1089/OMI.2011.0118/ASSET/IMAGES/LARGE/FIGURE1.JPEG.

145. S. Parekh, C. Ziegenhain, B. Vieth, W. Enard, and I. Hellmann. "The impact of amplification on differential expression analyses by RNA-seq". *Scientific Reports 2016 6:1* 6, 1 2016, pp. 1–11. ISSN: 2045-2322. DOI: 10.1038/srep25533.

146. G. G. Faust and I. M. Hall. "SAMBLASTER: fast duplicate marking and structural variant read extraction". *Bioinformatics* 30, 17 2014, pp. 2503–2505. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTU314.

147. Y. Liao, G. K. Smyth, and W. Shi. "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". *Bioinformatics* 30, 7 2014, pp. 923–930. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTT656.

148. M. I. Love, W. Huber, and S. Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome Biology* 15, 12 2014, pp. 1–21. ISSN: 1474760X. DOI: 10.1186/S13059-014-0550-8/FIGURES/9.

149. A. Zhu, J. G. Ibrahim, and M. I. Love. "Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences". *Bioinformatics* 35, 12 2019, pp. 2084–2092. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTY895.

150. N. Ignatiadis, B. Klaus, J. B. Zaugg, and W. Huber. "Data-driven hypothesis weighting increases detection power in genome-scale multiple testing". *Nature methods* 13, 7 2016, p. 577. ISSN: 15487105. DOI: 10.1038/NMETH.3885.

151. A. Scelfo, D. Fernández-Pérez, S. Tamburri, M. Zanotti, E. Lavarone, M. Soldi, T. Bonaldi, K. J. Ferrari, and D. Pasini. "Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities". *Molecular Cell* 74, 5 2019, 1037–1052.e7. ISSN: 10974164. DOI: 10.1016/j.molcel.2019.04.002.

152. S. Tamburri, E. Lavarone, D. Fernández-Pérez, E. Conway, M. Zanotti, D. Manganaro, and D. Pasini. "Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression". *Molecular Cell* 77, 4 2020, 840–856.e5. ISSN: 10974164. DOI: 10.1016/j.molcel.2019.11.021.

153. C. Moolenbeek and E. Ruitenberg. "The 'Swiss roll': a simple technique for histological studies of the rodent intestine". *Laboratory animals* 15:1, 1981, pp. 57–60.

154. N. Tanaskovic. "TUMOR SUPPRESSIVE ROLE OF THE POLYCOMB GROUP RING FINGER PROTEIN PCGF6 IN MYC-INDUCED LYMPHOMAGENESIS", 2020.

155. B. Stielow, F. Finkernagel, T. Stiewe, A. Nist, and G. Suske. "MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6". *PLOS Genetics* 14, 1 2018, e1007193. ISSN: 1553-7404. DOI: 10.1371/JOURNAL.PGEN.1007193.

156. Z. Gao, P. Lee, J. M. Stafford, M. V. Schimmelmann, A. Schaefer, and D. Reinberg. "AUTS2 confers gene activation to Polycomb group proteins in the CNS". *Nature* 516, 7531 2014, p. 349. ISSN: 14764687. DOI: 10.1038/NATURE13921.

157. W. Zhao, Y. Huang, J. Zhang, M. Liu, H. Ji, C. Wang, N. Cao, C. Li, Y. Xia, Q. Jiang, and J. Qin. "Polycomb group RING finger proteins 3/5 activate transcription via an interaction with the pluripotency factor Tex10 in embryonic stem cells". *The Journal of Biological Chemistry* 292, 52 2017, p. 21527. ISSN: 1083351X. DOI: 10.1074/JBC.M117.804054.

158. A. L. Roy, M. Meisterernst, P. Pognonec, and R. G. Roeder. "Cooperative interaction of an initiator-binding transcription initiation factor and the helix–loop–helix activator USF". *Nature* 354:6350, 1991, pp. 245–248.

159. A. R. Pengelly, Ö. Copur, H. Jäckle, A. Herzig, and J. Müller. "A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb". *Science* 339, 6120 2013, pp. 698–699. ISSN: 10959203. DOI: 10.1126/SCIENCE.1231382/SUPPL_FILE/1231382.PENGELLY.SM.PDF.

160. D. Pasini and L. D. Croce. "Emerging roles for Polycomb proteins in cancer". *Current Opinion in Genetics & Development* 36, 2016, pp. 50–58. ISSN: 0959-437X. DOI: 10.1016/J.GDE.2016.03.013.

161. D. Landeira, S. Sauer, R. Poot, M. Dvorkina, L. Mazzarella, H. F. Jørgensen, C. F. Pereira, M. Leleu, F. M. Piccolo, M. Spivakov, E. Brookes, A. Pombo, C. Fisher, W. C. Skarnes, T. Snoek, K. Bezstarosti, J. Demmers, R. J. Klose, M. Casanova, L. Tavares, N. Brockdorff, M. Merkenschlager, and A. G. Fisher. "Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators". *Nature Cell Biology 2010 12:6* 12, 6 2010, pp. 618–624. ISSN: 1476-4679. DOI: 10.1038/ncb2065.

162. E. M. Mendenhall, R. P. Koche, T. Truong, V. W. Zhou, B. Issac, A. S. Chi, M. Ku, and B. E. Bernstein. "GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells". *PLoS Genetics* 6, 12 2010, pp. 1–10. ISSN: 15537390. DOI: 10.1371/JOURNAL.PGEN.1001244.

163. W. Qin, P. Wolf, N. Liu, S. Link, M. Smets, F. La Mastra, I. Forné, G. Pichler, D. Hörl, K. Fellinger, et al. "DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination". *Cell research* 25:8, 2015, pp. 911–929.

164. K. J. Ferrari, A. Scelfo, S. Jammula, A. Cuomo, I. Barozzi, A. Stützer, W. Fischle, T. Bonaldi, and D. Pasini. "Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity". *Molecular cell* 53:1, 2014, pp. 49–62.

165. H.-G. Lee, T. G. Kahn, A. Simcox, Y. B. Schwartz, and V. Pirrotta. "Genome-wide activities of Polycomb complexes control pervasive transcription". *Genome research* 25:8, 2015, pp. 1170–1181.

166. I. Cohen, C. Bar, H. Liu, V. J. Valdes, D. Zhao, P. M. Galbo, J. M. Silva, H. Koseki, D. Zheng, and E. Ezhkova. "Polycomb complexes redundantly maintain epidermal stem cell identity during development". *Genes & Development* 35:5-6, 2021, pp. 354–366.