PhD degree in Molecular Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: MED/04

The role of enhanced Polycomb Repressive Complex2 activity in tumorigenesis

Laura Cedrone

Istituto Italiano di Tecnologia, Milan

Matricola n. R10317

Supervisor: Dr. Diego Pasini

Istituto Europeo di Oncologia, Milan

Anno accademico 2015-2016

TABLE OF CONTENTS

FIGURES AND TABLES INDEX

LIST OF ABBREVIATIONS

4-OHT: 4-hydroxytamoxifen

5caC: 5-carboxylcytosine

5fc: 5-formylcytosine

5hmc: 5-hydroxymethylcytosyne

5mc: 5-methyl cytosine

ALL: Acute lymphocytic leukemia

ALS: amyotrophic lateral sclerosis

AML: Acute myeloid leukemia

APL: Acute promyelocytic leukemia

ASXL1: Additional Sex Combs Like 1

Bcl: B-cell lymphoma

BCOR: BCL6 Corepressor

BRCA: Breast Related Cancer Antigen

BrdU: Bromo deoxy Uridine

CBP: CREB binding protein

ChIP: Chromatin immunoprecipitation

ChIP-Rx: Chromatin immunoprecipitation with reference exogenous genome

ChIP-seq: Chromatin immunoprecipitation-sequencing

CpG: Cytosine-guanine dinucleotides

CpGi: CpG island

DNMTs: De novo methyltransferases

Dpc: Days post coitum

DZNep: 3-Deazaneplanocin A

E[z]: Enhancer of Zeste

Eed: Embryonic Ectoderm Development

EtOH: ethanol

Ezh1/2: Enhancer Of Zeste

FACS: Fluorescence Activated Cell Sorting

GFP: Green Fluorescent Protein

H2AK119ub: Mono-ubiquitination of the lysine 119 on the histone H2a

H3K27me0: un-methylated histone H3 on lysine 27

H3K27me1:mono-methylation of histone H3 on lysine 27

H3K27me2:di-methylated of histone H3 on lysine 27

H3K27me3:tri- methylation of histone H3 on lysine 27

HAT: Histone Acetyltransferase

HDAC: Histone Deacetylases

HKMTs: Histone Lysine Methyltransferase

HMTs: histone methyltransferase

HP1: heterochromatin protein1

ICF: Immunodeficiency, centromeric instability, and facial dysmorphism syndrome

IGF2: Insulin Like Growth Factor 2

iPSCs: Induced pluripotent stem cells

IRES: internal ribosome entry site

KDM1A (or LSD1): Lysine Demethylase 1A

Kdm2b: Lysine Demethylase 2B

Klf4: Kruppel-like factor 4

LIF: Leukemia Inhibitory Factor

LOH: Loss of heterozygosity

MASPIN: Mammary serine protease inhibitor

MBD: Methyl binding domain

MDS: Myelodysplastic syndromes

Mecp2: Methyl-CpG binding protein 2

MEF: Mouse embryonic fibroblasts

mESC: Mouse embryonic stem cells

MLL: Mixed lineage leukemia

MS: Mass spectrometry

NSD: Nuclear Receptor Binding SET Domain Protein

PcG: Polycomb Group proteins

PCGF: Polycomb group RING fingers

Pcl: Polycomblike proteins

PEV: Position-effect variegation

PI: Propidium iodide

PLZF-RARα: Promyelocytic leukemia zinc finger-retinoic acid receptor α

PML-RARα: Promyelocytic leukemia-retinoic acid receptor α

PRC1: Polycomb repressive complex1

PRC2: Polycomb repressive complex2

PRE: Polycomb response element

Psc: Posterior sex combs

PTMs: Post-translational modifications

qPCR: Quantitative polymerase chain reaction

RAR: Retinoic acid receptor

RAREs: Retinoic acid response elements

Rb: retinoblastoma

RbAp46/48: Retinoblastoma protein associated protein 46/48

REST: RE1-Silencing Transcription factor

RIZ1: Retinoblastoma-interacting zinc-finger protein 1

RNA-seq: RNA-sequencing

RYBP: RING1 And YY1 Binding *Protein*

SAH: S-Adenosyl-L-homocysteine

SAM: S-adenosyl-L-methionine

Sox2: SRY-Box 2

Suz12: Suppressor Of Zeste 12 Protein Homolog

TET: Ten-eleven translocation

TF: Transcription factor

Trx-G: Trithorax Group proteins

TSS: Transcription start site

UT: untreated or Empty vector

VEGFR: Vascular endothelial growth factor receptor

WB: western blot

WT: wild type

YY1: Yin Yang 1 protein

ABSTRACT

Polycomb Group of proteins (PcGs), members of polycomb repressive complex 2 (PRC2) and polycomb repressive complex 1 (PRC1), are essential factors present in cells' nuclei. These multiproteins complexes are key repressive factors that regulate cellular differentiation during development, contributing to the correct establishment of lineage-specific transcriptional programs. Their repressive activities, linked to chromatin compaction and transcriptional repression, are regulated in a cell type-specific manner. They represent key factors of proliferation and a strong correlation between the deregulated levels or activity of PcG proteins and the development of several types of human cancers has been demostrated.

Recently, gain of function (GOF) heterozygous EZH2 mutations have been discovered. These mutations are present in 7% of follicular lymphomas and 21% of cases of human diffuse large B-cell lymphomas (DLBCL) accounting for about 30-40 % of newly diagnosed cases of non-Hodgkin lymphomas. These mutations occur mainly on Tyrosine 641 within the EZH2 catalytic SET domain, causing an aminoacidic substitution from tyrosine to phenylalanine, serine, histidine or asparagine (Y641F, Y641S, Y641H, and Y641N). These mutations are responsible for the enhanced activity of EZH2 towards H3K27me3 deposition when the mutated form of the enzyme is accompanied by the wild type (WT EZH2) counterpart in lymphoma cells. In particular, the increase in H3K27me3 deposition is accompanied by a decrease in H3K27me1 and H3K27me2 levels**.** Up to now little is known about the mechanism of action of this mutated enzyme, therefore the aim of this thesis is trying to unravel the tumorigenic mechanisms underlying these mutations.

To understand a general oncogenic role for this mutated enzyme, we took advantage of an alternative, simpler model system, represented by MEF, to perform our analyses. We then validated our results in lymphoma cell lines. We could show that, indeed, the increased

genome wide deposition of the H3K27me3 mark was valid also in an independent cellular setting. Surprisingly, this altered deposition was not accompanied by any relevant transcriptional alteration at steady state. To address this issue, we subjected MEFs to three different types of stimuli (starvation, myc upregulation and reprogramming to pluripotency) in the attempt to highlight a cooperative transcriptional de-regulation for mutant EZH2. Surprisingly, we found that the output of widespread trimethylation of K27 was evident only when cells transitioned from one cell fate to another, pointing to the blockade of upregulation of non-expressed genes during cell reprogramming. Thus, we hypothesized a sentinel role for the mutant EZH2 Y641N, in which the levels of the repressive mark H3K27me3 are increasingly deposited where the mark is already present at steady state. This could underlie mutant EZH2 functionality in DLBCL, where it could lock centroblasts in a "low-differentiation" state in the germinal center and impede the up-regulation of specific factors important for their differentiation. This, in combination with other tumorigenic mutations, would lead to the onset of lymphomas in patients. If proven true, these results would link the epigenetic control cellular differentiation to tumorigenesis, shedding light on new druggable targets for improved care of patients.

1. INTRODUCTION

1.1. Chromatin Properties

1.1.1. Overview of chromatin features.

DNA in eukaryotic cells undergoes multiple levels of hierarchical folding in order to be arranged into the nucleus. In this organelle the genome is highly organized and subjected to a high degree of compaction (from 10000 to 20000 fold) in order to fit in. The hierarchy, on the other hand, assures that all regions would be correctly accessible to be replicated in each cell cycle and to be subjected to the other DNA-based processes. The organization of DNA into chromatin ends up with the formation of chromosomes where the hyper-condensed chromatin structure is obtained through the tight association of DNA with histone and non-histone proteins [1]. The basic unit of the chromatin fiber is the nucleosome. The discovery of the nucleosome date back to early seventies thanks to seminal work by Luger and colleagues through the use of X-rays diffraction. In particular, the nucleosome is composed by 147bp DNA, left handedly coiled around a complex of histone proteins that are called H2A, H2B, H3 and H4, and organized in an octameric structure in which H3-H4 tetramer associates to two adjacent H2A-H2B dimers [2]. This structure was firstly observed by Kornberg in 1974 [3, 4]. The combination of these nucleoproteins, joined together through a region of so called linker DNA, compose a structure similar to the “beads on a string” visible with electron microscopy [5]. The linker DNA between nucleosome particles may be associated with the linker proteins belonging to the H1 histone family [6, 7]. They are usually present in close proximity to where the DNA enters and exits from nucleosomes. These packed nucleosome arrays in association with linker histones and other chromatin-associated proteins allow the higher order organization of the chromatin in the nucleus. This is possible thanks to amino acids present in the positively charged nucleosomes, which let them interact

with the negatively charged phosphate group of DNA. Moreover, all histones (except H4) are present in different variants and different post-translational modifications (PTMs) can occur on them, dictating nucleosome positioning and the overall organization of chromatin. In the end they represent fundamental modulators of many cellular DNA-template processes such as transcription, replication, repair and recombination. Based on the timing of expression, we can distinguish two categories of histones: replication-dependent and replacement variants. The first ones, encoded by multiple genes copies, are clustered in the genome, are devoid of introns, do not present a poly (A) tail and their 3'-end present a stem-loop conserved sequence responsible for the high rates of transcription of these proteins [8, 9]. They are expressed in S phase, in order to assure a major supply for incorporation events during replication, and become part of chromatin during DNA synthesis. They are represented by the H3.1 and H3.2 genes. The replacement histone variants are generally expressed in tissue-specific manner and throughout the cell cycle; their mRNA can contain introns and are characterized by the presence of 3'-polyadenylated tails. They are represented by the human variant H3.3 encoded by the H3.3A and H3.3B genes.

There are some molecules responsible for the incorporation of histone proteins in order to form the nucleosome particles. These factors are called histone chaperones [10]. Eukaryotes have evolved histone chaperones that can be specific for one or more histone variants. It has been shown that the same histones can have more than one chaperone in order to answer specifically to the processes taking place in a specific moment in the cell [11, 12]. For example, Asf1 is involved in chromatin assembly after double strand break repair [13], in chromatin disassembly before replication [14, 15] and chromatin disassembly during transcription [16-18]. FACT complex is responsible for the H2A–H2B dimer removal/ displacement [19-21]. Another chaperone is DAXX whose role is the incorporation of H3.3 at telomeres [22, 23].

Overall nucleosomes represent the key factor for DNA accessibility which is firstly determined by nucleosome position that allows the exposure of specific and precise portion

of DNA. Nucleosome positioning, determined by the deposition of PTMs on histones, is responsible for the higher or lower degree of compaction of chromatin affecting the levels of transcription. In particular, with respect to the degree of the compaction we can distinguish two chromatin types: euchromatin and heterochromatin. These two states of chromatin present differences from a cytological point of view being the euchromatin less intensely stained with respect to heterochromatin, the latter being characterized by a more compact form of DNA. Conversely the euchromatin is more relaxed and less folded, in line with its higher accessibility to the replication machinery. Indeed, heterochromatic configuration is commonly found on peripheral areas of the nucleus and correlates with a poorer transcriptional activity and, likely, with gene repression [24].

These configurations are plastic and interchangeable, with specific proteins that are responsible for the high dynamicity of the chromatin structure [25]. These factors, called collectively chromatin remodelers, can induce specific changes in the shape of the chromatin, modulating its accessibility in response to different stimuli.

1.1.2. Roles of the major histone and DNA modifications.

Allfrey and colleagues showed for the first time the possibility that histone could be subjected to modification to alter transcription [26]. In particular they focused on two modifications: acetylation and methylation. Histones are characterized by two structurally and functionally distinct domains: the globular portion is the one implicated in the formation of the core of the nucleosome and the unstructured tail protruding from this core. First indications on how PTMs on these tails could affect the interaction between nucleosomes came initially from the work of Luger and colleagues in 1997 [2].

Since then, there has been a massive effort to analyze, understand, characterize, and potentially to “use”, the enzymatic activities responsible for the deposition and removal of histones and DNA modifications. Histone modifications are able not only to regulate the

nucleosomes interaction but they are essential also for the recruitment of specific proteins, i.e. remodeling complexes. The PTMs and DNA modifications constitute the epigenetic regulation of cellular processes. The term epigenetics was coined by Conrad Waddington in 1942 even two years before Avery, MacLeod, and McCarty demonstrated that the DNA was the source of hereditary information. Waddington defined epigenetics as "*a branch of biology that study the casual interactions between genes and their products*". Riggs and colleagues defined epigenetics as "*the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence"* [27]. A more recent definition, proposed by Levenson and Sweatt [28] links epigenetics directly to chromatin modifications using this definition: "*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*". Indeed, several chromatin-modifying factors play essential roles in regulating epigenetic phenomena such as gene imprinting and X-chromosome inactivation [29].

The epigenetic world is composed by several characters identified as "writers" (and "erasers"), that catalyze the addition (removal) of chemical groups onto either histone tails or the DNA itself, "readers", responsible to recognize and bind specific modifications and "targets" of the modification.

Among the well-known "writers" there are the histone acetyltransferases (HAT) and lysine methyltransferases (KMT) about which I will discuss more deeply in next paragraph. Relative erasers are histone deacetylase and demethylases.

"Readers" are characterized by domains whose structure typically provides a cavity or surface groove to accommodate a specific epigenetic mark, such as lysine mono-, di- and trimethylation. One example is represented by CBX protein where its chromodomain can specifically recognize trimethylation on H3K27.

1.1.3. DNA methylation

DNA modifications can regulate the association and the downstream functions of factors able or induced to bind DNA. In this sense DNA can be methylated on the cytosine at position 5 (5mc) in the context of CG dinucleotides. This modification is heritable and linked with gene silencing due to the fact that its presence impedes the interaction between DNA and transcription factors [30]. It acts also as docking site for protein harboring methyl binding motif (MBD) that can be chromatin remodelers and transcriptional co-repressor factors [31]. On the other hand, it is also true that all along the genome there are regions rich of CpG dinucleotides. This CpG islands are indeed maintained free of DNA methylation. About 60-70% of promoters contain CpG islands whose methylation would result in stable transcriptional repression [32]. These modification is catalyzed by a family of enzymes called *de novo* methyltransferases (DNMT1, DNMT3A and DNMT3B). The levels and patterns of DNA methylation are regulated by both DNMT and others molecules that cooperate in cytosine demethylation, like the ten-eleven translocation (TET) family of dioxygenases (TET1, TET2 and TET3). The functional outputs of DNA methylation on chromatin and gene expression are due to specific protein 'reader', for example Methyl-CpG-binding domain (MBD). In this context the TET protein catalyze iterative oxidation on 5mc resulting in the formation of 5-hydroxymethylcytosine (5hmc), 5-formylcytosine (5fc) and 5-carboxylcytosine (5caC).

Taken together all these phenomena, histones and DNA together with all the modifications they are subjected to, work in a cooperative manner in a way that enable the chromatin to answer to environment, stimuli or stress through rapid changes in its organization in order to recruit or viceversa to remove specific effector proteins all along the entire genome.

1.1.4. Post-translational modifications

Among the most studied and well known PTMs, there are acetylation and methylation occurring on histone lysine residues, phosphorylation of serine, threonine and tyrosine,

ubiquitylation of lysine [33]. Other histone PTMs such as crotonilation, lysine butyrylation, lysine propionylation are less abundant (Table 1).

There are several complexes responsible for the deposition of these modifications (Table2). For example the so-called histone acetyl transferases (or HATs) are the ones capable of exert the acetylation of lysines using acetyl-CoA as cofactor to catalyze the transfer of an acetyl group to the Ɛ-amino group of lysine lateral chain. Several proteins belong to this group: CBP/p300, MYST and GNAT families. Acetylation neutralizes the positive charge of lysines, thus weakening interactions between histone and DNA, leading to an increased accessibility of DNA to the transcription machinery. Indeed, they are found to be enriched at actively transcribed regions. Histone acetylation has been demonstrated to be also important during DNA replication and be indispensable for origin firing [34, 35]. Moreover, acetylated lysines can be recognized by bromodomains of chromatin remodeler proteins, thus influencing the chromatin structure [35] High throughput approaches revealed that H3K27ac was enriched at promoter regions of transcriptionally active genes [36], but also Polycomb target sites. Indeed, H3K27ac has dynamic and complementary temporal deposition profiles during embryogenesis [37]. This is a highly dynamic modification since once added it can be removed by specific erasers called histone deacetylases (HDAC).

Also histone phosphorylation is highly dynamic. It takes place on serines, threonines and tyrosines, predominantly [38]. Its balance is controlled by two functionally opposite molecules that are kinases and phosphatases which respectively add and remove the modification [39]. The kinases act transferring phosphate group from ATP to the hydroxyl group of the target amino-acid side chain. When this modification is added to the histones, the final outcome is that they resulted characterized by an increased amount of negative charge, which in the end will affect the chromatin structure.

Chromatin Modifications	Residues Modified	Functions Regulated
Acetylation	**K**-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	**K**-me1 **K**-me2 **K**-me3	Transcription, Repair
Methylation (arginines)	**R**-me1 **R**-me2a **R**-me2s	Transcription
Phosphorylation	**S**-ph **T**-ph	Transcription, Repair, Condensation
Ubiquitylation	**K**-ub	Transcription, Repair
Sumoylation	**K**-su	Transcription
ADP ribosylation	**E**-ar	Transcription
Deimination	**R > Cit**	Transcription
Proline Isomerization	**P-cis > P-trans**	Transcription

Table 1: *Summary of residues subjected to post-translational modification and their functional outcome.*

From Kouzarides 2007 [40].

Acetyltransferase	
HAT1	H4 (K5, K12)
CBP/P300	H3 (K14, K18) H4 (K5, K8) H2A (K5) H2B (K12, K15)
PCAF/GCN5	H3 (K9, K14, K18)
TIP60	H4 (K5, K8, K12, K16) H3 K14
HB01 (ScESA1, SpMST1)	H4 (K5, K8, K12)
ScSAS3	H3 (K14, K23)
ScSAS2 (SpMST2)	H4 K16
ScRTT109	H3 K56
Deacetylases	
SirT2 (ScSir2)	H4 K16

Lysine Demethylases	
LSD1/BHC110	H3K4
JHDM1a	H3K36
JHDM1b	H3K36
JHDM2a	H3K9
JHDM2b	H3K9
JMJD2A/JHDM3A	H3K9, H3K36
JMJD2B	H3K9
JMJD2C/GASC1	H3K9, H3K36
JMJD2D	H3K9

Arginine Methlytransferases	
CARM1	H3 (R2, R17, R26
PRMT4	H4R3
PRMT5	H3R8, H4R3
Serine/Thrionine Kinases	
Haspin	H3T3
MSK1	H3S28
MSK2	H3S28
CKII	H4S1
Mst1	H2BS14
Ubiquitilases	
Bmi/Ring1A	H2AK119
RNF20/RNF40	H2BK120
Proline Isomerases	
ScFPR4	H3P30, H3P38

Lysine Methyltransferase	
SUV39H1	H3K9
SUV39H2	H3K9
G9a	H3K9
ESET/SETDB1	H3K9
EuHMTase/GLP	H3K9
CLL8	H3K9
SpClr4	H3K9
MLL1	H3K4
MLL2	H3K4
MLL3	H3K4
MLL4	H3K4
MLL5	H3K4
SET1A	H3K4
SET1B	H3K4
ASH1	H3K4
Sc/Sp SET1	H3K4
SET2 (Sc/Sp SET2)	H3K36
NSD1	H3K36
SYMD2	H3K36
DOT1	H3K79
Sc/Sp DOT1	H3K79
Pr-SET 7/8	H4K20
SUV4 20H1	H4K20
SUV420H2	H4K20
SpSet 9	H4K20
EZH2	H3K27
RIZ1	H3K9

Table 2: *List of enzymes that induce epigenetic modifications*

Adapted from Kouzarides 2007 [40].

1.1.5. Histone methylation

Histone methylation is a PTM that can occur on all basic aminoacidic residues such as arginines, histidines and lysines. It is possible to find monomethylated and symmetrically- or asymmetrically dimethylated arginines on their guanidinyl group. Histidines can be monomethylated. Lysines can be mono (me1)-, di (me2)- and tri-methylated (me3) on their Ɛ-group. Mass spectrometry and quantitative proteomic analyses [41] revealed that histone H3 has several basic residues that can be modified. In particular the best characterized arginine (R) sites that can be methylated include H3R2, H3R8, H3R17, H3R26 and H4R3. Instead, for lysines the most well-known and characterized sites are histone H3 lysine 4 (H3K4), H3K9, H3K27, H3K36, H3K79 and H4K20.

Compared to others PTMs, histone methyl groups are characterized by a slower turn over. Indeed this modification has been thought to be irreversible [42] until 2004 when Shi and colleagues reported firstly evidence of the existence of an "histone lysine demethylase", specifically lysine-specific demethylase 1A, KDM1A (or LSD1) [43]. Thereafter, a plethora of methyltransferases and demethylases have been discovered. These enzymes mediate the addition to, and removal of methyl groups from the lysine residues on histones [44]. It is also true that the lower or greater duration of the presence of the modification rely on the specificity of the ongoing process.

Methylated histones are recognized by the "readers" that mediate the recruitment of other molecules to alter chromatin structure and consequently transcription [45]. Indeed histone lysine methylation can promote both active transcription, such as in the case of the tri-methylation (me3) of H3K36 and H3K4, or transcriptional repression, as in case of H3K9me3 and H3K27me3. In this sense, the methylation mark can be suitable for dynamic regulation of gene expression, since it can be actively added and removed from histone tails. The enzymes responsible for the histone methylation on lysine residues are the histone lysine methyltransferases (HKMTs). They use S-adenylmethionine (SAM) to add methyl groups

to their specific substrate.

The HKMTs can be divided in two different families based on the characteristics of their catalytic domain. The first class is characterized by a SET domain. This is an evolutionary conserved sequence motif of 130 amino acids which was initially identified in Drosophila position-effect-variegation (PEV) suppressor gene *Su(Var)3-9*, in the Polycomb group (Pc-G) gene *Enhancer of Zeste* (E[z]) and in the activating Thritorax group (Trx-G) gene *Thritorax* [46]. It performs the transfer of a methyl group from S-adenosylmethionine (SAM) to the ε-amino group of the lysine (Fig. 1). The second family consists only in a single protein, Dot1L that is devoid of a SET domain [47].

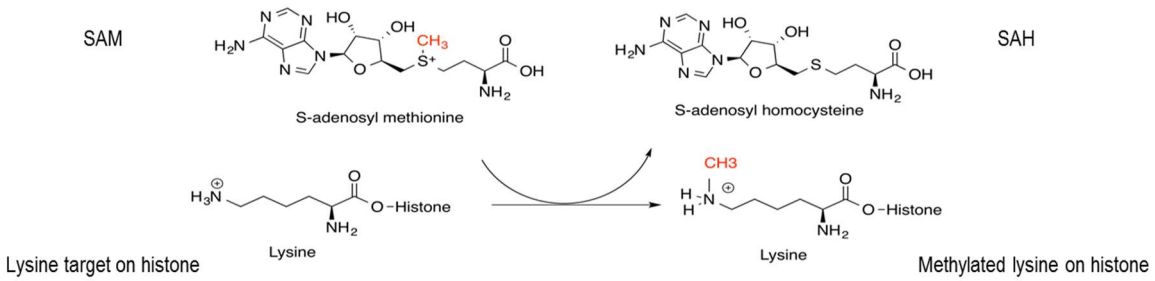


Figure 1: Mechanism of methylation.

The lysine amino group of the substrate histone polypeptide starts a reaction with the activated co-factor *S*-adenosylmethionine (SAM), resulting in the formation of an *N*-methylated lysine and *S*-adenosylhomocysteine (SAH). Adapted from Helin and Dhanak 2013 [48].

HKMTs 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) can play an essential role in controlling the state of methylation [49]. Taking into consideration their role in regulating transcription, among the most relevant lysine methylations, there are H3K4me1, H3K4me3 H3K9me3 and H3K27me. The latter PTM is deposited by Polycomb Repressive Complex2 (PRC2).

The first methyltransferase identified as able to selectively regulate H3K9 methylation was Suv39H1/2 [50], which deposits H3K9me3 mark. Subsequently, it was reported that the

H3K9 mono- and dimethylation is controlled by the G9a enzyme [51] and its paralogue Glp [52]. Suv39H1 acts preferentially on mono- or dimethylated H3K9 as a primary substrate [51]. The trimethylation of H3K9 and the subsequent binding of heterochromatin protein 1 (HP1) mediate the formation and maintenance of pericentromeric heterochromatin. Indeed during the mitosis numerous SUV39H1 factors are transiently accumulated at pericentromeric heterochromatin, suggesting that this KHMT may also play a structural role in these regions [53].

The first and unique enzyme identified in S.cerevisiae containing a SET domain able to specifically methylate lysine 4 on histone H3, a mark associated with transcriptionally active chromatin, was Set1 [54]. In mammals, instead, about ten enzymes exist which can contain a SET domain related to yeast Set1 or to Drosophila Trithorax. Together they are classified as MLL family, which includes six members. The presence of a number of H3K4 methyltransferases raises the question of why yeast can satisfy the task with just one enzyme while in mammals multiple isoforms are required [55]. It has also been shown that mutations occurring in different MLL genes are linked to different phenotypes. This suggest that these proteins are likely to be not fully redundant in their functions, but that they are all necessary in order to accomplish the complexity of the developmental process in mammals [56-58]. The MLL methyltranferases exist in a multiprotein complex with other subunits whose composition is highly similar to the Set1 complex in yeast [59]. Despite their association with several different subunits, a constant core made of three factors exists: WDR5, which interact with the histone H3 tail [60], presenting the chain to the other two proteins, RbPB5 and Ash2, that further methylate it [61]. The MLL family can form a complex with other partners like Menin, which is required for maintenance of Hox gene expression [59] or HCF-1, a regulator of cell proliferation involved in both cell growth and division processes [62]. MLL complex has the important role of antagonizing the repressive activity of Polycomb complexes by depositing methylation on H3K4 [63]. Indeed, *in vitro* studies demonstrated that when MLL proteins catalyze H3K4 trimethylation it impairs Polycomb activity in

depositing the H3K27me3 mark [64]. Several promoters of eukaryotic genes are characterized by the presence of H3K4me3. This modification is able to recruit nucleosome remodeling complexes and acetyltranferases [65, 66]. This results in the transcriptional activation of these promoters. Moreover, these sites have been found also enriched for RNA polymerase II [67].

In mouse embryonic stem cells there are several so-called bivalent promoters. The term "bivalent" is due to their contemporary features of both active and repressive chromatin since they are characterized by the presence of the two functionally opposite histone marks H3K4me3 and H3K27me3 [68]. A large proportion of these kind of loci corresponds to developmental genes encoding transcription factors and other regulators of the cellular state. These genes are mostly repressed in pluripotent cells but they can be rapidly turned on or stably inactivated according on the developmental stage that it is necessary to achieve or maintain.

Another important regulatory DNA region that is regulated by PTMs is represented by enhancers. The activation of these regions is able to greatly increase the transcription of gene promoters. Moreover these loci can regulate genes whose distance can range from hundreds of bases to megabases [69]. These regions are characterized by the deposition of several PTMs [70, 71], the most common of which is H3K4me1 [72, 73]. Besides this modification they can accommodate H3K27 acetylation. If they are marked by both of these modifications they are called active enhancers, while in the absence of the histone acetylation and/or presence of H3K27me3 they are called poised enhancers [74, 75].

Another important epigenetic mark linked to active transcription is H3K36me3. This modification is present all along the gene bodies. There are several enzymes responsible of this modification, such as Ash1L, NSD1, NSD2, NSD3 and Setd2 [76].

1.2 Polycomb group of proteins and their role on chromatin

1.2.1 Overview of Polycomb proteins

Polycomb group of proteins (PcG) were firstly described in *Drosophila Melanogaster* as master regulators of Hox genes, a set of transcription factors necessary for the correct establishment of proper segmentation along the anteroposterior axis of the body [77, 78]. In particular during *D. Melanogaster* embryonic development, PcG proteins regulate timing and spatial expression of the homeobox genes of the Hox cluster, thereby determining the proper activation of homeotic genes [79].

In particular PcG proteins ensure that Hox genes are kept silent outside of their expression domains, whereas another group of proteins, the Trithorax group (trxG), exert the opposite functional role to maintain these genes actively transcribed in their appropriate domains in accordance to a precise developmental timing [79]. In Drosophila, Hox genes are active during early embryogenesis and are maintained silenced during adult life via epigenetic mechanisms. The regulation of these genes is very important and continues along all the life-span of an organism. Indeed several lines of evidence highlighted that mutations in the Hox genes are linked to several developmental disorders and different kind of cancers. The Hox genes on one hand and the function of PcG and TrxG proteins on the other, as regulators of developmental genes, are strongly conserved in mammals [80].

The first homolog of PcG found in mammals was Bmi1, Psc in Drosophila, for its cooperation in Myc-induced lymphomagenesis [81, 82]. Other mammalian homologs were unraveled in 1997 thanks to functional analyses on murine genes [83]. Functional and biochemical studies showed that these proteins are present in all cell nuclei in mammals and, importantly, in the form of complexes.

These complexes (and their cooperation with TrxG) play fundamental roles in order to assure

the correct embryogenesis, tissue differentiation [84, 85], X-chromosome inactivation [86] and regulation of imprinted genes [87-89]. PcG complexes are also associated with nuclear reprogramming and chromatin remodeling [90].

The PcG complexes were called Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PCR2). These macromolecular complexes contain several Polycomb factors having different and still not fully understood functions. From their discovery, they have been linked to transcriptional repression but growing evidences highlight their possible involvement in transcriptional activation [91, 92].

PRC2 is able to deposit H3K27me1, H3K27me2 and H3K27me3 thanks to HKMT ability of its catalytic subunits Ezh1/Ezh2. These proteins are able to exert their functions through the cooperation with core and accessory proteins that together constitute the PRC2 complex. PRC1 is larger and more diversified with respect to the first complex, and is responsible for the deposition of ubiquitylation on K119 of histone H2A (H2AK119ub). Both activities are linked to chromatin compaction and transcriptional repression.

PcGs activity is regulated in a cell type-specific manner. Indeed, PcGs directly control the expression of cell-type specific gene-sets, contributing to the correct establishment of lineage-specific transcription programs [93].

The two PcG complexes share a large amount of common target genes, showed by a large number of chromatin immunoprecipitation (ChIP) experiments. This view has been recently challenged from different experiments coming from several laboratories which show that near the common sites there are other regions in which is possible to find only component of the PRC1 complex [94]. This became clearer with the discoveries of multiple possible PRC1 complexes.

1.2.2. Polycomb Repressive Complex 1

Among PcG complexes PRC1 is larger and more heterogeneous. There are several biochemically distinct PRC1 subcomplexes [95, 96]. Each of these subcomplexes contains the PRC1 catalytic subunit RING1A or RING1B able to perform monoubiquitylation on lysine 119 on histone H2A (H2AK119ub). On the other hand, they differ one from another for the presence of one specific PCGF subunit. These latter subunits play an essential role since they are crucial for the functionality of RING1 ligases. Indeed, a possible classification in six PRC1 sub-complexes based on which PCGF protein is present has been proposed. Another possible classification is based on the presence in the sub-complex of Polycomb Chromobox homolog (CBX) or Rybp/YAF2 factors. According to this classification, PRC1 complexes can be distinct in canonical and non-canonical. Canonical PRC1 are the ones containing CBX proteins able to recognize the H3K27me3 deposited by PRC2. Another important subunit necessary for the correct organization of this kind of complex and for the activity of RING1 is BMI1 (PCGF4). It can be substituted by Mel18 (PCGF2) even if *in vitro* this last subunit showed no ability to enhance PRC1 activity [97]. So there are the so-defined PRC1-PCGF2/4^{CBX} subcomplexes [92] or PRC1.2 and PRC1.4 [95] (Fig. 2). These complexes agree with the general view of PcG recruitment. PRC1 is recruited to specific chromatin loci through the action of CBX proteins that recognize the H3K27me3 histone mark established by PRC2. This has been pointed out by studies in which the depletion of PRC2 was able to induce a loss of Ring1b at specific common target sites [98].

On the other hand there are other PRC1 target loci where Ring1b and its mark are present despite PRC2 absence. These regions are the ones characterized by the presence of the so-named non-canonical PRC1. These other sub-complexes contain RYBP and its paralogue YAF2 instead of CBX proteins and can contain PCGF1, PCGF2 or PCGF4, PCGF3 or PCGF5, PCGF6 together with other subunits [95].

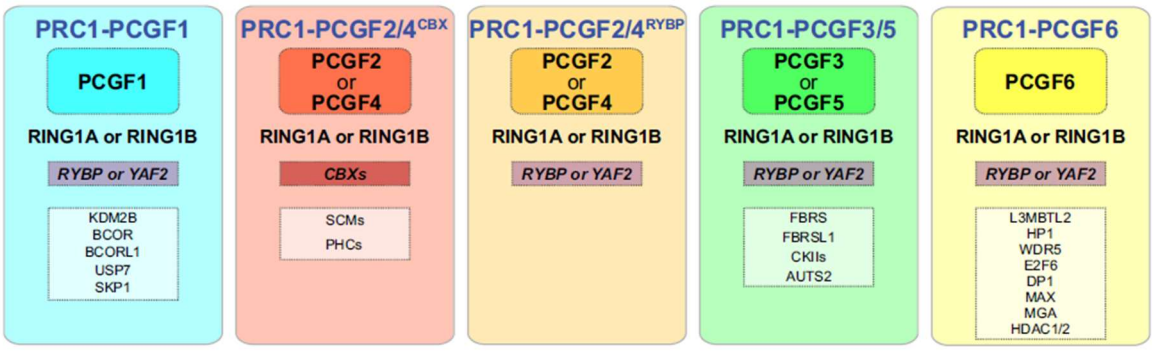


Figure 2: Biochemistry of the different PRC1 complexes.

The assigned nomenclature is from the original publication by Gao *et al*. giving that specific PCGF proteins associate with either CBXs or RYBP/YAF2, defining the functional and biochemical nature of the complexes. Adapted from Scelfo et al. 2015 [92]

RYBP and CBX-containing complexes share anyway a large amount of common targets genes [92, 99, 100]. More precisely, it seems that they bind the same sites in adjacent regions [95]. Moreover it has been shown that PRC1-PCGF2/4 complexes cannot deposit H2AK119ub if forcedly recruited on chromatin [94]. This can suggest that an already deposited H3K27me3 can be crucial to assure PRC1-PCGF2/4 activity and that the deposition of the H2AK119ub is largely dependent on the activity of RYBP/YAF2-containing complexes. The ways through which the PRC1 and the specific sub-complexes localize on specific regions on the chromatin still remain an open question. In this context a novel study uncovered a role for Kdm2b in recruiting PRC1 on chromatin [101]. Kdm2b, is a histone H3K36me3/2 demethylase mainly present in the complex PRC1-PCGF1, which is responsible for the large amount of H2AK119ub present in mESCs [95]. It contains a CXXC domain that has a high affinity towards CpG-rich DNA regions, in line with its localization to CpG-rich promoters [102]. Moreover Kdm2b depletion in mESCs leads to premature differentiation [103] similarly to what happens after *Ring1a/b* loss of function [104]. Also the non-canonical PRC1-PCGF6 has been showed to have a fundamental role in regulating ESC identity [95, 105]. Indeed it has been demonstrated that loss of subunits, L3mbtl2 and Wdr5 in particular, belonging to this complex, cause mESC premature differentiation [106,

107]. However, PRC1-PCGF6 is formed by subunits that are shared with independent complexes. Indeed, Wdr5 is also an essential protein of all COMPASS complexes that control all H3K4 methylation states [108]; Max is the well-known partner of Myc [109]; E2f6, can also associate with members of PRC2 and G9a/GLP complexes during proliferation [110, 111]; Hdac1/2 are partners of several different repressive complexes [112]; L3mbtl2 is found present also in the NuRD complex [107]. Why these proteins are incorporated in this PRC1 sub-complex and their role remain poorly characterized. For example it has been shown is that even if L3mbtl2-PRC1 complex can perform the deposition of H2AK119ubq on recombinant nucleosomes [105], in mESCs the ablation of *L3mbtl2* cause no change in the levels of this modification [107], thus possibly suggesting a specificity in the mark deposition for L3mbtl2. The PRC1.3/5 and are even more poorly characterized. What has been observed is that forcing the recruitment of Pcgf3 and Pcgf5 on the chromatin, this induces the deposition of the modification and PRC2 recruitment [94]. It is overall clear that what is known regarding molecular and biological roles of all these PRC1 sub-complexes is still poor and largely not understood.

1.2.3. Polycomb Repressive Complex 2

Polycomb Repressive Complex 2 is smaller than the other PcG complex and less diversified. This complex is composed of factors maintained conserved from Drosophila to mammals. The core subunits are suppressor of zeste 12 (Suz12), embryonic ectoderm development (Eed), Retinoblastoma associated protein 46/48 (RbAp46/48) and enhancer of zeste homolog (Ezh1 and Ezh2) which represent the catalytic subunits of the complex (Fig. 3). Ezh1/2 are in fact characterized by the SET domain responsible for lysine-methyltransferase activity. The two Ezh1 and Ezh2 proteins can exert the same activity but they are present with cell type specific expression pattern and they have different capabilities in binding chromatin. Each of the core proteins is crucial for the correct activity of the complex both *in*

vivo and *in vitro* [97, 113] and necessary for the correct development of an organism. Indeed it has been demonstrated that mice depleted of Eed, Suz12 and Ezh2 die at early embryo stage [114, 115].

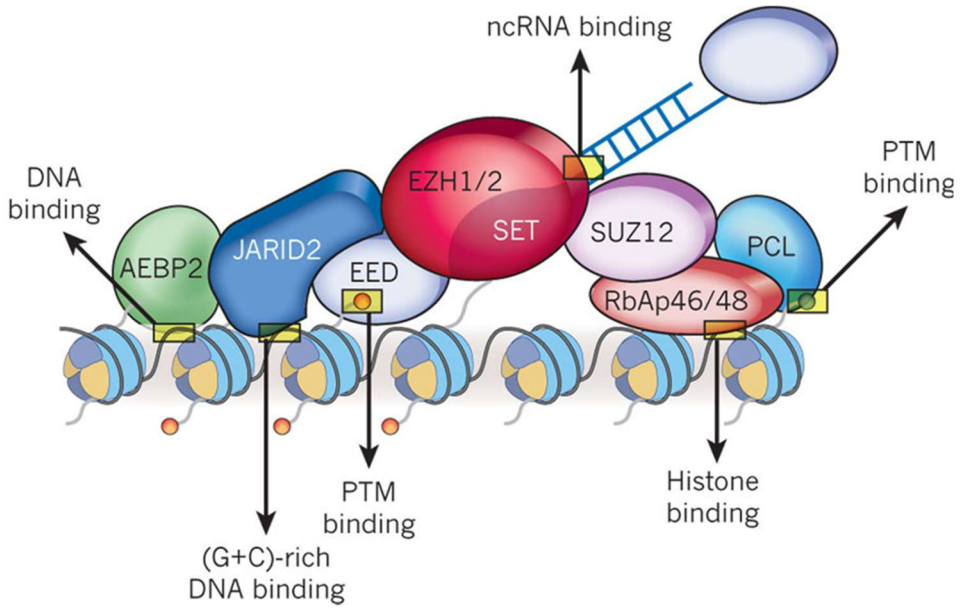


Figure 3: Principal members of PRC2 complex.

The illustration depicts the core members of PRC2, with other proteins and relative binding properties. Adapted from Margueron and Reinberg 2011 [116].

Others non-core proteins are PCL1/2/3, AEBP2, Jarid2. They can increase PRC2 binding, chromatin targeting and PRC2 activity [117].

PRC2 is able to perform mono-, di- and tri-methylation states on H3K27 (H3K27me1, H3K27me2 and H3K27me3). These modifications can be deposited in a stepwise manner going from the mono- to di-methylated, until the trimethylation of H3K27.

Studies *in vitro* analyzing the K_m of these reactions showed that the preferential substrates for PRC2 are most H3K27me0 and H3K27me1 with respect to H3K27me2 [118]. PRC2 is able to achieve different regulatory outcomes across different developmental stages and cell types thanks to the cooperation of accessory proteins. AEBP2 [97] is able to enhance PRC2 activity and colocalize with PcG targets. Also PCL, Polycomb-like Proteins (Pcl1-3)[119, 120], MTF2 [121, 122], PHF19 [123] are tissue specific and interact with EZH2. They are

characterized by a PHD tudor domain able to recognize the active H3K36me3 mark, possibly suggesting a function in silencing active genes [124].

Fundamental factors for Polycomb activity are represented by Jumonji and ARID containing protein Jarid2 found in association with Ezh2 [125, 126]. Jarid2 is important for the establishment of the correct development of mESC and its loss impairs the ability of PRC2 to localize to its specific target genes. Nevertheless, the link between Jarid2 and the H3K27me3 deposition is still controversial. In fact while some authors showed a reduction of this mark upon Jarid2 depletion [125, 127], others found no change nor enhanced mark amount [126, 128]. Since PCL and Jarid2 are mutually exclusive in the binding to the complex it is thought that they can be responsible for the specific targeting of PRC2 to certain genes [124].

PRC2 is responsible for the methylation of about 80% of bulk histone H3 in mESC. Our laboratory has demonstrated that each modification deposited by PRC2 is present in specific loci that are spatially mutually exclusive in mESC [129].

In more detail, we showed that H3K27me1 is localized in the gene bodies of actively transcribed genes where it is found in co-presence also of another transcriptionally active mark, H3K36me3, facilitating the active transcription; H3K27me2 is deposited in intergenic regions, spread in the genome where it exerts protective function preventing firing of non-cell type specific enhancers while H3K27me3 is present on promoters where it exerts repressive regulation of specific target promoters [129].

1.2.4. Polycomb recruitment to target loci

In *Drosophila*, PcG complexes are able to interact with chromatin through *cis*-elements known as “Polycomb response elements” (PREs) [130, 131]. There are numerous PRE in Drosophila genome [132-134] and genome-wide studies showed that PREs are described as

binding sites for both PRC1 and PRC2 subunits but also for other transcriptional factors that can play a role for the PcG binding itself. These analyses allowed the development of an algorithm whose aim was to find eventual novel PREs showing that some predicted elements have PRE proprieties *in vivo* [131]. Despite this, genome wide experiments show a very scarce overlap between PcG binding sites and the predicted elements [135]. Indeed although lots of studies have been made to predict PREs computationally [131, 136], it still is not possible to predict sequence of DNA having a kind of PRE activity.

Regarding PcG recruitment in mammals, since Drosophila DNA binding proteins resulted to be poorly conserved in mammalian cells, the application of such algorithm to mammalians genome fails to predict any potential PRE [131]. For example, in Drosophila PHO is one of the proteins that mediate the recruitment of PRC2 to chromatin. The mammalian homolog is Ying Yang 1 (YY1). Genome-wide analysis demonstrated that YY1 is not able at all to induce a recruitment of PRC2 in fact there are no correlations between the distribution of YY1 and PcGs. Moreover YY1 is devoid of any sequence known to be targeted by the PcGs. Consequently it has been demonstrated that YY1 exerts PcGs-indipendent functions in mESC [137].

For these reasons mechanisms by which PcG complexes are recruited to specific DNA sites are still poorly understood. Genome wide studies have shown that PcG proteins bind preferentially CG rich genomic regions. Nevertheless sequence analysis failed to identify consensus sequences [138] maybe also because PcG proteins can remain bound to their target loci for a very small time frame. Opposite to this, other works have identified mammalian genomic regions with putative PRE behavior [139] supporting a mechanistic conservation between flies and mammals.

Moreover, it has been demonstrated that PcG and TrxG proteins seem to compete in metazoan for the same regulatory pathways and binding sites. Despite the mechanisms of recruitment are poorly understood, deregulation of such equilibrium by either loss or gain of function of specific subunits may result in changes in cell identity that in the end could play

essential roles in pathogenesis. For example it has been shown that also proteins of Trx complexes are able to bind CpGi depositing the functionally opposite mark, H3K4me3, through the association via Zn-Finger-CXXC domain of these proteins. In this way it is possible a competitive binding between Trx and PcG at the same target loci.

Another factor that has been demonstrated to have a role not only at the PRC2 functional level but also for its targeting in mouse ES cells to chromatin is Jarid2 [125]. Despite Jarid2 interaction to PRC2 is not restricted to ES cells, the differences in target genes between cell types suggests tissue specific mechanisms of PcG recruitment [115, 140]. It has been also demonstrated that in absence of prior H3K27me3, PRC2 can be recruited by Jarid2 only if this protein has been previously methylated on its K116 by PRC2 itself.

It has been speculated that pre-existing H3K27me3 is necessary for the maintenance of the mark during replication but the deposition of the mark on new domains during differentiation require methylated Jarid2.

Apart from Jarid2, there are other factors that are able to localize PcGs to specific regions on the chromatin. Regarding PRC1, it has been shown that transcription factor REST is able to interact with CBX proteins inducing the recruitment of the complex independently of the PRC2-deposited mark and of unmethylated CpGi. Moreover Bmi1 has been purified with Runx1/Cbfβ that in turn is able to recruit PRC1 in a PRC2-independent manner. For what concerns PRC2, it has been shown that REST [141], Snail1 [142], PML-RARα [143], PLZF-RARα [144] can physically interact also with this complex. Besides these mechanisms, biochemical interaction studies showed the association of PcGs with long non coding RNA promoting Xist-mediated X-chromosome inactivation [145].

For what concerns the general view, the accepted model is based on the hierarchical recruitment PRC1-PRC2 dependent mechanism [101]. In this view PRC2 is targeted to specific target loci; once there it deposits the H3K27me3 mark that in turn recruit PRC1 because the chromodomain of its CBX proteins are able to bind the mark. This idea came from experiments showing that stable components of PRC1 like the CBX proteins can target

the entire PRC1 directly to chromatin through their ability to bind the H3K27me3 mark [146]. In support of this model, there are genome wide studies that identified subunits of both the complexes occupying the same regions rich in CGs [114, 146, 147].

On the other hand, the discovery of non-canonical PRC1 and more detailed studies on the PRC2 lead to the discovery of novel mechanisms of recruitment that continuously challenged the classical hierarchical one. In support of this new viewpoint it has been demonstrated that in ESCs not every region is co-occupied by both the complexes. In particular, 25% of regions were occupied by Eed and Suz12, components of PRC2 together with PRC1 subunits Phc1 and Ring1b [114]. A more recent work found that around 90% of sites occupied by Ring1b were also bound by Ezh2 and that just 50% of sites occupied by the latter subunit were also bound by Ring1b [148]. Moreover several groups made observations supporting the non-classical hierarchical model. Indeed it has been found that some loci are characterized by the presence of PRC1 and relative mark also in absence of PRC2 [98, 149, 150]. In addition Blackledge and colleagues showed that indeed it was possible to observe an opposite way of recruitment for the two complexes. In fact they demonstrated that variant PRC1 once occupied some loci, are able to induce the recruitment of PRC2. In this experiment they observed a PRC1-dependent PRC2 and H3K27me deposition. In particular they showed that PRC1 complexes containing Pcgf1, Pcgf3 and Pcgf5 after being forcedly recruited to chromatin, are able to greatly deposit H2AK119ub1 with subsequent nucleation of PRC2 complex which then, in turn deposits H3K27me3.

Concomitantly, in the case of PRC1 catalytic portion depletion in mESC, they observed a strong decrease in PRC2 deposition. It has also been previously observed that ESCs depleted of PRC2 do not undergo a change in the levels of the H2AK119ub. In particular it has been demonstrated that the abrogation of the entire H3K27me3 in *Eed* null mouse ES cells affects global Ring1b stability at chromatin but this does not affect the histone H2A lysine 119 mono-ubiquitination (H2AK119ub) levels [98]. These data indicate that, although PRC2 activity is required to stably localize the PRC1 at chromatin, H3K27me3 is generally

indispensable to maintain the global H2AK119ub levels. Moreover, the non-canonical PRC1.1 complex containing Pcgf1, Rybp, Bcor and the demethylase Kdm2b, is recruited to chromatin thanks to the Kdm2b domain that confers the affinity for CpGi. Indeed the demethylase is found associated to almost all CpGi (also to those that are non-PcG targets) to both active and bivalent genes that can become targeted during differentiation.

Finally, in a study the authors through an affinity purification assay using nuclear extract from mouse ESCs and Drosophila embryos showed that PRC2 results more able to bind nucleosomes harboring H2A when modified as H2AK119ub [151].

In conclusion, the recruitment models are more likely not mutually exclusive, acting in cell type specific manner and in synergy to help in addressing proper target modification during differentiation.

1.3. Epigenetics and diseases

Great advances in knowledge and technology allowed reaching extensive comprehension of the epigenetics and of its relevance. Many studies demonstrated the strict correlation between dysregulation of epigenetics-related factors and the onset of different kinds of diseases. In this context, lots of effort demonstrated a correlation between epigenetics and neurological disorders, autoimmune diseases and cancers.

Regarding neurological disorders it has been reported that many of these diseases are characterized by both hypo- and hyper-methylation of DNA. For example, patients affected by Parkinsons's disease overexpress TNFα because of a hypo-methylation of its promoter. Other diseases have been linked to the altered deposition of histone modifications, such as histone hypoacetylation in amyotrophic lateral sclerosis (ALS) patients. These patients are characterized by aggregates of the FUS protein in deposits of misfolded proteins in the cytoplasm. Since FUS factor is able to inhibit CBP binding it, it is responsible of the negative regulation of CREB targets [152, 153]. In this context, high amounts of FUS proteins lead

to histone hypoacetylation. Others examples are Huntington's disease and Friedreich's ataxia in which it is possible to observe a hypertrimethylation of H3K9 [154].

For what concerns autoimmune diseases, one of the best characterized is the immunodeficiency, centromeric instability and facial abnormalities (ICF) syndrome that is due to a heterozygous mutation occurring on DNMT3B [155, 156]. The effect of this mutation is that some DNA regions are found hypo-methylated. Another example is represented by systemic lupus erythematosus in which a correlation has been found between the disease and the alteration in the DNA methylation (hypomethylation) and histone deacetylases [157].

1.3.1 Epigenetics and cancers

Cancer is a heterogeneous disease that involves at the molecular level a multitude of different regulatory pathways and proteins. In general, at malignant stages, cancer cells acquire unlimited replicative potential, angiogenesis capabilities, the ability to evade apoptosis, self-sufficiency in growth signals, insensitivity to growth inhibition and invasive (metastatic) properties [158]. To achieve this, cells undergo a multistep process which gradually leads to the loss of cellular identity towards the acquisition of cancer cell capabilities. In both sporadic and hereditary cancers a large number of genetic alterations have been characterized. Such genetic mutations are required but not fully sufficient for cancer development. Indeed tumors undergo a latency period before the onset of the disease in order to accumulate additional genetic or epigenetic alterations to allow full cancer development [159].

Lots of evidences and observations have linked strictly the role of epigenetics to the development of cancer (Table 3). In particular it became gradually more evident that tumorigenesis can be linked to the distortion of epigenetic programs [160]. These alterations could lead to changes in the genetic expression patterns, with activation of oncogenes [161]

or silencing of tumor suppressor genes [162]. In this sense, epigenetic dysregulation can lead to the tumorigenesis. However, since the epigenetic regulatory machinery is reversible and dynamic, it possesses the potential to either cause or cure diseases, especially cancer itself. For these reasons lots of efforts are being exerted trying to dissect the role of epigenetics and key factors involved in the reversion of the epigenetic programs linked to tumorigenesis development to those typical of the normal condition. This makes epigenetics very promising for the development of new drugs in the future.

Histone regulator class	Epigenetic regulator	Function	Histone modification	Associated cancer	Alteration in cancer
Writer	DNMT1	DNMT	Methyl CpG	Various types	Overexpressed
Writer	DNMT3a	DNMT	Methyl CpG	Various types	Overexpressed
Writer	DNMT3b	DNMT	Methyl CpG	Various types	Overexpressed
Writer	p300	HAT	Multiple lysines	Leukemia, myelodysplasia	Translocation/inactivating mutation
Writer	CBP	HAT	Multiple lysines	Leukemia, myelodysplasia	Translocation/inactivating mutation
Writer	MOZ	HAT	Multiple lysines	Leukemia	Translocation
Writer	MORF	HAT	Multiple lysines	Leukemia	Translocation
Writer	HDAC1-3, 6	HDAC	General	Various types	Overexpressed
Writer	RIZ1	HMT	H3K9	Various types	Down-regulation/mutation
Writer	EZH2	HMT	H3K27	Various types	Overexpressed
Writer	MLL1	HMT	H3K4	Leukemia, lymphoma	Translocation
Writer	SMYD3	HMT	H3K4	Colorectal, hepatocellular carcinoma	Overexpressed
Writer	DOT1L	HMT	H3K79	Leukemia	Deregulated recruitment
Writer	NSD1	HMT	H3K36	Leukemia, hepatocellular carcinoma	Translocation/inactivating mutations
Writer	NSD2/MMSET	HMT	H3K36	Multiple myeloma	Translocation
Writer	NSD3	HMT	H3K36	Breast cancer	Translocation/overexpressed
Eraser	JMJD2C/GASC1	Histone demethylase	H3K9	Various types	Overexpressed
Reader	HP1	Methylated histone-binding protein	H3K9	Breast cancer, melanoma	Down-regulation
Reader	ING1-5	Methylated histone-binding protein	H3K4	Various types	Down-regulation/mutation
Reader	MBD1-4	Methyl-CpG-binding protein	Methyl CpG	Various types	Overexpressed
Reader	MeCP2	Methyl-CpG-binding protein	Methyl CpG	Various types	Overexpressed
Remodeler	INI1	SWI/SNF complex		Malignant rhabdoid tumor	Inactivating mutations
Remodeler	BRM	SWI/SNF complex		Various types	Inactivating mutations
Remodeler	BRG1	SWI/SNF complex		Various types	Inactivating mutations
Remodeler	ATRX	SWI/SNF complex		Myelodysplasia	Inactivating mutations

Table 3: Epigenetic regulators altered in cancer.

Adapted from Muntean and Hess 2009 [163].

1.3.1.1 DNA methylation

Regarding methylation, cancer cells can be characterized both by hyper and hypomethylation of DNA. In particular, a normal cell present overall genome-wide methylation with the exception of CpG islands, which are found unmethylated [164]. These methylation marks can affect the ability of certain transcriptional factors to bind specific target sites and, in the end, the transcriptional rate. Furthermore the methylation sites could

work as docking regions for methylated DNA binding proteins (MBD1, MBD2, MBD3, and Mecp2) that in turn could be bound by other histone modifying enzymes like histone deacetylases (HDACs). This can result in activating or repressing transcription [165].

In tumor cells an overall loss of methylation can lead to the hypomethylation of entire genome with the exception of CpG island promoters, which in contrast undergo hypermethylation [164].

In a normal cell DNA methylation is commonly present at telomeres, centromeres, inactive X-chromosome and at repeated sequences. It is therefore clear that the presence of hypomethylation observed in cancer cells has been linked to an increase in genomic instability inducing chromosomal rearrangements [166]. But this can also cause a reactivation of retrotransposons which translocate to other genomic regions, contributing to the propagation of genomic instability [167]. Another example is the hypomethylation of growth-promoting genes induces their activation. This is the case of Ras and of mammary serine protease inhibitor (MASPIN) for gastric carcinoma [168]. Moreover, hypomethylation can cause loss of imprinting, which is what happens to insulin like growth factor 2 (IGF-2) as seen in Wilms' tumor [169] and colorectal cancer [170].

Regarding the hypermethylation, it has been demonstrated that it spreads all along CpGi that are present at the promoter region. CpGi hypermethylation can induce tumorigenesis by shutting down the expression of tumor suppressor genes. This can be achieved by direct action over tumor suppressor genes but also through indirect silencing. One of the first studies regarding the direct association between hypermethylation and repression of tumor-suppressor genes identified that the hypermethylation of Rb promoter gene at its CpGi (retinoblastoma associated tumor suppressor gene) was able to promote retinoblastoma malignancy. p16 and BRCA1 are other tumor suppressor genes silenced trough the hypermethylation of their promoters [171]. Since these genes are very important for cell adhesion, apoptosis, and angiogenesis, their silencing facilitates the process of tumorigenesis. On the other hand, it is possible to observe an indirect function of CpG

promoter hypermethylation in silencing tumor suppressor genes. This is the case of several genes, such as RUNX3, GATA-4, and GATA-5 responsible respectively for esophageal, colorectal, and gastric cancers, whose silencing is responsible for the downstream target inactivation of the tumor suppressor genes, resulting in the propagation of cancer cells [162, 172]

1.3.1.2 Histone modifications

On the other hand, also the histone PTMs play a role in inducing the disruption of the epigenetic programs occurring during onset and development of cancers. In particular acetylation and methylation have been strongly associated with cancer [173].

1.3.1.2.1 Histone acetylation and cancer

The acetylation status of histones H3 and H4 is largely responsible for the fate of almost all the DNA-based processes such as chromatin assembly, transcription, and gene expression [174].

Deregulation of HAT activity has been reported in both hematological and solid cancers. Generally these alterations cause the inactivation of HAT activity either through gene mutation or by the activity of viral oncoproteins. Regarding the latter case for example, the adenovirus E1A and SV40 T antigen proteins are able to bind p300/CBP [175] wherefore it has been shown to play an important role leading to cellular transformation [176]. The underlying mechanism involves the binding of these viral proteins to p300 and CBP that induces a reduction in the acetylation of H3 lysine 18 (H3K18) [177] with subsequent relocalization of these HATs to the promoter regions of a limited number of genes that promote cell growth and division, favouring their expression [178]. Also chromosomal translocations involving HATs have been related to the onset and progression of acute leukemia [179]. For example this happens in the case of acute myeloid leukemia (AML) and

acute lymphoblastic leukemia (ALL), in which the translocation leads to the production of the fusion protein MLL-CBP. MLL can also be aberrantly fused to p300 in AML through the t(11;22)(q23;p13) translocation [180].

1.3.1.2.2 Histone deacetylation and cancer

The principal role of HDACs is to oppose the activity of HATs and regulate transcription through the removal of acetyl groups from lysine present on histone tails and also from non-histone substrates [181]. These enzymes are characterized by activities whose final effects can be different depending on the context and the cell type in which these activities are exerted. In general, while class I HDACs are mainly involved in cell proliferation and apoptosis, the class II HDACs are responsible for the regulation of cell migration and angiogenesis [182]. It is therefore clear that their deregulation can affect pathways that in the end can have a role in the onset or progression of cancers. Indeed, it has been shown that in an *in vitro* model of colon carcinoma, the knockdown of HDAC2 and HDAC1 is able to suppress the proliferation of these cells [183]. Moreover, knockdown of HDAC3 and HDAC2 induces DNA damage and concomitant apoptosis [184].

Class II and IV HDACs are found predominantly in the cytoplasm and may preferentially deacetylate non-histone proteins [181]. It has been shown that decreasing the expression of HDAC4 causes cell proliferation inhibition and induction of apoptosis [168]. And the knockdown of HDAC6 and HDAC10 was shown to be able to cause the depletion of VEGFR1 and VEGFR2, growth factor receptors implicated in angiogenesis [185].

Deregulation of HDAC activity by chromosomal translocations has been strongly implicated in aberrant gene silencing and the promotion of tumorigenesis. This is true for leukemias. For example, acute promyelocytic leukemia (APL) represents one of the most well-known connections between tumorigenesis and aberrant HDAC activity. The retinoic acid receptor (RAR) is a transcriptional regulator involved in myeloid differentiation through the binding to its heterodimerization partner RXR, which in turn binds to retinoic acid response elements

(RAREs) within the promoters of target genes [181]. In APL, the chromosomal translocations t(15;17) and t(11;17) result in production of fusion proteins RARα-PML (promyelocytic leukemia protein) and RARα-PLZF (promyelocytic zinc finger), respectively. These aberrant proteins maintain the ability to bind RAREs and HDACs with high affinity independently from the presence of retinoic acid. This aberrant process derepresses RAR-targeted genes and prevents cell differentiation [186]. Also some kinds of non-Hodgkin's lymphomas characterized by abnormal expression of the Bcl-6 display an irregular recruitment of HDACs. In this way, genes important for assuring cell cycle arrest and apoptosis are silenced [182, 187]. Also altered expression of individual HDACs in various types of tumor samples has been reported. For example, HDAC1 is overexpressed in prostate, gastric, colon, and breast carcinomas [188, 189]. HDAC2 is overexpressed in colorectal [190], cervical [191], and gastric cancer, whereas overexpression of HDAC6 is observed in breast cancer [192, 193]. Finally also sirtuins can facilitate tumor onset and/or progression. Indeed they regulate transcriptional activity of important proteins such as p53, p73, pRb, NF-κB [194, 195].

1.3.1.2.3 Histone Methylation and Cancer

With the advent of next generation sequencing at the chromatin and RNA level, the analysis of histone methylation marks showed that monomethylation of H3K27, H3K9, H4K20, H3K79, and H2BK5 is associated with gene activation, whereas trimethylation of H3K27, H3K9, and H3K79 is linked to gene repression. The subsequent binding of heterochromatin protein (HP1) to methylated H3K9, for example, promotes the formation of a transcriptional silent heterochromatin state through the association with DNMTs and HDACs [196]. It has been also demonstrated that mice in which HMT SUV39H has been knocked-out became more sensitive to tumorigenesis [197]. Several tumor cells have been reported to present an aberrant expression and deregulation of HMTs and demethylases, such as chromosomal

translocations involving MLL1, NSD1, NSD3, gene overexpression or amplification (i.e., EZH2, MLL2, NSD3, BMI1), gene silencing like RIZ1 and gene deletion MLL3 [192].

1.3.2 Polycomb and cancers

The role of PcGs in cancer is currently one of the most interesting topics and a lot of effort is being invested to try to dissect its contribution to this disease [48, 198]. The biological role of these proteins is becoming more and more controversial [199, 200].

The first evidence of a PcG protein having a direct role in cancer formation was the identification of BMI1 as a proto-oncogene that cooperate with MYC in the formation of B-cell lymphomas [100]. Since then, the attention at role of PcG proteins in human cancer gradually increased. Together with BMI1, the best-characterized PcG protein in human cancer is the catalytic subunit of the PRC2 complex EZH2. This subunit was identified as a direct downstream target of the pRB/E2F pathway and has been shown to be the most frequent over-expressed gene in malignant prostate cancer [201]. Regarding the PRC1 complexes, several and contradicting results have been presented, suggesting that the eventual PcGs oncogenic properties could be due to single subunits rather than the entire complexes. In addition few Polycomb proteins have been proposed as tumor suppressors in specific tumour types [202-204].

To date it is difficult to interpret how PcGs can cause tumorigenesis in haematological malignancies [205, 206]. In fact, for example, loss of Eed in the adult haematopoietic compartment induces long-term HSC exhaustion and pancytopenia [207], while several papers highlighted the presence of inactivating mutations on PRC2 subunits connected to the onset of leukemia and myelodysplastic syndrome (MDS) [208, 209].

Moreover, specific Ezh2 depletion in the haematopoietic system of adult mice is able to cause the onset of T-acute lymphoblastic leukaemias [204].

Furthermore it has been demonstrated the role of PRC2 activity in the development of MLL-AF9 acute myeloid leukaemias (AML) [210]. Indeed while it has been shown that the depletion of PRC2 activity facilitate the onset of MDS and leukaemias induced by ASXL1 mutations [211], it has also been reported that EZH2 loss can prevent the further development of MDS into AML [212].

More in general, it has been long proposed that the principal way through which PcGs favour tumour cell growth is due to their ability to repress the Ink4a/Arf locus, that negatively regulate the cell-cycle progression and for this reason it is a well-known non-cell-type specific tumour suppressor [213-215]. However, we demonstrated that mouse embryonic fibroblast genetically depleted of PRC1 or PRC2 activities became strongly affected in proliferation and transformation capabilities in an Ink4a/Arf-p53-pRb independent fashion [216]. Moreover we showed that PcGs can supervise and favour DNA replication by directly localizing at sites of ongoing DNA replication maybe being involved in the regulation of chromatin dynamics [216]. These observations pave the way to novel possibilities of treating cancer with EZH2 inhibitors despite the functionality of the pRb and p53 pathway, which is found very often inactivated in nearly all human tumours [217, 218].

1.3.2.1 Lymphoma and gain of function EZH2 mutations

Among tumors molecularly characterized by mutations on PcGs, lymphomas seem to be the best candidates for PcG inhibiting compounds as they are characterized by a strong expression of PcG subunits and by high PcG activity [219].

PcGs play important roles during germinal-center (GC) B-cell development. Indeed PRC2 catalytic subunit EZH2 undergoes a very dynamic expression since it results massively upregulated during the rapid proliferation of B cells and immunoglobulin affinity maturation on immune activation, and subsequently decreased once these processes terminate [220, 221]. This makes EZH2 a possible key factor in GC development.

EZH2 is expressed in a several B-cell neoplasms such as in Burkitt's lymphoma, mantle cell lymphomas (MCLs), follicular lymphoma (FL), and diffuse large B-cell lymphomas (DLBCLs) [222]. It is upregulated in MCLs as compared to the normal tissue from which the lymphoma originated [223]. Importantly the deregulation of EZH2 was linked to B-cell lymphomagenesis [222], correlating its expression levels with aggressiveness and unfavorable prognosis.

Recent studies have tried to better understand the role of this subunit in these pathologies. Two groups [220, 224] have independently showed Ezh2 importance for GCs and GC B-cell development using mice depleted for this gene. Indeed, EZH2 represses several genes including negative cell-cycle regulators and also key transcription factor IRF4 and BLIMP1/PRDM1 that are necessary for post GC B-cell development [224].

In 2010 Morin and colleagues characterize firstly a heterozygous EZH2 missense mutations found in B-cell lymphomas of GC origin, such as DLBCLs and FLs [225]. These mutations are due to a single aminoacidic substitution specifically in the catalytic SET domain of the enzyme. The most common mutation is a point mutation on the Tyr641 residue (using a short isoform of the enzyme to count its amino acids) found mutated to either Asparagine (N), Phenylalanine (F), Cysteine (C), Serine (S), or Histidine (H) in about 7% of FLs and 21% of DLBCLs [118, 225, 226]; two other rare EZH2 mutations, A677G and A687V , were reported in about 1%–3% of B-cell lymphoma cases [118, 225-227] and also in 2% of melanomas [228].

The first work that addressed the role of these particular EZH2 mutations considered them as loss of function variations, pointing at a role as tumor suppressor also in lymphomas [225]. Later studies demonstrated instead that they were acting as gain of function mutations, conferring an increased activity of the PcG complex towards the accumulation of H3K27me3 [229].

Several papers after the first one have shown that lymphomas harboring these mutations induce an alteration in the substrate specificity of the complex [118, 229, 230] (Fig. 4).

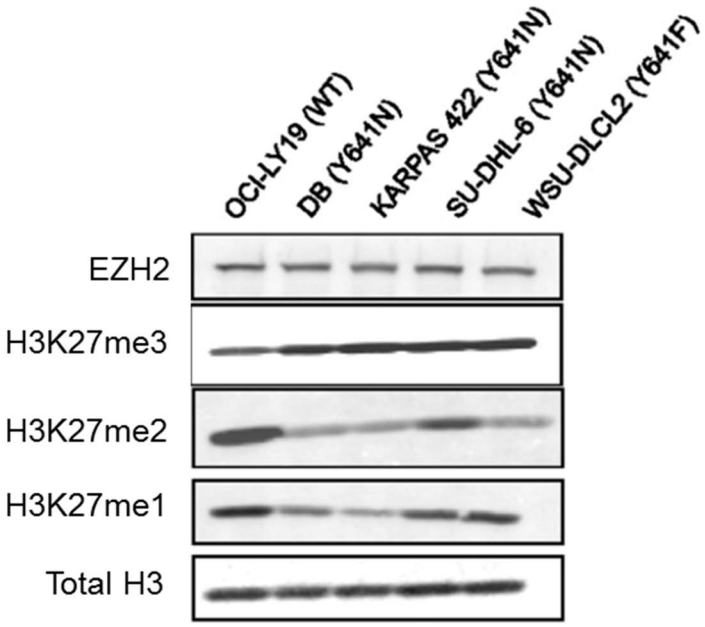


Figure 4: Patterns of H3K27 methylation status in lymphoma cell lines harboring a WT of heterozygous mutant form of EZH2.

Western blot analysis using the reported antibodies to unravel the H3K27 methylation states for lymphoma cell lines containing homozygous WT EZH2 or heterozygous for the indicated EZH2 Y641 mutation. (*Left to right*) (From Sneeringer 2010 [229]).

After the discovery of these mutations, kinetic studies tried to dissect the catalytic efficiency of the different complexes, the WT and the mutant ones, toward unmethylated (H3K27me0), mono-methylated (H3K27me1) and di-methylated (H3K27me2) substrates. These analyses showed that the WT complex can induce mono-, di-, and trimethylation of H3K27 but more in particular it exhibits the greatest catalytic efficiency in converting nonmethylated H3K27me0 to H3K27me1 and a lesser activity for subsequent (H3K27me1 to H3K27me2 and H3K27me2 to H3K27me3) reactions [118, 229-231]. On the other hand they showed that mutant PRC2 complexes bearing these mutations, such as EZH2 harboring the mutation form tyrosine to asparagine ,Y641N, display very limited ability to methylate H3K27me0, but once H3K27 is monomethylated, they can catalyze the turnover of H3K27me1 to H3K27me2 and, then, much more rapidly convert the H3K27me2-to-H3K27me3 (Fig. 5).

Since the WT and mutant complexes are characterized by different enzymatic capabilities, in several paper it has been hypothesized that WT and mutant EZH2 could cooperate to produce the abnormally high level of H3K27me3 seen in the lymphoma cells [118, 229-231].

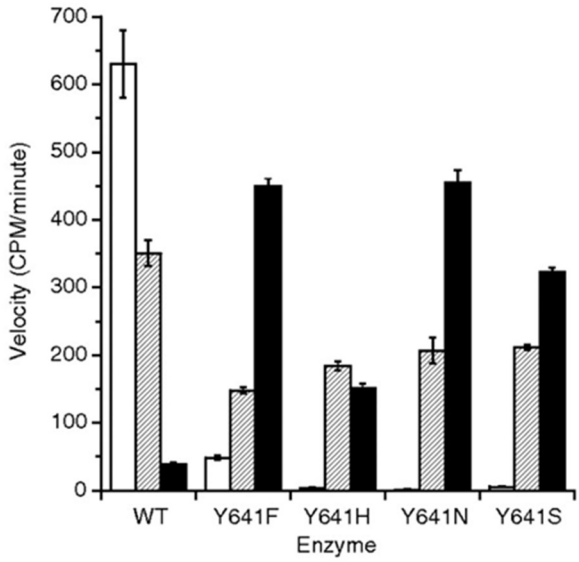


Figure 5: PRC2 complexes containing mutant EZH2 preferentially catalyze di- and trimethylation of histone H3K27.

In vitro methyltransferase activity of mutant and WT complexes on unmethylated peptide (white bars), monomethylated peptide (grey bars), and dimethylated peptide (black bars) (From Sneeringer 2010 [229]).

There are two other rare EZH2 SET domain mutations, A677G and A687V, that cause a change in the substrate specificity of PRC2 even if they are characterized by another different kinetic property [118, 226, 232, 233]. For example, the A677G is equally able to modify all the H3K27 substrates. On the other hand, the A687V mutation induces the formation of a complex able to methylate in the same manner H3K27me1 and H3K27me2 but unable of methylating H3K27me0 [118, 226, 232, 233]. It is therefore clear that all these gain-of-function mutations in lymphomas induce PRC2 hyperactivity using different molecular mechanisms but anyway leading to the same outcome that is the globally elevated H3K27me3 phenotype characterizing lymphoma patient. Some papers more focused their

attention on the structure of the WT and mutants complexes trying to understand the substrate specificity using homology modeling [118, 230, 231, 234]. An important role is played by a hydrogen bond formed by the vicinity between ε-amino lysine nitrogen of the H3K27 and the oxygen of Y641 of the WT complex. This hydrogen bond is necessary in order to recruit H3K27me0 as substrate and furthermore to form a pocket able to include H3K27me2 and reducing its ability, rotating, to accept a third methyl group from the methyl donor. In case of mutant PRC2 complex containing the mutated form EZH2 Y641N, the presence of asparagine induces the formation of a smaller substrate-binding pocket. The final outcome is that H3K27me2 substrate is able to rotate more freely, which accelerates the transfer of an additional methyl group. Similar effects were observed for other EZH2 Y641 mutations (Y641 to F, C, S, or H), but the most drastic effects were seen with EZH2 Y641N [118, 231, 235].

1.3.2.2 EZH2 as therapeutic target

Taken together, all these evidences highlight a possible role of EZH2 enzymatic activity in driving tumorigenesis.

For this reason several laboratories and pharmaceutical companies are investing to find novel promising compounds able to specifically inhibit EZH2 activity. Indeed preclinical results have been collected and human phase 1 trials is under investigation with very promising results.

EZH2 inhibition has been pursued in two different ways. On one hand, the interference with the formation of PRC2 complex; on the other hand the direct targeting of its enzymatic activity.

Regarding the first approach, that is the possibility to act preventing the formation of the PRC2 complex, a peptide called stabilized alpha-helix of EZH2 (SAH-EZH2) whose structure resembles the portion of EZH2 that interacts with EED [236] was developed. SAH-

EZH2 in this way disrupts the EZH2-EED complex, can decrease EZH2 protein levels and consequently inhibits H3K27me3 in a dose-dependent fashion. This peptide has been demonstrated to be effective in both EZH2-mutant lymphoma cells and EZH2-dependent MLL-AF9 leukemia with no effect on nontransformed and EZH2–wild-type controls *in vitro*.

Regarding the second approach, lots of compounds have been produced and analyzed. Firstly 3-deazaneplanocin (DZNep) has been widely used in laboratories also in a human promyelocytic leukemia cell line [237]. This compound is able to repress the activity of SAM–dependent histone lysine methyltransferase activity by increasing the *S*-adenosyl-l-homocysteine hydrolase (SAH). This means that even if DZNep induces significant antitumor activity in various tumor types [238], the induced impairment in histone methylation activity is not specific for EZH2, resulting in toxicity in animal models [239].

Subsequent studies have been aimed at developing compounds more specifically direct toward EZH2 activity. To achieve this purpose, several independent groups performed high-throughput biochemical screens, obtaining different potent inhibitors such as EPZ005687, GSK126, EI1, UNC1999, EPZ-6438.

EPZ005687 has been found able to bind to both WT and Y641mutant EZH2 showing 500-fold selectivity for EZH2 compared to 15 other human methyltransferases and 50-fold over EZH1 [240]. EPZ005687 shows dose-dependent inhibition of H3K27me3 deposition in EZH2 WT and Y641. This molecule is able to inhibit also the A677-mutant EZH2 in lymphoma cells and also in cell lines of breast and prostate cancer.

GSK126 showed to be even more specific, having 1000-fold greater selectivity with respect to other methyltransferases and 150-fold over EZH1 [231, 241]. In addition GSK126 succeeded in impairing the growth of lymphomas *in vivo* [231]. Another promising compound, EI1, is able to bind both WT and mutant form of EZH2 showing selectivity over 10,000-fold with respect to other methyltransferases and 90-fold over EZH1. It is able to affect both H3K27me2 and H3K27me3 and to impair the growth of cells harboring the

mutation but also to induce cell cycle arrest and apoptosis [242]. Many of the drugs described so far have to be frequently administrated.

Another compound, UNC1999, demonstreated to be effective also against EZH1 (10-fold less with respect to EZH2), so it can be used to target EZH2 and EZH1 together.

Finally a compound that entered a phase 1/2 clinical trials is EPZ-6438 to treat patients presenting B cell lymphomas or advanced solid tumors. It presents greater potency and better pharmacokinetics with respect to EPZ005687 [240]. 9 of 15 patients with Non-Hodgkin lymphoma and a patient presenting mutant EZH2 Y646H give back partial or complete responses when treated with this compound (as reported at the following link: http://www.epizyme.com/wp-content/uploads/2014/11/Ribrag-ENA-FINAL.pdf and http://www.epizyme.com/wp-content/uploads/2015/07/ICML-Slides-Presented-062015-v2.pdf) [243].

Taken together, these evidences show that compounds directed against EZH2 could be a promising tool for tumor targeting. In order to refine the design of new drugs for treatment of tumor-affected patients, it would be important to elucidate the mechanism of action through which the mutated form of EZH2 is able to contribute to tumorigenesis.

1.4. Aim of the thesis

My thesis is aimed at the dissection of the role of the mutated PRC2 catalytic subunit EZH2 in the regulation of tumorigenesis. I focused my attention in particular on EZH2-Y641N that is one the most frequently mutations in a particular subset of diffuse large B cell lymphomas and follicular lymphomas. In particular we previously reported that EZH2 is responsible for the establishment of all the methylation states occurring on H3K27 (H3K27me1, H3K27me2, H3K27me3), each of them belonging to specific genomic domains and having specific roles [129]. My experimental hypothesis is that the mutated form of EZH2 can cause an alteration in the domains of deposition of these histone marks, possibly conferring an

oncogenic potential to cells. Here I show experiments aimed at shedding light on the molecular mechanism underlying this kind of mutation.

2. MATERIALS AND METHODS

2.1 Cell culture and manipulation.

2.1.1 Cell lines generation and growing conditions.

Mouse Embryonic Fibroblasts (MEFs) are primary cells derived from mice embryos. These cells are commonly used as proliferating differentiated cells to study a broad range of biological process such as cellular proliferation, cellular senescence, DNA damage response, etc. Conditional alleles are commonly used for inducible knock-out, allowing alleles deletion after cells derivation from the animal (in this case from the embryos). CRE recombinase is the enzyme required to delete the conditional allele, this enzyme could be fused with a mutated ligand binding site of the Estrogen Receptor (ER T2), this leads the Cre recombinase inactive (because is retained in the cytoplasm) until the ligand (4-hydroxytamoxifen) is provided.

MEFs Ink4a/Arf-/- Ezh2 fl/fl derived from 13.5 dpc embryos from Rosa26 CRE-ERT2, Ink4a/Arf knockout mice, used in this work have been described elsewhere [244]. These cells were grown in DMEM medium supplemented with 10% 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) in a CO_2 incubator (5% CO_2) with reduced oxygen tension (3% oxygen). MEFs were passaged every 2 day. The confluent condition consists in maintaining MEF in culture for 10 days subjected only to change medium every two days. To induce CRE-ERT2 nuclear translocation, cells were treated with 500 nM of 4-hydroxytamoxifen (4-OHT, Sigma) dissolved in absolute ethanol (Panreac). In starvation condition MEFs have been cultured with 1% or 0.1% of 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).

MEF 3T3 fibroblasts expressing a conditional Myc-oestrogen receptor chimaera ($3T3^{MycER}$) immortalized following the 3T3 protocol, were cultured in DMEM medium supplemented with 10% fetal calf serum (Euroclone), 2 mM glutamine (Gibco) and 100 U/ml penicillin and 0.1 mg/ml streptomycin (Gibco), 0.1 mM non-essential aminoacids (Gibco), 1 mM Na-Pyruvate (Gibco) and maintained in an incubator at normoxia condition (21% oxygen). MEFs were passaged every 2 day. To induce nuclear translocation, cells were treated with 20nM and 100nM for 24 and 48 hours of 4-hydroxytamoxifen (4-OHT, Sigma) dissolved in absolute ethanol (Panreac). These cells were provided by our collaborator in Bruno Amati and Stefano Campaner group.

OKSM MEFs have been described elsewhere [245, 246]. These "reprogrammable MEFs", were derived from mice carrying a double knock in of the four transgenes in the 3'-UTR of the Col1a1 locus under the transcriptional control of a doxycycline-responsive promoter. In these mice, also, the reverse tetracycline-controlled transactivator (M2-rtTA) is expressed from the Rosa26 locus, allowing for induction of expression upon doxycycline administration [245, 246]. They were grown in DMEM medium supplemented with 10% fetal bovine serum (Euroclone), non-essential amino acids (Gibco), 1% glutamine (Gibco), sodium pyruvate (Gibco) and 1% penicillin/streptomycin (Gibco), in a CO2 incubator (5% CO2) with reduced oxygen tension (3% oxygen). MEFs were passaged every 2 day. These cells were provided by our collaborator in Giuseppe Testa's group.

OCI-LY7: B cell lymphoma established from the peripheral blood sample of a 48-year-old man with B-cell non-Hodgkin lymphoma (B-NHL, diffuse large B-cell lymphoma, DLBCL) carrying the t(8;14) translocation inducing the fusion MYC-IGH and a p53 point mutation. These cells were grown in suspension in IMDM medium supplemented with 15% fetal bovine serum (Euroclone), non-essential amino acids (Gibco), 1% glutamine (Gibco),

sodium pyruvate (Gibco) and 1% penicillin/streptomycin (Gibco), in normoxia incubator (21% CO2).

WSU-DLCL2: B cell lymphoma derived from the pleural effusion of a 41-year-old Caucasian man with B-cell non-Hodgkin lymphoma (B-NHL, diffuse large cell lymphoma) carrying EZH2 Y641F mutation. These cells were grown in suspension in IMDM medium supplemented with 15% fetal bovine serum (Euroclone), non-essential amino acids (Gibco), 1% glutamine (Gibco), sodium pyruvate (Gibco) and 1% penicillin/streptomycin (Gibco), in normoxia incubator (21% CO2).

SU-DHL-6: B cell lymphoma derived from the peritoneal effusion of a 43-year-old man with B-cell non-Hodgkin lymphoma (B-NHL) carrying EZH2 Y641N mutation; assigned to GCB-like lymphoma subtype (germinal center B-cell). These cells were grown in suspension in IMDM medium supplemented with 15% fetal bovine serum (Euroclone), non-essential amino acids (Gibco), 1% glutamine (Gibco), sodium pyruvate (Gibco) and 1% penicillin/streptomycin (Gibco), in normoxia incubator (21% CO2).

All lymphoma cells were provided by Stefano Casola's group.

293T cells were grown in DMEM medium supplemented with 10% fetal bovine serum (Euroclone), non-essential aminoacids (Gibco), sodium pyruvate (Gibco) and penicillin/streptomycin (Gibco), in a CO2 incubator (5% CO2) with standard oxygen tension (21% oxygen). These cells have been used for the virus production to infect MEFs.

2.1.2 Transfection

It is a technique that allows the introduction of DNA into cells. It is performed creating Calcium Phosphate-DNA co-precipitates that permeate the cell membrane to permit the entry of material. The Calcium Phosphate procedure is based on the combination of HBS 2X (HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) 21 mM, 0,7 mM Na2HPO4, 137 mM NaCl, 5 mM KCl, 6 mM Dextrose pH 7,10) with a mix containing

calcium chloride (2.5M) and the DNA that has to be delivered. HEPES and calcium chloride form a precipitate that binds the DNA.

Transfection protocol can be exploited to make lentiviral transductions. In this case the transfection was performed to obtain medium containing lentiviral particles suitable to transduce 293T cell line.

Transfection was set up using a mix of 8-10 µg of DNA, 6 µg of psPAX2, 3 µg of pMD2.G. psPAX2 is a packaging plasmid for the production of viral particles. It has a CAG promoter for the expression of Gag, Pol, RRE, REV and TAT. pMD2.G is a VSVG envelope expressing plasmid. It has a SV40 promoter.

The mix of these DNA was combined with 0.125M CaCl2 and free nuclease water to a final volume of 500 µl. The mix was then vortexed adding 500 µl of 2X HBS drop-by-drop and finally added to cell plates.

2.1.3 Lentiviral infection

It is a process by which exogenous DNA is transduced into host cells by a virus. MEFs cells were infected using different lentiviral vectors.

In this process the medium of the transfected cells that contains the virus was harvested and centrifuged in order to discard the eventual pellet formed by detached cells and filtrated using 0,45 uM filters. The virus-containing medium was then added to the plates where the cells that have to be infected were seeded in presence of 5 µg/ml polybrene to increase transduction efficiency. A cycle of transduction was repeated after about 24h. To select transduced cells puromycin (2ug/mL) was added to cell plates.

2.1.3 Lentiviral vectors

The used lentiviral vectors are: p-lenti EF1 expressing WT or mutated Y641F or Y641N

EZH2, p-lenti pCDH-EF1-MCS-BGH-PGK-GFP-T2A-Puro cDNA Cloning and Expression Vector that has been bought from SBI contain a GFP cassette with PGK promoter downstream of EF1α strong promoter and a multiple cloning site (MCS) that offer the possibility to clone whatever cDNA of interest (indeed I will use it to clone EZH2 cDNA to repeat the reprogramming experiments using the same constructs for every cell lines).

pEF1-EZH2-ZSgreen has been obtained cloning the EZH2 cDNA devoid of its own ATG downstream of EF1α promoter. For the cDNA of WT EZH2 has been used a short isoform of the enzyme to count its amino acids. The same cDNA has been modified on the 641 amino acids in order to get the point mutation from Tyrosine to Phenylalanine (F) or Asparagine (N). The protein is fused with a Zs green cassette and an Internal Ribosome Entry Site (IRES) that allows the expression of the Zsgreen. These vectors were provided by our collaborator in Stefano Casola' s group.

p-lenti for MEF Ink4a/Arf -/- Ezh2 fl/fl has been obtained cloning the EZH2 cDNA devoid of its own ATG downstream of EF1α promoter. For the cDNA of WT EZH2 has been used a short isoform of the enzyme to count its amino acids. The same cDNA has been modified on the 641 amino acids in order to get the point mutation from Tyrosine to phenylalanine (F) or Asparagine (N). The protein is fused with a FLAG-tag and an Internal Ribosome Entry Site (IRES) that allows the expression of the PuroR gene. These vectors were provided by our collaborator in Stefano Casola' s group.

2.2 Techniques used for protein detection

2.2.1 Immunoblot analysis

This method, commonly known as Western blot analysis (WB), allows the detection of protein of interest and relative post translational modifications in a protein extract. Protein extracts were obtained from cell pellets after lysis with high salt lysis buffer (20mM Tris-HCl, pH 7.6, 300mM NaCl, 10% glycerol, 0.2% (v/v)) Igepal (Sigma-Aldrich, cat. CA 630),

incubating for 30 minutes on ice, sonicated with Bioruptor Diagenode (B435) with 10 cycles of pulse 30'' ON+30" OFF and centrifuging at 10000 x g for 30' at 4°C, and recovering the supernatant. Total extracted proteins were quantified with Bradford assay (Bio- Rad, cat.30500-0006). Then Laemmli sample buffer was added and samples boiled 10 minutes at 95 °C. Usually 30-40 µg of protein extracts were loaded onto each lane of acrylamide/bisacrylamide gel, and a sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed. Gel-separated proteins were transferred to a Protran Nitrocellulose Membrane (Whatman), one hour and 20 minutes at 4 °C, at 100 Volts. Membranes were blocked with a solution of 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 one hour 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 (enhanced chemiluminenscence, Amersham) or ECL-plus (Amersham).

Detection of histone PTMs were performed with the same method.

2.2.2 Immunofluorescence.

This is a standard technique that allows to spatially locating in single cells the antigens you are interested in (such as proteins or nucleotide analogue in genomic DNA in our case). Moreover using the confocal microscope it is possible to test the co-localization of 2 or more antigens on the same focal plane with a resolution of 200-400nm.

Indicated MEFs lines were seeded on 0.1% gelatinized glass coverslips, and nuclei were eventually prepared by treatment with cold pre-extraction buffer (10 mM Tris HCl pH 7,6, 100 mM NaCl, 2 mM $MgCl_2$, 0,3 M sucrose and 0,25% Igepal) for 10 min at 4°C. Whole

cells or nuclei were fixed at -20°C with 100% methanol for 10 min. Then, fixed cells were incubated with primary antibodies diluted in 0.1% tween-TBS for 1 hour at RT, washed and incubated with fluorophore-conjugated secondary antibodies. Nuclei were counterstained with DAPI and embedded in anti-fade containing glycerol (DABCO). Images were acquired using a Leica SP2 confocal microscope. Mender's co-localization coefficient was calculated on the entire Z-stacks images using the jacop tool of the Image J software.

2.2.3 Mass Spectrometry analysis

Analysis of proteins isolated from tandem purifications was done by mass spectrometry (MS), an analytical technique generating spectra of the masses of the atoms or molecules present in a sample of material. The sample is ionized by an ion source, and then ions are separated on the basis of their mass/charge ratio, and detected by the detector component which converts received ions into spectra of the relative abundance of ions as a function of the mass-to-charge ratio. The atoms or molecules can be identified by correlating known masses to the identified masses. To have higher resolution, two steps in mass analysis are made (tandem mass spectrometry or MS/MS). In our proteomic approach, briefly, proteins were separated by gel electrophoresis and enzymatically in gel-digested, producing a peptides mixture which is separated by liquid chromatography before its injection into the mass spectrometer. Protein digestion produces peptides earning each one a typical spectrum; in addition, some peptides are produced uniquely by digestion of specific proteins, therefore indicating without any doubt the presence of such proteins in the sample. Combining these features with computational tools, we were able to match peptide spectra with the proteins to which they belong.

To achieve this, eluted proteins obtained by tandem purifications were separated by 1D SDS–PAGE, using 4–12% NuPAGE Novex Bis–Tris gels (Invitrogen, cat. NP0321BOX) and NuPAGE® MES SDS running buffer (Invitrogen, cat. NP0002) according to

manufacturer's instructions; gel was stained with Coomassie Blue using a Colloidal Blue Staining Kit (Invitrogen, Cat. LC6025). Samples were digested with trypsin (Promega). The gel bands were cut and then washed four times with 50mM ammonium bicarbonate, 50% ethanol and incubated with 10mM DTT in 50mM ammonium bicarbonate for 1 h at 56°C for protein reduction. Alkylation step was performed incubating the sample with 55mM iodoacetamide in 50mM ammonium bicarbonate for 1 h at 25°C in the dark. Gel pieces were washed two times with a 50mM ammonium bicarbonate, 50% acetonitrile solution, dehydrated with 100% ethanol and dried in a vacuum concentrator. Digestion was performed using 12.5 ng/ml trypsin in 50mM ammonium bicarbonate and incubated for 16 h at 37°C for protein digestion. Supernatant was transferred to fresh tube, and the remaining peptides were extracted by incubating gel pieces two times with 30% acetonitrile (MeCN) in 3% trifluoroacetic acid (TFA), followed by dehydration with 100% acetonitrile. The extracts were combined, reduced in volume in a vacuum concentrator, desalted and concentrated using RP-C18 StageTip columns and the eluted peptides used for mass spectrometric analysis.

Peptide mixtures were separated by nano-LC/MSMS using an Agilent 1100 Series nanoflow LC system (Agilent Technologies), interfaced to a 7-Tesla LTQ-FT-Ultra mass spectrometer (ThermoFisher Scientific, Bremen, Germany). The nanoliter flow LC was operated in one column set-up with a 15-cm analytical column (75 µm inner diameter, 350 mm outer diameter) packed with C18 resin (ReproSil, Pur C18AQ 3 mm, Dr Maisch, Germany). Solvent A was 0.1% FA and 5% ACN in ddH2 O and Solvent B was 95% CAN with 0.1% FA. Samples were injected in an aqueous 0.1% TFA solution at a flow rate of 500 nl/min. Peptides were separated over a gradient of 0–40% Solvent B over 90 min followed by a gradient of 40–60% for 10 min and 60–80% over 5 min at a flow rate of 250 nl/min. The mass spectrometer was operated in a data-dependent mode to automatically switch between MS and MS/MS acquisition. In the LTQ-FT full scan MS spectra were acquired in a range of m/z 300–1700 by FTICR with resolution r =100 000 at m/z 400 with a target value of 1

000 000. The five most intense ions were isolated for fragmentation in the linear ion trap at a target value of 5000. The nanoelectrospray ion source (Proxeon, Odense, Denmark) was used with a spray voltage of 2.4 kV. Sing wide band activation mode was 35%. Normalized collision energy was set to 35%, and activation time to 10 ms; spray voltage, 2.2 kV; no sheath and auxiliary gas flow; heated capillary temperature, 275ºC; predictive automatic gain control (pAGC) enabled, and an S-lens RF level of 65%. For all full-scan measurements with the Orbitrap detector, a lock mass ion from ambient air (m/z 445.120024) was used as an internal calibrant as described previously [247].

2.2.4 Acidic extraction of histones and Immunopurification

Ink4a/Arf -/- Ezh2 fl/fl MEFs were harvested, resuspended in N-Buffer (15 mM HEPES, pH 7.5, 10% sucrose, 0.5% Triton X-100, 0.5 mM EGTA, 60 mM KCl, 15 mM NaCl, 30 ug/ml Spermine, 30 ug/ml Spermidine, 1 mM DTT, 3 mM NaButyrate, 5 mM NaF, 5 mM Na-Pyrophosphate, 5 mM ß-glycerophosphate) with fresh addition of a protease inhibitor cocktail (Calbiochem) at a cell density of 125x106 cells/ml and lysed 10 min at 4ºC with gentle stirring. Lysate was centrifuged onto a 10% sucrose cushion (107 cells/cushion) at 4000 rpm for 30 min and washed twice in ice cold 1x PBS. Histones were extracted in 0.4 N HCl overnight at 4ºC with gentle stirring, extensively dialyzed in 0.1 M acetic acid and dried out. Histone pellets were re-suspended in water and quantified with Bradford assay and denature in Laemmli sample buffer (Invitrogen).

2.2.5 In-gel digestion of histones for MS analysis

Histones were separated on 1D SDS-PAGE, using 4–12% NuPAGE® Novex Bis–Tris gels (Invitrogen) and NuPAGE® MES SDS running buffer (Invitrogen) according to manufacturer's instructions. The gel was stained with Coomassie Blue using Colloidal Blue Staining Kit (Invitrogen). Protein band corresponding to histone H3 was excised from the

gel, de-stained with 50% acetonitrile (MeCN) diluted in H_2O to be then chemically alkylated by incubation with acetic anhydride-D6 (Sigma 175641) 1:9 ratio in 1M NH_4HCO_3 as previously described [248]. After 3h at 37 °C with strong shaking (1400 rpm), gel pieces were washed by increasing concentration of MeCN. In-gel digestion was performed with 7.5 ng/µl trypsin (Promega V5113) in 50 mM NH_4HCO_3 at 37 °C overnight. Supernatant was transferred to fresh tube, and the remaining peptides were extracted by incubating gel pieces two times with 30% MeCN in 3% trifluoroacetic acid (TFA), followed by dehydration with 100% MeCN. The extracts were pooled, reduced in volume in a vacuum concentrator, desalted and concentrated using a combination of in house-made RP-C18/Carbon and a strong cation exchange (SCX) solid phase extraction (SPE) StageTip [249]. Briefly, digested peptides loaded on combined RP-C18/Carbon and SCX StageTip were eluted with high organic solvent (80% MeCN) and NH_4OH, respectively. Eluted peptides were lyophilized, re-suspended in 0.1% TFA and 0.5% acetic acid in ddH_2O, pooled and subjected to liquid chromatography and tandem mass spectrometry (LC-MS/MS).

2.2.6 Data analysis for histone PTMs MS/MS

The mass spectrometric raw data were analyzed with the MaxQuant software (version 1.1.1.25) [250, 251]. A false discovery rate (FDR) of 0.01 for proteins and peptides and a minimum peptide length of 6 amino acids were required. In order to improve mass accuracy of the precursor ions, the time-dependent recalibration algorithm of MaxQuant was used. The MS/MS spectra were searched by Andromeda engine against the IPI human database (containing 87,061 entries) combined with 262 common contaminants and concatenated with the reversed versions of all sequences [252]. Enzyme specificity was set to Arg-C and maximum of three missed cleavages were allowed. Peptide identification was based on a search with an initial mass deviation of the precursor ion of up to 7 ppm. The fragment mass tolerance was set to 20 ppm on the m/z scale. Variable modifications included: deuterate

acetylation (D3-acetylation) (+45.0294 Da) on Lysine, Lysine mono-methylation (calculated as the sum of the masses of D3-acetylation (+45.0294) and mono-methylation (+14.016 Da)), dimethylation (+ 28.031 Da) and tri-methylation (42.046 Da), Lysine acetylation (+ 42.010 Da), Methionine oxidation (+ 15.995 Da) and N-terminal protein acetylation. D3-acetyl chemical alkylation results in a delta mass of 45.0294 Da for each group added either to the unmodified or mono-methylated Lysine, allowing the discrimination of isobaric modified peptides. Output table from MaxQuant were filtered with the following criteria: peptides with a low score (cut-off score value, 60) [253] and with more than 5 putative PTMs per peptide were removed. Redundant peptides were filtered so that only the peptide with the highest Andromeda score among peptides with the same identification was included. The filtered data were then subjected to manual validation using Qual Browser version 2.2 (ThermoFisher Scientific). Estimation of relative species percentage (RS%) calculated Extracted ion chromatograms (XIC) were constructed for precursor ions with mass tolerance of 10 ppm and mass precision up to 4 decimal places using Qual Browser version 2.2. Peak areas for both unmodified and modified peptide species were measured within the same retention time interval.

Relative species percentage (RS%) for (27-40) peptide derived from MEF Ink4a/rf, Ezh2 fl/fl transduced with empty vector or overexpressing EZH2 WT or EZH2 Y641N were calculated dividing the peak area relative to each peptide species divided by the sum of peak areas for all peptide species sharing the same amino acid sequence. Specific delta masses relative to modified sequence of (27-40) peptide from H3 was further included in Andromeda configuration module (AndromedaConfig.exe).

2.2.7 Cell cycle analysis with FACS

Indicated MEFs lines were subjected to cell cycle analysis by means of Fluorescence Activated Cell Sorting (FACS) taking advantage of specific dye intercalating the DNA (Propidium Iodide) and nucleotide analogue (BrdU) previously pulsed into cells.

Cell events were acquired at FACSCalibur flow cytometer (BD Biosciences) and analyzed with FlowJo 8.5.3 software (Tree Star). For cell cycle analysis, MEF Ink4a/rf, Ezh2 fl/fl were cultured as previously described and pulsed with 33 μM 5-bromo-2-deoxyuridine (BrdU) for 45 min. Cells were washed with PBS, fixed with 75% ethanol, denatured with 2N HCl and neutralized by 0.1 M Borax. Cell staining was performed in blocking buffer solution using mouse anti-BrdU as primary antibodies; anti-mouse IgG-FITC conjugate was used as secondary antibodies. Cells were finally re-suspended in PBS containing 2.5 ug/mL Propidium Iodide in presence of RNase A and analyzed with FACSCalibur flow cytometer. Collected data were analyzed with Cell Quest Pro (BD Biosciences) and the single cell fluorescence intensity values of the gated populations were retrieved using FCS Assistant software.

2.3 Assays for detection of DNA modifications and protein binding to DNA.

2.3.1 Chromatin Immunoprecipitation (ChIP)

ChIP (chromatin immunoprecipitation) technique is used to investigate protein-DNA interaction studies. The fundamental principle is the cross-linking between DNA and DNA-associated proteins that can be achieved by treating (“fixing”) cells with formaldehyde or UV rays. Cross-linked chromatin is sheared by sonication to generate fragments of 300 - 1000 base pairs (bp) in length. Through immunoprecipitation, proteins of interest coupled to DNA are isolated by means of antibodies. 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. Briefly, 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 ChIP on H3K27me2 and H3K27me3 modifications and of 500-1000 bp to make proteins ChIP and immunoprecipitated ON with 5 µg of indicated antibodies per milligram of sonicated chromatin. Then protein A sepharose beads (GE Healthcare, cat. 170780-01; 30 µl slurry per milligram of sonicated chromatin; 1 mg of chromatin was used for each protein precipitation; 500 ug were used for histone PTMs) were added and incubated 2 h at 4°C, followed by three washes with "low salt" wash buffer (20 mM TRIS/HCl pH8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% TRITON X-100), and one in "high salt" wash buffer (20 mM TRIS/HCl pH8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% TRITON X-100) were performed. De-crosslinking was made at 65°C ON. Eluted DNA was purified using QIAquick PCR purification kit (Qiagen).

2.3.2 ChIP Rx: Reference exogenous genome ChIP

Briefly, 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 ChIP on H3K27me2 and H3K27me3 modifications. Then after to have added 5% of drosophila S2 cells chromatin to the samples subjected to immunoprecipitation, the 1% input is taken. Subsequently ChIP is performed overnight with 5 µg of indicated antibodies per milligram of sonicated chromatin. The following day protein A sepharose beads (GE Healthcare, cat. 170780-01; 30 µl slurry per 500 ug chromatin were used for histone PTMs) were added and incubated 2 h at 4°C,

followed by three washes with “low salt” wash buffer (20 mM TRIS/HCl pH8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% TRITON X-100), and one in “high salt” wash buffer (20 mM TRIS/HCl pH8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% TRITON X-100) were performed. De-crosslinking was made at 65°C overnight. Eluted DNA was purified using QIAquick PCR purification kit (Qiagen).

2.3.3 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 our internal genomic facility. DNA 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.

2.3.4 ChIP-Quantitative Real Time PCR (RT-qPCR)

The real-time quantitative polymerase chain reaction (qPCR), allows the detection and relative quantification of a specific DNA sequence in a sample. Unspecific fluorescent dye (SYBR green) intercalates with double-strand DNA, specifically amplified through PCR made with specific oligonucleotide probes (primers). 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 (such as target genes in a ChIP experiment), or the abundance of a cDNA derived from an RNA sample. All the qPCRs were carried out using Fast SYBR Green (Applied Biosystem) as dye. Lists of primers used are available in Table 4.

2.3.5 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 favouring 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. Irrespective of modification, normalized reads with or without normalization factor for their respective sample within each bin were computed and represented as heatmap.

2.4 Methods for RNA analysis

2.4.1 RNA sequencing (RNA-seq)

High throughput RNA sequencing is a technique allowing the whole transcriptome profiling of a cell population. In our cases, a selection for poly adenilated mRNA was performed. Total RNAs from indicated MEFs lines were extracted using the Qiagen RNeasy Plus RNA extraction kit. Retrieved RNA was checked for integrity on a Bioanalyzer instrument by picoRNA Chip (Agilent) and then converted into libraries of double stranded cDNA suitable for next generation sequencing on the Illumina platform. At this purpose, the Illumina TruSeq v.2 RNA Sample Preparation Kit was used following manufacturer's recommendations. Briefly, 5 µg of total RNA were subjected to mRNA purification by means of poly-T oligo-attached magnetic beads, then fragmentation was performed exploiting divalent cations contained in the Illumina fragmentation buffer and high temperature. First, cDNA strand was synthesized with random oligos by Reverse Transcriptase SuperScript III (Invitrogen). Second, cDNA strand synthesis was performed by DNA polymerase I and Rnase H. Then, DNA fragments were blunt ended and adenylated

at 3' extremities before ligating specific Illumina oligonucleotides adapters. Resulting fragments were enriched performing 15 cycles of PCR reaction using proprietary Illumina primers mix. Prepared libraries were quality checked and quantified using Agilent high sensitivity DNA assay on a Bioanalizer 2100 instrument (Agilent Technologies).

2.4.2 RNA sequencing data analysis.

Sequencing data was aligned to mouse reference genome (mm9) using Tophat (PMID 23618408). Differentially expressed genes (DEGs) were identified using DESeq (PMID 20979621). We regarded genes as differentially expressed with fold change greater than or equal to 3 and with adjusted p-value less than 0.05.

2.4.3 Real Time quantitative PCR

Total RNA was extracted with the Qiagen RNeasy Plus RNA extraction kit. The RNA is subsequently retro-transcribed usingThe ImProm-II™ Reverse Transcription System that is a kit that allows obtaining first-strand cDNA in preparation for PCR amplification according to the manufacturer's instructions. Reverse-transcrption real-time quantitative PCR (RT-qPCR) was carried out using Fast Sybergreen as previously described. Primers are listed in Table 5.

2.5 Reprogramming

2.5.1 Infection and selection of OKSM MEFs

OKSM MEFs were infected as in 3.1.3 with either the Empty vector (expressing also the GFP) or vectors expressing WT or mutant EZH2 (expressing also the ZsGreen protein), with two independent vector productions for each EZH2 condition. After 5 days cells were sorted

on a BD influx (BD) on the basis of GFP or ZsGreen expression, selecting cells that fluoresced at least 1 log over the background (uninfected MEFs were used as negative control).

2.5.2 Reprogramming experiment

Plates were incubated with gelatin 0.1% for 30' at 37°C before mytomycin-C inactivated MEFs were plated (333.000 cells/well of a six-well plate) in MEF medium. The day after OKSM MEFs were plated on top of this feeder-layer (5000 cells/well of a six well plate) in MEF medium. In an additional well we plated the same amount of cells without feeder-layer was used, in order to perform count at the day of induction. The day after, after counting cells with a burker chamber, we induced reprogramming by incubating cells with iPSC medium. Doxycicline (2 µg/mL) was added in all conditions except for the UT control (untreated), and medium was changed every day until day 14.

miPSC medium			
Product	**Cat. Nr.**	**Producer**	**Final mix (600 ml)**
Advanced D-MEM/F-12 (1X) liquid	12634010	Thermo Fisher	500 mL
Knockout™ Serum Replacement	10828028	Thermo Fisher	100 mL
Sodium Pyruvate MEM 100 mM, liquid	11360039	Thermo Fisher	5 mL
MEM Non Essential Amino Acids (100X), liquid without L-Glutamine.	11140035	Thermo Fisher	5 mL
Penicillin-Streptomycin, liquid	15140122	Thermo Fisher	5 mL
L-Glutamine 200 mM (100X), liquid	25030-024	Thermo Fisher	5 mL
2-Mercaptoethanol, 50 mM (1000X)	31350-010	Thermo Fisher	0.5 mL
mLIF	n.a.	Transgenic Facility	1.2 mL

2.5.3 Assessment of GFP+ colonies

One well from each condition was analyzed under a fluorescence microscope (EVOS, Thermo Fisher Scientific) to assess the fraction of GFP+ colonies. Colonies were manually counted and the fraction was assessed as the number of GFP+ colonies on the total of colonies in the well.

2.5.4 Alkaline phosphatase staining

A solution made of Citrate, Acetone and Formaldehyde was used to fix cells for 45 seconds after a wash in DPBS. Cells were rinsed in ddH2O for 30 seconds and the staining solution (Vector RED Alkaline phosphatase substrate kit, Vector Labs) was added to cells for 30 minutes. Cells were rinsed in ddH_2O and imaged for the presence of a red precipitate. Colonies were counted manually and the efficiency of reprogramming was computed by dividing the number of alkaline phosphatase positive colonies by the number of counted cells the day of induction.

2.6 Antibodies

2.6.1 Antibodies used for Immunoblot and immunoprecipitation analyses

For Western blots the following antibodies were used: mouse anti-Vinculin (Sigma-Aldrich, Cat. V9131), mouse anti-Eed [254], goat anti-Suz12 (Santa Cruz, Cat. sc-46264), rabbit anti-FLAG (Sigma, Cat. F7425), goat anti-Lamin B (Santa Cruz, Cat. sc6216), mouse anti-Ezh2 [254].For histone PTMs detection of the following antibodies were used: rabbit anti-H3 (Abcam, Cat. 1791), mouse anti-H3K27me1 (Active Motif, Cat. 61015), rabbit anti-H3K27me2 (Cell Signaling, Cat. 9728), rabbit anti-H3K27me3 (Cell signaling, Cat. 9733).

2.6.2 Antibodies used for ChIP analyses

In ChIP experiments, the following antibodies were used: rabbit anti-Suz12 (Cell signaling, Cat. 3737), rabbit anti-H3K27me2 (Cell Signaling, Cat. 9728), rabbit anti-H3K27me3 (Cell signaling, Cat. 9733).

2.6.3 Antibodies used for immunofluorescence staining and FACS

For immunofluorescence and FACS analysis, the following antibodies were used: rabbit anti-Suz12 (Cell Signaling, Cat. 3737), rabbit anti-H3K27me3 (Cell Signaling, Cat. 9733), mouse anti-BrdU (BD Biosciences, Cat. 347580), Alexa fluor 488 conjugated donkey anti-rabbit (Invitrogen, Cat. A21206), FITC conjugated donkey anti-mouse (Jackson Immunoresearch, Cat. 715095150).

2.7 Primers

ChIP qPCR primers	
oligo name	**sequence (5'-3')**
WT1 forward	GTCGGAGCCCATTTGCTG
WT1 reverse	CAGTGAGACGAGGCTCCCAC
Hoxd9 forward	GGATAATCGCCTAGGTGTGACTTAG
Hoxd9 reverse	CATCTCTTCTTGCCTCTCTGGG
Utp6 forward	AGCTAGGCAGCAGTCACCAT
Utp6 reverse	CAGTTGCGCAATAGTGTCGT
c-Myc forward	GGAGTGGTTCAGGATTGGGG
c-Myc reverse	AAGTTCACGTTGAGGGGCAT
Reg1 forward	GACCATGTCCAGAAGACAGAAG
Reg1 reverse	CTAAGGCTCAGGAAACAGTGTAG
Reg2 forward	GGCCTGCTGGGTTTAGATAAT
Reg2 reverse	TCTGAGCTGAAAGCACATCC
Reg3 forward	AGGAATGGCTATGTGGTTAAGG
Reg3 reverse	TGCCCTGGAAATGGAGTTATG
Reg4 forward	GCCCAAGGTACATGTTTGTATTG
Reg4 reverse	TATCTCTGTCGAGCCCTGTAA
Reg5 forward	CCCGTGCTCTTACACATTCT
Reg5 reverse	CCTGACCTCTTCTGCTTTAGTT
Reg6 forward	GAGCCTGCACAACTCTTCTAT
Reg6 reverse	AGAATTTCCCACCACTCCTTAC
chr5 forward	TGTCTGTTTCCTGCCGGTTA
chr5 reverse	AGCCAAGCCATAAAAATTCGCA
Chr13#1 forward	CGGAAACTTTTCCAGTGGCG
Chr13#1 reverse	CAGTCCAGAGACCCCCTCAT
Chr13#2 forward	CGTTTCCACCGCACCAATTT
Chr13#2 reverse	TGCGTTTGACAGCTATGACCT
Chr11 forward	CCAGCACACACAGCAATCAG
Chr11 reverse	AAGCTGGGAGAAAAGCAGGC
Chr7 forward	CCCCTTCCCATTGCATGGAT
Chr7 reverse	CTTTTCACAGCGACTGGCAC

Table 4: Primers used in ChIP qPCR.

Reverse transcription qPCR primers	
oligo name	**sequence (5'-3')**
Gata6 forward	TGGAAGACACCCCAATCTCG
Gata6 reverse	ACATGGCCCCACAATTGACA
FoxA2 forward	GCGGCCAGCGAGTTAAAGTAT
FoxA2 reverse	GTTGCTCACGGAAGAGTAGC
Tgm2 forward	TAAGAGTGTGGGCCGTGATG
Tgm2 reverse	TTTGTTCAGGTGGTTGGCCT
Asb4 forward	GGAGACGAGCCATTCCTGAT
Asb4 reverse	GACAGGTGCATGAGGGTTCT
Gnao1 forward	AAGTCCCCACTCACCATCTGC
Gnao1 reverse	TACTGCCCTTGGATGTGAGCC
Gsn forward	CCAAAGTCGGGTGTCTGAGG
Gsn reverse	CAAACTTCTCCACACGCCAG

Table 5: Primers used in RT-qPCR.

3. RESULTS

3.1 MEF model system

In order to characterize and dissect the molecular mechanisms underlying the described gain of function mutations occurring on the Polycomb subunit EZH2 (EZH2-Y641N/F) we decided to use mouse embryonic fibroblasts (MEFs) as model system. MEF are indeed easy to manage for experimental purpose. I used MEFs, derived from E13.5 conditional *Rosa26* CRE-ERT2, *Ezh2 fl/fl*, *Ink4a/Arf* knockout embryos, that can be maintained in culture for long period when cultivated at 3% of oxygen. The other reason was to characterize EZH2 putative oncogenic potential independently of the B-cell context in a tumor-prone background, since the *Ink4a/Arf* locus and its downstream targets *pRB* and *p53*, are frequently lost in tumors by genetic mutations or loss of heterozygosity (LOH). To evaluate the role of EZH2 gain of function mutations, I infected MEFs with lentiviral vectors, expressing either the human wild type form of EZH2 (EZH2 WT) or EZH2 carrying a mutation within the SET domain where tyrosine 641 is converted into either a phenylalanine or an asparagine (Y641F and Y641N mutants, respectively). As transduction control, I used an empty vector. All constructs express also Puromycin resistance gene, by means of an Internal Ribosome Entry Site (IRES), to allow selection of stably infected cells (Fig. 6).

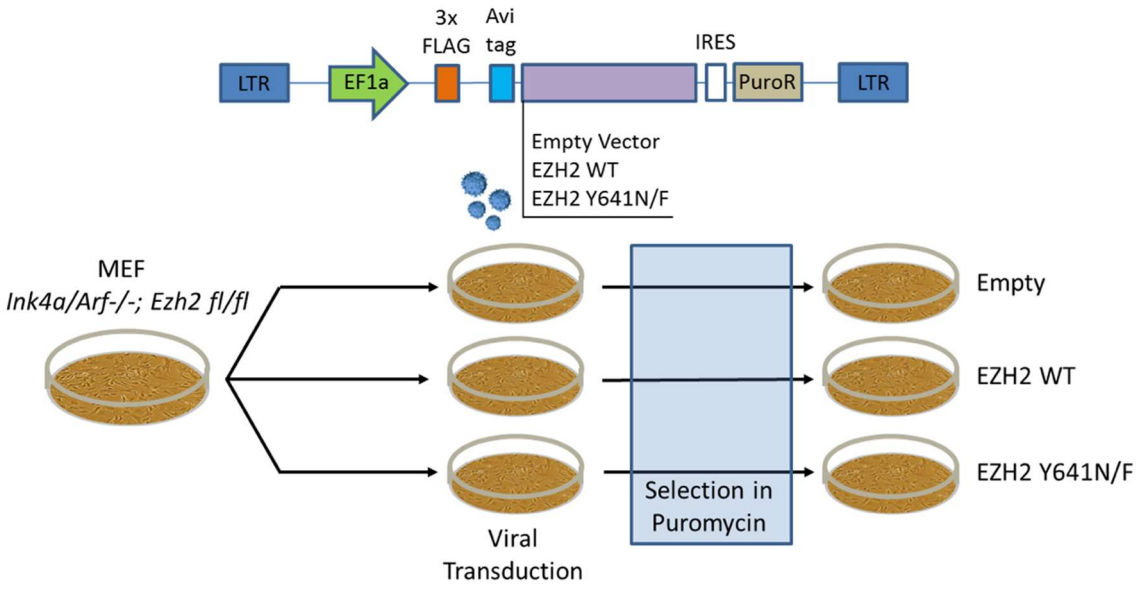


Figure 6: Schematic representation of the employed lentiviral vectors and experimental approach.

Top panel: Schematic representation of lentiviral vectors harboring either the human form of EZH2 WT or the mutated form of EZH2 on Y641 from tyrosine to asparagine (Y641N) or from tyrosine to phenylalanine (Y641F) or no transgene (Empty vector) (purple box), under the control of the EF1α promoter. The protein is fused with a FLAG-tag and an Internal Ribosome Entry Site (IRES) allows the expression of the PuroR gene.

Bottom panel: scheme of the viral transduction protocol in MEFs and selection in puromycin.

3.2 MEF harboring Y641N/F EZH2 show imbalance in H3K27 methylation status as in lymphoma cells

To assess the impact of EZH2 GOF mutants on global levels of H3K27 modifications, I performed western blot analyses for the three H3K27 methylation states on total protein extracts obtained from these cells lines. I showed that only cells expressing EZH2 GOF mutants (EZH2 Y641F/N), but neither EZH2 WT, nor MEFs harboring the empty vector, had increased H3K27me3 levels with concomitant decrease in H3K27me2 and H3K27me1 levels (Fig. 7). Importantly, this pattern of histone modifications caused by the ectopic expression of EZH2 GOF mutants mirrors the one observed in the WSU-DLCL2 lymphoma cell line, naturally carrying the Y641F mutation reported in Sneeringer paper [229].

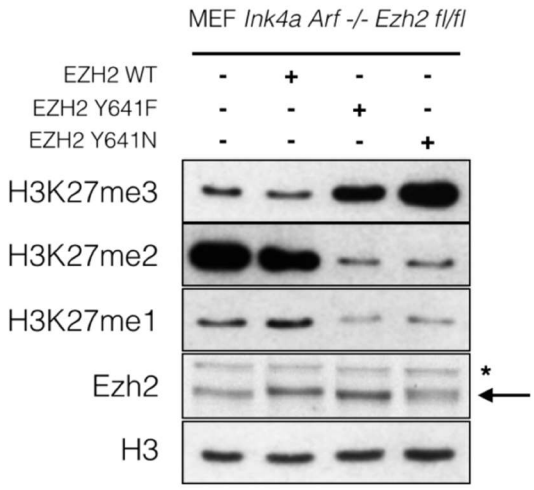


Figure 7: MEFs overexpressing mutant EZH2 rearrange H3K27 methylation balance.

Immunoblot using the indicated antibodies on total protein extracts obtained from MEFs expressing the empty vector (first lane), EZH2 WT or the mutants EZH2 Y641F and EZH2 Y641N. H3 served as loading control. The asterisk indicates an aspecific band while the arrow corresponds to EZH2 band.

To check whether this effect was due to the co-presence of both the WT and the mutated form of EZH2, I depleted the endogenous Ezh2 allele by CRE-ERT2 treating cells with 500 nM of 4-hydroxytamoxifen (4-OHT). Upon deletion, I could show that, only in the presence of a WT Ezh2 allele, the EZH2 Y641N can lead to the accumulation of H3K27me3 (Fig. 8).

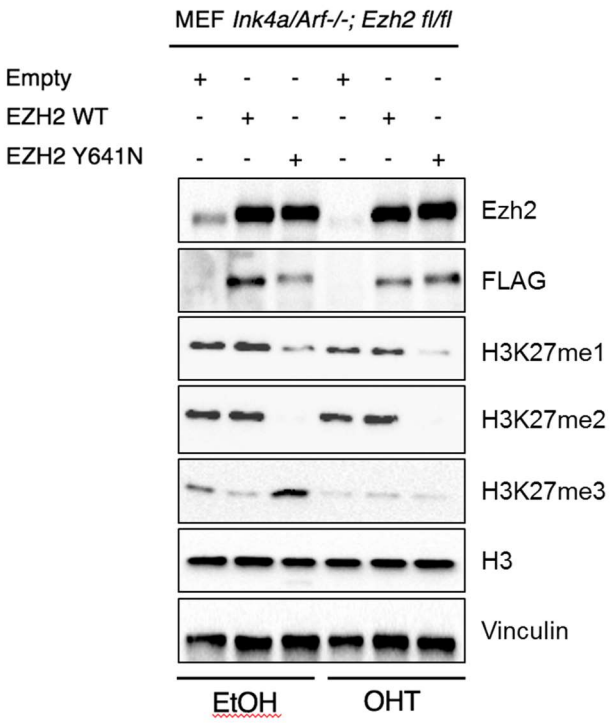


Figure 8: EZH2 Y641N mutant needs its WT counterpart to alter the methylation status of H3K27.

Immunoblots using the indicated antibodies with protein extracts obtained from MEFs expressing EZH2 WT or the mutant Y641N in the presence of the endogenous WT copy of Ezh2 (EtOH-treated, Ezh2fl/fl), or in the absence of it (OHT-treated, Ezh2 -/-). H3 and Vinculin served as loading controls.

In order to obtain a more quantitative evaluation of H3K27 modifications, meaning their relative abundance, in collaboration with Tiziana Bonaldi's group I performed on two biological replicates, in which I firstly assessed the amount of H3K27me2 and H3K27me3 (Fig. 9A), a quantitative mass spectrometry analysis on the histone H3 bulk in cells expressing different EZH2 mutations. Indeed I could score a reduction in H3K27me2 and a concomitant increase in H3K27me3 levels in cells expressing the mutant form of EZH2 (Fig. 9B and 9C).

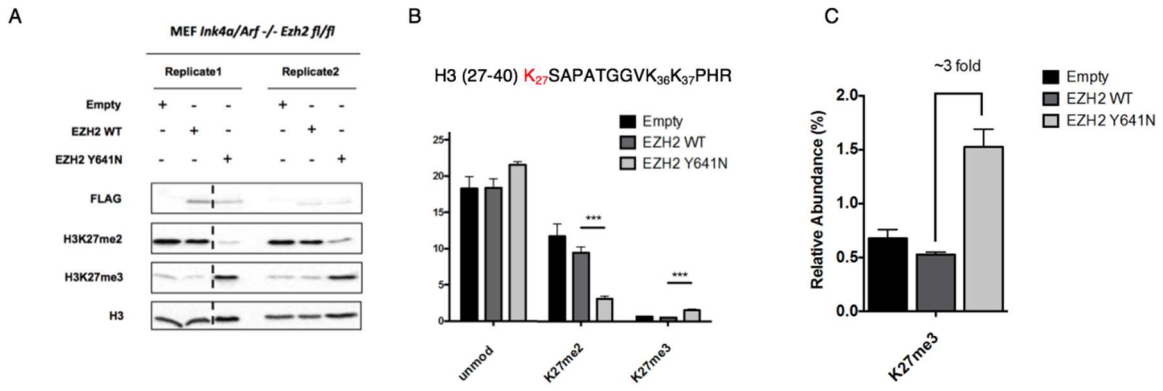


Figure 9: Quantification by mass spectrometry of H3K27 methylation status.

A: Immunoblots using the indicated antibodies with total protein extracts obtained from two biological replicates of MEFs expressing the empty vector or either EZH2 WT or the indicated mutant EZH2 Y641N subjected to MS. H3 served as loading control. **B:** Barplot showing the quantification of the relative abundance of the modifications by mass spectrometry (MS/MS). **C:** Magnification of H3K27me3 barplot as in B.

3.3 Distribution of H3K27 modifications and PRC2 localization

In order to characterize the distribution of the increased levels of H3K27me3, I performed ChIP analyses in MEFs expressing the WT and the mutated forms of EZH2 (Y641N) using H3K27me3 and Suz12 specific antibodies. First of all, I looked at known PcGs target genes by real-time quantitative PCR (ChIP-qPCR). Even if these cell lines present a clear global specific increase of H3K27me3 and decrease in H3K27me2, as assessed by Western blot analyses (Fig. 7), surprisingly by ChIP-qPCR I found that H3K27me3 and Suz12 levels were reduced at the HoxD9 and WT1 promoters (Fig. 10A) that are typical Polycomb targets. I then coupled ChIP to high-throughput DNA sequencing (ChIP-seq) to assess genome-wide the distribution of H3K27me3 mark and Suz12. ChIP-seq analysis showed indeed a loss in H3K27me3 in TSS enriched in H3K27me3 both for the modification and for PRC2 occupancy (Fig. 10B). To validate the results of the sequencing, I analyzed by ChIP-qPCR regions shown to be enriched for H3K27me3 mark in our ChIP-seq analysis and indeed I

could confirm regions characterized by an increase in H3K27me3 compared to WT-EZH2 expressing MEFs (Fig. 10C). Coherently, also Suz12, a core member of the PRC2 complex, is overall displaced from transcription start sites (Fig. 10A and 10B).

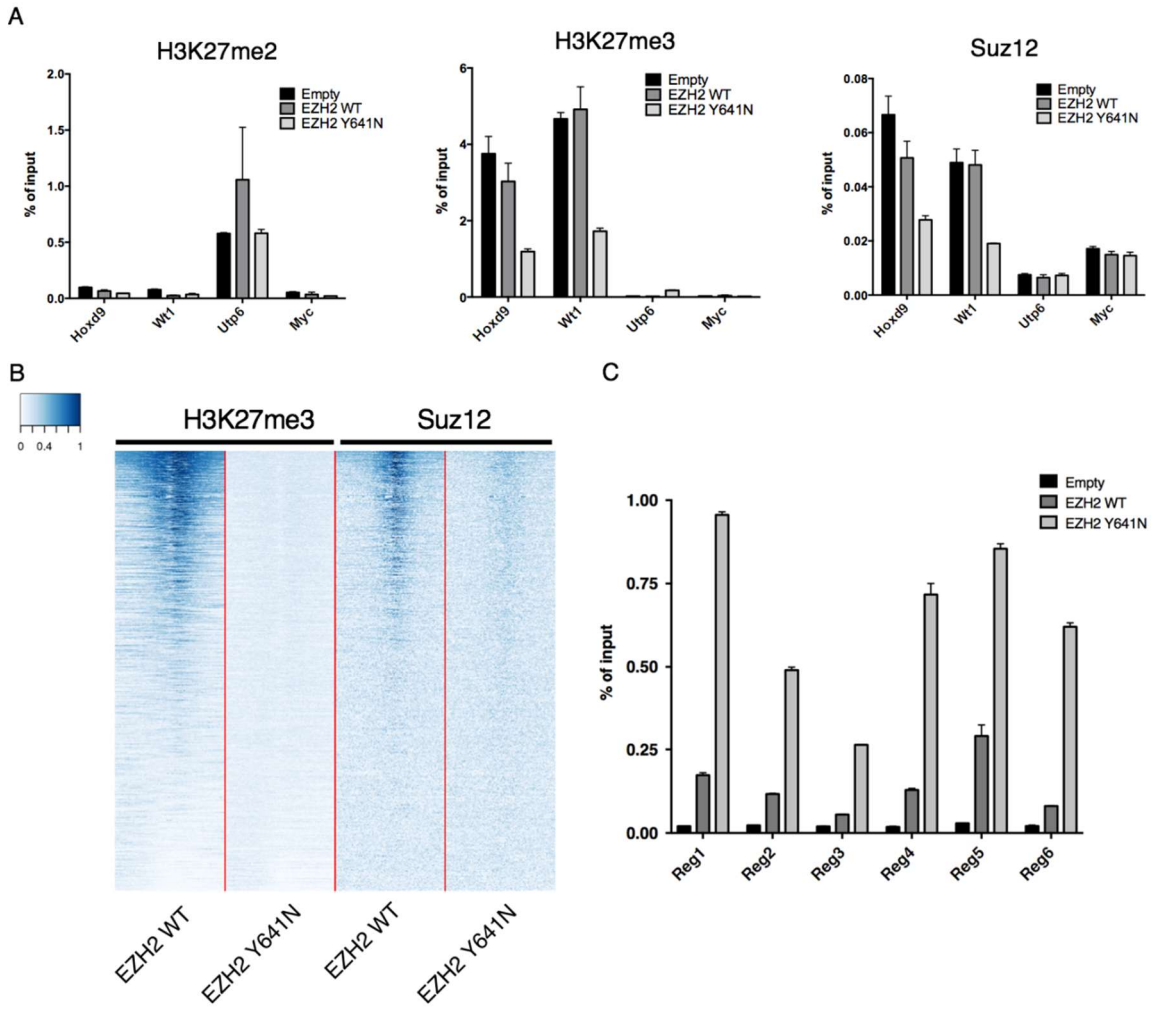


Figure 10: Traditional ChIP-seq analysis is not able to reveal an increase in H3K27me3 upon expression of the Y641N mutant.

A: ChIP-qPCR analyses in MEFs expressing the empty vector or EZH2 WT or mutated EZH2 Y641N performed for the indicated marks on Hoxd9 and Wt1 (as typical Polycomb targets), Utp6 (known to be enriched in H3K27me2) and Myc (negative region) loci. ChIP enrichments are normalized to input. **B:** Heat map of the normalized H3K27me3 and Suz12 ChIP-seq signals in MEFs harboring EZH2 WT or EZH2 Y641N mutation for H3K27me3-enriched promoters +/- 5kb from TSS. **C:** Validation of ChIP-seq by ChIP-qPCR analyses in regions enriched in H3K27me3. ChIP enrichments are normalized to input.

I hypothesized that this controversial result could imply a saturation problem in ChIP analyses because the same amount of antibody is applied in the two conditions. Indeed WT MEFs have a defined number of H3K27me3 enriched regions, while, according to the previous results, a higher number of (or more enriched) targets should be present upon expression of EZH2 mutations. In those circumstances, the antibody binds a larger amount of targets, with a reduction in the local number of ChIP-seq aligned reads with respect to the WT EZH2.

To solve these problems, I decided to investigate the genome-wide distribution of H3K27me3 using a quantitative ChIP with reference exogenous genome using specific antibodies against H3K27me3 (ChIP-Rx).

ChIP-Rx [255] is an innovative method that allows performing genome-wide quantitative comparisons of histone modification status across cell populations using a constant quantity of a reference epigenome or "spike-in". In our case I mixed the chromatin of MEFs expressing the WT or mutated Y641N form of EZH2 cell lines with an amount of Drosophila S2 cell chromatin corresponding to the 5% of the total amount of chromatin obtained from MEFs in order to enable correction of ChIP efficiency. After sequencing and mapping, ChIPseq reads are normalized to the percentage of reference genome reads in the sample. ChIP-Rx indeed enables the discovery and quantification of dynamic epigenomic profiles across mammalian cells that would otherwise remain hidden using traditional normalization methods. Indeed by this approach, we computed the cumulative H3K27me3 signal and could show that H3K27me3 levels were increased upon proper normalization, in agreement with the Western blot results. Overall this analysis clearly highlights, with all its potential limitations, the fundamental role of the "spike-in" normalization to allow a more accurate quantification of ChIP-seq results. Moreover, we could observe a general increase of deposition of this modification not only at promoters (Fig. 11A) but also at genome-wide level as demonstrated by the analysis along the entire lengths of chromosomes (Fig. 11B), highlighting the role of mutant EZH2 Y641N in increasing the global amount of the

H3K27me3 mark.

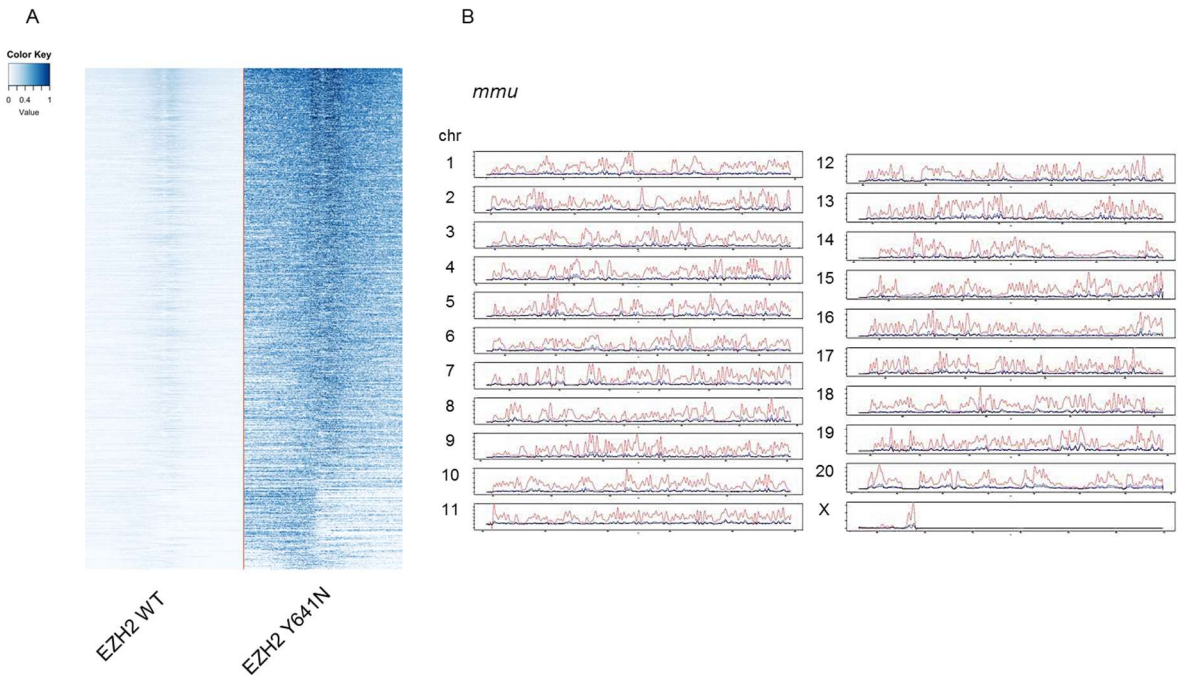


Figure 11: ChIP-Rx shows gain of H3K27me3 in MEFs overexpressing mutant EZH2.

A: Heat map of the normalized H3K27me3 ChIP-seq signal in MEFs harboring EZH2 WT or EZH2 Y641N mutation for H3K27me3-enriched promoters +/- 5kb from TSS. Drosophila S2 spike-in normalization has been applied. **B:** Genome-wide deposition of the H3K27me3 modification along chromosomes lengths in MEFs expressing the empty vector (in blue) or EZH2 WT (in black) or EZH2 Y641N mutation (in red).

3.4. Relocalization

In order to see whether genome wide deposition of the H3K27me3 mark could results in altered nuclear localization of target genes, I performed immunofluorescence analysis of the H3K27me3 modification in the MEFs expressing the WT or the mutant form of EZH2. By confocal microscopy, I observed a strong accumulation of H3K27me3 at the nuclear periphery in cells ectopically expressing the Y641N mutated form of the enzyme compared to the normal dispersed localization pattern observed in cells expressing the WT form (Fig. 12, top panel). Indeed in MEF expressing the mutant form of the enzyme the modification colocalizes with the nuclear lamina (as assessed by Lamin B staining) that typically occupies

the periphery of the nucleus and is preferentially associated with gene-poor regions of chromosomes. This phenomenon occurred despite the usual distribution/localization of the subunit Suz12 (Fig. 12, bottom panel), thus also of the PRC2 complex, both in cells ectopically expressing the WT or the mutated form of the enzyme (Fig. 12).

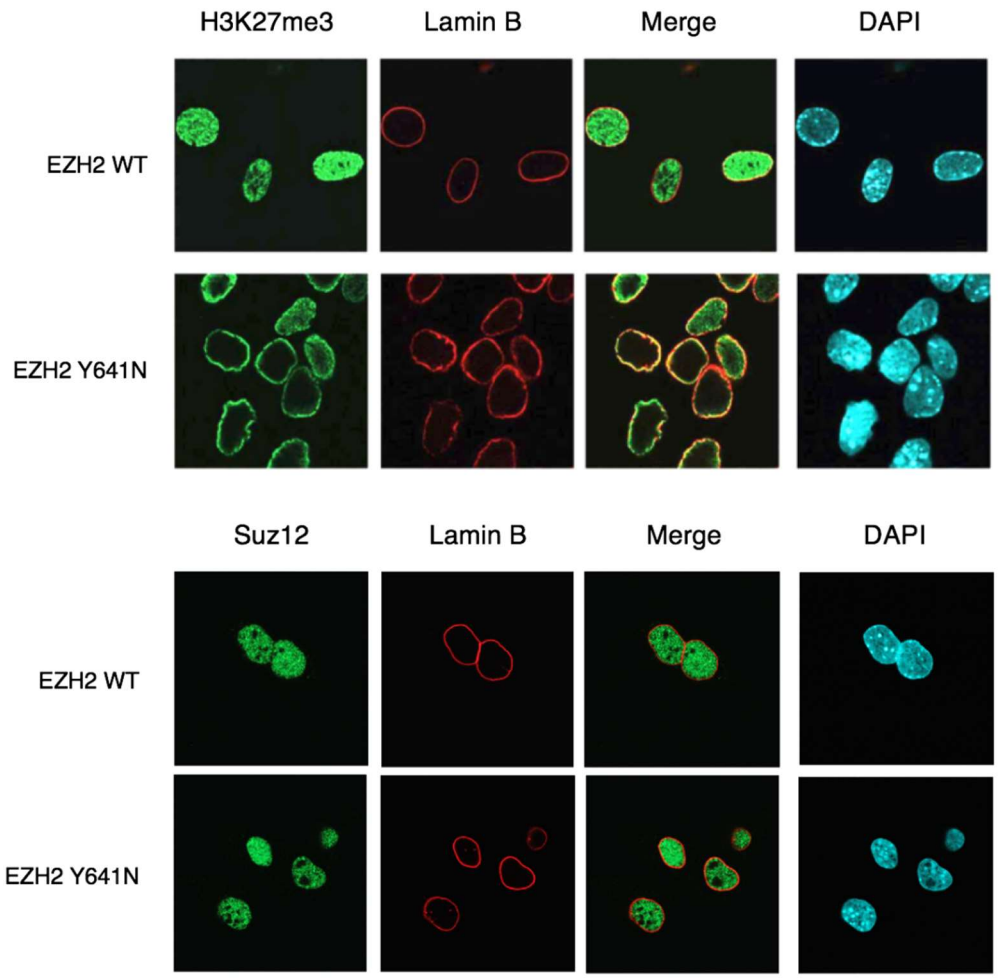


Figure 12: H3K27me3 mark is relocalized to nuclear periphery upon overexpression of mutant EZH2.

Top panel: Confocal analysis of immunofluorescence staining with H3K27me3 (green) and Lamin B (red) specific antibodies in MEFs expressing EZH2 WT or EZH2 Y641N. Cell nuclei were stained with DAPI (blue). **Bottom panel:** Confocal analysis of immunofluorescence staining with Suz12 (green) and Lamin B (red) specific antibodies in MEFs expressing EZH2 WT or EZH2 Y641N. Cell nuclei were stained with DAPI (blue).

3.5. Transcriptional analyses

Since PRC2 is a well-known transcriptional repressor that acts through the deposition of the H3K27me3 mark and by lentiviral infection I overexpressed a subunit of the PRC2 complex, EZH2, responsible of the methyltransferase activity of the complex, I wanted to analyze eventual changes in the expression profile in these cells. I extracted RNA from cells expressing WT or the mutated form of EZH2 and performed RNAseq analysis on two biological replicates. Surprisingly, RNA-seq data showed no expression differences between the two samples (Fig. 13).

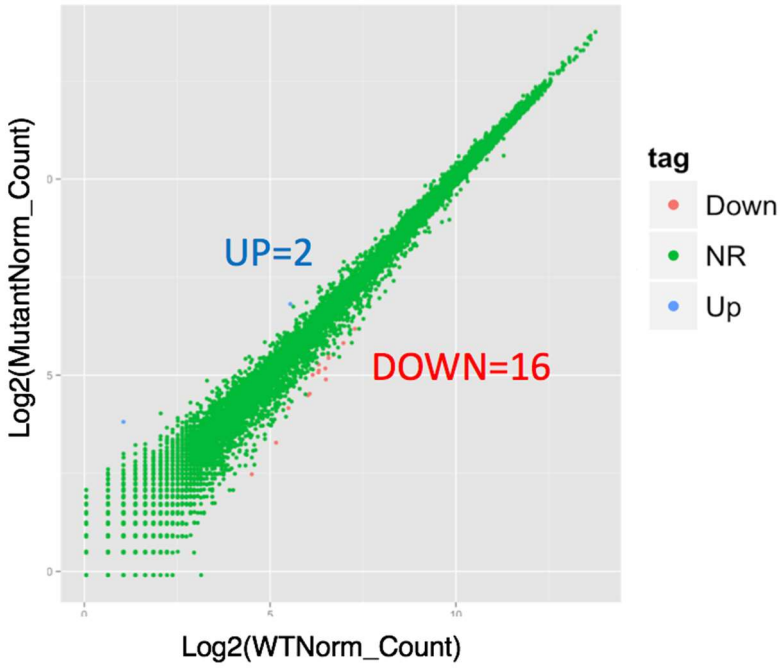


Figure 13: Expression profiling on MEFs expressing EZH2 WT and EZH2 Y641N does not show any differences between samples.

Expression profile performed on MEF *Ink4a/Arf Ezh2 fl/fl* in two biological replicates. Downregulated genes are depicted in red, upregulated genes are depicted in blue while genes that do not change in the two conditions are depicted in green.

Since I noticed that, when growing to confluence, MEFs expressing the mutated form of EZH2 started to behave in a transformed-like manner, changing their morphology and starting to grow without contact inhibition, I decided to investigate the expression profile of

these cells when growing at full confluence. So, with respect to the normal condition in which the cells are passaged every 2 days, from here on called "growing" condition, after a first passage cells were left 10 days in culture subjected only to medium change, from here on called "confluent" condition. With this setup, firstly I checked by western blot analysis that the H3K27me3 mark would not show any difference in the two culturing condition (Fig. 14A). Then, I extracted RNA from cells expressing WT or the mutated form of EZH2 in the two culturing conditions and performed RNAseq analysis on two biological replicates. At the level of gene expression, I could now appreciate a strong difference between MEFs expressing the WT and the Y641N form of EZH2 (Fig. 14B).

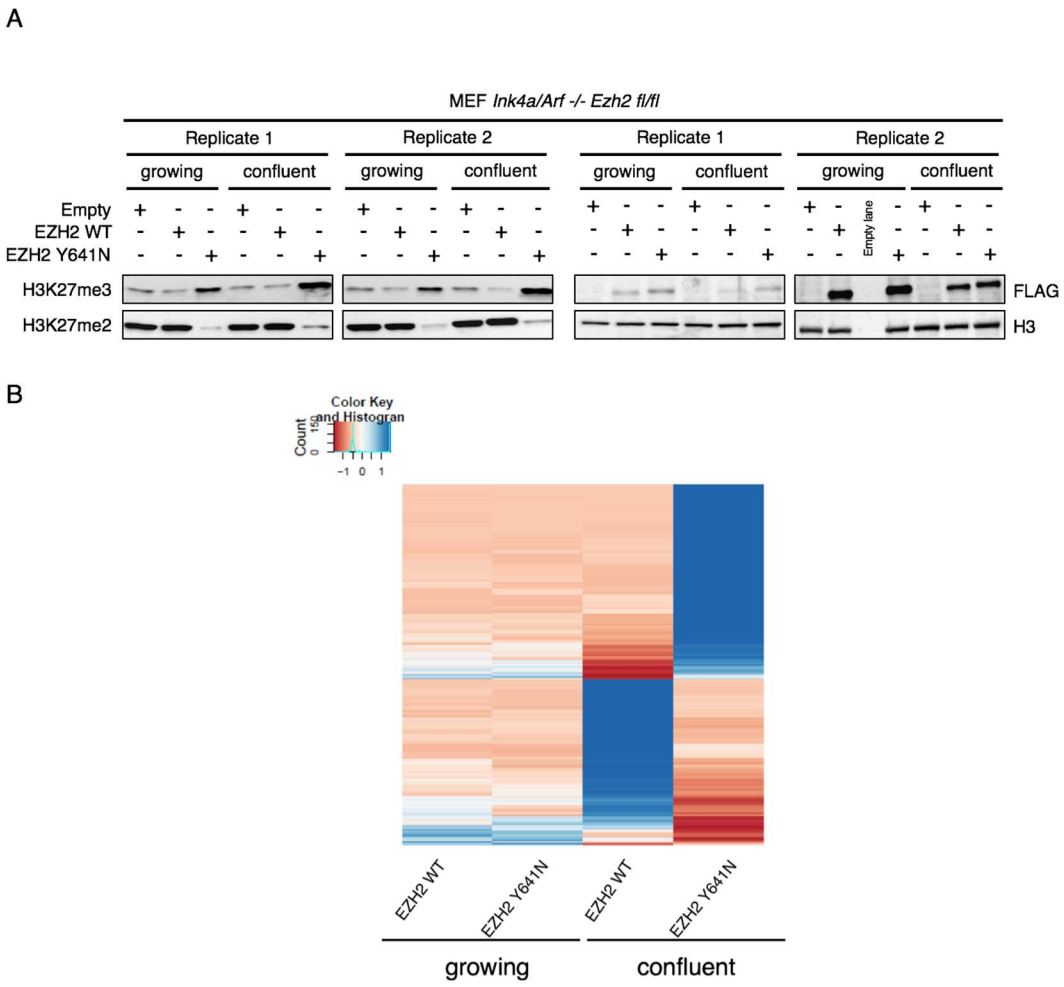


Figure 14: Expression profile on MEFs expressing EZH2 WT and EZH2 Y641N shows differences when cultured at confluency.

A: Western blot analysis using the indicated antibodies on protein extracts obtained from two biological

replicates of MEFs expressing EZH2 WT or EZH2 Y641N mutant. The cells are taken at normal condition of growing passaged every 2 days ("growing") and after being left for 10 days in culture ("confluent"). H3 served as loading control. Empty lane indicates a lane in which no lysate was loaded. **B:** Heat-map of the expression profile averaged between replicates performed on "growing" and "confluent" condition. Downregulated genes are depicted in red, upregulated genes are depicted in blue.

I then performed gene ontology enrichment analysis on the deregulated genes and found that up-regulated genes in the mutant were mainly implicated in epithelial cell differentiation, while down-regulated genes were involved into cell-cell adhesion, in line with the phenotype observed in the cells in culture (Fig. 15).

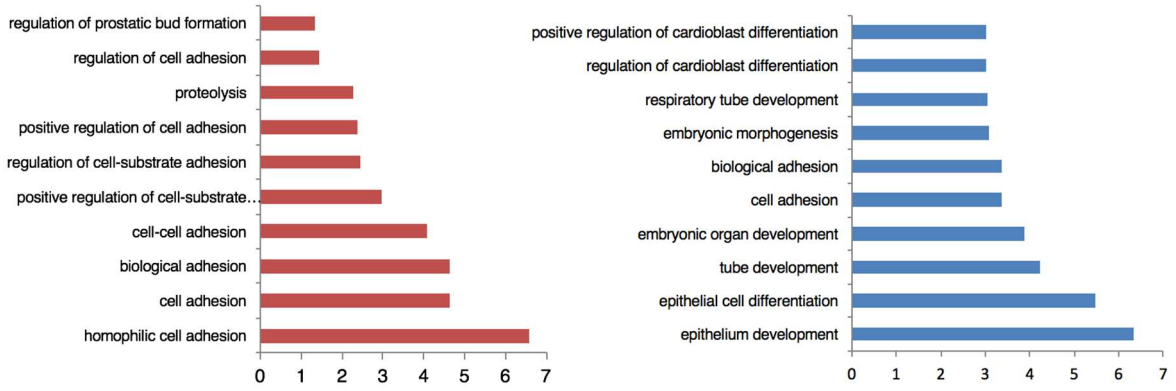


Figure 15: Gene-ontology enrichment analyses reveals that upon confluency, there is downregulation of genes involved in cell adhesion and upregulation of genes involved in epithelial differentiation.

GO-enrichment analysis on the obtained down-regulated (left panel) and up-regulated (right panel) gene sets in MEFs overexpressing the mutant form EZH2 Y641N compared to the WT one grown at confluence condition.

I validated the RNA-seq data through reverse transcription RT-qPCR for some of the deregulated genes, such as Gata6, Foxa2, Tgm2 (upregulated in the mutant), and Asb4, Gnao1 and Gsn (downregulated in the mutant) (Fig. 16).

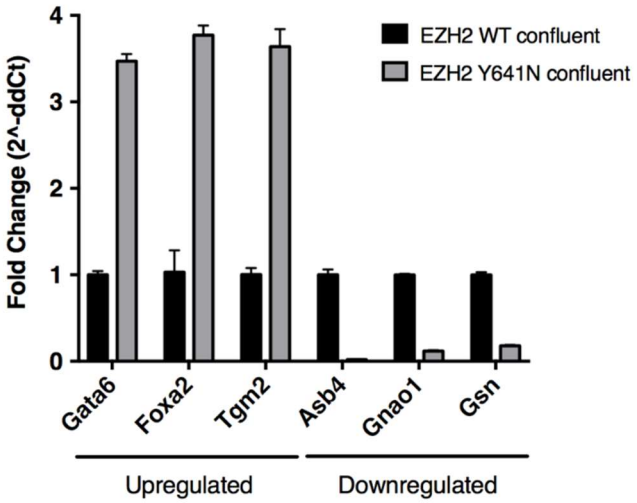


Figure 16: Validation of up- and down-regulated genes in MEFs expressing EZH2 Y641N.

Validation of RNA-seq by RT-qPCR in regions resulted upregulated or downregulated in MEFs cells expressing mutated form of EZH2 Y641N compared to cells expressing EZH2 WT grown upon full comfluence condition of growth.

These results suggest that the transcriptional outcome of the aberrant deposition of H3K27me3 mark becomes evident when cells are subjected to stress.

3.6 The effect of EZH2 GOF mutations in response to stimuli

Given this kind of response, I reasoned that the transcriptional effects due to increased levels of H3K27me3 could become transcriptionally apparent only when cells have to epigenetically adapt to a particular stimulus. So I decided to test our cellular model with different stimuli in order to verify whether this increased amount of repressive mark could improve or on the contrary to injure the ability of the MEF to face to specific environmental change. Specifically I analyzed three different "stimuli": i) the cooperation with an oncogene; ii) starvation of cells; iii) reprogramming to pluripotency.

3.6.1 Myc-driven Polycomb activity

MYC (also known as c-MYC), described more than 30 years ago as a cellular homologue of the avian retroviral oncogene *v-myc* [256] is part of to the proto-oncogenic transcription factor family that comprise, MYC, NMYC and LMYC.

The c-*myc* proto-oncogene product, MYC, is a transcription factor that binds thousands of genomic loci [257]. Its targets are represented by active promoters and enhancers in the genome. The result is that it induces an amplification of transcription [258]. Moreover, several studies have shown that MYC is deregulated in Burkitt's lymphomas [259, 260] and in non-Hodgkin B-cell lymphomas [261].

I wondered whether the expression of an oncogene could synergistically act with the mutated EZH2 in regulating transcription in cells.

I took advantage of mouse 3T3 fibroblasts expressing a conditional Myc-oestrogen chimaeric receptor ($3T3^{MycER}$). In this way I could induce Myc translocation into the nucleus upon tamoxifen administration, where it can exert its function. I proceeded by stably integrating EZH2-expressing lentiviral constructs also in these cells (Fig. 17).

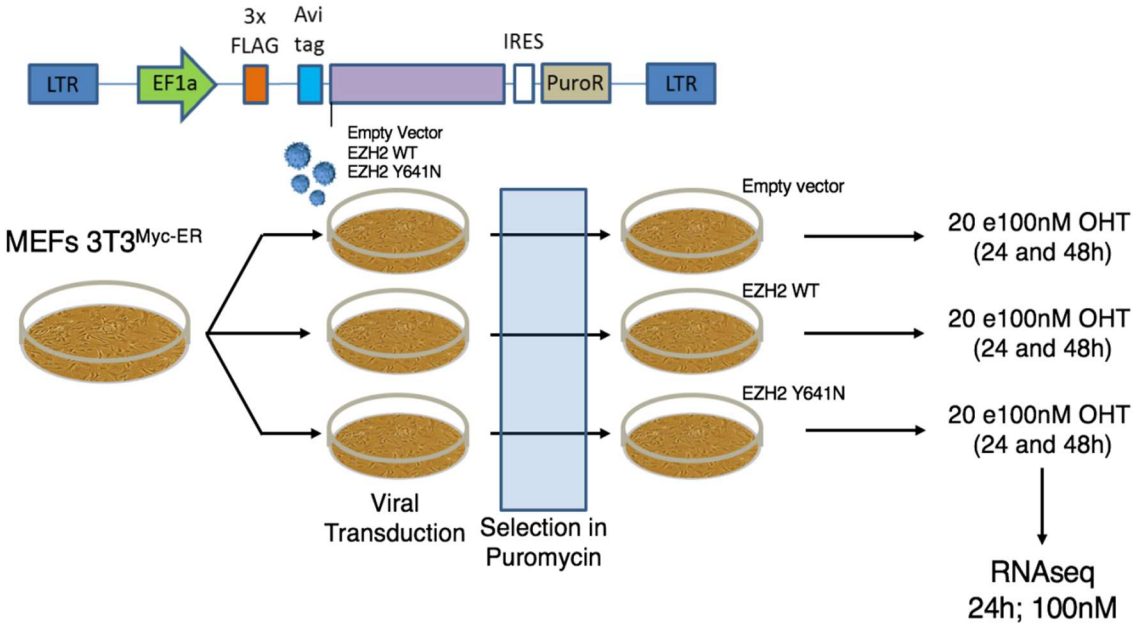


Figure 17: Schematic representation of the employed lentiviral vectors and experimental approach.

Top panel: Schematic representation of the lentiviral vector harboring either the human form of EZH2 WT or the mutated form of EZH2 Y641N or no transgene (Empty vector) (purple box), under the control of the EF1α promoter. The protein is fused with a FLAG-tag and an Internal Ribosome Entry Site (IRES) allows the expression of the PuroR gene. **Bottom panel:** scheme of the viral transduction protocol in MEFs Myc-inducible, $3T3^{MycER}$, and selection in puromycin and description of the followed experimental procedure.

First of all I assessed the level of ectopic expression of the enzyme and I tested whether the expression of the mutant also in these cells was able to increase the global level of H3K27me3. I performed western blot analyses on total extracts obtained from these cells lines and again I could show that only EZH2 gain of function mutants (EZH2 Y641N), but neither EZH2 WT, nor MEFs harboring the empty vector, showed increased H3K27me3 levels with concomitant decrease in H3K27me2 independently from the myc induction upon administration of 4-OH T (Fig.18).

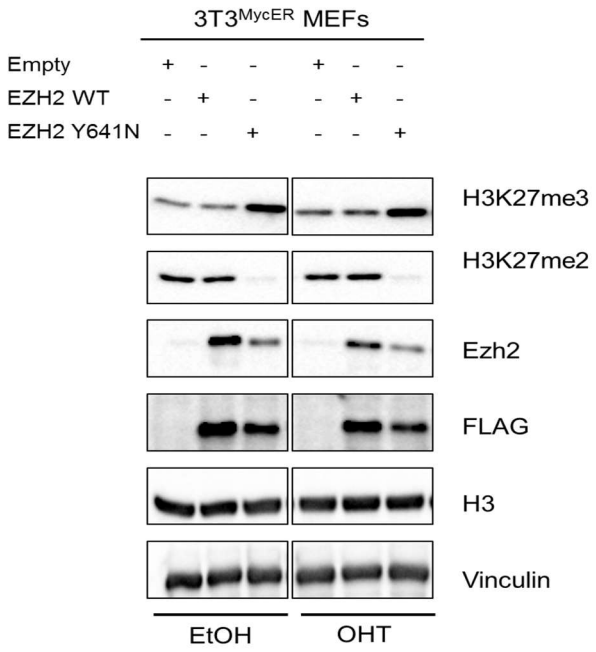


Figure 18: The mutant EZH2 Y641N alters methylation states on H3K27 also in 3T3^{MycER} MEF.

Immunoblots using the indicated antibodies with protein extracts obtained from 3T3^{MycER} MEFs expressing the empty vector or EZH2 WT or the mutant Y641N before (EtOH) and after (100 nM OHT) Myc induction. H3 and Vinculin served as loading controls.

Then I treated these cell lines with ethanol (vehicle control) or different concentration of 4-OHT, 20 and 100 nM for 24 and 48 hours as schematically described in Fig. 17 observing the typical change in morphology encountered by cells Myc-induced driven observe in Petri-dishes.

I then, extracted RNA from these MEFs and checked the upregulation of classical Myc targets [262] upon Myc-induction, such as REEP6 and SMPD, by RT-qPCR (Fig.19).

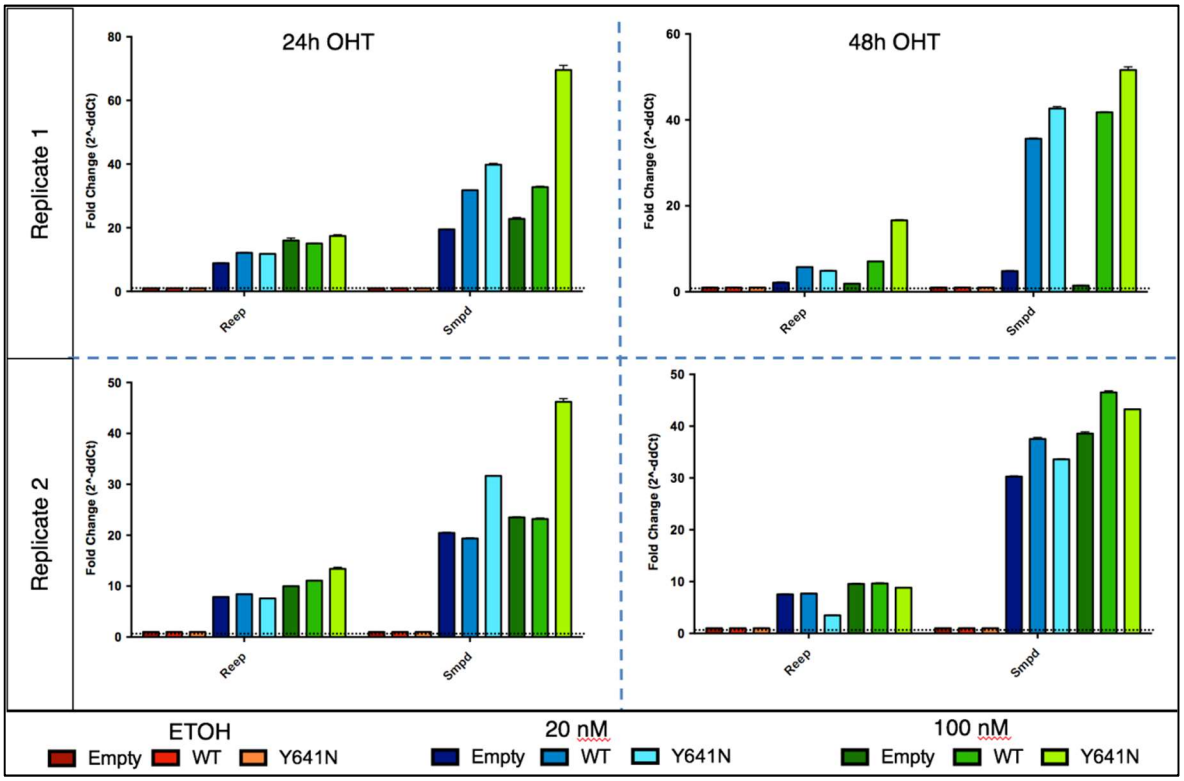


Figure 19: Time and dose dependent-induction of Myc in MEFs expressing WT or mutated form of EZH2.

Analyses of reverse transcription followed by qPCR in $3T3^{MycER}$ MEFs expressing the empty vector or EZH2 WT or mutated EZH2 Y641N without (EtOH treatment) or upon (OHT treatment) Myc induction in Myc target regions.

To avoid saturation, I selected 24h and 100 nM concentration as the condition to be subjected to RNA-seq.

In this case, I could detect very few genes differentially regulated between the MEF expressing the mutant EZH2 in respect to WT form upon induction of Myc (Figure 20).

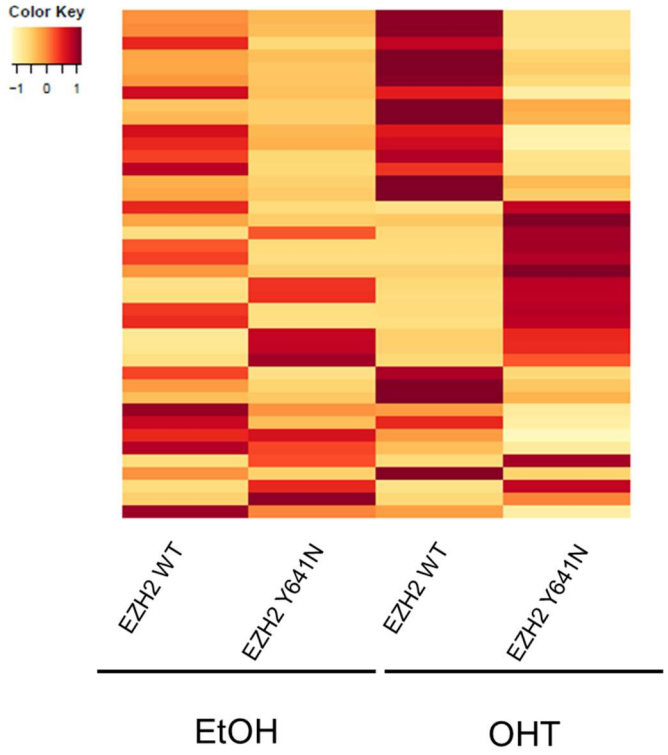


Figure 20: Expression profile on 3T3^{MycER} MEFs expressing EZH2 WT or the mutant Y641N

Heat-map of the expression profile of 3T3^{MycER} MEFs expressing the EZH2 WT or the mutant Y641N performed with induction of myc with 100nM OHT for 24 hours. Upregulated genes are depicted in red, downregulated genes are depicted in yellow.

3.6.2 Starvation-driven Polycomb activity

As a new kind of stimulus I analyzed the behavior of *Ink4a/Arf-/-, Ezh2 fl/fl* MEFs when grown in starvation condition.

To address this point I passaged normally the cells but using different concentration of serum put in the culturing medium. In particular I used 0.1%, 1% and standard 10% serum concentration. I firstly tested the ability of the cells to live with lower amount of serum. In particular I tested these concentration until 72 hours (data not shown) and then I chose 24 hours as end-time point, since afterwards cells tended to die. So, I left the cells growing for 24 hours with different concentrations of serum and verified the exit from the cell cycle induced by the starvation process by FACS analysis (Fig. 21A). To do so, cells were stained for DNA content with PI and for replicative DNA synthesis with anti-BrdU antibody.

Indeed, I could observe that cells grown with lower amount of serum underwent starvation since the majority of cells exit from S-phase to enter in G_0 phase (Fig. 21B).

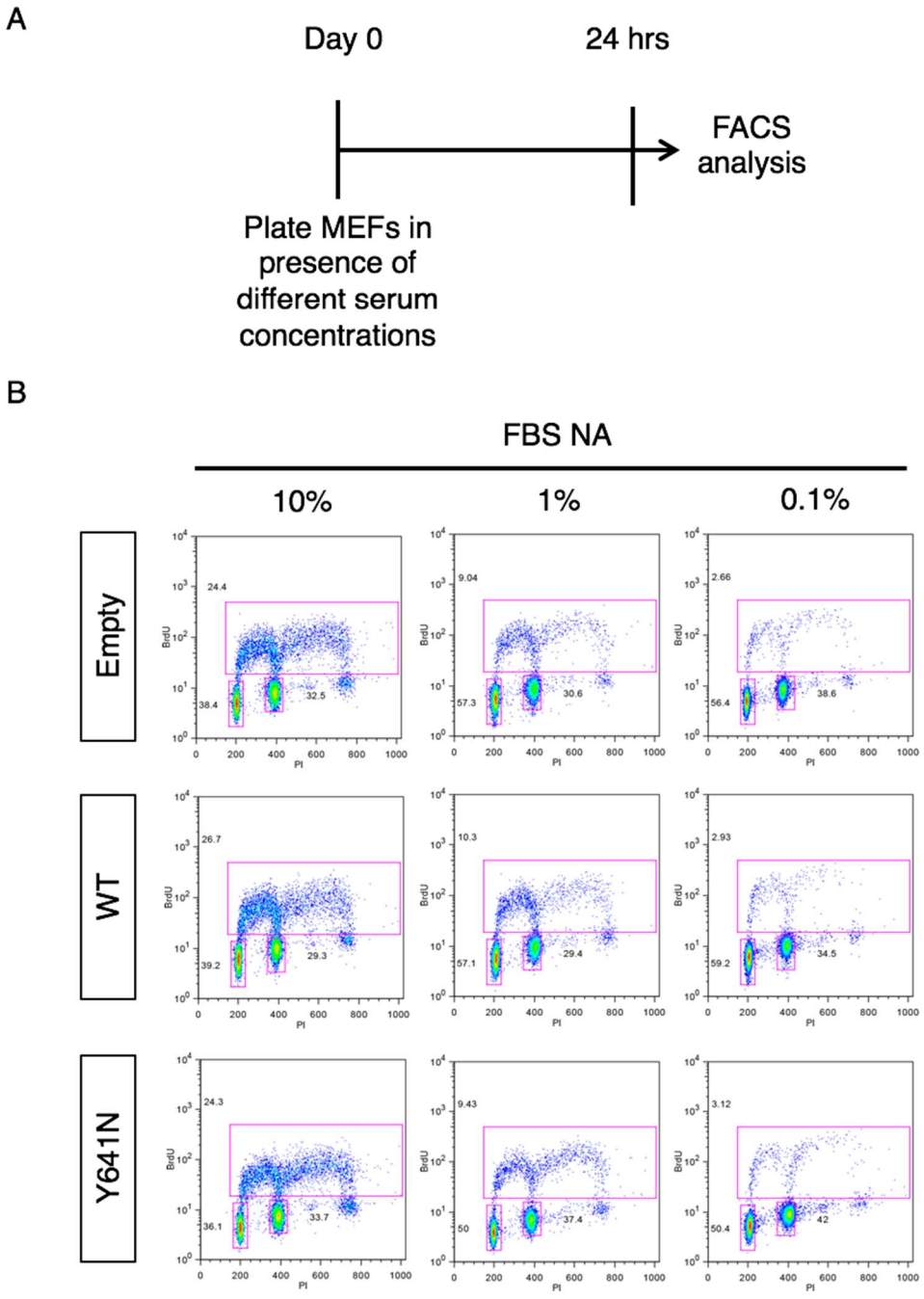


Figure 21: FACS analyses show exit from the cell cycle when cells are grown in starvation condition.

A: Scheme of the followed experimental procedure. MEFs expressing EZH2 WT or Y641N cultured with 10%, 1% or 0.1% serum and 24h later were subjected to FACS analysis. **B:** FACS analyses on MEFs transduced with empty vector or vectors expressing EZH2 WT or EZH2 Y641N cultured in presence of different concentrations of serum (10% or 1% or 0.1%). The analyses are representative of two biological replicates.

Since at 0.1% of serum I could obtain the highest fraction of cells in G_0 phase, I used this condition and the normal serum concentration as a control to perform western blot analyses on total extracts. I analyzed the levels of H3K27me2 and H3K27me3 states and I could show that also in this condition only EZH2 gain of function mutants (EZH2 Y641N), but neither EZH2 WT, nor MEFs harboring the empty vector, showed increased H3K27me3 levels with concomitant decrease in H3K27me2 (Fig. 22). Moreover, starved MEFs, so cells mainly in the G_0 phase, were characterized by lower amount of proteins such as Suz12 and Eed as previously observed in literature [201].

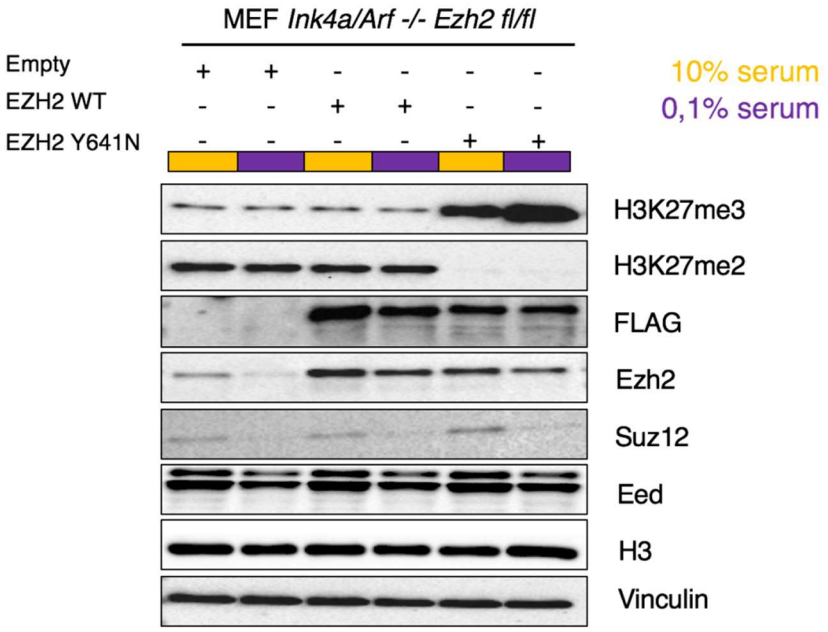


Figure 22: Mutant EZH2 induces further accumulation of H3K27me3 mark in starved cells.

Western blot analysis using the indicated antibodies on protein extracts obtained from MEFs *Ink4a/Arf Ezh2 fl/fl* expressing EZH2 WT or Y641N cultured with 10% (lanes depicted in yellow in the upper part of the panel) or 0.1% serum (lanes depicted in purple in the upper part of the panel). The analyses are representative of two biological replicates.

After this, I extracted RNA from these cells to performed RNA-seq analysis. Also in this case, the expression profile of the mutant starved cells compared to 10% serum-grew cells showed no differences in the expression profiles (Fig. 23).

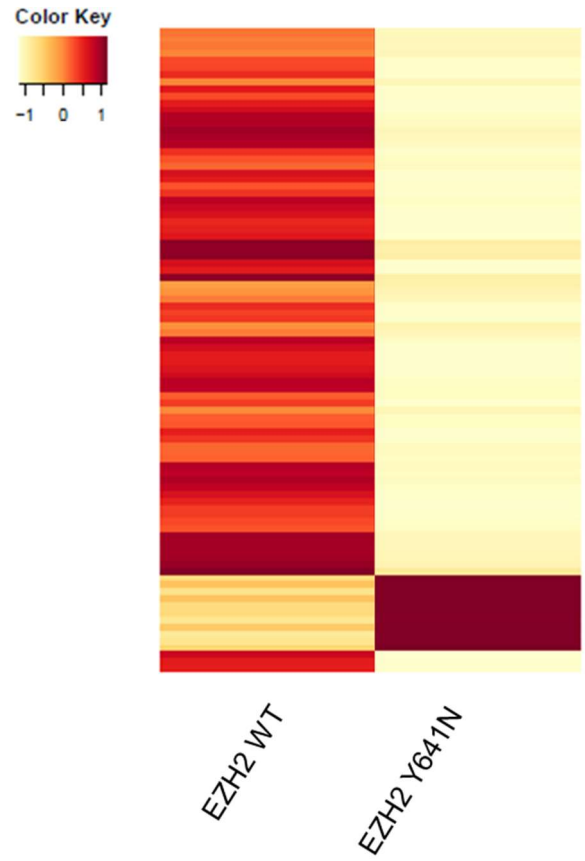


Figure 23: Expression profile on MEFs expressing EZH2 WT and EZH2 Y641N shows few differences in starvation condition.

Heat-map of the expression profile performed at starvation condition between MEF *Ink4a Arf-/- Ezh2 fl/fl* expressing EZH2 WT or mutant EZH2 Y641N grown at 0.1% serum. Upregulated genes are depicted in red, downregulated genes are depicted in yellow.

3.6.3 Reprogramming

I sought to determine whether genome-wide deposition of H3K27me3 could alter another induced cellular process, that is reprogramming to pluripotency.

Indeed, in the presence of mutant EZH2 Y641N, the H3K27me3 mark spreads across the whole genome, so I hypothesized that this phenomenon could impair epigenetic resetting upon induction of cell fate reassignment by the Yamanaka factors (OCT4, SOX2, KLF4, c-MYC) [263-265]. Also, upon spreading of this mark, pluripotency genes show increased

marking on their body/promoter. This is true both for early and late markers of reprogramming [266] (Fig. 24).

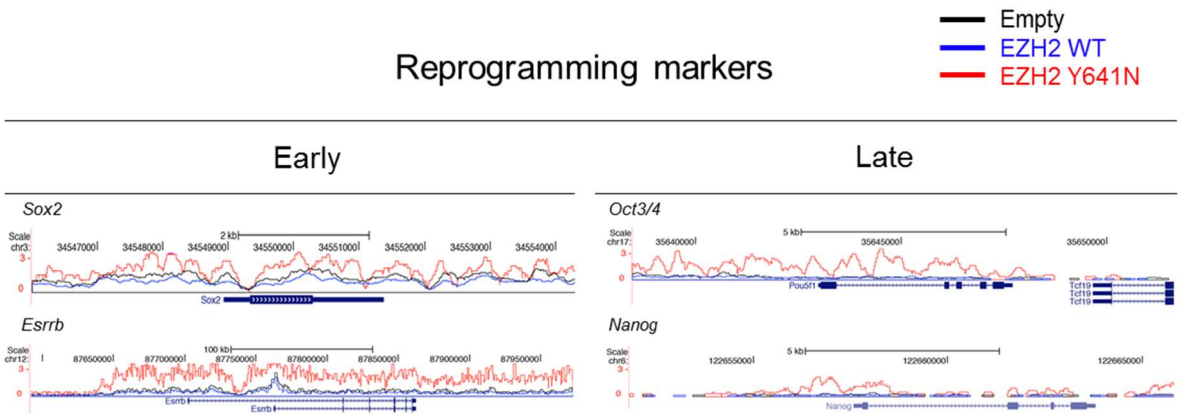


Figure 24: Genes involved in early and late phases of reprogramming are enriched for H3K27me3 upon expression of mutant EZH2 Y641N in MEFs.

Genomic snapshots of H3K27me3 ChIP-seq experiments from MEF *Ink4a/arf -/- Ezh2 fl/fl* transduced with empty vector (in black) or EZH2 WT (in blue) or mutant EZH2 Y641N (in red).

So I wondered whether the increased presence of this repressive mark could have some effects on the reprogramming capabilities of cells expressing the mutated form of EZH2, like the impairment of the process itself.

To address this point, I made use of "reprogrammable MEFs", which were derived from mice carrying a double knock in of the four transgenes in the 3'-UTR of the *Col1a1* locus under the transcriptional control of a doxycycline-responsive promoter. In these mice, also, the reverse tetracycline-controlled transactivator (M2-rtTA) is expressed from the *Rosa26* locus, allowing for induction of expression upon doxycycline administration [245, 246] (Fig.25).

I infected these MEFs with vectors expressing the empty vector or either the WT form of EZH2 or the Y641F/N mutants. These vectors also carry all an independent Zs-green fluorescent protein expression cassette, but the empty vector carrying a Green fluorescent protein cassette, that allowed us to sort these cells to purity (Fig. 25).

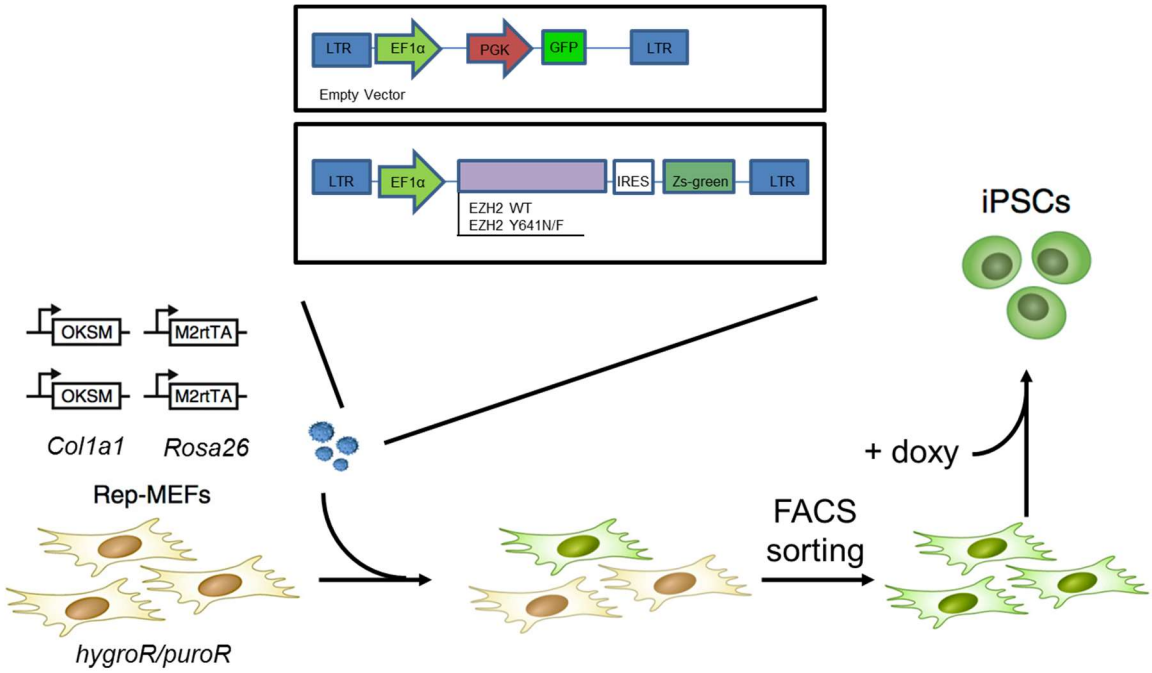


Figure 25: Schematic representation of the employed lentiviral vectors and experimental approach for reprograming.

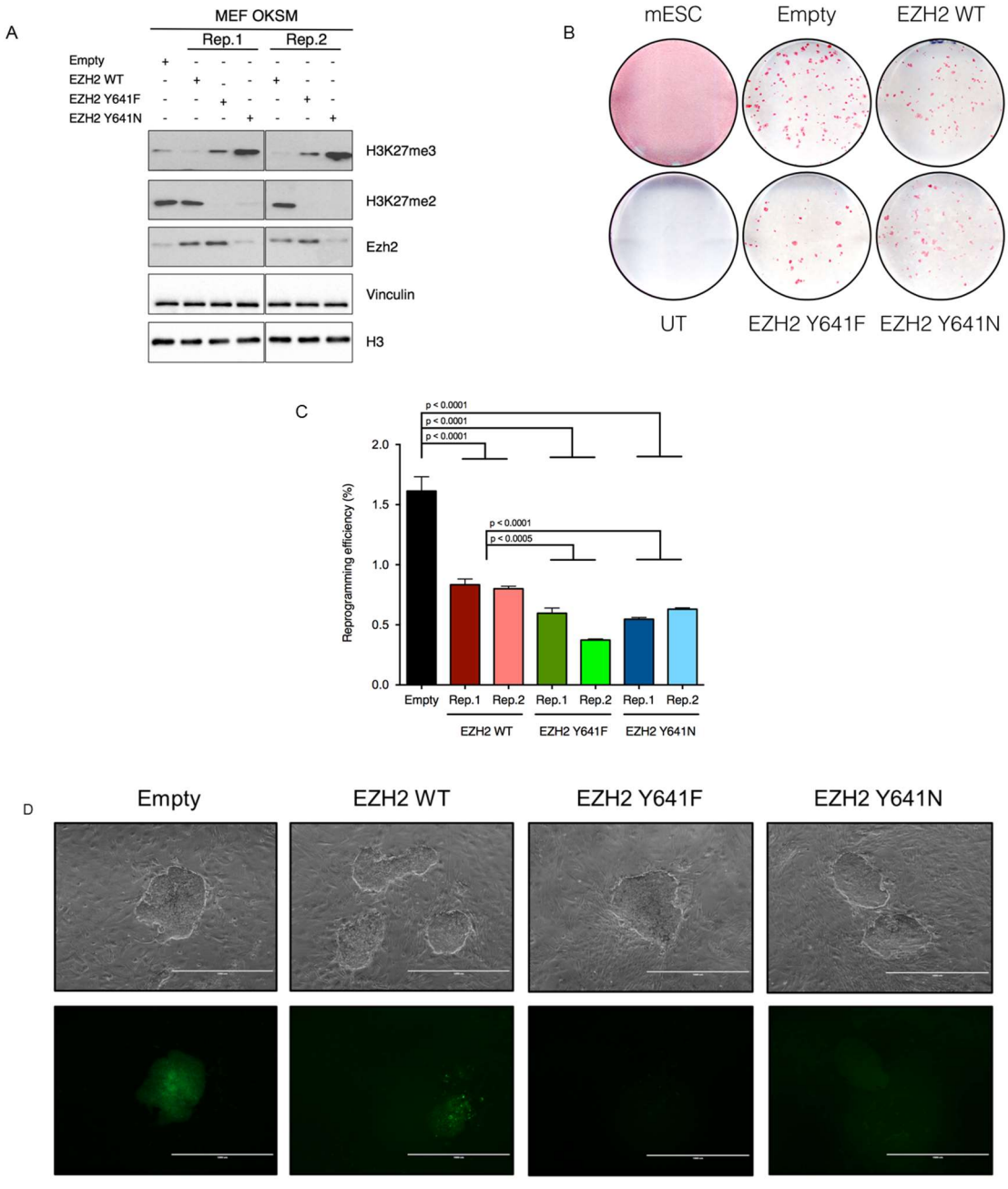
Top panel: Schematic representation of lentiviral vectors harboring either the human form of EZH2 WT or the mutated EZH2Y641F or EZH2 Y641N forms (purple box), under the control of the EF1α promoter and in frame to Zs-green cassette to use as selection of integrated lentiviral vector. An empty vector carrying a GFP cassette was used as selection control of integrated lentiviral vector. **Bottom panel:** scheme of the viral transduction protocol in reprogrammable OKSM (Oct3/4, Klf4, Sox2, c-Myc) MEFs used for the reprogramming and FACS selection of positive cells with GFP/Zs-green signal.

I verified the accumulation of the H3K27me3 mark by western blot analyses on protein extracts derived from sorted MEFs OKSM (Fig. 26A).

I then induced reprogramming in these cells through the use of doxycycline, and after 12 days performed an alkaline phosphatase staining to assess reprogramming efficiency. Indeed, I could score a reduced efficiency in the presence of the overexpression of WT EZH2, and this reduction was even more marked in the presence of each of the two mutants (Fig. 26B-C) .

Interestingly, colonies that emerged from the mutant EZH2-expressing MEFs showed a remarkable reduction in the fraction of GFP+ colonies, suggesting either silencing of the

vector or reprogramming of cells that were either lowly or not expressing the transgene, despite the initial sorting (Fig. 26D and relative quantification Fig. 26E). These results suggest also that the impairment in GFP+ colonies result even more strong cells expressing the mutant enzyme.



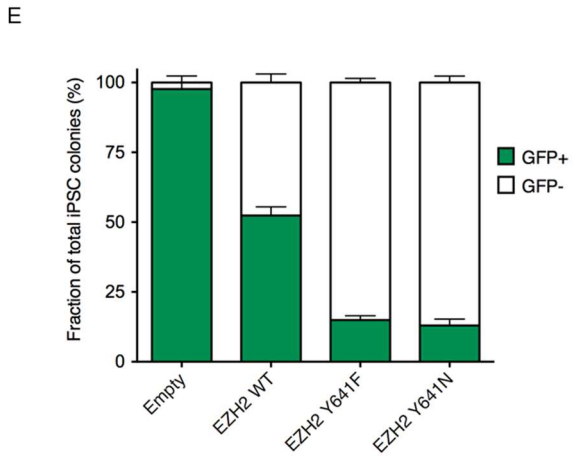


Figure 26: Mutant EZH2 impairs reprogramming.

A: Western blot analyses of OKSM MEFs expressing the expty vector or the ones expressing EZH2 WT or mutated EZH2 Y641F and EZH2 Y641N. **B:** Alkaline Phosphatase staining of emerging iPSC. The assay has been performed on MEF OKSM transduced with empty vector or EZH2 WT or mutated EZH2 Y641F and EZH2 Y641N, mESC were used as positive control, UT represent MEF OKSM not induced with doxycycline. **C:** Reprogramming efficiency calculated as the ratio between number of plated cells and the formed colonies. **D:** Bright field (top) and green-fluorescence (bottom) pictures of colonies after 12 days of induction. Scale bar 1000 µm. **E:** Fraction of GFP+ colonies calculated in respect of total formed colonies.

3.7 Validation in B-cell Lymphomas

3.7.1 Lymphoma cells ectopically expressing or physiologically harboring Y641N/F EZH2 show imbalance in H3K27 methylation status

Finally, I wanted to validate our results in a more relevant system, diffuse large B cell lymphoma (DLBCL). There are different human DLBCL cells, characterized by either a normal activity of PRC2 enzyme EZH2 or the natural occurrence of the mutated form of EZH2. As representative for the first type, I used OCI-LY7 cells that are described as DLBCL, characterized by a reciprocal chromosomal translocation that transposes the coding exon 2 and 3 of the c-myc locus from chromosome 8 to chromosome 14 where it is present

the immunoglobulin heavy chain locus and presenting a point mutation on p53. As representative of the second type I have used WSU-DLCL2 that naturally harbors Y641F mutated form of EZH2.

Firstly I performed western blot analysis to check the levels of H3K27 modifications in the different described lymphoma cells. I first used Ly7 cells. I transduced them with the same lentiviral vectors expressing the EZH2 WT or EZH2 mutated form Y641F and Y641N described in Fig. 6.

I could score an increase in H3K27me3 and concomitant reduction of H3K27me2 levels (Fig. 27A). Then I analyzed WSU-DLCL2 cells that pathophisiologically harbor Y641F mutated form of EZH2 and could saw the same pattern of H3K27 methylation with increased level of H3K27me3 and concomitant decreased amount of H3K27me2 (Fig. 27B).

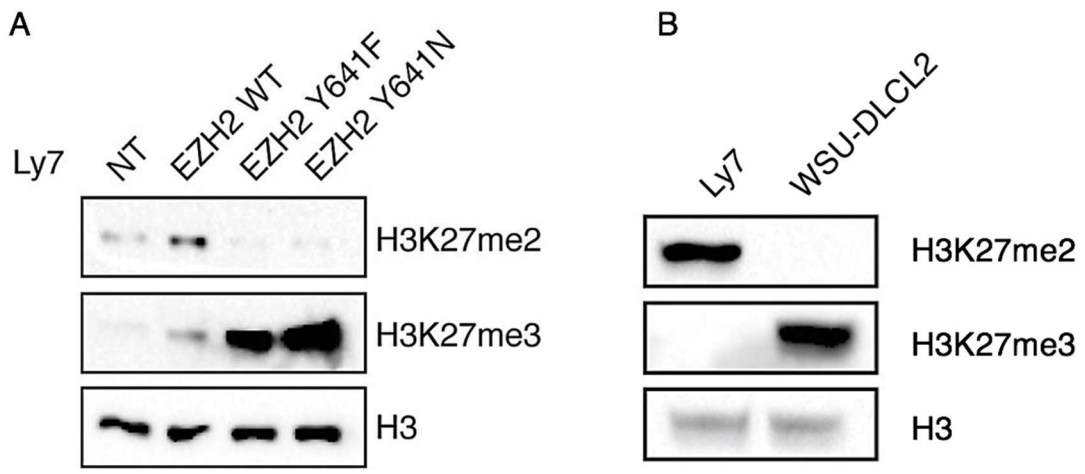


Figure 27: Lymphoma cells overexpressing or harboring mutant EZH2 rearrange H3K27 methylation states.

A: Immunoblots using the indicated antibodies with protein extracts obtained from lymphoma cells Ly7 ectopically expressing EZH2 WT or the Y641F and Y641N mutation. H3 served as loading control.

B: Immunoblots using the indicated antibodies with protein extracts obtained from lymphoma cells Ly7 having a WT form of EZH2 or WSU-DLCL2 lymhphoma physiologically harboring Y641F mutation. H3 served as loading control.

3.7.2 Distribution of H3K27 modifications in lymphoma cells

I decided to investigate the effect of H3K27me3 accumulation by ChIP-qPCR also in the lymphoma system. I also performed ChIP for Suz12 subunit in order to analyze the occupancy of the PRC2 in these conditions in these cells. Testing the Ly7 cells expressing the empty vector (UT) or EZH2 WT (WT) or mutated EZH2 Y641N (Fig. 28), I could show that H3K27me3 and Suz12 levels were reduced at several promoter regions that are typical Polycomb targets in lymphomas cells. This result resembles what I previously observed in MEFs (Fig. 10A).

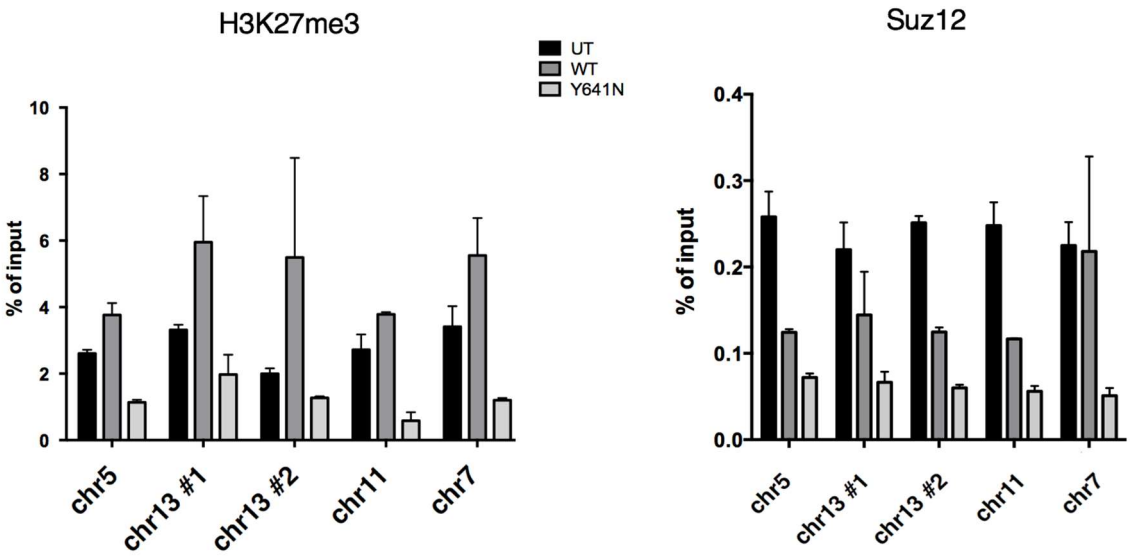


Figure 28: Traditional ChIP qPCR analysis is not able to reveal increase in H3K27me3 upon expression of the mutant EZH2 Y641N.

qPCR of ChIP analyses in Ly7 expressing the empty vector (UT) or EZH2 WT (WT) and mutated EZH2 Y641N performed with the indicated antibodies on top of the histogram on typical Polycomb targets. ChIP enrichments are normalized to input.

At this point, I performed the same analyses also in WSU-DLCL2 (Fig. 29) and SU DHL6 (Fig. 30) cells (that pathophisiologically harbor EZH2 Y641F and EZH2 Y641N mutated form of EZH2 respectively) checking the same promoter regions by ChIP-qPCR for both the deposition of H3K27me3 and the localization of PRC2 performing a Suz12 ChIP. In these

experiments I used WT Ly7 cells as control since normally these latter cells present a WT form of EZH2. Also here I could appreciate a reduction in the deposition of the repressive mark and a displacement of PRC2 complex (Fig. 29-30) as happen in Ly7 transduced with mutant EZH2 Y641N (Fig. 27) and in MEF cells (fig.10A).

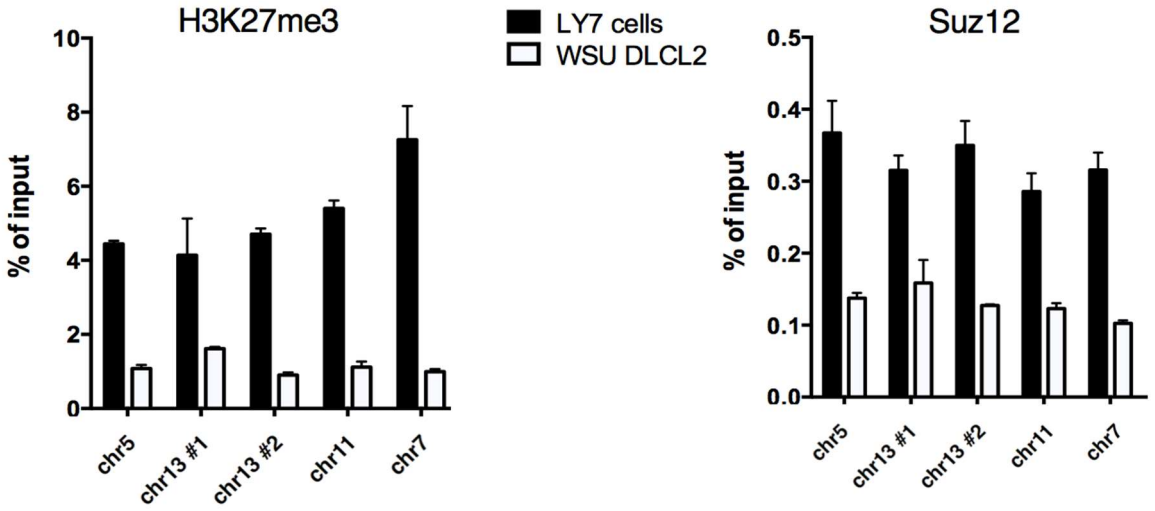


Figure 29: ChIP qPCR analysis is not able to reveal increase in H3K27me3 in lymphoma cells naturally expressing mutant EZH2 Y641F enzyme.

ChIP-qPCR analyses in Ly7 characterize by a WT form of the enzyme (used as control) and WSU-DLCL2 physiologically harboring Y641F mutation, performed with the indicated antibodies on top of the histogram on typical Polycomb targets. ChIP enrichments are normalized to input.

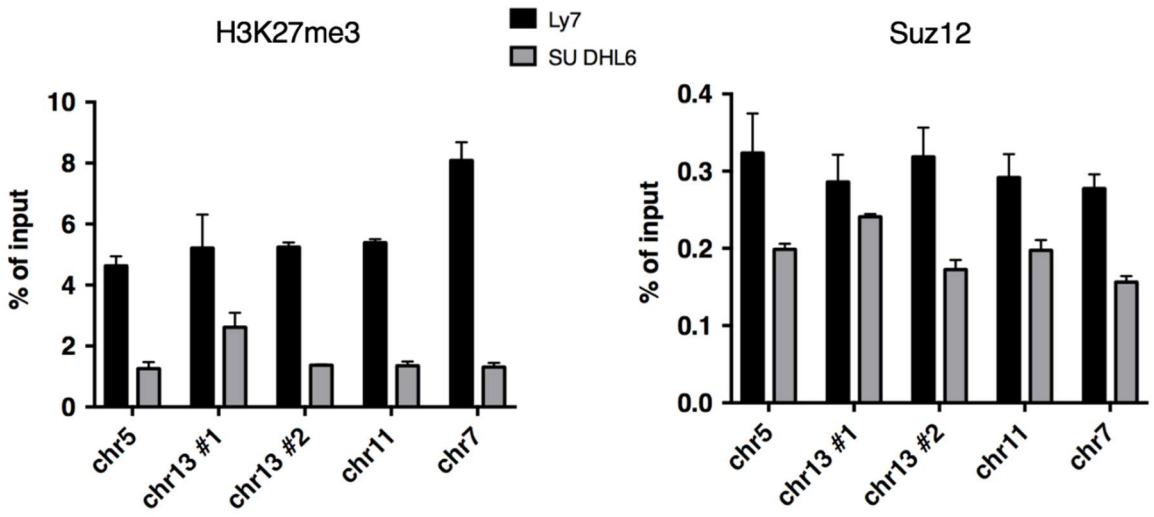


Figure 30: ChIP qPCR analysis is not able to reveal increase in H3K27me3 in SU DHL6 cells that naturally express mutant EZH2 Y641N enzyme.

ChIP-qPCR analyses in Ly7 characterize by a WT form of the enzyme and SU-DHL6 physiologically harboring Y641N mutation, performed with the indicated antibodies on top of the histogram on typical Polycomb targets. ChIP enrichments are normalized to input.

I performed also ChIP-Rx analysis on Ly7 ectopically expressing the WT or mutated form of EZH2. We computed the cumulative H3K27me3 signal and observed that, after the normalization, H3K27me3 levels resulted increased, in agreement with the Western blot results and in contrast to what observed previously without the spike-in correction. I then analyzed the genome-wide deposition of the H3K27me3 modification along chromosomes length in these cells. I could observe a general increase of deposition of this modification not only at the promoter level (Fig. 30) but also genome wide (Fig. 31), as obtained also in MEFs (Fig. 11B), supporting the global increase amount of the H3K27me3 mark.

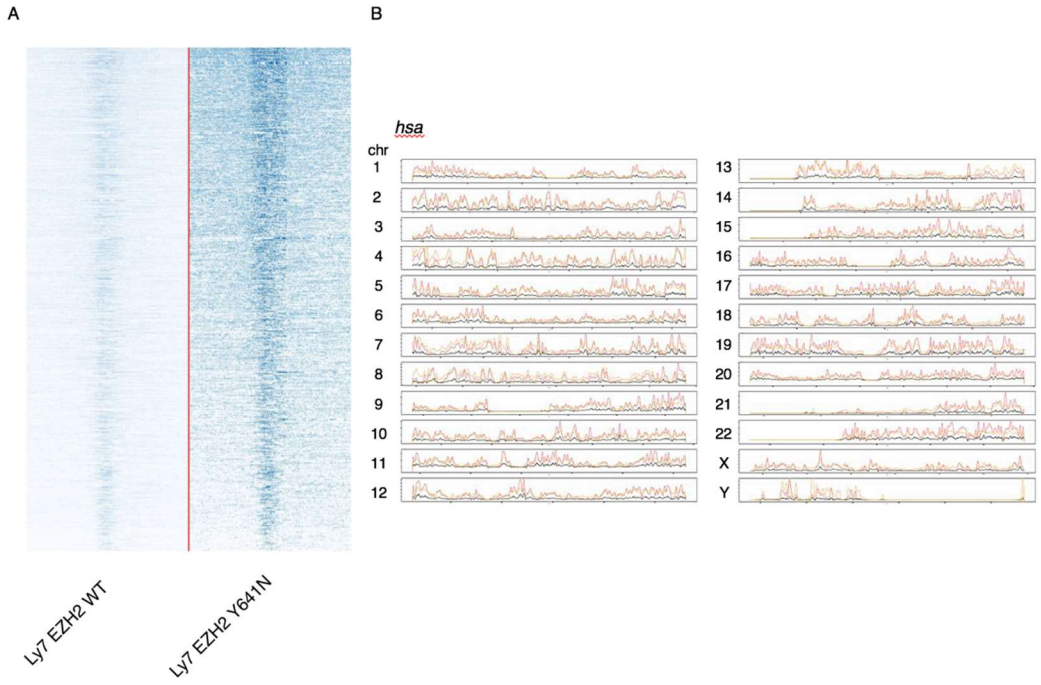


Figure 31: Y641N EZH2 mutant expression causes the increased deposition of H3K27me3 in DLBCL.

A: Heat map of the normalized H3K27me3 ChIP-seq (ChIP-Rx) signal in lymphoma Ly7 cells ectopically expressing EZH2 WT or EZH2 Y641N mutation for H3K27me3-enriched promoters. Drosophila S2 spike-in normalization has been applied. **B:** Genome-wide deposition of the H3K27me3 modification along chromosomes in lymphoma cells ectopically expressing the empty vector (in yellow and blue) or EZH2 Y641N (in red) or physiologically harboring EZH2 Y641F mutation (in orange).

4. DISCUSSION

Among cancers affecting lymphoid tissues, the most frequent disease is represented by diffuse large B cell lymphoma (DLBCL), accounting for about 30-40 % of the newly diagnosed cases of non-Hodgkin lymphomas. DLBCL is a very heterogenous disorder. It comprises two major subgroups: germinal center B cell (GCB) and activated B cell (ABC) lymphomas, each characterized by specific oncogenic pathways, translocations, recurrent mutations and gains and losses of genetic material and of activity of specific enzymes. GCB and ABC derive from subsequent stages of differentiation of germinal center B cells. In particular, the GCB subtype arises from centroblasts, the proliferative cells in the germinal center, while the ABC from a plasmablastic cell just before its egress from germinal center. The second most diffused type of non-Hodgkin's lymphoma is the follicular lymphoma (FL). These kinds of diseases present some features that resemble the ones of the germinal center like DLBCL.

Among the key player in the regulation of germinal B cells there is EZH2. The level of expression of this factor is higher in lymphoid progenitors, then it decreases in resting B cells while it increases again in germinal centers where B cells undergoes high proliferation and immunoglobulin affinity maturation.

The diseases affecting these kinds of cells are different and are characterized by different clinical features and multiple responses to the different possible treatments. Of note, 20% of DLBCLs and 10% of FLs display heterozygous somatic mutations of EZH2 [225]. These mutations enable EZH2 to exhibit increased activity toward the accumulation of H3K27me3. For these reasons lots of efforts have being done in several laboratories to better understand these disorders in order to find novel therapies able to manage this heterogeneity.

The work presented in this thesis aimed to better understand the molecular mechanisms underlying these disorders using molecular and biochemical approaches.

We decided to use MEF as a model to investigate these mutations in a simpler alternative system. We reasoned that since EZH2 is found deregulated in several cancers, we could use a different epigenetic context to understand the general oncogenic activity of this enzyme. Moreover, this same mutation has been found also in a small percentage of melanoma solid tumor where this mutation greatly contributes to tumorigenesis [228].

Regarding the deposition of H3K27me3 mark, we demonstrated that our model system recapitulates the global increase of this mark in respect to a decrease in H3K27me2 mark (Fig. 7). We showed, therefore, that this phenotype occurs also in a different developmental and epigenetic context in respect to B-cells such as MEFs. Furthermore, we validated our results in lymphoma cells. This is true for lymphomas harboring naturally a WT EZH2 (OCI-Ly7) transduced with a mutated form of EZH2 Y641 or presenting its own mutated version of the enzyme: EZH2 Y641F in WSU DLCL2 (Fig. 27). That is in total accordance with what has been shown in literature [118, 224, 229, 230]. We showed that all the cell lines transduced with vectors containing the point mutation on EZH2 SET domain or taken from patients have increased levels of the H3K27me3 modification with respect to cells presenting the WT form of the enzyme [267]. Importantly, we demonstrated for the first time in our MEF model system that mutant EZH2 needs its WT counterpart to exert its increased activity toward accumulation of H3K27me3.

Indeed, we took advantage of OHT-inducible *Rosa26* CRE-ERT2, *Ezh2 fl/fl*, *Ink4a/Arf* MEF that allowed depletion of the WT alleles. The result is that without the WT enzyme, the cell is not more able to perform hyper-trimethylation of H3K27 and the accumulation of the mark is impaired (Fig. 8). In literature this has been demonstrated only *in vitro* through methyltransferase assay [229] while *in vivo* it has been shown that lymphoma cells always display a WT copy of the enzyme in presence of the mutant; [224]. To this regard, different authors showed only that by treating mutant EZH2 Y641 harboring lymphomas with compounds able to inhibit PRC2 activity, such as GSK343 [224] or GSK126 [231], the global level of H3K27me3 results drastically reduced. But these compounds affect

presumably both WT and mutant EZH2 present in cells thus confirming only that they are necessary together to have the hyper-trimethylation. Moreover, our data demonstrated that in presence of the hyper-activating mutation, Ezh1 is not able to compensate for Ezh2 functionality after its depletion. This highlights for the first time, that specifically Ezh2 WT is necessary in order to achieve H3K27 hyper-trimethylation.

The first ChIP analyses revealed a result that was in disagreement with the accumulation of H3K27me3, as shown by western blot analyses and MS (Fig. 9). Indeed both ChIP-qPCR and ChIP-seq data indicated a lower presence of the H3K27me3 both on Polycomb targets and on H3K27me3-enriched promoters, accompanied by a displacement of PRC2 (Fig. 10). We thought this was possibly due to: i) the choice of regions that by chance could be characterized by lower deposition of this mark, in the case of qPCR analyses; ii) the saturation of the antibody, regarding the ChIP-seq data. We hypothesized that using a certain amount of antibody able to recognize a certain level of antigen H3K27me3 present in a cell characterized by a WT EZH2, that same amount could be not sufficient to bind all the H3K27me3 antigens present in a cell with a mutant EZH2. This could explain the results of our ChIP-seq data. In order to assess this, we took advantage of a quantitative ChIP-seq analysis using *Drosophila* chromatin as reference genome [255]. By ChIP-Rx correction, we could show that indeed an overall spreading of the H3K27me3 mark was present in mutant-expressing MEF. Moreover, the deposition was increased not only on the TSS of typical PRC2 targets but also along the entire length of all chromosomes (Fig. 11). This means that when we analyze and compare the tracks coming from cells harboring the WT or mutated form of EZH2, we can find the spread of the mark also in the gene body. Our lab [129] has demonstrated that PRC2 complex in embryonic stem cells is responsible for H3K27 methylation on 80% of H3. The presented ChIP-seq data showed also that each H3K27 methylation state is confined into spatially defined domains. In particular we demonstrated that H3K27me1 present at intra-genic regions favors transcription, H3K27me2 representing the constitutive activity of the PRC2 complex is present at inter- and intra-genic region

where it exerts protective function preventing firing of non-cell type specific enhancers and H3K27me3 is present on the TSS of genes that have to be repressed. In our system, we have a model that shows that in co-presence of WT and mutant form of the enzyme the cells are characterized by a spread of H3K27me3 mark genome-wide at the expenses of H3K27me2 that is the default activity of the complex. Since the H3K27me3 mark is linked to transcriptional repression, we wanted to investigate the functional outcome of this widespread mark. To address this point, we performed genome-wide transcriptomic analyses. Surprisingly, the first RNA-seq analyses did not point out any difference upon the expression of the mutant form of the enzyme (fig. 13). Our collaborators in Stefano Casola's group performed a microarray analysis on lymphoma cells harboring either the WT or the mutant form of EZH2. The comparison between the samples also in this case did not highlight any difference. Unexpectedly, the mutant harboring cells were not characterized by an increased level of gene repression in accordance with the increased levels of the repressive mark. In literature different groups compared WT and mutant EZH2 harboring lymphoma cells trying to find the correlation between the increased PRC2 activity and the cancer development. McCabe and colleagues, for example, showed very few differences in different lymphoma cells harboring WT or mutant form of the enzyme and varying significantly according to the analyzed cell line [231]. Among them the authors found the down-regulation of genes responsible for the maturation of GC B cells toward a subsequent step of maturation and differentiation. In this way the cells are more prone to stably express a stem phenotype toward to lymphoid transformation. The same has been reported in the work of Beguelin and colleagues [224]. They used a lymphoma cell line, BCL1, characterized by a WT form of EZH2 that transduced with WT EZH2 or EZH2 Y641F or EZH2 Y641N, and performed ChIP-seq and RNA-seq analysis. They found (but did not show) a concordance between the spreading of the H3K27me3 mark and the effective consequent transcriptional repression. Interestingly, among negatively regulated genes, they found some involved in the termination of the GC reaction, such the ones IRF4-induced, and

in negative regulation of cell cycle, so genes involved in differentiation induction. The final output of this deregulation is that B-cells remain in a pro-proliferative state.

The only case in which we could observe a difference in the transcription profile upon expression of the mutant form of the enzyme (EZH2 Y641N) was when cells have been subjected to a stressful growth condition. Upon confluence, indeed, they activate novel and distinct pathways of regulation to manage the stress condition (Fig. 14). What we could observe was the down-regulation of genes involved in the cell adhesion that is indeed what we could observe in Petri dish where the cells tend to lose the cell-to cell inhibition, resembling a transformed phenotype (Fig. 15). This behavior of the MEFs can be linked to the potential oncogenic response to this stressful stimulus.

Besides those results, we could also observe the localization on the lamina of the H3K27me3 characterizing MEFs transduced with the mutant EZH2 Y641N (Fig. 12). The nuclear lamina is composed of Lamin B and Lamin A that code for Lamin A and Lamin C. These protein components undergo PTMs and their role is to act as scaffold for the correct organization of the genome thus providing a structural support to the nucleus. They can interact with several distinct factors including proteins involved in regulating chromatin structure and transcription. Several papers demonstrated that Lamins cover important roles contributing, during cell differentiation, to the re-localization of portions of chromosomes in the nuclear space correlating with the expression of specific genes according to the developmental stage [268, 269]. In mammalian cells, from the transcription point of view, the nuclear periphery is considered as a transcriptionally inactive region [270-272]. Indeed, several papers showed that when the portion of chromatin located in the nuclear periphery should undergo transcription, it is rearranged and delocalized into a more inner part of nuclei. For example it has been demonstrated that when mouse embryonic stem cells differentiate toward neural fate, a large portion of chromosome 10 moves from the nuclear periphery to the inner part of nuclei [270] and in consequence of this there is a strong transcriptional activation of Mash1 gene, that code specifically for a neural transcription factor. In our experiments, we

scored, by immunofluorescence assay (IF), the re-localization of this increased amount of H3K27me3 to the nuclear periphery in cells expressing the mutated EZH2 Y641N. We speculate that the H3K27me3-enriched chromatin could occupy a different localization in the nuclei interacting in alternative way with nuclear lamina and this could be associated to a functional role of this increased marking. In order to study chromosome architecture with respect to nuclear lamina in our system, we can take advantage of Dam-ID technology. This method is based on the fusion of Escherichia coli DNA adenine methyltransferase (dam) to the protein of interest, in our case LaminB, in order to map the interaction of this protein to its specific target sites [273]. The expression of this fusion protein cause a preferential methylation of adenine in the regions surrounding the interaction site of the dam fused-to-LaminB. The resulting adenine methylation, being not performed endogenously in most eukaryotes, represent a unique tag to map the protein of interest interaction sites [273]. For these reasons it can be a system that allows us to better investigate the newly acquired interaction sites of the lamina upon expression of mutant EZH2 Y641 that underlies the massive re-localization of H3K27me3-enriched chromatin.

It would be interesting to further investigate this result by mass spectrometry analysis after having isolated the mutant complex with respect to the WT one, in order to highlight possible novel interactors that could be responsible of this delocalization. To further address this point, we will take advantage of a technique capable of identifying long range interactions, namely Hi-C, that could help to decipher the relationship between histone mark localization and genome activity [274]. We will also use this approach in DLBCL cells to understand if this is a general mechanism or if it is specific of MEFs. In this context, it would be interesting to analyze also the deposition of marks that positively correlate with active transcription and the localization of RNA polymerase II. Indeed, it has been demonstrated [275] that in the nuclear periphery, regions of chromatin DNase I-sensitive are present suggesting that these regions can accomodate poised or active genes. In this sense it would be possible the

presence of regions silenced but responsive to developmental cues to become activated or repressed.

For these reasons, it would be very interesting to investigate the conformation of the chromatin solving the position of the nucleosomes and to map binding regions for specific transcription factor. These aims can be solved through the use of different technologies. These techniques include chromatin digestion using the micrococcal nuclease, that is able to cut nucleosome-free regions, coupled to paired-end sequencing in order to obtain a deep characterization of the positioning and occupancy of nucleosome. Another very useful technique would be the Assay for Transposase-Accessible Chromatin coupled to sequencing (ATAC-seq). This technique allows identifying chromatin accessible regions, to map nucleosome position and to find transcription factor binding sites [276].

In our model, indeed we observed a genome-wide deposition of the H3K27me3. This mark is associated to repression of transcription because it is able to compact the chromatin in a state that is less prone to be bound by transcription factors. Indeed what I will expect is to find more stable nucleosomes, less accessible to interact with the different possible factors. To test these hypotheses, I would like to better characterize also the localization of the elongating Polymerase II and the deposition of the H3K27Ac and H3K36me3 through ChIP-seq data. These two transcriptional active marks correlate usually with the presence of H3K27me1. Since our observations indicate a loss in the deposition of this modification, I will test by ATAC-seq whether upon expression of the mutant form of EZH2 Y641N, it is possible to highlight a change in the positioning of other main key player of transcription. Another interesting level of analysis would be methylation of DNA. Indeed, overall a normal cell presents a certain degree of DNA methylation present on the genome except in CpG islands that are on the other hand very diffuse on the TSS of promoters and can be bound by PcGs. In the case of transformed cells, it is possible to find methylated CpG islands at promoters which in turn can cause altered gene expression. In this regard would be very interesting to investigate if it is possible to score a difference in the deposition of this mark

upon expression of mutant PRC2 and which kind, where eventually present, of relationship can be present between methylated DNA and the given spreading of the H3K27me3. Also this analysis can be very informative and can help elucidating the group of factors that can interact with the complex harboring together the WT and mutant copy of EZH2.

Since our initial hypothesis was based on the fact that the hyper-active mutation could silently perturb the steady state equilibrium of the cell through the accumulation of the repressive mark we thought to challenge the cells using different stimuli in order to investigate the ability of the cells to answer them. We asked whether upon expression of the mutant form of the enzyme the cells could answer to the stress in the same or different way in respect to the ones expressing the WT one. We hypothesized that in response to the stress triggered, cells would have induced a re-deposition of the H3K27me3.

To address this point we analyze the ability of the cell to face firstly a deprivation of nutrients in the starvation condition (Fig. 21). Surprisingly, we could not observe any huge difference in the EZH2 mutant expressing cells compared to WT (Fig. 23) Even if this result surprised us it led us to think that despite the spreading of the H3K27me3 mark, this could not be detrimental for the cells. Indeed cells expressing mutant EZH2 are able to front the serum-deprivation at least shortly after the beginning of the stress as cells expressing the WT form of the enzyme. What we can assume is that despite the spreading of this repressive mark this does not impair pathways that the cells would activate to respond to this particular stimulus. So the mutation seems not to interfere with the common ability of the cells to front nutrient deprivation.

The other analysis we performed addressed an eventual cooperation between EZH2 and oncogenes, in this specific case with c-Myc (Fig. 17). The idea was to understand whether the widespread deposition of H3K27me3 mark could interact with oncogenic signals, mediated by the induction of myc, interfering or synergizing with it towards tumor transformation.

We took advantage of a cellular system that allows the induction of c-myc by addition of OHT to the medium of the cells. The RNA profiles obtained from these cells, comparing cells expressing WT or mutant form of EZH2, showed very few differences (Fig. 20). This, in agreement with the previous analyses on the starvation, suggest that also in this case it seems that the abundant presence of the mark does not influence in any case the transcriptional programs and the pathway that the cells have taken in presence or not of this hyper-activating mutation. These results taken together suggest that the induction of the oncogene myc induce the activation of the same pathways independently of the presence of EZH2 WT or mutant EZH2 Y641N and that also in this case the spread of the repressive mark does not impact on the ability of the cells to activate specific pathways in response to a certain stimulus. It is also possible that spreading is not stochastic but more probably occupies specific regions that in general do not affect the normal behavior and responsiveness of the cells harboring the mutation at least in these conditions.

Lastly we decided to challenge the cells with a functional assay such as reprogramming to pluripotency. We took advantage of the so-called reprogrammable OKSM MEFs harboring the 4 Yamanaka factors, Oct4-Klf4-Sox2-Myc under the control of a doxycycline-responsive promoter. In this way, we could induce reprogramming by addition of doxycycline to the culturing medium (Fig. 25). Before starting, we controlled whether in the presence of mutant EZH2 the H3K27me3 mark was deposited also on genes that are responsible for the establishment of pluripotency. We looked at the ChIP-seq data derived from *Ink4a/Arf-/-* MEF expressing the WT and mutant EZH2 Y641N form and we could observe an increased deposition of the repressive mark on these genes from the portion upstream their promoter and continuously present on their gene body (Fig. 24). So we hypothesized that the mutant harboring cells could be impaired in reprogramming, since it seemed that mutant EZH2 established additional epigenetic control. Indeed we observed a lower efficiency of reprogramming when cells express the mutant form of EZH2. Surprisingly, this effect was evident even in the presence of the WT form of this gene, even

if at a lower extent. Of note, we found also that, as compared to cells infected with the empty vector, where we can observe a 95% of green fluorescent colonies, only ~ 50% of emerged colonies were still expressing the green marker in cells over-expressing the WT form of the enzyme. This fraction was even lower (only ~12% of the colonies) in the mutant-expressing conditions (Fig. 26). This can be attributed to two possible explanations: i) the cell is forced to silence the vector to avoid the maintenance of the H3K27 hypertrimethylation or ii) the only cells that were reprogrammed failed to trimethylate H3K27 genome-wide. These results point to a barrier effect to the reprogramming process in the presence of widespread deposition of the H3K27me3 mark. To address these points, we will characterize the presence of H3K27me3 in the formed colonies. What we expect is that the emerged colonies would not present accumulation of the repressive mark, even in colonies expressing the fluorescent protein.

Also several papers demonstrated that the reprogramming process consists of different steps. The early step is characterized by increased proliferation, metabolic changes, change in histone marks, activation of RNA processing in a stochastic process until the so-called deterministic, hierarchical phase take place, characterized by activation of core of pluripotency starting from Sox2 through Sall4 and Esrrb until Oct4 Nanog and Ezh2 itself and complete epigenetic resetting [266]. In our setting the experiment took 12 days. So the idea was that if these genes were covered by the repressive mark, it could be possible that they could not be activated properly, impairing in the process. Indeed we found a defective process in presence of already extra copies of EZH2 WT that became even more evident in presence of extra-copies of the mutated allele coding for EZH2 enzyme.

Also the reverse experiment would be interesting. Indeed it could be interesting to infect mESC/mIPSC with EZH2-expressing vectors, verify that it would not impair the maintenance of pluripotency, and to perform differentiation experiments toward specific differentiated lineages to investigate whether the presence of the mutant can affect the differentiation potential of these cells.

Taken together these results point to a role of mutant EZH2 in "locking" cells in their characteristic epigenetic/transcriptional status. We speculate that the presence of mutant EZH2 could cause an accumulation of H3K27me3 genome-wide only in sites that are not transcriptionally active in that moment in the cell. This could be due to the hindrance of RNA-polII only in active genes. This would explain the lack of: i) down-regulation of genes at steady state and ii) of strong gene expression changes when cells are challenged with specific stimuli but are still retaining their fate, such as in the case of starvation, myc induction and growth in confluency.

The only case in which we could see an altered response was during the reprogramming experiment. MEFs need to undergo radical transcriptional changes in order to become pluripotent cells and the "transcriptional lock" caused by the deposition of H3K27me3 could block the upregulation of pluripotency-related genes that are normally silent in MEFs. In this setting RNA-polII and transcription factors would fail to bind to promoters of pluripotency-related genes to promote their transcription.

We speculate that this phenomenon would be occurring also in another cell fate determination setting, namely differentiation. Indeed, we will perform differentiation experiments in mESC/iPSC expressing the WT or mutant form of EZH2 (either by lentiviral infection or gene targeting strategies by means of the CRISPR/Cas9 platform [277, 278]), with the hypothesis that also in this case master regulators of differentiation will fail to be expressed, impairing the process.

If proven true, this would explain the "low-differentiation" phenotype in lymphomas. We hypothesize that centroblasts harboring the mutant form of EZH2 would be impaired in their differentiation program, failing to upregulate crucial genes for B-cell maturation, and in combination with tumorigenic mutations this would cause uncontrolled cell proliferation, giving rise to tumor in patients. Indeed also in our setting, cell-to-cell contact inhibition

escape is induced in mutant EZH2-expressing MEFs, pointing to an additional role of this mutation in cell transformation.

The results of this work, in accordance to several lines of research pointing to a re-framing of cancer as a developmental disease, will shed light on the molecular mechanisms that link epigenetic control of cell fate to tumorigenesis, paving the way to the identification of druggable pathways for improved care of lymphoma patients.

5. BIBLIOGRAPHY

1. Luger, K., *Structure and dynamic behavior of nucleosomes.* Curr Opin Genet Dev, 2003. **13**(2): p. 127-35.
2. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
3. Kornberg, R.D., *Chromatin structure: a repeating unit of histones and DNA.* Science, 1974. **184**(4139): p. 868-71.
4. Kornberg, R.D. and J.O. Thomas, *Chromatin structure; oligomers of the histones.* Science, 1974. **184**(4139): p. 865-8.
5. Olins, A.L. and D.E. Olins, *Spheroid chromatin units (v bodies).* Science, 1974. **183**(4122): p. 330-2.
6. Marino-Ramirez, L., et al., *Histone structure and nucleosome stability.* Expert Rev Proteomics, 2005. **2**(5): p. 719-29.
7. Maxson, R., et al., *Distinct organizations and patterns of expression of early and late histone gene sets in the sea urchin.* Nature, 1983. **301**(5896): p. 120-5.
8. Chioda, M., et al., *Histone mRNAs do not accumulate during S phase of either mitotic or endoreduplicative cycles in the chordate Oikopleura dioica.* Mol Cell Biol, 2004. **24**(12): p. 5391-403.
9. Dominski, Z. and W.F. Marzluff, *Formation of the 3' end of histone mRNA.* Gene, 1999. **239**(1): p. 1-14.
10. Burgess, R.J. and Z. Zhang, *Histone chaperones in nucleosome assembly and human disease.* Nat Struct Mol Biol, 2013. **20**(1): p. 14-22.
11. Black, B.E., et al., *Centromere identity, function, and epigenetic propagation across cell divisions.* Cold Spring Harb Symp Quant Biol, 2010. **75**: p. 403-18.

12. Elsaesser, S.J. and C.D. Allis, *HIRA and Daxx constitute two independent histone H3.3-containing predeposition complexes.* Cold Spring Harb Symp Quant Biol, 2010. **75**: p. 27-34.

13. Kim, J.A. and J.E. Haber, *Chromatin assembly factors Asf1 and CAF-1 have overlapping roles in deactivating the DNA damage checkpoint when DNA repair is complete.* Proc Natl Acad Sci U S A, 2009. **106**(4): p. 1151-6.

14. Groth, A., et al., *Regulation of replication fork progression through histone supply and demand.* Science, 2007. **318**(5858): p. 1928-31.

15. Jasencakova, Z., et al., *Replication stress interferes with histone recycling and predeposition marking of new histones.* Mol Cell, 2010. **37**(5): p. 736-43.

16. Adkins, M.W., et al., *The histone chaperone anti-silencing function 1 stimulates the acetylation of newly synthesized histone H3 in S-phase.* J Biol Chem, 2007. **282**(2): p. 1334-40.

17. Adkins, M.W. and J.K. Tyler, *The histone chaperone Asf1p mediates global chromatin disassembly in vivo.* J Biol Chem, 2004. **279**(50): p. 52069-74.

18. Korber, P., et al., *The histone chaperone Asf1 increases the rate of histone eviction at the yeast PHO5 and PHO8 promoters.* J Biol Chem, 2006. **281**(9): p. 5539-45.

19. Belotserkovskaya, R., et al., *FACT facilitates transcription-dependent nucleosome alteration.* Science, 2003. **301**(5636): p. 1090-3.

20. Kireeva, M.L., et al., *Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription.* Mol Cell, 2002. **9**(3): p. 541-52.

21. Orphanides, G., et al., *The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins.* Nature, 1999. **400**(6741): p. 284-8.

22. Drane, P., et al., *The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3.* Genes Dev, 2010. **24**(12): p. 1253-65.

23. Goldberg, A.D., et al., *Distinct factors control histone variant H3.3 localization at specific genomic regions.* Cell, 2010. **140**(5): p. 678-91.

24. Mattout, A., D.S. Cabianca, and S.M. Gasser, *Chromatin states and nuclear organization in development--a view from the nuclear lamina.* Genome Biol, 2015. **16**: p. 174.

25. Li, G. and D. Reinberg, *Chromatin higher-order structures and gene regulation.* Curr Opin Genet Dev, 2011. **21**(2): p. 175-86.

26. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis.* Proc Natl Acad Sci U S A, 1964. **51**: p. 786-94.

27. Russo, V.E., R.A. Martienssen, and A.D. Riggs, *Epigenetic mechanisms of gene regulation*. 1996: Cold Spring Harbor Laboratory Press.

28. Levenson, J.M. and J.D. Sweatt, *Epigenetic mechanisms in memory formation.* Nat Rev Neurosci, 2005. **6**(2): p. 108-18.

29. Keverne, B., *Monoallelic gene expression and mammalian evolution.* Bioessays, 2009. **31**(12): p. 1318-26.

30. Watt, F. and P.L. Molloy, *Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter.* Genes Dev, 1988. **2**(9): p. 1136-43.

31. Rothbart, S.B. and B.D. Strahl, *Interpreting the language of histone and DNA modifications.* Biochim Biophys Acta, 2014. **1839**(8): p. 627-43.

32. Blackledge, N.P. and R. Klose, *CpG island chromatin: a platform for gene regulation.* Epigenetics, 2011. **6**(2): p. 147-52.

33. Jenuwein, T., *Re-SET-ting heterochromatin by histone methyltransferases.* Trends Cell Biol, 2001. **11**(6): p. 266-73.

34. Unnikrishnan, A., P.R. Gafken, and T. Tsukiyama, *Dynamic changes in histone acetylation regulate origins of DNA replication.* Nat Struct Mol Biol, 2010. **17**(4): p. 430-7.

35. Zeng, L. and M.M. Zhou, *Bromodomain: an acetyl-lysine binding domain.* FEBS Lett, 2002. **513**(1): p. 124-8.

36. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nat Genet, 2008. **40**(7): p. 897-903.

37. Tie, F., et al., *CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing.* Development, 2009. **136**(18): p. 3131-41.

38. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

39. Oki, M., H. Aihara, and T. Ito, *Role of histone phosphorylation in chromatin dynamics and its implications in diseases.* Subcell Biochem, 2007. **41**: p. 319-36.

40. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.

41. Tan, M., et al., *Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification.* Cell, 2011. **146**(6): p. 1016-28.

42. Byvoet, P., et al., *The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells.* Arch Biochem Biophys, 1972. **148**(2): p. 558-67.

43. Shi, Y., et al., *Histone demethylation mediated by the nuclear amine oxidase homolog LSD1.* Cell, 2004. **119**(7): p. 941-53.

44. Greer, E.L. and Y. Shi, *Histone methylation: a dynamic mark in health, disease and inheritance.* Nat Rev Genet, 2012. **13**(5): p. 343-57.

45. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nat Struct Mol Biol, 2007. **14**(11): p. 1025-40.

46. Jenuwein, T., et al., *SET domain proteins modulate chromatin domains in eu- and heterochromatin.* Cell Mol Life Sci, 1998. **54**(1): p. 80-93.

47. Nguyen, A.T. and Y. Zhang, *The diverse functions of Dot1 and H3K79 methylation.* Genes Dev, 2011. **25**(13): p. 1345-58.

48. Helin, K. and D. Dhanak, *Chromatin proteins and modifications as drug targets.* Nature, 2013. **502**(7472): p. 480-8.

49. 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, 2005. **280**(7): p. 5563-70.

50. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.* Genes Dev, 2002. **16**(22): p. 2893-905.

51. Peters, A.H., et al., *Partitioning and plasticity of repressive histone methylation states in mammalian chromatin.* Mol Cell, 2003. **12**(6): p. 1577-89.

52. Tachibana, M., et al., *G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis.* Genes Dev, 2002. **16**(14): p. 1779-91.

53. Melcher, M., et al., *Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression.* Mol Cell Biol, 2000. **20**(10): p. 3728-41.

54. Briggs, S.D., et al., *Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae.* Genes Dev, 2001. **15**(24): p. 3286-95.

55. Ruthenburg, A.J., C.D. Allis, and J. Wysocka, *Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark.* Mol Cell, 2007. **25**(1): p. 15-30.

56. Glaser, S., et al., *Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development.* Development, 2006. **133**(8): p. 1423-32.

57. Lee, S., et al., *Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15392-7.

58. Yu, B.D., et al., *Altered Hox expression and segmental identity in Mll-mutant mice.* Nature, 1995. **378**(6556): p. 505-8.

59. Yokoyama, A., et al., *Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression.* Mol Cell Biol, 2004. **24**(13): p. 5639-49.

60. Milne, T.A., et al., *Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors.* Proc Natl Acad Sci U S A, 2005. **102**(3): p. 749-54.

61. Ruthenburg, A.J., et al., *Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex.* Nat Struct Mol Biol, 2006. **13**(8): p. 704-12.

62. Wysocka, J., et al., *Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1.* Genes Dev, 2003. **17**(7): p. 896-911.

63. Schwartz, Y.B. and V. Pirrotta, *Polycomb silencing mechanisms and the management of genomic programmes.* Nat Rev Genet, 2007. **8**(1): p. 9-22.

64. Schmitges, F.W., et al., *Histone methylation by PRC2 is inhibited by active chromatin marks.* Mol Cell, 2011. **42**(3): p. 330-41.

65. Pray-Grant, M.G., et al., *Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation.* Nature, 2005. **433**(7024): p. 434-8.

66. Santos-Rosa, H., et al., *Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin.* Mol Cell, 2003. **12**(5): p. 1325-32.

67. Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells.* Cell, 2007. **130**(1): p. 77-88.

68. Voigt, P., W.W. Tee, and D. Reinberg, *A double take on bivalent promoters.* Genes Dev, 2013. **27**(12): p. 1318-38.

69. Lettice, L.A., et al., *A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly.* Hum Mol Genet, 2003. **12**(14): p. 1725-35.

70. He, J., et al., *Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly.* Nucleic Acids Res, 2011. **39**(10): p. 4464-74.

71. Junion, G., et al., *A transcription factor collective defines cardiac cell fate and reflects lineage history.* Cell, 2012. **148**(3): p. 473-86.

72. Heintzman, N.D. and B. Ren, *Finding distal regulatory elements in the human genome.* Curr Opin Genet Dev, 2009. **19**(6): p. 541-9.

73. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.

74. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

75. Rada-Iglesias, A., et al., *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-83.

76. Wagner, E.J. and P.B. Carpenter, *Understanding the language of Lys36 methylation at histone H3.* Nat Rev Mol Cell Biol, 2012. **13**(2): p. 115-26.

77. 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, 1982. **102**(1): p. 49-70.

78. Lewis, E.B., *A gene complex controlling segmentation in Drosophila.* Nature, 1978. **276**(5688): p. 565-70.

79. Schuettengruber, B. and G. Cavalli, *Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice.* Development, 2009. **136**(21): p. 3531-42.

80. Morey, L. and K. Helin, *Polycomb group protein-mediated repression of transcription.* Trends Biochem Sci, 2010. **35**(6): p. 323-32.

81. Haupt, Y., et al., *Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice.* Cell, 1991. **65**(5): p. 753-63.

82. van Lohuizen, M., et al., *Sequence similarity between the mammalian bmi-1 proto-oncogene and the Drosophila regulatory genes Psc and Su(z)2.* Nature, 1991. **353**(6342): p. 353-5.

83. Schumacher, A. and T. Magnuson, *Murine Polycomb- and trithorax-group genes regulate homeotic pathways and beyond.* Trends Genet, 1997. **13**(5): p. 167-70.

84. Pietersen, A.M. and M. van Lohuizen, *Stem cell regulation by polycomb repressors: postponing commitment.* Curr Opin Cell Biol, 2008. **20**(2): p. 201-7.

85. Sparmann, A. and M. van Lohuizen, *Polycomb silencers control cell fate, development and cancer.* Nat Rev Cancer, 2006. **6**(11): p. 846-56.

86. Casanova, M., et al., *Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells.* Development, 2011. **138**(8): p. 1471-82.

87. Jullien, P.E., et al., *Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis.* Curr Biol, 2006. **16**(5): p. 486-92.

88. Makarevich, G., et al., *Different Polycomb group complexes regulate common target genes in Arabidopsis.* EMBO Rep, 2006. **7**(9): p. 947-52.

89. Schubert, D., et al., *Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27.* EMBO J, 2006. **25**(19): p. 4638-49.

90. Golbabapour, S., M.A. Abdulla, and M. Hajrezaei, *A concise review on epigenetic regulation: insight into molecular mechanisms.* Int J Mol Sci, 2011. **12**(12): p. 8661-94.

91. Aranda, S., G. Mas, and L. Di Croce, *Regulation of gene transcription by Polycomb proteins.* Sci Adv, 2015. **1**(11): p. e1500737.

92. Scelfo, A., A. Piunti, and D. Pasini, *The controversial role of the Polycomb group proteins in transcription and cancer: how much do we not understand Polycomb proteins?* FEBS J, 2015. **282**(9): p. 1703-22.

93. Bracken, A.P. and K. Helin, *Polycomb group proteins: navigators of lineage pathways led astray in cancer.* Nat Rev Cancer, 2009. **9**(11): p. 773-84.

94. Blackledge, N.P., et al., *Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation.* Cell, 2014. **157**(6): p. 1445-59.

95. Gao, Z., et al., *PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes.* Mol Cell, 2012. **45**(3): p. 344-56.

96. Vandamme, J., et al., *Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells.* Mol Cell Proteomics, 2011. **10**(4): p. M110 002642.

97. Cao, R. and Y. Zhang, *SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex.* Mol Cell, 2004. **15**(1): p. 57-67.

98. Tavares, L., et al., *RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3.* Cell, 2012. **148**(4): p. 664-78.

99. Morey, L., et al., *RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells.* Cell Rep, 2013. **3**(1): p. 60-9.

100. van Lohuizen, M., et al., *Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging.* Cell, 1991. **65**(5): p. 737-52.

101. Barrero, M.J. and J.C. Izpisua Belmonte, *Polycomb complex recruitment in pluripotent stem cells.* Nat Cell Biol, 2013. **15**(4): p. 348-50.

102. Blackledge, N.P., et al., *CpG islands recruit a histone H3 lysine 36 demethylase.* Mol Cell, 2010. **38**(2): p. 179-90.

103. He, J., et al., *Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes.* Nat Cell Biol, 2013. **15**(4): p. 373-84.

104. Endoh, M., et al., *Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity.* Development, 2008. **135**(8): p. 1513-24.

105. Trojer, P., et al., *L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure.* Mol Cell, 2011. **42**(4): p. 438-50.

106. Ang, Y.S., et al., *Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network.* Cell, 2011. **145**(2): p. 183-97.

107. Qin, J., et al., *The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development.* Cell Stem Cell, 2012. **11**(3): p. 319-32.

108. Trievel, R.C. and A. Shilatifard, *WDR5, a complexed protein.* Nat Struct Mol Biol, 2009. **16**(7): p. 678-80.

109. Amati, B., et al., *Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max.* Nature, 1992. **359**(6394): p. 423-6.

110. Attwooll, C., et al., *A novel repressive E2F6 complex containing the polycomb group protein, EPC1, that interacts with EZH2 in a proliferation-specific manner.* J Biol Chem, 2005. **280**(2): p. 1199-208.

111. Ogawa, H., et al., *A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells.* Science, 2002. **296**(5570): p. 1132-6.

112. Hayakawa, T. and J. Nakayama, *Physiological roles of class I HDAC complex and histone demethylase.* J Biomed Biotechnol, 2011. **2011**: p. 129383.

113. Ketel, C.S., et al., *Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes.* Mol Cell Biol, 2005. **25**(16): p. 6857-68.

114. Boyer, L.A., et al., *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. **441**(7091): p. 349-53.

115. Mohn, F., et al., *Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.* Mol Cell, 2008. **30**(6): p. 755-66.

116. Margueron, R. and D. Reinberg, *The Polycomb complex PRC2 and its mark in life.* Nature, 2011. **469**(7330): p. 343-9.

117. van Kruijsbergen, I., S. Hontelez, and G.J. Veenstra, *Recruiting polycomb to chromatin.* Int J Biochem Cell Biol, 2015. **67**: p. 177-87.

118. 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, 2012. **109**(8): p. 2989-94.

119. Cao, R., et al., *Role of hPHF1 in H3K27 methylation and Hox gene silencing.* Mol Cell Biol, 2008. **28**(5): p. 1862-72.

120. Sarma, K., et al., *Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo.* Mol Cell Biol, 2008. **28**(8): p. 2718-31.

121. 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, 2010. **6**(2): p. 153-66.

122. Zhang, Z., et al., *PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming.* Stem Cells, 2011. **29**(2): p. 229-40.

123. Boulay, G., et al., *Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes.* Biochem J, 2011. **434**(2): p. 333-42.

124. Ballare, C., et al., *Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity.* Nat Struct Mol Biol, 2012. **19**(12): p. 1257-65.

125. Pasini, D., et al., *JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells.* Nature, 2010. **464**(7286): p. 306-10.

126. Peng, J.C., et al., *Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells.* Cell, 2009. **139**(7): p. 1290-302.

127. Landeira, D., et al., *Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators.* Nat Cell Biol, 2010. **12**(6): p. 618-24.

128. Shen, X., et al., *Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells.* Cell, 2009. **139**(7): p. 1303-14.

129. Ferrari, K.J., et al., *Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity.* Mol Cell, 2014. **53**(1): p. 49-62.

130. Kassis, J.A. and J.L. Brown, *Polycomb group response elements in Drosophila and vertebrates.* Adv Genet, 2013. **81**: p. 83-118.

131. Ringrose, L., et al., *Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster.* Dev Cell, 2003. **5**(5): p. 759-71.

132. Schuettengruber, B., et al., *Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos.* PLoS Biol, 2009. **7**(1): p. e13.

133. Schwartz, Y.B., et al., *Genome-wide analysis of Polycomb targets in Drosophila melanogaster.* Nat Genet, 2006. **38**(6): p. 700-5.

134. Tolhuis, B., et al., *Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster.* Nat Genet, 2006. **38**(6): p. 694-9.

135. Negre, N., et al., *Chromosomal distribution of PcG proteins during Drosophila development.* PLoS Biol, 2006. **4**(6): p. e170.

136. Hauenschild, A., et al., *Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species.* PLoS Biol, 2008. **6**(10): p. e261.

137. Vella, P., et al., *Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells.* Nucleic Acids Res, 2012. **40**(8): p. 3403-18.

138. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.

139. Sing, A., et al., *A vertebrate Polycomb response element governs segmentation of the posterior hindbrain.* Cell, 2009. **138**(5): p. 885-97.

140. Squazzo, S.L., et al., *Suz12 binds to silenced regions of the genome in a cell-type-specific manner.* Genome Res, 2006. **16**(7): p. 890-900.

141. Dietrich, N., et al., *REST-mediated recruitment of polycomb repressor complexes in mammalian cells.* PLoS Genet, 2012. **8**(3): p. e1002494.

142. Herranz, N., et al., *Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor.* Mol Cell Biol, 2008. **28**(15): p. 4772-81.

143. Villa, R., et al., *Role of the polycomb repressive complex 2 in acute promyelocytic leukemia.* Cancer Cell, 2007. **11**(6): p. 513-25.

144. Boukarabila, H., et al., *The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation.* Genes Dev, 2009. **23**(10): p. 1195-206.

145. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.* Science, 2008. **322**(5902): p. 750-6.

146. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing.* Science, 2002. **298**(5595): p. 1039-43.

147. Wang, L., et al., *Hierarchical recruitment of polycomb group silencing complexes.* Mol Cell, 2004. **14**(5): p. 637-46.

148. Ku, M., et al., *Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains.* PLoS Genet, 2008. **4**(10): p. e1000242.

149. Pasini, D., et al., *The polycomb group protein Suz12 is required for embryonic stem cell differentiation.* Mol Cell Biol, 2007. **27**(10): p. 3769-79.

150. Schoeftner, S. and M.A. Blasco, *Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II.* Nat Cell Biol, 2008. **10**(2): p. 228-36.

151. Kalb, R., et al., *Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression.* Nat Struct Mol Biol, 2014. **21**(6): p. 569-71.

152. Ishigaki, S., et al., *Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions.* Sci Rep, 2012. **2**: p. 529.

153. Schwartz, J.C., et al., *FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2.* Genes Dev, 2012. **26**(24): p. 2690-5.

154. Punga, T. and M. Buhler, *Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation.* EMBO Mol Med, 2010. **2**(4): p. 120-9.

155. Hansen, R.S., et al., *The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome.* Proc Natl Acad Sci U S A, 1999. **96**(25): p. 14412-7.

156. Lana, E., et al., *DNA replication is altered in Immunodeficiency Centromeric instability Facial anomalies (ICF) cells carrying DNMT3B mutations.* Eur J Hum Genet, 2012. **20**(10): p. 1044-50.

157. Jeffries, M.A. and A.H. Sawalha, *Epigenetics in systemic lupus erythematosus: leading the way for specific therapeutic agents.* Int J Clin Rheumtol, 2011. **6**(4): p. 423-439.

158. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.

159. Karakosta, A., et al., *Genetic models of human cancer as a multistep process. Paradigm models of colorectal cancer, breast cancer, and chronic myelogenous and acute lymphoblastic leukaemia.* J Exp Clin Cancer Res, 2005. **24**(4): p. 505-14.

160. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer.* Cell, 2007. **128**(4): p. 683-92.

161. Jones, P.A. and P.W. Laird, *Cancer epigenetics comes of age.* Nat Genet, 1999. **21**(2): p. 163-7.

162. Bhattacharjee, D., S. Shenoy, and K.L. Bairy, *DNA Methylation and Chromatin Remodeling: The Blueprint of Cancer Epigenetics.* Scientifica (Cairo), 2016. **2016**: p. 6072357.

163. Muntean, A.G. and J.L. Hess, *Epigenetic dysregulation in cancer.* Am J Pathol, 2009. **175**(4): p. 1353-61.

164. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics.* Nat Rev Genet, 2008. **9**(6): p. 465-76.

165. Curradi, M., et al., *Molecular mechanisms of gene silencing mediated by DNA methylation.* Mol Cell Biol, 2002. **22**(9): p. 3157-73.

166. Eden, A., et al., *Chromosomal instability and tumors promoted by DNA hypomethylation.* Science, 2003. **300**(5618): p. 455.

167. Howard, G., et al., *Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice.* Oncogene, 2008. **27**(3): p. 404-8.

168. Wilson, A.S., B.E. Power, and P.L. Molloy, *DNA hypomethylation and human diseases.* Biochim Biophys Acta, 2007. **1775**(1): p. 138-62.

169. Ogawa, O., et al., *Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour.* Nature, 1993. **362**(6422): p. 749-51.

170. Cui, H., et al., *Loss of IGF2 imprinting: a potential marker of colorectal cancer risk.* Science, 2003. **299**(5613): p. 1753-5.

171. Baylin, S.B., *DNA methylation and gene silencing in cancer.* Nat Clin Pract Oncol, 2005. **2 Suppl 1**: p. S4-11.

172. Long, C., et al., *Promoter hypermethylation of the RUNX3 gene in esophageal squamous cell carcinoma.* Cancer Invest, 2007. **25**(8): p. 685-90.

173. Seligson, D.B., et al., *Global histone modification patterns predict risk of prostate cancer recurrence.* Nature, 2005. **435**(7046): p. 1262-6.

174. Grunstein, M., *Histone acetylation in chromatin structure and transcription.* Nature, 1997. **389**(6649): p. 349-52.

175. Eckner, R., et al., *The adenovirus E1A-associated 300-kD protein exhibits properties of a transcriptional coactivator and belongs to an evolutionarily conserved family.* Cold Spring Harb Symp Quant Biol, 1994. **59**: p. 85-95.

176. Rasti, M., et al., *Recruitment of CBP/p300, TATA-binding protein, and S8 to distinct regions at the N terminus of adenovirus E1A.* J Virol, 2005. **79**(9): p. 5594-605.

177. Horwitz, G.A., et al., *Adenovirus small e1a alters global patterns of histone modification.* Science, 2008. **321**(5892): p. 1084-5.

178. Ferrari, R., et al., *Epigenetic reprogramming by adenovirus e1a.* Science, 2008. **321**(5892): p. 1086-8.

179. Davis, Z.A., et al., *Divergence from the germ-line sequence in unmutated chronic lymphocytic leukemia is due to somatic mutation rather than polymorphisms.* Blood, 2003. **102**(8): p. 3075.

180. Ida, K., et al., *Adenoviral E1A-associated protein p300 is involved in acute myeloid leukemia with t(11;22)(q23;q13).* Blood, 1997. **90**(12): p. 4699-704.

181. Bolden, J.E., M.J. Peart, and R.W. Johnstone, *Anticancer activities of histone deacetylase inhibitors.* Nat Rev Drug Discov, 2006. **5**(9): p. 769-84.

182. Witt, O., et al., *HDAC family: What are the cancer relevant targets?* Cancer Lett, 2009. **277**(1): p. 8-21.

183. Weichert, W., et al., *Class I histone deacetylase expression has independent prognostic impact in human colorectal cancer: specific role of class I histone deacetylases in vitro and in vivo.* Clin Cancer Res, 2008. **14**(6): p. 1669-77.

184. Bhaskara, S., et al., *Deletion of histone deacetylase 3 reveals critical roles in S phase progression and DNA damage control.* Mol Cell, 2008. **30**(1): p. 61-72.

185. Park, J.H., et al., *Class II histone deacetylases play pivotal roles in heat shock protein 90-mediated proteasomal degradation of vascular endothelial growth factor receptors.* Biochem Biophys Res Commun, 2008. **368**(2): p. 318-22.

186. Zelent, A., et al., *Translocations of the RARalpha gene in acute promyelocytic leukemia.* Oncogene, 2001. **20**(49): p. 7186-203.

187. Bereshchenko, O.R., W. Gu, and R. Dalla-Favera, *Acetylation inactivates the transcriptional repressor BCL6.* Nat Genet, 2002. **32**(4): p. 606-13.

188. Choi, J.H., et al., *Expression profile of histone deacetylase 1 in gastric cancer tissues.* Jpn J Cancer Res, 2001. **92**(12): p. 1300-4.

189. Halkidou, K., et al., *Upregulation and nuclear recruitment of HDAC1 in hormone refractory prostate cancer.* Prostate, 2004. **59**(2): p. 177-89.

190. Zhu, P., et al., *Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis.* Cancer Cell, 2004. **5**(5): p. 455-63.

191. Huang, B.H., et al., *Inhibition of histone deacetylase 2 increases apoptosis and p21Cip1/WAF1 expression, independent of histone deacetylase 1.* Cell Death Differ, 2005. **12**(4): p. 395-404.

192. Ellis, L., P.W. Atadja, and R.W. Johnstone, *Epigenetics in cancer: targeting chromatin modifications.* Mol Cancer Ther, 2009. **8**(6): p. 1409-20.

193. Ellis, L., H. Hammers, and R. Pili, *Targeting tumor angiogenesis with histone deacetylase inhibitors.* Cancer Lett, 2009. **280**(2): p. 145-53.

194. Michan, S. and D. Sinclair, *Sirtuins in mammals: insights into their biological function.* Biochem J, 2007. **404**(1): p. 1-13.

195. Wong, S. and J.D. Weber, *Deacetylation of the retinoblastoma tumour suppressor protein by SIRT1.* Biochem J, 2007. **407**(3): p. 451-60.

196. Fuks, F., et al., *The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation.* J Biol Chem, 2003. **278**(6): p. 4035-40.

197. Peters, A.H., et al., *Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability.* Cell, 2001. **107**(3): p. 323-37.

198. Baylin, S.B. and P.A. Jones, *A decade of exploring the cancer epigenome - biological and translational implications.* Nat Rev Cancer, 2011. **11**(10): p. 726-34.

199. Hock, H., *A complex Polycomb issue: the two faces of EZH2 in cancer.* Genes Dev, 2012. **26**(8): p. 751-5.

200. Mochizuki-Kashio, M., G.R. Wendt, and A. Iwama, *Tumor suppressor function of the polycomb group genes.* Cell Cycle, 2012. **11**(11): p. 2043-4.

201. Bracken, A.P., et al., *EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer.* EMBO J, 2003. **22**(20): p. 5323-35.

202. Ntziachristos, P., et al., *Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia.* Nat Med, 2012. **18**(2): p. 298-301.

203. Oguro, H., et al., *Lethal myelofibrosis induced by Bmi1-deficient hematopoietic cells unveils a tumor suppressor function of the polycomb group genes.* J Exp Med, 2012. **209**(3): p. 445-54.

204. Simon, C., et al., *A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia.* Genes Dev, 2012. **26**(7): p. 651-6.

205. Lund, K., P.D. Adams, and M. Copland, *EZH2 in normal and malignant hematopoiesis.* Leukemia, 2014. **28**(1): p. 44-9.

206. Score, J., et al., *Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms.* Blood, 2012. **119**(5): p. 1208-13.

207. Xie, H., et al., *Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner.* Cell Stem Cell, 2014. **14**(1): p. 68-80.

208. Puda, A., et al., *Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies.* Am J Hematol, 2012. **87**(3): p. 245-50.

209. Ueda, T., et al., *EED mutants impair polycomb repressive complex 2 in myelodysplastic syndrome and related neoplasms.* Leukemia, 2012. **26**(12): p. 2557-60.

210. Neff, T., et al., *Polycomb repressive complex 2 is required for MLL-AF9 leukemia.* Proc Natl Acad Sci U S A, 2012. **109**(13): p. 5028-33.

211. Abdel-Wahab, O., et al., *ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression.* Cancer Cell, 2012. **22**(2): p. 180-93.

212. Sashida, G., et al., *Ezh2 loss promotes development of myelodysplastic syndrome but attenuates its predisposition to leukaemic transformation.* Nat Commun, 2014. **5**: p. 4177.

213. Bracken, A.P., et al., *The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells.* Genes Dev, 2007. **21**(5): p. 525-30.

214. Dietrich, N., et al., *Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus.* EMBO J, 2007. **26**(6): p. 1637-48.

215. Gil, J., et al., *Polycomb CBX7 has a unifying role in cellular lifespan.* Nat Cell Biol, 2004. **6**(1): p. 67-72.

216. Piunti, A., et al., *Polycomb proteins control proliferation and transformation independently of cell cycle checkpoints by regulating DNA replication.* Nat Commun, 2014. **5**: p. 3649.

217. Dawson, M.A. and T. Kouzarides, *Cancer epigenetics: from mechanism to therapy.* Cell, 2012. **150**(1): p. 12-27.

218. Serrano, M., *The INK4a/ARF locus in murine tumorigenesis.* Carcinogenesis, 2000. **21**(5): p. 865-9.

219. Piunti, A. and D. Pasini, *Epigenetic factors in cancer development: polycomb group proteins.* Future Oncol, 2011. **7**(1): p. 57-75.

220. Caganova, M., et al., *Germinal center dysregulation by histone methyltransferase EZH2 promotes lymphomagenesis.* J Clin Invest, 2013. **123**(12): p. 5009-22.

221. van Galen, J.C., et al., *Distinct expression patterns of polycomb oncoproteins and their binding partners during the germinal center reaction.* Eur J Immunol, 2004. **34**(7): p. 1870-81.

222. Wang, G.G., K.D. Konze, and J. Tao, *Polycomb genes, miRNA, and their deregulation in B-cell malignancies.* Blood, 2015. **125**(8): p. 1217-25.

223. Visser, H.P., et al., *The Polycomb group protein EZH2 is upregulated in proliferating, cultured human mantle cell lymphoma.* Br J Haematol, 2001. **112**(4): p. 950-8.

224. Beguelin, W., et al., *EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation.* Cancer Cell, 2013. **23**(5): p. 677-92.

225. Morin, R.D., et al., *Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin.* Nat Genet, 2010. **42**(2): p. 181-5.

226. Majer, C.R., et al., *A687V EZH2 is a gain-of-function mutation found in lymphoma patients.* FEBS Lett, 2012. **586**(19): p. 3448-51.

227. Bodor, C., et al., *EZH2 mutations are frequent and represent an early event in follicular lymphoma.* Blood, 2013. **122**(18): p. 3165-8.

228. 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, 2015. **6**(5): p. 2928-38.

229. Sneeringer, C.J., et al., *Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas.* Proc Natl Acad Sci U S A, 2010. **107**(49): p. 20980-5.

230. 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, 2011. **117**(8): p. 2451-9.

231. McCabe, M.T., et al., *EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations.* Nature, 2012. **492**(7427): p. 108-12.

232. Ott, H.M., et al., *A687V EZH2 is a driver of histone H3 lysine 27 (H3K27) hypertrimethylation.* Mol Cancer Ther, 2014. **13**(12): p. 3062-73.

233. Swalm, B.M., et al., *Reaction coupling between wild-type and disease-associated mutant EZH2.* ACS Chem Biol, 2014. **9**(11): p. 2459-64.

234. Antonysamy, S., et al., *Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain.* PLoS One, 2013. **8**(12): p. e84147.

235. Xu, B., et al., *Selective inhibition of EZH2 and EZH1 enzymatic activity by a small molecule suppresses MLL-rearranged leukemia.* Blood, 2015. **125**(2): p. 346-57.

236. Kim, W., et al., *Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer.* Nat Chem Biol, 2013. **9**(10): p. 643-50.

237. Glazer, R.I., et al., *3-Deazaneplanocin: a new and potent inhibitor of S-adenosylhomocysteine hydrolase and its effects on human promyelocytic leukemia cell line HL-60.* Biochem Biophys Res Commun, 1986. **135**(2): p. 688-94.

238. Tan, J., et al., *Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells.* Genes Dev, 2007. **21**(9): p. 1050-63.

239. Miranda, T.B., et al., *DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation.* Mol Cancer Ther, 2009. **8**(6): p. 1579-88.

240. Knutson, S.K., et al., *A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells.* Nat Chem Biol, 2012. **8**(11): p. 890-6.

241. Verma, S.K., et al., *Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2.* ACS Med Chem Lett, 2012. **3**(12): p. 1091-6.

242. Qi, W., et al., *Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation.* Proc Natl Acad Sci U S A, 2012. **109**(52): p. 21360-5.

243. Kim, K.H. and C.W. Roberts, *Targeting EZH2 in cancer.* Nat Med, 2016. **22**(2): p. 128-34.

244. Campaner, S., et al., *Cdk2 suppresses cellular senescence induced by the c-myc oncogene.* Nat Cell Biol, 2010. **12**(1): p. 54-9; sup pp 1-14.

245. Beard, C., et al., *Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells.* Genesis, 2006. **44**(1): p. 23-8.

246. Stadtfeld, M., et al., *A reprogrammable mouse strain from gene-targeted embryonic stem cells.* Nat Methods, 2010. **7**(1): p. 53-5.

247. Olsen, J.V., et al., *Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap.* Mol Cell Proteomics, 2005. **4**(12): p. 2010-21.

248. Bonaldi, T., J.T. Regula, and A. Imhof, *The use of mass spectrometry for the analysis of histone modifications.* Methods Enzymol, 2004. **377**: p. 111-30.

249. Rappsilber, J., M. Mann, and Y. Ishihama, *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.* Nat Protoc, 2007. **2**(8): p. 1896-906.

250. Cox, J. and M. Mann, *Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap.* J Am Soc Mass Spectrom, 2009. **20**(8): p. 1477-85.

251. Cox, J., et al., *A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics.* Nat Protoc, 2009. **4**(5): p. 698-705.

252. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.

253. Thakur, S.S., et al., *Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation.* Mol Cell Proteomics, 2011. **10**(8): p. M110 003699.

254. Pasini, D., et al., *Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity.* EMBO J, 2004. **23**(20): p. 4061-71.

255. Orlando, D.A., et al., *Quantitative ChIP-Seq normalization reveals global modulation of the epigenome.* Cell Rep, 2014. **9**(3): p. 1163-70.

256. Hayward, W.S., B.G. Neel, and S.M. Astrin, *Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis.* Nature, 1981. **290**(5806): p. 475-80.

257. Eilers, M. and R.N. Eisenman, *Myc's broad reach.* Genes Dev, 2008. **22**(20): p. 2755-66.

258. Nie, Z., et al., *c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells.* Cell, 2012. **151**(1): p. 68-79.

259. Brown, J.R., et al., *Integrative genomic analysis implicates gain of PIK3CA at 3q26 and MYC at 8q24 in chronic lymphocytic leukemia.* Clin Cancer Res, 2012. **18**(14): p. 3791-802.

260. Dave, S.S., et al., *Molecular diagnosis of Burkitt's lymphoma.* N Engl J Med, 2006. **354**(23): p. 2431-42.

261. Zhang, X., et al., *Myc represses miR-15a/miR-16-1 expression through recruitment of HDAC3 in mantle cell and other non-Hodgkin B-cell lymphomas.* Oncogene, 2012. **31**(24): p. 3002-8.

262. Sabo, A., et al., *Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis.* Nature, 2014. **511**(7510): p. 488-92.

263. Papp, B. and K. Plath, *Reprogramming to pluripotency: stepwise resetting of the epigenetic landscape.* Cell Res, 2011. **21**(3): p. 486-501.

264. Papp, B. and K. Plath, *Epigenetics of reprogramming to induced pluripotency.* Cell, 2013. **152**(6): p. 1324-43.

265. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.

266. Buganim, Y., D.A. Faddah, and R. Jaenisch, *Mechanisms and models of somatic cell reprogramming.* Nat Rev Genet, 2013. **14**(6): p. 427-39.

267. Dubois, S., et al., *Immunohistochemical and genomic profiles of diffuse large B-cell lymphomas: implications for targeted EZH2 inhibitor therapy?* Oncotarget, 2015. **6**(18): p. 16712-24.

268. Bickmore, W.A. and B. van Steensel, *Genome architecture: domain organization of interphase chromosomes.* Cell, 2013. **152**(6): p. 1270-84.

269. Schneider, R. and R. Grosschedl, *Dynamics and interplay of nuclear architecture, genome organization, and gene expression.* Genes Dev, 2007. **21**(23): p. 3027-43.

270. Williams, R.R., et al., *Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus.* J Cell Sci, 2006. **119**(Pt 1): p. 132-40.

271. Zhou, J., et al., *Replication and subnuclear location dynamics of the immunoglobulin heavy-chain locus in B-lineage cells.* Mol Cell Biol, 2002. **22**(13): p. 4876-89.

272. Zink, D., et al., *Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei.* J Cell Biol, 2004. **166**(6): p. 815-25.

273. Greil, F., C. Moorman, and B. van Steensel, *DamID: mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase.* Methods Enzymol, 2006. **410**: p. 342-59.

274. Belton, J.M., et al., *Hi-C: a comprehensive technique to capture the conformation of genomes.* Methods, 2012. **58**(3): p. 268-76.

275. Hutchison, N. and H. Weintraub, *Localization of DNAase I-sensitive sequences to specific regions of interphase nuclei.* Cell, 1985. **43**(2 Pt 1): p. 471-82.

276. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.

277. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.* Cell, 2013. **154**(6): p. 1380-9.

278. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-308.

6. ACKNOWLEDGEMENTS

I want strongly to thank the Istituto Italiano di Tecnologia for sustaining me in these four years of research. I would like to thank my boss, Diego, for giving me the opportunity to work on this project in this excellence place and for helping me out to take my first steps in the mammalian epigenetics field. I would like to thank our entire lab, past members included, for these intense four years.

I would like to thank our collaborators in Tiziana Bonaldi's group, especially Alessandro Cuomo and Alessio Silvola, for helping me with the mass spectrometry analyses. I would like to thank our collaborators in Amati's and Campaners's group, especially Sara Rohban and Mirko Doni, for providing me the $3T3^{MycER}$ MEFs. I would like to thank our collaborators in Stefano Casola's group, in particular, Mahshid Rahmat, Federica Mainoldi and Federica Zanardi, for sharing with me reagents and expertise on lymphoma cells. I would like to thank our collaborators in Giuseppe Testa's group, especially Pietro Lo Riso for providing the MEF OKSM and for introducing me into reprogramming world.

I would like to thank my internal advisor Dr. Susanna Chiocca and my external advisor Prof. Adrian Bracken, for giving me punctual insights on the project and for their helpful comments on my work.

A big thank goes to Sriganesh Jammula for having performed all the bioinformatic analyses in the project and for his kindness together with Andrea Piunti with whom I started and Andrea Scelfo with whom I continued this project. I would like to thank my friends and colleagues Andrea Scelfo and Elisa Lavarone for the continuous support, the big help, the exciting scientific conversations and for their essential and unavoidable friendship (and these words..”they're not enough”).

A huge thank you is dedicated to my relatives: to my father that would have been so cheerful and proud for this achievement, and to my mother without whom I would not be here writing these lines. Big thanks to my precious brother and sister in law.

The final thank goes to Pietro, with whom I could share ideas, who let me approach this period in a peaceful way and supports me every day in a special way (“linvincciugadà).