University of Udine

PhD Course "Molecular and Cellular Medicine" (XXIX Cycle)

Department of Medical and Biological Sciences

PhD Thesis

Exploiting the pathophysiological role of the Polycomb mutational landscape by CRISPR/Cas9 genome editing

PhD Student:
Elisa Lavarone

Supervisor:
Prof. Giuseppe Damante

Co-supervisor:
Dott. Diego Pasini

Year of final Exam 2017

Table of Contents

2

Figures index

Tables Index

Abstract

Polycomb group of proteins (PcGs) are essential multiprotein complexes that regulate, through chromatin repression and compaction, cell identity and cell-fate transitions ensuring the correct establishment of lineage-specific transcriptional programs. Due to their key roles in cellular homeostasis and proliferation, it is not surprising at all that their deregulation, in terms of expression levels or activity, has been linked to the development and sustainment of several types of human cancers. Aberrations affecting subunits of both Polycomb Repressive Complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2), the two major PcGs complexes, have been reported. EZH2, the catalytic subunit of PRC2 responsible for its methylation activity on lysine 27 of histone H3 (H3K27), is often over-expressed in human cancers. This correlates with global increased H3K27 trimethylation (H3K27me3) levels and with tumor prognosis. Recently, mutations affecting critical residues within EZH2 catalytic SET domain and thus impairing its activity have been described. Interestingly, both hyper-activating and inactivating mutations have been shown to affect EZH2 histone methylation activity. A complex scenario in which EZH2 can act as oncogene or tumor-suppressor depending on the cell-context, is emerging. Up to now very little is known about the biological role of the mutated forms of EZH2 and subsequent alteration of methylation patterns and, therefore, the aim of this thesis is to try to unravel the molecular mechanisms underlying these tumorigenic mutations. I took advantage of mouse embryonic stem cells (mESCs), representing a simple model system where PcGs activity is well characterized, and of the new powerful CRISPR/Cas9 genome-editing tool. At the beginning of this project CRISPR/Cas9 technology had just emerged as a versatile and powerful tool to perform highly efficient genome-editing in a variety of cell-types. I applied this approach to mESC

to obtain relevant genetic cellular models to study the role of mutations affecting EZH2 activity. I generated isogenic mESC lines harboring physiological EZH2 Y726D and R685C-inactivating aminoacidic substitution. Moreover, a cellular model for K27M mutation, that affects EZH2 substrate histone H3.3 thus inhibiting its enzymatic activity, was obtained. *Ezh2* and *Ezh1* knock-out cells combined with homozygous EZH2 Y641N expressing cells allowed me to clarify several aspects regarding EZH1 and EZH2 interplay and cooperation within PRC2 activity, suggesting a context-dependent EZH1-compensative role. My preliminary results demonstrate that the differentiation capabilities of mESCs rely on H3K27me3 deposition whereas PRC2 recruitment to target loci occurs in an H3K27me3-independent manner. I coupled differentiation assays with location analyses (ChIP-qPCR and ChIP-seq), aimed to map chromatin association of Polycomb components and specific deposition of histone modifications, to elucidate the molecular mechanisms by which distinct mutations affect the activity of PRC2.

1. Introduction

1.1 Overview on chromatin

1.1.1 Chromatin structure

Genetic material in eukaryotes is compacted through several hierarchical levels of organization, in order to fit into the small volume of the nucleus. The first level of packaging is achieved through the association of genomic DNA to specialized proteins, called histones, that leads to the formation of chromatin, a highly-ordered regularly repeated structure. The basic unit of chromatin is the nucleosome, described for the first time by Kornberg in 1974[1,2] that consists in 147 bp of left handedly DNA wrapped around an octamer of histones proteins, comprising a central core made of a tetramer of histones H3-H4 flanked by two dimers of histones H2A-H2B. In the nucleosome context the basic charge of histones neutralizes the negative charges of the DNA phosphate backbone allowing the formation of "beads on a string" fibers, visible by electron microscopy[3]. A multitude of protein-protein interactions within the histone octamer and numerous electrostatic and hydrogen bonds between protein and DNA contribute to the stabilization of the nucleosome structure[4-6]. The core histones (H2A, H2B, H3 and H4) are small, positively charged proteins characterized by a histone fold domain, composed of three α-helices. While the majority of the interactions are between the structured regions of histones and DNA, the flexible tails of histones protrude away from the nucleosome core and engage interactions with neighboring nucleosomes or with nuclear factors. These less-structured C- and N-terminus tails can be post-translationally modified thus affecting histones-DNA interactions and therefore modulating chromatin structure and functions[7,8]. Consecutive nucleosomes are separated from each other by

unwrapped 10-80 bp linker DNA and can be associated with the linker histone H1. Unlike core histone proteins building a nucleosomal core on which DNA is wrapped, histone H1 is attached outwardly at the sites where DNA enters and exits the nucleosome core[9]. Histone H1 is connected to the DNA located on the histone core to stabilize both the nucleosome and the linker DNA. This can promote the assembly of higher-order chromatin structures, leading to an increased degree of chromatin compaction[10,11]. A long time ago, cytologists discovered the non-uniform distribution of DNA within the nucleus. Indeed, the staining with dyes that specifically bind DNA revealed the existence of more and less densely compacted areas[12]. Early studies established the association between the degree of compaction of genomic regions and their transcriptional status[13]. From the functional point of view, chromatin is commonly divided into two major states, euchromatin and heterochromatin. Euchromatin is characterized by a more relaxed conformation that provides a more permissive environment and defines genomic regions containing actively transcribed genes or potentially active ones[14,15]. Heterochromatin, instead, is tightly compacted and refers to transcriptionally inactive and highly condensed genomic regions. Heterochromatin, depending on whether it is established in a cell-type specific manner or not, can be further subdivided into facultative or constitutive. The first one consists of genomic regions containing genes that are differentially expressed through development and/or differentiation and which then become silenced. The second, is typically gene-poor and includes permanently silenced genes in genomic regions such as centromere, peri-centromeric and telomeric repeats[16]. If on one hand DNA has to be compacted and compartmentalized within the nucleus, at the same time specific genomic regions need to remain accessible for interaction with different regulatory and transcription factors in order to timely achieve the diverse key cellular biological processes. Chromatin structure is indeed highly dynamically

12

modulated to ensure the accessibility and the recruitment of specific regulatory factors to DNA. The chromatin plasticity is finely modulated by different epigenetic mechanisms that include the action of chromatin remodeling complexes, the incorporation of histone variants and covalent post-translational modifications of the histone tails or of DNA. Indeed, this "epigenetic information" provides an additional informative layer beyond the underlying genomic sequence. The increased understanding of this complex molecular scenario has tightly linked the chromatin environment with the concept of epigenetics. The meaning of the word "*epigenetics*" has been matter of debate in the last decade and a widely accepted consensus definition is still missing. The term was originally coined by Conrad Waddington in 1942[17,18] that, in a developmental biology context, defined epigenetics as "t*he branch of biology that study the casual interactions between genes and their products*", referring in this way to all the molecular pathways somehow modulating the expression of a genotype into a particular phenotype. After many different attempts, Wu and Morris[17] in 2001, proposed this new definition of the term, adding the concept of heritability to Waddington's original one: "*the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail change in DNA sequence.*" More recent definitions define epigenetics as "*molecular factors and processes around DNA that regulate genome activity independent of DNA sequence and are mitotically stable*" and "*the mechanism for the stable maintenance of gene expression that involves physically "marking" DNA or its associated proteins*" allowing "*genotypically*" *identical cells to be phenotypically distinct*", introducing a direct link to chromatin modification[19]. Nowadays the term epigenetics is frequently used to describe, in general, the study of chromatin biology and it can be defined as "*the study of heritable changes in gene expression that occur independent of changes in the primary DNA sequence*". According to this comprehensive definition, epigenetic mechanisms and

players responsible for the regulation of chromatin organization and packaging comprise chromatin remodelers, histone variants incorporation and covalent modifications of DNA and histones.

1.1.2 Chromatin remodeling

The establishment and maintenance of specific chromatin states is crucial during development and during all the DNA template-based processes that require rapid rearrangements of chromatin structure. Such chromatin dynamicity is achieved through the activity of different remodeling complexes. Among them a crucial role is exerted by adenosine triphosphate (ATP)-dependent chromatin remodeling complexes. These are large multi-complexes, highly conserved within eukaryotes, that use energy from ATP hydrolysis to modify chromatin assembly. They are all characterized by the presence of an ATPase subunit homologous to ATP-binding helicase of the DEAD/H box-containing family[20]. These ATPases can be classified in at least 4 major families SWI/SNF, ISWI (imitation switch), NURD/Mi-2/CHD (chromodomain helicase, DNA-binding) and INO80/SWR1 (Snf2-related CREBBP activator protein) based on the additional presence of unique domains within or adjacent to the ATPase domain[21]. The presence of many accessory non-catalytic subunits, ranging from two in some ISWI complexes to 11 or more in the SWI/SNF complexes, guarantees a combinatorial assembly of these complexes that allows to diversify their biochemical properties and functions. Chromatin-remodeling enzymes catalyze a broad range of chromatin transformations that include sliding the histone octamer across the DNA, changing the conformation of nucleosome DNA and changing the composition of the histone octamer, leading to a fine modulation of chromatin structure and functionality[22,23]. In general, the nucleosome

14

remodeling reaction can result in increased or reduced accessibility of a genomic site thus leading to transcriptional activation or repression[24,25].

1.1.3 Histone variants

An additional layer of epigenetic control is exerted by the incorporation of histone variants within the nucleosome. In eukaryotes, to increase the complexity of chromatin-mediated signaling, a number of histone variants have evolved. These variants resemble their canonical counterparts but show differences in expression pattern, genomic organization, deposition and also in terms of functional outcome[26,27]. Indeed, histones can be classified in canonical and non-canonical[28]. Canonical histones are also known as replication-dependent histones since they are deposited in a replication-coupled manner and have a specific expression peak during cell cycle S-phase[26,27,29]. They are expressed from large gene clusters, where multiple gene copies are organized in tandem and do not present introns. Moreover, their mRNAs are not polyadenylated but instead have a unique 3' stem-loop structure which is key for mRNA stability and translation[30]. Non-canonical histones, also known as "replacement" histones, are expressed from single or low copy genes that include introns and a polyadenylated mRNA, in a replication-independent manner throughout the phases of cell cycle[31,32]. Histone variants have been identified for all histones, except for histone H4. The human histone H3 family includes eight different proteins: the core histones H3.1, H3.2, the ubiquitous replacement variant H3.3, the centromeric CENP-A (also referred to as cenH3), the primate-specific H3.X, H3.Y, and the testis-specific histones H3t and H3.5[33]. Replacement of canonical histones with histone variants has important effects on nucleosome stability and organization and so, by creating functionally distinct chromatin domains, plays critical roles in the

regulation of a range of DNA-based cellular processes. Histones dynamics, in terms of faithful and accurate chromatin positioning and incorporation of histone variants, is timely and spatially regulated by ATP-dependent chromatin remodeling complexes and by histones chaperones. This latter is a family of specialized protein complexes involved in the histone trafficking as well as in histones deposition or eviction into/from the nucleosomes. The biochemical diversity introduced by different histone variants, by adding another level of complexity and a distinct way of modulating chromatin, highlights how histones are more than merely structural proteins but represent real epigenetic platforms that can act as fundamental regulators of cellular processes.

1.1.4 Introduction on covalent modifications of DNA and histones

Among the several factors that influence chromatin architecture, both locally and globally, covalent modifications, of either DNA or histones, represent probably the most influential epigenetic determinants. These modifications determine the accessibility to DNA by physically modulating the non-covalent interactions between histones and between histones and DNA. Moreover, they provide a fundamental informative platform for the recruitment of epigenetic regulators. These latter can be broadly classified in epigenetic writers, erasers and readers[34]. Epigenetic writers are chromatin enzymes able to deposit covalent modifications on histones or DNA, while epigenetic erasers are the ones responsible for their removal. Many chromatin regulators can act also as epigenetic readers: through specialized domains they are able to recognize and bind to distinct covalent modifications within the nucleosome and, by recruiting additional effector proteins, to trigger downstream signaling pathways[35]. Chromatin modifying enzymes deposit and remove covalent modifications in a highly regulated manner and the

information conveyed by these epigenetic modifications plays a critical instructive role in the regulation of all cellular DNA-based processes.

1.1.4.1 DNA modifications

Four different DNA modifications have been described in mammals so far, including 5-methyl cytosine (5mC), 5-hydroxymethylcytosine (5hmC) and the oxidation products 5-formylcytosine (5fC) and 5-carboxylcytosine (5aC)[36,37]. The first two represent the major epigenetic modifications of DNA and are relatively stable and abundant across mammalian genomes[38,39]. 5fC and 5aC, in contrast, are extremely rare, transient and considered as active DNA demethylation intermediates[39]. DNA cytosine methylation (5mC) is one of the best-characterized epigenetic modifications and implies the covalent addition of a methyl group to the 5-position of the cytosine ring. This stable and heritable epigenetic modification is usually associated to chromatin repressive states and is involved in retrotransposon silencing, mammalian genomic imprinting, X-chromosome inactivation, repetitive elements suppression, lineage-specific gene-expression regulation and maintenance of epigenetic memory[40-42]. This modification occurs preferentially on the cytosines of CpG dinucleotides. Indeed, genome-wide studies in mammalian genomes revealed that 5mC is widespread throughout the mammalian genome and post-replicatively marks approximately 70–80% of CpG dinucleotides, accounting in human somatic cells for approximately 1% of all DNA bases. Surprisingly, CpG islands (CGIs), specific CpG-rich regions occurring at almost two-thirds of mammalian gene promoters, are refractory to this modification and are generally unmethylated[43]. DNA methylation is performed by specific DNA methyltransferases (DNMTs). New DNA methylation patterns are established by the so called *de novo*

DNMTs, Dnmt3a and Dnmt3b[44,45] while the faithful maintenance of the methylation pattern during DNA replication is ensured by Dnmt1[46-48]. DNA methylation in essential for mammalian development, in fact deletion of Dnmt1 or Dnmt3b results in embryonic lethality, whereas homozygous Dnmt3a knockout mice die 4 weeks after birth[44,49]. DNA methylation has been classically associated to gene silencing through the physical blocking of the interaction with DNA-binding regulating factors and by serving as binding site of methyl-binding domain containing proteins. To highlight the crucial role of DNA methylation in epigenetic control, the deregulation of DNMTs expression and abnormal DNA methylation have been linked to several human diseases including cancer. The DNA methylation scenario is far from being completely uncovered. DNA methylation patterns represent indeed an additional layer of epigenetic control that regulate chromatin architecture.

1.1.4.2 Histone post-translational modifications (HPTMs)

Histones are subjected to an astonishing number of PTMs that do not occur only on the flexible N- and C- terminus tails protruding from nucleosomes, but also on the histone globular domain and at the DNA-histone interface thus regulating histone-histone and histone-DNA interactions[39,50]. In addition to well characterized histone modifications such as lysine acetylation, lysine/arginine methylation and serine/threonine/tyrosine phosphorylation[35,51], recent studies uncovered a surprising number of novel modifications including lysine crotonylation, butyrylation, propionylation and succinylation[52]. Many of these newly identified modifications are extremely low in abundance and their functional role has just started to be elucidated. Histone covalent modifications are deposited by specific "chromatin writers" and removed by "chromatin

erasers" in a highly dynamic and regulated manner. Histone PTMs exert their mechanism through two main mechanisms[53]. First of all, they can directly and physically influence the chromatin structure. Histone acetylation and phosphorylation can effectively reduce the positive charge of histones thus disrupting the electrostatic DNA-histones interactions, leading to a more open and accessible chromatin structure. Moreover, also ubiquitylation, by the addition of large molecules, induces a direct overall change in the nucleosome conformation. Histone modifications can also act through a second mechanism that involves the selective recruiting of specific chromatin factors also defined as "readers". Many chromatin regulators possess in fact specialized domains that allow the surveying and reading of the chromatin landscape and the docking at specifically marked genomic regions. This further induces the recruitment of other effectors or regulatory factors in order to remodel the chromatin environment and regulate all nuclear processes. In this context, an extra-level of complexity is achieved through the cross-talk between the different modifications. The communication between modifications, even if present on different histone tails, may occur at several levels: 1) antagonism between different modifications that are mutually exclusive; 2) the binding of a protein can be positively influenced or impaired by the adjacent modifications; 3) the catalytic activity of an enzyme can be impaired by the modification of its substrate[53].

1.1.4.2.1 Histone acetylation

Histone acetylation, from its identification in 1964 by Allfrey and colleagues[54], is probably the most well studied among all the histone modifications. It occurs on lysine residues and is classically associated with an open chromatin conformation and activation of transcription. This modification is dynamically regulated by the action of

two families of enzymes: histone acetyltransferases, the writers (HATs) and histone deacetylases, the erasers (HDACs). HATs use acetyl CoA as cofactor, catalyzing the transfer of an acetyl group to the ε-amino group of lysine side chains, thus neutralizing its positive charge. In general, HATs acetylate multiple sites within the histone tails neutralizing lysine's positive charge thus disrupting the DNA-histones electrostatic interactions, leading to an increased accessibility of DNA. However, also residues of the globular histone core, such as H3K56, can be acetylated[55]. Lysine acetylation is reversed by HDACs. This action restores the positive charge of the residue thus potentially stabilizing local chromatin structure. In general, promoters of active genes are highly acetylated, whereas the coding regions of genes need to be maintained in a hypo-acetylated state to prevent the aberrant initiation of transcription. In addition to the direct induction of changes into chromatin structure, histone acetylation can have also a recruitment function. Importantly, acetylated lysines can be primarily recognized by proteins containing an evolutionarily conserved binding motif, the bromodomain. More than 40 human bromodomain-containing proteins have been described and comprise chromatin remodelers, histone acetyltransferases, histone methyltransferases and transcriptional coactivators[56].

1.1.4.2.2 Histone methylation

Histone methylation, first described in 1964 by Murray[57], mainly occurs on lysine and arginine residues at the N-terminus tail of histones H3 and H4. Indeed, methylation can occur also on glutamine, aspartic acid and proline residues. Unlike acetylation and phosphorylation, histone methylation does not alter the charge of the histone protein. Histone lysine methyltransferases (HKMTs) are the enzymes responsible for the transfer

of methyl groups from S-adenosylmethionine (SAM) to a lysine's ε-amino group[58]. Different degrees of modification are possible, lysines in fact can be mono-, di- or tri-methylated. Except for Dot1 enzyme, responsible for the methylation of H3K79 residue within the histone globular core, all HKMTs enzymes contain a specific evolutionarily conserved catalytic domain, the SET domain. HMTs retain high specificity for the substrate and some of them are even specific for a given methylation state. Indeed, inside the SET catalytic pocket domain an aromatic residue (a tyrosine or a phenylalanine) plays an essential role in controlling the state of methylation[59]. A plethora of histone demethylases, responsible for the removal of methyl groups from histone lysine residues, have been characterized. Histone methylation plays a role in transcriptional regulation at many levels, including chromatin architecture modulation and interaction with initiation and elongation factors. In addition, it can also influence RNA splicing processing. H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20 represent the most studied histone methylation sites. Depending on the site and the degree of modification, lysine methylation can result in either activation or repression of transcription[35]. For example, H3K4me2/me3, H3K36me3, and H3K79me3, are associated with active genes and transcriptional activation, whereas H3K9me2/me3, H3K27me2/me3, and H4K20me3 are associated to transcriptional repression[60]. Among the activatory modifications, methylation of lysine 4 on histone H3 (H3K4me) is extremely important for transcriptional initiation. Trimethylated H3K4 was found to be enriched at many promoters of eukaryotic genes. In particular it marks the actively transcribed ones[61]. Methylation of lysine 36 on histone H3 is another modification related to transcriptional activation and specifically to transcriptional elongation[62]. Methylation of histone H3 on lysine 9 (H3K9me) is a well-conserved histone mark associated to transcriptional silencing[53]. Another mark related to transcriptional repression is methylation of lysine

27 (H3K27me) on histone H3. Polycomb repressive complex 2 (PRC2), through its catalytic subunit Ezh2/Ezh1, is able to perform all the 3 different methylation states of lysine 27, mono-, di- and tri-methylation[63], thus defining discrete genomic domains. The activity of this repressive complex will be one of the focus of this thesis. Importantly, the combinatorial pattern of histone methylation provides an important regulatory epigenetic platform. Interestingly, in mESC, about 20% of promoters are simultaneously marked by H3K4me3 and H3K27me3. These peculiar domains, characterized by the co-occurrence of both active and repressive marks, mark loci corresponding to developmental genes and regulators of the cellular state[64].

1.2 Polycomb proteins

1.2.1 Overview on Polycomb proteins

Polycomb group proteins (PcG) are essential epigenetic regulators that play key roles in development and cell-fate specification and transitions. They control many critical processes ensuring gene transcriptional repression in a highly cell-type specific manner. They were first described in *Drosophila melanogaster* almost 40 years ago where, in concert with the counteracting activating activity of Trithorax group proteins (TrxG), they were shown to repress homeotic (Hox) genes. This crosstalk ensures the correct establishment of proper segmentation along the anteroposterior axis of the fly's body[65-67]. This function of PcGs in negatively regulating developmental genes is strongly conserved also in mammals[68]. Mammalian orthologues began to be described in the early 1990s with the identification of Bmi1 (Psc in Drosophila) and the discovery of its direct role in cancer development[69,70]. Indeed, PcG proteins have been identified in all metazoans, exhibiting a remarkable degree of evolutionary conservation from *Drosophila*

to humans. In mammals, PcG proteins are part of several large multi-protein repressive complexes, termed Polycomb Repressive Complexes (PRCs)[71], whose composition is variable and context-dependent. Despite the great diversity of PRCs, two main classes of complexes have been extensively characterized: Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2). From a mechanistic point of view, when targeted to specific genomic loci, both groups of proteins promote gene repression via modification of histone tails and chromatin compaction. Indeed, at the molecular level, PRC1 catalyzes the monoubiquitylation of lysine 119 on histone H2A (H2AK119ub)[72] while PRC2 is responsible for the mono-, di- and tri-methylation of lysine 27 of histone H3 (H3K27me1/me2/me3)[73-75]. In the last years, much effort has been done to deeply characterize PRCs composition, activity, interplay and recruitment, however many aspects underlying their molecular mechanisms are still elusive. The importance of dissecting these mechanisms is highlighted by the strong involvement of these proteins in human diseases, including cancer.

1.2.1.2 PRC1

PRC1 is the complex with the largest number of reported subunits and is present in several sub-complexes that are biochemically distinct from each other with potentially different biological functions[76]. The *Drosophila* PRC1 core complex is composed by four proteins: Polycomb (Pc), Polyhomeotic (Ph), Posterior sex combs (Psc) or the closely related Suppressor of zeste 2 (Su(z)2 and Sex combs extra (Sce, also known as Ring)[68]. Mammalian homologues have been discovered for each of the core PRC1 proteins. Indeed, the composition of the mammalian PRC1 complex is much more diverse and depends on the cellular context. In general, it consists of Polycomb group ring finger

proteins (PCGF, Psc homolog), really interesting new gene 1A/B (Ring1A/B, Sce homolog), chromobox (CBX, Pc homolog) and polyhomeotic-like protein (PHC, Ph homolog). Moreover, in mammals several paralogues exist for all the subunits: five Pc (CBX2, CBX4, CBX6, CBX7, and CBX8), two Sce (RING1/RING1A and RING2/RING1B), three Ph (PHC1, PHC2, and PHC3) and six Psc, (PCGFs)[68]. All PRC1 complexes contain the core Ring1A or Ring1B (also known as Rnf1 and Rnf2, respectively) E3-ligases[77,78] that catalyze the monoubiquitylation of lysine 119 on histone H2A (H2AK119ub)[79]. Indeed, the monoubiquitin mark is one of the most abundant modifications and decorates almost 10% of endogenous H2A histone. Different sub-complexes can be defined according to the stable interaction of the catalytic subunit with one of the six different PCGF proteins, subunits that are essential for the E3-ligases functionality. Another classification takes into account the presence of CBX or Rybp/YAF2 (RING1/YY1-binding protein) subunits thus subdividing PRC1 complexes in two major classes. The so called "canonical" PRC1 complexes contain Cbx subunit that is able to recognize and bind H3K27me3, deposited by PRC2. They are characterized by the presence of PCGF proteins 2 or 4 (BMI1 and Mel18, respectively) and can be defined PRC1-PCGF2/4[CBX] subcomplexes[80] or PRC1.2 and PRC1.4[81] according to the nomenclature of the different classifications. Due to the presence of CBX subunit, canonical complexes are thought to act together with PRC2 to repress target genes expression. This mechanism is supported by the large overlap between PRC2 and RING1B target genes and by the global loss of RING1B chromatin binding at those target sites in the absence of PRC2 activity. However, upon loss of PRC2 activity, the global H2AK119ub levels remain largely unaffected suggesting that PRC1 enzymatic activity does not depend on PRC2[82]. On the other hand, "non-canonical" PRC1 rather that Cbx contain Rybp or its paralogue YAF2 subunit in addition to PCGF1, PCGF2 or PCGF4, PCGF3 or PCGF5, PCGF6 together with other subunits. In this case PRC1

24

complex acts in a PRC2-independent manner, in fact at the identified target genes RING1B and its mark are present despite PRC2 absence[82]. Interestingly, genomic profiling has shown that, in addition to a common subset of genes, these two kinds of complexes regulate different target gene sets[83] thus highlighting that also different recruitment mechanisms may exist. The different PRC1 sub-complexes are thought to have diverse intrinsically biochemical properties that ensure the targeting of specific subsets of genes in a highly context-dependent manner. Despite this intriguing hypothesis, our current knowledge about the biological and molecular functions of the different PRC1 sub-complexes is still largely not understood.

1.2.1.3 PRC2

PRC2 composition is indeed simpler than PRC1. *Drosophila* PRC2 core complex is formed by Enhancer of zeste [E(z)], Suppressor of zeste [Su(z)] and Extra sexcombs (Esc). In mammals core PRC2 complex is composed of Ezh2 (or Ezh1, homologs of E(z)), Suz12 (homolog of Su(z)) and Eed (Embryonic ectoderm development; homolog of Esc). Ezh2 is the catalytic subunit of the complex and through its SET domain performs the methylation of lysine 27 on histone H3[84]. These three core subunits are present in the complex in a 1:1:1 stoichiometry[85] and comprise the minimal composition necessary for the catalytic activity of PRC2. Suz12 and Eed proteins are necessary for the enzymatic activity of the complex[86,87], as demonstrated both *in vitro and in vivo* and are indeed essential for the correct development of an organism, consistent with the early embryonic lethality of mice deficient for Eed and Suz12[87,88]. Eed protein, in particular, is able to bind H3K27me3 and establish a positive feedback loop enhancing lysine methyl transferase (KTM) activity of the complex. PRC2 activity is boosted up to 7 fold in the

presence of H3K27me3 peptide, and consistently, the complex is more efficient in trimethylating K27 when in presence of pre-existing H3K27me3, thus suggesting a mechanism to maintain K27me3 levels through cell cycle progression[89].

Interestingly, Ezh2 is the only core subunit that has a paralog, Ezh1. The loss of Ezh2 causes early embryonic lethality whereas mice that lack Ezh1 are phenotypically normal and fertile, which indicates that all vital EZH1 functions can be carried out by EZH2[90]. Despite Ezh1 can partially complement Ezh2 in mediating H3K27 methylation activity, the two proteins show dissimilar expression and are found in diverse complexes, in terms of composition and function. Generally, Ezh2 forms a core together with both Eed and Suz12, whereas Ezh1 has been found alone or in a complex together with Suz12[91]. They show a partial redundancy in catalytic activity and localization but, whereas Ezh2 is generally considered to deposit di-and tri-methylation of H3K27 on repressed genomic loci, Ezh1 seems associated with mono-methylation of H3K27 on regions with active transcription[91,92]. The ratio between Ezh1 and Ezh2-containing PRC2 changes during cell differentiation, with Ezh2 levels decreasing and Ezh1 levels increasing upon differentiation[84,91,92]. Furthermore, Ezh2 is mainly expressed in proliferating tissues, whereas Ezh1 expression is found in dividing and differentiated cells[93]. This suggests that Ezh2-containing PRC2 complexes might establish the H3K27me3 repressive mark, whereas Ezh1-containing PRC2 complexes might contribute to the restoration of the H3K27me3 methylation profile after histone demethylation or histone exchanges. Despite H3K27me3 is widely considered PRC2 operative mark in terms of gene repression and is indeed the best characterized, PRC2 can perform all the three subsequent methylations of K27 (H3K27me1, H3K27me2 and H3K27me3)[63] in a stepwise manner. Recently, it has been shown that each different H2K27 methylated form has a peculiar deposition pattern along the genome thus ensuring a specific

26

functional outcome[63]. Several accessory subunits, dispensable for its intrinsic enzymatic activity, can sub-stoichiometrically associate to the core complex in order to ensure a finely regulated repression activity, by modulating the catalytic activity or the binding and recruitment of PCR2 to the target loci. They include: the histone chaperone RbAp48/46 (Retinoblastoma associated protein 46/48), required for the catalytic activity of the complex *in vivo*[94]; AEBP2, a zinc finger protein that enhances KTM activity in vitro[95]; the three Polycomb-like proteins (PCLs; PHF1, MTF2, PHF19) that retain a TUDOR domain, able to recognize H3K36me3 suggesting a role in initiating silencing of actively transcribed genes[96] and Jarid2 (Jumonji and ARID domain containing protein) that is able to bind GC-GA rich DNA elements and is involved in PRC2 recruitment at target genes and for proper ESC differentiation[80,97]. PRC2 structure is represented in Figure 1.
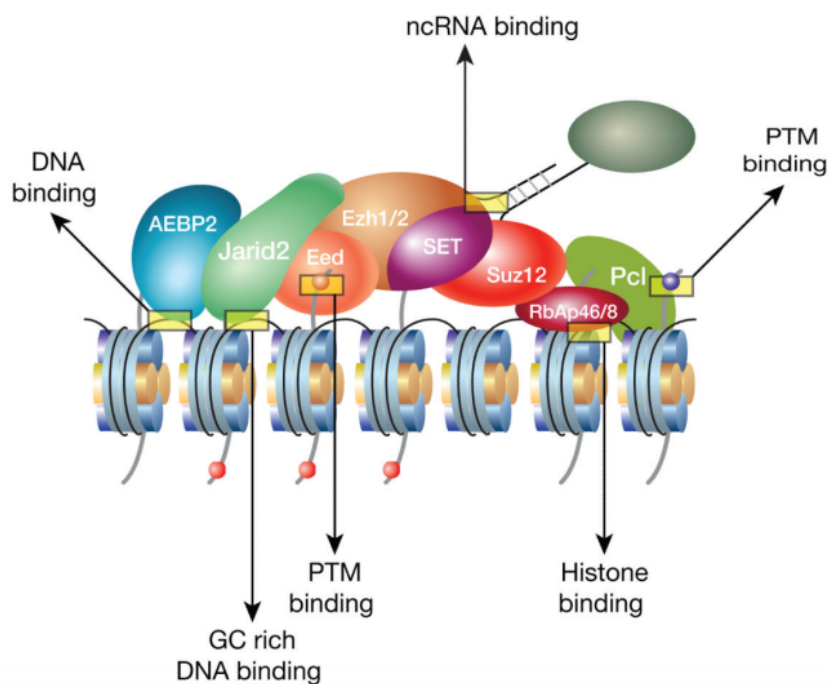
Figure 1:Principal components of PRC2.

Schematic representation of PRC2 where Ezh1/2, Suz12 and Eed core subunits as well as accessory subunits are depicted. Putative interactions responsible for PRC2 recruitment to chromatin are indicated. Adapted from Margueron and Reinberg, Nature, 2011[98].

1.2.2 Polycomb proteins recruitment

Polycomb complexes, to exert their repressive function, must be targeted to specific genes in specific cell types in a highly space and time regulated manner. In *Drosophila*, PcG recruitment to target loci is accomplished by Polycomb Response Elements (PREs), as first demonstrated at Hox genes level[99,100]. These are distal cis-regulatory elements, almost few hundred base pairs long, devoid of nucleosomes that can be present tens of kilobases (Kb) upstream or downstream of the target promoter, within introns or, in many cases, close to the transcription start site. The recruitment mechanisms in mammals appear more diverse and complex, representing a major research issue. Despite PcG core subunits are highly conserved, PRE-binding proteins seem to be absent in mammals. In general, it has been demonstrated that PRCs preferentially bind CpG rich promoters of their target genes[101] and that CpG-rich sequences mediate PRC2 recruitment[102]. Moreover, CpG islands lacking 5-methylcytosine are largely overlapping with H3K27me3, PRC2 and PRC1[102-104]. Following the classical hierarchical model of action, PRC1 and PRC2 cooperate in their repressive activity on target genes. This implies a two-step process: PRC2 is recruited as first and methylates H3K27, that is then recognized by PRC1, through its Cbx subunit. This model can be true for some targets but it is not comprehensive at all, taking into consideration the huge amount of non-canonical PRC1 complexes identified so far. First, the generally broad distribution of H3K27me3 does not fit the localized binding of PRC1[66,105]. Moreover, there are examples of PRC1 binding to sites without any apparent H3K27 methylation[82]. Furthermore, even

PRC1 dependent H2AK119ub1 has been suggested to play a key role in PRC2 recruitment thus highlighting the complexity of the recruitment mechanisms and the great crosstalk between PRCs[106]. Indeed, several different recruitment mechanisms for PRCs have been suggested, including association with near-stoichiometric interaction partners (such as PCL1-3, AEBP2, JARID2) or with specific transcription factors, chromatin signatures (CpG islands, histone modifications, histone variants), long non-coding RNAs (lncRNAs) and possibly the status of RNA polymerase II (Pol II). Since components of PRC1 and PRC2 generally do not have DNA binding properties, for both complexes it has been proposed that specific recruitment relies on crucial interacting proteins and/or transcription factors. For example, both PRC1 and PRC2 have been shown to interact and to be potentially recruited at promoters by the transcription factor REST[107]. Regarding PRC2, it has been suggested that Jumonji/ARID domain-containing protein Jarid2[97] and the members of the Polycomb-like family, the Pcl proteins, are responsible for PRC2 recruitment to target genes in mammals[96,108]. In addition to the association with specific interactors, also the methylation state of CpG islands can modulate the occupancy of different Polycomb complexes. Indeed, the mechanisms by which the PRC1 and in non-canonical sub-complexes are recruited on chromatin still remain an important open issue. As already reported, non-canonical PRC1 complexes, lacking Cbx subunit, must be recruited in an H3K27me3-independent manner. Recently, a role for Kdm2b, a histone H3K36me3/2 specific demethylase, in PRC1 complexes recruitment on chromatin has been proposed[109]. Further studies are necessary to characterize the molecular link between DNA methylation and PcG occupancy. Chromatin signatures in terms of histone modifications can indeed modulate PRCs recruitment. Pre-existing histone modifications such as H3K27me3, H2AK119ub, and H3K9me3 can facilitate PRC2 recruitment. For example, PRC2 was shown to bind to its

own catalytic product, H3K27me3, by the aromatic cage of Eed[89]. This interaction results in allosteric activation of the complex and propagation of the mark, as was shown *in vitro* and in *Drosophila*. Recently, long non-coding RNAs (lncRNAs) and the status of RNA polymerase II (Pol II) have also been involved in PRCs recruitment[110]. The discovery of the affinity of PRCs for RNA molecules has favored a model in which recruitment to target loci is mediated by ncRNAs. The involvement of ncRNAs in PRCs recruitment has been extensively studied in the context of X chromosome inactivation. Here, the accumulation of H3K27me3 is dependent on XIST expression, and the A repeats (RepA) of XIST have been shown to bind PRC2[111]. Several studies showed the association of PRC2 members despite differences in the kind of RNA and PRC2 component involved, thus the exact role of ncRNAs in PRC2 recruitment remains unclear[112,113]. Moreover, also the phosphorylation status of the serine residues within the carboxyl-terminal domain (CTD) of Pol II has been shown to correlate with the occupancy of PcG at certain target genes. The phosphorylation at Ser5 normally occurs right after initiation (Ser5P) while the phosphorylation at Ser2 and Ser7 positions are generally associated with productive elongation. Recently it has been shown that in mESC development-related PRC targets are generally associated with poised Pol II while PRC targets involved in metabolism are characterized with transcriptional active Pol II[110]. So, several different alternative and/or complementary mechanisms regulate PRC1 and PRC2 recruitment to chromatin whose complete understanding remains a matter of debate but is essential for our comprehension of PRCs-mediated transcriptional repression.

1.2.2 Polycomb proteins in mouse embryonic stem cells (mESCs)

Embryonic stem cells (ESCs), derived from the inner cell mass of the pre-implantation

embryo at the stage of blastocyst (3.5 days post coitum (dpc)), represent an excellent model system to study the molecular mechanisms underlying cell-fate establishment and transitions. Indeed, if cultured under the proper conditions, ES cells can self-renew and can be subsequently differentiated, thus recapitulating the same hierarchical steps that occur *in vivo*. Polycomb repressive complexes are highly expressed in ESCs where they preferentially bind CpG-rich promoters of genes encoding for transcription factors and molecules controlling development[90,114]. ESCs are characterized by pluripotency, the capability to give rise to all cell types of an embryo except the trophoectoderm. Pluripotency is maintained by a finely regulated balance between differentiation (prevented) and proliferation (favored). In general, ESCs display an open and permissive chromatin structure with low levels of DNA methylation and a huge abundance of activating histone modifications, such as H3K4me3 and histone acetylation. The hyperdynamic and open chromatin environment of ESCs must be finely regulated by chromatin modifiers thus, on one hand, balancing self-renewal and pluripotency and on the other to timely orchestrate the specific differentiation programs[115]. Upon differentiation, the overall chromatin structure shifts toward a tighter configuration with decreased transcriptional activity and concomitant accumulation of H3K27me3. In the last years, several knockout mice and derived ESC lines have been generated in order to investigate PcG function. Indeed, both PRC1 and PRC2 have been shown to be essential for proper ESCs differentiation and lineage commitment[116,117], whereas they do not seem to be required for embryonic stem cells self-renewal. PRC2 has a major role in the early steps of embryogenesis indeed Ezh2, Suz12 and Eed knockout mice display embryonic lethality soon after implantation. mESCs lacking Eed, Suz12, or Ezh2 can be derived from knockout embryos, yielding similar phenotypes with retention of self-renewal and pluripotency capacity but loss of H3K27me2/3 and *in vitro* differentiation defects. In fact

ES cells lacking Suz12, Eed or Ezh2 show aberrant de-repression of lineage-specific genes and aberrant differentiation[118]. Regarding PRC1, only Ring1b is not dispensable for mouse development. Indeed knockout of the PRC1 E3 ligase Ring1B causes gastrulation arrest and results in embryonic lethality. Ring1b-deficient mESCs maintain the expression of pluripotency markers[118], but show reduced levels of H2AK119ub1, a slight deregulation of some target genes and a loss of differentiation potential[78]. Ring1a/Ring1b double knockout mESCs lose the ability to self-renew after few passages and show defects in cell-cycle regulation, suggesting PRC2-independent roles of Ring1a/Ring1b[119]. Notably, depletion of both Ring1A and Ring1B impairs ESC self-renewal, indicating that Ring1 proteins (and hence PRC1) are essential for ESC identity[119]. Moreover, PRC1 seems to be critical for stem cell maintenance. For instance, embryonic neural stem cells require Ring1b for maintaining proper undifferentiated features[120]. PRC1 composition, in terms of Cbx protein, varies when ESCs start to differentiate. Cbx7 is the main component of canonical PRC1 in self-renewing ES whereas is substituted by Cbx2 and Cxb4 in differentiating cells[121,122]. Thus, Cbx2 and Cbx4 appear to replace Cbx7 in differentiating cells, thereby targeting PRC1 to a different set of genes. PRC1 and PRC2 epigenetically regulate the expression of developmental loci, impeding their expression through their repressive marks. Genome-wide studies have deeply tried to characterize ESCs epigenetic landscape. A large proportion of PcGs target genes marked by H3K27me3 and H2aK119ub present also the H3K4me3 active mark, deposited by the Trithorax/MLL complex[64,67]. These so-called "bivalent domains" are responsible for the regulation of many developmental and tissue-specific genes This peculiar signature, characterized by the co-occurrence of both active and repressive marks, is thought to poise the target genes in order to rapidly induce their expression upon a differentiation stimulus[101]. Indeed, these loci are silent or expressed at low levels

32

in ESCs and appear to be in a poised state ready for activation since the RNA polymerase II phosphorylated at serine 5, associated with paused transcription, also localizes to these loci[72]. In differentiating ESCs these domains tend to disappear and, depending on the expression of the gene in the cell context, retain the activating mark H3K4me3 or the repressive H3K27me3 one. Interestingly, Mohn and colleagues[123] have reported that new bivalent domains are acquired upon differentiation of ESCs in neural progenitors, indicating that epigenetic regulation mediated by PcG complexes is highly dynamic and cell type specific. According to this model new bivalent domains are acquired upon differentiation. This is supported by the fact that bivalent domains seem not to be restricted to ES cells but are present also in others multipotent stem cells, such as mesenchymal stem cells and hematopoietic stem and progenitor cells[124,125]. Recently, the methylation pattern of H3K27 in mouse embryonic stem cells (mESCs) has been deeply investigated. A proteomic analysis showed that more that 80% if H3K27 is indeed methylated, mainly di-methylated (ca. 70%), suggesting that the main enzymatic product of PRC2 relies on the deposition of H3K27me2. Its diffused deposition in intra- and inter-genic domains has a protective function in preventing the aberrant activation of enhancer-like elements counteracting H3K27 acetylation (ac)[63]. *In vitro* studies showed that H3K27me0 and H3K27me1 are better substrates for PRC2 than H3K27me2, thus supporting H3K27me2 as the major product of PRC2 activity[63]. H3K27me3, accounts for approximately 7% and is preferentially deposited at specific loci in correspondence to CpG-rich DNA regions, largely correspond to TATA-less gene promoters. In contrast to H3K27me2 and me3 that are associated to transcriptional silent regions, H3K27me1 (accounting for approximately 4% of all H3K27) is instead deposited throughout the gene bodies of actively transcribed genes in correspondence to H3K36me3 enrichment[63]. It has been proposed that H3K27me1 domains are formed via H3K36me3-mediated in

cis inhibition of PRC2-dependent H3K27me1 conversion to H3K27me2, while H3K27me3 deposition is achieved only upon stable interaction of PRC2 with chromatin, necessary to compensate its low enzymatic efficiency in methylating H3K27me2[63].

1.3 Epigenetics and cancer

1.3.1 Overview on involvement of epigenetic regulators in cancer

The chromatin environment is fundamental in the establishment and maintenance of cell identity. Accordingly, protein complexes that modulate chromatin structure are important for many aspects of mammalian development and stem cell function. It is indeed not surprising that their deregulation is involved in the pathogenesis of different human diseases, in particular cancer. Cancer is an heterogeneous disease and is generally considered a multistep process that, through the involvement of many different pathways and molecular actors, gradually leads to the loss of cellular identity towards the acquisition of cancer cell features[126]. It is now irrefutable many of the hallmarks of cancer, such as malignant self-renewal, differentiation blockade, evasion of cell death, angiogenesis capability, self-sufficiency in growth signals and tissue invasiveness are profoundly influenced by epigenetic aberrations, in addition to the well-characterized genetic ones. Indeed, epigenetic actors are affected at several different layers and in a highly context-dependent manner in cancer. Intense research has tried to elucidate the mechanisms by which chromatin modifiers and modifications promote cancer development or progression and to understand if these alterations could have a direct role in promoting tumorigenesis. The deep comprehension of the molecular mechanisms underlying epigenetic aberrations will allow the developing of novel selective inhibitors thus improving cancer therapies in terms of enhanced efficacy and

reduced side effects. The link between epigenetics and cancer was first suggested by gene expression and DNA methylation studies that, although purely correlative, highlight for the first time this possible connection[127]. These early observations have been strengthened by recent whole-genome sequencing studies that have identified recurrent somatic mutations in numerous epigenetic regulators in a vast array of cancers thus supporting the causative role of epigenetic aberrations in tumorigenesis[128,129]. In general, the cancer epigenome is characterized by aberrations and mis-regulation at multiple layers and that involve all the epigenetic actors. Indeed, changes in DNA methylation and histone modification patterns as well as altered expression levels or activity of chromatin-modifying enzymes can be observed in cancer. Furthermore, almost all human cancers display a deregulation in DNA methylation. As already reported, almost 70% of mammalian promoters harbor CpG islands whose methylation state has an important role in transcriptional regulation. Indeed, in normal cells despite an overall genome-wide methylation CpG islands are usually unmethylated[130]. In contrast, cancer cells exhibit two apparently opposing changes in the DNA methylation pattern. Despite a global hypomethylation is commonly observed in malignant cells, 5-10% of unmethylated, CpG promoter islands become abnormally methylated during malignant transformation[40]. Epigenetic hallmarks of cancer include global DNA hypomethylation and locus-specific hypermethylation of CpG islands. The overall hypomethylation can lead to genomic instability favoring chromosomal rearrangements[131] and reactivation of retrotransposons[132], in addition to the possible aberrant reactivation of some genes. Among the determinants of this mis-regulation in DNA methylation, the increased expression of DNMT enzymes and mutations affecting their activity have been widely investigated. CpG islands hypermethylation has quite clear effects, leading to the aberrant inactivation of key genes, such as tumor-suppressors thus promoting

tumorigenesis. On the contrary, the role of global DNA hypomethylation remains to be fully elucidated. For these reasons, understanding the role of aberrant DNA methylation is a key area of interest. Indeed, the reactivation of aberrantly silenced genes, through hypomethylating agents, represent an appealing therapeutic strategy. In addition to DNA methylation, also the histone modification patterns result affected in cancer. Many of the histone modifying enzymes and chromatin readers are found deregulated in cancer. Several are the examples of recurrent chromosomal translocations (MLL-CBP[133] and MOZ-TIF2[134]) or coding mutations (p300/CBP[135]) affecting HATs in a broad range of solid and hematological malignancies. Moreover, HDACs expression levels are also found altered in numerous malignancies. Much effort has been done to develop specific histone acetyltransferase inhibitors (HAT-I) and on the other hand, inhibitors of histone deacetylases (HDAC-I). Among HDAC inhibitors, two of them, pan-HDAC inhibitors Vorinostat and Romidepsin, have been granted FDA approval for clinical use in patients with cutaneous T cell lymphoma. Another suitable therapeutic target is represented by histone acetylation readers, in particular targeting their bromodomain. Highly specific small molecules against the BET family (BRD2, BRD3, BRD4, and BRDt) of bromodomain proteins have been recently developed[136,137]. So far, BET inhibitors have shown efficacy in NUT-midline carcinoma, characterized by recurrent translocations involving BRD3/4[137] and in a range of hematological malignancies[136,138,139]. Recurrent mutations in histone methyltransferases, demethylases, and methyl-lysine binders have also been identified in a large number of cancers. These mutations alter the catalytic activity of the enzymes or impair the ability of chromatin readers to recognize and bind specific epigenetic modifications. Indeed, recurrent translocations and/or coding mutations have been shown to affect histone methyltransferases, including EZH2. EZH2 role in cancer is an emerging and intriguing area of interest. In addition to its previously reported

36

overexpression in different tumor types[98], recent studies suggest a complex scenario in which EZH2 can act as an oncogene or as a tumor-suppressor depending on the cellular context. Histone phosphorylation pattern appears also deregulated in cancer. Recurrent mutations in signaling kinases are one of the most frequent oncogenic events found in cancer. Indeed, activating and inactivating mutations of histone kinases have been described in a range of malignancies. Despite the idea to therapeutically target specific epigenetic aberrations is highly intriguing, some issues must be solved. It is now quite clear that hematopoietic malignancies are more vulnerable to epigenetic therapies compared solid malignancies, that appear to be vastly more complex also from a genetic point of view. Moreover, hematopoietic niche offers a peculiar environment also in terms of drug exposure. This strongly suggests that not all cancers are equally susceptible to epigenetic therapies thus requiring a deep research of the best epigenetic therapeutic targets and combined strategies (also with conventional chemotherapies) in order to maximize the therapeutic efficacy of these compounds and also to reduce the likelihood of drug resistance.

1.3.2 Polycomb and cancer

Polycomb group proteins are transcriptional repressors that are crucial in regulating cell differentiation and determining cell identity. Thus, it is not surprising that their deregulation is frequently implicated in human pathogenesis. A strong link exist between PcG proteins and cancer, indeed several cancer types exhibit mis/deregulation of their expression and/or function[140]. Through the aberrant inhibition of tumor suppressors and activation of proto-oncogenes, cells may acquire a malignant phenotype characterized by loss of cell identity, persistent proliferative ability, resistance to cell

death mechanisms, bypass of cellular senescence programs and increased migratory/invasive potential. Indeed, there is increasing evidence that PRC complexes have a role in tumor progression and development. Overexpression of PcG genes has been observed in hematological malignancies and also in solid tumors, including medulloblastomas, and tumors originating from liver, colon, breast, lung, penis, and prostate[141]. In addition, PcG proteins have been shown to be recruited to target genes by physically interacting with a number of chimeric fusion proteins involved in leukemia, such as PLZF-RARA and TMPRSS2-ERG[142,143]. Moreover, PcG proteins activity can be impaired or enhanced by mutations in residues critical for their enzymatic activities. Although the involvement of chromatin repressive complexes in cancer is indisputable, their precise role in tumorigenesis is still incompletely understood. It is now clear that altering the homeostasis of PcG activity, by adversely affecting cellular identity, promotes tumorigenesis in a tissue-specific and/or even cell type–specific manner. Indeed, both the catalytic and non-catalytic functions of PcG proteins are involved. Even if both PRC1 and PRC2 have been found somehow deregulated in different kind of tumors, primarily PRC2 rather than PRC1, seems to have an essential role. Deciphering the precise role of repressive complexes in specific cancer types is necessary to improve anti-tumor therapies. In fact, keeping in mind the reversibility of epigenetic modifications, they represent suitable and attractive therapeutic targets.

1.3.2.1 PRC1 in cancer

One of the first indications of the link of PcG proteins with cancer was the identification, from a genetic screen, of Bmi1 as a c-Myc-collaborating oncogene in lymphomagenesis[70]. Indeed, many PRC1 genes are aberrantly expressed in a broad spectrum of human

cancers. Bmi1, for example, is overexpressed in a range of hematological cancers and solid tumors including squamous cell[144]. Moreover, its increased expression seems to correlate with poor prognosis[145,146]. This suggests that Bmi-1 has an important role in PRC1 action in addition to supporting the Ring1B-catalyzed enzymatic action at chromatin. The oncogenic function of Bmi1 and of the other PRC1 components has mainly been attributed to their repression of the CDKN2A locus. Conversely, the Bmi1 paralog Pcgf2/Mel18 expression, known for its tumor suppressive function, is frequently lost in different types of human cancers[147]. Both oncogenic and tumor suppressor roles have been proposed for CBX7 in the hematopoietic system[148,149] and some solid tumors[150,151]. In the context of targeting PRC1, application of small-molecule BMI1 inhibitors has been shown to reduce global H2AK119ub1 in colorectal cancer cells and decrease tumor mass in transplanted mice through a depletion of cancer initiation cells[152]. Another potential approach to targeting PRC1 implies the use of chromodomain inhibitors to target the CBX-component. Recently, Simhadri and collaborators reported the development of a chromodomain antagonist with 10- to 400-fold selectivity for CBX7 over other CBX family members[153].

1.3.2.2 PRC2 in cancer

PRC2 components are even more involved and compromised in cancer. Indeed, EZH2, SUZ12 and PCL3/PHF19 are frequently overexpressed in a variety of different human cancers[144]. Overexpression of EZH2 was first found in prostate and breast cancer. Then, it has also been reported in a large number of other solid tumors including bladder and ovarian cancers[154]. Furthermore, EZH2 overexpression is linked to aggressive and advanced metastatic stages of the disease and is strongly associated with poor clinical

outcome and prognosis[155]. The increased levels of EZH2 may result from different mechanisms. They can be associated to gene amplification or to the activation of different transcriptional signals and pathways, such as MEK-ERK-ELK1 or pRb-E2F pathways, which are often activated in cancer. However, recent studies suggest a much more complex scenario in which EZH2 can act as an oncogene or as a tumor-suppressor depending on the cellular context. Indeed, both hyperactivating and inactivating mutations have been described in different kind of tumors. Interestingly, EZH2 function may be impaired in different ways and it seems that the type of EZH2 dysregulation often correlates with the malignancy. EZH2 overexpression is mainly found in solid tumors, whereas activating or inactivating mutations are preferentially identified in hematologic malignancies. Heterozygous missense mutations resulting in the substitution of tyrosine 641 (Y641) within the SET domain of EZH2 have been recently found in 22% of patients with diffuse large B-cell lymphoma and in almost 7% of follicular lymphoma[156]. These mutations lead to an hyperactive form of EZH2 yielding increased levels of H3K27me3[156,157] thus supporting the Ezh2 role as an oncogene[158]. In contrast, loss-of-function mutations in EZH2 gene have been described in some myeloid malignancies[159,160] and T-ALL[161] thus suggesting a tumor-suppressive role of Ezh2 in these contexts. Since EZH2 plays a prominent role in tumorigenesis it is soon emerged as a potential therapeutic target. Indeed, the search of EZH2 pharmacological inhibitors yielded several promising molecules. In particular, different molecules including EPZ005687, EPZ-6438, EI1, UNC1999 and GSK126 act as EZH2 S-adenosylmethionine (SAM) competitive inhibitors. They all bind to SAM pocket of catalytic SET domain thus inhibiting H3K27 methyltransferase activity. They showed promising results in reducing H3K27me3 levels, decreasing proliferation and increasing apoptosis in lymphoma cell lines carrying SET domain mutations and markedly increased survival in mouse

40

xenograft models[162-164]. Since EZH2 requires the association with Eed and Suz12 to exert its enzymatic activity, another strategy to inhibit aberrant EZH2 methyltransferase activity is to disrupt PRC2 assembly by targeting EED-EZH2 interface using a hydrocarbon-stapled peptide that mimics the α-helical EED-binding domain of EZH2 (SAH-EZH2 peptide). This strategy has been shown to impair mutant PRC2 complex formation and H3K27me3 accumulation and to induce growth arrest and differentiation of MLL-AF9 driven leukemic cells[165]. The identification of Ezh2 mutations with opposing outcomes on H3K27me highlights the importance of the balance of this histone mark for cell homeostasis and emphasize the need for a complete and deep comprehension of altered epigenomes in order to develop novel therapeutic molecules and to select which cancer subtypes could benefit from these drugs.

1.3.2.2.1 EZH2 hyperactivating and inactivating mutations

Morin and colleagues in 2010[156] described for the first time the occurrence of missense mutations on EZH2 SET domain in a significant percentage of lymphomas. Indeed, heterozygous somatic missense mutations of *EZH2* have been described in almost 7% of follicular lymphoma and up to 22% of germinal center B-cell-like diffuse large B-cell lymphoma (GCB-DLBCL)[156], the two most prevalent types of mature B-cell lymphomas. These mutations affect preferentially the SET domain of EZH2 at tyrosine 641 (Y641N, F, S, or H) but also, less frequently (1%–3% of B-cell lymphomas) at alanine 677 (A677G) and alanine 687 (A687V) residues[163,166,167]. Despite Y641 mutations were initially considered loss-of-function due to a bias in the *in vitro* methylation assay, now their gain-of-function (GOF) nature is widely accepted. In fact these aminoacidic substitutions result in an hyperactive form of EZH2 thus leading to a genome-wide increase in the

H3K27me3 deposition[156]. Structural studies have revealed that these mutations affect critical residues within the catalytic pocket of the SET domain thus impairing the normal enzymatic activity of EZH2. While wild type EZH2 efficiently monomethylates H3K27 but has weaker activity for the subsequent reactions that lead to di- and tri-methylation, Y641 mutations impair the ability of EZH2 to modify unmethylated substrates while enhancing its efficiency in converting H3K27me2 into H3K27me3. In general this results in a broad accumulation of H3K27me3 as well as a reduction of H3K27me2 levels[156,158]. These mutations normally occur in an heterozygous state, in fact due to the their reduced H3K27me1/me2 activities, EZH2-Y641 mutants cooperate with their wild type counterparts in order to achieve the increased conversion of H3K27me0 to H3K27me3[158,168]. EZH2 expression is finely regulated during B-cell maturation within the germinal center thus establishing the suitable chromatin landscape for cells expansion. Then EZH2 activity is reduced thus leading to gene activation and allowing cells to exit the GC compartment and accomplishing terminal differentiation. The presence of Y641 mutation results in an enhanced H3K27 trimethylation and so an increased silencing of EZH2 targets thus blocking differentiation and promoting proliferation and tumorigenesis. Much effort has been done to understand the oncogenic properties of these peculiar mutations. Indeed, transgenic mice expressing the EZH2-Y641F mutant in lymphocytes displayed a global increase in trimethylated H3K27 in spleen cells and developed lymphomas when combined with Eμ-Myc expression[168]. Recently, Souroullas and colleagues[169] have tried to elucidate the molecular mechanisms underlying Y641F mutation through the generation of a mouse model permitting the conditional expression of the mutant protein 'knocked-in' to the native locus. Indeed, they tested the ability of this mutation in promoting lymphoid and solid malignancies by itself, and in co-operation with other oncogenic events. Their results showed that Y641F

42

mutation does not only increase H3K27me3, but rather redistributes the H3K27me3 mark across the genome. This redistribution leads to complex effects on transcription with some loci enriched for the mark and so silenced and some others activated. Indeed, the distribution pattern shows a preference for H3K27me3 deposition over broad regions—at gene bodies and intergenic regions—at the expense of focal peaks near the TSS compared to wt. Interestingly, the authors showed also that, at least in the specific case of Y641F mutation, wild type EZH2 function is dispensable for tumorigenesis in cells harboring Ezh2 Y641F. Recently, the crystal structure of the yeast (*C. Thermophilus*) PRC2 complex has been solved[170]. This structure suggests that the Y826 aa (homolog to Y646) participates in the formation of the substrate channel. Wild type form of EZH2 is indeed allosterically activated when EED binds H3K27me3. Mutations in the Y646 or in adjacent residues can therefore alter the complex' substrate specificity. The presence of Y646 mutation may reduce the need for H3K27me3-EED stimulation of the enzymatic complex thus creating a more promiscuous complex. It has to be pointed out that the precise mechanisms underlying these gain-of-function mutations have just started to be elucidated in detail. Most DLBCL cell lines harboring EZH2 gain-of-function mutations are sensitive to EZH2 inhibitors so gain-of-function EZH2 represents an attractive therapeutic target for lymphoma. Recently, EZH2 Y641 mutations were also identified in melanomas, albeit at low frequencies (ca. 2%). A recent study by Barsotti and colleagues[171] started to evaluate Y641 mutations role in melanoma tumor biology. The endogenous or ectopic expression of mutant EZH2 in skin or melanoma cell lines led to a global increase in H3K27me3 and was associated to huge morphological changes when cells were grown in 3D culture. Indeed, mutant cells showed a decrease in cell contractility and an increase in cell migration. Moreover, the aggressive 3D morphology of EZH2 GOF-expressing melanoma cells was partially reversed by EZH2 catalytic

inhibition thus highlighting EZH2 critical role in this pathogenetic model. The discovery of gain-of-function EZH2 mutations in non-hematopoietic malignancies but also in solid tumors, such as melanoma, suggests that the oncogenic proprieties behind aberrant H3K27me3 may not be restricted to a specific cell-type and/or context. However, the molecular mechanisms underlying these mutations and the causal link between H3K27me3 accumulation and cell transformation remain largely unknown. According to a possible role of EZH2 as tumor suppressor in certain cellular context, inactivating EZH2 mutations have been identified in about 10%-23% of various subtypes of myelodysplastic syndromes and myeloproliferative neoplasms, (MDS/MPN)[156,159,160]. Moreover, also a small fraction of chronic myelomonocytic leukemias and chronic myelomonocytic leukemia-derived acute myeloid leukemias harbor mutations in EZH2[161]. Interestingly, loss of function mutations are not limited to aberrant myelopoiesis. Indeed, Ntziachristos and colleagues[161] identified EZH2 or SUZ12 mutations in 25% of primary T-acute lymphoblastic leukemia (T-ALL) samples. Ernst and collaborators[159] in 2010 screened a cohort of 614 patients affected by different myeloid disorders and identified recurrent somatic frameshift, nonsense and missense mutations throughout EZH2 gene. The truncating mutations were dispersed throughout the gene, whereas the missense mutations targeted highly conserved residues within SET and adjacent CXC domains, both required for histone methyltransferase (HMT). By Western Blot they assessed in two cell lines, ELF-153 and SKM-1, endogenously expressing mutant EZH2 (R509G and Y646C, respectively), the almost complete absence of H3K27me3 compared to control lines without mutations. Then the authors focused their attention on 4 CXC-SET domain mutants identified in patients: Y731D, C576W, Y646C and R690C. Immunoprecipitated mutant complexes from Sf9 insect cells, previously infected with baculoviruses expressing Flag-tagged *EZH2* mutants as well as

44

EED and *SUZ12,* were assayed for H3K27 methyltransferase activity. All the four mutations were shown to abrogate or greatly reduce EZH2 catalytic activity *in vitro.* Surprinsingly, Y646 residue, typically mutated in a gain-of-function manner in lymphomas, is found mutated in C in MDS where shows an inactivating effect on EZH2 catalytic function. In general, the identified mutations can be monoallelic or biallelic and are all predicted to inactivate the methyltransferase activity of EZH2. Indeed, they are all associated with reduced H3K27me3 and derepression of EZH2 classical targets. Myelodysplastic syndromes represent a heterogeneous group of clonal pre-malignant hematopoietic disorders characterized by cytopenia, ineffective hematopoiesis, and increased risk of progression to secondary acute myeloid leukemia. Mutations in *EZH2* seem to be associated with worsened overall survival, independent of known clinical predictors of worsened survival in MDS. *EZH2* has indeed a tumor suppressor function in myeloid malignancies. A detailed elucidation/dissection of the molecular mechanism underlying EZH2 inactivating mutations is critical for designing suitable therapeutic approaches for these patients.

1.3.2.2.2 Histone mutations

Although mechanisms could be different, misregulation of H3K27 methylation is very common in tumorigenesis. Indeed, EZH2 activity may be impaired also in an indirect way, through mutations affecting its specific substrate, histones. In 2012, two studies simultaneously described for the first time the occurrence of recurrent somatic heterozigous mutations resulting in amino acid substitutions at two critical residues in the tail of histone H3 (K27M, G34R/G34V) in up to one-third of pediatric HGG (high grade glioblastoma) including glioblastoma multiforms (GBM) and diffuse intrinsic

pontine gliomas (DIPG)[172,173]. Both groups reported mutations in the H3F3A gene, coding for histone variant H3.3. One of the groups also reported mutations in the HIST1H3B gene, encoding histone H3.1[173]. Indeed, up to 78 % of diffuse intrinsic pontine gliomas (DIPGs) carry lysine-to-methionine substitution at lysine 27 (K27M) whereas 36% of non-brainstem gliomas have been shown to carry either K27M or glycine-to-ariginine or glycine-to-valine (G34R/V) mutations at position[174]. The mutated residue associates with the tumor topology within the central nervous system. In fact H3.3 K27M is predominantly found in gliomas of midline brain locations (spinal cord, thalamus, pons, brainstem) whereas G34R/V mutations characterize tumors located in cerebral hemispheres[175,176]. Additionally, H3.3 K27M pediatric glioblastoma are diagnosed at earlier ages and have shorter overall survival rates in comparison with glioblastomas bearing mutations at the H3.3 G34[172,176]. These data suggest that K27M and G34R/V mutations may arise from independent cellular precursors and niches within the brain, and possibly at different developmental time points. Moreover, mutations affecting histone H3 have been identified at high frequencies also in chondroblastoma (K36M) and in giant cell tumor of bone (G34W/L), diseases of adolescents and young adults[177]. 95% of chondroblastomas present lysine 36 to-methionine (K36M) substitutions whereas G34W/L substitution characterizes ca. 92% of giant cell tumors of the bone. Notably, nearly all K36M mutations occur in H3F3B (~90 %) gene, rather than in H3F3A as happens for K27M mutation in glioblastoma. This discrepancy cannot be explained by differences in codon usage between H3F3A and H3F3B and so other mechanisms may be involved. These two H3.3 mutations exhibit remarkable tumor type specificity: G34/W/L for chondroblastoma and K36M for giant cell tumor of bone. Interestingly no H3.3 K27M mutations have been detected in samples of all types of bone tumors. This strongly suggests that the molecular basis of the onset and development of bone tumors triggered

46

by H3.3 mutations is independent from that of brain tumors. These mutations affect three specific aminoacids within the N-terminus tail of histone H3 that is usually subjected to a variety of post-translational modifications. In general, they alter the ability of critical residues of being properly modified by histone writers and recognized by specific readers, leading to alterations in the chromatin structure and transcription. Indeed, K27 and K36 are well known functionally critical targets for PTMs. G34 is not substrate for any known PTM but its mutations may affect epigenetic modifications at the adjacent residue K36, important target for activating methylation. Starting from Lewis and colleagues in 2013[178], many studies have shown that the K27M mutation acts in a dominant-negative manner by competitively inhibiting the methyltransferase activity of EZH2. Despite the fact that mutant H3 represents only ca.5% of the total H3 pool and that only one allele among the 30 alleles encoding H3 in the cell is mutated, H3K27M gliomas exhibit a dramatic deficiency of global H3K27 methylation, with a pronounced reduction in H3K27me2 and me3 levels[178,179]. From a mechanistic point of view, H3K27M appears to bind the SET domain of EZH2, thus poisoning it and directly interfering/blocking its catalytic function. Moreover, Bender and collaborators[179] showed that H3K27M leads also to a loss of DNA methylation at many sites that, in co-occurrence with decreased H3K27me3, further leads to transcriptional activation. Despite the global reduction in K27methylation in gliomas expressing K27M H3.3, ChIP-seq analysis has revealed that some genomic loci escape this effect and can accumulate high levels of K27me2/3 marks[179,180]. H3.3K27M mutation alters the global epigenetic landscape thus affecting gene expression programs that may drive to gliomagenesis. The molecular basis for tumorigenesis associated with the H3.3 G34W/L mutation in skeletal cells or H3.3 G34R/V mutation in brain cells is less clear since G34 residue is not subjected to known PMTs. Nevertheless, due to the proximity to K36, H3.3 G34

mutations are thought to interfere with the action of the SETD2 methyltransferase, responsible for H3K36 tri-methylation[178]. Indeed, it has been demonstrated that the introduction of G34R/V mutant histone results in substantial decreases in K36me3 on the same, and nearby, nucleosomes[178,180]. The substitution of G34 with a bulky, hydrophobic or charged amino acid residue seems to perturb the nucleosome environment thus compromising the accessibility of the adjacent K36 residue. Regarding H3.3 K36M mutation, it was reported that the presence of the mutation was associated to a global reduction of H3K36me3 di- and trimethylation of histone H3 at lysine-36 (H3K36me2 and H3K36me3) presumably through the inhibition of SETD2 methyltransferase[178]. This year Fang and colleagues[181] confirmed that the presence of K36M, in human chondroblastomas and in chondrocytes harboring the same genetic mutation, results in a global reduction in H3K36 methylation due to the inhibition of at least two H3K36methyltransferases, MMSET and SETD2. This reduction in H3K36me2 and H3K36me3 at gene bodies has a significant correlation with changes in gene expression. Moreover, chondrocytes harboring K36M mutation exhibit several hallmarks of cancer cells, including increased ability to form colonies, resistance to apoptosis, and defects in differentiation. Although H3F3A and H3F3B have identical amino acid sequences, the H3.3K36M mutation occurs predominantly in H3F3B gene whereas the other mutations are almost exclusive to H3F3A[182]. Furthermore, these different mutations also appear to correlate with tumor topology or even tumor-type. For instance, the K27M mutation has been found only in gliomas in midline whereas 34R/V mutations predominantly associate with pediatric glioblastoma multiforme (GBM) in the cerebral hemispheres. The majority of K36M mutations has been found in chondroblastoma, while G34W and G34L mutations have been found only in giant-cell tumor of bone[182]. Interestingly, to date, no K27M mutations have been observed in bone

48

or cartilage tumors, and the K36M mutant has not been found in glioblastoma. It is widely accepted that H3.3 mutations play a driver role in oncogenesis by reshaping the epigenome through global and local alterations of histone modification patterns. However, the molecular mechanisms underlying these mutations and their biological outcomes have only started to be elucidated and many questions remain to be answered. More in depth studies are needed to dissect their functional roles in order to provide more powerful therapeutic approaches/treatments for these pediatric patients

1.4 CRISPR/Cas9 system

1.4.1 Overview on CRISPR/Cas system

Genetic manipulation is a powerful and versatile tool to create suitable *in vitro* and *in vivo* models thus allowing the study of gene function in a given cellular context. CRISPR/Cas-based gene editing system has revolutionized the world of cell biology. Indeed, this system is highly versatile and enables many diverse types of genome engineering approaches overcoming the limitations of traditional genetic manipulation tools. Notably, at the beginning of my PhD, this technology was relatively new and just emerged as tool to induce precise genetic alterations. The CRISPR system was first discovered in bacteria and archaea as an "adaptive immune system", a defense mechanism against foreign genetic material[183]. This system is based on the action of an RNA-guided endonuclease (Cas protein) that can target and cleave foreign DNA by means of RNA molecules that serve as guides. A functional CRISPR-Cas system is made of two components: the CRISPR locus/array located in the host genome and the effectors, the Cas proteins. This defense system comprises a multistep process that can be broadly divided into 3 stages. During the first stage, named acquisition or adaptation, the host

incorporates specific small fragments of foreign genomic elements into the endogenous CRISPR (clustered regularly interspaced short palindromic repeats) locus. It consists of an array of partially palindromic DNA repeats (30 bp) spaced by specific unique DNA sequences, called protospacers, that have been acquired by bacteria from previous rounds of infection. These spacer sequences are indeed used in future invasions to target and destroy invading nucleic acids (CRISPR RNA or crRNA). In fact this creates a sort of memory surveillance system. In the second stage, named expression, in presence of a specific stimulus (i.e. reinfection), these spacers are transcribed from the CRISPR locus and processed into small non-coding RNAs that guide, through a sequence-specific recognition process, Cas protein complexes to the recognition and subsequent cleavage of invading foreign genetic material (third stage, interference)[183,184]. On the basis of different features including phylogeny, sequence, locus organization and Cas proteins, three types of CRISPR/Cas system have been distinguished: Type I, Type II and Type III. In type I and II systems, for proper targeting the protospacer has to be flanked by a system-specific, highly conserved CRISPR motif, the protospacer adjacent motif (PAM)[185]. Indeed, PAM allows the recognition of foreign DNA thus, only the PAM-bearing invading sequence are targeted for destruction. Type II CRISPR-Cas9 system is the simplest among the three and is the one that has been engineered for research usage.

1.4.2 CRISPR/Cas9 system: mechanism of action and design

The CRISPR/Cas system, in particular type IIA from *S. pyogenes,* has been engineered in order to obtain a versatile molecular tool to perform eukaryotic genome engineering. In type II system a single Cas protein, Cas9, uses only two small RNAs (a mature crRNA (CRISPR-RNA) and a trans-acting tracrRNA (trans-activating crRNA)) for sequence-

specific DNA cleavage[184]. It forms a complex with this "guide RNA" molecule and localizes to a target DNA sequence following simple guide RNA-genomic DNA base-pairing rules thus leading to specific DNA cleavage. To date, the Cas9 has been the most broadly used Cas protein for CRISPR/Cas experiments. For most applications, a 20-nt chimeric single guide RNA (sgRNA) recapitulating the structure and function of the tracrRNA –crRNA complex[186] is used for the Cas9 targeting. Cas9 protein contains two independent endonuclease domains: HNH endonuclease and the other one to the RuvC endonuclease[186]. Each domain can cut one strand of DNA at the target site: the HNH domain cleaves the complementary DNA strand (the strand forming the duplex with gRNA) whereas the RuvC-like domain cleaves the non-complementary DNA strand. The target DNA sequence must be both complementary to the guide RNA and be followed by Cas9-specific PAM sequence[185,187]. Theoretically, Cas9 could be programmed to target and cleave any genomic region of interest, the only requirement for the selection of target sites, in addition to the complementarity with the 20 nt sgRNA, is the presence of the appropriate PAM immediately downstream of the target site. Indeed, Cas9 enzyme, guided by a proper gRNA, introduces double-stranded breaks (DSBs) at specific target sequences within the genome. DSBs are typically repaired by non-homologous end-joining (NHEJ). In this error-prone pathway, the ends of a DSB are processed by endogenous DNA repair machinery and rejoined, which can result in random insertions/deletions (indel) mutations at the site of junction. Indel mutations occurring within the coding region of a gene can result in frameshifts or in the creation of a premature stop codons, thus resulting in gene knockout. Alternatively, DSBs can be repaired also by homology directed repair (HDR), that by using a proper DNA repair template, allows a high fidelity and precise editing. For the introduction of small changes, i.e. point mutations, a single-stranded oligo DNA (ssODN) may be used. Notably, the

mechanism by which the DSB induced by Cas9 is repaired determines which kind of modification is introduced at the target site[188]. This RNA-based targeting system intrinsically offers a lot of different experimental options. Interestingly, Cas9 can be converted into a nickase that is able to create only SSB (single strand breaks) by inactivating one of the 2 domains via point mutations. Both domains can be mutated thus converting Cas9 into a nickase: D10A in RuvC and H840A in HNH[189-191]. In particular, Cas9n mutant has been generated mutating the RuvC (D10A) domain. This mutated form of the enzyme, acting as a nickase, is usually used in combination with 2 guides in a so called "double nickase approach" that allows to improve the specificity of the system. The nickase is guided by a pair of appropriately spaced and oriented sgRNAs to simultaneously introduce single stranded nicks on both strands of the target DNA site. DSB occurs only if both sgRNAs locate within a defined space. This strategy doubles the number of bases that need to be specifically recognized at the target site thus increasing specificity. Moreover, the pair of guides must be designed such as 5' overhangs are generated upon nicking and with a typical offset, defined as distance from the end of guides, between -4 and 20 bp in order to get adjacent nicks on opposite strands[192]. The ideal genome-editing tool would edit any genomic locus with high efficiency and high sequence specificity without undesired effects. Despite the great effort in the field, unfortunately, a perfect tool has not yet been developed. Indeed, multiple factors, many of which are beyond experimental controls, determine the success of CRISPR experiments, such as the quantity of Cas9 proteins and gRNA, chromatin accessibility of the targeting loci and cellular response to CRISPR-induced DNA lesions. The goodness of a CRISPR/Cas9 experiment relies on two parameters: specificity and efficiency. Specificity can be defined as the probability that Cas9 will target the designed locus compared to other undesirable loci (off-target effects) whereas efficiency is defined as

the probability of the locus of interest to be modified by Cas9 in the context of a pool of available targets. The goal is for sure to minimize the off-target effect and maximize the on-target effect. Despite its huge powerfulness, the system has some intrinsic limitations. Indeed, it has to be taken into consideration that it could have off-targets sequences, that can be cut as well as the designed target sites, or even with greater efficiency. In general, mismatches in the first 12 nucleotides of the gRNA (seed sequence) and the DNA target are not well tolerated thus suggesting the critical role of the PAM-proximal region for the targeting specificity. Mismatches beyond this region can be indeed compatible with efficient cleavage[189,193]. In order to plan a CRISPR/Cas9 experiment several bioinformatic tools are now available for the selection of proper sgRNAs with minimal sequence homology across the genome. In general, given a particular locus, the species of interest and on the basis of the experiment (i.e Cas9 vs Cas9n) they rank guide RNAs by computationally predicted specificity and suggest likely off-target sites taking into consideration the usually tolerated number of mismatches by Cas9. The developing of new methods of off-targets detection and quantification has become an important research focus. Off-target mutations represent the major limitation of CRISPR/Cas9 system. Notably, off-target activities depend on several parameters, including the amount of Cas9 protein, the structure and nature of the sgRNA sequence, the targeted cell type and the cellular state and are not easy to predict. Many attempts have been done to improve the system specificity and reduce off-targets. For instance, the use of truncated gRNA was shown to decrease up to several orders of magnitude displayed off-target effects[193]. Several strategies have been employed to improve the precision of the system, for instance the use of a "paired nickases" approach that requires two separate specific binding events in order to induce DNA cleavage[192] or the titration of the ratio between srRNAs and Cas9 protein and the use of novel Cas9 variants[194,195]. Morevoer,

the use of orthogonal Cas9 will allow to perform multiplex genome editing simultaneously[196,197]. Another strategy to improve the specificity of Cas9 is to decrease its activity or lifetime in cells without its ability to modify the target locus. For example, the direct delivery of Cas9: sgRNA ribonucleotide protein complexes (RNPs) to cells, which results in transient Cas9 activity, rather than plasmid transfection, can increase the ratio of on-target genome editing to off-target genome editing thus limiting the Cas9 window of action[198]. Another issue is the control of the repair process to ensure the desired genetic correction. For many applications, where a precise editing is required, a major limitation to the use of this tool is the introduction of stochastic indels at the site of DSBs due to NHEJ. HDR efficiency depends on many factors, including cell type, cell state and location of the target DNA, but is indeed typically quite low (<5%) and is competitive with NHEJ[199]. Different strategies have been applied in order to improve HDR such as the inhibition of endogenous repair components of NHEJ system[200-202]. Another option is to chemically synchronize cells to arrest in G-phase before the targeting[203]. However, these exogenous treatments, being perturbative to cells and cell-specific, have only limited relevance in a therapeutic context. Interestingly, the design of the donor DNA template can influence HDR-based genome-editing efficiency. Indeed, an asymmetrically ssODN can improve HDR-mediated genome-editing rates[204]. Moreover, the re-processing of the desired HDR product can be avoided by the introduction, on the template DNA, of silent mutations affecting the PAM or the adjacent residues thus impeding the Cas9 to cut the target after the repair[205]. A new strategy, called "base editing", has been recently proposed to introduce point mutations in an RNA-programmed manner that does not rely on HDR or double-stranded DNA breaks[206] and implies the fusion to dCas9 of a cytidine deaminase enzyme. Improving the specificity of CRISPR-based genome-editing in terms of the use of new enzymes, a better selection of guide RNAs, new delivery

54

protocols and the development of novel off-target detection methods, is indeed mandatory for its future use as research tools and even more for their use in therapeutics.

1.4.3 CRISPR applications

CRISPR/Cas9 system has been widely used as gene-editing tool for the mechanistic interrogation of the genome in diverse types of cells and organisms. Beyond classical knock-out and knock-in applications, it is becoming more and more evident its intrinsic versatility. Indeed, many are its possible applications including transcriptional control and epigenetic regulation, large-scale genetic screens, generation of animal models, genomic imaging and lineage tracing[206]. Regarding transcriptional regulation, nuclease-dead dCas9 has been generated by introducing point mutations into the HNH and RuvC domains to abolish endonuclease activity. This modified enzyme provides a platform to transcriptionally and epigenetically manipulate the genome, without altering its sequence. It functions as an RNA-guided DNA-binding protein that can be fused to transcriptional effector proteins to perform gene interference (CRISPRi) and activation (CRISPRa)[207-209]. This option is particularly important in those cases in which the KO is lethal for the cell or the organism thus allowing a finely tunable repression system[210]. Different strategies have been employed to perform CRISPRa including dCas9 fusion to a multimeric peptide array (SunTag) that allows the recruiting of multiple copies of the VP64 domain[211]. Moreover, dCas9 fused to epigenetic-modifying enzymes has been used to introduce locus-specific epigenetic modifications in the genome. One example is the fusion of dCas9 to the catalytic domain of the human acetyltransferase p300 (p300core), which allowed acetylation of histone H3 Lys27 (H3K27) and upregulation of genes when

binding to proximal or distal enhancers[212]. One of the powerful applications of CRISPR/Cas9 technology is the genomic high-throughput functional screening. A library of hundreds of thousands of sgRNAs can be used in combination with Cas9 or dCas9 fusion proteins in order to systematically knock out, repress, or activate genes on a large scale thus allowing, for instance, the identification of genes or regulatory elements responsible for cell growth and drug resistance. The capability of CRISPR/Cas system to perform sequence-specific gene editing in many organisms combined to the availability of genetic information regarding human diseases led to the generation of disease cell and animal models thus enabling the modeling and the functional study of many even complex genetic diseases such as cancer. Indeed, CRISPR/Cas9 technology has been widely used for the generation of suitable cellular and animal models recapitulating the genetic mutations or aberration found in patients. Moreover, many are the examples of CRISPR-edited induced pluripotent stem cells (iPSCs), isogenic models that enable the study of human genetic variants in a wide variety of tissues in cell culture[213-215]. For the generation of cellular models, Cas9, sgRNA and ssODN for HDR can be easily introduced into the target cells using transient transfection of plasmids. Additionally, the multiplexing capabilities of the approach offer a promising approach for studying common human complex diseases. CRISPR/Cas9 can be applied also *in vivo* for generating genetically-engineered animal models crucial for the study of complex cellular and pathophysiological processes. To date, the CRISPR/Cas9 gene-editing approach has been established not only in mouse but also in many other animal models, including fly, rabbit, and monkey[186]. Compared to previous genome-editing tools, whose design was usually very challenging, CRISPR/Cas9 system provides an easier approach to obtain transgenic animals. In fact, nucleic acids encoding the Cas9 protein, target-specific sgRNAs and oligos (for HDR) can be directly injected into fertilized zygotes to

56

generate gene-modified animals[215-217]. Indeed, the use of CRISPR/Cas9 platform allows to bypass the typical ES cell targeting stage thus dramatically reducing the time needed for the generation of mutant animals. Given the enormous potential of CRISPR/Cas9 technique, the subsequent step is for sure its use for therapeutic purposes. Regenerative medicine has the goal to replace unhealthy or diseased cells with healthy ones. One of the possible approaches is cell therapy, in which primary cells are genetically manipulated and then implanted into patients in order to cure the disease. Indeed, CRISPR/Cas system was used to correct several disease-associated genetic lesions in a variety of human disease including Duchenne muscular dystrophy (DMD)[218], Fanconi anemia[219], hemophilia[220], cystic fibrosis[221], and beta thalassemia[222]. Moreover, CRISPR can be used, not only for corrective purposes, but also to obtain engineering "therapeutic cells". Patient's cells, once edited *ex-vivo* can be reinfused in order to facilitate the treatment, for example enhancing immunological anti-tumor response. As a powerful, versatile, quite simple and robust gene-editing and regulation tool, CRISPR/Cas9 technology is driving a revolution both in research and therapeutics. The field is rapidly progressing, with improvements in guide RNA selection, protein and guide engineering, use of novel enzymes, better delivery methods to increase the specificity of the system thus reducing off-target effects. Furthermore, for its application in genomic therapy also ethical issues have to be considered. However, CRISPR/Cas9 has the potential to completely revolutionize human disease therapy, enabling the correction of disease-causing genetic aberrations. Interestingly, at the end of October of this year a Chinese group reported for the first time the application of CRISPR/Cas9 approach in humans. Immune cells from patient's blood were modified ex-vivo in order to disable PD-1 gene and then reinjected into the patient, who has metastatic non-small-cell lung cancer, part of a clinical trial the West China Hospital. The goal is to enhance immune-cells response against cancer. This

represents a real milestone towards the therapeutic application of CRISPR-Cas9 and

nowadays several trials are waiting for ethical approval.
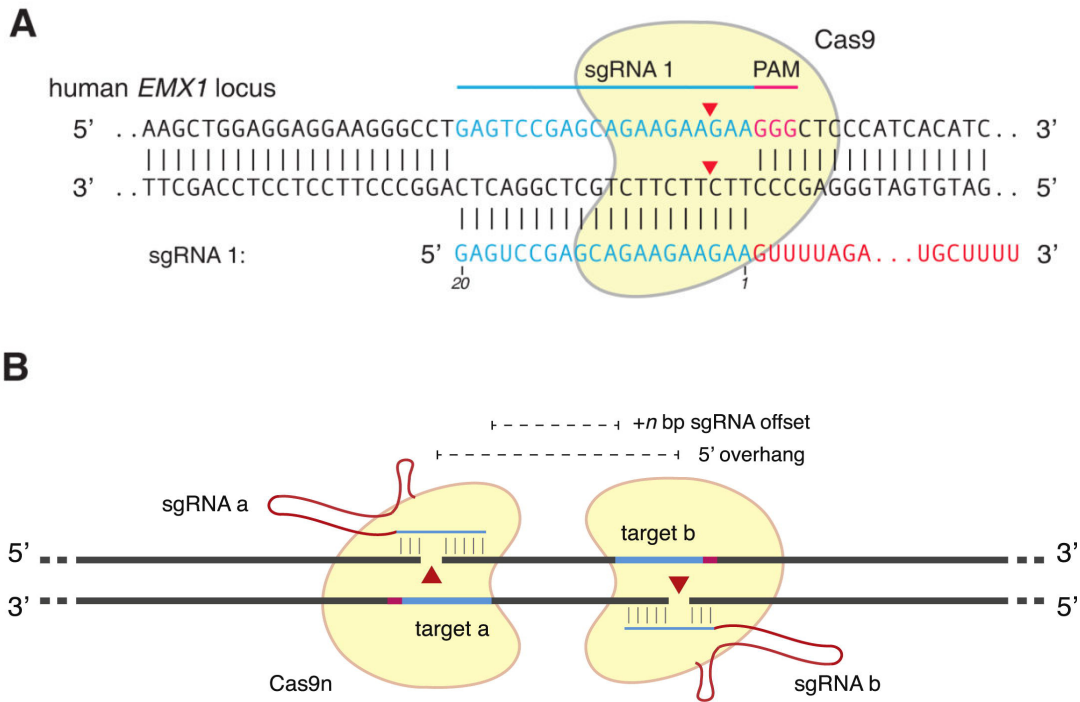
Figure 2: CRISPR/Cas9 approach.

A: Schematic representation of wt Cas9 targeted by 20-nt sgRNA to human EMX1 locus. DSB site and PAM

sequence are indicated. B: Schematic illustration of double nickase approach using a pair of Cas9 D10A

nickases (Cas9n), each one targeted by a specific sgRNA. The D10A mutation renders Cas9 able to nick the

strand complementary to the sgRNA. Adapted from Ran et al., Cell, 2013[192].

2. Aim of the thesis

My thesis was aimed to develop relevant genetic mESC models through CRISPR/Cas9 technology in order to dissect the role of *Ezh2*-affecting mutations in tumorigenesis. I focused my attention on 1) *Ezh2*-Y641N hyper-activating mutation, 2) *Ezh2*-Y731D/R690C inactivating mutations and, 3) H3.3K27M that inhibits PRC2 enzymatic activity. Overall, these cellular models combined with both ChIP/RNA-seq analyses and the broad differentiation potential of mESCs, will allow me to dissect the mechanisms by which these different mutations alter the transcriptional proprieties of the PRC2 complex to drive tumor development in distinct cellular contexts.

3. Materials and methods

3.1 Plasmids

3.1.1 Generation of mutant-EZH2 expression plasmids

The expression constructs for mutant-Ezh2 were generated by LR recombination of the EZH2 coding sequences from a pCR8 Gateway entry vector, already available in the laboratory, into a pCAG-Flag-Avi-ires-Puromycin Gateway compatible destination vector using LR recombinase (Invitrogen). Coding sequences for the different EZH2 mutants were obtained by mutagenesis PCR on the EZH2-containing pCR8 Gateway entry vector. Mutagenesis PCR was performed with specific primers (listed in Table 4) with Phusion® High-Fidelity DNA Polymerase (NEB) according to manufacturer's instructions and checked on 1% agarose gel. pCR8 Gateway entry Ezh2-containing vectors were obtained through TOP10 competent cells (Invitrogen) transformation and subsequent plating onto Luria Broth (LB) agar plates containing 100 µg/mL spectinomycin. LR recombination reaction was set up according to manufacture 's instructions and was used to transform TOP10 competent cells (Invitrogen). Transformed bacteria were then plated onto Luria Broth (LB) plates containing 100 µg/mL ampicillin until the growth of visible colonies. After proper culture, plasmid DNA was extracted with NucleoSpin Plasmid and NucleoBond Xtra Maxi kits fromm Macherey-Nagel. All the constructs were checked by sequencing at Nucleic Acid Unit of the European Institute of Oncology.

3.2. Cell culture and manipulation

3.2.1 Mouse embryonic stem cells: culturing and manipulation

Mouse embryonic stem (ES) cells belonging to ES-E14TG2a cell line, provided by Cogentech transgenic facility, were used. These cells, commonly called E14, are derivative of one of several embryonic stem cell (ES) lines developed by Smith and Hooper[20]. All the genome-edited cell lines used in this project were derived from ES-E14TG2a cells through CRISPR/Cas9 technology. All mESC lines were grown at 37°C, in a CO_2 incubator (5% CO2) with standard oxygen tension (21% oxygen), on 0.1% gelatinized tissue culture dishes in GMEM supplemented with 20% fetal calf serum (Euroclone), 2 mM glutamine (Gibco), 100 U/ml penicillin and 0.1 mg/ml streptomycin (Gibco), 0.1 mM non-essential aminoacids (Gibco), 1 mM Na-Pyruvate (Gibco), 50 μM ß-mercaptoethanol-phosphate-buffered saline (PBS; Gibco) and Leukemia Inhibitory Factor (produced in house). Moreover, mESC growth medium was completed with 2 selective GSK3β & Mek 1/2 inhibitors purchased from ABCR GmbH. Mek 1/2 inhibitor and GSK3β inhibitor were used at 1 μM and 3 μM final concentrations respectively, in order to enhance cell viability and the maintenance of pluripotency. Where indicated, cells were treated with 10 μM MG132 or DMSO as vehicle.

3.2.2 Embryoid bodies formation assay

In order to study mESC capability to differentiate an embryoid bodies formation assay was performed. Indeed, this method allows the formation of cell aggregates differentiating towards the three germ layers. Undifferentiated mESC were induced to differentiate into embryoid bodies (EBs) by LIF removal in hanging drops containing

1,000 cell/20 μl drop on the lid of 15 cm petri dish for 48 h. EBs were then collected from the drops and stimulated between day 2 and 5 with 0.5 μM all-trans-retinoic acid (ATRA). EBs were left in culture in non-coated petri dishes usually to day 9 in ES medium without LIF. Medium was replaced every second day.

3.2.3 Transfections

Transfections were all performed with lipofectamine 2000 reagent (Invitrogen) according to manufacturer's instructions. Lipofectamine 2000 is a cationic liposome-based reagent that allows the formation of liposomal particles containing the negative charged nucleic acid molecules that have to be transfected. In this way, the DNA-containing liposomes, positively charged on their surface, can fuse with cell plasma-membrane thus allowing the entering of DNA into the cell. In particular, a reverse transfection approach, was used to transfect mESC in 10 cm dishes. Briefly, two 1 ml reactions were prepared by mixing the proper amount of plasmidic DNA (10-20 μg for 10 cm dish) and 40 μl lipofectamine 2000 reagent, respectively, with Opti-MEM medium (Gibco). After 5 minutes incubation, the reactions were mixed, thus allowing the formation of the liposomes, and incubate for 30 minutes. Then, the mixed reaction was added to previously gelatinized 10 cm plates and 5-7 x 106 mESC were plated. Medium was replaced after 6-8 hours to avoid cell toxicity. Depending on the experiment, transfected cells were then subjected to selection (puromycin 2 ug/ml) or FACS-sorted.

3.2.4 CRISPR/Cas9

CRISPR/Cas9 procedure was based on published protocols from Zhang's lab[188,192].

<u>Reagents</u>

pSpCas9(BB)-2a-Puro (PX459)V2.0 and pSpCas9n(BB)-2A-GFP(PX461) plasmids were obtained from Addgene. Once selected the precise genomic region to target, suitable sgRNAs were designed using Target Finders Zhangs lab, DNA 2.0 and Benchling.com tools. All the sgRNAs used in this study are listed in Table 1. 20 nucleotide-long RNA guides were ordered as common destalted oligos from Life Technologies.

If not already present, a G was added at 5' position in order to ensure efficient U6 transcription of sgRNA. Moreover, specific sequences were added (CACCG to the 5' of forward oligo and CAAA to the 5' of the reverse oligo) to clone these sequences into BbsI-opened Cas9/Cas9n-plasmid. After proper phosphorylation and annealing, oligos were diluted and a ligation reaction with BbsI opened and dephosphorylated Cas9/Cas9n-expressing plasmid was set up. Ligation was performed following Quick ligation kit protocol (NEB). Ligation mix was used to transform TOP10 competent cells that were subsequently plated onto ampicillin LB plates. After bacteria cultures, plasmid DNA was extracted by Macherey-Nagel extraction kit. Correct integration of the guides was checked by Sanger sequencing using pGEX primer. For HDR-based experiments, 200 bp ssODN template, designed as antisense, was purchased from IDT. For the design we referred to the instructions published by Ran and collaborators[188,192].

<u>Transfections</u>

Reverse transfections were performed in 10 cm plate with 5-7 x 106 mESC using lipofectamine 2000 reagent as described in 2.3 section. For each reaction 16 ug of total

plasmidic DNA were used (8 μg plasmid a + 8 μg plasmid 2 for double guide approach).
For HDR 3-6 μl of 100 μM ssODN were co-transfected with sgRNA,Cas9/Cas9n-expressing plasmids. After transfections cells were FACS-sorted or subjected to puromycin selection (2 μg/ml) for 24 hours. Then cells were counted and plated in a number of 300-600 for 15 cm dish (dilution cloning). Medium was replaced every 2 days till the appearance of visible colonies (ca. 10 days). At this point, clones were picked and expanded for subsequent analysis

<u>Screening</u>

Screening procedure for KO was based on WB analysis and/or PCR. Genomic DNA was extracted with DNeasy Blood & Tissue Kit (Qiagen) and PCR was performed with Phusion® High-Fidelity DNA Polymerase (NEB) according to manufacturer's instructions. PCR Primers are listed in Table 1. Where HDR had to be assessed, PCR products were subjected to restriction digestion. EcoRI, XmnI, BamHI and ScaI enzyme were purchased from NEB. The presence of the desired mutations was then confirmed by Sanger sequencing.

3.3 Techniques used for protein detection and protein-protein interactions assessment

3.3.1 Immunoblot analysis

This method, commonly known as Western blot analysis (WB), allows the detection of proteins of interest and relative post translational modifications in a protein extract.

This approach implies a first protein mass-based separation step of the extracts by

denaturing gel electrophoresis. Then, separated proteins are transferred to a nitrocellulose membrane, which, after being blocked, is incubated with a solution containing a specific antibody against the protein of interest. After that, a secondary antibody, conjugated to a detection system, recognizes the first one, therefore detecting also the protein of interest. Moving to practice, after medium removal, cells were washed twice with PBS (phosphate buffered saline: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, KH_2PO_4 1.8 mM, pH 7.4) and 0.25% trypsin, 0.53 mM EDTA solution (Lonza) was added to the plates in order to detach cells. After few minutes of incubation at 37%, trypsin action was blocked by addition of growth medium and cells were recovered by mechanical pipetting. Cells were centrifuged for 5 min at 1200 rpm at 4°C and after two washes with PBS, cellular pellets were obtained. Cellular pellets can be directly lysed or frozen at -80 for subsequent analysis. Pelleted cells were lysed with high salt lysis buffer (20mM Tris-HCl, pH 7.6, 300mM NaCl, 10% glycerol, 0.2% (v/v) Igepal). After 30 minutes incubation on ice, extracts were sonicated for 10 cycles (30" on- 30" off) with Diagenode Bioruptor® sonicator. Clarified lysates were then centrifuged at 13000 x g for 30' at 4°C. The recovered supernatant was quantified by Bradford assay with Bio-Rad Protein Assay reagent (Bio- Rad, cat. 500-0006). Then, Laemmli sample buffer was added and samples boiled 10 minutes at 95 °C. Usually 20-40 μg of total protein extracts were loaded onto each lane of a acrylamide, bisacrylamide gel, and a sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed. Gel-separated proteins were transferred to a Amersham Protran Nitrocellulose Membrane (GE Healthcare Life Sciences), one hour and 20 minutes at 4 °C, at 100 volts. Membranes were blocked with a solution of in TRIS-buffered saline (TBS: 20mM TRIS/HCl, pH 7.4, 137 mM NaCl, 2.7 mM KCl) plus 0.1% Tween (TBS-T) containing 5% non-fat dried milk. The same milk/TBS-T solution was prepared to dilute primary antibodies, which were incubated for usually 1-

2 hours at room temperature or ON at 4 °C. After three washes with TBS-T, a secondary HRP (horseradish peroxidase)-conjugated antibody (BioRad) was diluted in the same solution and incubated for one hour at room temperature. Following three further washes in TBS-T, the bound secondary antibody was revealed by ECL method (Biorad) using ChimiDoc XRS+.

3.3.2 Cellular fractionation

Cytoplasmic and nuclear cell fractions were isolated re-suspending fresh whole cell pellets in Nuclear Prep Buffer (10 mM Tris HCl pH 8, 100 mM NaCl, 2 mM MgCl2, 0.3 M Sucrose, 0.2% Igepal) followed by centrifugation. Then, Laemmli sample buffer was added to supernatants, the cytoplasmic extracts. Nuclear pellets were washed once with nuclear prep buffer, lysate in high salt lysis buffer (20mM Tris-HCl, pH 7.6, 300mM NaCl, 10% glycerol, 0.2% (v/v) Igepal), resuspendeded in Laemmli sample buffer, briefly sonicated and boiled 10 minutes at 95 °C. All the fractions were then analyzed by Western blot.

3.3.3 Immunoprecipitation

Total Protein extracts were obtained from fresh cell pellets after lysis with high salt lysis buffer (20mM Tris-HCl, pH 7.6, 300mM NaCl, 10% glycerol, 0.2% (v/v) Igepal). After incubation for 30 minutes on ice, lysates were centrifuged at 13000 x g for 30' at 4°C and the superantant was recovered and quantified with Bradford assay using Biorad reagent (Bio- Rad, cat.30500-0006). Immunoprecipitations were performed by incubation of protein extracts (1 mg) for 3-4 hours with home-made Ezh2 Ab (AC22) - crosslinked-PA

sepharose beads (30 ul slurry for each IP, Healthcare, Cat. 170780-01) at 4°C on a rotating platform. After 3-4 washes with lysis buffer the immunoprecipitated proteins were eluted by Laemmli sample buffer addition to the beads and analyzed by WB.

3.4 Assays for detection of DNA modifications and protein binding to DNA

3.4.1 Chromatin Immunoprecipitation (ChIP)

ChIP (chromatin immunoprecipitation) assay is a powerful technique used to investigate epigenetic modifications and protein-DNA interactions. Briefly, DNA and DNA-associated proteins are cross-linked with formaldehyde (or UV rays or other chemical agents i.e. DSG). in order "to fix" these interactions. Cross-linked chromatin is then sheared by sonication to generate fragments of 300 - 1000 base pairs (bp) in length, depending of the analysis required. Through immunoprecipitation, proteins of interest coupled to DNA are isolated by means of specific antibodies. Indeed, chemical cross-linking is reversible, thus DNA can be separated from associated proteins and analyzed, both by high throughput sequencing and Real Time quantitative PCR. As previously described[46], 1% formaldehyde cross-linked chromatin was resuspended in IP buffer (70 mM TRIS/HCl pH 8.0, 5 mM EDTA, 100 mM NaCl, 0.3 % sodium dodecyl sulfate or SDS, 1.7% TRITON X-100), fragmented by sonication using a Digital sonifier 450 (Branson) to an average size of 200–500 bp to perform HPTM ChIP and of 500-1000 bp to make proteins ChIP and immunoprecipitated ON with 2-5 µg of the indicated antibodies. 0.5-1 mg of chromatin was used for each protein precipitation; 200-500 ug were used for histone PTM. Then protein A sepharose beads (GE Healthcare, cat. 170780-01; 30 µl slurry per IP) were

added and incubated 2-3 h at 4°C, followed by three washes with "low salt" wash buffer (20 mM TRIS/HCl pH=8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% TRITON X-100), and one in "high salt" wash buffer (20 mM TRIS/HCl pH=8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% TRITON X-100). De-crosslinking with proper de-crosslinking solution (0.1 M NaHCO3, 1% SDS) at 65°C ON was performed. Eluted DNA was finally recovered and purified using QIAquick PCR purification kit (Qiagen).

3.4.2 Chromatin with reference exogenous genome (ChIP-Rx)

For H3K27me3 ChIP-seq analysis this new correction method has been applied. (Orlando et al. 2014). This approach implies the use/addition of a constant amount of reference or "spike-in" epigenome within IP samples thus allowing to perform genome-wide quantitative comparisons of histone modification status across different sample. It has been shown that ChIP-Rx enables the discovery and quantification of dynamic epigenomic profiles across mammalian cells that would otherwise remain hidden using traditional normalization methods[223]. Moving to practice, sonicated cross-linked chromatin to be used for immunoprecipitation was mixed with 5% of drosophila S2 cells chromatin (already crosslinked and sonicated). Then, ChIP was performed following the protocol described in section 2.4.1.

3.4.3 High throughput ChIP sequencing (ChIPseq)

The DNA retrieved from ChIP experiments were used for ChIPseq libraries preparation with the Illumina ChIPSeq Sample Prep kit (IP-102-1001) and multiplexing oligonucleotide kit (PE-400-1001) by Campus IFOM-IEO internal genomic facility. DNA

68

libraries were quantified using a high sensitivity DNA Chip on Bioanalyzer instrument (Agilent) and used for cluster generation and sequencing using the HiSeq 2000 platform (Illumina) following the protocol of the manufacturer.

3.4.4 Quantitative Real Time PCR (RT-qPCR)

Purified DNA coming from ChIP experiments was analyzed by qPCR carried out using GoTaq qPCR Master Mix (Promega) on CFX96 Real-Time PCR Detection System (Biorad). Primers used for qPCR analysis are listed in Table 2.

3.4.5 ChIP-sequencing data analysis

Each ChIP-seq data with spike-in was aligned to mouse (mm9) and drosophila (dm6) reference genome using Bowtie (PMID 19261174) separately. Alignment was executed favoring only unique alignments, and duplicates were removed for downstream analysis. Peak calling was for all samples was performed with macs2 (PMID 18798982). Broader peaks were generated by enabling the --broad option.

We derived a normalization factor for individual dataset as described in publication (PMID 25437568), where α, for each sample, is such that the resulting drosophila signal was equilibrated across all samples. The mathematical derivation of the normalization factor α is as follows:

Let:

• α = normalization factor

• β = reference signal from the reference sample (drosophila)

• Nd = total number of reads from a sample aligning to the reference genome

• r = percentage of the sample comprised of reference sample

as the reference signal should always be the same, it can be written as

$$\beta = \alpha \frac{Nd}{r}$$

And, since β is always the same, it can be arbitrarily set it to any value, and for convenience it was set to 1.

$$1 = \alpha \frac{Nd}{r}$$

Then above equation can be reformed as

$$\alpha = \frac{r}{Nd}$$

Since r is the same for all experiments in this work, it can be further simplified to

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

Therefore, the normalization constant used is 1 over the number of reads mapping to drosophila per million. This is applied to all samples. For profiling, we considered center of target region and extended to defined length both up and downstream. Extended region was further broken down into smaller bins of 50 bp in size. Normalized reads with or without normalization factor for their respective sample within each bin were computed and represented as average distribution profiles or as heatmaps.

3.5. Methods for RNA analysis

3.5.1 Real Time quantitative PCR

The real-time quantitative polymerase chain reaction (or quantitative polymerase chain reaction, qPCR), allows the detection and relative quantification of a specific DNA sequence in a sample. It implies the use of an unspecific fluorescent dye (i.e. SYBR green) that intercalates with double-strand DNA that is amplified through a PCR reaction. The specificity of the amplification step relies on the use of gene-specific oligonucleotide probes (primers). During amplification, the fluorescent dye is incorporated in the nascent DNA thus allowing the detection of amplified DNA. For a short period of the reaction, DNA amplification is exponential, therefore it can be described by a mathematical function, allowing DNA quantification. This technique can be used both to detect the amount of a DNA sequence within a sample (such as target genes in a ChIP experiment), or the abundance of a cDNA derived from an RNA sample. Total RNA was extracted from cellular pellets with the Quick-RNA™ MiniPrep extraction kit (Zymo Research) and retro-transcribed with ImProm-II™ Reverse Transcription System (Promega) according to the manufacturer's instructions. Real-time quantitative PCR (qPCR) was carried out using GoTaq qPCR Master Mix (Promega) on CFX96 Real-Time PCR Detection System (Biorad). Expression primes used for qPCR analysis are listed in Table 3.

3.6. Antibodies

3.6.1 Antibodies used for Immunoblot and immunoprecipitation analyses

For Western blots the following antibodies were used: mouse anti-Vinculin (Sigma-Aldrich, V9131), mouse anti-Oct3/4 (Santa Cruz, sc-5279), mouse anti-Nanog (Novus Biologicals, Cat. 100-587A, rabbit anti-Ezh2 (Cell signaling, sc-25383), goat anti-Suz12 (Santa Cruz, sc-46264), homemade mouse anti-Ezh2 BD43[87], homemade mouse anti-Eed[87], rabbit anti-beta-Tubulin (Santa Cruz, sc-9104), mouse anti-lamin A/C (Santa Cruz, sc-7292), homemade mouse anti-p53 (Amati's group) and rabbit anti-flag (Sigma, F7425). For immunoprecipitation experiments, home-made AC22-crosslinked PA sepharose and HA-crosslinked PA sepharose beads were used. For histone PTMs detection the following antibodies were used: rabbit anti-H3K27me3 (Cell signaling, 9733), rabbit anti-H3K27me2 (Cell Signaling, 9728), mouse anti-H3K27me1 (Active Motif, 61015), rabbit anti-H3K27ac (Active Motif, 39133), rabbit anti-H2AK119ub (Cell Signaling, 8240S), rabbit anti-H3 (Abcam, 1791), rabbit anti-H2A (Cell Signaling, 12349).

3.6.2 Antibodies used for ChIP analyses

In ChIP experiments, the following antibodies were used: rabbit anti-Suz12 (Cell signaling, 3737), home-made rabbit anti-Ring1b, rabbit anti-H3K27me3 (Cell signaling, 9733), rabbit anti- H3 (Abcam, 1791) and rabbit anti-H2AK119ub (Cell Signaling, 8240S). Rabbit IgG (Sigma, I5006) was used as negative control.

3.7 Primers and oligos

CRISPR/Cas9 oligos and ssODNs		
Application	**Oligo name**	**Sequence (5' – 3')**
Ezh2 Y641N	sgRNA1 forward	CACCGCATCTCAGAATACTGTGGGG
Ezh2 Y641N	sgRNA1 reverse	AAACCCCCACAGTATTCTGAGATGC
Ezh2 Y641N	sgRNA2 forward	CACCGATGAATTCATTTTTCTGTAC
Ezh2 Y641N	sgRNA2 reverse	AAACGTACAGAAAAATGAATTCATC
Ezh2 Y641N	Y641N ssODN	GGCCTTGTCCAGTGAAATCTATCCAAAACAATGCAAGCTGCTAC TATAAACAAGTCAACTGGAATGCACGAGTATGTCTTACTTCCCC ACAGTTTTCTGAGATGAACTCATTCTTCTGCACAGGATCTTTGA TAAAGATGCCCCAGCCTGCCACATCAGACGGTGCCAGCAGTAAG TGCTAGAGAGTAAGCAGTCACATT
Ezh1 KO	sgRNA3 forward	CACCGACTTCCCGCTGCATTCCATG
Ezh1 KO	sgRNA3 reverse	AAACCATGGAATGCAGCGGGAAGTC
Ezh1 KO	sgRNA4 forward	CACCGTATGTGGCAAATTTTGCAA
Ezh1 KO	sgRNA4 reverse	AAACTTGCAAAATTTGCCACATAC
Ezh2 KO	sgRNA5 forward	CACCGACACGCTTCCGCCAACAAAC
Ezh2 KO	sgRNA5 reverse	AAACGTTTGTTGGCGGAAGCGTGTC
Ezh2 KO	sgRNA6 forward	CACCGACTTCTGTGAGCTCATTGCG
Ezh2 KO	sgRNA6 reverse	AAACCGCAATGAGCTCACAGAAGTC
Ezh2 Y726D	sgRNA7 forward	CACCGCAGGTTGGTAAAATACACAA
Ezh2 Y726D	sgRNA7 reverse	AAACTTGTGTATTTTACCAACCTGC
Ezh2 Y726D	Y726D ssODN	TAGTGACTGGTCAGTTAAAAGAATGCACCCTCCAATGATGGCAG ACCTGTCAAAAAACTTACAAACAGCCTTGGATCCAAGCCCCATAG TTTCAGAAGGGAAAACTTTTGTGTATTTTACCAACCTGTCATCAA AAAACAACTCTTCACCAGTCTGGATAGCCCTCTTAGCAAAGATG CCTATCCTGTGGTCACCATTAA
Ezh2 R685C	sgRNA8 forward	CACCGTGTGGTGGATGCAACCCGAA
Ezh2 R685C	sgRNA8 reverse	AAACTTCGGGTTGCATCCACCACAC
Ezh2 R685C	R685C ssODN	AAATGGAACTGTTGCATCAAACTGTCAATTCCAACTTAAAAGCT TACCTTTTGCATAGCAGTTTGGATTTACTGAATGATTAGCAAAA CAAATCTTATTGCCTTTTCGGGTTGCATCCACCACAGAATTCAA AGTGAAAAACATAGATAATCCAGTGACTTATTTTCAGTCATAGA CCAAGGTTATTATGGCTTTAGAAA
H3.3 K27M	sgRNA9 forward	CACCGATTTCTAAAACGTCGAGCAG
H3.3 K27M	sgRNA9 reverse	AAACCTGCTCGACGTTTTAGAAATC
H3.3 K27M	K27M ssODN	CCCATAGCAAAAAGTATGACTATCTTCCTGTTAAATTACACGGA CATACTGCTTAGACACTATCTTGCTGCTCGACGTTTTAGAAATA CCTGTAACGATGAGGTTTCTTCACCCCTCCAGTACTAGGCGCAC TCATGCGAGCGGCTTTTGTAGCCAGTTGTTTCCTGGGTGCTTTA CCACCGGTGGATTTGCGGGCAGTC

Table 1: List of sgRNAs and ssODNs used in CRISPR/Cas9 experiments.

ChIP-qPCR primers	
Oligo name	**Sequence (5' – 3')**
Wnt5a forward	CTCTGAGTTGAGTCGCCACC
Wnt5a reverse	TTCTTCCTTTCTTCGGGTTAACC
HoxA9 forward	TTGATGTTGACTGGCGATTTTC
HoxA9 reverse	ATCTGTATGCCTAGTCCCGCTCC
Hoxd9 forward	GGATAATCGCCTAGGTGTGACTTAG
Hoxd9 reverse	CATCTCTTCTTGCCTCTCTGGG
Wt1 forward	GTCGGAGCCCATTTGCTG
Wt1 reverse	CAGTGAGACGAGGCTCCCAC
Zic2 forward	TACAAACTGGTCAACCACATCC
Zic2 reverse	TTGTGGATCTTGAGGTTCTCG
HoxA11 forward	TTCTTGTCCCCGGGGTAGTC
Hoxa11 reverse	GACCAGTTTTTCGAGACGGC
Foxb1 forward	AGAGAGCTGCCCATGGTAGT
Foxb1 reverse	GAGTACAAGATGCCTGGGGG
Utp6 forward	AGCTAGGCAGCAGTCACCAT
Utp6 reverse	CAGTTGCGCAATAGTGTCGT
c-Myc forward	GGAGTGGTTCAGGATTGGGG
c-Myc reverse	AAGTTCACGTTGAGGGGCAT

Table 2: List of primers used for ChIP-qPCR analyses.

RT-qPCR primers	
Oligo name	**Sequence (5' – 3')**
Oct4 forward	CGAGAACAATGAGAACCTTC
Oct4 reverse	CCTTCTCTAGCCCAAGCTGAT
Nanog forward	CCTCCAGCAGATGCAAGAACTC
Nanog reverse	CTTCAACCACTGGTTTTTCTGCC
Pax3 forward	TCCCATGGTTGCGTCTCTAAG
Pax3 reverse	CTCCACGTCAGGCGTTGTC
Mausashi forward	CCATGCTGATGTTCGACAAAAC
Mausashi reverse	TCAAACGTGACAAATCCAAACC
Nestin forward	GCTCCCTATCCTAAAAATGCAGAG
Nestin reverse	GTAGAACTGGGCACTGTGGCC
Gapdh forward	CATCTTCTTGTGCAGTGCCAG
Gapdh reverse	GGCAACAATCTCCACTTTGCC
Actin forward	TACAATGAGCTGCGTGTGGC
Actin reverse	GTACATGGCTGGGGTGTTGA

Table 3: List or primers used for RT-qPCR analyses.

Mutagenesis PCR primers	
Oligo name	**Sequence (5' – 3')**
C576W forward	CCAGGTAGCACGGCCACTGCTTGGTGT
C576W reverse	ACACCAAGCAGTGGCCGTGCTACCTGG
C593Y forward	CAGCGGCTCCATAAGTAAGACAGAGGTCAGGGT
C593Y reverse	ACCCTGACCTCTGTCTTACTTATGGAGCCGCTG
R690A forward	CCGAATGATTTGCAAAAGCAATTTTGTTACCCTTGCGGGTTGC
R690A reverse	GCAACCCGCAAGGGTAACAAAATTGCTTTTGCAAATCATTCGG
R690H forward	TTTACCGAATGATTTGCAAAATGAATTTTGTTACCCTTGCGGG
R690H reverse	CCCGCAAGGGTAACAAAATTCATTTTGCAAATCATTCGGTAAA
R690C forward	CCGAATGATTTGCAAAACAAATTTTGTTACCCTTGCGGGT
R690C reverse	ACCCGCAAGGGTAACAAAATTTGTTTTGCAAATCATTCGG
H694A forward	GTTTGGATTTACCGAAGCATTTGCAAAACGAATTTTGTTACCCTTGCG
H694A reverse	CGCAAGGGTAACAAAATTCGTTTTGCAAATGCTTCGGTAAATCCAAAC
G630S forward	TCTGACGTGGCAGGCTGGAGCATTTTTATCAAAGATCCTG
G630S reverse	CAGGATCTTTGATAAAAATGCTCCAGCCTGCCACGTCAGA
H694K forward	GCAAGGGTAACAAAATTCGTTTTGCAAATAATTCGGTAAATCCAAA
H694K reverse	TTTGGATTTACCGAATTATTTGCAAAACGAATTTTGTTACCCTTGC
Y646C forward	AGAAAAATGAATTCATCTCAGAATGCTGTGGAGAGATTATTTCTCAAG
Y646C reverse	CTTGAGAAATAATCTCTCCACAGCATTCTGAGATGAATTCATTTTTCT
Y731D forward	GCCTGGCTGTATCTGTCATCAAAAAACAGCTCTTCG
Y731D reverse	CGAAGAGCTGTTTTTTGATGACAGATACAGCCAGGC
Y731F forward	GCCTGGCTGTATCTGAAATCAAAAAACAGCTCTTCGC
Y731F reverse	GCGAAGAGCTGTTTTTTGATTTCAGATACAGCCAGGC
R732K forward	ATCAGCCTGGCTGTATTTGTAATCAAAAAACAGCTCTTCGC
R732K reverse	GCGAAGAGCTGTTTTTTGATTACAAATACAGCCAGGCTGAT

Table 4: list of primers used for EZH2 mutagenesis PCRs.

4. Results

4.1 Generation of EZH2 Y641N mESCs through CRISPR/Cas9 system

4.1.1 Strategy used to obtain EZH2 Y641N expressing mESCs

In order to unravel the molecular mechanisms underlying both hyper-activating and inactivating mutations of *EZH2*, I took advantage of the CRISPR/Cas9 approach to develop suitable mESC models. The first goal was to obtain mESC models to investigate Y641N aminoacidic substitution affecting the EZH2 SET domain both in heterozygosity and in homozygosity. EZH2 Y641N heterozygous aminoacidic substitution has been found in a high percentage of DLBC lymphomas and follicular lymphomas, where it has been shown to generate a hyperactive form of EZH2 with enhanced H3K27me3 activity. A physiological model of this mutation is indeed necessary for the elucidation of the molecular mechanisms underlying the accumulation of H3K27me3 observed in disease. EZH2 Y641N homozygous substitution instead has not been found in patients yet, but is expected to inactivate the enzymatic activity of EZH2. This other model would allow me to study the different biochemical aspects of PRC2 activity in presence of an inactive catalytic subunit. I used the CRISPR/Cas9 system to physiologically introduce the Y641N aminoacidic substitution by properly mutating *Ezh2* gene in mESCs. EZH2 aminoacidic sequence is highly conserved between human and mouse. Indeed, aminoacidic sequence alignment between the two species revealed that the murine residue where the mutation occurs is the Y641, taking into consideration the longest EZH2 protein isoform (746 aa). A double nickase approach was used to enhance the specificity and to reduce the

potential off-target effects of the CRISPR/Cas9 system[192]. This approach relies on the activity of a mutated form of the Cas9 enzyme, Cas9n. This enzyme contains a mutation on one of the two catalytic domains, RuvC, and compared to the wild type (wt) Cas9 it is not able to induce a DSB but only to nick one strand of the DNA. Only when two nicks occur near to each other a DSB is generated. Two sgRNAs, annealing to the opposites DNA strands adjacent to the target site (Y641), have been designed using Feng Zhang lab's Target Finder and CRISPR-DNA2.0 tools. In particular, the two designed guides displayed a -2 bp offset generating 5' overhangs upon nicking (Fig. 2B). To achieve a specific HDR, a proper single stranded template was designed according to published protocol instructions[192]. For the purpose of introducing small modifications, such as point mutations, a single stranded template is recommended. Indeed, I used a 200 bp long single stranded template (ssODN) with 82/81 bp homology arms to ensure homologous recombination. This was properly designed to assure the mutation of some critical residues compared to the wt sequence. In addition to the desired substitution (Y641N) that was obtained by the TAC/CAC mutation, other residues were silently mutated: the CCT PAM was mutated in CTT; the AGG PAM couldn't be silently mutated so some other adjacent residues were mutated: TTT/CTT and TTC/CTC. PAMs have to be disrupted in order to avoid a re-targeting by Cas9 enzyme after the recombination event. Another silent mutation was introduced to disrupt an EcoRI restriction site (GTA/GTT) while an XmnI restriction site was also introduced (GAANNNNTTC) for screening proposes (Fig. 3A-B). The absence/presence of restriction sites is very useful for the subsequent clone screening. mESC were co-transfected with px461 plasmids expressing Cas9n carrying two single guides RNA together with the ssODN as template. Transfection was performed according to manufacture's instructions and growth medium was replaced the day after. After 24 hours cells were subjected to FACS analysis. Px461

78

plasmid expresses also GFP thus allowing the selection of transfected cells. GFP-positive cells were counted and diluted into 300-600 cells for 15 cm dish. The medium was replaced every second day until the clones started to emerge and were ready to be picked (ca. 10 days). Single clones were expanded and collected for subsequent screening. Clones screening was based on a specific PCR amplification followed by digestion with EcoRI. Wild type 907 bp-long PCR product was cut into two fragments of 390 bp and 517 bp, respectively, by EcoRI whereas, the restriction site was lost when precise HDR occurs. However, the loss of EcoRI site may also be caused by a deletion. To exclude this possibility, clones positive for the first "negative" screening step were then subjected to XmnI digestion to confirm the acquisition of this new restriction site. If correct HDR was achieved, the PCR product could be cut by XmnI enzyme. PCR products were subsequently Sanger-sequenced in order to confirm the recombination and/or deletion events. Among almost 100 clones screened I was not able to identify any heterozygous clone for the desired aminoacidic substitution. Notably, I found 2 clones harboring the Y641N on both *Ezh2* alleles, from now referred as Y641N#1 and Y641N#2 (Y641N#1 sequence is shown in Fig. 3B). This is not surprising since it is well known that many factors including locus-specificity and cell-type can influence the HDR efficiency. The CRISPR-based editing was efficient since I obtained many heterozygous clones for the mutation but displaying indels on the other allele. Additionally, clones with small deletions nearby Y641 residue were also identified during the screening procedure and two of them were used for subsequent analysis. These ΔSET clones are referred as ΔEzh2#1 and ΔEzh2#2.

A

Wild type Ezh2 sequence
...
GGCCTTGTCCAGTGAAATCTATCCAAAACAATGCAAGCTGCTACTATAAACAAGTCAACTGGAATGCACGAG
TATGTCTTACCTCCCCACAGTATTCTGAGATGAATTCATTTTTCTGTACAGGATCTTTGATAAAGATGCCCC
AGCCTGCCACATCAGACGGTGCCAGCAGTAAGTGCTAGAGAGTAAGCAGTCACATT...

200 bp ssODN

GGCCTTGTCCAGTGAAATCTATCCAAAACAATGCAAGCTGCTACTATAAACAAGTCAACTGGAATGCACGAG
TATGTCTTACTTCCCCACAGTTTTCTGAGATGAACTCATTCTTCTGCACAGGATCTTTGATAAAGATGCCCC
AGCCTGCCACATCAGACGGTGCCAGCAGTAAGTGCTAGAGAGTAAGCAGTCACATT

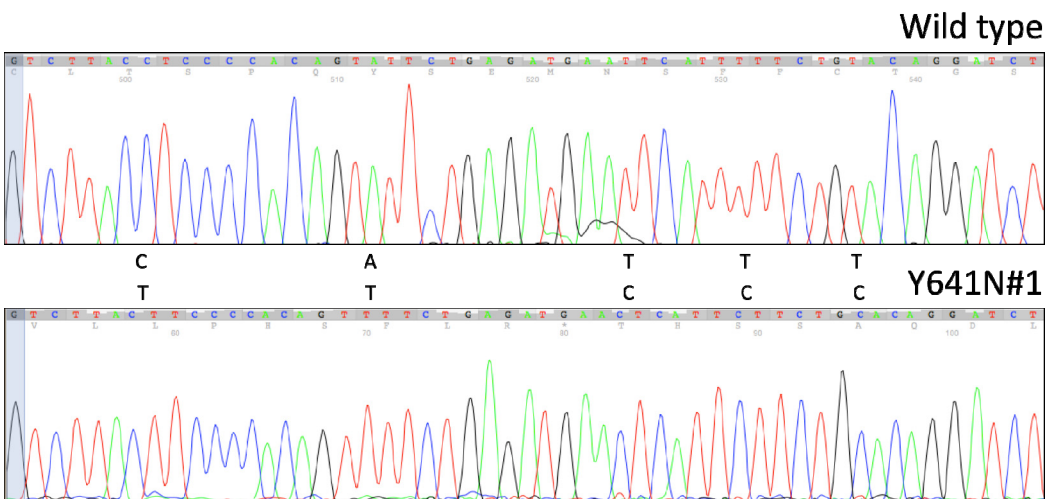|		
PAM sequence	EcoRI site	**N** mutated base
Y641/N641 residue	XmnI site	NNNNNNNNNNNNNNNNNNNNN sgRNA

B

Figure 3: Generation of EZH2 Y641N mESC through CRISPR/Cas9 system.

A: Schematic representation of sgRNAs and ssODN design. In the upper panel, partial wild type *Ezh2* sequence is reported with sgRNAs indicated as underlined. Y641 residue and PAMs sequences are highlightened (light blue and orange, respectively). EcoRI and XmnI restriction sites are highlightened in green and violet, respectively. In ssODN sequence (bottom panel), mutated bases are indicated in bold red.

B: Sanger sequencing analysis on specific PCR products obtained from amplification of genomic (gDNA) extracted from wild type or Y641N#1 cells. Mutations introduced by HDR-based CRISPR/Cas9 are indicated.

Despite the previously described strategy didn't allowed me to obtain cell lines heterozygous for Y641N to model the hyper-activating form of EZH2, the two obtained homozygous Y641N clones could be used to model catalytically inactivating point mutations of EZH2.

4.1.2 Characterization of EZH2 Y641N mESC clones

To assess the inactivating potential of homozygous Y641N aminoacidic substitution, total protein lysates from N-mutant clones (Y641N#1 and Y641N#2), ΔSET clones (ΔEzh2#1 and ΔEzh2#2) as well as wild type cells were subjected to Western Blot analysis. In both ΔSET-EZH2 clones H3K27me3 and H3K27me2 deposition was severely impaired compared to wt cells, with just a little residue of H3K27me2 still detectable. H3K27me1 levels were not affected by the deletion within EZH2 SET domain. The deletion was confirmed by a lower band corresponding to EZH2 protein (Fig. 4A). Both N-mutant clones resembled the effects observed with ΔSET clones considering H3K27me1 (unaffected) and H3K27me2 (almost abolished), however, some residual H3K27me3 was still present. Moreover, N-EZH2 displayed lower expression levels, comparable to those of ΔSET-EZH2, respect to wt. The SUZ12 levels appeared also slightly reduced in both kinds of modified clones thus suggesting some overall effects of the mutant EZH2 over the PRC2-subunits' expression. Finally, ES pluripotency seemed not to be affected neither by the deletion in the SET domain nor by the Y641N aminoacidic substitution as confirmed by Oct3/4 marker expression whose levels resulted indeed unaffected (Fig. 4A). PRC2 has been demonstrated to perform all the three distinct methylation states of the H3K27 residue[63]. Among them, it remains unclear whether H3K27me1 is rather dependent on EZH2 or EZH1 activity. Indeed, in my models ΔSET deletion or Y641N

aminoacidic substitution affecting EZH2 did not result in a loss of H3K27me1 deposition suggesting that EZH1 is the major responsible for this activity. Moreover, also the residual levels of H3K27me2 may be due to an EZH1-dependent activity of PRC2. Finally, the residual H3K27me3 levels observed in Y641N-clones suggest either a potential incomplete loss-of-function effect exerted by EZH2 Y641N and/or potential partial compensations by EZH1.
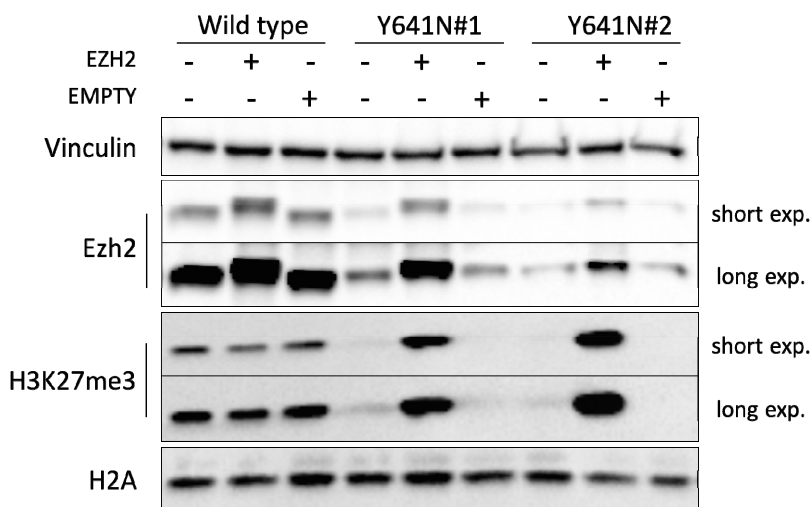
A
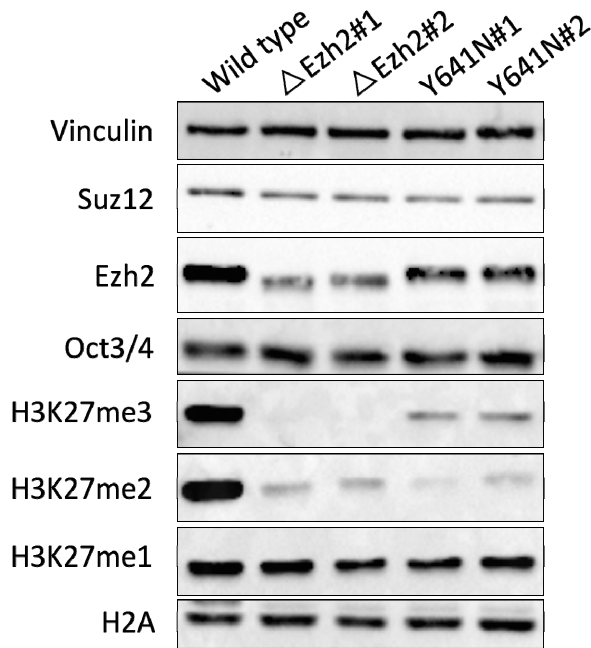
B

Figure 4: Y641N aminoacidic substitution affects H3K27me2 and H3K27me3 PRC2 activity.

A: Immunoblot analysis using the indicated antibodies on whole protein extracts obtained from wild type or mESCs expressing ΔSET- (ΔEzh2#1 and ΔEzh2#2) or homozygous EZH2 Y641N (Y641N#1 and Y641N#2). Vinculin and H2A served as loading controls. B: Immunoblot analysis using the indicated antibodies on whole protein extracts obtained from untransfected wild type or homozygous EZH2 Y641N expressing mESCs or from the same cells upon transient transfection with empty vector (EMPTY) or vector carrying hEZH2 (EZH2). Vinculin and H2A served as loading controls.

As a proof of concept, I reintroduced a human wild type EZH2 form in both wt, Y641N#1 and Y641N#2 clones. The transfection of an hEZH2-expressing plasmid and its empty counterpart, used as control, was performed using lipofectamine 2000 reagent according to manufacture's instructions. After 24 hours of puromycin selection, cells were collected and total protein lysates were obtained. As can be observed in Figure 4B, re-expression of wt EZH2 in N-clones led to a recovery and furthermore to increased levels of H3K27me3 when compared to wild type cells. The observed increase is specific since no effects on H3K27me3 were observed when cells were transfected with the empty vector. Moreover, H3K27me3 levels were unaffected in wild type cells, independently on the type of plasmid used for transfection (EZH2-expressing or empty). As reported in literature for EZH2 Y641N, also in our experimental conditions EZH2 Y641N cooperates with its wt counterpart leading to an increased accumulation of H3K27me3. Moreover, this result confirms the integrity of the genome-edited Y641N-harboring alleles.

4.1.3 EZH2 Y641N reduced protein levels are not attributable to a proteasome-dependent degradation

As shown in figure 4A, EZH2 protein levels were reduced in both N-mutant clones. To get insight into this result, I asked whether this could be due to an increased proteasome

degradation of the protein. Therefore, I treated wild type and Y641N#2 cells with 10 µM MG132 (or DMSO as vehicle) for 2, 5 or 8 hours. At the end of the treatment cells were collected and total protein extracts were obtained. The ubiquitin–proteasome pathway is the major molecular mechanism that regulates the concentration of specific proteins in the cells by their ubiquitination for subsequent proteasome-mediated degradation[224]. It has been already reported that treatment with the proteasome inhibitor MG132 results in a depletion of monoubiquitinated histone[225]. As shown in the WB analysis in Figure 5, MG132 treatment resulted in a time-dependent depletion of H2AK119ub and in a concomitant stabilization of p53 protein levels, used as positive control[226]. Indeed, EZH2 Y641N protein levels were not stabilized upon MG132 treatment. According to this, no recovery in H3K27me3 deposition was observed. Moreover, qPCR analysis (data not shown) demonstrated that *Ezh2* expression levels were not indeed affected by the presence of Y641N aminoacidic substitution. These results show that the reduced levels of EZH2 Y641N protein are neither attributable to a transcriptional defect nor to a proteasome-dependent degradation. This means that, other post-transcriptional or post-translational mechanisms must be involved.
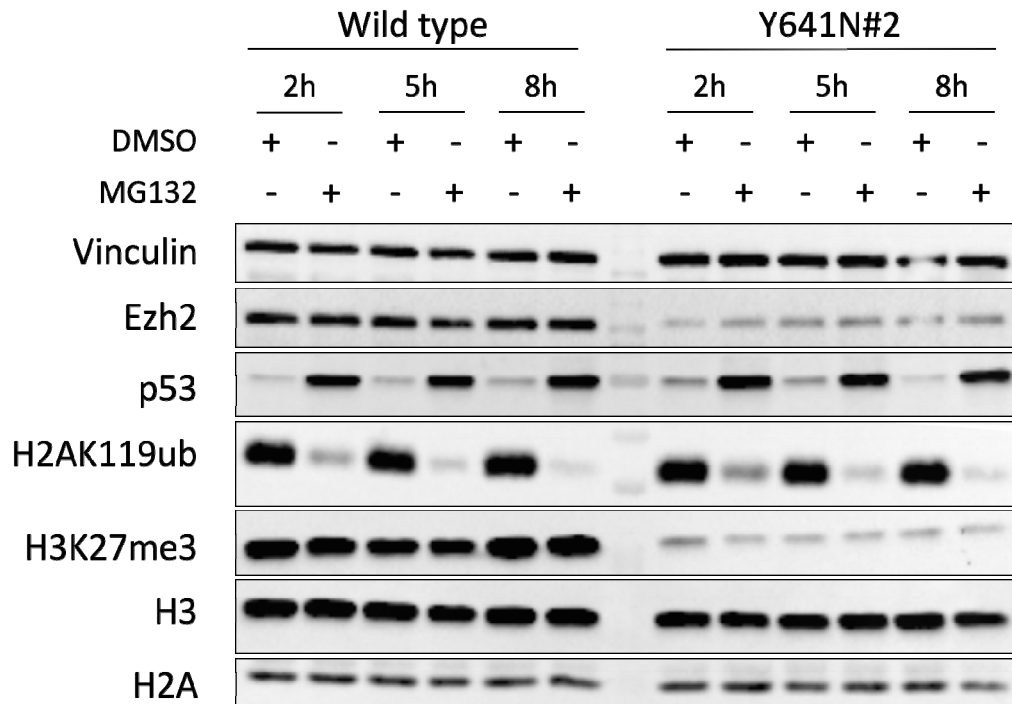
Figure 5: Homozygous EZH2 Y641N protein destabilization is not due to an enhanced proteasome-dependent degradation.

Immunoblot analysis with the indicated antibodies on whole protein extracts obtained from wild type or homozygous EZH2 Y641N expressing mESCs upon treatment with 10 μM MG132 or DMSO (as vehicle) for 2, 5 or 8 hours. Vinculin, H3 and H2A served as loading controls.

4.1.4 EZH2 Y641N has a nuclear localization and is able to complex with SUZ12 and EED and to bind chromatin

At this point I asked whether this reduction in EZH2 Y641N protein levels was also associated to a relocalization of the protein. To address this issue I performed nucleus/cytoplasm fractionation by using Nuclear Prep Buffer. The results of WB analysis performed on total protein lysates, nuclear protein fractions and soluble protein fractions of wild type and N-mutant clones are shown in Figure 6A. β-tubulin and vinculin were used as markers for soluble fractions whereas H3 as marker for the

nuclear fraction. First, I confirmed that EZH2 Y641N protein was less expressed compared to the wild type enzyme. However, its localization was not affected. In fact, both EZH2 Y641N and wt EZH2 had a preferential nuclear localization independently of expression levels.

A

Total lysates   Nuclear fractions   Cytoplasmic fractions

B

INPUT (2.5%)   HA-IP   Ezh2-IP
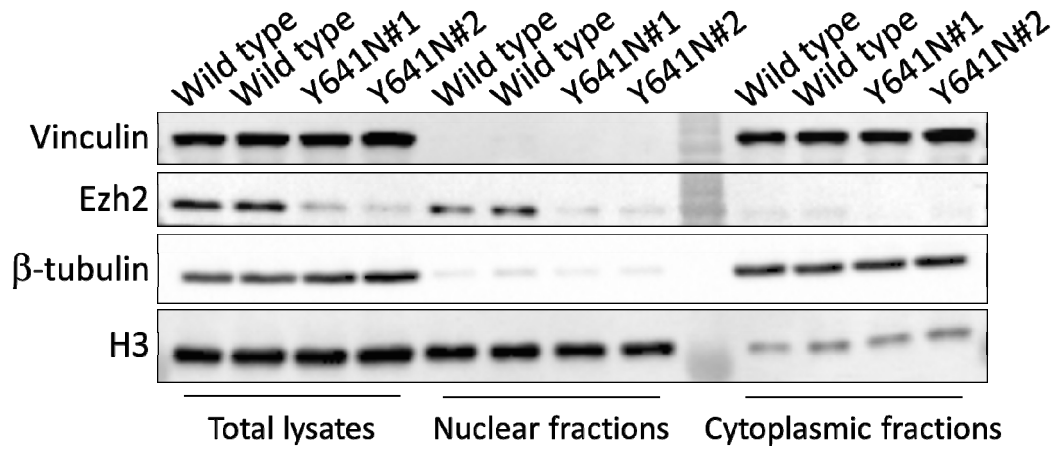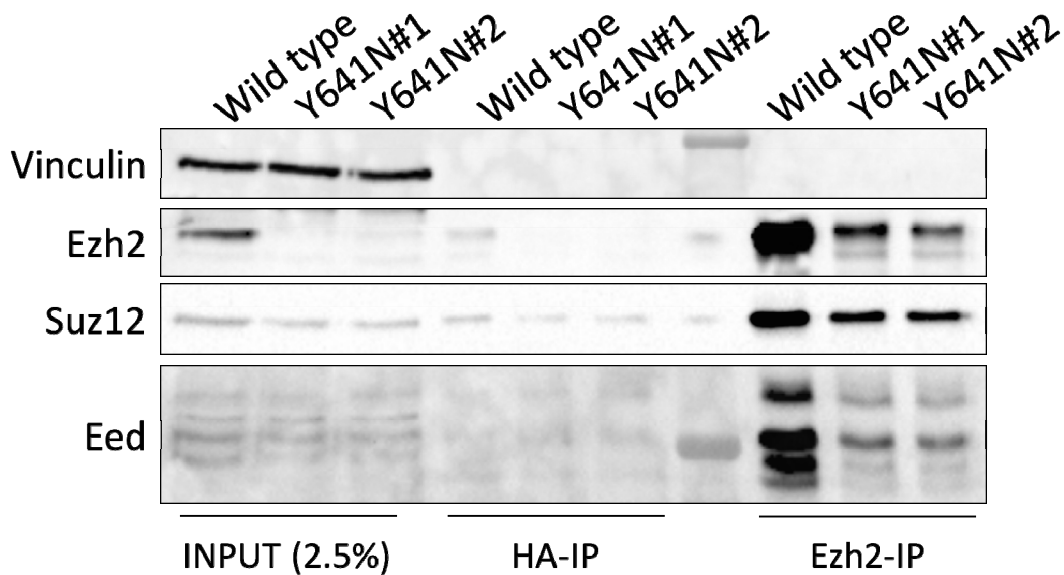
C

Figure 6: Homozygous EZH2 Y641N is a nuclear protein able to associate with SUZ12 and EED and to bind chromatin.

A: Immunoblot analysis with the indicated antibodies on whole protein extracts or nuclear/cytoplasmic fractions obtained from wild type or homozygous EZH2 Y641N expressing mESCs (Y641N#1 and Y641N#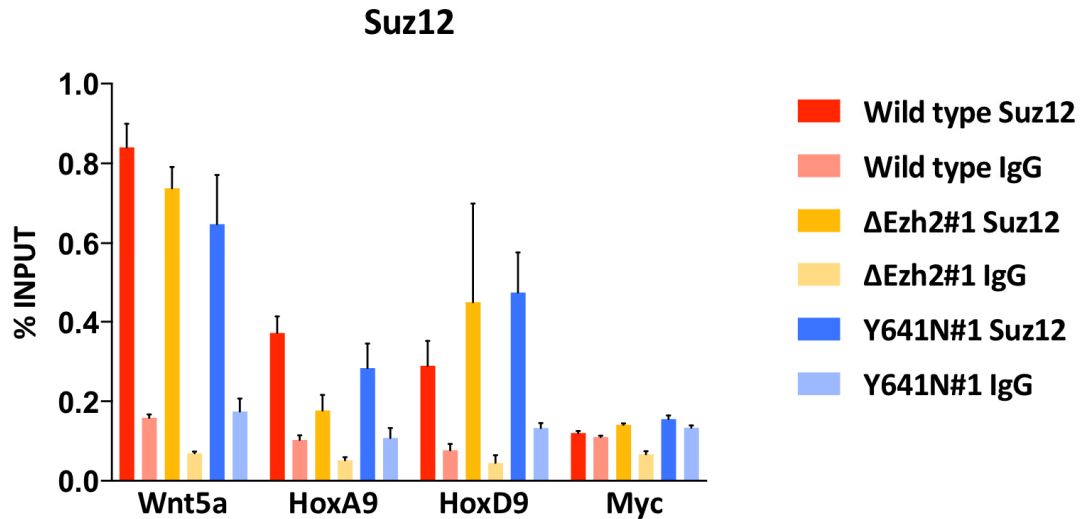2) after cell fractionation. Vinculin and β-tubulin served as cytoplasmic markers whereas H3 was used as nuclear marker. B: Immunoblot analysis with the indicated antibodies on INPUTS (2.5%) and immunoprecipitated proteins from total wild type or homozygous EZH2 Y641N expressing mESCs (Y641N#1 and Y641N#2) lysates with Ezh2 (AC22)-crosslinked Protein A sepharose beads. HA-crosslinked Protein A sepharose beads were used as unrelated IP control. C: ChIP-qPCR analyses in wild type, ΔEzh2#1 and Y641N#1 mESCs performed with anti-Suz12 antibody at the indicated loci. SUZ12 ChIP enrichments are normalized to input. ChIPs with rabbit IgG were made as negative control. Wnt5a, HoxA9 and HoxD9 were used as typical Polycomb targets whereas Myc served as negative control region.

Since EZH2 requires the association with SUZ12 and EED to exert its catalytic activity, I evaluated EZH2 Y641N capability to form a normal PRC2 complex.

Immunoprecipitations with Protein A (PA) Sepharose beads crosslinked with a anti-Ezh2 antibody were performed on total protein extracts of N-mutant (Y641N#1 and Y641N#2) as well as of wild type cells. Western Blot analysis (Fig. 6B) shows that, as already observed in Figure 4A, the SUZ12 protein levels were indeed slightly decreased in presence of EZH2 Y641N thus suggesting a strong interplay between the two subunits that can influence each other activity and/or stability. Importantly, EZH2 Y641N was able to bind SUZ12 and EED subunits as the wt enzyme did. This result strongly suggests that EZH2 Y641N containing-PRC2 is still able to bind chromatin. To address this point, I performed ChIP-qPCR for SUZ12 in both wt, ΔEzh2#1 and Y641N#1 cells. The SUZ12 enrichment was determined at established Polycomb targets and is presented in Figure 6C. A negative region (Myc) was also included into the analysis. SUZ12 binding was maintained at Wnt5a and HoxD9 TSS regions whereas it was partially displaced from HoxA9 TSS region, in the ΔSET-EZH2 clone ΔEzh2#1. Its binding was not affected at all the three analyzed regions in Y641N#1 cells. EZH2 Y641N presence doesn't seem to affect PRC2 capability to bind chromatin, at least at some Polycomb typical targets. Overall, these results show that EZH2 Y641N, despite less expressed, is a nuclear chromatin bound protein, exactly as its wt counterpart.

4.1.5 EZH2 Y641N does not impair mESC differentiation in EBs

To further characterize Y641N mutation I evaluated the differentiation capability of cells expressing EZH2 Y641N. ΔEzh2#1 and wild type cells were also included in the embryoid bodies (EB) formation assay. This assay implies the formation of three-dimensional aggregates of cells differentiating into the three germ layers upon LIF removal and addition of retinoic acid thus recapitulating development *in vitro*. As

presented in Figure 7A, ΔEzh2#1-derived EBs at day 9 displayed a disorganized structure and were very small compared to wt. On the other hand, Y641N-derived EBs showed both morphology and size comparable to wt and a normal repression of pluripotency genes and activation of differentiation markers (Fig. 7A-B). The differentiation defects observed in ΔSET clones-derived EBs seem to be indeed H3K27me3-dependent. The residual H3K27me3 in EZH2 Y641N clones detected by WB analysis, is probably sufficient to sustain an almost-normal differentiation of mESC into EBs.

A

d9

Wild type   △Ezh2#1   Y641N#1   Y641N#2

2X

4X

B

Pluripotency

Oct4

Nanog

Wild type
ΔEzh2#1
Y641N#1
Y641N#2

Mesoderm

Pax3

Ectoderm

Mausashi

Nestin
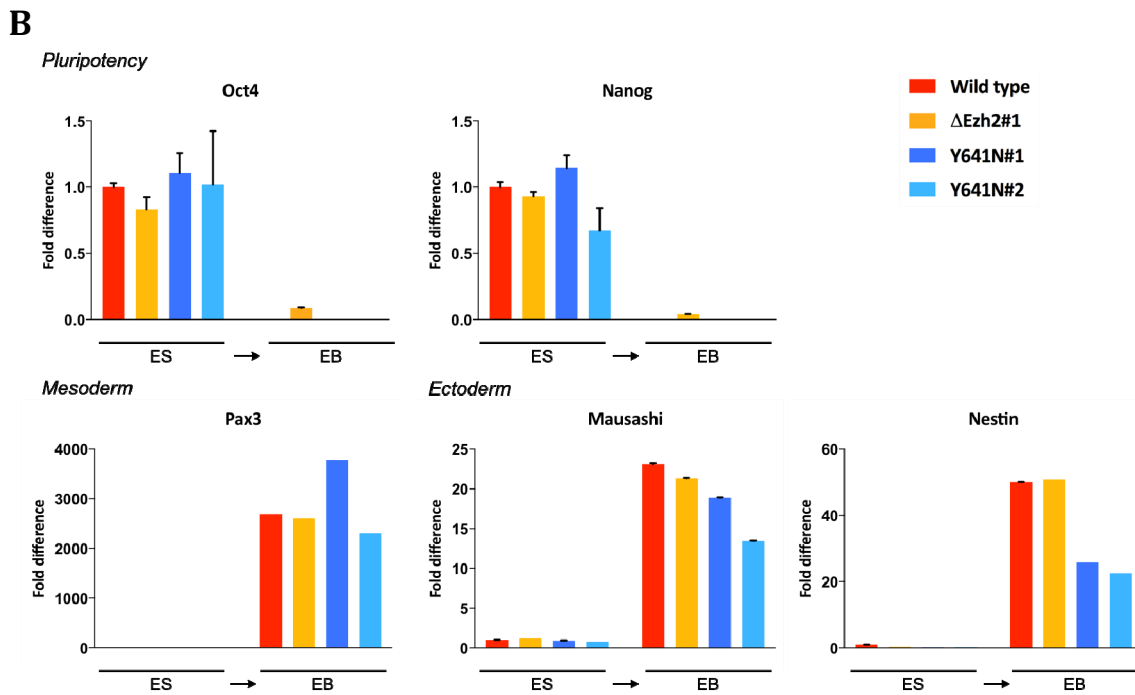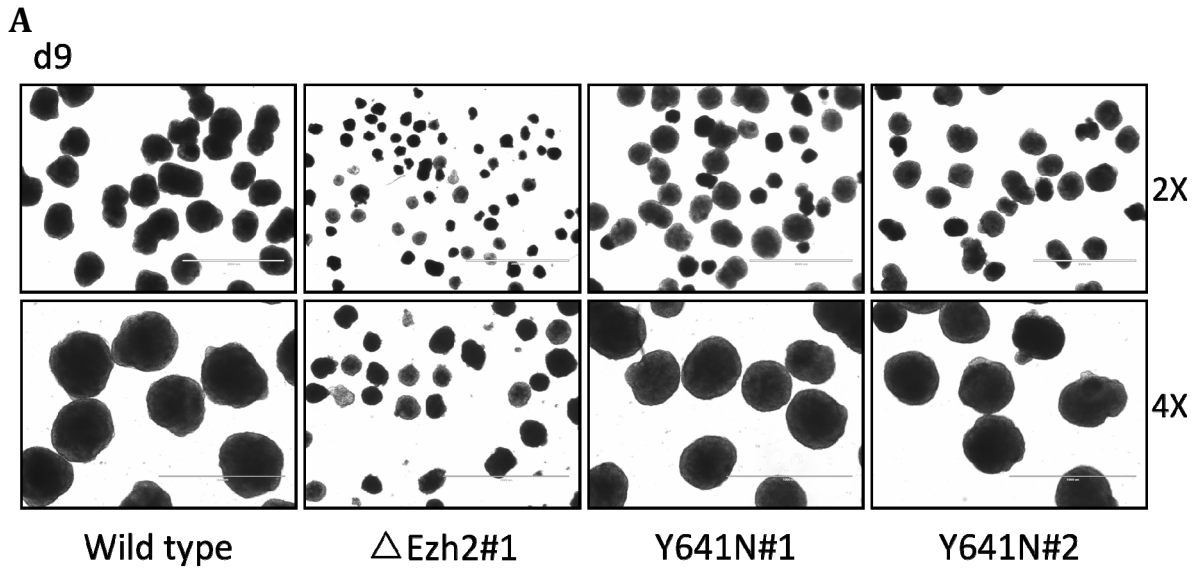
Figure 7: Homozygous EZH2 Y641N does not impair mESCs differentiation capability.

A: Pictures at day 9 of EBs derived from wild type, ΔEzh2#1, Y641N#1 and Y641N#2 mESCs. Two different magnifications are shown. B: Relative expression of the indicated pluripotency/differentiation markes determined in wild type, ΔEzh2#1, Y641N#1 and Y641N#2 mESCs before (ES) and after 9 days of differentiation (EB). Gapdh served as normalizing expression control.

4.1.6 Strategy to obtain heterozygous EZH2 Y641N expressing mESCs

During the screening I identified a particular clone that was chosen to generate an isogenic mESC model for heterozygous EZH2 Y641N mutant. This clone presented one allele with the desired Y641N mutation and the other allele with only a 1 bp deletion (Fig. 8A). I decided to re-target the deleted allele to reintroduce, through a HDR-based strategy, the missing G nucleotide. In this way, I will reconstitute a functional wildtype allele thus perfectly mimicking the heterozygosity of Y641N observed in human diseases. As a proof of principle, reintroduction of wt hEZH2, reconstituting an "heterozygous condition", led to hyper-trimethylation of H3K27 (Fig. 8B).

A   **B**
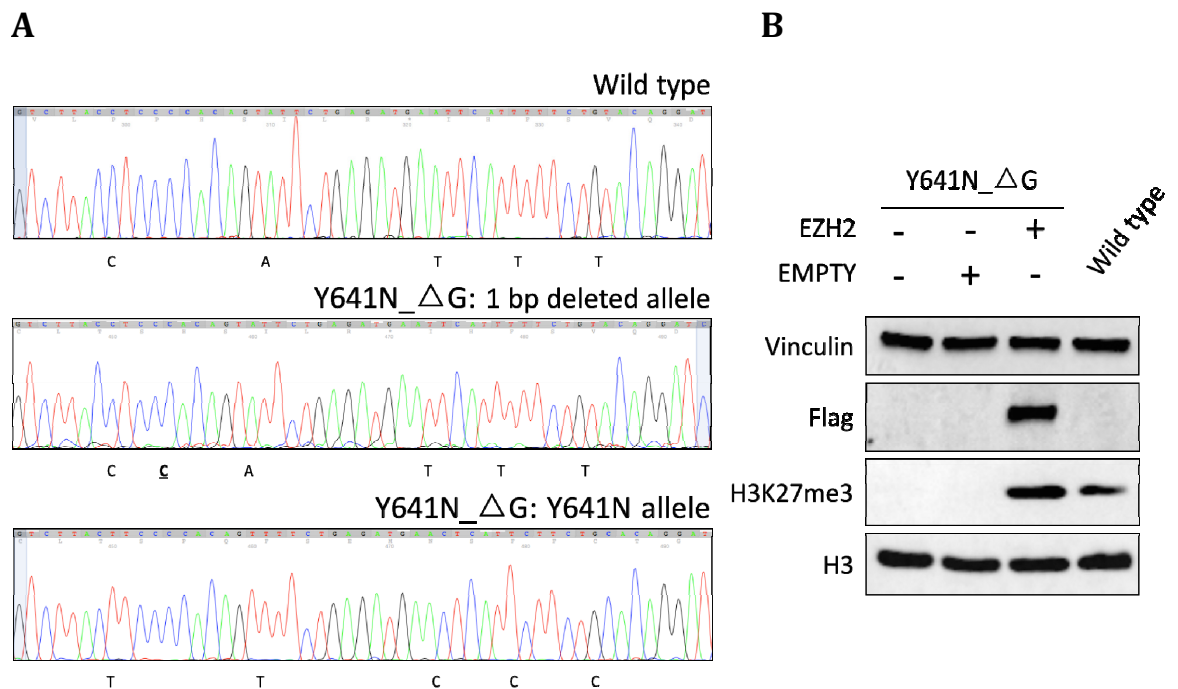
Figure 8: Future strategy to obtain heterozygous EZH2 Y641N mESCs.

A: Sanger sequencing analysis on specific PCR products obtained from amplification of gDNA extracted from wild type or Y641N_ΔG mESCs. Sequences of both alleles from Y641N_ΔG cells have been compared to wild type sequence. Differential bases are indicated. B: Immunoblot analysis using the indicated antibodies on whole protein extracts obtained from untransfected wild type cells or Y641N_G-*Ezh2* expressing mESCs

before and after transient transfection with vector carrying hEZH2 (EZH2) or empty vector (EMPTY). Vinculin and H3 served as loading controls.

The previous results are quite puzzling. Homozygous Y641N substitution acts as loss of function, however the effect is incomplete since some residual H3K27me3 in N-mutant cells can still be detected. EZH2 Y641N is indeed able to form PRC2 complex with SUZ12 and EED and to bind chromatin; moreover, it doesn't trigger any major differentiation defect as observed from EBs assay results. Despite I couldn't discriminate if the reduced H3K27me3 enzymatic activity is simply due to the reduced levels of protein expression, I asked whether this residual enzymatic activity can be attributable to an EZH1-dependent compensation process.

4.2 Generation of *Ezh1* KO mESC

4.2.1 Strategy used to obtain *Ezh1* KO mESCs

In order to verify whether the apparent incomplete loss of function features of EZH2 Y641N are due to an EZH1-compensation mechanism, I applied CRISPR/Cas9 technology to knock out the *Ezh1* gene in different cellular backgrounds. Indeed, a double guide approach was used on wt, ΔEzh2#1, Y641N#1 and Y641N#2 cells, previously generated. This approach implies the use of wt Cas9 in combination with two (or more) sgRNAs to generate macro-deletion thus enhancing knock out efficiency. I designed two independent RNA guides targeting exons 2 and 4. With this approach I could obtain a 1554 bp-long macro-deletion when Cas9 is targeted and cuts at both target sites. RNA guides were cloned into px459 plasmid that, in addition to wt Cas9 enzyme, carries a puromycin resistance gene. Transfection with lipofectamine 2000 was performed as

described in the Materials and Methods section. After 24 hours from transfection cells were subjected to puromycin selection. Resistant cells were then subjected to dilution cloning until the growth of visible clones. My initial idea was to screen clones by Western blot for the loss of EZH1 protein expression. Unfortunately, despite many attempts, I was not able to set a reproducible immunoblot screening strategy since three distinct antibodies that I have tested did not recognize endogenous EZH1 (data not shown). The clones were therefore screened by genomic PCR to confirm the allelic deletion (Fig. 9A), followed by Sanger sequencing. An *Ezh1* verified KO clone for each cellular background was selected for further analysis.

A
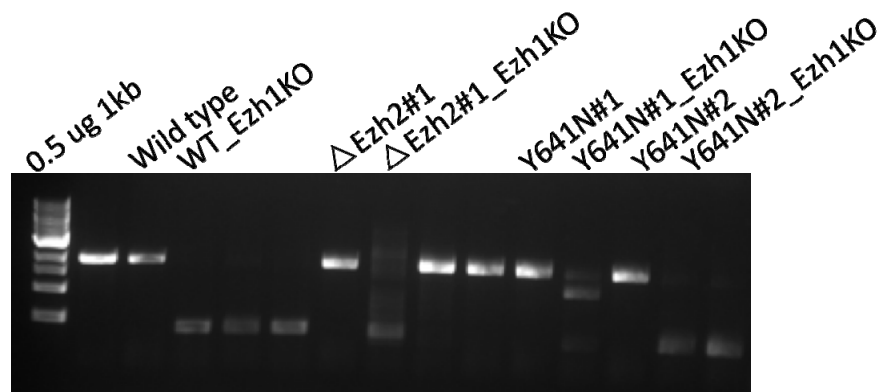
B
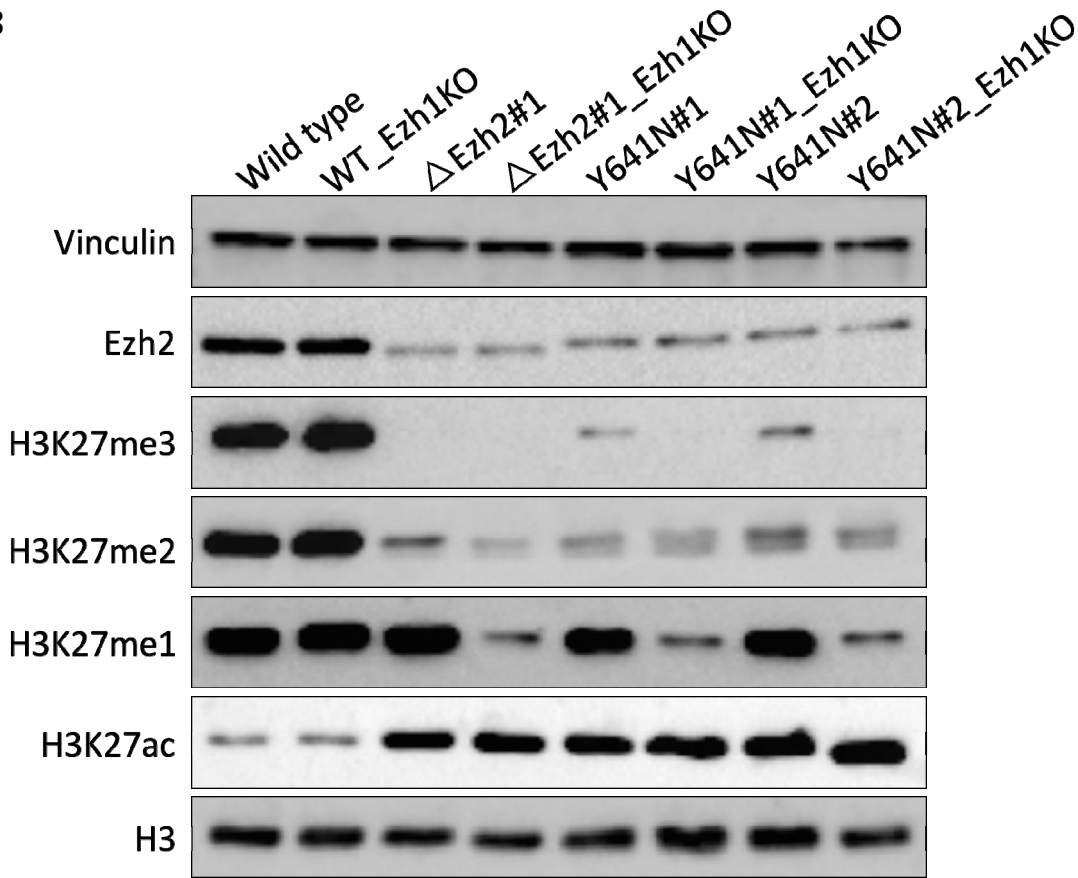
Figure 9: Ezh1 KO strongly impairs H3K27me1 deposition in both ΔSET- and homozygous Y641N – EZH2 expressing mESCs.

A: PCR analyses using specific primers showing *Ezh1* locus-specific deletions in wild type (WT_Ezh1KO), ΔEzh2#1 (ΔEzh2#1_Ezh1KO), Y641N#1 (Y641N#1_Ezh1KO) and Y641N#2 (Y641N#1_Ezh1KO) mESCs. B: Immunoblot analysis using the indicated antibodies on whole protein extracts obtained from wild type cells, ΔEzh2#1, Y641N#1 and Y641N#2 before and after Ezh1 KO induction. Vinculin and H3 served as loading controls.

4.2.2 Characterization of *Ezh1* KO mESC by Western Blot analysis

Indeed, WT_Ezh1KO, ΔEzh2#1_Ezh1KO, Y641N#1_Ezh1KO and Y641N#2_Ezh1KO were firstly characterized by Western Blot analysis performed on total protein extracts. As

shown in Figure 9B, EZH2 ΔSET and the Y641N mutation in an *Ezh2*-null background strongly impaired H3K27me1 deposition. This confirms a major role of EZH1 in depositing the mono-methylation on H3K27 also in my models. H3K27me1 resulted unaffected in a wt EZH2 background thus suggesting that EZH2 is perfectly able to compensate for the loss of EZH1. Moreover, *Ezh1* KO induced the complete loss of the residual H3K27me3 mark in EZH2 Y641N expressing cells. No changes were observed in H3K27me2 while H3K27ac was increased already in the absence of EZH2 activity. These results confirmed that in the KO clones EZH1 catalytic activity was indeed impaired. Moreover, these data further demonstrate that EZH1 is partially able to compensate H3K27me3 activity of EZH2 in presence of EZH2 Y641N but not in a EZH2 ΔSET condition. Moreover, EZH1, whose activity is completely compensated by wt EZH2, has a crucial role in H3K27me1 deposition.

4.2.3 EZH1 loss has no effects on mESC differentiation into EBs

I have already demonstrated that deletion within EZH2 SET domain strongly impairs mESC differentiation capability as shown in Figure 5. In order to elucidate EZH1 role in differentiation an EBs formation assay was performed on wild type and WT_Ezh1KO cells. As shown in panels A and B of Figure 10, WT_Ezh1KO-derived EBs were comparable to wt in terms of morphology, size and activation of differentiation markers expression. My data suggest that wt EZH2 is perfectly able to compensate for EZH1 function. This can possibly take place also during mESC differentiation.

A
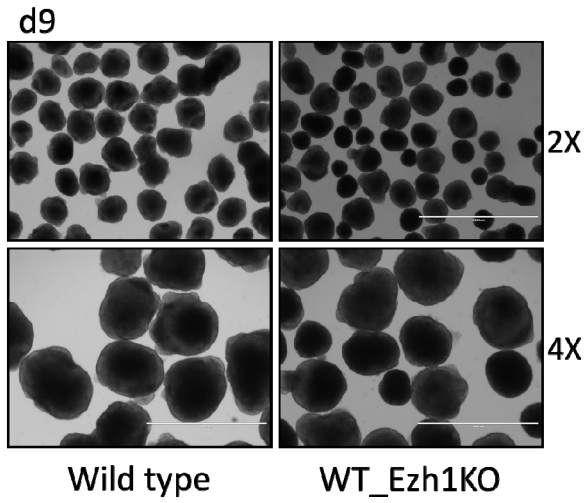
B

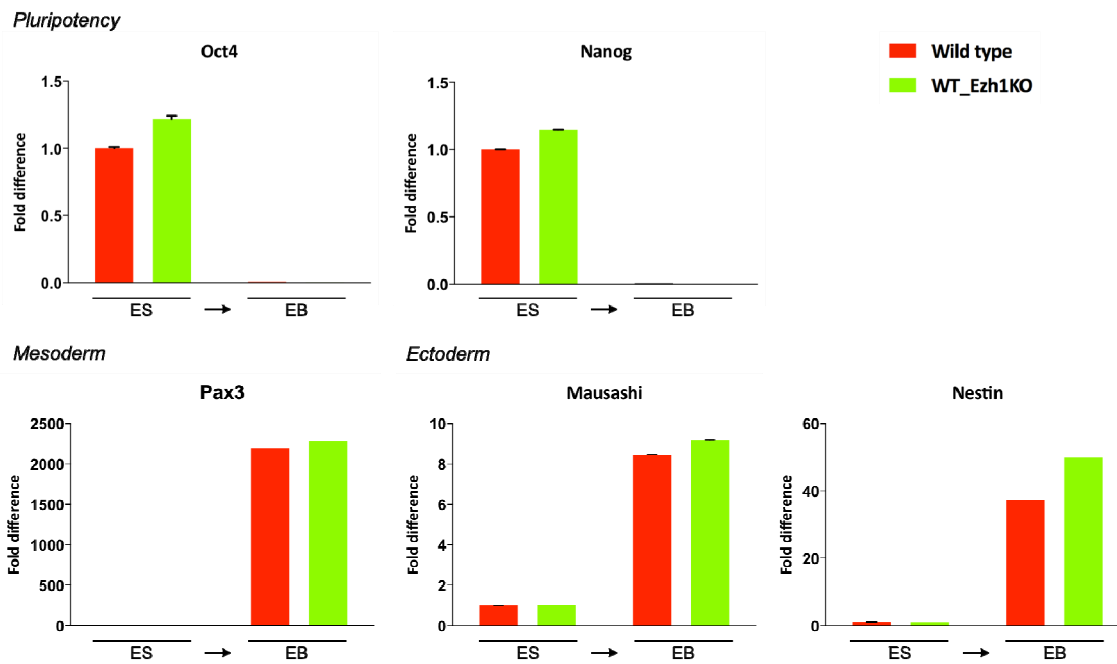*Figure 10: Ezh1 KO does not impair mESCs differentiation into EBs.*

A: Pictures at day 9 of EBs derived from wild type and WT_Ezh1KO mESCs. Two different magnifications are shown. B: Relative expression of the indicated pluripotency/differentiation markes determined in wild type and WT_Ezh1KO mESCs before (ES) and after 9 days of differentiation (EB). Gapdh served as normalizing expression control.

4.2.4 EZH1 compensatory role in absence of EZH2 catalytic activity

To evaluate EZH1 and EZH2 specific contribution in H3K27me3 deposition, I performed H3K27me3 ChIP-qPCR analysis on wt, WT_Ezh1KO, ΔEzh2#1 and ΔEzh2#1_Ezh1KO cells. Panels A and B of Figure 11 show H3K27me3 enrichment (relative to % of INPUT or H3, respectively) at 2 classical Polycomb targets (Wnt5a and HoxA9 TSS) and at one negative region (Utp6). H3K27me3 deposition was maintained at Wnt5a TSS but almost lost at HoxA9 TSS, in ΔEzh2#1. This correlates with the locus-specific maintenance of SUZ12 binding at some Polycomb targets as observed in Figure 6C. *Ezh1* KO abolished H3K27me3 enrichment at both targets thus suggesting that H3K27me3 deposition in ΔEzh2#1 is indeed EZH1-dependent, at least at these two targets. No changes were observed in H3K27me3 deposition upon *Ezh1* KO in wt cells further confirming that wt EZH2 is perfectly able to compensate EZH1 loss.
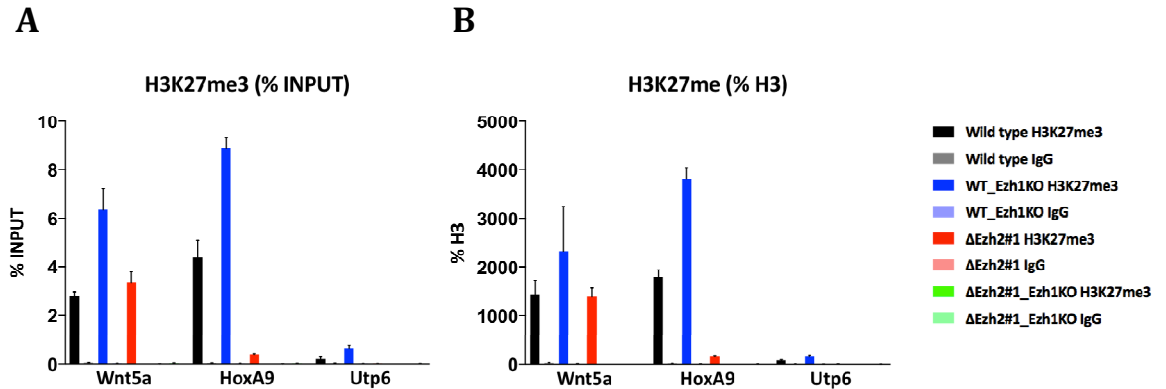
Figure 11: ΔSET-EZH2 expressing mESCs maintain H3K27me3 enrichment at typical Polycomb targets.
ChIP-qPCR analyses in wild type, WT_Ezh1KO#1, ΔEzh2#1 and ΔEzh2#1_Ezh1 KO mESCs performed with anti-H3K27me3 antibody at the indicated loci. H3K27me3 ChIP enrichments are normalized to input (A) or to histone H3 density (B). ChIPs with rabbit IgG were made as negative control. Wnt5a and HoxA9 were used as typical Polycomb targets whereas Utp6 served as negative control region.

4.2.5 H3K27me3 ChIP-RX

In order to elucidate EZH1 and EZH2 contribution in genome-wide H3K27me3 deposition, I performed H3K27me3 ChIP-seq analysis on the same cell lines used for the previous ChIP-qPCR assay (WT, WT_Ezh1KO, ΔEzh2#1 and ΔEzh2#1_Ezh1KO). In particular ChIP with reference exogenous genome (ChIP-Rx) approach was applied. This relatively new method implies the use of a defined amount of reference exogenous epigenome (namely "spike-in") in order to perform genome-wide quantitative comparisons of histone modification status across cell populations[223]. Chromatin derived from wt, WT_Ezh1KO, ΔEzh2#1 and ΔEzh2#1_Ezh1KO cells was mixed with an amount of Drosophila S2 cell chromatin corresponding to the 5% of the total chromatin amount and H3K27me3 ChIP was performed following a standard protocol, as reported in Material and Methods section. After sequencing and mapping, ChIP-seq reads were normalized to the percentage of reference genome reads in the sample. Peak calling returned about 8272 peaks in wt cells. Since the role of H3K27me3 deposition at promoters is quite well characterized, I decided to perform a promoter-based analysis. I focused my attention on approximately 5903 H3K27me3-enriched promoters, accounting for almost 72% of the total peaks that were identified. This confirmed (right panel of Fig. 12) the complete loss of H327me3 mark observed by WB (Fig. 9B) since no positive regions (in neither analysis) were detected in the ΔEzh2#1_Ezh1KO condition whereas ChIP-seq experiment was indeed successful for the other three cell types. Promoter-based analysis revealed an almost complete overlap between H3K27me3-enriched promoters in wt and *Ezh1* KO (WT_Ezh1KO) conditions thus suggesting that, if EZH1 may have a role in H3K27me3 deposition, this activity is entirely compensated by EZH2. Interestingly, a consistent amount of H3K27me3-marked promoters (half of

common between wt and WT_Ezh1KO) were maintained also in ΔSET-EZH2 cells. (ΔEzh2#1, Fig. 12). To gain insight into the nature of these "positive" regions retained in ΔEzh2#1, the average coverage of H3K27me3 around the TSS (+/-10 kb) was calculated. The analysis was performed taking into account the H3K27me3-enriched promoters identified in the wt cells. Figure 14A shows the genomic distribution of ChIP-seq peaks (promoter-based analysis) at common (n° 2525) and not common targets (n° 2286), before and after the application of the spike-in correction. Traditional analysis (no spike-in) gave us a falsified result with ΔEzh2#1 distribution profile standing out among the other three conditions, showing a great increase in term of intensity at common promoters (Fig. 14A, left upper panel). As shown by Orlando and colleagues, normalizing to a reference exogenous genome rectifies the HPTMs genome occupancy signals thus allowing also the revealing of subtle epigenomic changes [223]. Indeed, normalization with the Drosophila reference (normalized reference-adjusted RPM [RRPM]) allowed us to obtain the real H3K27me3 distribution scenario across TSS (Fig. 14A, bottom panels). Both wt and WT_Ezh1KO showed a strong enrichment of H3K27me3 around the TSS. In both cases the H3K27me3 distribution shape resembled what is usually observed in ES, two peaks with a sharp dip around the TSS, corresponding to the position of the nucleosome-depleted zone[227]. The two profiles were indeed comparable in terms of shape, distribution and intensity. A striking decrease in H3K27me3 signal across TSS was observed in ΔEzh2#1 cells. However, the signal appeared to be significant and had the classical shape observed in wt cells (Fig. 14A, bottom panels). This result is perfectly represented also by the heatmaps (Fig. 14B). ΔEzh2#1cells maintained some specific, despite less intense, H3K27me3 enrichment at promoter regions. Genomic snapshots retrieved by UCSC genome browser of H3K27me3 enrichment at typical Polycomb targets reflect exactly the same scenario, with H3K27me3 enrichment strongly reduced

but still retained in ΔEzh2#1 cells. Moreover, peaks appear much more narrow compared to wt (Fig. 13). This result could suggest again some compensatory role exerted by EZH1 in absence of EZH2. EZH1-containing PRC2 is able to deposit a certain level of H3K27me3 mark, but this is not enough to favor the spreading of the signal thus creating the typical H3K27me3-marked domains at Polycomb targets.
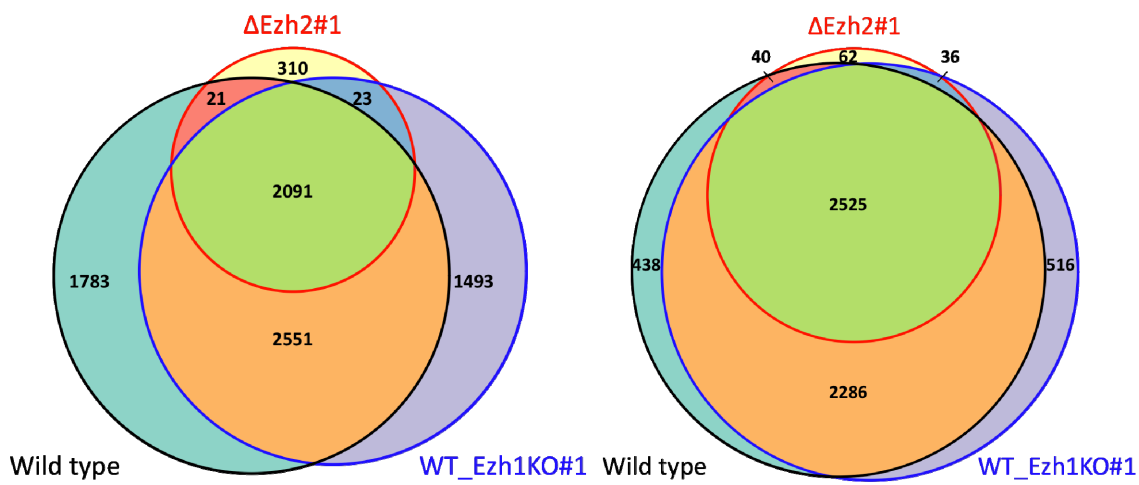
Figure 12: ΔSET-EZH2 expressing mESCs maintain significant genome-wide H3K27me3 enrichment.

Venn diagrams showing the extent of overlap of H3K27me3-enriched peaks (A) or promoters (B) between Wild type, WT_Ezh1KO, ΔEzh2#1 and ΔEzh2#1_Ezh1 KO mESCs.
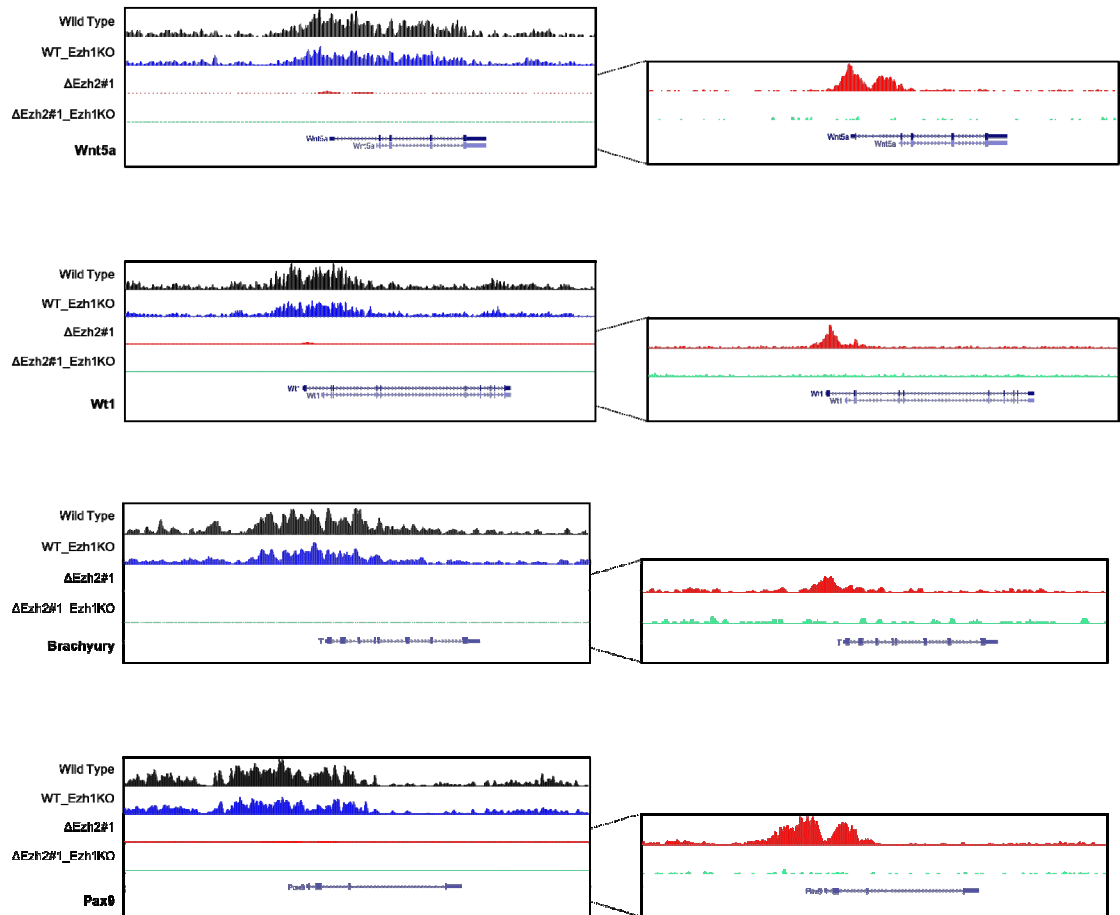
Figure 13: H3K27me3 peaks are less intense and narrower in ΔEzh2#1 (ΔSET-EZH2) expressing cells compared to wild type.

Genomic snapshots of H3K27me3 ChIP-seq experiments from wild type, WT_Ezh1KO, ΔEzh2#1 and ΔEzh2#1_Ezh1 KO mESCs. Scale was enlarged in order to appreciate differential H3K27me3 deposition between ΔEzh2#1 and ΔEzh2#1_Ezh1 KO conditions (left panels).
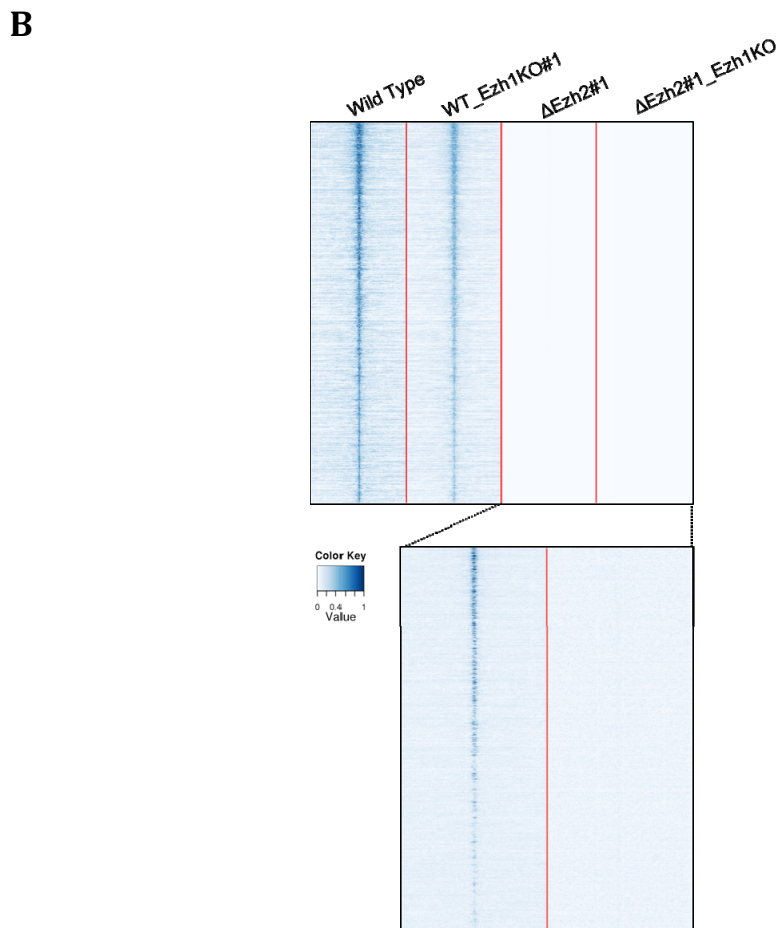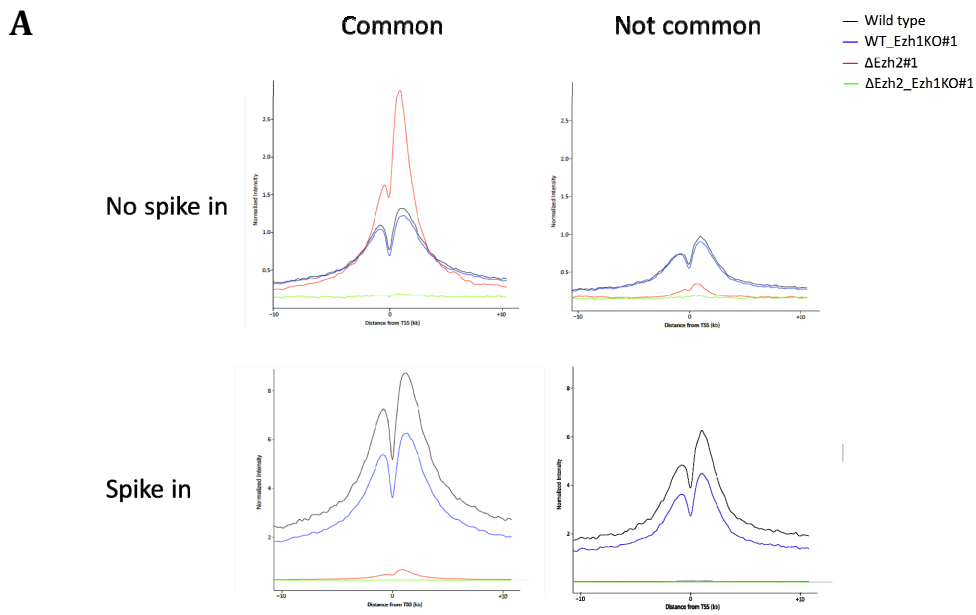
Figure 14: ChIP-RX approach reveals the real H3K27me3 differential distribution at TSS in the different analyzed cell populations.

A: Cumulative distribution of H3K27me3 enrichment at TSS (+/-10kb) at common (left panels) and not common (right panels) H3K27me3-enriched promoters before (upper panels) and after (bottom panel) spike-in correction in wild type, WT_Ezh1KO, ΔEzh2#1 and ΔEzh2#1_Ezh1 KO mESCs. B: Heatmaps showing normalized H3K27me3 ChIP-seq signal in wild type, WT_Ezh1KO, ΔEzh2#1 and ΔEzh2#1_Ezh1 KO mESCs for H3K27me3-enriched promoters +/- 10kb from TSS. Drosophila S2 spike-in normalization has been applied. The last two columns were re-analyzed (and normalized onto ΔEzh2#1) in order to appreciate differential H3K27me3 enrichment between the two conditions.

4.3 Generation of N-terminus (N-ter) deleted EZH2 mESCs

A this point I wanted to confirm my results on EZH1/EZH2 and to better investigate their interplay in depositing the different H3K27 methylations. To achieve this goal I decided to generate another kind of EZH2 inactivating mutant targeting the N-terminal part of EZH2 protein. The CRISPR/Cas9 strategy used was the same used for *Ezh1* KO. Indeed, a double-guide targeting was used to induce a macro-deletion around *Ezh2* ATG in wt and in WT_Ezh1KO cells. This deletion was firstly confirmed by PCR (Fig. 15A) and then by Sanger sequencing.

A

B   **C**
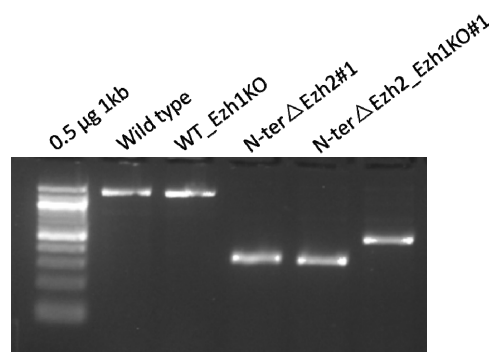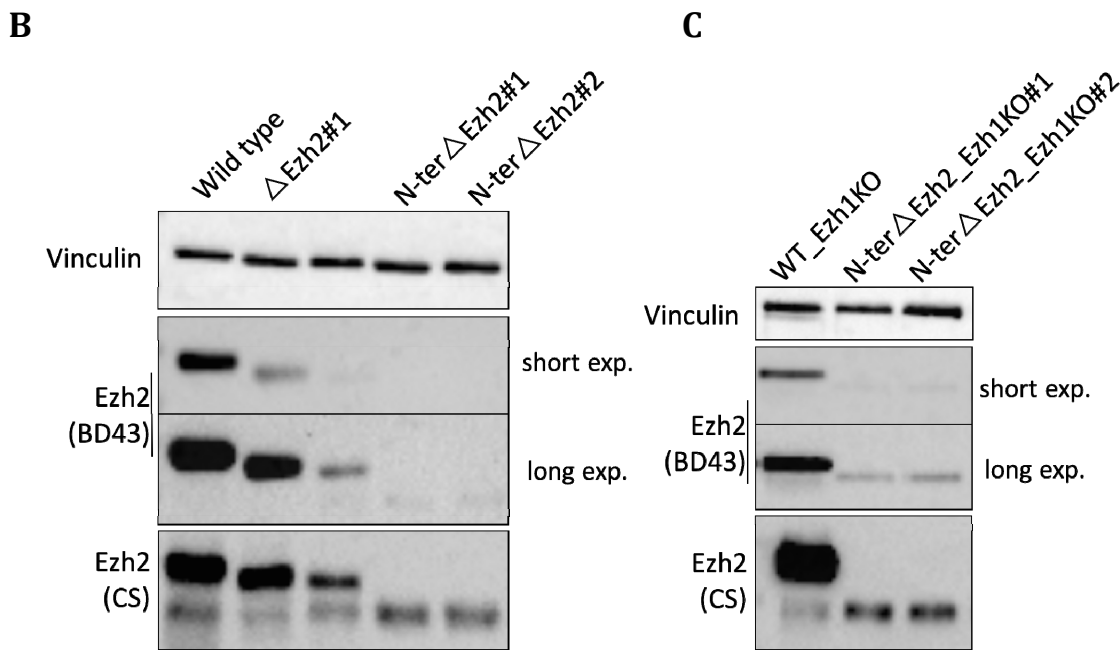
Figure 15: Screening strategy used for selecting Ezh2 KO mESC clones.

A: PCR analyses using specific primers showing Ezh2 locus-specific deletions in wild type (N-terΔEzh2#1)

and WT_Ezh1KO (N-terΔEzh2_Ezh1KO#1) mESCs. *B-C* Immunoblot analyses using the indicated antibodies

on whole protein extracts after CRISPR/Cas9-based Ezh2 KO induction in wild type (*B,* N-terΔEzh2#1 and

N-terΔEzh2#2) and Ezh1 KO mESCs (*C,* N-terΔEzh2_Ezh1KO#1 and N-terΔEzh2_Ezh1KO#2). Vinculin

served as loading control.

4.3.1 Characterization of N-ter deleted EZH2 mESCs

Even though the macro-deletion chopped out one crucial ATG, this strategy didn't

allowed me to generate a real KO. In fact WB analysis revealed the presence of a specific

lower band corresponding to EZH2, thus suggesting the formation of a truncated form of

the protein in all the targeted clones. Notably, this truncated form of EZH2 was not

recognized by the used anti-Ezh2 antibody, designed against the first 286 N-terminal aa

of the protein (Fig. 15B-C). Two deleted clones for each cellular background (WT or *Ezh1*

KO), named N-terΔEzh2#1, N-terΔEzh2#2, N-terΔEzh2_Ezh1KO#1 and N-

terΔEzh2_Ezh1KO#2, were characterized by WB analysis. As it is possible to appreciate from the WB analysis, H3K27me2 and H3K27me3 deposition was completely abolished upon N-terminal deletion of EZH2 with a concomitant increase in H3K27ac levels in both wt-*Ezh*1 and *Ezh1* KO background (Fig. 16). This result confirms that, even if a truncated form of EZH2 is still present, its activity is impaired. Moreover, H3K27me1 levels were not affected by the deletion in a wt EZH1 background. By contrast, as already showed in an EZH2 ΔSET background (Fig. 9B), the EZH1 loss concomitant with the N-terminal deletion of EZH2 abolished also H3K27me1 levels. *Ezh2* deletion and *Ezh1* KO, alone or in combination did not affect the pluripotency state of the cells, in fact Oct3/4 levels didn't change (Fig. 16). In addition to what observed in a ΔSET EZH2 background, I was able to show that EZH1 has a crucial role in performing H3K27me1 also in a different genetic context using both C-terminal and N-ter deletion of EZH2.
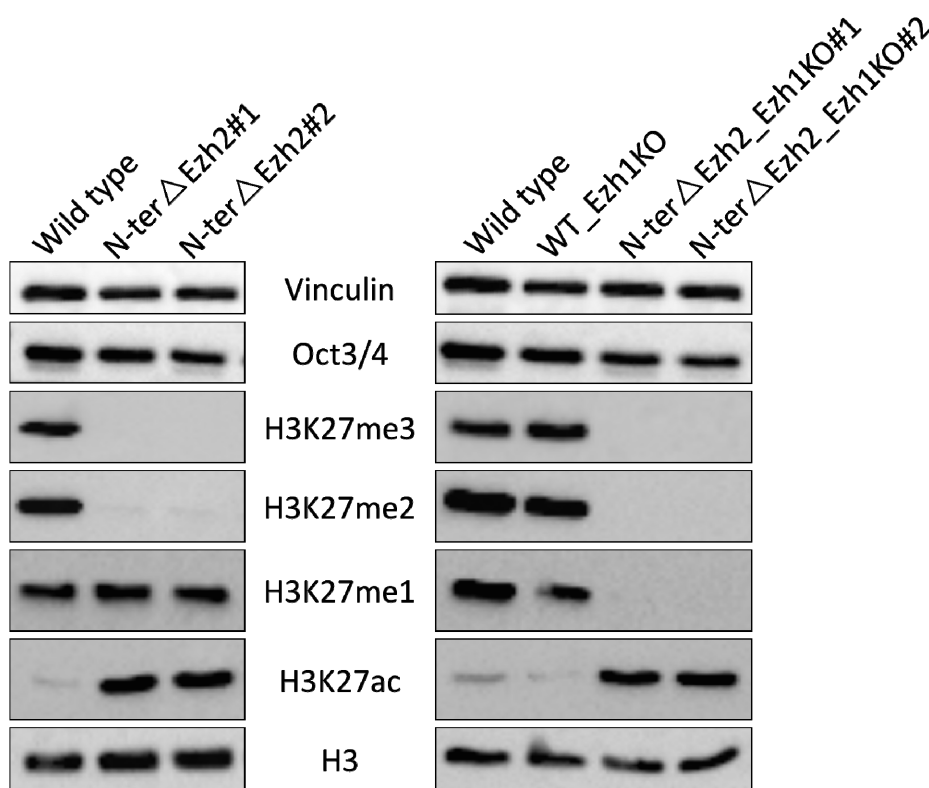
Figure 16: H3K27me1 deposition is severely impaired upon Ezh1 KO in N-terΔEzh2 mESCs.

Immunoblot analyses using the indicated antibodies on whole protein extracts after *Ezh2* KO in wild type (A, N-terΔEzh2#1 and N-terΔEzh2#2) and WT_Ezh1KO (B, N-terΔEzh2_Ezh1KO#1 and N-terΔEzh2_Ezh1KO#2) mESCs.Vinculin and H3 served as loading controls.

To evaluate more deeply EZH2 and EZH1 roles in mESCs differentiation, I performed an EBs formation assay on the new generated cell lines. As shown in panels A and B of Figure 17 (already shown in Fig. 10), *Ezh1* KO had no impact on mESCs capability to differentiate, indeed WT_Ezh1KO-derived EBs were comparable to wt in terms of morphology and activation of differentiation markers expression. As previously seen for ΔEzh2#1cells, also N-terΔEzh2#1 cells were strongly defective in the formation of EBs. Indeed, compared to the wt, N-terΔEzh2#1-derived EBs were smaller and irregular. Therefore, the observed phenotype seems to be H3K27me3-dependent. Notably, the defective phenotype appeared worse when both EZH1 and EZH2 were compromised, thus highlighting a possible role of EZH1 in differentiation that usually can be compensated by EZH2.

A

4X

Wild type   WT_Ezh1KO   N-ter△Ezh2#1   N-ter△Ezh2_Ezh1KO#1

B

Pluripotency

Oct4

Nanog

- Wild type
- WT_Ezh1KO
- N-terΔ_Ezh2#1
- N-terΔEzh2_Ezh1KO#1

Mesoderm

Pax3

Ectoderm

Mausashi

Nestin
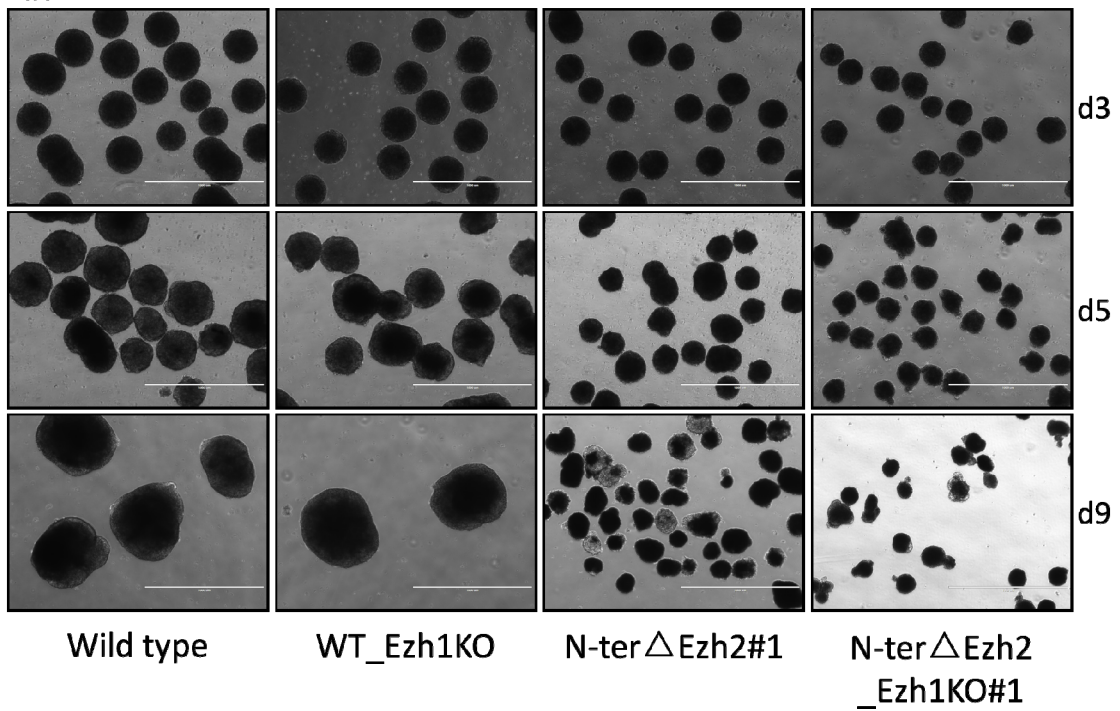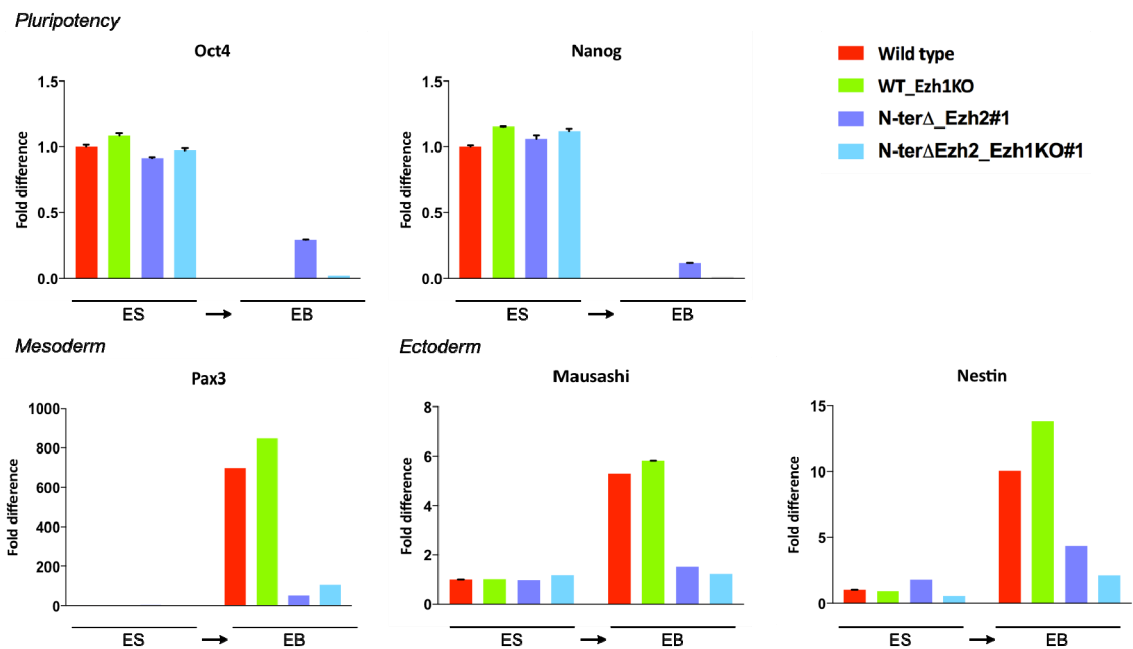
Figure 17: N-terminal deletion of EZH2 strongly impairs mESCs differentiation into EBs.

A: Pictures at day 3, day 5 and day 9 after differentiation of EBs derived from wild type, WT_Ezh1KO, N-terΔEzh2#1 and N-terΔEzh2_Ezh1KO#1 mESCs. B: Relative expression of the indicated pluripotency/differentiation markers determined in wild type, WT_Ezh1KO, N-terΔEzh2#1 and N-terΔEzh2_Ezh1KO#1 mESCs before (ES) and after 9 days of differentiation (EB). Actin served as normalizing expression control.

To understand if EZH1 can compensate for EZH2 compromised activity, I performed ChIP-qPCR for SUZ12 and H3K27me3 modification in wt, N-terΔEzh2#1 and N-terΔEzh2_Ezh1KO#1 background. Indeed, in N-terΔEzh2#1 cells SUZ12 was still enriched at classical Polycomb targets with normal H3K27me3 deposition. The presence of SUZ12 (and so of PRC2) and the deposition of the repressive mark were lost upon *Ezh1* KO, suggesting that, at least at these specific analyzed targets, EZH1 can compensate for EZH2 catalytic activity (Fig. 18A-C). Additionally, N-terΔEzh2_Ezh1KO#1 cells represent a suitable model to study PRC2 re-recruitment in a H3K27me3-free context. Indeed, PRCs recruitment to target loci is matter of great debate and many different regulatory mechanisms have been proposed. Following the hierarchical model of action of PRC2 and PRC1, H3K27me3 has been considered, for long time, the driver of PRC1 recruitment through the recognition by Cbx subunit. Indeed, human wt EZH2 was reintroduced, upon acute transfection, in N-terΔEzh2_Ezh1KO#1 cells. H3K27m3 ChIP-qPCR analysis was performed on wild type, N-terΔEzh2_Ezh1KO#1 and EZH2-transfected N-terΔEzh2_Ezh1KO#1 cells. As can be appreciated from Figure 19, H3K27me3 deposition at classical Polycomb targets was completely rescued to levels comparable to the ones of wt cells, upon reintroduction of a functional form of hEZH2. This first result suggests that de-novo recruitment of the PRC2 complex is independent of H3K27me3 deposition.

A

B
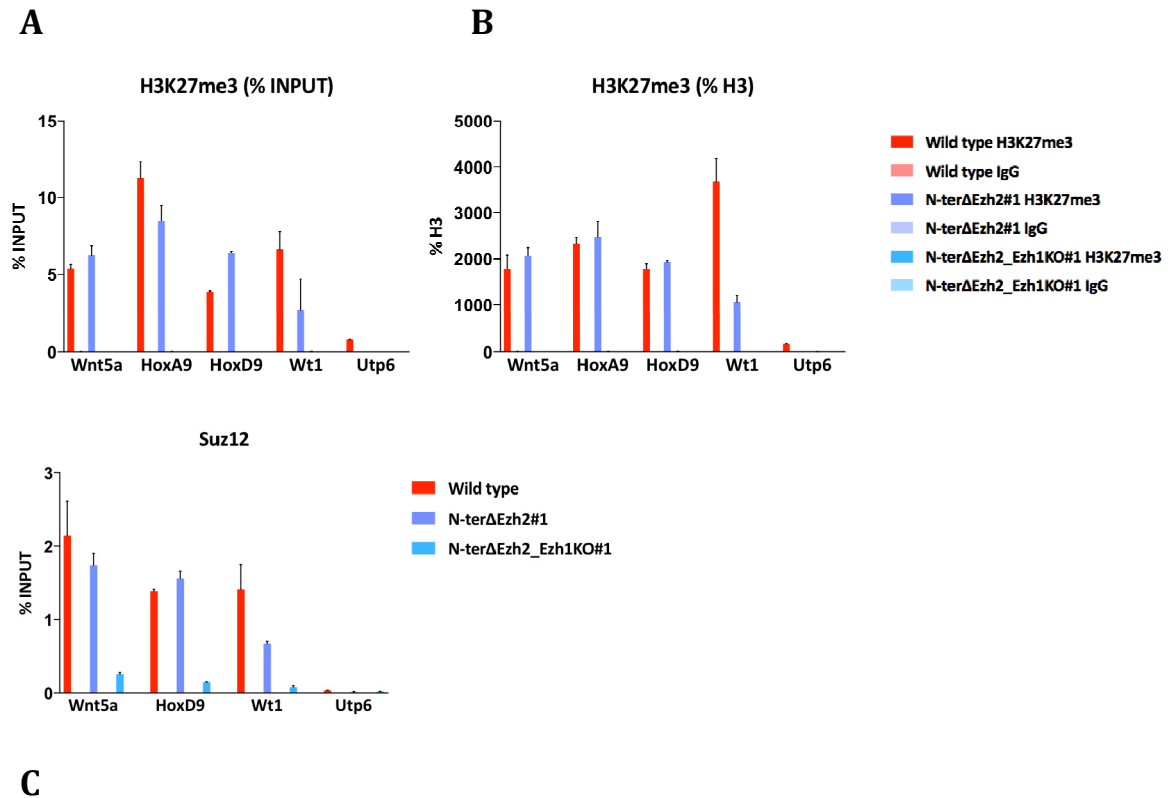
Figure 18: H3K27me3 enrichment and SUZ12 binding at typical Polycomb targets are lost upon Ezh1 KO in mESCs expressing N-terminal deleted EZH2.

ChIP-qPCR analyses in wild type, N-terΔEzh2#1 and N-terΔEzh2_Ezh1KO#1 mESCs at the indicated loci. H3K27me3 ChIP enrichments are normalized to input (A) or to histone H3 density (B). Suz12 ChIP enrichments are normalized to input (C). ChIPs with rabbit IgG were made as negative control. Wnt5a, HoxA9, HoxD9 and Wt1 were used as typical Polycomb targets whereas Utp6 served as negative control region.

Through the generation of several genetic models of mESCs I started to define the interplay between EZH1 and EZH2 in regulating H3K27 methylation states in different genetic backgrounds. I showed that, in the presence of EZH2 Y641N or ΔSET, EZH1 is partially able to compensate for EZH2 thus ensuring a certain level of H3K27me3 deposition and that EZH1 major contribute is to perform H3K27me1. Despite H3K27me3 deposition is essential for mESCs differentiation into EBs, PRC2 recruitment to its typical

targets seems to be H3K27me3-independent.

H3K27me3 (% INPUT)
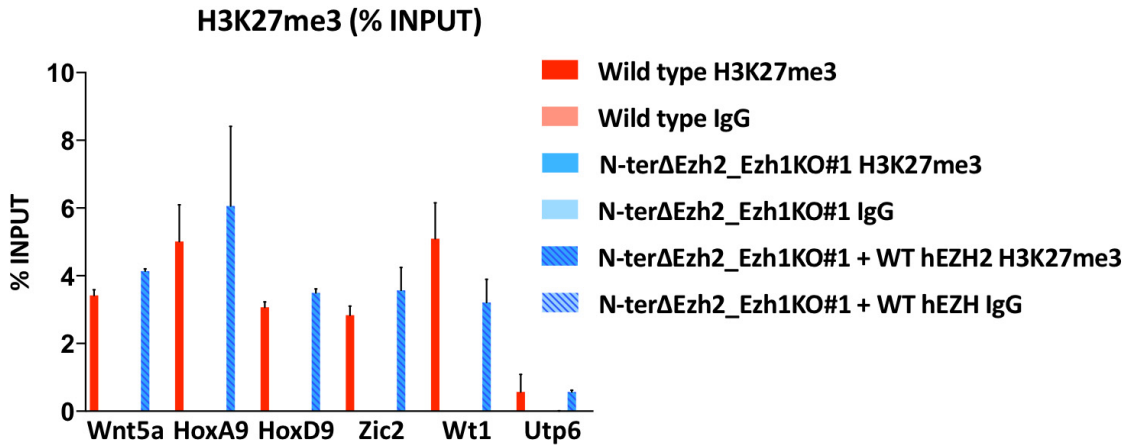
Figure 19: H3K27m3 deposition at typical Polycomb targets is completely restored after re-introduction of WT hEZH2 in N-terΔEzh2_Ezh1KO#1 mESCs.

ChIP-qPCR analyses in wild type, N-terΔEzh2_Ezh1KO and N-terΔEzh2_Ezh1KO mESCs after re-introduction of hEZH2 performed with anti-H3K27me3 antibody at the indicated loci. H3K27me3 ChIP enrichments are normalized to input. ChIPs with rabbit IgG were performed as negative control. Wnt5a, HoxA9, HoxD9, Zic2 and Wt1 were used as typical Polycomb targets whereas Utp6 served as negative control region.

4.4. Generation of mESC models for inactive EZH2

4.4.1 Screening of putative inactivating *EZH2* mutations

EZH2 Y641N expressing cells represented a useful model to study EZH1/EZH2 interplay and redundancy. However, they resulted not perfectly suitable as model for EZH2 inactivating mutations. In fact, Y641N-EZH2 was less expressed compared to wt protein and no major defects in EBs differentiation were observed in presence of EZH2 Y641N, thus suggesting a compensatory effect exerted by EZH1. Finally, homozygous Y641N

aminoacidic substitution has not yet been found in human diseases. For all these reasons, I decided to generate other models for EZH2 inactivating mutations in an *Ezh1*-null background in order to avoid any possible compensation from EZH1. Starting from published data I focused my attention on 7 putative critical residues, all encompassing the CXC or SET Ezh2 domains. Some aminoacidic substitutions at these residues were characterized by *in vitro* studies[75,228] whereas some others were directly identified in patients affected by MDS[159] (listed in Table 5).

Characterization	*D. Melanogaster*	*M. Musculus*
In vitro	C545Y	C588Y
In vitro	R699A/H	R685A/H
In vitro	H703A/K	H689A/K
Characterization	*H. Sapiens*	*M. Musculus*
MDS	C576W	C571W
MDS	Y646C	Y641C
MDS	R690C	R685C
MDS	Y731D	Y726D

Table 5: List of putative Ezh2 inactivating mutations.

Putative EZH2 inactivating mutations identified by *in vitro* studies or found in MDS patients.

To select the most promising mutations I performed a preliminary transfection-based screening. I mutated the residues of interest into the human *EZH2* sequence cloned into a pcr8 entry vector by mutagenesis PCR. Then, a Gateway (GW) reaction was performed in order to express the different mutants of EZH2 as flagged proteins in a suitable vector containing pCAG promoter and a puromycin resistance gene. The presence of the desired mutations was confirmed by Sanger sequencing. A total of 12 constructs expressing 12 different *EZH2* mutants was generated (Table 5). ΔEzh2#1_Ezh1KO cells were transiently transfected with these 12 vectors in addition to a wt *EZH2*-expressing

plasmid used as rescue control. After 24 hours of puromycin selection, cells were collected and total protein lysates were obtained. The experiment was performed in a background where both EZH1 and EZH2 activities are impaired. The capability of mutant forms of EZH2 to restore H3K27me3 levels in ΔEzh2#1_Ezh1KO cells is shown in Figures 20A-C. All the EZH2 mutant proteins were expressed at levels almost comparable to wt EZH2 with the exception of C576W, as assessed by Ezh2 and flag immunoblots. C593Y and H694A-EZH2 were completely able to restore H3K27me3, at levels comparable to wt condition (wild type cells or ΔEzh2#1_Ezh1KO cells where WT hEZH2 was reintroduced) whereas G630S and R690A-EZH2 proteins led only to a modest restoration of H3K27me3. Surprisingly, H694K and R732K substitutions caused an hyper-activation of EZH2 activity as confirmed by the enhanced levels in the deposition of H3K27me3. Finally, reintroduction of R690C/H, Y731D/F and Y646C–EZH2 did not restore H3K27me3 levels thus confirming the inactivating nature of these aminoacidic substitutions (Fig. 20A-C). H3K27me2 levels, on the same protein lysates, were also evaluated. As shown in Figures 20D-E, the majority of mutant-EZH2 (G630S, R690A/C, C593Y, H694A/K, R732K) was still able to perform H3K27me2 whose levels were indeed comparable to wt condition. R690H and Y731F-EZH2 were able to deposit some H3K27me2 mark although with less efficiency than wt EZH2. Notably, Y731D and Y646C substitutions completely impaired EZH2 activity towards H3K27me2 (Fig. 20D-E).
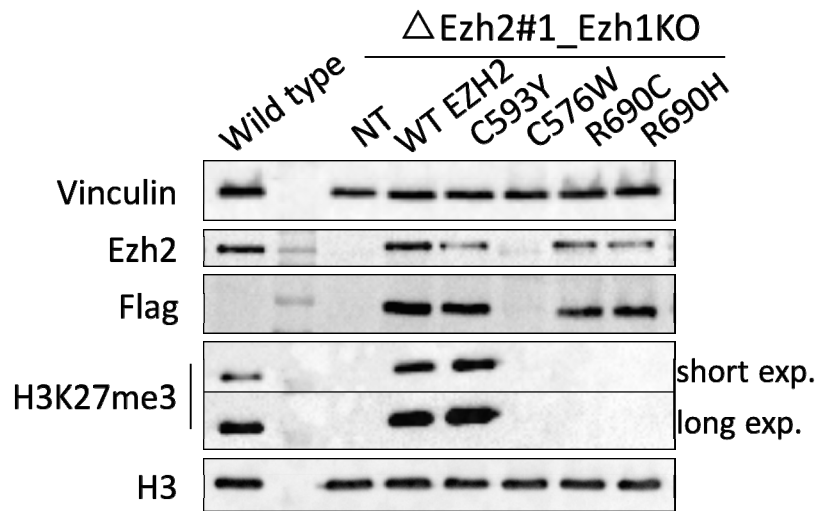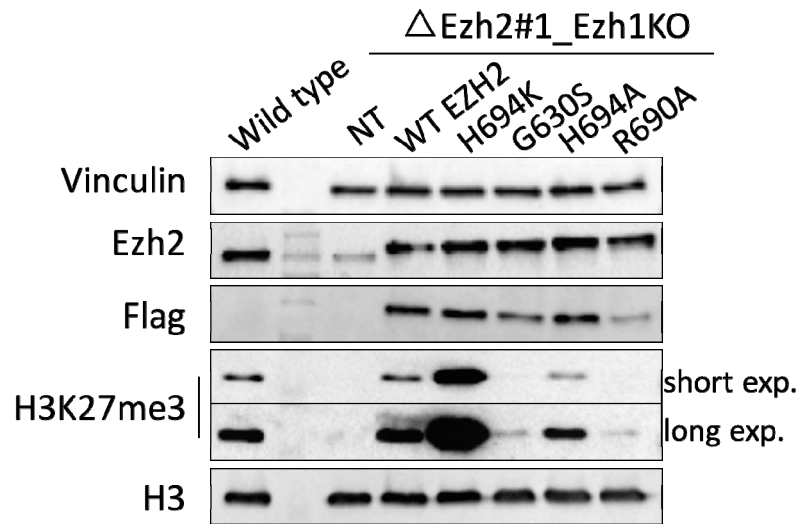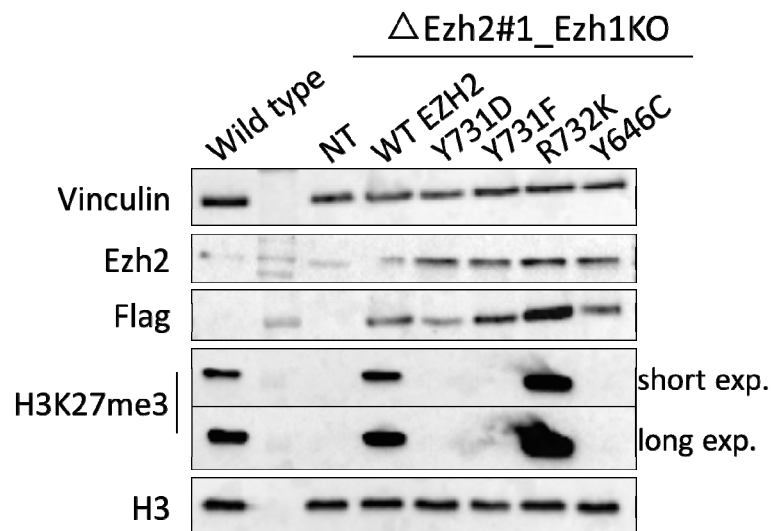
D
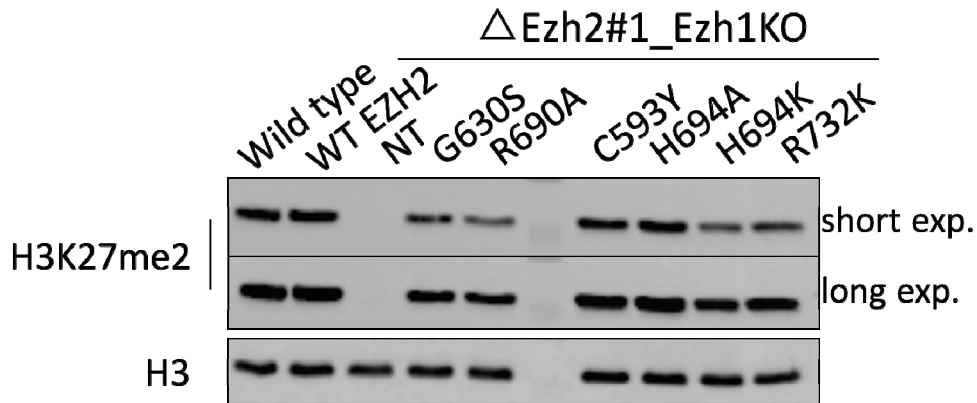
E
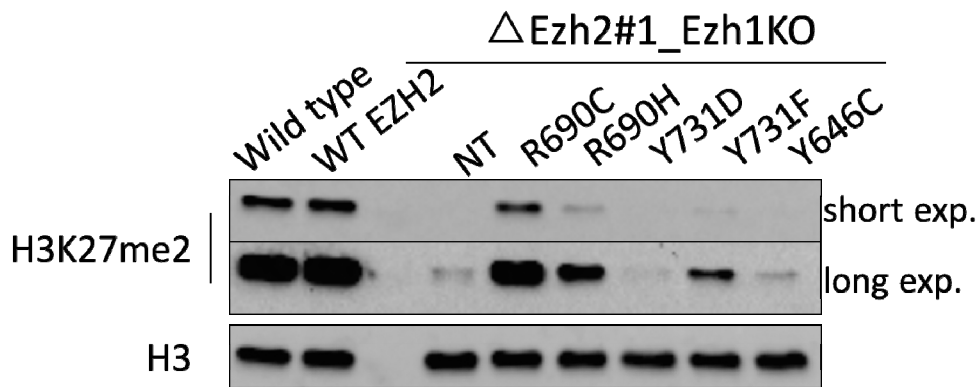
Figure 20: Screening of different mutant forms of EZH2 in ΔEzh2_Ezh1KO#1 mESCs.

Immunoblot analyses using the indicated antibodies on whole protein extracts obtained from untransfected wild type and ΔEzh2_Ezh1KO#1 mESCs or from ΔEzh2_Ezh1KO#1 mESCs upon transient transfection with WT hEZH2 or the indicated mutated forms of hEZH2. In A-C the H3K27me3 recovery capability of the indicated hEZH2 mutants has been tested whereas in D-E the H3K27me2 recovery capability has been evaluated. Vinculin and H3 served as loading controls.

Taking all these results into consideration, I decided to focus my attention on two mutants: R690C and Y731D-EZH2. Both displayed H3K27me3 impaired activity but only EZH2 Y726D resulted also unable to deposit H3K27me2. A more complete characterization of these mutants is presented in Figure 21A were the same lysates from

ΔEzh2#1_Ezh1KO transfected cells were subjected to further WB analysis. To exclude possible antibody-linearity problems, undiluted lysates as well as 1:2 and 1:4 dilutions were loaded. As already showed in Figure 20, H3K27me3 deposition was severely impaired in both conditions, with a little residual signal visible at high exposure in the R690C-EZH2 condition. Regarding the H3K27me2 levels, EZH2 Y731D was strongly impaired, whereas EZH2 R690C was not. The same effect was evident also for H3K27me1. EZH2 Y726D reintroduction was characterized by an H3K27methylaton status comparable to the one of ΔEzh2#1_Ezh1KO untransfected cells. In fact, H3K27me3 as well as H3K27me2 deposition were abolished concomitant with an increase in H3K27ac levels whereas, some residual H3K27me1 was maintained (Fig. 21A). The EZH2 re-expression experiment was performed also in a different cellular model. R690C, Y731D as well as wt EZH2-expressing plasmids were transiently transfected in N-terΔEzh2_Ezh1KO#1 cells. The results are shown in Figure 21B. Protein lysates derived from ΔEzh2#1_Ezh1KO transfected cells were analyzed by WB. ΔEzh2#1_Ezh1KO untransfected or transfected with an empty vector were used as negative controls whereas ΔEzh2#1_Ezh1KO in which wt EZH2 was reintroduced served as positive control of the experiment. Wt as well as mutant forms of EZH2 were expressed at comparable levels, as shown by Ezh2 or flag immunoblots. WB analysis confirmed that EZH2 Y726D was strongly impaired in performing all the three methylation states of H3K27 whereas EZH2 R690C was compromised only in its H3K27me3 activity (Figure 21B). Notably, both the selected mutations are of clinical relevance since they have been identified in MDS patients' samples.

A

△Ezh2#1_Ezh1KO

Wild type · Wild type 1:2 · NT · R690C · R690C 1:2 · R690C 1:4 · Y731D · Y731D 1:2 · Y731D 1:4

H3K27me3		short exp.
		long exp.
H3K27me2		short exp.
		long exp.
H3K27me1		short exp.
		long exp.
H3K27ac		
H3		

B

N-ter△Ezh2_Ezh1KO#1

NT · EMPTY · WT EZH2 · R690C · Y731D

- Vinculin
- Flag
- Ezh2
- Oct3/4
- H3K27me3
- H3K27me2
- H3K27me1
- H3K27ac
- H3
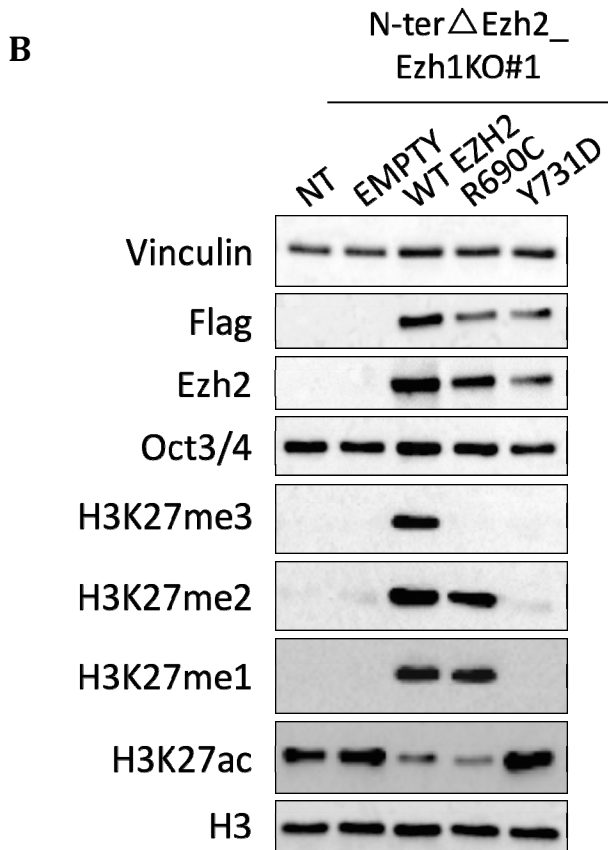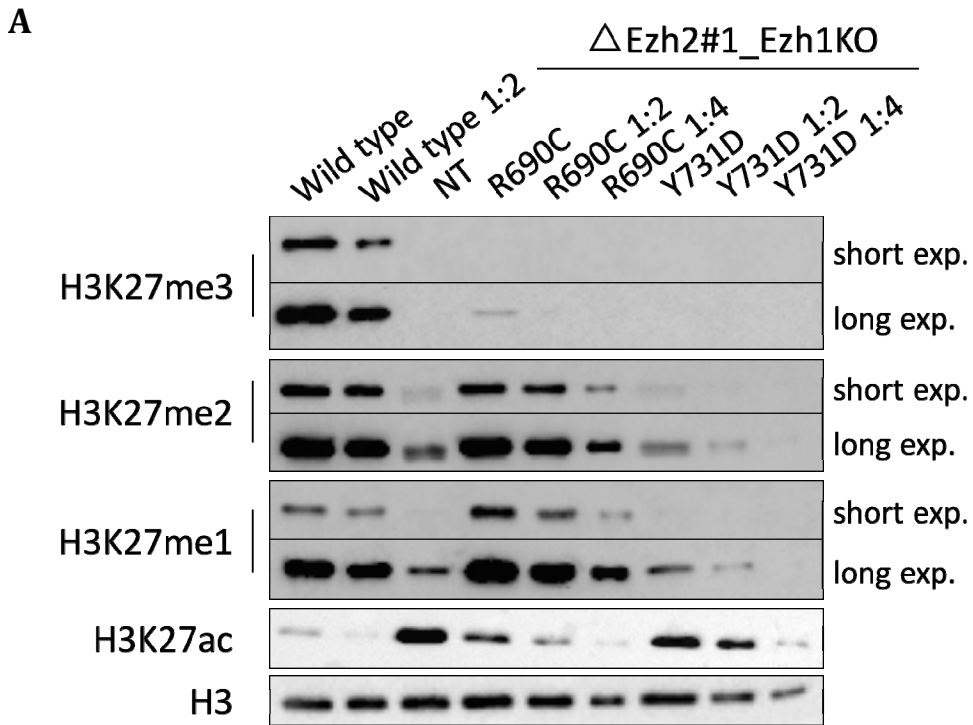
Figure 21: R690C and Y731D aminoacidic substitutions impair hEZH2-H3K27me3 activity.

116

A: Immunoblot analysis using the indicated antibodies on whole protein extracts obtained from untransfected wild type and ΔEzh2_Ezh1KO#1 cells or from ΔEzh2_Ezh1KO#1 upon transient transfection with WT hEZH2 or R690C/Y731D-hEZH2. Protein extracts have been diluted as indicated to test antibodies linearity. H3 served as loading controls. B: Immunoblots analysis using the indicated antibodies on whole protein extracts obtained from untransfected N-terΔEzh2_Ezh1KO#1 cells or from N-terΔEzh2_Ezh1KO#1 upon transient transfection with empty, WT hEZH2 or-hEZH2 R690C/Y731D expressing vectors. Vinculin and H3 served as loading controls.

4.4.2 Generation of mESC models for inactive EZH2 through CRISPR/Cas9 approach

CRISPR/Cas9 approach was used to introduce these mutations within mouse *Ezh2* sequence in WT_Ezh1KO cells. An *Ezh1*-null background was chosen to avoid any possible compensation effect from EZH1 thus allowing the elucidation of the precise effects of these mutations on EZH2 activity. Once the corresponding sequences to mutate in mouse *Ezh2* were identified (Table 5), RNA guides, one for each targeting, were designed using Feng Zhang lab's Target Finder. In both cases, a classical single-guide approach was used to introduce a specific DSB adjacent to the sequence to mutate. sgRNAs were cloned into px459 plasmid and transfected in WT_Ezh1KO cells in combination with a properly designed ssODN to obtain the introduction of the desired mutations through HDR. After 24 hours from transfection, performed with lipofectamine 2000 reagent as described in Materials and Methods section, cells were subjected to puromycin selection. After 24 hours, resistant cells were subjected to dilution cloning until the growth of visible clones. After expansion, clones were screened for the presence of the desired mutations. The screening strategy, for both mutants, was based on PCR amplification followed by restriction digestion analysis. Template ssODNs were in fact

designed to introduce, in addition to the mutations of interest, also one new restriction site. The introduction of the desired mutations was then confirmed by Sanger sequencing of the PCR products after TA-cloning in pcr8 vector. For Y726D substitution (the mouse equivalent of human Y731D) a 200 bp template ssODN with 71/72 bp homology arms was designed. In addition to the desired substitution (Y726D) obtained by TAC/GAC mutation, also AGG PAM was disrupted by AAG/AAA substitution. Moreover, a BamHI site was introduced (GGATCC). Specific primers were designed to amplify a 471 bp region encompassing the mutated sites. After successful HDR, this PCR product could be digested with BamHI enzyme thus producing two fragments, 242 and 229 bp long, respectively (panel A of Fig. 22). R685C substitution (the mouse equivalent of human R690C) was introduced using a 200 bp with 70/89 bp homology arms in which the desired mutation was obtained by CGT/TGT mutation. AGG PAM was disrupted by converting AAG in AAA. However, another AGG PAM was still present. For this reason some following residues were also silently mutated: AAC into AAT and AAA into AAG. Furthermore, an EcoRI site was introduced (GAATTC). A 557 bp PCR product is cut by EcoRI into 2 fragments, 276 and 281 bp long, when the Cas9-induced DSB is properly repaired by HDR (panel A of Fig. 23). Among more than 100 clones screened for each kind of mutation, 2 Y726D-EZH2 and 1 R685C-EZH2 expressing clones were identified. These clones were named Y726D_Ezh1KO#1, Y726D_Ezh1KO#2 and R685C_Ezh1KO#1, respectively. Both Y726D_Ezh1KO clones appeared heterozygous from PCR BamHI digestion, since an uncut band was present after PCR digestion (Fig. 22A). Sanger sequencing of the PCR products, after cloning into pcr8 entry vector (Invitrogen), confirmed that these two clones are indeed heterozygous for BamHI site, due to an incomplete homologous recombination, but both harbor homozygous Y726D substitution (Fig. 22B). As shown in Figure 23B, R685C_Ezh1KO#1 clone sequence was

118

indeed homozygous for all the desired mutations.

A   **B**
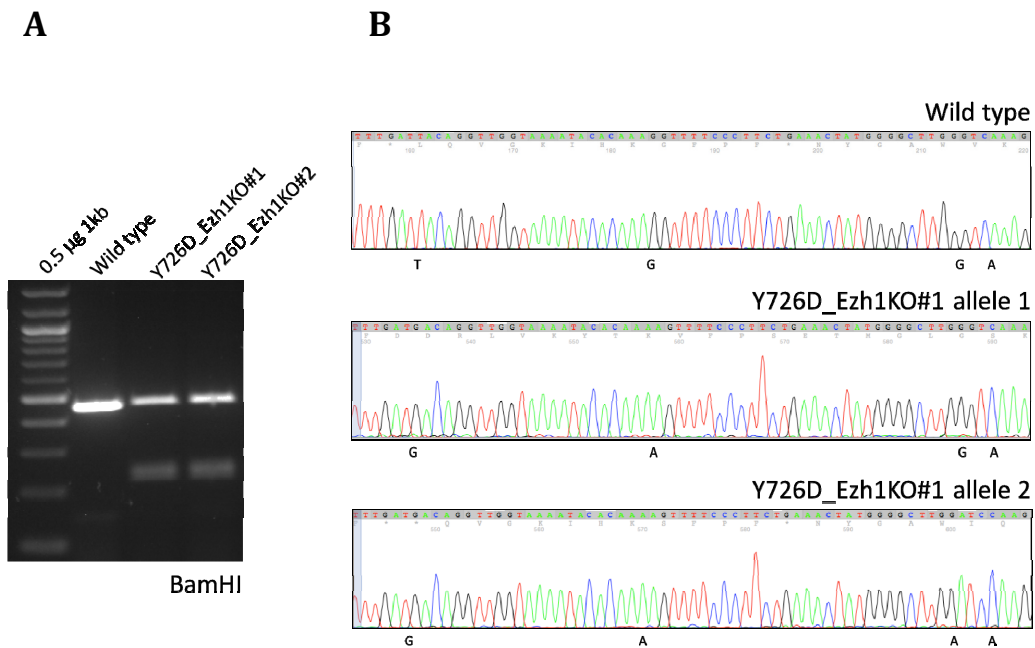
Figure 22: Screening strategy to identify mESCs harboring EZH2 Y726D aminoacidic substitution.

A: Restriction analysis on specific PCR products obtained from amplification of gDNA extracted from wild type, Y726D_Ezh1KO#1 and Y726D_Ezh1KO#2 mESCs. BamHI enzyme was used for PCR products digestion. B: Sanger sequencing analysis on specific PCR products obtained from amplification of gDNA extracted from wild type or Y726D_Ezh1KO#1 mESCs. Sequences of both alleles from Y726D_Ezh1KO#1 cells have been compared to wild type sequence. Introduced mutations are indicated.

A   **B**
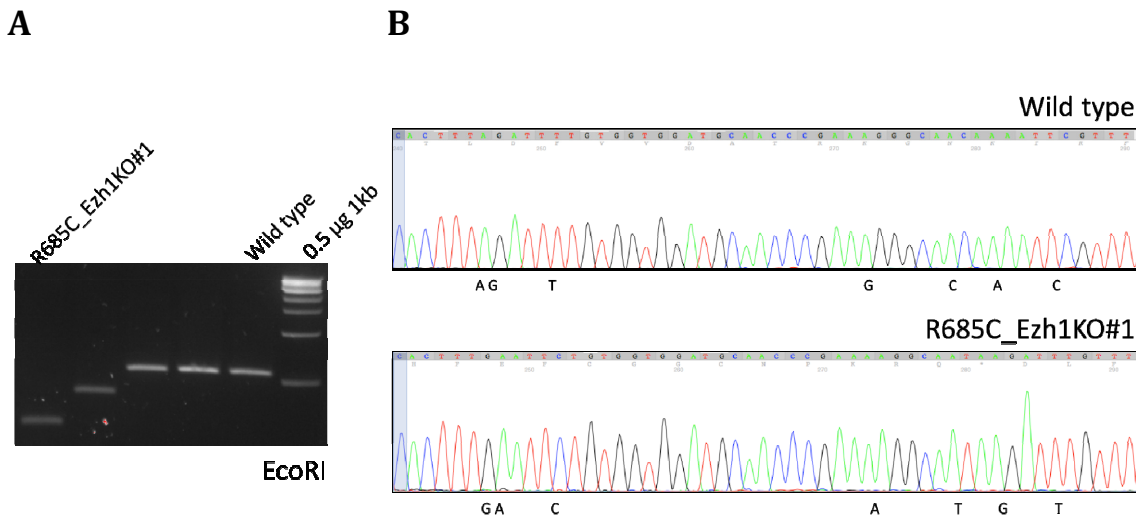
Figure 23: Screening strategy to identify mESCs harboring EZH2 R685C aminoacidic substitution

A: Restriction analysis on specific PCR products obtained from amplification of gDNA extracted from wild type, R685C_Ezh1KO#1 mESCs. EcoRI enzyme was used for PCR products digestion. B: Sanger sequencing analysis on specific PCR products obtained from amplification of gDNA extracted from wild type or R685C_Ezh1KO#1 mESCs. Sequences from wild type and R685C_Ezh1KO#1 mESCs have been compared. Introduced mutations are indicated.

4.4.3 Characterization of EZH2 R685C and Y726 expressing mESCs

Total protein lysates from Y726D_Ezh1KO#1, Y726D_Ezh1KO#2 and R685C_Ezh1KO#1 cells were analyzed by Western Blot as shown by Figure 24A. Wild type mESC and WB_Ezh1KO lysates were used as positive controls, whereas ΔEzh2#1, ΔEzh2_Ezh1KO#1, N-terΔEzh2#1and N-terΔEzh2_Ezh1KO#1 as negative ones. First of all, the presence of mutated (Y726D or R685C) EZH2 forms did not affect the pluripotency state of cells, as shown by Oct3/4 immunoblot. ΔSET and N-ter Δ deleted-EZH2 expressing cells characterization was already presented in the previous sections (Fig. 4 and 16). Briefly, ΔSET deletion led to a slightly smaller form of EZH2 that was less expressed compared to the wt; N-ter deletion resulted in an even smaller form of EZH2

that was indeed much less expressed compared to the previous deleted mutant. Both deletions completely impaired EZH2 ability to perform H3K27me3 and partially compromised also H3K27me2 deposition. The residual H3K27me2 and H3K27me1 marks, whose levels were unaffected by both deletions, were abolished after *Ezh1* knock-out. EZH2 Y726D resulted expressed in both clones to comparable levels with respect to wt protein, whereas R685C mutation highly destabilized EZH2 protein that resulted almost not expressed (Fig. 24A). The deposition of all the three methylation states of H3K27 was indeed abolished in EZH2 Y726D expressing cells concomitant with an increase of H3K27ac levels. Notably, H2AK119ub levels were also slightly decreased in presence of EZH2 Y726D/R685C. The absence of H3K27 methylation observed in EZH2 R685C expressing cells was likely due to the lack of EZH2 protein expression in an *Ezh1*-null background (Fig. 24A). From this first evidence EZH2 Y726D expressing cells seem actually a suitable model to study EZH2 inactivating mutations.

A
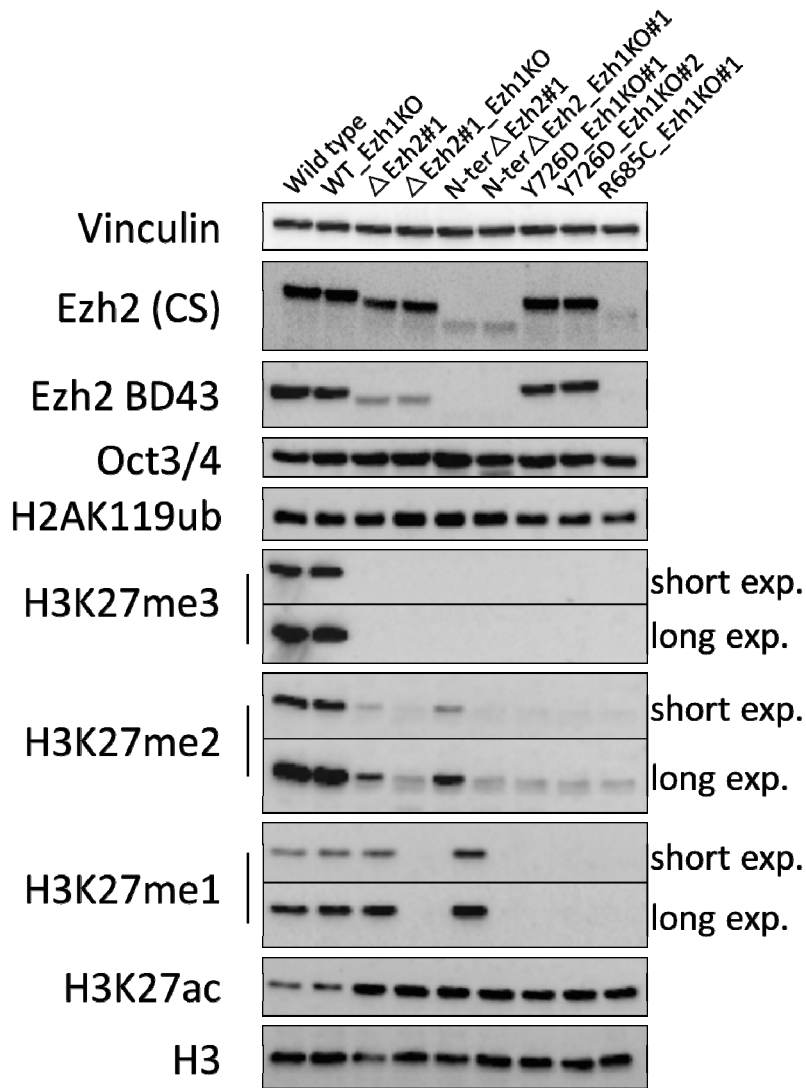
B
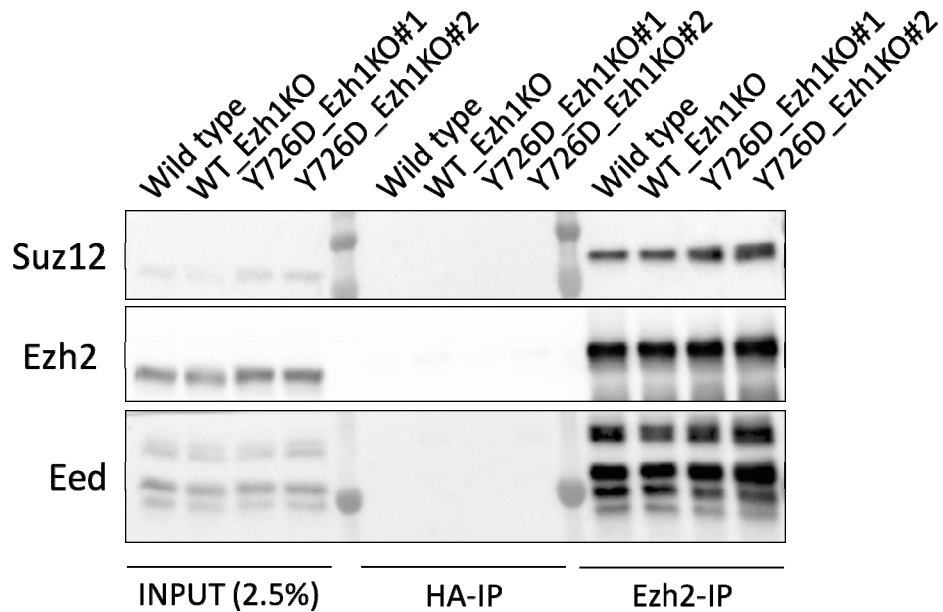
Figure 24: Y726D aminoacidic substitution impairs EZH2 H3K27 methylation activity.

A: Immunoblot analysis using the indicated antibodies on whole protein extracts obtained from the reported mESC lines. Vinculin and H3 were used as loading controls. B: Immunoblot analysis with the indicated antibodies on INPUTS (2.5%) and immunoprecipitated proteins from total wild type, WT_Ezh1KO, Y726D_Ezh1KO#1 and Y726D_Ezh1KO#2 mESCs lysates with Ezh2 (AC22)-crosslinked Protein A sepharose beads. HA- crosslinked Protein A sepharose beads were used as unrelated IP control.

In order to verify whether EZH2 Y726D protein is still able to form PRC2 complex by association with SUZ12 and EED subunits, immunoprecipitation with PA Sepharose beads crosslinked with anti-Ezh2 antibody was performed on total protein extracts of wt, WT_Ezh1KO, Y726D_Ezh1KO#1 and Y726D_Ezh1KO#2 cells. As you can see from WB analysis in Figure 24B, EZH2 Y726D was able to complex with the crucial SUZ12 and EED subunits in the same way the wt protein did. This result suggests that the lack of H3K27 methylation activity of EZH2 Y726D is not due to its inability to form PRC2 complex. At this point we wanted to evaluate how this absence of H3K27 methylation affects mESC

differentiation. So, an EBs formation assay was performed on wt, WT_Ezh1KO, N-terΔEzh2_Ezh1KO#1, Y726D_Ezh1KO#1 and R685C_Ezh1KO#1. As shown in Figure 25A, EBs formation was strongly impaired in N-terΔEzh2_Ezh1KO#1 cells as already showed in Figure 17 and also in Y726D_Ezh1KO#1 and R685C_Ezh1KO#1 cells. Indeed, Y726D_Ezh1KO#1 and R685C_Ezh1KO#1- derived EBs were irregular and very small. Moreover, the Y726D_Ezh1KO#1-derived EBs phenotype seemed to be even worse compared to N-terΔEzh2_Ezh1KO#1 and R685C_Ezh1KO#1-derived ones. R685C_Ezh1KO#1-derived EBs behave in a N-terΔEzh2_Ezh1KO#1 similar way in terms of differentiation markers expression (Fig. 25B). Y726D_Ezh1KO#1-derived EBs showed a worse defect also in terms of pluripotency/differentiation markers expression. Indeed, they did not completely shut down pluripotency markers expression and they failed to activate mesoderm and ectoderm-specific markers expression (Fig. 25B). Even if Y726D_Ezh1KO#1-derived EBs defective phenotype seems to be more pronounced compared to N-terΔEzh2_Ezh1KO#1 (Fig. 25 and data not shown of replicated experiment), these results support our hypothesis that this defective differentiation phenotype is indeed H3K27me3-dependent.

A

Wild type   WT_Ezh1KO   N-ter△Ezh2_Ezh1KO#1   Y726D_Ezh1KO#1   R685C_Ezh1KO#1

d3   d5   d7

B

Pluripotency

Oct4   Nanog

Wild type
WT_Ezh1KO
N-terΔEzh2_Ezh1KO#1
Y726D_Ezh1KO#1
R685C_Ezh1KO#1

Mesoderm

Pax3

Ectoderm

Mausashi   Nestin
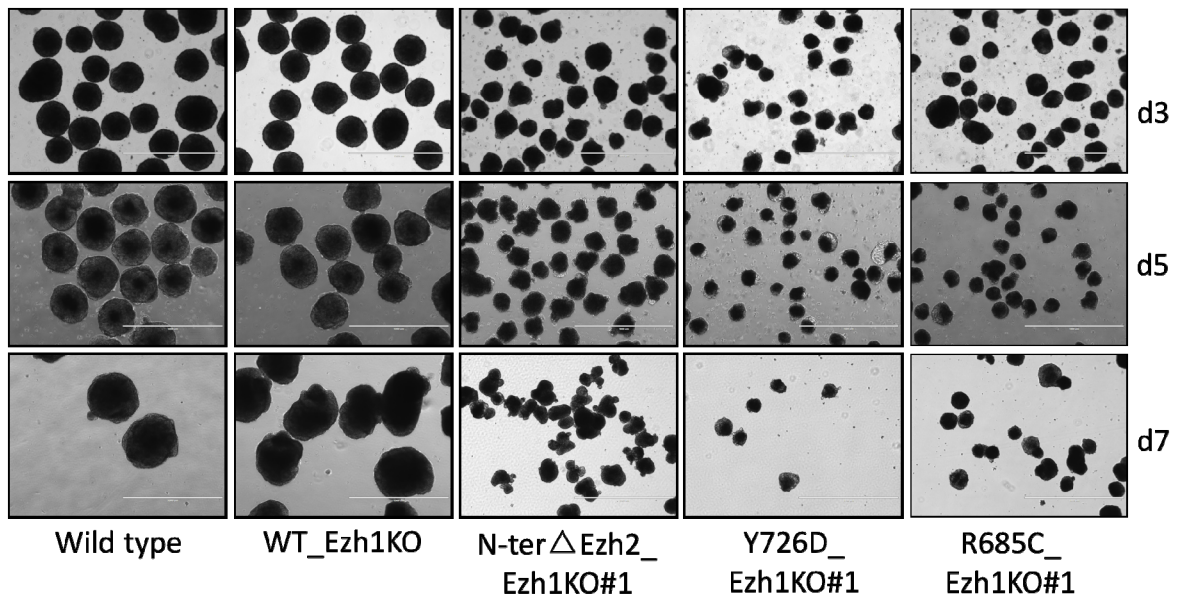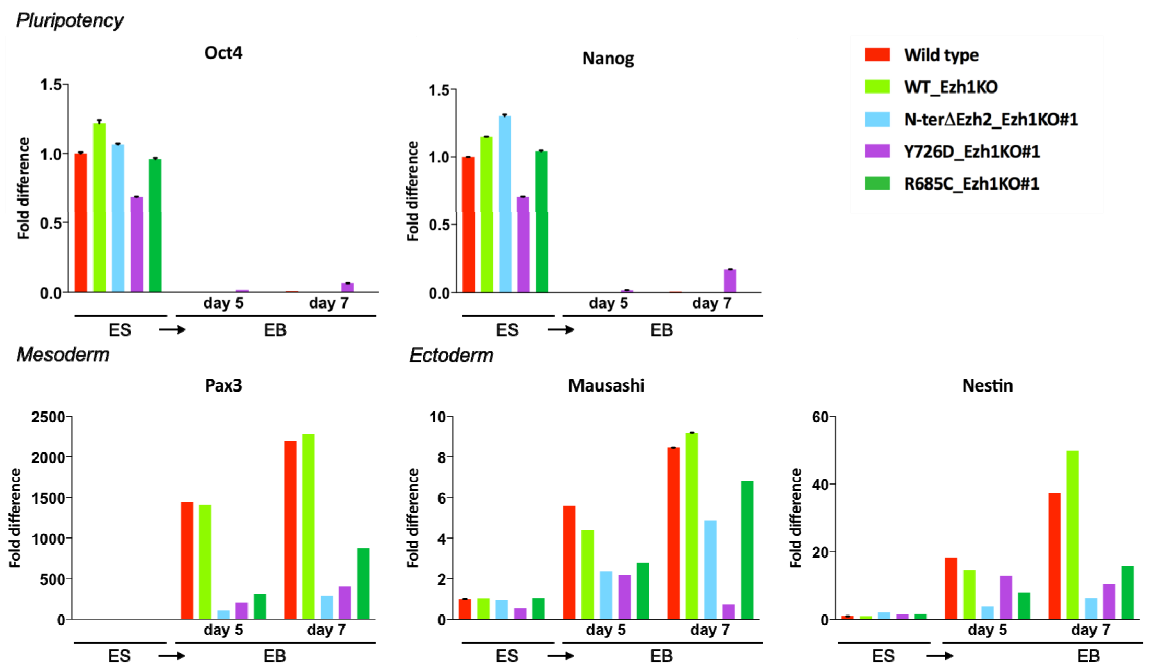
Figure 25: Y726D aminoacidic substitution severely impairs mESCs differentiation into EBs.

A: Pictures at day 3, day 5 and day 9 after differentiation of EBs derived from wild type, WT_Ezh1KO, N-terΔEzh2_Ezh1KO#1, Y726D_Ezh1KO#1 and R685C_Ezh1KO#1 mESCs. B: Relative expression of the

indicated pluripotency/differentiation markes determined in wild type, WT_Ezh1KO, N-terΔEzh2_Ezh1KO#1, Y726D_Ezh1KO#1 and R685C_Ezh1KO#1 mESCs before (ES) and after 9 days of differentiation (EB). Gapdh served as normalizing expression control.

To understand how this huge H3K27 methylation loss impacts on PRC2 and PRC1 binding to chromatin, SUZ12 and RING1B ChIP-qPCRs at classical Polycomb targets were performed on WT_Ezh1KO, Y726D_Ezh1KO#1 and Y726D_Ezh1KO#2 cells. Moreover, to investigate the impact on "canonical-PRC1" HoxA11 and Foxb1 regions were also included in the analyses. As depicted in panels A and B of Figure 26, both SUZ12 and RING1B resulted displaced, despite non-completely, from all the analyzed regions in EZH2 Y726D expressing cells. Even if more deeper investigations are required, these results suggest that PRC2 and, to less extent, PRC1 binding to chromatin is possible also in absence of H3K27me3. Moreover, even if a slightly decrease in H2AK119ub was observed by WB analysis (Fig. 24A), ChIP-qPCR on classical Polycomb targets did not reveal differential deposition of the mark in EZH2 Y726D expressing cells compared to wt (Fig. 26C).

126

A

B

C
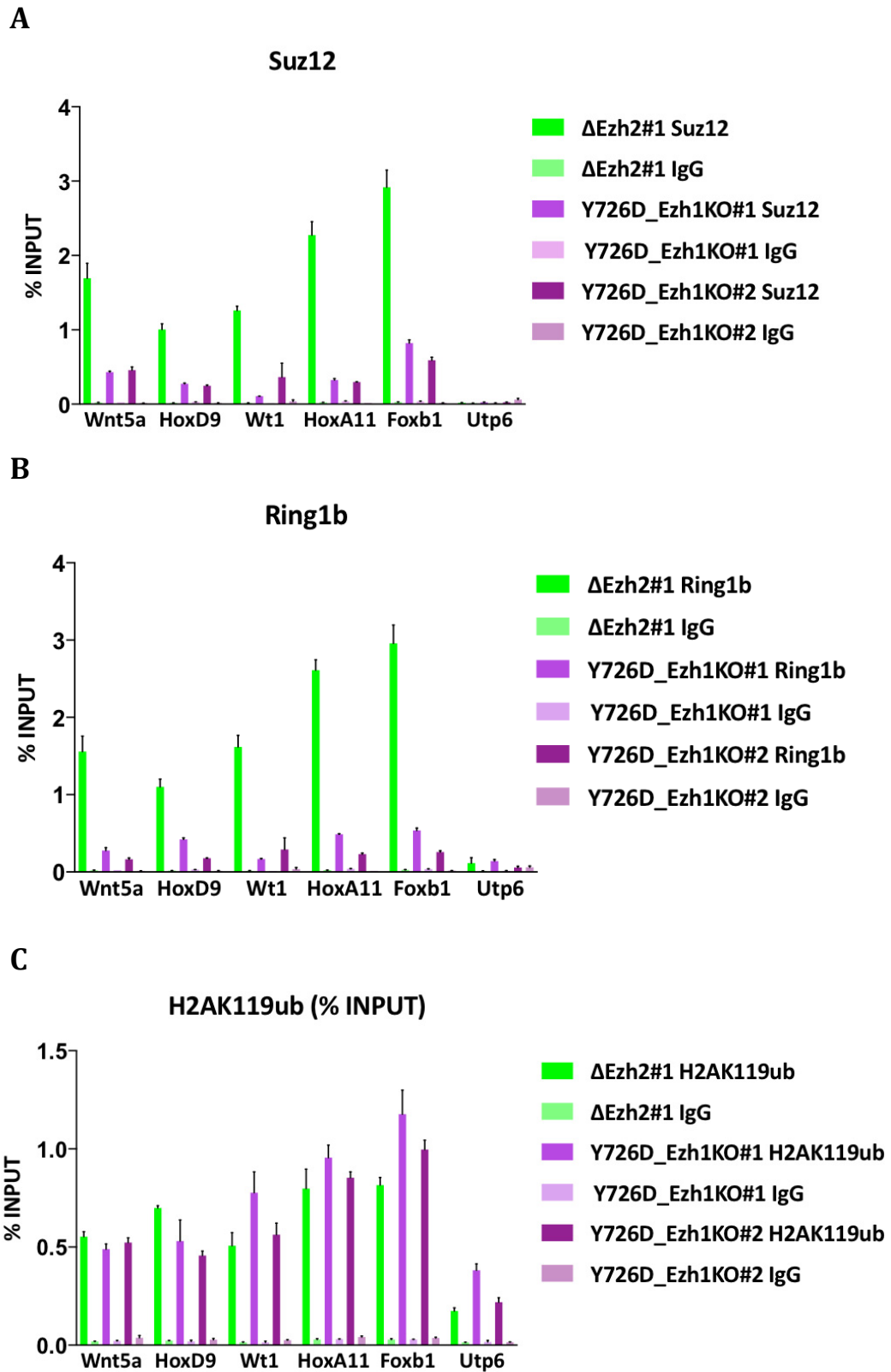
Figure 26: SUZ12 and RING1B are significantly but not completely displaced from typical Polycomb targets in Y726D_Ezh1KO#1 mESCs.

ChIP-qPCR analyses in WT_Ezh1KO, Y726D_Ezh1KO#1 and Y726D_Ezh1KO#2 mESCs performed with anti-Suz12 (A), anti-Ring1b (B) and anti-H2AK119ub (C) antibodies at the indicated loci. Suz12, Ring1b and H2AK119ub ChIP enrichments are normalized to input. ChIPs with rabbit IgG were performed as negative control. Wnt5a, HoxD9, Wt, HoxA11 and Foxb1 were used as typical Polycomb targets whereas Utp6 served as negative control region.

4.5 Generation of H3.3 K27M mutant mESCs

As reported in the introduction EZH2 activity can be impaired also by mutations affecting its substrate. Indeed, CRISPR/Cas9 technology was used to obtain mESC expressing K27M H3.3. For the targeting a single guide approach was used to induce a DSB nearby the site to mutate within *H3f3a* gene and HDR was obtained by co-transfection of a properly designed template ssODN. All the reagents are described in Materials and Methods section. Briefly, the screening was based on PCR digestion analysis with ScaI enzyme, whose site was added into the ssODN. The introduction of the mutation of interest was confirmed by Sanger sequencing of specific PCR products. Among more than 100 screened clones, 2 positive clones for the mutation were identified and named H3.3K27M#1 and H3.3K27M#2. Sanger sequencing revealed indeed the homozygous nature of the introduced mutation in both clones. Moreover, H3.3K27M expression, in the selected clones, was evaluated also by WB analysis performed with a specific K27M antibody on nuclear protein extracts (Fig. 27B). Western Blot analysis confirmed the inactivating nature of the mutation also in my cellular models (Fig. 27A). The presence of mutant H3.3 did not impact pluripotency features of mESC, in fact Oct3/4 and NANOG expression levels were not affected as well as EZH2. H3K27me3 deposition was completely abolished in H3.3K27M-expressing cells and H3K27me2 resulted strongly reduced, even if not completely. Interestingly, H3K27me1

deposition appeared slightly increased in presence of H3.3K27M. These cells represent another complementary useful model for the study of EZH2 activity.

A　　　　　　　　　　　　**B**
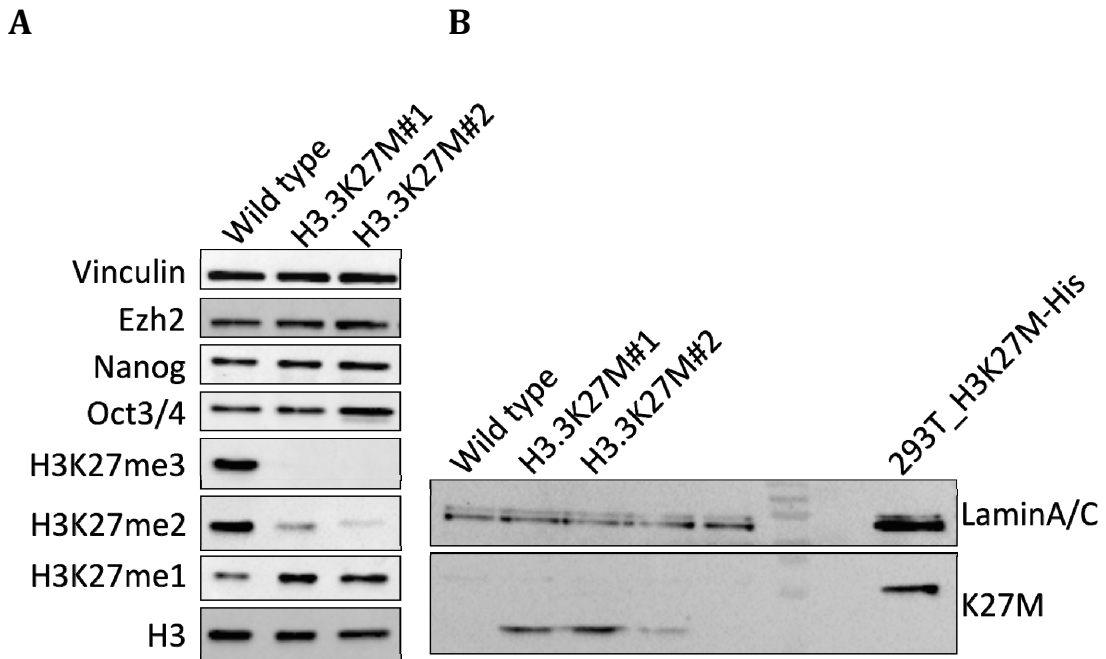
Figure 27: K27M H3.3 expression impairs PRC2 H3K27me2 and H3K27me3 activities.

A: Immunoblot analysis using the indicated antibodies on whole protein extracts obtained from wild type or K27M H3.3- expressing mESCs (H3.3K27M #1 and H3.3K27M#2). Vinculin and H3 were used as loading controls. B: Immunoblot analysis using the indicated antibodies on nuclear protein extracts obtained from wild type or K27M H3.3- expressing mESCs (H3.3K27M #1 and H3.3K27M#2). 293T cells stably expressing His-tagged K27M H3.3 were used as positive control. LaminA/C was used as loading control.

5. Future Perspectives

Taken together, these novel isogenic genome-edited mESC models represent unprecedented tools that will allow me to investigate and get insight into several crucial aspects of PRCs biochemistry and activity. Their characterization at a molecular level will not only increase the understanding about the mechanisms regulating PcGs activity during development and differentiation but will also be fundamental for the dissection of the molecular mechanism behind EZH2 pathological mutations in human cancer.

In the close future, using these models, it will be possible to address several unsolved issues related to the activity of Polycomb proteins. This will involve:

1 The interplay between EZH1 and EZH2 activities and the compensation effects on H3K27 methylation in normal and pathological contexts, both at basal state and during differentiation.

2 The direct role of H3K27 methylation in controlling PRCs recruitment and stability to their genomic targets focusing the attention on both canonical and non-canonical complexes.

3 The molecular and biological role of EZH2 hyper-activating and inactivating mutations as well as H3.3-affecting mutations, identified in human cancer.

The specific contribution of the different H3K27 methylation states in terms of transcriptional and biological outcome will be investigated coupled to differentiation assays by genome-wide expression (RNA-seq) and location analyses (ChIP-seq). This will allow to map chromatin association of Polycomb components and specific deposition of histone modifications.

Overall, this will allow me to dissect the mechanisms by which these different mutations alter the transcriptional proprieties of the PRC2 complex to drive tumour development in distinct cellular contexts.

6. Discussion

Polycomb group of proteins (PcGs) are well known essential regulators of cellular homeostasis and cell fate transitions. Due to this crucial regulatory role, their involvement in cancer development is not surprising. Members of both PRC2 and PRC1, the two major PcGs complexes, are frequently found deregulated in many different kinds of human cancers, including both solid tumors and hematopoietic malignancies[229]. EZH2, the catalytic subunit of PRC2, has been widely implicated in cancer onset and progression. It is frequently over-expressed in several cancers including prostate and breast cancer and it has been also correlated to disease stage and poor prognosis[155,230]. Despite the impressive amount of disease-association data, the role of EZH2 in cancer is still poorly characterized. Indeed, the extent to which different kind of tumors actually depend on EZH2 deregulated levels and/or activity is still unclear. Moreover, the role of catalytic or non-catalytic functions in both physiological and pathological contexts is even less understood. The dissection of the tumorigenic role of EZH2 has become even more complicated once both hyper-activating and inactivating mutations affecting its catalytic activity have been identified. Hyper-activating mutations have been recently identified in a consistent percentage of DLBC-lymphomas (21%) and follicular lymphomas (7%). These mutations result in aminoacidic substitutions, affecting critical residues within EZH2 SET domain (Y641 and less frequently A677 and A687) thus leading to the hyper-activation of the enzyme towards H3K27me3 deposition[156-158]. Moreover, the identification of Y641 mutations in a solid tumor-context, specifically in melanoma[171] has raised new questions regarding the oncogenic potential of these mutations that were classically considered as "hematologically-restricted". The most peculiar feature of these mutations is that they occur in heterozygosity, with the wt form

of the enzyme being considered essential for the mutant form to exert its oncogenic activity. However, recently Souroullas and collaborators showed that the wt *Ezh2* allele is dispensable for EZH2 Y641F oncogenic role. The presence of these mutations lead to increased levels of the repressive H3K27me3 mark, concomitant with a decrease in H3K27me2 mark[169]. Despite this evidence, very little is known about the molecular mechanisms underlying these mutations and the biological consequences of H3K27me3 accumulation. Even less characterized are *EZH2* inactivating mutations, that have been described in several kinds of myelodysplastic syndromes[159-161]. Except for few *in vitro* studies, very little is known about the oncogenic role of mutations that negatively affect EZH2 enzymatic activity. Notably, many evidences suggest a clear cell-type specific tumorigenic role of *EZH2* that can act as oncogene or as oncosuppressor depending on the context. A deeper clarification of the tumorigenic proprieties of *EZH2*-affecting mutations is therefore necessary to understand the underlying molecular mechanisms in order to develop more targeted and efficient therapeutic strategies.

CRISPR/Cas9 as a powerful genome-editing tool in mESCs.

To do so, it is mandatory to generate suitable cellular and animal models in order to better characterize the role of Ezh2 harboring hyper-activating and inactivating mutations. To achieve this essential goal, I took advantage of mESCs that represent a simple cellular model in terms of culturing and handling. In addition to this, the function and role of Polycomb proteins has been widely investigated and characterized in mESCs and a number of different well-established biochemical assays and differentiation protocols are available. To precisely edit mESCs genome in order to knock out *Ezh2/Ezh1* genes or to introduce specific mutations in *Ezh2*, I took advantage of the

CRISPR/Cas9 approach. CRISPR/Cas9 technology has emerged as a highly powerful and versatile genome-editing tool that has radically revolutionized biology research field. Beyond classical knock-outs and knock-in solutions it allows many different and intriguing applications including the easy and rapid generation of animal models and the possibility to perform specific transcriptional regulation by means of a catalytically inactive form of Cas9 enzyme[186]. Nowadays it represents indeed an indispensable tool to precisely modify and interrogate the genome and is suitable to several different research fields, including Polycomb world.

In this project the introduction of specific point mutations was achieved by co-transfection of Cas9 (orCas9n)/sgRNA-expression plasmids in combination with a ssODN to obtain HDR into mESCs. Globally, despite the experimental approach was very efficient in term of indels introduction, I noticed that HDR-based CRISPR/Cas9 editing efficiency was highly gene/locus-specific. For example, among more than 100 clones screened in my EZH2 Y641N-experiment, two homozygous clones were identified. Indeed, I was not able to find any heterozygous clone for Y641N substitution. This was not surprising since it is well known that many different variables may affect HDR efficiency, including some out of experimental control. In my hands, a high percentage of clones that displayed the correct mutation on one allele, presented indels on the other. It was already known that HDR can be frequently corrupted by undesired indels. Paquet and collaborators[205] reported for the first time the generation of iPSC harboring heterozygous (in addition to homozygous) dominant early onset Alzheimer's disease-causing mutations affecting amyloid precursor protein (*APP*) and presenilin I genes (*PSEN1*). By analyzing the mutational status of the two alleles in individual iPSC clones they were able to isolate both homozygous and heterozygous clones for the desired

mutations. In addition to this, as in my case, in heterozygous clones the "non-HDR" allele always contained indels. Recently, to overcome HDR relative inefficiency, many attempts have been done including the use of asymmetric ssODN and the inhibition of members of NHEJ pathway[200-202,204]. However, in my case the question was how to obtain monoallelic mutations. Regarding this, Paquet and collaborators[205] showed that a monotonic inverse correlation exists between rate of mutation incorporation and distance from the cleavage site. According to this finding, cut-to-mutation distance should be minimized when introducing biallelic mutations whereas the frequency of mono-allelic mutations should increase at greater distances. This so called "distance effect" has to be taken into consideration as ssODN design could indeed influence the success of an HDR-based experiment. Another strategy could be the reduction of Cas9-sgRNA or ssODN amount used for the experiments, but this would probably lead also to a reduction in terms of efficiency thus requiring much more effort in culturing and screening an enormous number of clones. An additional possibility is represented by the use of a mixture of wt and mutated-ssODNs as templates[205]. The existence of a "distance effect" was indeed corroborated by my results on EZH2 Y726D aminoacidic substitution. Following the previously described experimental workflow, I was indeed able to introduce a heterozygous mutation. The two selected clones harbored Y726D substitution on both alleles whereas were heterozygous for the restriction site introduced into the template for screening purposes. This highlighted the fact that the system was intrinsically able to introduce a heterozygous modification. Indeed, looking at the ssODN design, I noticed that the restriction site was indeed far away from the desired mutation located almost in the middle of the template. A proper design of the single stranded template could indeed modulate HDR efficiency[205]. Since hyper-activating mutations affecting *EZH2* represent a relevant clinical issue I definitely want to obtain mESC physiologically expressing

heterozygous EZH2 Y641N in order to analyze in depth its tumorigenic properties. That's why I planned to generate heterozygous EZH2 Y641N expressing cells by retargeting one clone identified during the screening procedure (Y641N_ΔG). This specific clone is heterozygous for Y641N aminoacidic substitution. In fact, one allele had correctly recombined and harbors the mutation of interest whereas the other one displayed a 1 bp deletion. By transient transfection I demonstrated that the N-modified allele was indeed functional since reintroduction of wt EZH2 led to accumulation of H3K27me3 compared to wt or non-transfected cells. I already designed a proper guide RNA to specifically target the deleted allele (PAM sequence is mutated on the other so it won't be recognized) and a wt template where the missing base is present. In this way I aim to reconstitute a functional wild type allele. The experiment is already ongoing and soon I will obtain properly modified mESCs. The availability of such a model will allow the comprehension of the molecular and transcriptional proprieties of this mutation and to determine whether it has a direct role as driving cancer mutation.

Regarding the generation of KO lines, an intrinsic strength of CRISPR/Cas9 tool in terms of indels introduction clearly emerged from my results. Nevertheless, indels introduction do not necessarily lead to gene KO. Indeed, in my hands, specifically targeting the ATG was not enough to generate a complete KO for *Ezh2*. In fact, my targeting resulted in a catalytically impaired truncated form of the protein. A careful design of the targeting strategy is therefore necessary to ensure the success of the genome-editing. For the next CRISPR/Cas9-based experiments I plan to use at least 2 sgRNAs simultaneously to induce large macro-deletions thus chopping out, if possible, the entire gene. In case of longer genes I will target the ATG and critical domains of the protein. In addition, a strategy implying subsequent targetings can also be adopted. Taken together my results

136

support the great power of CRISPR/Cas9 technology for genome-editing in mESCs. In fact, I was able to target and genome-edit different genes in a quite simple manner and with high efficiency. My idea is to take advantage of this precious tool to expand my cellular models panel. I want to re-generate independent *Ezh2* and *Ezh1* KOs, to generate other points mutants for *Ezh2* but also for *Suz12* and *Eed*[231]. The combination of all these models with the already generated ones will allow me to elucidate the role of PRC2 subunits in cancer pathogenesis and to shed light on many different aspects regarding PRC2 activity and regulation.

EZH1 may exert a context-dependent compensatory role.

Despite the fact that the targeting used strategy didn't led to the generation of mESCs carrying heterozygous EZH2 Y641N, I thought to use homozygous cells as model for studying *EZH2* inactivating mutations, since cells expressing EZH2 Y641N were characterized by a huge decrease in H3K27me2 and me3 levels. Even if EZH2 Y641N was still able to form the complex with SUZ12 and EED and to bind to chromatin, unfortunately its expression levels were affected by the mutation. I demonstrated that protein destabilization could not be attributed to a transcriptional effect neither to an enhanced proteasome-dependent degradation. Moreover, in comparison to △SET-EZH2 cells, some residual H3K27me3 was still detectable, thus suggesting incomplete loss of function features of homozygous Y641N or a possible compensation by EZH1. Intriguingly, no H3K27me3 compensation was visible in △SET-EZH2 background or at least it was under the detection limit of Western Blot analysis. Furthermore, in both EZH2 contexts (Y641N or △SET) some residual H3K27me2 activity was still detectable. It is known that EZH2 preferential substrate is represented by H3K27me 1and that more

that 70% of total pool of histone H3 in mESC is indeed dimethylated[63]. Residual H3K27me3 in presence of EZH2 Y641N can be explained by an EZH1-dependent compensation as confirmed by Western blot analysis. In fact, upon *Ezh1* KO residual H3K27me3 levels were abolished. Upon targeting of the SET domain (△SET-Ezh2), EZH2 catalytic activity was abrogated. Despite not detectable by WB analysis, the presence of residual H3K27me3 mark couldn't be excluded. In fact ChIP-seq analysis confirmed the maintenance of H3K27me3-enriched regions in △SET-EZH2 expressing cells. Moreover, residual H3K27me2, observed in both genetic background, could be explained by a compensatory mechanism ruled out by EZH1 or by the action of other novel uncharacterized di-methyltransferase activities. Interestingly, in both cases global H3K27me1 deposition was not affected suggesting its EZH2-independent regulation. Subsequent WB analysis confirmed that both residual H3K27me2 and H3K27me3 levels were indeed EZH1-dependent since they were abolished upon *Ezh1* KO. Despite the intrinsic detection limit of WB, at least regarding H3K27me3, I could conclude that the mark is completely removed from it targets in △SET-Ezh2 cells after *Ezh1* KO, as assessed by ChIP-seq analysis. In this genetic background, all residual H3K27me3 was indeed EZH1 dependent. I will verify if this is the case also in presence of EZH2 Y641N. Genome wide analysis regarding the deposition of three methylation states occurring on H3K27 in both the genetic backgrounds will clarify EZH1 role and compensatory activity in absence of an active form of EZH2 or unveil the existence of novel complementary methyltransferase activities. An intriguing scenario emerges from our results; EZH1-compensatory role seems to be indeed context-dependent. Even if data regarding H3K27me3 genome-wide distribution in EZH2 Y641N expressing cells are still missing, WB analysis suggested that more H3K27me3 mark is retained in presence of EZH2

Y641N compared to △SET-EZH2. Thus it will be important to assess the causative relation of this result to a different EZH1 role in the two contexts or if other molecular determinants could be involved.

EZH1-dependent compensation and differentiation.

mESCs differentiation into embryoid bodies was highly impaired in △SET-EZH2 and N-ter deleted EZH2 cells whereas EZH2 Y641N expressing cells-derived EB, surprisingly, did not show major defects in terms of morphology or lineage specific markers activation. I hypothesized that the residual H3K27me3 observed in presence of EZH2 Y641N was indeed enough to sustain an almost normal mESC differentiation into EBs. Interestingly, ChIP-seq analysis revealed the maintenance of several H3K27me3-enriched regions in △SET-EZH2 expressing mESC. The same scenario was suggested also by ChIP-qPCR performed on N-ter deleted EZH2 expressing clones that displayed the same differentiation defects previously seen in △SET-cells. A deeper analysis of H3K27me3 genome-wide distribution is needed to clarify whether EZH1 has actually different capabilities to sustain some H3K27me3 activity in different mutant-*Ezh2* contexts. What I can conclude from my data is that mESCs differentiation capability is indeed H3K27me3-dependent and a certain amount of residual mark is enough to sustain EZH2 Y641N expressing cells differentiation into EBs. H3K27me3 genome-wide distribution analysis in EZH2 Y641N-expressing cells will eventually confirm my hypothesis. Moreover, it will be interesting to understand which are the differential retained H3K27me3-marked regions in △SET-EZH2 and N-ter△EZH2 cells compared to EZH2 Y641N expressing cells. The clue of the different displayed phenotype could indeed reside there. In presence of EZH2 Y641N, EZH1 may still be able to correctly regulate

crucial genes for differentiation. Further experiments are needed to clarify if and how EZH1 and EZH2 may cooperate and/or sustain each other activity in different "EZH2-contexts". Despite EZH1 and EZH2 have been classically considered as mutual exclusive within PRC2[232] their possible interaction within the same complex has also been suggested[84,232]. Indeed, the existence of a dimeric PRC2 has been proposed by Davidovich and colleagues[233]. Interestingly, results from Shen and collaborators suggested a possible role of EZH2 in regulating EZH1-containing PRC2 by promoting EZH1-EED protein interactions[232]. In line with these observations, EZH1 activity may be diversely modulated by Y641N- or △SET-EZH2 depending indeed on the non-enzymatic activities of EZH2. Taken this into consideration, it could be that EZH2 Y641N retains some of this ancillary function towards EZH1 that is indeed lost or more impaired in the △SET and N-ter△EZH2 mutant cells. These observations could help me explaining, at least in part, the strikingly different phenotype observed in the EBs formation assay. Mutations abrogating EZH1-EED protein interactions could be induced in the two different EZH2 Y641N contexts in order to verify if actually EZH2 Y641N is still partially able to regulate EZH1-containing PRC2. Moreover, in order to address all these issues I am planning different experiments: i) H3K27me3 ChIP-seq on EZH2 Y641N-expressing cells before and after *Ezh1* KO to understand which genomic regions are still H3K27me3-enriched upon KO of *Ezh2* and if they get lost upon *Ezh1* KO; ii) differentiation assay in EZH2 Y641N after *Ezh1* KO to assess if EZH1 is responsible for the retained differentiation capability of these cells; iii) H3K27me2 and H3K27me1 ChIP-seq on all the cellular models to globally understand what happens to EZH2/EZH1 methyltransferase activities and if and how and why the mutants differ from each other. Moreover, to assess EZH2/EZH1 specific roles differentiation towards neural precursor

cells followed by astrocytes, oligodendrocytes and neurons differentiation taking advantage of well-established protocols will also be performed as well as *in vivo* teratoma formation assays.

EBs formation assays showed that when EZH2 catalytic activity is severely impaired, as in △SET-EZH2 and N-ter△EZH2 cells, but also in EZH2 Y726D/R685C expressing cells, differentiation is highly impaired. The residual mark maintained in EZH2 Y641N expressing cells seems enough to sustain differentiation and this effect is indeed EZH1-dependent. It has to be clarified if and how the EZH1-related basal level of mark is enough to sustain differentiation in the presence of EZH2 Y641N. Moreover, if EZH1 is actually responsible for this compensatory effect it remains to be elucidated why this is not happening also in different genetic contexts including SET-EZH2 and EZH2 Y726D/R685C expressing cells.

H3K27me1 global levels were unaffected in the presence of △SET or Y641N mutation, whereas they were strongly reduced upon *Ezh1* KO in these two genetic backgrounds. My results confirmed a partial capability of EZH1 to compensate for EZH2-H3K27me3 and me2 activities. Recently, a work from our group showed that PRC2 controls the deposition of all forms of H3K27 methylation (me1, me2 and me3) demonstrating that these modifications are deposited in mutually exclusive, spatially defined, domains. While H3K27me2 is broadly deposited in both intra-genic and inter-genic domains where it prevents the aberrant acetylation of non-cell-type specific enhancers, H3K27me1 is set down at highly transcribed genes through an H3K36me3-dependent mechanism[63]. These results point to the fact that H3K27me1 can be considered an

important intermediary product EZH1-PRC2 that not only constitutes the substrate for subsequent H3K27me2 but also prevents H3K27 acetylation.

Despite EZH1 and EZH2 have been shown to co-occupy the same set of genes in ESCs and embryonic carcinoma cells[84,232] and EZH1 has been shown to partially compensate for EZH2 absence, several observations suggest that they cannot completely compensate each other. Mice lacking *Ezh1* are viable and fertile whereas *Ezh2* KO leads to lethality at early stages of mouse development[234]. Moreover, EZH2 is highly expressed in the embryo but is barely detectable in adult tissues. while EZH1 is ubiquitously expressed and represents the predominant homolog upon terminal differentiation[84,235]. Stokic and colleagues[236] studied the interplay of the two (EZH1 or EZH2) PRC2 complexes in skeletal muscle differentiation demonstrating that EZH2-PRC2 is predominant in proliferating myoblasts whereas EZH1-PRC2 is specific of post-mitotic myotubes. The switch between the two complexes was shown to be essential for the proper activation of differentiation-specific genes. Indeed, many findings suggest that EZH1 is more than just a substitute for EZH2 and that the two repressive activities are differentially coordinated during tissue development and homeostasis. A role for EZH1 in maintaining adult hematopoietic stem cells (HSC) in a slow cycling, undifferentiated state thus protecting them from senescence has been proposed[237]. The authors proposed also a role of H3K27me1 in maintaining heterochromatin condensation and gene silencing, a mechanism very similar to the one related to Arabidopsis ATXR5 and ATXR6 histone methyltransferases[238]. In this context, EZH1 was shown to perform both H3K27me1 and me2, preventing proliferation and senescence in adults HSC. Interestingly, in the context of myoblast differentiation into myotubes they showed that EZH1 genome-wide mapping, despite a modest overlap (14%) with H3K27me3, presented a more extensive

142

overlap with H3K4me3 and RNA polymerase II regions[92]. In line with this finding, EZH1 was shown to associate with transcribed genes in both myoblasts and myotubes and its knockdown led to a decreased association of elongation form of PolII (phospho-Ser2-positive) with EZH1 target genes. In addition to the well-established compensatory role in transcriptional silencing reported in the absence of EZH2 in mESCs[84,232], these surprising results suggest a newly intriguing function for EZH1 in promoting transcription by a putative direct regulation of PolII. In perfect accordance with my results, *Ezh1* knockdown has been shown to have no consequences on H3K27me3/me2 levels[84,232]. Also in my hands *Ezh1* KO mESC display a normal H3K27 methylation pattern, H3K27me3 genome wide distribution comparable with wild type cells and normal differentiation into EBs. According to these results, loss of EZH1 seems to be perfectly compensated by EZH2 at least in terms of H3K27 methylation. Notably, EBs formation assay can only give a global idea of the differentiation capabilities of mESC. Indeed, deeper and more refined analyses are needed to unveil a potential regulatory role of EZH1 during cell-type specification. A global transcriptomic analysis on EBs may reveal even subtle differences in activation/repression of differentiation-related key genes. Moreover, more cell-specific differentiation protocols, including toward mesodermal and neural lineages as well as *in vivo* teratoma formation assays will be performed. Globally, my results demonstrate that EZH2 is completely able to compensate for EZH1 loss in terms of H3K27 methylation, as showed by WB and H3K27me3 ChIP-seq analysis and of differentiation ability of mESC. The major role of EZH1 in performing H3K27me1 was confirmed by my findings. Indeed, *Ezh1* KO in △SET- or N-ter△- or EZH2 Y641N backgrounds led to a strong reduction in H3K27me1 deposition. Residual H3K27me1 deposition observed upon *Ezh1* KO in △SET or EZH2 Y641N expressing cells

can be due to the "imperfect" nature of our *Ezh1* KO. In fact I did not confirm the loss of EZH1 protein. It is possible that, as observed for EZH2, a truncated form of the protein is still present within cells and retains some modest H3K27me1 activity. To solve this issue I am going to generate a novel *Ezh1* KO in order to obtain the complete loss of EZH1 protein. Then, I will investigate if some H3K27me1 is still retained in absence of EZH1 protein. If this will be the case, the existence of novel and uncharacterized methyltransferase activities able to perform H3K27 mono-methylation could be plausible. In that case, it will be interesting to further investigate these novel activities by means of a peptide pull-down assay followed by MS analysis. Moreover, new KOs for *Ezh2* will also be generated in order to obtain relevant models to better understand EZH1/EZH2 crosstalk, in which not only EZH2 activity is compromised but also EZH2 protein is not present anymore.

My data so far showed that EZH2 is perfectly able to compensate for EZH1 loss, also in H3K27me1 deposition, whereas EZH1 is not. Moreover, a cell-context dependent role for EZH1 also emerges as well as an enzymatic-independent role of EZH2 in supporting EZH1-PRC2 activity. Taken all together, it is highly unlikely that EZH1 is just a simple less active phenocopy of EZH2. Despite the EZH2-PRC2 function in mediating gene repression has been well characterized, EZH-PRC2 role remains controversial. Notably, the new generated cellular models represent unique tools in order to decipher EZH1 role in epigenetic regulation.

EZH1-dependent genome-wide H3K27me3 deposition.

H3K27me3 ChIP-sequencing analysis revealed that a high percentage of H3K27me3-

enriched promoters were retained in △SET-EZH2 expressing cells compared to wild type or *Ezh1* KO cells. H3K27me3 deposition at these genomic regions was indeed EZH1-dependent since it is completely abolished upon *Ezh1* KO. My analysis revealed that H3K27me3-enriched promoters identified in △SET-EZH2 expressing cells correspond to real enrichment peaks, despite much less intense and narrower compared to wild type ones. These features were already shown by Shen and collaborators[232]. As in my setting, they showed that *Ezh2-/-* (△SET) ESCs display a dramatic reduction in H3K27me2 and me3 but retain H3K27me1. Moreover, ChIP-chip analysis revealed that a consistent number of H3K27me3 peaks are indeed retained in Ezh2-/- mESCs. Interestingly, target genes of H3K27me3 and of EZH1 were largely overlapping (~80%). They also noticed that H3K27me3 and bfEzh1 peaks in *Ezh2-/-* cells were narrower compared to those in wild type cells. Finally, they observed how loss of EZH2 negatively affects EZH1 association to chromatin thus supporting the idea that EZH1-PRC2 ability to bind DNA is at least partially dependent on EZH2. In the same year, Margueron and colleagues[84] also addressed Ezh1 and Ezh2 roles in chromatin repression. They demonstrated that EZH2/EZH1-PRC2 share an overlapping set of target genes and that EZH1-PRC2 significantly represses transcription, even more efficiently than EZH2-PRC2. Unfortunately, they were unable to perform successful gene expression profile upon knockdown of *Ezh1, Ezh2* or both. My study, not only confirmed these observations but provided a consistent technological improvement. Indeed, I refined them by means of more advanced techniques: I used CRISPR/Cas9 technology to obtain *Ezh2/Ezh1* KOs and the ChIP-RX approach for H3K27me3 analysis to sustain and highlight my findings. ChIP-RX approach is emerging as an essential tool for HMT ChIP-seq analysis. As shown by Orlando and collaborators[223] the use of spike-in correction by adding a defined

amount of reference epigenome to ChIP samples, allows the correct comparison of HMTs deposition between different cellular populations. This approach permits, indeed, to obtain more accurate ChIP-sequencing analysis since it is able to guarantee to reveal even subtle differences in H3K27me3 enrichment between all the analyzed samples compared to already published datasets. The fact that in the absence of EZH2 catalytic activity EZH1 is able to specifically deposit some H3K27me3 at promoter level opens different scenarios. I can hypothesize the existence of a threshold-effect according to which EZH1 is able to start the deposition of H3K27me3 but EZH2 is required for its accumulation thus ensuring the spreading of the mark. According to this hypothesis, EZH2 is fundamental to create the typical H3K27me3 domains observed at promoter level of Polycomb targets. This is in perfect concordance with the fact that the retained H3K27me3 peaks are not only less intense but also narrower. Genome-wide deposition of H3K27me3 in EZH2 Y641N expressing cells in presence or absence of EZH1 may add another layer of comprehension, highlighting possible differences in EZH1 H3K27me3-activity depending on EZH2 status. It will be also interesting to understand why only some promoters retain H3K27me3 and which is the biological function of its specific maintenance at some, but not all target genes. Immunoblot analysis on EZH2 Y641N expressing cells suggested that much more H3K27me3 deposition is maintained compared to $\triangle$SET-EZH2 cells. As already postulated this can be explained considering non-catalytical roles of EZH2 that can somehow support EZH1 activity in a context-dependent manner. Unveiling the differences in retained H3K27me3 promoters in the two previous conditions may help me to explain the different differentiation phenotypes observed and importantly to identify new layers of epigenetic control related to EZH1/EZH2 activities. Furthermore, I will investigate if and how the maintenance of some deposition of H3K27me3 at specific targets could impact the transcriptional

146

features of the controlled genes. To do this, in addition to the repetition of the analysis on new *Ezh1, Ezh2* and *Ezh1/Ezh2* KOs, I will also characterize the genome-wide deposition of chromatin modifications involved in controlling RNApol II transcriptional elongation. To this purpose, ChIP-seq analysis for H3K36me3, H3K79me2 and H2Bub and for RNApolII occupancy will be performed. Moreover, for a more complete picture of EZH1/EZH2 interplay in my cellular models, H3K27me2 and H3K27me1 genome-wide analyses are needed. Additionally, RNA-sequencing and/or nascent RNA analyses will allow me to link the observed epigenetic profiles in our cellular models to the transcriptional status of the related genes.

A suitable model to investigate Polycomb proteins recruitment.

The recruitment of Polycomb proteins to specific target genes is matter of great debate in the epigenetic field. Many different mechanisms have been proposed to regulate PcGs binding to genomic loci including specific transcription factors or accessory proteins, chromatin features and long non-coding RNAs. PRC1 and PRC2 are known to associate in large part to the same genomic loci[239] and a possible interaction between the two has just recently been proposed[240]. CBX proteins can bind H3K27me3, but this mechanism involves only the canonical-PRC1 complexes despite being considered the base of the classical hierarchical model of Polycomb proteins recruitment. Moreover, PRC2 is also able to bind H3K27me3 through the WD40 domain of EED, thus suggesting the existence of a sort of auto-regulating mechanism[89]. The fact that the three states of H3K27 methylation are deposited in spatially defined domains raises several open questions regarding the specific recruitment of PcG to these regions. Importantly, it is known that H3K27me1 and me2 deposition does not require a stable association of PRC2 with

chromatin[63]. To address the specific contribution of H3K27me3 in mediating PRC2 recruitment I took advantage of double KO cells (N-ter△-Ezh2_Ezh1KO). ChIP-qPCR in this background confirmed that typical Polycomb targets are devoid of the H3K27me3 mark. In these cells, I observed also that upon re-introduction of wild type EZH2, H3K27me3 was recovered at all the analyzed targets to levels comparable to wt cells thus suggesting that PRC2 loaded with exogenous EZH2 was perfectly able to go back to its targets and deposit there H3K27me3. To better investigate this aspect, the generation of new double KOs (*Ezh1* and *Ezh2*) is necessary to obtain a pure model system to study PRCs recruitment. In fact, despite loss of H3K27me3 was observed at all the typical Polycomb targets, some residual SUZ12 binding was still detectable there. I will therefore perform again the experiment by re-introducing a tagged form of EZH2 in a completely H3K27me3-free system. ChIP-seq analyses for H3K27me1, me2, me3 as well as for tagged-EZH2 will be performed. Despite a deeper characterization is needed, my preliminary results clearly suggest that PRC2 recruitment to chromatin is indeed H3K27me3-independent. The study will then be extended also to PRC1 recruitment, focusing on both canonical and non-canonical PRC1 complexes. The availability of a novel H3K27me3-free model system will be very useful to elucidate the mechanism regulating PRC2 and PRC1 recruitment to chromatin.

A focus on EZH2 inactivating mutations.

As already reported, homozygous EZH2 Y641N expressing cells didn't represent a suitable study model for *Ezh2* inactivating mutations. So, a cellular model for these uncharacterized mutations is still missing. Starting from published data, both deriving from *in vitro* characterizations and from clinical studies[75,159,160,228], I selected some

critical residues within CXC or SET-EZH2 domains that when properly mutated putatively inactivate EZH2 catalytic activity. After an initial screening, two specific mutations were selected for their ability to impair H3K27me3 activity: Y726D (Y731D in human) and R685C (corresponding to R690C in human). Notably, both mutations have been previously identified in human diseases and so are of clinical interest. CRISPR/Cas9 technology, by means of an HDR strategy, was used to generate precisely genome-edited mESCs in an *Ezh1*-null background. This choice was made in order to avoid any EZH1-dependent compensation thus allowing me an unbiased focus on EZH2 specific activity. Despite EZH2 R685C protein failed to be expressed suggesting that the mutation affects somehow protein stability, EZH2 Y726D was normally expressed in mESCs and able to form a possibly active complex with SUZ12 and EED. Cells harboring this mutated form of *Ezh2* were severely impaired in their global methylation activity on H3K27. All the three methylation states of H3K27 were indeed affected by the presence of the mutation. Moreover, these cells displayed huge defects in terms of differentiation. One possible explanation, in agreement with what observed before for △SET or N-ter△-EZH2 expressing cells, is that differentiation capability of mESCs highly depends on H3K27me3 deposition at specific targets genes. In fact, when the deposition of the mark is massively lost, mESCs fail to properly differentiate into EBs. Indeed, EZH2 Y726D-derived EBs phenotype seemed even worse compared to N-ter△_Ezh1KO-expressing cells-derived EBs. Moreover, I observed a temporally earlier defective differentiation process in this condition. These results must be confirmed and replicated after the generation of new KOs for *Ezh2*, for correct comparison. Anyway, if this result will be confirmed this would suggest that the presence of an inactive form of EZH2 (but expressed and in complex) is indeed worse than its absence or truncation. To unveil eventual differences between N-

ter△_Ezh1KO/△SET_Ezh1KO and Y726D-Ezh2-derived EBs a deeper characterization is needed. A time-course RNA-seq analysis will help me to clarify differences, even subtle, in the differentiation capabilities of our mESC models. I previously attributed defects of △SET-Ezh2 expressing cells in differentiation to the huge loss of H3K27me3, that was indeed worse compared to EZH2 Y641N expressing cells. Indeed, both in that context and in EZH2 Y726D and R685C-expressing cells also H3K27me2 was strongly impaired. I cannot exclude a role of all the three methylation states in ensuring differentiation capabilities to mESCs. In presence of EZH2 Y726D and R685C, EBs phenotype seemed even worse, thus suggesting that also EZH1-dependent H3K27me1 deposition may have a role in the failure of the differentiation process. Genome-wide studies on H3K27me3, me2 and me1 deposition in these two new cell lines, combined with the same analyses conducted on single and double KOs (*Ezh1/Ezh2*) will allow me to better understand the role of the three methylations states and of the two homologous proteins, EZH2 and EZH1, in differentiation. Cell-lineage-specific differentiation protocols (i.e. towards neuronal cell types) as well as *in vivo* teratoma formation assays and tetraploid complementation will be performed in order to examine in depth PRC2 and H3K27 methylation roles in differentiation. So far, very little is known about the biological and pathological roles of inactivating mutations affecting EZH2. Focusing on Y726D mutation, ChIP-qPCR analysis revealed that both SUZ12 and RING1B, critical subunits of PRC2 and PRC1, respectively, were indeed displaced from typical Polycomb targets. Anyway, this displacement was not complete thus suggesting that both PRC2 and PRC1 were still able to bind, to some extent, their targets. Despite Western Blot analysis revealed slightly decreased levels of H2AK119ub in EZH2 Y726D expressing cells compared to wild type, ChIP-qPCR showed no loss of the mark at the analyzed targets. These preliminary results

suggest that despite a global change in the methylation state of H3K27 both PRC2 and PRC1 are still partially able to bind to their specific loci. Moreover, even if displaced from chromatin, RING1B is still able to sustain H2AK119ub to levels comparable to control cells. Obviously, genome-wide analyses are required to better understand what happens to PRC2/PRC1 binding to chromatin in absence of detectable H3K27 methylation levels. H3K27me3, me2, me1 ChIP-seq experiments will show if some enriched regions are indeed retained in presence of mutant EZH2 Y726D. The eventual maintenance of these marks at specific regions can support the existence of additional H3K27 methyltransferase apart from EZH1 and EZH2. Moreover, the transcriptional outcomes of this huge epigenetic resetting will be assessed by RNA-seq/nascent RNA analyses. The availability of this cellular model expressing catalytically inactive-EZH2 in combination with H3.3 K27M expressing mESCs, where EZH2 catalytic activity is impaired by mutations affecting its substrate, will allow a deeper understanding of EZH2 pathophysiological role. Furthermore, the comparison between all these novel mESCs models will allow me to discriminate between enzymatic activity-dependent and non-dependent functions of EZH2 within PRC2 function but also in relation with PRC1. Related to this, the fact that EBs phenotype in presence of EZH2 Y726D seemed to be more strongly affected compared to that of EBs derived from△SET-Ezh2 expressing cells may suggest that complete, even if mutated, EZH2 protein could drive additional functions that a truncated protein cannot.

My study highlights the power of CRISPR/Cas9 approach for the generation of suitable knock-out or knock-in mESCs models. Indeed, I generated several genetic models in order to examine in depth EZH1 and EZH2 roles in mESC homeostasis and to understand what goes wrong in pathology, when *Ezh2* or histone H3.3 are indeed mutated. A deeper

151

characterization of the genomic and transcriptional landscapes of all these cellular models will shed light on several aspects regarding EZH1/EZH2 crosstalk and PRC2/PRC1 interplay. More in general I will be able to investigate which is the specific contribution of the different H3K27 methylations in PcG-related transcriptional and biological activities and to clarify the direct role of H3K27 methylations in mediating PRC1 and PRC2 chromatin association. Moreover, the availability of mESCs models of disease-related mutations affecting *Ezh2* (both hyperactivating and inactivating) and H3.3 represent an indispensable tool to understand the molecular mechanisms underlying these mutations giving the possibility to better investigate and understand the oncogenic/tumor suppressive role of EZH1/2 proteins. This will lead to the opportunity to find novel druggable pathways to improve the treatment of these diseases.

7. Bibliography

1   Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871 (1974).

2   Kornberg, R. D. & Thomas, J. O. Chromatin structure; oligomers of the histones. *Science* **184**, 865-868 (1974).

3   Olins, A. L. & Olins, D. E. Spheroid chromatin units (v bodies). *Science* **183**, 330-332 (1974).

4   Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).

5   Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145-150, doi:10.1038/nature01595 (2003).

6   Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248-1253, doi:10.1038/nature08473 (2009).

7   Choi, J. K. & Howe, L. J. Histone acetylation: truth of consequences? *Biochem Cell Biol* **87**, 139-150, doi:10.1139/O08-112 (2009).

8   Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41-45, doi:10.1038/47412 (2000).

9   Ramakrishnan, V. Histone H1 and chromatin higher-order structure. *Crit Rev Eukaryot Gene Expr* **7**, 215-230 (1997).

10   Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett* **589**, 2914-2922, doi:10.1016/j.febslet.2015.05.016 (2015).

11   Hansen, J. C. Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annu Rev Biophys Biomol Struct* **31**, 361-392, doi:10.1146/annurev.biophys.31.101101.140858 (2002).

12   Davie, J. R. & Candido, E. P. DNase I sensitive chromatin is enriched in the acetylated species of histone H4. *FEBS Lett* **110**, 164-168 (1980).

13   Hebbes, T. R., Thorne, A. W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J* **7**, 1395-1402 (1988).

14   Arney, K. L. & Fisher, A. G. Epigenetic aspects of differentiation. *J Cell Sci* **117**, 4355-4363, doi:10.1242/jcs.01390 (2004).

15   Kosak, S. T. & Groudine, M. Form follows function: The genomic organization of cellular differentiation. *Genes Dev* **18**, 1371-1384, doi:10.1101/gad.1209304 (2004).

16   Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147-1155, doi:10.1038/nsmb.2669 (2013).

17   Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev* **23**, 781-783, doi:10.1101/gad.1787609 (2009).

18   Deichmann, U. Epigenetics: The origins and evolution of a fashionable topic. *Dev Biol* **416**, 249-254, doi:10.1016/j.ydbio.2016.06.005 (2016).

19   Levenson, J. M. & Sweatt, J. D. Epigenetic mechanisms in memory formation. *Nat Rev Neurosci* **6**, 108-118, doi:10.1038/nrn1604 (2005).

20   Singleton, M. R. & Wigley, D. B. Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol* **184**, 1819-1826 (2002).

154

21   Hargreaves, D. C. & Crabtree, G. R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* **21**, 396-420, doi:10.1038/cr.2011.32 (2011).

22   Cosma, M. P. Ordered recruitment: gene-specific mechanism of transcription activation. *Mol Cell* **10**, 227-236 (2002).

23   Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147-151, doi:10.1038/nature01763 (2003).

24   Sudarsanam, P. & Winston, F. The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. *Trends Genet* **16**, 345-351 (2000).

25   Tyler, J. K. & Kadonaga, J. T. The "dark side" of chromatin remodeling: repressive effects on transcription. *Cell* **99**, 443-446 (1999).

26   Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res* **21**, 421-434, doi:10.1038/cr.2011.14 (2011).

27   Zlatanova, J. & Thakar, A. H2A.Z: view from the top. *Structure* **16**, 166-179, doi:10.1016/j.str.2007.12.008 (2008).

28   Hake, S. B. & Allis, C. D. Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". *Proc Natl Acad Sci U S A* **103**, 6428-6435, doi:10.1073/pnas.0600803103 (2006).

29   Bonisch, C. & Hake, S. B. Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic Acids Res* **40**, 10719-10741, doi:10.1093/nar/gks865 (2012).

30   Dominski, Z. & Marzluff, W. F. Formation of the 3' end of histone mRNA. *Gene* **239**, 1-14 (1999).

31   Marzluff, W. F. & Duronio, R. J. Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr Opin Cell Biol* **14**, 692-699 (2002).

32   Rasmussen, T. P. *et al.* Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing. *Nucleic Acids Res* **27**, 3685-3689 (1999).

33   Zink, L. M. & Hake, S. B. Histone variants: nuclear function and disease. *Curr Opin Genet Dev* **37**, 82-89, doi:10.1016/j.gde.2015.12.002 (2016).

34   Dawson, M. A., Kouzarides, T. & Huntly, B. J. Targeting epigenetic readers in cancer. *N Engl J Med* **367**, 647-657, doi:10.1056/NEJMra1112635 (2012).

35   Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395, doi:10.1038/cr.2011.22 (2011).

36   Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930, doi:10.1126/science.1169786 (2009).

37   Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* **286**, 35334-35338, doi:10.1074/jbc.C111.284620 (2011).

38   Blackledge, N. P. & Klose, R. CpG island chromatin: a platform for gene regulation. *Epigenetics* **6**, 147-152 (2011).

39   Rothbart, S. B. & Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta* **1839**, 627-643, doi:10.1016/j.bbagrm.2014.03.001 (2014).

156

40 Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726-734, doi:10.1038/nrc3130 (2011).

41 Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21, doi:10.1101/gad.947102 (2002).

42 Robertson, K. D. DNA methylation and human disease. *Nat Rev Genet* **6**, 597-610, doi:10.1038/nrg1655 (2005).

43 Illingworth, R. S. & Bird, A. P. CpG islands--'a rough guide'. *FEBS Lett* **583**, 1713-1720, doi:10.1016/j.febslet.2009.04.012 (2009).

44 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

45 Okano, M., Xie, S. & Li, E. Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res* **26**, 2536-2540 (1998).

46 Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-1764, doi:10.1126/science.1147939 (2007).

47 Hermann, A., Gowher, H. & Jeltsch, A. Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol Life Sci* **61**, 2571-2587, doi:10.1007/s00018-004-4201-1 (2004).

48 Hermann, A., Goyal, R. & Jeltsch, A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* **279**, 48350-48359, doi:10.1074/jbc.M403427200 (2004).

49   Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* **3**, 662-673, doi:10.1038/nrg887 (2002).

50   Cosgrove, M. S., Boeke, J. D. & Wolberger, C. Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* **11**, 1037-1043, doi:10.1038/nsmb851 (2004).

51   Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128**, 669-681, doi:10.1016/j.cell.2007.01.033 (2007).

52   Arnaudo, A. M. & Garcia, B. A. Proteomic characterization of novel histone post-translational modifications. *Epigenetics Chromatin* **6**, 24, doi:10.1186/1756-8935-6-24 (2013).

53   Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705, doi:10.1016/j.cell.2007.02.005 (2007).

54   Phillips, D. M. The presence of acetyl groups of histones. *Biochem J* **87**, 258-263 (1963).

55   Yuan, J., Pu, M., Zhang, Z. & Lou, Z. Histone H3-K56 acetylation is important for genomic stability in mammals. *Cell Cycle* **8**, 1747-1753, doi:10.4161/cc.8.11.8620 (2009).

56   Chung, C. W. & Witherington, J. Progress in the discovery of small-molecule inhibitors of bromodomain--histone interactions. *J Biomol Screen* **16**, 1170-1185, doi:10.1177/1087057111421372 (2011).

57   Murray, K. The Occurrence of Epsilon-N-Methyl Lysine in Histones. *Biochemistry* **3**, 10-15 (1964).

58   Helin, K. & Dhanak, D. Chromatin proteins and modifications as drug targets. *Nature* **502**, 480-488, doi:10.1038/nature12751 (2013).

158

59   Collins, R. E. *et al.* In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J Biol Chem* **280**, 5563-5570, doi:10.1074/jbc.M410483200 (2005).

60   Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837, doi:10.1016/j.cell.2007.05.009 (2007).

61   Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).

62   Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20**, 2922-2936, doi:10.1101/gad.1477006 (2006).

63   Ferrari, K. J. *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol Cell* **53**, 49-62, doi:10.1016/j.molcel.2013.10.030 (2014).

64   Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326, doi:10.1016/j.cell.2006.02.041 (2006).

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

66   Schuettengruber, B. & Cavalli, G. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* **136**, 3531-3542, doi:10.1242/dev.033902 (2009).

67   Schuettengruber, B., Martinez, A. M., Iovino, N. & Cavalli, G. Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* **12**, 799-814, doi:10.1038/nrm3230 (2011).

68   Morey, L. & Helin, K. Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci* **35**, 323-332, doi:10.1016/j.tibs.2010.02.009 (2010).

69   Haupt, Y., Alexander, W. S., Barri, G., Klinken, S. P. & Adams, J. M. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. *Cell* **65**, 753-763 (1991).

70   Haupt, Y., Bath, M. L., Harris, A. W. & Adams, J. M. bmi-1 transgene induces lymphomas and collaborates with myc in tumorigenesis. *Oncogene* **8**, 3161-3164 (1993).

71   Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. *Sci Adv* **1**, e1500737, doi:10.1126/sciadv.1500737 (2015).

72   Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**, 1428-1435, doi:10.1038/ncb1663 (2007).

73   Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043, doi:10.1126/science.1076997 (2002).

74   Czermin, B. *et al.* Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185-196 (2002).

75   Muller, J. *et al.* Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**, 197-208 (2002).

160

76   Gao, Z. *et al.* PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* **45**, 344-356, doi:10.1016/j.molcel.2012.01.002 (2012).

77   de Napoles, M. *et al.* Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* **7**, 663-676, doi:10.1016/j.devcel.2004.10.005 (2004).

78   Leeb, M. & Wutz, A. Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J Cell Biol* **178**, 219-229, doi:10.1083/jcb.200612127 (2007).

79   Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873-878, doi:10.1038/nature02985 (2004).

80   Scelfo, A., Piunti, A. & Pasini, D. The controversial role of the Polycomb group proteins in transcription and cancer: how much do we not understand Polycomb proteins? *FEBS J* **282**, 1703-1722, doi:10.1111/febs.13112 (2015).

81   Allis, C. D. *et al.* New nomenclature for chromatin-modifying enzymes. *Cell* **131**, 633-636, doi:10.1016/j.cell.2007.10.039 (2007).

82   Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-678, doi:10.1016/j.cell.2011.12.029 (2012).

83   Morey, L., Aloia, L., Cozzuto, L., Benitah, S. A. & Di Croce, L. RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. *Cell Rep* **3**, 60-69, doi:10.1016/j.celrep.2012.11.026 (2013).

84   Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell* **32**, 503-518, doi:10.1016/j.molcel.2008.11.004 (2008).

85   Smits, A. H., Jansen, P. W., Poser, I., Hyman, A. A. & Vermeulen, M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* **41**, e28, doi:10.1093/nar/gks941 (2013).

86   Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* **15**, 57-67, doi:10.1016/j.molcel.2004.06.020 (2004).

87   Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J* **23**, 4061-4071, doi:10.1038/sj.emboj.7600402 (2004).

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

89   Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762-767, doi:10.1038/nature08398 (2009).

90   Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353, doi:10.1038/nature04733 (2006).

91   Xu, B. *et al.* Selective inhibition of EZH2 and EZH1 enzymatic activity by a small molecule suppresses MLL-rearranged leukemia. *Blood* **125**, 346-357, doi:10.1182/blood-2014-06-581082 (2015).

92   Mousavi, K., Zare, H., Wang, A. H. & Sartorelli, V. Polycomb protein Ezh1 promotes RNA polymerase II elongation. *Mol Cell* **45**, 255-262, doi:10.1016/j.molcel.2011.11.019 (2012).

93   Ezhkova, E. *et al.* Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* **136**, 1122-1135, doi:10.1016/j.cell.2008.12.043 (2009).

94   Ketel, C. S. *et al.* Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. *Mol Cell Biol* **25**, 6857-6868, doi:10.1128/MCB.25.16.6857-6868.2005 (2005).

95   Cao, R. *et al.* Role of hPHF1 in H3K27 methylation and Hox gene silencing. *Mol Cell Biol* **28**, 1862-1872, doi:10.1128/MCB.01589-07 (2008).

96   Ballare, C. *et al.* Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat Struct Mol Biol* **19**, 1257-1265, doi:10.1038/nsmb.2434 (2012).

97   Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306-310, doi:10.1038/nature08788 (2010).

98   Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349, doi:10.1038/nature09784 (2011).

99   Busturia, A. & Bienz, M. Silencers in abdominal-B, a homeotic Drosophila gene. *EMBO J* **12**, 1415-1425 (1993).

100   Sengupta, A. K., Kuhrs, A. & Muller, J. General transcriptional silencing by a Polycomb response element in Drosophila. *Development* **131**, 1959-1965, doi:10.1242/dev.01084 (2004).

101   Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242, doi:10.1371/journal.pgen.1000242 (2008).

102   Mendenhall, E. M. *et al.* GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* **6**, e1001244, doi:10.1371/journal.pgen.1001244 (2010).

103   Lynch, M. D. *et al.* An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J* **31**, 317-329, doi:10.1038/emboj.2011.399 (2012).

104   Riising, E. M. *et al.* Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* **55**, 347-360, doi:10.1016/j.molcel.2014.06.005 (2014).

105   Enderle, D. *et al.* Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. *Genome Res* **21**, 216-226, doi:10.1101/gr.114348.110 (2011).

106   Schwartz, Y. B. & Pirrotta, V. Ruled by ubiquitylation: a new order for polycomb recruitment. *Cell Rep* **8**, 321-325, doi:10.1016/j.celrep.2014.07.001 (2014).

107   Dietrich, N. *et al.* REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genet* **8**, e1002494, doi:10.1371/journal.pgen.1002494 (2012).

164

108   Walker, E. *et al.* Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* **6**, 153-166, doi:10.1016/j.stem.2009.12.014 (2010).

109   Farcas, A. M. *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* **1**, e00205, doi:10.7554/eLife.00205 (2012).

110   Brookes, E. *et al.* Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell* **10**, 157-170, doi:10.1016/j.stem.2011.12.017 (2012).

111   Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756, doi:10.1126/science.1163045 (2008).

112   da Rocha, S. T. *et al.* Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell* **53**, 301-316, doi:10.1016/j.molcel.2014.01.002 (2014).

113   Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol* **20**, 1250-1257, doi:10.1038/nsmb.2679 (2013).

114   Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**, 1123-1136, doi:10.1101/gad.381706 (2006).

115   Orkin, S. H. & Hochedlinger, K. Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145**, 835-850, doi:10.1016/j.cell.2011.05.019 (2011).

116   Chamberlain, S. J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **26**, 1496-1505, doi:10.1634/stemcells.2008-0102 (2008).

117   Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* **27**, 3769-3779, doi:10.1128/MCB.01432-06 (2007).

118   Richly, H., Aloia, L. & Di Croce, L. Roles of the Polycomb group proteins in stem cells and cancer. *Cell Death Dis* **2**, e204, doi:10.1038/cddis.2011.84 (2011).

119   Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**, 1513-1524, doi:10.1242/dev.014340 (2008).

120   Roman-Trufero, M. *et al.* Maintenance of undifferentiated state and self-renewal of embryonic neural stem cells by Polycomb protein Ring1B. *Stem Cells* **27**, 1559-1570, doi:10.1002/stem.82 (2009).

121   Morey, L. *et al.* Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47-62, doi:10.1016/j.stem.2011.12.006 (2012).

122   O'Loghlen, A. *et al.* MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell* **10**, 33-46, doi:10.1016/j.stem.2011.12.004 (2012).

123   Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30**, 755-766, doi:10.1016/j.molcel.2008.05.007 (2008).

166

124   Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560, doi:10.1038/nature06008 (2007).

125   Oguro, H. *et al.* Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1. *Cell Stem Cell* **6**, 279-286, doi:10.1016/j.stem.2010.01.005 (2010).

126   Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).

127   Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nat Rev Cancer* **4**, 143-153, doi:10.1038/nrc1279 (2004).

128   Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39**, D945-950, doi:10.1093/nar/gkq929 (2011).

129   Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719-724, doi:10.1038/nature07943 (2009).

130   Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-476, doi:10.1038/nrg2341 (2008).

131   Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**, 455, doi:10.1126/science.1083557 (2003).

132   Howard, G., Eiges, R., Gaudet, F., Jaenisch, R. & Eden, A. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* **27**, 404-408, doi:10.1038/sj.onc.1210631 (2008).

133   Wang, J. *et al.* Conditional MLL-CBP targets GMP and models therapy-related myeloproliferative disease. *EMBO J* **24**, 368-381, doi:10.1038/sj.emboj.7600521 (2005).

134   Huntly, B. J. *et al.* MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell* **6**, 587-596, doi:10.1016/j.ccr.2004.10.015 (2004).

135   Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189-195, doi:10.1038/nature09730 (2011).

136   Dawson, M. A. *et al.* Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* **478**, 529-533, doi:10.1038/nature10509 (2011).

137   Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073, doi:10.1038/nature09504 (2010).

138   Delmore, J. E. *et al.* BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904-917, doi:10.1016/j.cell.2011.08.017 (2011).

139   Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524-528, doi:10.1038/nature10334 (2011).

140   Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683-692, doi:10.1016/j.cell.2007.01.029 (2007).

141   Raaphorst, F. M. Deregulated expression of Polycomb-group oncogenes in human malignant lymphomas and epithelial tumors. *Hum Mol Genet* **14 Spec No 1**, R93-R100, doi:10.1093/hmg/ddi111 (2005).

168

142   Boukarabila, H. *et al.* The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev* **23**, 1195-1206, doi:10.1101/gad.512009 (2009).

143   Yu, J. *et al.* An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* **17**, 443-454, doi:10.1016/j.ccr.2010.03.018 (2010).

144   Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**, 846-856, doi:10.1038/nrc1991 (2006).

145   Mohty, M., Yong, A. S., Szydlo, R. M., Apperley, J. F. & Melo, J. V. The polycomb group BMI1 gene is a molecular marker for predicting prognosis of chronic myeloid leukemia. *Blood* **110**, 380-383, doi:10.1182/blood-2006-12-065599 (2007).

146   Shafaroudi, A. M. *et al.* Overexpression of BMI1, a polycomb group repressor protein, in bladder tumors: a preliminary report. *Urol J* **5**, 99-105 (2008).

147   Schwartz, Y. B. & Pirrotta, V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat Rev Genet* **14**, 853-864, doi:10.1038/nrg3603 (2013).

148   Klauke, K. *et al.* Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat Cell Biol* **15**, 353-362, doi:10.1038/ncb2701 (2013).

149   Scott, C. L. *et al.* Role of the chromobox protein CBX7 in lymphomagenesis. *Proc Natl Acad Sci U S A* **104**, 5389-5394, doi:10.1073/pnas.0608721104 (2007).

150   Forzati, F. *et al.* CBX7 is a tumor suppressor in mice and humans. *J Clin Invest* **122**, 612-623, doi:10.1172/JCI58620 (2012).

151   Shinjo, K. *et al.* Expression of chromobox homolog 7 (CBX7) is associated with poor prognosis in ovarian clear cell adenocarcinoma via TRAIL-induced apoptotic pathway regulation. *Int J Cancer* **135**, 308-318, doi:10.1002/ijc.28692 (2014).

152   Kreso, A. *et al.* Self-renewal as a therapeutic target in human colorectal cancer. *Nat Med* **20**, 29-36, doi:10.1038/nm.3418 (2014).

153   Simhadri, C. *et al.* Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). *J Med Chem* **57**, 2874-2883, doi:10.1021/jm401487x (2014).

154   Volkel, P., Dupret, B., Le Bourhis, X. & Angrand, P. O. Diverse involvement of EZH2 in cancer epigenetics. *Am J Transl Res* **7**, 175-193 (2015).

155   Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* **100**, 11606-11611, doi:10.1073/pnas.1933744100 (2003).

156   Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* **42**, 181-185, doi:10.1038/ng.518 (2010).

157   McCabe, M. T. *et al.* Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc Natl Acad Sci U S A* **109**, 2989-2994, doi:10.1073/pnas.1116418109 (2012).

158   Sneeringer, C. J. *et al.* Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3

170

(H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci U S A* **107**, 20980-20985, doi:10.1073/pnas.1012525107 (2010).

159   Ernst, T. *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* **42**, 722-726, doi:10.1038/ng.621 (2010).

160   Nikoloski, G. *et al.* Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet* **42**, 665-667, doi:10.1038/ng.620 (2010).

161   Ntziachristos, P. *et al.* Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat Med* **18**, 298-301, doi:10.1038/nm.2651 (2012).

162   Knutson, S. K. *et al.* Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. *Mol Cancer Ther* **13**, 842-854, doi:10.1158/1535-7163.MCT-13-0773 (2014).

163   McCabe, M. T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108-112, doi:10.1038/nature11606 (2012).

164   Qi, W. *et al.* Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation. *Proc Natl Acad Sci U S A* **109**, 21360-21365, doi:10.1073/pnas.1210371110 (2012).

165   Kim, W. *et al.* Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. *Nat Chem Biol* **9**, 643-650, doi:10.1038/nchembio.1331 (2013).

166   Bodor, C. *et al.* EZH2 mutations are frequent and represent an early event in follicular lymphoma. *Blood* **122**, 3165-3168, doi:10.1182/blood-2013-04-496893 (2013).

167   Majer, C. R. *et al.* A687V EZH2 is a gain-of-function mutation found in lymphoma patients. *FEBS Lett* **586**, 3448-3451, doi:10.1016/j.febslet.2012.07.066 (2012).

168   Yap, D. B. *et al.* Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* **117**, 2451-2459, doi:10.1182/blood-2010-11-321208 (2011).

169   Souroullas, G. P. *et al.* An oncogenic Ezh2 mutation induces tumors through global redistribution of histone 3 lysine 27 trimethylation. *Nat Med* **22**, 632-640, doi:10.1038/nm.4092 (2016).

170   Jiao, L. & Liu, X. Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science* **350**, aac4383, doi:10.1126/science.aac4383 (2015).

171   Barsotti, A. M. *et al.* Epigenetic reprogramming by tumor-derived EZH2 gain-of-function mutations promotes aggressive 3D cell morphologies and enhances melanoma tumor growth. *Oncotarget* **6**, 2928-2938, doi:10.18632/oncotarget.2758 (2015).

172   Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226-231, doi:10.1038/nature10833 (2012).

172

173   Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet* **44**, 251-253, doi:10.1038/ng.1102 (2012).

174   Kallappagoudar, S., Yadav, R. K., Lowe, B. R. & Partridge, J. F. Histone H3 mutations--a special role for H3.3 in tumorigenesis? *Chromosoma* **124**, 177-189, doi:10.1007/s00412-015-0510-4 (2015).

175   Bjerke, L. *et al.* Histone H3.3. mutations drive pediatric glioblastoma through upregulation of MYCN. *Cancer Discov* **3**, 512-519, doi:10.1158/2159-8290.CD-12-0426 (2013).

176   Sturm, D. *et al.* Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22**, 425-437, doi:10.1016/j.ccr.2012.08.024 (2012).

177   Behjati, S. *et al.* Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat Genet* **45**, 1479-1482, doi:10.1038/ng.2814 (2013).

178   Lewis, P. W. *et al.* Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* **340**, 857-861, doi:10.1126/science.1232245 (2013).

179   Bender, S. *et al.* Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. *Cancer Cell* **24**, 660-672, doi:10.1016/j.ccr.2013.10.006 (2013).

180   Chan, K. M. *et al.* The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes Dev* **27**, 985-990, doi:10.1101/gad.217778.113 (2013).

181   Fang, D. *et al.* The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. *Science* **352**, 1344-1348, doi:10.1126/science.aae0065 (2016).

182   Lan, F. & Shi, Y. Histone H3.3 and cancer: A potential reader connection. *Proc Natl Acad Sci U S A* **112**, 6814-6819, doi:10.1073/pnas.1418996111 (2015).

183   Barrangou, R. RNA events. Cas9 targeting and the CRISPR revolution. *Science* **344**, 707-708, doi:10.1126/science.1252964 (2014).

184   Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* **45**, 273-297, doi:10.1146/annurev-genet-110410-132430 (2011).

185   Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* **64**, 475-493, doi:10.1146/annurev.micro.112408.134123 (2010).

186   Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278, doi:10.1016/j.cell.2014.05.010 (2014).

187   Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740, doi:10.1099/mic.0.023960-0 (2009).

188   Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

189   Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

174

190   Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-2586, doi:10.1073/pnas.1208507109 (2012).

191   Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

192   Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389, doi:10.1016/j.cell.2013.08.021 (2013).

193   Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279-284, doi:10.1038/nbt.2808 (2014).

194   Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771, doi:10.1016/j.cell.2015.09.038 (2015).

195   Zhang, Y. *et al.* Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. *Mol Cell* **50**, 488-503, doi:10.1016/j.molcel.2013.05.001 (2013).

196   Briner, A. E. *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* **56**, 333-339, doi:10.1016/j.molcel.2014.09.019 (2014).

197   Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-350, doi:10.1016/j.cell.2014.11.052 (2015).

198   Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **33**, 73-80, doi:10.1038/nbt.3081 (2015).

199   Miyaoka, Y. *et al.* Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep* **6**, 23549, doi:10.1038/srep23549 (2016).

200   Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**, 543-548, doi:10.1038/nbt.3198 (2015).

201   Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**, 538-542, doi:10.1038/nbt.3190 (2015).

202   Vartak, S. V. & Raghavan, S. C. Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. *FEBS J* **282**, 4289-4294, doi:10.1111/febs.13416 (2015).

203   Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766, doi:10.7554/eLife.04766 (2014).

204   Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339-344, doi:10.1038/nbt.3481 (2016).

205   Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125-129, doi:10.1038/nature17664 (2016).

176

206   Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell*, doi:10.1016/j.cell.2016.10.044 (2016).

207   Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451, doi:10.1016/j.cell.2013.06.044 (2013).

208   Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588, doi:10.1038/nature14136 (2015).

209   Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* **10**, 977-979, doi:10.1038/nmeth.2598 (2013).

210   Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661, doi:10.1016/j.cell.2014.09.029 (2014).

211   Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635-646, doi:10.1016/j.cell.2014.09.039 (2014).

212   Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510-517, doi:10.1038/nbt.3199 (2015).

213   Liao, J. *et al.* Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat Genet* **47**, 469-478, doi:10.1038/ng.3258 (2015).

214   Smith, C. *et al.* Efficient and allele-specific genome editing of disease loci in human iPSCs. *Mol Ther* **23**, 570-577, doi:10.1038/mt.2014.226 (2015).

215   Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370-1379, doi:10.1016/j.cell.2013.08.022 (2013).

216   Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-918, doi:10.1016/j.cell.2013.04.025 (2013).

217   Yang, H., Wang, H. & Jaenisch, R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc* **9**, 1956-1968, doi:10.1038/nprot.2014.134 (2014).

218   Ousterout, D. G. *et al.* Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat Commun* **6**, 6244, doi:10.1038/ncomms7244 (2015).

219   Osborn, M. J. *et al.* Fanconi anemia gene editing by the CRISPR/Cas9 system. *Hum Gene Ther* **26**, 114-126, doi:10.1089/hum.2014.111 (2015).

220   Park, C. Y. *et al.* Functional Correction of Large Factor VIII Gene Chromosomal Inversions in Hemophilia A Patient-Derived iPSCs Using CRISPR-Cas9. *Cell Stem Cell* **17**, 213-220, doi:10.1016/j.stem.2015.07.001 (2015).

221   Schwank, G. *et al.* Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**, 653-658, doi:10.1016/j.stem.2013.11.002 (2013).

222   Xie, F. *et al.* Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res* **24**, 1526-1533, doi:10.1101/gr.173427.114 (2014).

223   Orlando, D. A. *et al.* Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep* **9**, 1163-1170, doi:10.1016/j.celrep.2014.10.018 (2014).

224   Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem* **78**, 477-513, doi:10.1146/annurev.biochem.78.081507.101607 (2009).

225   Dantuma, N. P., Groothuis, T. A., Salomons, F. A. & Neefjes, J. A dynamic ubiquitin equilibrium couples proteasomal activity to chromatin remodeling. *J Cell Biol* **173**, 19-26, doi:10.1083/jcb.200510071 (2006).

226   Zhang, L., Hu, J. J. & Gong, F. MG132 inhibition of proteasome blocks apoptosis induced by severe DNA damage. *Cell Cycle* **10**, 3515-3518, doi:10.4161/cc.10.20.17789 (2011).

227   Young, M. D. *et al.* ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res* **39**, 7415-7427, doi:10.1093/nar/gkr416 (2011).

228   Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**, 2893-2905, doi:10.1101/gad.1035902 (2002).

229   Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27, doi:10.1016/j.cell.2012.06.013 (2012).

230   Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624-629, doi:10.1038/nature01075 (2002).

231   Kim, K. H. & Roberts, C. W. Targeting EZH2 in cancer. *Nat Med* **22**, 128-134, doi:10.1038/nm.4036 (2016).

232   Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* **32**, 491-502, doi:10.1016/j.molcel.2008.10.016 (2008).

233   Davidovich, C., Goodrich, K. J., Gooding, A. R. & Cech, T. R. A dimeric state for PRC2. *Nucleic Acids Res* **42**, 9236-9248, doi:10.1093/nar/gku540 (2014).

234   Ezhkova, E. *et al.* EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair. *Genes Dev* **25**, 485-498, doi:10.1101/gad.2019811 (2011).

235   Laible, G. *et al.* Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at S. cerevisiae telomeres. *EMBO J* **16**, 3219-3232, doi:10.1093/emboj/16.11.3219 (1997).

236   Stojic, L. *et al.* Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells. *Epigenetics Chromatin* **4**, 16, doi:10.1186/1756-8935-4-16 (2011).

237   Hidalgo, I. *et al.* Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest. *Cell Stem Cell* **11**, 649-662, doi:10.1016/j.stem.2012.08.001 (2012).

238   Jacob, Y. *et al.* Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* **466**, 987-991, doi:10.1038/nature09290 (2010).

239   Bracken, A. P. & Helin, K. Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat Rev Cancer* **9**, 773-784, doi:10.1038/nrc2736 (2009).

240   Cao, Q. *et al.* The central role of EED in the orchestration of polycomb group complexes. *Nat Commun* **5**, 3127, doi:10.1038/ncomms4127 (2014).

List of Publications

Ferrari, K.J., **Lavarone, E.**, Pasini, D. The dual role of EPOP and Elongin BC in controlling transcriptional activity. Mol Cell. 2016. 64(4):637-638.

Mio, C., **Lavarone, E.**, Conzatti, K., Baldan, F., Toffoletto, B., Puppin, C., Filetti, S., Durante, C., Russo, D., Orlacchio, A., Di Cristofano, A., Di Loreto, C., Damante, G. *MCM5 as a target of BET inhibitors in thyroid cancer cells.* Endocr Relat Cancer. 2016. 23(4):335-47.

Baldan, F., Mio, C., **Lavarone, E.**, Di Loreto, C., Puglisi, F., Damante, G., Puppin, C. *Epigenetic bivalent marking is permissive to the synergy of HDAC and PARP inhibitors on TXNIP expression in breast cancer cells.* Oncol Rep. 2015. 33(5):2199-206.

Baldan, F., **Lavarone, E.**, Di Loreto, C., Filetti, S., Russo, D., Damante, G., Puppin, C. *Histone post-translational modifications induced by histone deacetylase inhibition in transcriptional control units of NIS gene.* Mol Biol Rep. 2014. 41(8):5257-65.

Puppin, C., Durante, C., Sponziello, M., Verrienti, A., Pecce, V., **Lavarone, E.**, Baldan, F., Campese, A.F., Boichard, A., Lacroix, L., Russo, D., Filetti, S., Damante, G. *Overexpression of genes involved in miRNA biogenesis in medullary thyroid carcinomas with RET mutation.* Endocrine. 2014. 47(2):528-36.

Sponziello, M., **Lavarone, E**., Pegolo, E., Di Loreto, C., Puppin, C., Russo, M.A., Bruno, R., Filetti, S., Durante, C., Russo, D., Di Cristofano, A., Damante, G. *Molecular differences between human thyroid follicular adenoma and carcinoma revealed by analysis of a murine model of thyroid cancer.* Endocrinology. 2013. 154(9):3043-53.

Lavarone, E., Puppin, C., Passon, N., Filetti, S., Russo, D., Damante, G. *The PARP inhibitor PJ34 modifies proliferation, NIS expression and epigenetic marks in thyroid cancer cell lines.* Mol Cell Endocrinol. 2013. 5;365(1):1-10.

Passon, N., Puppin, C., **Lavarone, E**., Bregant, E., Franzoni, A., Hershman, J.M., Fenton, M.S., D'Agostino, M., Durante, C., Russo, D., Filetti, S., Damante, G. *Cyclic AMP-response element modulator inhibits the promoter activity of the sodium iodide symporter gene in thyroid cancer cells.* Thyroid. 2012. 22(5):487-93.

Passon, N., Gerometta, A., Puppin, C., **Lavarone, E**., Puglisi, F., Tell, G., Di Loreto, C., Damante, G. *Expression of Dicer and Drosha in triple-negative breast cancer.* J Clin Pathol. 2012. 65(4):320-6.

Puppin, C., Passon, N., **Lavarone, E**., Di Loreto, C., Frasca, F., Vella, V., Vigneri, R., Damante, G. *Levels of histone acetylation in thyroid tumors.* Biochem Biophys Res Commun. 2011. 12;411(4):679-83.

Acknowledgements

I would like to thank my Supervisor Prof. Giuseppe Damante and my co-Supervisor Dr. Diego Pasini for giving me the opportunity to join this excellence research place and to work on such a challenging but exciting project.

I would like to thank all the Pasini's lab members for these three intense years of work, and life.

Special thanks to Laura and Andrea, friends more than colleagues. I wouldn't be able to survive this PhD without you.

Thanks to Cecilia, Daria and Simone for supporting and understanding me, especially in the last month.

Thanks to Catia, our friendship means a lot to me.

Thanks to my family, for their unlimited support.

The final huge thank goes to Federico, for being always here, right next to me.

"Twenty years from now you will be more disappointed by the things you didn't do than by the ones you did do. So throw off the bowlines. Sail away from the safe harbor.
Catch the trade winds in your sails.
Explore. Dream. Discover."
Mark Twain