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: Bio/11

Novel functions of Polycomb proteins and Ogt glycosyltransferase in chromatin regulation and transcriptional control

Andrea Scelfo

IEO, Milan

Matricola n. R09409

Supervisor: Dr. Diego Pasini

European Institute of Oncology, IEO, Milan

Anno accademico 2013-2014

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

5-carboxylcytosine (5caC)

5-formylcytosine (5fC)

5-hydroxymethylcytosyne (5hmC)

5-methyl cytosine (5mC)

Acute lymphocytic leukemia (ALL)

Acute myeloid leukemia (AML)

All-trans-retinoic acid (ATRA)

Alzheimer's disease (AD)

Bromo deoxy Uridine (BrdU)

Chromatin immunoprecipitation (ChIP)

CpG island (CpGi)

CREB binding protein (CBP)

Cytosine-guanine dinucleotides (CpG)

Days *post coitum* (dpc)

De novo methyltransferases (DNMTs)

Deacetylase enzymes (HDACs).

Dimethylated Lysine 27 of histone H3 (H3K27me2)

DNA immunoprecipitation (DIP)

Embryoid bodies (EBs)

Endoribonuclease-prepared siRNAs (esiRNAs)

Fluorescence Activated Cell Sorting (FACS)

Hematopoietic stem cells (HSC)

Hexosamine biosynthetic pathway (HBP)

Histone acetyltransferases (HATs)

Immunoprecipitation (IP)

Inactive X chromosome (Xi)

Induced pluripotent stem cells (iPSCs).

Ingenuity pathway analysis (IPA)

Isocytrate dehydrogenase (IDH)

Knockout (KO)

Leukemia Inhibitory Factor (LIF)

Long non coding RNA (lncRNA)

Low methylated regions (LMRs)

Lysine methyltransferases (HKMTs)

Mass spectrometry (MS)

Methyl binding domain (MBD)

Methyl-lysine analogue (MLA)

Monomethylated Lysine 27 of histone H3 (H3K27me1)

Mono-ubiquitination of the lysine 119 on the histone H2a (H2aK119Ubq)

Mouse embryonic fibroblast (MEF)

Mouse embryonic stem cells (mESC)

Myelodisplastic syndrome (MDS)

Myeloproliferative neoplasm (MPN)

O-(2-Acetamido-2-deoxy-D-glucopyranosylidenamino) N-phenylcarbamate

(PUGNAc)

O-GlcNAcase (OGA)

O-linked glycosyltransferase (Ogt)

Polycomb group protein (PcG)

Polycomb Repressive Complex 1 (PRC1)

Polycomb Repressive Complex 2 (PRC2)

Polycomb response Elements (PREs)

Post translational modification (PTMs)

Principal component analysis biplot (PCA)

Real time quantitative PCR (qRT-PCR)

RNA sequencing (RNAseq)

S-adenosylmethionine (SAM)

Short hairpin RNA (shRNA)

Ten-eleven translocation (TET)

Thymidine DNA glycosilase (TDG)

Transcription end site (TES)

Transcription factor (TF)

Transcription start site (TSS)

Trimethylated Lysine 27 of histone H3 (H3K27me3)

Unmethylated regions (UMRs)

Western blot (WB)

Wheat germ agglutinin (WGA)

Wild type (WT)

LIST OF FIGURES

ABSTRACT

During my PhD course, I have been involved in studies aiming to identify novel properties of proteins having fundamental roles in the regulation of chromatin modifications and gene transcription.

One of the two project was focused on the functional characterization of the products of the catalytic activity of Polycomb Repressive Complex 2 (PRC2). In spite of the wide knowledge about PRC2-dependent trimethylation of Lysine 27 of histone H3 (H3K27me3), the other forms of methylation on H3K27, namely mono (me1) and di-methylation (me2), are still poorly characterized. Using mouse embryonic stem cells (mESC) as model system, we were able to provide an extensive characterization of the functional properties of these methylation forms, highlighting their differential deposition along the genome, their fundamental role in the mechanisms of transcriptional regulation in mESC, and their potential implications during differentiation program. Our data demonstrated that while H3K27me1 was required for efficient transcription of genes and positively correlated with the deposition of H3K36me3, H3K27me2 was broadly deposited and protects the genome from aberrant firing of non specific cell type enhancers.

My second project focused on the activity of the O-linked glycosyltransferase Ogt which is the only enzyme capable to catalyze O-linked GlcNAcylation within the cell. Sxc protein, which is the *Drosophila* orthologue of mammalian Ogt, is essential for Polycomb (PcG) function. Sxc null embryos showed lethal phenotypes at early developmental stages like those observed in PcG null embryos. Starting from this observation we found appealing to investigate whether Sxc function in *Drosophila* were conserved in mammals. In particular, we were interested in the discovery of

possible role for Ogt in the context of PcG recruitment to chromatin and the mechanisms governing transcriptional regulation in mESC. With our study we have identified a novel partnership between Ogt and ten-eleven translocation (TET) proteins, which catalyze hydroxylation of methylated cytosine. We have shown that Tet1 protein recruits Ogt to target genes in proximity of transcription start sites (TSS) rich in cytosine-guanine dinucleotides (CpG). Tet1-Ogt co-localization at target genes correlated with low levels of cytosine modification, suggesting a role in the regulation of CpG island (CpGi) methylation.

Chapter 1 - INTRODUCTION

1.1. Chromatin properties

1.1.1. O*verview of chromatin characteristics and functions.*

Genomes of Eukaryotic require a high degree of compaction (from 10000 to 20000 fold[4]) in order to fit within the small volume of the nucleus. Indeed, genomic DNA is tightly associated to specialized protein allowing the formation of chromatin fibers which are characterized by a hierarchical folding within the cell nucleus. The basic unit of the chromatin fiber is the nucleosome. The seminal discovery of the nucleosome, made by means of X-rays diffraction on whole cell nuclei, date back to early seventies[5], fostering a great interest in the investigation about biochemical and functional properties of chromatin. In particular, the nucleosome core particle consists in 147 bp DNA left handedly coiled around a octameric complex of histone proteins. This protein core consists of four core histones: H2A, H2B, H3, and H4 in an arrangement where H3–H4 tetramer binds to two adjacent H2A–H2B dimers. Linker DNA regions connect these nucleoprotein particles forming a string of nucleosomes. Linker histones (histones of the H1 family) bind to these regions in close proximity to the sites of DNA entry and exit to the nucleosome protein core[6]. Linker histones have fundamental role in the three-dimensional organization of chromatin. DNA between two adjacent nucleosomes is named linker DNA and the average distance in nucleotides between the centers of adjacent nucleosomes over a genome of a given cell type or organism is variable and is known as the nucleosome repeat length[7]. Biochemical properties of histone proteins allow the formation of a very stable structure. Nucleosomes are positively charged, due to the abundance of aminoacids

with basic lateral chain capable to establish ionic bonds with negatively charged phosphate group of DNA[8].

All histones, except H4, exist in different variants thus providing further ways of regulation of chromatin functions along with histone post-translational modifications (PTMs) and DNA nucleotide modifications. Indeed, histone variants have been shown to influence *in vitro* nucleosome stability, their interaction with protein partners, the formation of high order chromatin structures; notably, the primary structure of histone variants are able to influence their PTMs[9]. The biochemical diversity among histone proteins mirrors the fact that they are not merely structural proteins but act as fundamental modulators of many cellular functions governed by DNA such as gene expression, replication, repair, and recombination. According to their function, the histone variants can be ascribed into two categories: replicative and replacement. The replicative histones are encoded by multiple gene copies, are organized in tandem and are devoid of introns; they retain a peak of expression in S phase and their incorporation into chromatin is coupled to DNA synthesis; in humans are represented by H3.1 and H3.2[10]. Replacement histone variants are often expressed in a tissue specific manner, do not have a peak of expression in S phase but are transcribed throughout the cell cycle. The best characterized replacement histone variants in humans are H3.3, encoded by H3.3A and H3.3B genes, and largely considered as marker of transcriptional activity[10]. Functional differences of histone variants have been widely demonstrated; for example, in mouse embryonic stem cells (mESC), mutation of the endogenous H3.3B gene to the canonical H3.2 sequence alters its genome-wide enrichment patterns, supporting the importance of the amino acid sequence of H3.3 in determining its final distribution[11]. Incorporation of histones onto chromatin is ensured by specific histone chaperones which can be either specific

or unspecific for the different variants; in addition, the same histone variants can be differentially deposited to specific genomic loci by diverse chaperons, like in the case of H3.3 which can be found at transient nucleosome free regions (mediated by HIRA complex) or at pericentric heterochromatin (through the action of DAXX/ATRX)[12].

Overall, nucleosomes are the main determinant of DNA accessibility. The characteristics of an individual nucleosome depend on the DNA sequence incorporated, and this can have functional significance for specific gene promoters[13]. Another feature influencing nucleosome stability and positioning is the combination of the PTMs which characterize incorporated histones[8]. Nucleosomes can be more or less compacted, thus the folding of chromatin can influence gene transcription. Indeed, two kinds of chromatin can be detected on the basis of their compaction status and transcriptional activity: euchromatin and heterochromatin. Euchromatin shows a more relaxed state thus being considered more accessible to transcription factors which favour transcription, while heterochromatin is more compacted and likely correlates with transcriptional repression. However, chromatin structure is very dynamic, and it can be speculated that, through the action of particular proteins, namely chromatin remodelers, its degree of folding can be readily changed (such as upon specific stimuli sensed by cell) in order to allow the necessary adaptations in terms of gene expression. Moreover, regions of constitutive heterochromatin also exist: it is the case, for example, of pericentromeric chromatin.

1.1.2. *Roles of the major histone and DNA modifications.*

Evidences that histone proteins are subject of post translational modifications at their N-terminal tails date back to early sixties when Allfrey and colleagues conducted seminal studies in the field[14]. Since this discovery, over the years, about 100 histone modifications have been identified and described at a rapid pace. Such modifications

are able to influence the chromatin structure, as demonstrated for the first time by solving the X-ray structure of the nucleosome[8]. Besides governing nucleosome interactions, histone PTMs were shown to be fundamental in recruiting specific proteins, such as remodeling enzymes to reposition nucleosomes throughout the genome, which is a way of regulation of gene expression. Among the most studied and well characterized histone PTMs there are acetylated and methylated lysines, phoshorilated threonine, tyrosine and serine; newly identified, but less abundant PTMs are crotonilation[15], lysine butyrylation and proprionylation, lysine 5-hydroxylation and serine/threonine O-GlcNAcylation[16]. The majority of PTMs occurs at the flexible N- and C-terminal "tail" domains protruding from the nucleosome core particle, but a significant fraction of modification takes place also in the globular domain of the histones, which regulates histone-histone and histone-DNA interactions[17]. An extensive list of known histone PTMs is shown in table 1.

In recent years, the develop and improvement of mass spectrometry (MS) techniques and chromatin immunoprecipitation (ChIP) assay, coupled to high throughput DNA sequencing technology (ChIP-seq) has provided a powerful tool to investigate the nature and the patterns of deposition of several histone PTMs in the genome of different organisms, and helped to unravel their functions through unbiased approaches. The several generated datasets allowed to associate particular histone modifications with gene activation or repression as well as with genomic features, including promoters, transcribed regions, enhancers and insulators. The combination of different histone PTMs mediates specific function on chromatin, for example through their chemical properties they can influence nucleosome structure. Moreover, particular histone PTMs combinations can drive the selective recruitment of effector proteins. The histone PTMs can function both sinergically and

antagonistically, having also different roles if asymmetrically present within the nucleosome[18].

Histone modification	Modified aminoacid	Modified residues
Phosphorilation	Serine, thyrosine, threonine	H1S171, H1S172, H1S17, H1S186, H1S188, H1S26, H1T10, H1T137, H1T145 H1T153, H1T154, H1T17,H1T30, H2AS137 H2AS139, H2AS1, H2AT120ph, H2AY142, H2BS14, H3S10, H3S28, H3S31, H3S6,H3T11, H3T3, H3T45, H3T6, H3Y41, H4S1
Methylation	Lysine, arginine	H1K186, H1K25, H2BK5, H3K27,H3K36, H3K4, H3K79, H3K9, H4K20; H2AR3, H3R17, H3R26, H3R2, H3R8, H4R3,
Acetylation	Lysine	H1K25, H2AK5, H2AK9, H2BK120, H2BK12, H2BK15, H2BK16, H2BK20, H2BK46, H2BK5, H3K14,H3K18, H3K23, H3K27, H3K36, H3K4, H3K56, H3K9, H4K12, H4K16, H4K5, H4K8, H4K91
O-linked glycosilation	Serine, threonine	H3S10, H4S37, H2AT101, H2B36, H2BS91, H2BS112, H2BS123
ADP-ribosylation	Arginine, glutamate	H1E2, H1R33, H2AK13, H2BK30, H3K27a H3K37, H4K16
Deimination (citrullination)	Arginine	H2AR3, H3R17, H3R26, H3R2, H3R8, H4R3
Isomerization	Proline	H3P30, H3P38
Ubiquitylation	Lysine	H2AK119 H2AK121, H2BK120, H4K91
SUMOylation	Lysine	H2AK126, H2B6,7
Biotinilation	Lysine	H2AK125, H2AK127, H2AK129, H2AK13,H 2AK9, H3K18, H3K9, H4K12, H4K8
Clipping	Lysine	H3K20

Table 1. Table summarizing histone PTMs.

During the work that will be presented in the result section of this thesis, the following histone PTMs were subject of investigation: the methylation of lysine 27 (H3K27me), lysine 36 (H3K36me), and lysine 4 (H3K4me) of histone H3, and the acetylation of lysine 27 of histone H3 (H3K27ac).

1.1.2.1. Histone acetylation.

Histone acetylation was the first modification described[19]. 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, acetylated histones were found to be enriched at actively transcribed regions. Interestingly, studies have demonstrated that is the charge neutralization, rather than the acetylation of specific lysines, to influence the transcriptional status of interested

genes[20]. Histone acetylation have been demonstrated to be also important during DNA replication and be indispensable for origin firing[21]. Lysine acetylation is the enzymatic product of histone acetyltransferases (HATs), which generally have low substrate specificity, and acetylate both cytoplasmic free and nucleosomal histones; they are present at sites of active transcription to facilitate polymerase transit by weakening DNA – histones interactions[22]. Acetylation can be removed by the activity of histone deacetylase enzymes (HDACs). These enzymes restore the positive charge of lysines, and are associated with transcriptional repression. Moreover, acetylated lysines can be recognized by bromodomains of chromatin remodeler proteins, thus influencing the chromatin structure[23]. Among the acetylated lysine residues of histone H3, the lysine 27 (H3K27ac) results particularly interesting. It was firstly identified in yeast[24], and in later studies in mouse and human specimens[25]. High throughput approaches revealed that H3K27ac was enriched at promoter regions of transcriptionally active genes[26], including Polycomb target sites. Indeed, H3K27ac, which is deposited by both p300 and CREB binding protein (CBP), have been shown to prevent trimethylation of H3K27 Polycomb target genes, with H3K27ac and H3K27me3 having dynamic and complementary temporal deposition profiles during embryogenesis[27]. H3K27ac has also important function in the enhancer element regulation. Indeed it discriminates active (H3K27ac positive) from poised (H3K27ac negative) enhancers on the basis of expression of proximal genes[28].

1.1.2.2. Histone methylation.

Methylation of histones occurs at side chain of arginine and lysine and, contrarily to acetylation, it does not lead to charge changes of histone protein. Methylations can be catalyzed in different manners, providing many combinations and related functional outcomes: indeed, arginines can be mono, symmetrically and asymmetrically

dimethylated[29], while lysines can be mono, di, or trimethylated. However, the effects on transcriptional regulation and nucleosome dynamics seem to be exerted almost completely by lysines methylation rather than methylated arginines[20]. Histone lysine methylations can be either associated to repression (like H3K9 and H3K27) and activation (such as H3K36 and H3K4) of transcription. The enzymes responsible for these PTMs are the histone lysine methyltransferases (HKMTs), and the first identified one was SUV39H1 which is specific for H3K9[30]. All methyltranferases, except DOT1L[31], endow a SET domain, which is responsible for the transfer of a methyl group from S-adenosylmethionine (SAM) to the ε-amino group of the lysine. Contrarily to HATs, HKMTs retain high substrate specificity and some of them are even specific for a given methylation state. In particular, it seems that an aromatic residue within the SET catalytic pocket domain (a tyrosine or a phenylalanine) is crucial to control the state of methylation[32]. Among the most relevant lysine methylations in the context of regulation of transcription are H3K4me1/me3 and H3K27me. The latter PTMs, which is the catalytic product of Polycomb Repressive Complex 2 (PRC2), will be extensively discussed in the next section of this thesis.

The first and unique H3K4 methyltransferase to be identified in *S. cerevisiae* was Set1[33]; in mammals about ten enzymes exist and can have SET domain related to yeast Set1 or to *Drosophila* Trithorax, overall classified as MLL family, which includes six members[34]. MLL complex has antirepressor functions and antagonizes the repressive activity of Polycomb complexes[35]; *in vitro* experiments showed that H3K4me3 deposition impairs the PRC2 complex to deposit H3K27me3[36]. Trimethylated H3K4 was found to be enriched at many promoters of Eukaryotic genes, in particular it marks the actively transcribed ones[37], thus it positively regulates transcription by recruiting nucleosome remodeling enzymes and histone

acetyltransferases[38 39]. Moreover, H3K4me3-marked promoters are also enriched for RNA polymerase II[40]. In mESC , interstingly, a fraction of promoters bearing high CpG content (about 20%) is both marked by H3K4me3 and H3K27me3, two histone marks with opposite functions. These promoters have been termed “bivalent” as they have characteristics of both active and repressive chromatin. A large proportion of these loci corresponds to developmental genes encoding transcription factors and other regulators of cellular state[41]. These genes are mostly repressed in pluripotent cells, are almost completely devoid of RNA polymerase binding[42] but can be rapidly induced or stably inactivated, depending on the developmental stages[41].

Monomethylated H3K4 was identified as marker of enhancer regulatory elements in human cells in association with other PTMs[43]. However, the mechanism by which H3K4me1 is established at these sites remains unknown. According to the general accepted view, the signature of enhancers’ PTMs might indicate general genome accessibility or chromatin dynamics at these sites.

Another histone PTM correlating with active gene transcription is H3K36me3. Enrichment of such modification occurs at gene bodies, and is a highly conserved phenomenon. At least eight distinct mammalian enzymes have been described to methylate H3K36, such as NSD1, NSD2, NSD3, MES-4, Setd2 and Ash1L[44]. H3K36me3 deposition by the KMT Setd2 is mediated by its interaction with elongating RNAPII[45-48]. This interaction is conserved in different organisms from yeast to human[49,50] and it is regulated by the phosphorylation of Serine 2 (pS2) in the RNAPII C-terminal Domain (CTD). pS2 is indicative of elongation and determines Setd2 recruitment and H3K36me3 accumulation at transcribed genes[51]. The role of H3K36me3 in the transcriptional elongation is still not fully understood. In yeast, H3K36me3 recruits the Rpd3 deacetylase complex to deacetylate newly incorporated histones[52-54].

Alternatively, H3K36me3 could regulate pre-mRNA splicing as H3K36me3 was shown to preferentially accumulate on exons of transcribed genes and to associate with spliceosome components[55,56]. Intron-less genes have significantly lower levels of H3K36me3 than the intron-containing genes where H3K36me3 deposition is proportional to transcriptional activity, in agreement with the reduction of H3K36me3 observed upon splicing inhibition.

1.1.2.3. DNA modifications.

In addition to DNA sequence, and in cooperation with histone PTMs, also DNA modifications drive and regulate the association and the downstream functions of factors that bind DNA. Eukaryotic DNA can be methylated at 5 position of cytosine (5mC) in the context of CG dinucleotides; DNA methylation is heritable and have been correlated with gene silencing: this can occur through direct impairment of interaction between DNA and transcription factors[57], and/or acting as docking sites for protein bearing methyl binding domain (MBD) such as chromatin remodelers and transcriptional corepressor complexes[58]. However, regions with high density of CpG di-nucleotides, called CpG islands (CpGi), are present in the genome and are generally maintained free of DNA methylation: about 60-70% of promoter regions contains a CpGi, whose methylation would result in stable transcriptional repression[59]; this is highlighted in some types of human cancer, where specific promoters of tumor suppressor genes get hyper-methylated at CpGi leading to their silencing. Cytosine methylation is catalyzed by a family of enzymes called *de novo* methyltransferases (DNMTs): specific DNMTs introduce cytosine methylation at previously unmethylated CpG sites, whereas other ones maintain such modification copying pre-existing methylation patterns onto the new DNA strand during DNA replication[60].

It was recently demonstrated that the ten-eleven translocation (TET) family of enzymes catalyze iterative 5mc oxidation to 5hmC[61], 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[62]. These two cytosine modifications are substrate for thymidine DNA glycosilase (TDG)-mediated base excision repair, thus suggesting an active role for TET proteins in the DNA demethylation process independently from replication process[63].

Establishment, maintenance and removal of DNA methylation have been shown to be influenced by the recognition of specific patterns of histone PTMs[58]

Overall, histone and DNA modifications function in a cooperative manner to rapidly change the chromatin organization and facilitate the recruitment of effector proteins throughout the genome.

1.2. Polycomb group proteins and their activity on chromatin

1.2.1. General overview of Polycomb group proteins

Polycomb group proteins (PcGs) were firstly described in *D. Melanogaster* as important players of fly development and tissue morphogenesis[64]. The homologs of these proteins were readily identified in mammals[65,66]. The first mammalian homolog identified was Bmi1 (Psc in fly), whose involvement as co-operative oncogene in Myc induced lymphomagenesis was soon unraveled[67,68]. Other mammalian homologs were promptly identified[69], and these studies led to the biochemical and functional characterization of these proteins present in all the mammalian cell nuclei as multi-protein complexes[70]. These complexes were named Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2). These two distinct complexes are formed by several Polycomb proteins with different and still not fully understood functions[71]. The PRC1 is the larger complex, recently shown to do not exist as unique

defined complex, but as several different PRC1 sub-complexes with different biological functions[72]. The two main Polycomb proteins that characterize the PRC1 complexes are Ring1b and Ring1a. These proteins are able to catalyse the mono-ubiquitination of the lysine 119 on the histone H2a (H2aK119Ubq)[73] that is totally dependent on PRC1 activity[74,75]. The PRC2 is a smaller complex and possesses the ability to methylate the lysine 27 on the histone H3. This PRC2 activity is exerted by the Ezh2 and Ezh1 polycomb proteins[3]. As the PRC1 is responsible for the entire H2aK119Ubq in the cells, the global methylation levels of the lysine 27 on the histone H3 is dependent on the PRC2 activity. The catalytic activity of both the complexes is also dependent on the presence of the other Polycomb proteins. Ring1a/b ubiquitin-ligase activity is strongly dependent *in vitro* and *in vivo* on the presence of Pcgf2 and Pcgf4[76,77] while Ezh2/1 methyltransferase activity is totally dependent on Suz12 and Eed[78-80]. Both PRC1 and PRC2 mainly localize at genomic loci enriched in genes involved in cellular differentiation and proliferation[81,82]. PcGs activity on chromatin is traditionally associated to the transcriptional repression[83]. Polycomb proteins regulate the catalytic activity of the two complexes, some others like Rybp, Kdm2b and Jarid2 regulate PRC1 and PRC2 stability and recruitment to the chromatin[84-87]. The complete understanding of the PRC1 and PRC2 recruitment to the chromatin in mammalian cells is still matter of debate. Emerging evidences show how biochemical links between PcGs and long non coding RNA transcripts can be at the basis of PcGs recruitment at specified target in the genome. PcGs functions were generally studied in two main directions: the role in cellular differentiation and development and the role in cellular proliferation and tumorigenesis[70,88]. Indeed, genetic depletion of different Polycomb proteins results in mouse embryonic lethality or in developmental defects, and PcGs overexpression is a negative prognostic factor in several different

tumors where PcGs inhibition is generally regarded as a potential strategy for tumour treatment[70].

1.2.2. Polycomb Repressive Complex 2: biochemical features and enzymatic activities on chromatin.

The Polycomb Repressive complex 2 (PRC2), through its enzymatic activity exerted on histones, is one of the most important regulator of gene expression within the cell. Its main activity is to tri-methylate the Lys27 on histone H3 (H3K27me3)[80,89,90], which is deposited at CpG dense genomic regions (mostly promoters), and it has been widely demonstrated to have a role in keeping gene transcriptional repression. The functional PRC2 complex, whose composition is conserved from *Drosophila* to mammals and across cell types, is characterized by four core protein components: Suppressor of zeste (Suz12), Embryonic ectoderm development (Eed), Retinoblastoma protein associated protein 46/48 (RbAp46/48), and Enhancer of Zeste homolog (Ezh1/2). The latter proteins retain SET domain responsible for Lysine methyil-transferase activity. In mammals, owing to duplication events, two different Ezh proteins are present (Ezh1/2) with cell type specific expression patterns, and different chromatin binding capabilities[91]. Each of the core protein is fundamental to achieve proper functional activity of the complex on the histone substrate both *in vivo* and *in vitro*[92,93], as demonstrated by early embryonic lethality of mice deficient for Eed, Suz12 or Ezh2[94-96]. This is consistent with the role of PRC2 in repressing genes involved in lineage specification[42,97-99]. Eed protein has the ability to bind H3K27me3 and establish a positive feedback loop enhancing lysine methyl transferase (KTM) activity of the complex[100]. KTM activity is boosted up to 7 fold in the presence of H3K27me3 peptide, and, consistently, PRC2 is more efficient in trimethylating K27 when in presence of pre existing H3K27me3: this suggests a

mechanism to maintain K27me3 levels through cell cycle progression[100,101]. The K27 residue on histone H3 can be post translationally modified by EZH1/2 in a stepwise manner from the mono-methylated to the tri-methylated form (H3K27me1/2/3), with likely different functional outcomes. *In vitro* KTM activity analyses of PRC2 demonstrated H3K27me0 and H3K27me1 to be better substrates than H3K27me2[102]. The different regulatory functions of PRC2-dependent H3K27me3 can be achieved thanks to the differential association of the core complex to accessory proteins (dispensable for intrinsic KTM activity) among cell types and across developmental stages; indeed, the core complex is unable by itself to target and mediate gene repression. Several additional proteins have been identified in the last years: 1) AEBP2[92], a zinc finger protein that enhances KTM activity and show a certain degree of colocalization at PcG genomic loci[103]; 2) three mammalian homologs of *Drosophila* Polycomblike proteins (Pcl1-3, also known as PHF1[104,105], MTF2[106,107], PHF19[108,109]) are tissue specific and interacts with Ezh2 and retain a tudor domain PHD capable to recognize H3K36me3 mark, suggesting a possible function in silencing active genes[110]. In particular MTF2 plays a role in the self renewal circuit in mESC, preventing the expression of self-renewal genes and controlling developmental regulators[106]. 3) Jumonji and ARID containing protein Jarid2 was found in association to Ezh2[86,111-114], it is able to directly bind GC-GA rich DNA elements, its loss impairs PRC2 recruitment at target genes, and it is necessary for proper mESC differentiation process[86]. The effects of Jarid2 loss on H3K27me3 are controversial: as observed in[86] [115] [116] H3K27me3 levels decreased upon Jarid2 depletion, while in[114] and[113] authors showed no change or enhanced H3K27me3 levels, respectively.

Pcl and Jarid2 proteins mutually bind to PRC2 core complex, thus these different accessory proteins can target PRC2 to specific genes in a context dependent manner[110].

Recent findings, based on mass spectrometry analyses on immunopurified complexes, revealed a physical interaction between PRC1 and PRC2 members in mESC and cancer cell lines, thus providing additional mechanisms of regulation of PcG activities. PRC1 proteins found to be associated to Eed almost belong to CBX containing PRC1 sub-complexes, in particular those with PCGF4 (BMI1) and PCGF2 (MEL18); association to RYBP and to the other PCGFs were found at very low levels. In this work authors also show EED to be the recruiter of PRC1-RYBP variants to PcG target loci in prostate cancer cell lines[117]. This phenomenon could be the explanation to the residual level of ubiquination observed at histone H2A in PRC2 null mESC[84], thus suggesting an enzymatic activity of RYBP/YAF2 PRC1 variants that is independent from Eed (PRC2). Core proteins and principal accessory proteins of PRC2 are depicted in figure 1.2.1.

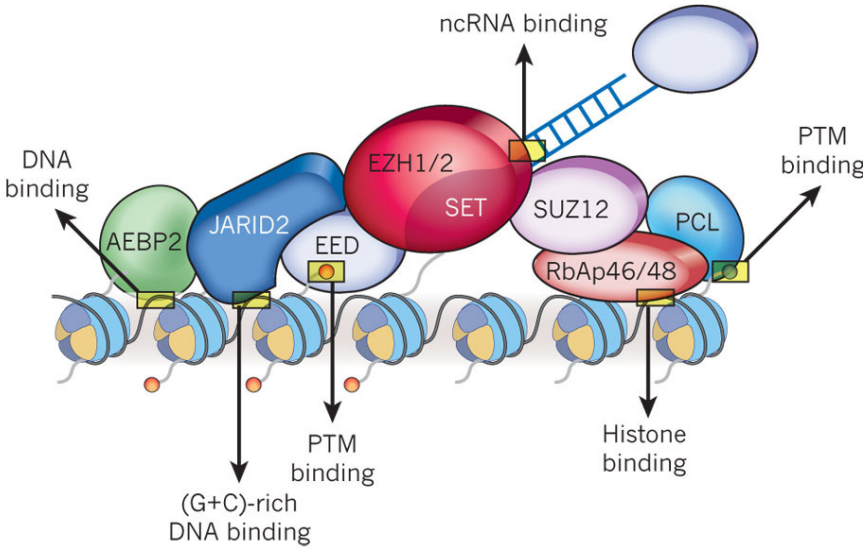


Figure 1.2.1. Principal members of PRC2 complex. The illustration, adapted from Margueron and Reinberg [3], depicts the core members of PRC2, along with accessory proteins and relative binding properties.

Altered expression of PRC2 proteins has been ascribed to be involved in tumoral malignancies such as leukemia, melanoma, lymphoma, breast and prostate cancer. Of particular interest is the role of EZH2, often found at high levels in primary tumors, ascribed as a *bona fide* oncogene[118]. Besides the fact that *EZH2* locus is amplified in tumors, it can bear specific mutation conferring an oncogenic potential. *EZH2* was found to be mutated in lymphomas of germinal center origin in a heterozygous manner. The described mutations (i.e. Y641, A687V, A677), initially found in a cohort of non-Hodgkin and diffuse-large B-cell lymphoma patient, affects the SET domain of the protein[119,120] leading to an enhanced catalytic activity on the substrate. This, in concert with wild type EZH2 protein, produces abnormal high level of H3K27me3 (and reduced H3K27me2) which in turn mediates aberrant repression of PcG target genes and the onset of lymphoma disease[121-125].

1.2.3. Polycomb Repressive Complex 1: biochemical features and general activities

Mammalian cells contain different PRC1 sub-complexes biochemically distinct one from each other[72,126]. All the PRC1 sub-complexes contain Ring1b or Ring1a core protein which determines the PRC1 catalytic activity[73]. The different sub-complexes are defined by the mutually exclusive presence of one of the six Pcgf subunits, while other components like Phc or Cbx proteins are present in different sub-complexes[72]. Phc and Cbx proteins are part of the canonical PRC1, originally identified in human cells[127]; canonical PRC1 contain Bmi1 (Pcgf4) or Mel18 (Pcgf2), named PRC1.4 and PRC1.2, respectively[72]. The canonical PRC1 is recruited on the chromatin through the ability of the Cbx proteins to recognize the H3K27me3 deposited by the PRC2 like showed by global loss of Ring1b binding at common (PRC2 and PRC1) target sites upon PRC2 depletion[84]. Although, the global H2AK119ub1 remains largely unaffected,

suggesting a PRC2 independent PRC1 activity on chromatin[84]. RYBP containing PRC1 is responsible for PRC1 activity in a PRC2 independent manner[84]. RYBP (and its homolog YAF2) is present in all the PRC1 sub-complexes, underlining its crucial role in PRC1 functions[72]. CBX proteins and RYBP presence in the PRC1 are mutually exclusive[72,84]. CBX and RYBP containing PRC1 share common target genes [72,128] even though they bind adjacent regions[72]. A recent work demonstrated that both PRC1.2 and PRC1.4 are unable to deposit H2Aub1 when forcedly recruited on chromatin, suggesting that pre-existing H3K27me3 could be essential for canonical complexes activity[129]. It remains although unclear why RYBP containing "sub-canonical complexes" are enzymatically silent in those conditions. Novel findings recently uncovered a role for Kdm2b in recruiting the PRC1 on chromatin[130]. Kdm2b belongs mainly to the PRC1.1 sub-complex[72] and it is a histone H3K36 demethylase[131]. Another PRC1 sub-complex that is likely to participate to mESC cell identity is the PRC1.6 (also named PRC1L4)[72,132]. The loss of some members of this sub-complex such as L3mbtl2 and Wdr5 lead to mESC premature differentiation[133,134]. PRC1.6 is characterized by subunits with a high degree of promiscuity in terms of complexes affinity. Wdr5 is a member of the COMPASS complex[135], Max is the notorious partner of Myc[136], E2f6 can be found with PRC2 members during proliferation[137], Hdac1/2 are partners of different repressive complex[138] and L3mbtl2 is also present in the NuRD complex[133]. Sub-complexes PRC1.3 and PRC1.5 are poorly characterized and the subunits role are not still determined. Forced recruitment of Pcgf3 and Pcgf5 on chromatin leads to H2Aub1 monoubiquitination and PRC2 recruitment[129]. Overall, the current knowledge about PRC1 sub-complexes roles and functions is still poor.

1.2.4. Mechanisms of Polycomb proteins recruitment to target loci

The investigation about mechanisms underlying Polycomb proteins targeting at genomic loci is one of the most debated but still undefined topic. Such mechanisms, leading to gene repression and establishment of correct transcriptional programs, have been well addressed and characterized in *Drosophila*, but they are not conserved in vertebrate systems in which different models of PRC1 and PRC2 recruitment take place. At the light of recent discoveries made by means of genetic, biochemical and genome-wide approaches, some alternative ways for PcG targeting to chromatin can be proposed; this is reasonable, given the high number of PcGs target genes, which in addition have to be specified in a cell type specific manner.

In *Drosophila* system, studies about PcG mediated Hox gene cluster repression led to the identification of Polycomb response Elements (PREs), genetic elements associated to promoters acting as *cis*–regulatory elements[139-141]. These PREs, which are almost devoid from nucleosomes, bear consensus sites for a number of DNA binding proteins which in turn are able to recruit PRC1 and PRC2 complexes[142,143]. The principal PREs binding factor is PHO, a specific transcription factor mediating direct recruitment of E(z) containing PRC2 complex, which in turn deposits H3K27me3 mark[144]. This mechanism, relying on PREs binding, is not conserved in mammals where genetic elements resembling the PREs are absent, and the mammalian PHO homologue, i.e. Ying Yang 1 (YY1), does not mediate PRC2 recruitment. This is demonstrated by the lack of correlation between genome wide distribution of YY1 and PcGs, and the lack of YY1 binding sites in sequences known to be targeted by PcGs[145]; accordingly, YY1 exerts PcG-independent function in mESC[146].

Among the proposed recruitment models, one is based on direct biochemical interactions between PcGs and transcription factors (TFs) which retain sequence

specific DNA binding elements for motifs present at promoters of target genes. The identification of such TFs, initiating PcGs nucleation on chromatin, is complicated by the fact that they could be only transiently bound to genomic loci. Transcription factor REST, interacting with Cbx protein, is able to target PRC1 at some genomic loci independently from H3K27me3 deposition and unmethylated CpG islands (CpGi)[147,148]. Bmi1 protein was biochemically purified along with Runx1/Cbfβ, which is able to recruit PRC1 in a PRC2 independent manner[149]. Evidences for PRC2 targeting *via* physical interaction with TFs were also reported. As shown for PRC1, REST can also interact with PRC2 complex[147]; other examples are Snail1[150], PML-RARα[151], and PLZF-RARα.[152].

In addition, recruitment *via* biochemical interactions occurs also through PcG association to long non coding RNA (lncRNA), as demonstrated by the phenomenon of X chromosome inactivation mediated by Xist non coding transcript[153].

A classical model elucidating PcGs recruitment relies on a hierarchical scenario in which PRC1 is targeted to chromatin in a PRC2 dependent manner. Genome-wide analyses of PRC2 (and its enzymatic product H3K27me3) and PRC1 correlate their occupancy with CpG islands (CpGi)[154], 1-2 Kb long genomic region characterized by a non methylated state surrounding gene promoters[155].

Experiments of genome engineering revealed the ability of CpG rich stretches, depleted of activating motifs, to ectopically recruit PRC2 in mESC[156]. Once targeted on its target loci, PRC2 catalyzes tri-methylation on H3K27, which is sampled and recognized by chromodomain of Cbx proteins, part of canonical PRC1 complex[157,158]. The final outcome is the recruitment of PRC1 and the consequent establishment of low amount of H2AK119 mono ubiquitilation driving chromatin compaction and gene repression. The hierarchical recruitment model is inadequate in proving the residual

level of Ring1b chromatin association and relative H2AK119ub1 present in PRC2 null mESC[84,159]. These evidences suggest that hierarchical recruitment is an over simplified model and additional PRC2-independent mechanisms underlying the targeting of PRC1 complexes to chromatin.

The non canonical PRC1.1 variant containing Pcgf1, Rybp, Bcor and demethylase Fbxl10/Kdm2b proteins exerts important function for the deposition of H2AK119ub1[72], and shows a direct recruitment to target chromatin. Fbxl10 retains a Zn finger CXXC domain that confers high affinity binding to CpGi, and the consequent PRC1 complex recruitment[85,160]. Fbxl10 is localized within almost all CpGi, associates to both active and bivalent genes (being also non PcGs targets), but that can become targeted during differentiation process[85].

The mechanisms underlying PRC2 localization at CpGi in absence of DNA methylation and active transcription can be provided by some models. One of them relies on the physical association between PRC2 protein and accessory regulatory subunits, such as Jarid2 and Aebp2, containing a Zn CXXC domain. Both proteins have affinity for CG elements, and show large colocalization with PRC2 occupancy[86,103,114]. Other PRC2 subunits that control PRC2 binding to chromatin are Pcls, interacting with core PRC2 proteins *via* PHD finger domain. In particular, more than half of PRC2 target sites are lost upon deletion of Pcl3[161]. Pcl3, besides having CpGi affinity, endows also a Tudor domain conferring binding property to H3K36me2/3, a histone mark know to inhibit PRC2 activity. This constitutes an additional mechanism that controls transcriptional repression and PRC2 selective binding on CpGi: indeed, PRC2 can be targeted to genes that have to be switched from an active state to a repressed one. Similarly, Trx complexes are able to bind and deposit H3K4me3 at CpGi *via* association with Zn

Finger-CXXC domain containing proteins[162], suggesting a competitive binding between Trx and PcG binding at target loci.

Novel evidences have reported a strict link between transcription and PRC2 recruitment. In mESC, ectopic recruitment of PRC2 (and novel deposition of H3K27me3) upon inhibition of transcription machinery was observed, and this occurs at CpGi of PRC2 targets annotated in differentiated cells. This model implies the existence of other factors allowing for the set up of gene silencing, and PRC2 displacement upon gene reactivation[163].

Besides PRC2 independent recruitment of PRC1 to chromatin, and opposite to the classic hierarchical model proposed for PcGs targeting, an alternative scenario in which PRC1 dependent H2AK119ub1 plays a key role in PRC2 recruitment was described by recent reports. Ectopic recruitment of different PRC1 variants at an artificial chromosome, randomly inserted in the mESC genome and lacking CpGi, were assayed. Analyses demonstrated that non canonical PRC1 variants containing Pcgf 1, 3 and 5, once ectopically targeted to chromatin, led to robust deposition of H2AK119ub1 followed by nucleation of PRC2 complex and H3K27me3 deposition; this phenomenon does not take place with PRC1 canonical variants. PRC2 recruitment was also dependent on the PRC1.1 variant associated to Kdm2b protein, thus linking recruitment of both PcG complexes to a subset of target genes[129]. This novel recruitment hierarchy was also reported by an independent study performed in mESC showing how, in absence of DNA methylation and in cooperation of other histone PTMs, both PRC1 and PRC2 are redirected to non canonical sites such as pericentric heterochromatin. Tethering Kdm2b containing PRC1 variant at pericentric heterochromatin led to H2AK119ub1 deposition and consequent PRC2 recruitment[164]. Another work performed in *Drosophila* embryos and mESC, has

recently reported that Jarid2-Aebp2 containing PRC2 complex binds H2AK119ub1 on nucleosomes, and has enhanced enzymatic activity. On the basis of these results, authors envisaged a positive feedback loop occurring at PcG target gene: boosted deposition of H3K27me3 is able to further recruit PRC1 through hierarchical CBX mediated mechanism[165]. The identification of factors able to link H2Aub and PRC2 will be very worthy to further elucidate the mechanisms governing PcGs targeting. Although, given that PRC1 null ESCs bear normal level of H3K27me3, the maintenance of this mark is independent from PRC1 activity[166].

Overall, all the recruitment models proposed are not mutually exclusive, but act in a cell type specific manner and in synergy to ensure proper target specification during differentiation.

1.2.5. Polycomb functions in stem cells and cellular differentiation.

Stem cells endow both self renewal and differentiation capabilities; the balance between these two conditions is regulated by means of integrated signaling and transcription cues. A single genome can indeed be modified and organized at different levels in order to give rise to cellular diversity; among the mechanisms underlying regulation of genes governing stem cell identity and cell fate determination, there are epigenetic modifications, with the orchestrated action of Tritorax and Polycomb proteins having a pivotal role. PcGs are able to establish heritable chromatin states that preserve gene silencing patterns in a cell type specific manner.

In vitro and *in vivo* evidences have shown that Polycomb proteins are able to mediate gene repression through different processes. Two main mechanisms of PcG-mediated gene repression have been proposed: 1) chromatin compaction; 2) impairment of transcription machinery functions. The ability of PRC1 in compacting nucleosome arrays was described in *Drosophila*, and this allows to repress the Posterior Sex

Combs genomic region[167]; functional domains of compacted chromatin were also observed as foci within nuclei, and were named PcG bodies[168]. Although these data from *Drosophila* claimed that PRC1 E3 ligase activity is needed for chromatin compaction, more recent studies done in mESC unveiled a dispensable role for PRC1-mediated H2Aub in chromatin condensation at *Hox* gene cluster; nevertheless, H2Aub was indispensable for proper repression of these target genes[169]. In general, a more compacted state of the chromatin, mediated by PcGs, is less accessible to chromatin remodelers (e.g. SWI/SNF complex) and transcription factors which eventually would lead to transcriptional activation[170]. Moreover, densely packed nucleosomes have been shown to stimulate PRC2 activity on H3K27, thus generating a positive feedback loop[171].

Another mechanism of PcG repression implies the inhibition of transcription machinery functions. Genome wide data in mESC displayed a drastically reduced levels of RNA Pol II occupancy at bivalent promoters of PcG target genes[172]. RNA Pol II present at promoters characterized by PRCs mediated repression markers is in the paused form (phosphorilated at Ser 5 of its C-terminal domain), and deletion of E3 ligases Ring1a and Ring1b causes the switch to its elongating form (phosphorilated at Ser 2)[173,174]. This suggests that PcG occupancy at bivalent promoters is able to hold and stall RNA Pol II at their transcription start sites.

In mESC, derived from tissue culture adaptation of cells from the inner cell mass of blastocyst, self-renewal is achieved through maintained expression of pluripotency genes (namely *Oct4, Sox2* and *Nanog*) and repression of lineage specific genes[175]. PcG proteins are expressed at high levels in mESC and are enriched at CpG rich TSSs of key developmental regulators such as *Hox* gene clusters, *Sox, Tbx* and *Wnt* gene families[97,175,176]. Normal levels of pluripotency genes are expressed in PRC2 depleted

cells, suggesting its dispensable role in self renewal process[177,178]. Differently, an essential role for PRC1 in maintaining self renewal property has been demonstrated, but the scenario is complicated by the different sub complexes formed by Ring1 proteins with several accessory factors. *Ring1b* KO mESC are able to self renew and to be maintained in *vitro* without striking morphology defects, although displaying reduced H2AK119ub levels[179], while co-depletion of both Ring1a and Ring1b proteins leads to cell cycle abnormalities and loss of self renew capacity after some *in vitro* culture passages[75]. The role of Cbx7-associated PRC1 complex in maintaining pluripotency is controversial, while its over-expression enhances self renewal of mESC[180,181]. Upon induction of differentiation by *in vitro* stimuli (such as retinoic acid), both PRC2 and PRC1 KO mESC fail to activate proper lineage commitment program owing to lack of de repression of lineage specific genes. *Eed* KO cells are able to form chimaeric embryos which die at early stages of development due to gastrulation failure[177]; *Suz12* KO and *Ezh2* KO mice are lethal at early post implantation stage (days *post coitum* E7.5-E8.5) being impaired to give rise to endodermal layer[178] and mesoendoderm lineage, respectively[182]. In addition, also depletion of PRC2 regulatory subunits, e.g. Jarid2 and Pcl2, have effects on the establishment of proper differentiation programs. Jarid2 depleted cells have altered H3K27me3 deposition pattern and it is required for correct differentiation: Jarid2 was shown to regulate recruitment of PRC2 at a large fraction of target genes[86] and to control recruitment of RNA Pol II at bivalent genes[115]. Pcl2 KO mice are viable, although being characterized by growth defects[183], and Pcl2 depleted mESC fail to undergo differentiation due to the maintenance of high level of pluripotency factors Nanog and Oct4[106].

Similarly to the *in vivo* phenotypes observed upon PRC2 KO, the depletion of *Ring1b* leads to severe gastrulation defects and embryonic death at early gastrulation stages (dpc 10.5)[184], while *Ring1b/a* double KO zygotes are not able to proceed beyond the first division[185]. mESC lacking Ring1b are not able to properly differentiate and show de repression of target genes[179]. Alternative PRC1 sub complexes containing RYBP and L3MBTL2 proteins have important functions in early developmental stages, as demonstrated by embryonic lethality due to aberrant gastrulation process[133,186]; knockdown in mESC does not allow correct embryoid bodies generation, mirroring altered cell proliferation and H2AK119ub levels[72,133]. The depletion of other PRC1 proteins causes developmental defects and mice are viable. Among the Pc proteins, the worst phenotype is observed in Cbx2 KO mice among which male individuals have female gonads, and female have no ovaries[187]. Similarly, a male karyotype was reported in a female individual carrying mutations at *CBX2* gene[188]. While Cbx7 is the main Cbx associated to PRC1 in mESC, a shift in expression and association towards Cbx2/4/8 occurs during differentiation stages; indeed, Cbx2/4/8-PRC1 exert repression of pluripotency factors favoring differentiation[180]. Deletion of Pcgf protein Mel18 (Pcgf2) or Bmi1 (Pcgf4) alone does not lead to embryonic lethality, but mice show homeotic transformation of axial skeleton and immune deficiency[189,190]; co-deletion of both Pcgf proteins causes embryonic lethality at dpc 9.5, suggesting compensatory functions exerted by these two homologs[191].

In addition, depletion of Ring1b protein in a *Eed* null background impairs differentiation capability of targeted mESC, suggesting independent functions for PRC1, rather than merely being downstream to PRC2 function[159].

Besides having key roles in differentiation processes in mESC, PcG proteins retain fundamental functions also in multipotent cell lineages and adult stem cells. Early B

cell development strictly depends on Ezh2 functions, shown to be required for IgH gene rearrangement controlled by proper pattern of H3K27 methylation deposition[192]; similarly, mutant *Eed* null hematopoietic stem cells (HSC) are not able to give rise to mature blood cells, and *Eed* deletion leads to exhaustion of adult HSCs pools[193]. Moreover, also PRC1 member Bmi1 is needed for HSCs maintenance as it represses the *Ink4a/Arf* locus[194]. Similarly to mESC, also in HSCs there is a tight control of Cbx proteins which associate to Ring1b to form PRC1 sub complexes. Cbx7 is present at high levels in HSCs and over expression causes leukemia; Cbx protein levels decrease through differentiation in favor of Cbx2/4/8 which, if aberrantly expressed in HSCs lead to their exhaustion [195]. Adipocytes formation from precursors is impaired in absence of functional PCR2, through an altered mechanism implying the lack of Wnt pathway suppression [196]. PRC2 is also required for proper myogenesis: premature transcription of specific muscle genes is prevented by a PcG complex, and Ezh2 expression gets progressively reduced upon external differentiation stimuli, thus allowing activation of myogenesis genes [197]. Epidermis formation from the basal layer of multipotent progenitors requires PRC2 mediated repression of AP1, acting as transcriptional activator of genes driving epidermal differentiation. The action of Ezh2 on AP1 prevents its premature recruitment to specific epidermal genes, thus ensuring proper time-controlled activation of lineage specific genes [198].

The role of PcG proteins is also important in reprogramming of terminal differentiated cells into the so called induced pluripotent stem cells (iPSCs). This process implies profound re organization of chromatin state exterted by different factors. By means of different approaches, both PRC1 and PRC2 complexes were shown to be essential for reprogramming of B cells [199] and human fibroblast [200].

Induction of pluripotent state in mouse embryonic fibroblast (MEF) does not require enzymatic activity of Ezh2, probably due to compensatory effects mediated by Ezh1, and co-deletion of Eed protein further lowers H3K27me3 levels at developmental genes and impairs reprogramming[201].

1.3. Roles of OGT glycosyltransferase in transcriptional regulation.

1.3.1. Characteristics of OGT and its counterpart OGA: roles in the cellular metabolism pathway and diseases.

Increasing evidences showed how the "epigenome" is able to integrate nutrient information: indeed, nutrient intake and metabolic disorders influence cancer onset through modification of the epigenome[202]. One of the most important element connecting nutrient levels and epigenetic is the OGT glycosyltransferase which can also act as chromatin modifier[202].

O-GlcNAcylation is a reversible post translational modification of cytosolic, nuclear and mitochondrial proteins consisting in the covalent attachment of N-acetylglucosamine (GlcNAc) to serines and threonines. This modification is exerted by an unique enzyme, highly evolutionary conserved, OGT (O-GlcNAc transferase) which uses the UDP-GlcNAc as donor substrate[203-205]. This modification is reverted by OGA (O-GlcNAcase) the only enzyme able to hydrolyzes the residue[206]. Cellular O-GlcNAcylation of proteins is strictly dependent on the levels of UDP-GlcNAc, which is the most abundant nucleotide within the organism after the ATP[2]. This nucleotide-sugar is the product of the hexosamine biosynthetic pathway (HBP) which originates from metabolism of different molecules such as glucose, glutamine, fatty acids and ATP (figure 1.3.1). For this reason, O-GlcNAcylation and UDP-GlcNAc are considered

the sensors of the metabolic state of the organism[207]. Notably, the serine and threonine residues that are targets of O-GlcNAcylation are also often phosphorylated by kinases, indicating a competition between these activities.

OGT expression and activity have been widely linked to cancer etiology and metabolic disorders. For example, OGT activity influences invasiveness of cancer cells and their metabolic reprogramming; in addition, several proteins coded by oncogenes and oncosuppressor genes have modified activities after OGT mediated O-GlcNAcylation (like shown for p53, β catenin, NF-κB)[208]. Besides being involved in cancer onset, alteration of OGT activity leads to metabolic pathologies, such as diabetes. The insulin signaling pathway is also regulated by OGT, highlighted by the fact that OGT overexpression in muscle and fat tissues causes mice to develop type II-like diabetes[209]. This is further supported by chromosomal deletion of the OGA locus (*MGEA5*) associated with type II diabetes, which results in increased level of O-GlcNAcylated proteins[210].

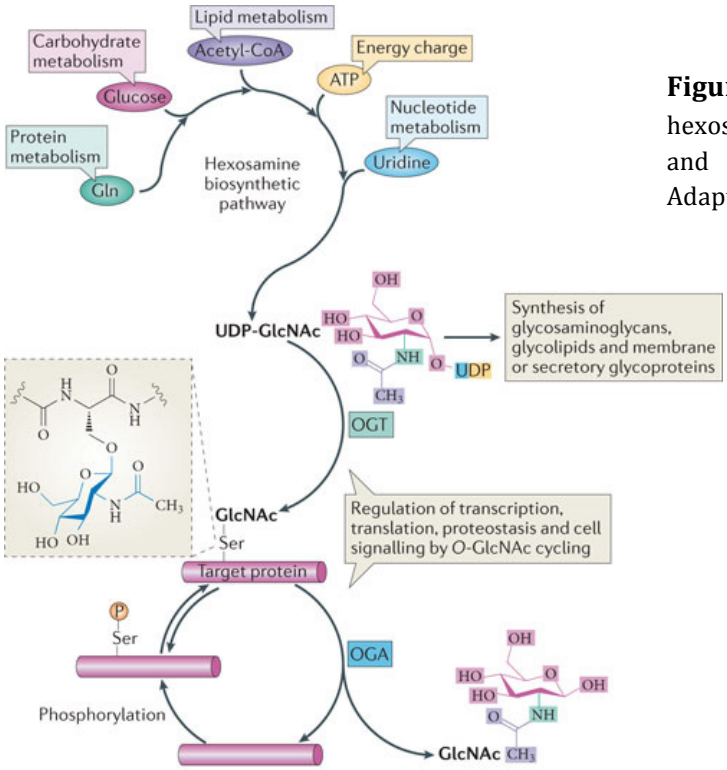


Figure 1.3.1. Schematization of the hexosamine biosynthetic pathway (HBP) and protein O-GlcNAcylation process. Adapted from Hanover el al. [2]

1.3.2. OGT roles as regulator of transcription and chromatin dynamics.

OGT and OGA can regulate gene expression both directly and indirectly. OGT, glycosylating target proteins localized in all cellular compartment, exerts pleiotropic functions and influences different cellular processes. OGT modifies enzymes involved in metabolic pathways, like members of glycolysis and insulin signaling, protein chaperones and components of the proteasome machinery, being indispensable for protein folding and stability[210]. Within the nucleus, OGT exerts its enzymatic activity on proteins that control transcription, thus it could directly link the metabolic state of the cell to transcriptional regulation.

Among the proteins that are regulated by O-GlcNAcylation there is HCF1, an epigenetic transcriptional regulator which undergoes proteolitic cleavage: the resulting protein fragments control cell cycle progression[211]. The HCF1 cleavage process was shown to be dependent on OGT glycosylation through the different phases of the cell cycle[212]. The N-terminal fragment of HCF1 (HCF1N) is strictly involved in transcriptional regulation; indeed it can associate with both transcriptional activators (like SET1–ASH2 histone methyltransferase complex) and transcriptional repressors (such as the switch-independent 3A (SIN3A) histone deacetylase complex); in addition it can be complexed to OGT itself. This suggests that OGT activity regulates not only the cleavage of HCF1 but has also effects in the regulation of HCF1 binding partners[2]. Moreover, OGT can interact with the transcriptional co-repressor SIN3A independently of HCF1 in a complex with other proteins, among which are HDACs[213]. The presence of OGT in the SIN3A complex along with HDACs implies that O-GlcNAcylation has a role in transcriptional repression. In addition, OGT physically interacts with the co-activator-associated Arg

methyltransferase1 (CARM1) and MMSET, which are important executors of epigenetic regulation[2].

Important OGT functions have been elucidated in the context of pluripotency maintenance. Indeed, OGT physically interacts with and O-GlcNAcylates components of core pluripotency factors of ES cells: Nanog, SOX2, OCT4 and cMYC[214]. Notably, the Oga promoter is co-occupied by SOX2, OCT4 and NANOG, allowing Oga gene expression in mESC: this suggests that the balance of O-GlcNAcylation is important for mESC physiology[215]. This is consistent with the suppression of mESC differentiation driven by excessive O-GlcNAcylation[216] and the decreased levels during differentiation process. Moreover, loss of OGT functions impairs self-renewal of mESC and reprogramming capabilities: Oct4 activity requires its O-glycosylation and the mutation of its glycosylation site (threonine 228) impairs Oct4 transcriptional properties[214].

Recently, O‑GlcNAc modifications of RNA Pol II has been identified. In particular, an O-GlcNAcylation site has been mapped to Thr4 of C-terminal tail of subunit 7. *In vitro,* the phosphorylation at serine 2 and 5 inhibits CTD O-GlcNAcylation and *vice versa*[217]. However, it remains to be understood if this mechanism regulates *in vivo* RNA Pol II activity.

Recent findings suggest that all four core histones are modified by O-GlcNAc, firstly reported in HeLa cells. Reported histone O-GlcNAcylation of histone H4, H2A and H2B occur at globular domains at known phosphorilation sites; some of them are in close contact with nucleosomal DNA and have previously been implicated in the dimerization of H2A and H2B and the association of these dimers with H3–H4 tetramers[218]. The histone H3 serine 10 (H3S10) glycosylation has been proposed as mechanism preventing H3 phosphorylation and impairing both cell cycle progression

(serine 10 is target of Aurora B kinase) and the transcriptional effects exerted by serine 10 phosphorylation which is linked to lysine 9 trimethylation, marker of repressed heterochromatin. O-GlcNAcylation of H3S10 was also associated with active chromatin (H3K4me3)[219].

Early studies conducted in *D. melanogaster* identified a new homeotic gene, named *Sxc* which, at the light of its feature, was characterized as belonging to polycomb group proteins[220]; years later, this gene was shown to be coding for Ogt[221], suggesting that O-GlcNAcylation is an essential function of the PcG machinery. The same study identified a number of O-GlcNAcylated proteins localizing at Polycomb Repressive Elements (PRE) and reported the polyhometotic (PH) protein to be O-GlcNAcylated. Recently, the direct interaction between EZH2 and OGT has been demonstrated and this regulates the stability of this PRC2 core member; in addition EZH2 was shown to be a direct of OGT enzymatic activity stabilizing the levels of H3K27me3[222]. However, the functional consequences of this PcG-OGT partnership need to be further elucidated.

Owing to its pleiotropic function, OGT depletion in mammals is not compatible with life, as demonstrated by embryonic lethality at day 5 *post coitum* observed in Ogt knockout mice[223]. Moreover, the maternal loss of Ogt, whose gene resides in the X chromosome, also leads the heterozygous embryos to death as consequence of developmental defects involving the extra-embryonic tissues in which the paternal X chromosome has been inactivated[224].

1.4. TET proteins and 5-hydroxymethylcytosine.

1.4.1. Characteristics of TET proteins and enzymatic activity on DNA.

The identification of enzymatic activity of TET protein family is quite recent (2009)[61,225]. These proteins owe their name to the ten to eleven translocation

(t(10;11)(q22;q23)) which characterizes very rare forms of acute myeloid (AML) and lymphocytic leukemia (ALL) and leads to the fusion of *MLL1* gene, located in chromosome 10, with *TET1* gene on chromosome 11[226,227]. The family comprehends three members: TET1, TET2 and TET3, often expressed in a tissue specific manner. They are Fe^{2+} - 2-oxoglutarate dependent dioxygenases having 5-methylcytoine as substrate which is iteratively oxidized into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[62] (figure 1.4.1). TET proteins exert their function in many biological processes like regulation of gene transcription, embryonic development, cancer onset and stemness maintenance.

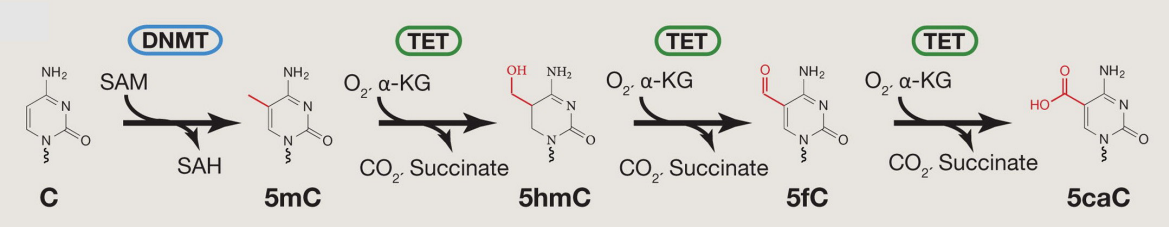


Figure 1.4.1. Schematization of the cytosine methylation process an its iterative oxydation. Adapted from Wu el al. [1]

Concerning the structure of such enzymes, having different aminoacidic sequences, metazoans TETs contain a N-terminal CXXC domain, able to bind DNA, and a C-terminal catalytic domain, the core of which is Cys rich and folds into a double stranded β helix[225]. TET2 protein is an exception as a chromosomal inversion led to the loss of the sequence coding the CXXC domain which became an independent protein called IDAX, encoded by a separate gene[225]. CXXC domains of TET1 TET3 showed a preferential *in vitro* binding for unmethylated CpG stretches[228], and *in vitro* TET1 associates to CpGi and CpG rich promoter regions[229]. Curiously, CXXC domains of TET3 and IDAX exert regulatory functions on their own expression: IDAX negatively regulates TET2 expression like demonstrated by the lack of physiological

TET2 downreguation during differentiation process of mESC upon IDAX depletion[230]. Moreover, TET3 has an auto-inhibitory mechanism that has been postulated to be dependent on interaction between CXXC and catalytic domain of TET3[231].

5hmC levels vary in a cell type manner, with the highest one found in the brain (reaching about 25% of total 5mC)[62]; in mESC, 5hmC is about 5-10% respect to the total level of 5mC[61]. 5fC and 5caC forms were found at very low level respect to 5mC in mESC (0.03% and 0.01%, respectively)[62]. This is reasonable, given the fact that they are intermediate enzymatic products leading to their removal through base excision mediated by thymidine DNA glycosilase (TDG)[232].

1.4.2. Oxidized forms of 5mC: genomic distributions and effects on gene expression.

Since their identification, many efforts aiming to unravel biological functions of Tet proteins have been made. Nowadays, the main biological role envisaged for TETs is DNA demethylation, but this is very challenging to univocally demonstrate due to the differences among the results achieved by the different approaches adopted in the studies. However, the understanding of TETs functions relies on the identification of their genome-wide binding sites and deposition patterns of oxidized forms of 5mC by means of unbiased approaches.

Two main categories of techniques allowing the genome-wide mapping of modified cytosine have been developed: affinity methods (such as DNA immunoprecipitation (DIP), and base resolution mapping methods such as bisulphate sequencing (BS-seq). The first method integrates signals from multiple cytosines present at a certain locus, and detects also cytosines with low levels of modification; on the other hand, the second method is more sensitive for detecting single modified cytosine, but requires very high sequencing depth. The usage of these two approaches in independent

studies, along with the different analysis strategies adopted, led to discrepancies among published results[1].

Oxidized cytosines can be deposited both at promoter, gene bodies and enhancers sites. About 70% of mammalian promoters are characterized by CpGi which are devoid of methylation, are H3K4me3 positive, and at these sites, the histone variant H2A.Z prevents binding of DNMTs[233]. This explains the lack of 5hmC at these promoters[234]. However, TETs are found to be enriched at CpG rich active promoters, suggesting functions in favoring transcription which are independent from their enzymatic activity, for example the stabilization or the recruitment of chromatin modifiers proteins[1]. In addition, promoters of gene with low CpG density and expressed at medium-low level, show higher 5hmC enrichment[234]. This is consistent with the 5hmC enrichment at low methylated regions (LMRs)[235]. Moreover, TETs and cytosine oxidized forms are envisaged to play a role in the maintenance of lineage specific promoters in a inactive- poised state by regulation of demethylation[1].

5hmC deposited at gene bodies, with the highest accumulation at 3' portion, has been correlated to active transcription in all of the analyzed cell types; also 5fC and 5caC were reported to accumulate at gene bodies of transcribed genes[236]. However, the function of such enrichments is still unclear. One hypothesized explanation is that they could positively contribute to mRNA processing[1].

Enhancer regulatory elements are regions of low CpG density with reduced DNA methylation respect to other genome loci[235]; in all mammalian cell lines in which 5hmC profiling have been studied was reported strong accumulation of such modification at enhancer sites[237]. In differentiating mESC 5hmC levels rise as soon as they become activated; indeed, poised enhancers show high 5hmC, 5fC and 5caC levels, suggesting that they are pre-marked in favour of subsequent demethylation

during development[238]. However, 5hmC levels in enhancers drop down in correspondence of transcription factors binding sites[239], probably because transcription factors prevent DNMT and TETs from binding those loci[236].

Although TETs regulate the levels of DNA methylation at promoters, only about 10% of genes show altered transcription upon Tet1 depletion in mESC[229], with the number of upregulated genes being higher than the downregulated ones, indicating a role for Tet1 also in repression of transcription. This is supported by the occupancy of some of those genes by PRC2 proteins[229,240]. One possible scenario is that TET1, decreasing methylation levels at PRC2 target sites, facilitates PRC2 chromatin binding. In addition, TET1, preventing stable methylation of Polycomb target genes, mostly related to cell-fate commitment, could play important role in keeping these gene ready to be activated during development process[241]. Transcriptional repression exerted by TET1 is likely to occur through its association to SIN3A repressor complex, as demonstrated by the large overlap among their target genes, whose expression increased in mESC knockout for the three DNMT (DNMT TKO). This suggests catalytic-independent functions of TET1 in transcriptional regulation[229].

TET proteins are proposed to maintain DNA methylation fidelity: indeed, it has been proposed that high levels of H3K4me3 at CpGi prevents DNA methylation by inhibiting the recruitment of DNMTs. Sporadically, upon deregulation of H3K4 methylation or aberrant DNMT activity, DNA methylation of CpG islands could occur. Aberrant methylation would be oxidized by TET proteins, leading to passive or active demethylation, thus restoring unmethylated CpGi. Overall, this hypothesis is supported by the global decrease of 5hmC in favour of 5mC observed in different tumors, along with the frequent mutations of TET proteins found in hematopoietic malignancies and colorectal tumors[241].

1.4.3. General roles of Tet proteins in development and cancer.

In spite of the abundance of 5hmC in mESC, TET1 and TET2 have been demonstrated to do not exert essential roles in mouse development. Indeed, *Tet*1$^{-/-}$ and *Tet1*$^{-/-}$-*Tet2*$^{-/-}$ mice develop all three germ layers, reach birth and survive[242,243]. However, Tet1 deficient mice showed defects in neuronal development[244]. Tet2 KO mice are also viable and fertile, although spontaneously developed hematological tumors within few months, owing to an aberrant size of the hematopoietic stem/progenitor cell pool[245]. However, there is variability among the costituitive Tet1/Tet2 double KO mice which have an incompletely penetrant lethal phenotype: indeed, some mice die soon after birth due to various defects, such as in exencefaly development, growth retardation, hemorragy[243]. This leads to speculate that some mice acquired excessive methylation at promoters of early development genes leading to death, while other ones did not. However, these mice retain global and not-specific hypermethylation in several tissue[243]. In addition, Tet3-null embryos are neonatal lethal, whereas germ-line-specific deletion of Tet3 from primordial germ cells affects female fertility and compromises embryo development[246].

In general, DNA of tumoral cells are characterized by loss of 5hmC compared to normal counterparts. Although, if this phenomenon is a cause or a consequence of tumorigenesis is still not understood. The role of TET proteins in cancer is quite controversial, as different reports have documented both oncosuppressor and oncogenic roles[247]. As already mentioned, TET1 was found as MLL fusion partner in Acute Myeloid Leukemia (AML)[226,227], and the lack of TET1 activity led to increased methylation and silencing of target genes. In particular, TET1 is targeted and overexpressed by MLL fusion proteins[248]. TET2 loss of function mutations were identified in myeloproliferative neoplasm (MPN) and myelodisplastic syndrome

(MDS), where it was shown to be a critical oncosuppressor[249]. *Tet2*$^{-/-}$ hematopoietic stem cells (HSC) are not able to differentiate and have higher proliferative capacity respect to normal cells, thus leading to the enlargement of HSC pool which cause hepatomegaly, splenomegaly, monocytosis and extramedullary hematopoiesis. In these cells, the levels of 5hmC were globally decreased in favour of CpGi methylation[245]. This suggests that lineage determination in hematopoietic system is in part controlled by cytosine methylation and hydroxymethylation.

Decreased expression of TET1 and TET3 were reported in some solid tumors, above all colorectal cancer (CRCs): loss of 5hmC regarded about 70% of analyzed CRCs, and TET1 was observed to be downregulated *via* oncogene-dependent cellular transformation[250].

Altered TETs activities in cancer can also be the consequence of mutations affecting metabolic enzymes necessary to produce the cofactors needed for proper TETs enzymatic activity; indeed, loss of function mutations of isocytrate dehydrogenase proteins (IDHs) lead to the production of 2-hydroxyglutarate rather than 2-oxoglutarate, thus inhibiting dioxygenases activity of TET proteins[251]. It has been observed that in leukemia, IDH and TET2 mutation are mutually exclusive, and it has been demonstrated that that Tet2 oncogenic activity is independent from IDH mutations[252].

Overall, the strict connection between 5hmC and cancer development has still to be univocally demonstrated. An exception is Tet2 in the hematopoietic system, because the loss of TET2 *per se* is able to induce neoplasm.

Chapter 2 – RESULTS

2.1. Polycomb dependend H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity.

2.1.1. Polycomb Repressive Complex 2 differentially controls the three methylation states of Lys27 on histone H3.

One of the aims of my PhD work was to assess the distribution of the post-translational modifications (PTMs) of histone tails within nuclei of embryonic stem cells. Thus, a proteomic study performed together with our collaborators[253], was able to determine the relative abundance of PTMs of Lysine 27 (K27) in the Histone variants H3.3 and H3.2. From the analysis of these data, retrieved by a tandem Mass Spectrometry procedure, was evident that, overall, more than 80% of the total histone H3 accounts for the different methylated forms of H3K27, the unmodified H3K27 represents in average 16% of total H3, while H3K27ac is limited to approximately 2% of the total H3. Analyzing the data relative to the different methylated forms of H3K27, we observed that more than 70% of H3K27 is present in the dimethylated form (H3K27me2), while the trimethylation (H3K27me3) and the monomethylation (H3K27me1) account for 7% and 4% of total H3, respectively (Figure 2.1.1A). Looking at the distributions among the H3 isoforms, H3K27me3 shows a preferential accumulation in the H3.3 variant: this datum corroborates the major deposition of H3.3 found at the level of promoter regions of both silent and expressed genes[11]. Overall, these data show that in mESC the vast majority of histone H3K27 is post-translationally modified, with the H3K27me2 being the most abundant modification.

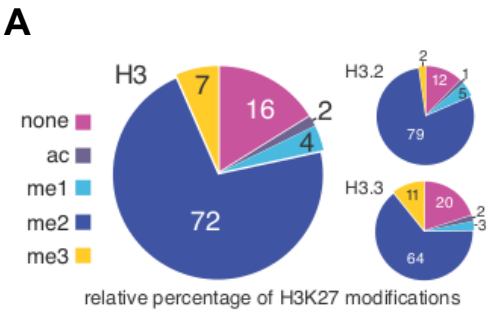
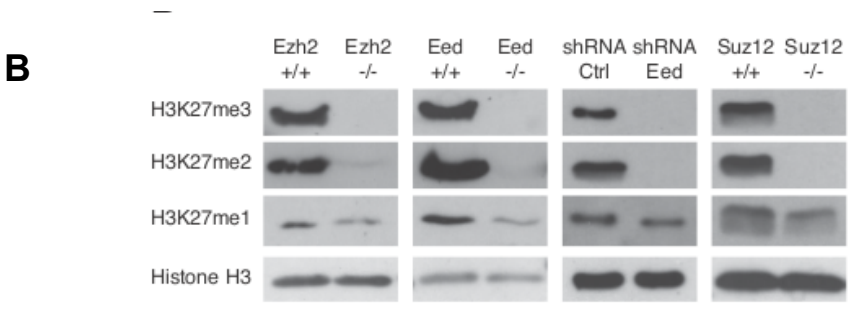


Fig.2.1.1A Relative percentage of H3K27 post-translational modifications in mESC. Data from (Jung et al., 2010) were determined by tandem Mass Spectrometry analyses. The larger pie graph depicts the average abundance calculated between Histone H3.2 and H3.3 isoforms. The smaller pies show the specific abundances detected on Histone H3.2 and H3.3 variants.

Reports present in the literature raised the issue if H3K27me1 was a PTM dependent on PRC2 catalytic activity, or the result of an active de-methylation process[3]. To address this issue, we performed Western blot (WB) analyses on histones extracted from mESC lysates, both wild type (WT) and knock out (KO) for PRC2 core members Ezh2, Eed and Suz12; as further control, we performed the same analyses in mESC lysates acutely knocked down for the protein Eed by means of lentiviral transduction of sequence specific short hairpin RNA (shRNA). The levels of H3K27me2 and H3K27me3 were quite completely abolished in absence of PRC2 core component, while H3K27me1 levels were significantly reduced (figure 2.1.1B, C).



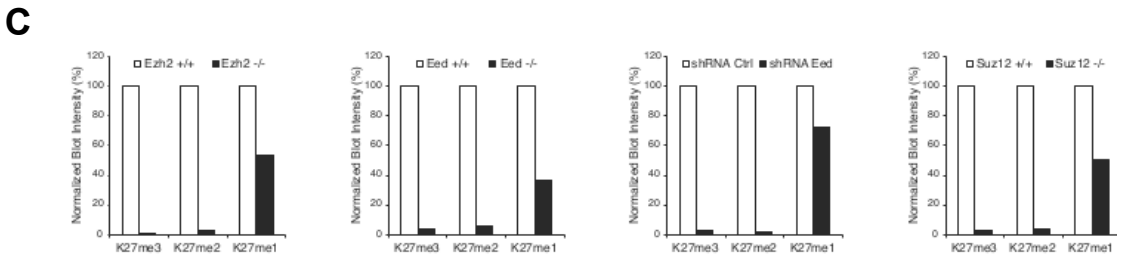


Fig. 2.1.1B, C. Levels of methylation on H3K27 decrease upon loss of PRC2 activity in mESCs. (B) Western blot analyses using the indicated antibodies of protein extracts obtained from WT (+/+) and *Eed*, *Ezh2* and *Suz12* KO (-/-) mESC lines or from E14 mESC stably transduced with *Eed* specific shRNAs. A scrambled shRNA sequence was used as negative control. Histone H3 served as loading control. (C) Quantification of western blot signals for H3K27 methylation forms showed in Figure 1B using ImageJ software. Western blot signal for total histone H3 was used to normalize blot intensities, and reductions in methylation signals upon PRC2 components depletion are expressed as percentage of the normalized blot intensities obtained in wild type (+/+) or control (Ctrl) mESC lines.

To fully demonstrate that H3K27me1 and H3K27me2 depositions were dependent on PRC2 activity, we wanted to exclude a possible enzymatic role of the novel identified H3K9 specific lysine methyltransferases (KMT) Glp and G9a on H3K27. These enzymes, *in vivo*, sinergically act as a complex[254-256]. At this purpose we depleted at the same time Glp and G9a proteins in mESC by using specific short interfering RNAs (siRNA). The loss of the two KTMs did not lead to global change in H3K27 methylation levels (figures 2.1.1D, 1E). Moreover, the expression level of both G9a and Glp, measured by RT-qPCR, was unaltered in *Eed* KO mESC cells (data not shown), thus excluding the involvement of these KTM proteins in the regulation of H3K27 methylation forms catalyzed by PRC2.

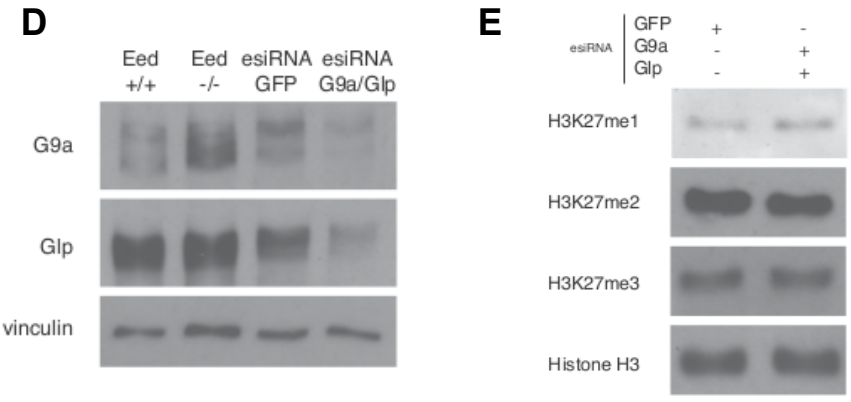


Fig 2.1.1 D, E. Knock down of G9a and Glp in mESC do not affect methylation states on H3K27. (D) WB analysis using antibodies anti-G9a and anti-Glp of protein extracts prepared from mESC *Eed* WT (Eed +/+), *Eed* KO (Eed -/-) and mouse E14 ES cells simultaneously interfered for *G9a* and *Glp* (G9a/Glp) or unrelated protein (GFP). Vinculin served as loading control.
(E) Western blot analysis using specific H3K27me1, H3K27me2, H3K27me3 antibodies of protein extracts prepared from mESC interfered for *G9a* and *Glp* or unrelated protein (GFP). Histone H3 served as loading control.

2.1.2. The three methylation states of Lys27 on histone H3 form mutually exclusive genomic domains.

To investigate the genome-wide distribution of the different forms of H3K27 methylationl we took advantage of chromatin immunoprecipitation (ChIP) coupled to high-throughput DNA sequencing (ChIP-seq) using specific antibodies against H3K27me1, H3K27me2, H3K27me3 and total H3. Since these antibodies must accurately discriminate single methylation moieties on the same lysine residue, we tested antibodies specificity on different sets of modified peptides. Antibodies resulted to be specific for each methylation states of K27 and did not cross-react with other methylated residues on the histone N-terminal tail (figure 2.1.2A, 2B).

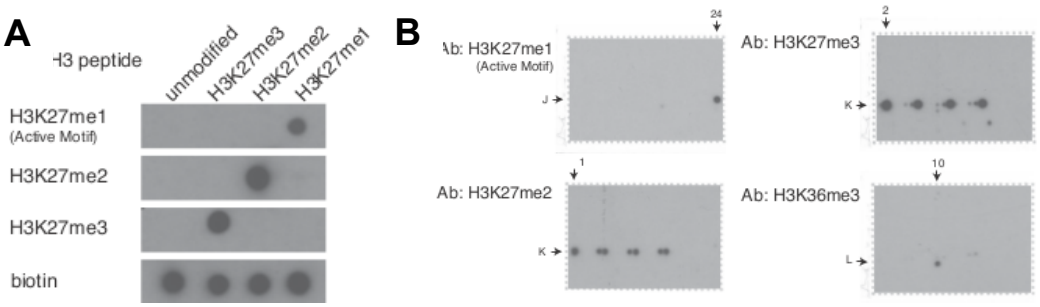


Fig. 2.1.2A,B Test of antibodies specificity. (A) Immuno-dot blot analysis showing specificity of the indicated antibodies tested against biotinylated histone H3 peptides bearing specific modification on Lys27. Biotin served as loading control. (B) Cross reactivity screening of the indicated antibodies using a peptide array bearing 384 unique histone modification combinations. Arrows indicate the coordinates on the array of the modified peptide matching the modification to which the antibody avidity is directed. The complete list of spotted peptides is availableonline at www.activemotif.com/documents/MODified_Histone_Peptide_Array_grid.xls.

The analyses of ChIP sequencing data performed in WT mESC showed that the different methylated forms of H3K27 are deposited in mutually exclusive genomic domains, as shown by the genomic snapshots retrieved by UCSC genome browser (figure 2.1.2C). More in detail, H3K27me2 spread across both intergenic and intragenic regions, while H3K27me1 was preferentially enriched within gene bodies (figure 2.1.2C). We compared this pattern of H3K27me1 deposition, with a published

ChIP-seq dataset of H3K36me3[257], obtaining a significant overlap between H3K27me1 and H3K36me3 deposition at the same intragenic sites (figure 2.1.2C).

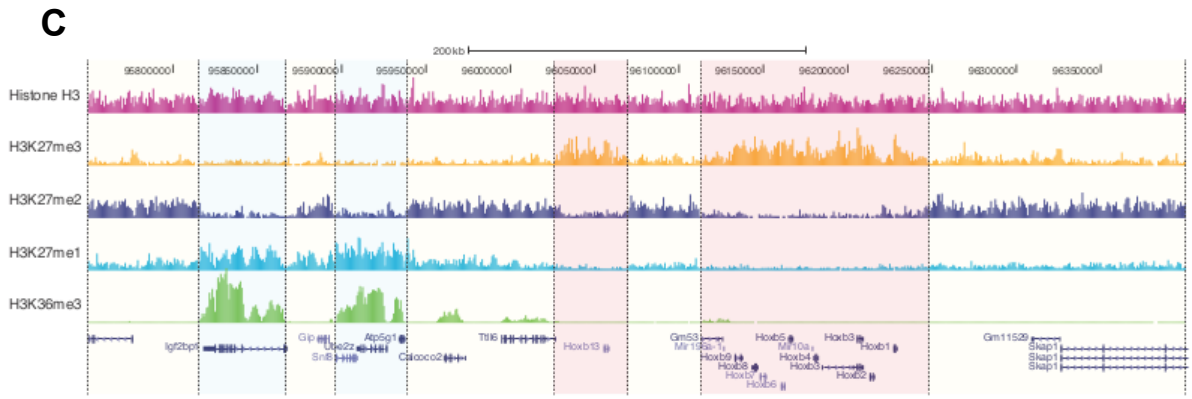
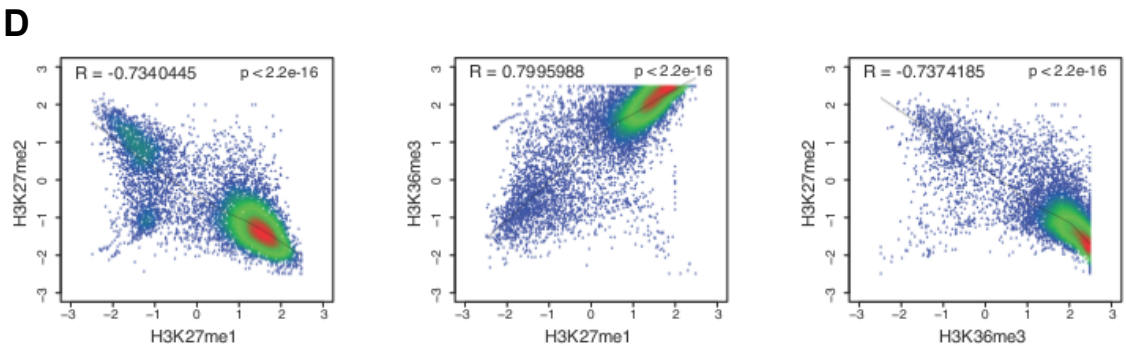


Fig. 2.1.2C Genomic snapshot of ChIPseq data analyses performed in mESC using the indicated antibodies. H3K27me1 enrichment domains are highlighted in blue, while H3K27me3 domains are depicted in red.

To prove these observations at a genome wide level, we performed correlative analyses between pair of ChIPseq datasets, and represented them as scatter plot graphs. These correlative analyses of the intragenic signals for H3K27me2, H3K27me1 and H3K36me3, relative to all RefSeq annotated genes, highlighted that, while H3K27me2 negatively correlated with H3K27me1 and H3K36me3 deposition, the accumulation of H3K27me1 positively correlated with H3K36me3 (figures 2.1.2D, E).



E

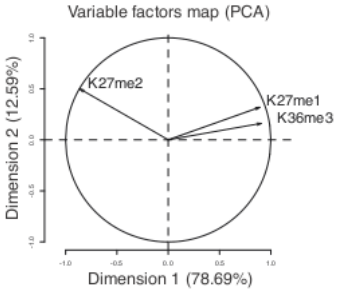
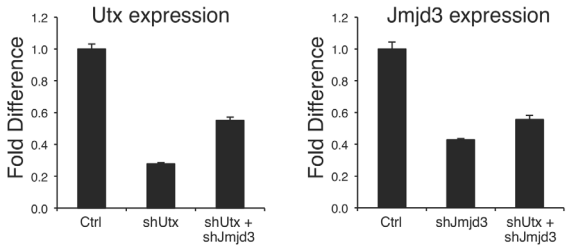


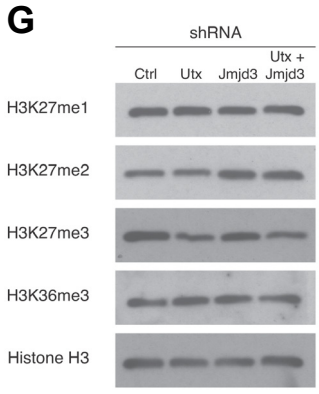
Fig. 2.1.2D, E Correlation plots between the indicated PTMs. (D) Scatter plots showing the correlation between the enrichments normalized to the histone H3 density of K27 and K36 PTMs among all RefSeq annotated genes. The R coefficient indicates Pearson correlation value. (E) Principal component analysis biplot (PCA) indicating the extent of correlation between PTMs in gene bodies of all RefSeq annotated genes.

We raised the question whether the differential deposition patterns were consequence of H3K27 de-methylation activity exerted by the demethylases Utx and Jmjd3 proteins. Thus, we knocked down *Utx* and *Jmjd3* by means of lentiviral transduction of specific shRNAs (figure 2.1.2F), or by chemical inhibition of demethylases activity on H3K27 through hypoxia induction by $CoCl_2$ (figure 2.1.2I). Both approaches had no effect on the global and site-specific H3K27 methylations levels, as shown by immunoblot (figure 2.1.2G) and ChIP-qPCR analyses at selected target loci (figure 2.1.2H, J).

F



G



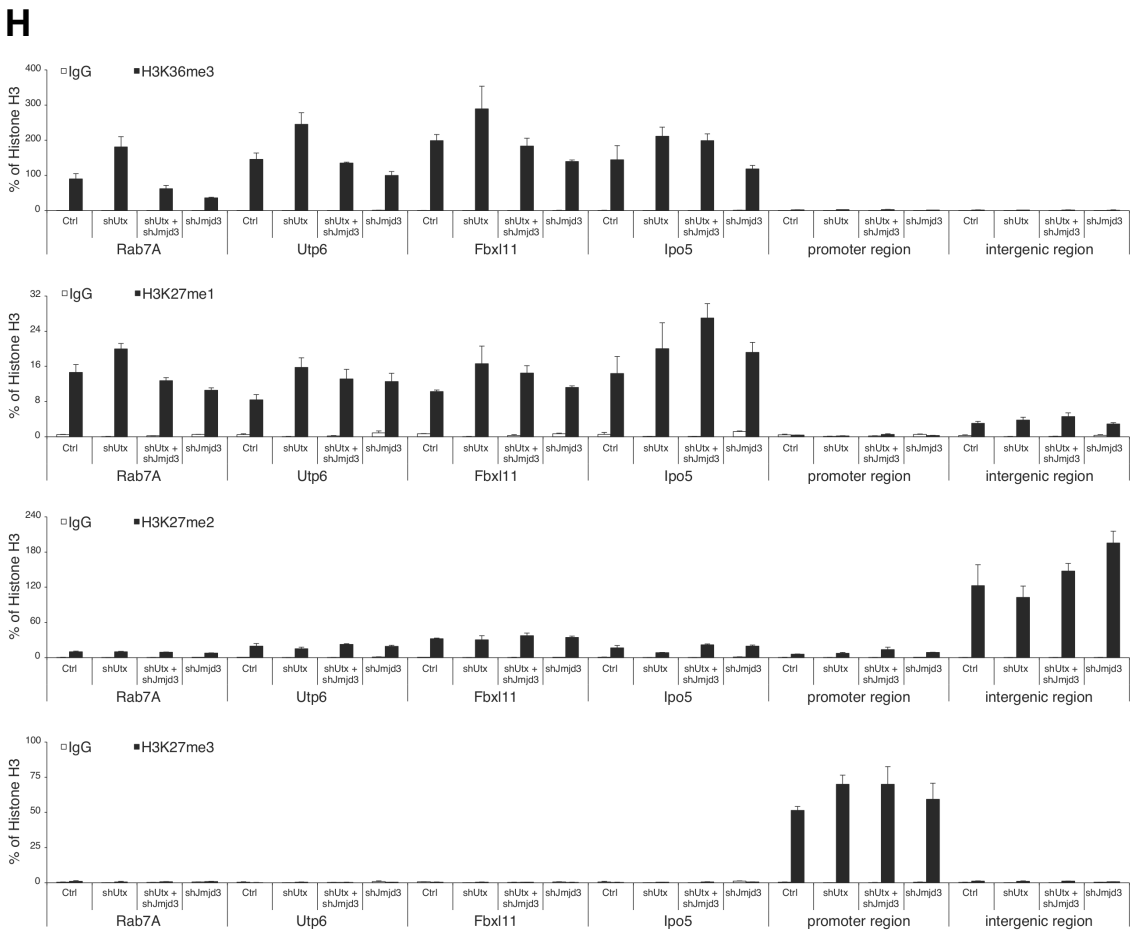


Fig. 2.1.2F-H Analyses of demethylases activity on H3K27 methylated forms. (F) Relative expression levels of *Utx* and *Jmjd3* determined by qRT-PCR analyses in cells expressing scrambled (Ctrl) or *Utx* and/or *Jmjd3* specific shRNAs (shUtx, shJmjd3 and shUtx + shJmjd3). *Gapdh* served as normalizing expression control. (G) WB analyses using the indicated antibodies of protein extracts obtained from cells described in (F). A scrambled shRNA sequence was used as negative control (Ctrl). Histone H3 served as loading control. (H) qRT-PCR of ChIP analyses in cells described in (F) using the indicated antibodies for showed regions: Rab7A, Utp6, Fbxl11, Ipo5 are intragenic H3K27me1 enriched regions; promoter region corresponds to the TSS of Wnt5a gene, a PRC2 specific target locus; intergenic region corresponds to a H3K27me2 enriched domain. ChIPs with rabbit IgG served as negative control. ChIP enrichments are normalized to histone H3 density.

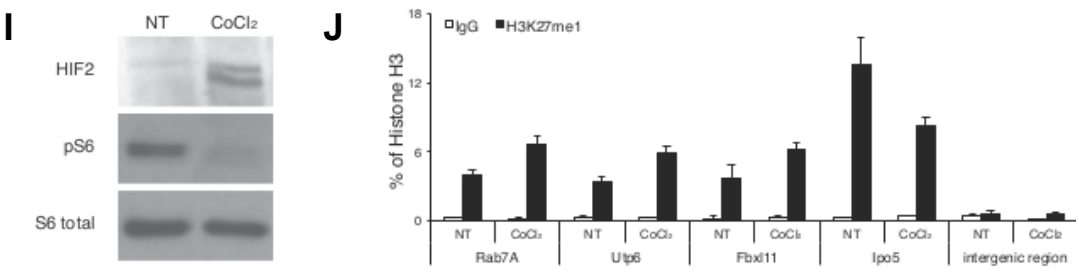


Fig. 2.1.2 I, J Effects of inhibition of demethylases activity on H3K27 methylated forms. (I) WB analyses using anti-Hif2 and anti-phospho S6 Ribosomal protein (pS6) specific antibodies of protein extracts obtained from mESC treated with 100 μM CoCl2 for 48h. H2O was used as vehicle control (NT). Total S6 Ribosomal protein served as loading control. (J) qRT-PCR of H3K27me1 ChIP analyses in cells described in (I). Analyzed loci are described in 2.1.2H

2.1.3. The distinct PRC2-dependent H3K27 methylation domains functionally correlate with transcription states.

H3K36me3 is a histone PTM enriching at gene bodies of gene undergoing transcriptional elongation[45-48]. We wanted to have insights about the correlation between the deposition of the analyzed histone marks and the transcriptional status of genes, so we divided all annotated RefSeq genes in three main groups on the basis of their H3K27me2 and H3K27me1 levels in the intragenic regions, and we correlated these three groups respect to their transcriptional status. Genes transcription data were retrieved from microarray expression analyses previously generated and published under the accession number GSE19076[159]. As shown in the relative box plot graph, intragenic regions of highly transcribed genes bore high level of H3K27me1, while H3K27me2 accumulated at genes with lower transcriptional rates (figure 2.1.3A). RefSeq genes, ranked on the basis of their transcriptional rate were correlated with enrichment of H3K27me1 and H3K27me2 which were evaluated along the whole genes length, from transcription start site (TSS), to transcription end site (TES): this analysis showed that around 90% of all highly expressed genes were enriched in H3K27me1, whereas 90% of genes with low expression accumulated H3K27me2 throughout their gene bodies (figures 2.1.3B, C). We further confirmed these correlations performing the same analyses with high throughput transcriptome profiling through RNA sequencing (RNAseq) (figures 2.1.3D, E).

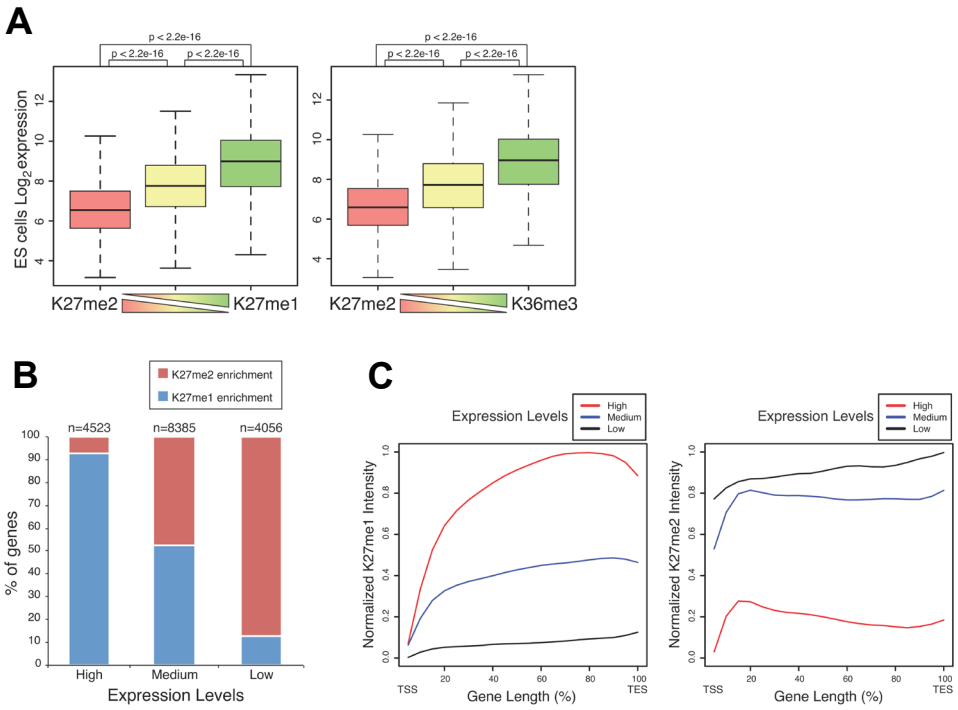


Fig. 2.1.3A-C Correlation analyses between K27 methylation enrichment and gene transcription rate. (A) Expression levels from microarray analyses (Leeb et al., 2010) of all RefSeq genes divided in three groups relative to the abundance of H3K27me2 and H3K27me1 within their gene bodies. Pearson correlation coefficient are indicated. (B) Percentage of K27me1 and K27me2 enriched genes relative to each class of expression levels. (C) Enrichment profiles of H3K27me1 and H3K27me2 deposition along the gene bodies (length expressed in %) for all the three classes of gene sets analyzed.

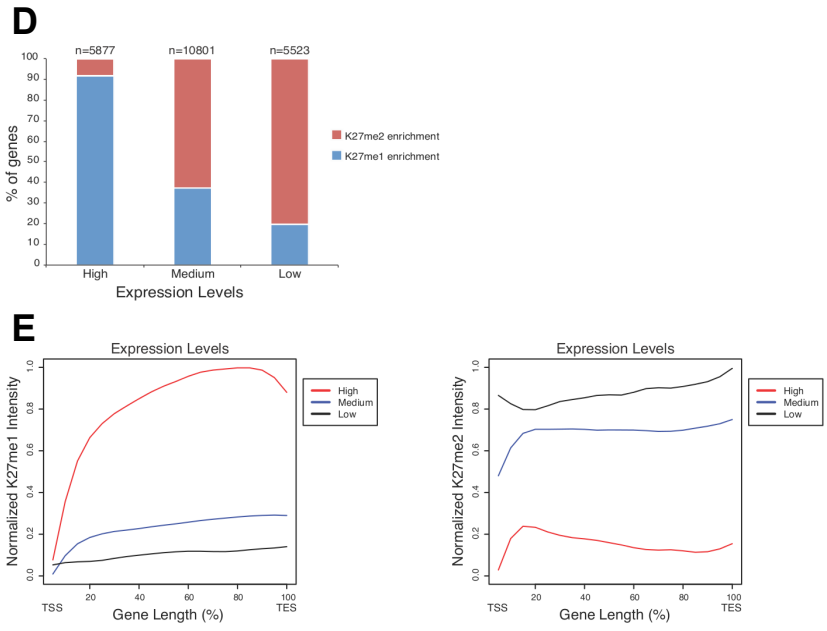


Fig. 2.1.3D, E. Correlation analyses between K27 methylation enrichment and gene transcription rate assessed by RNA-seq. (D) Relative proportion of genes representing the enrichment for either H3K27me1/H3K27me2within each class of expression generated by RNAseq. High correspond to expression > 3rdquantile; Medium correspond to expression < 3rd and > 1st quantile; Low correspond to expression < 1st quantile. (E) Average profiles of H3K27me1 and H3K27me2 deposition along the gene bodies (length expressed in %) for all three classes of gene sets shown in (D).

Altogether, these data indicate a tight control of H3K27 methylation forms exerted by PRC2; methylation domains throughout the genome are mutually exclusive and specifically correlate with transcriptional activity of genes.

2.1.4. Genome-wide deposition of H3K27me1 and H3K27me2 depends on PRC2 activity upon transient chromatin association.

The observation that H3K27 methylation forms are differentially enriched throughout the genome and that PRC2 associates preferentially with regions of H3K27me3 accumulation[97-99,156], suggested us that H3K27me1 and H3K27me2 deposition is exerted in the absence of a stable association of PRC2 to target loci.

Evidences in literature demonstrated that core PRC2 subunits associate with sites of ongoing replication[101], thus we could envisage a PRC2- dependent deposition of H3K27me1 and H3K27me2 immediately after the replication of DNA strands during the S phase of cell cycle.

At first instance we wanted to demonstrate that H3K27me1 and H3K27me2 are deposited after nucleosomes assembly to chromatin. To prove this, we performed immunoprecipitations (IP) of non-nucleosomal histone H3 from mESC soluble extracts, and analyzed H3K27 methylated forms by immunoblot analysis. H3K27me1, me2 and me3 were not detected in non-nucleosomal H3 fraction, suggesting that these PTMs are catalyzed only upon H3 incorporation into chromatin (figure 2.1.4A). These data are consistent with what previously reported in HeLa cell line[258-260].

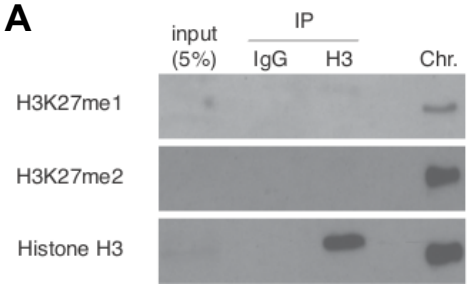


Fig. 2.1.4A H3K27 methylations forms are present only in chromatin incorporated H3 histones. WB analyses of immunopurified H3 histones from mESC soluble extract with the indicated antibodies. Urea extracted chromatin fraction served as positive control. IgG were used as control IP. Histone H3 is presented as loading control.

To gain insights about the deposition of H3K27me2 on nucleosomal H3 contemporarily to DNA synthesis, we decided to take advantage of Fluorescence Activated Cell Sorting (FACS) to analyze at the same time cell cycle phases and H3K27me2. To verify the specificity and the performance of the H3K27me2 antibody in immunofluorescent staining with cytometry analysis we stained and acquired at flow cytometer wild type and Eed KO mESC (figure 2.1.4B). Once the antibody specificity was verified, we pulsed wild type mESC with Bromo deoxy Uridine (BrdU) to define the S-phase boundaries, and we stained the cells with a H3K27me2 specific antibody to measure the accumulation of H3K27me2 throughout the different phases of the cell cycle (figure 2.1.4C). The correlation between the mean fluorescence values of H3K27me2 and the corresponding DNA content (represented by Propidium Iodide fluorescence intensity) allowed us to show that, in S-phase, H3K27me2 accumulated linearly respect to the increase of the DNA content. The correlation between DNA synthesis and H3K27me2 accumulation displayed a strong correlation (R^2=0.94) (Figure 2.1.4D). These data suggested that the PRC2 complex controls the modification of a large fraction of histone H3 during S-phase progression.

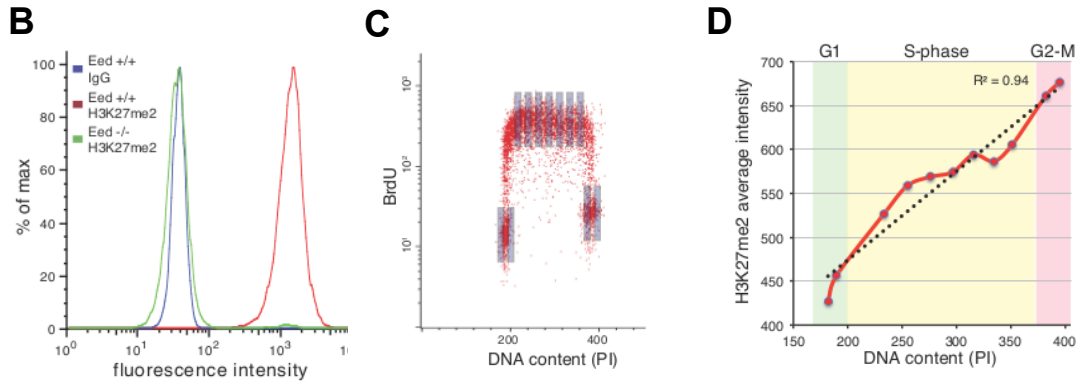
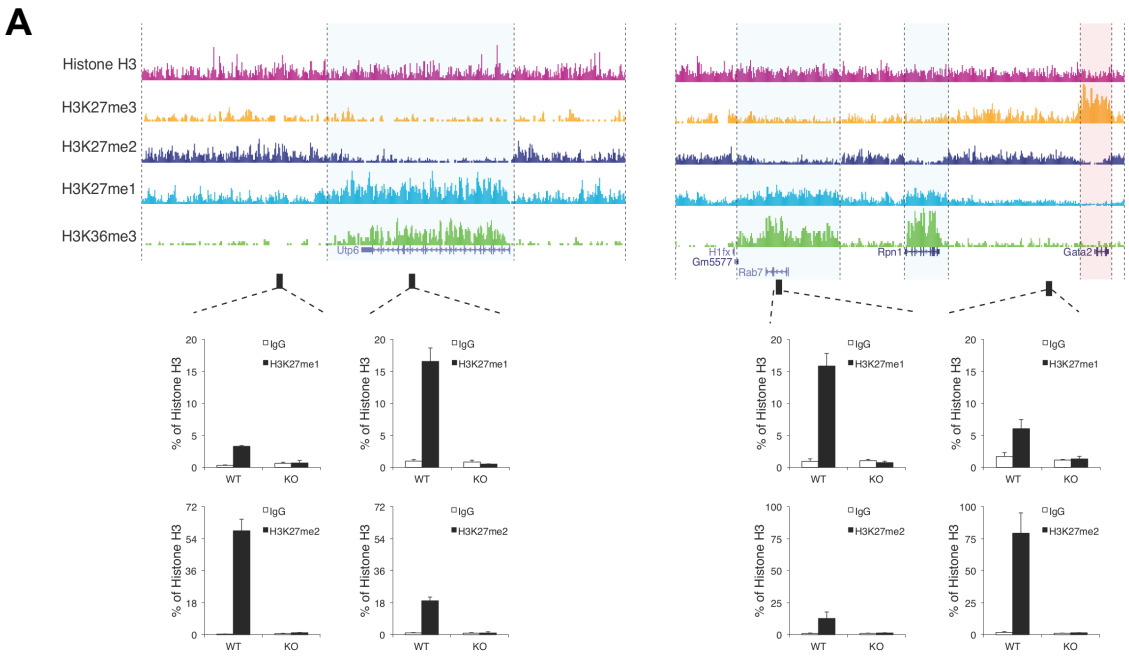
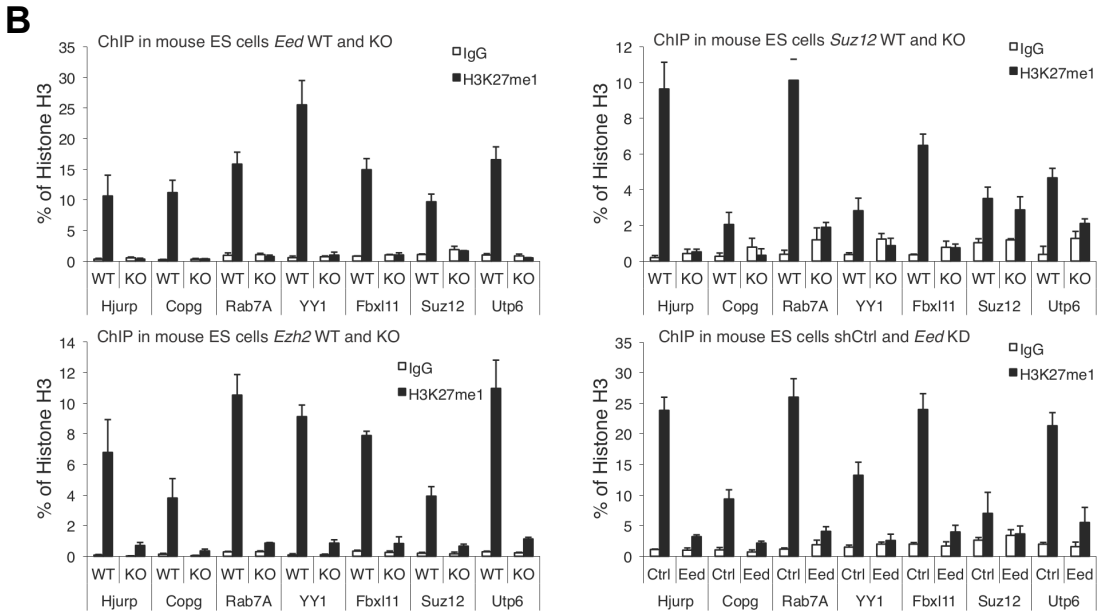


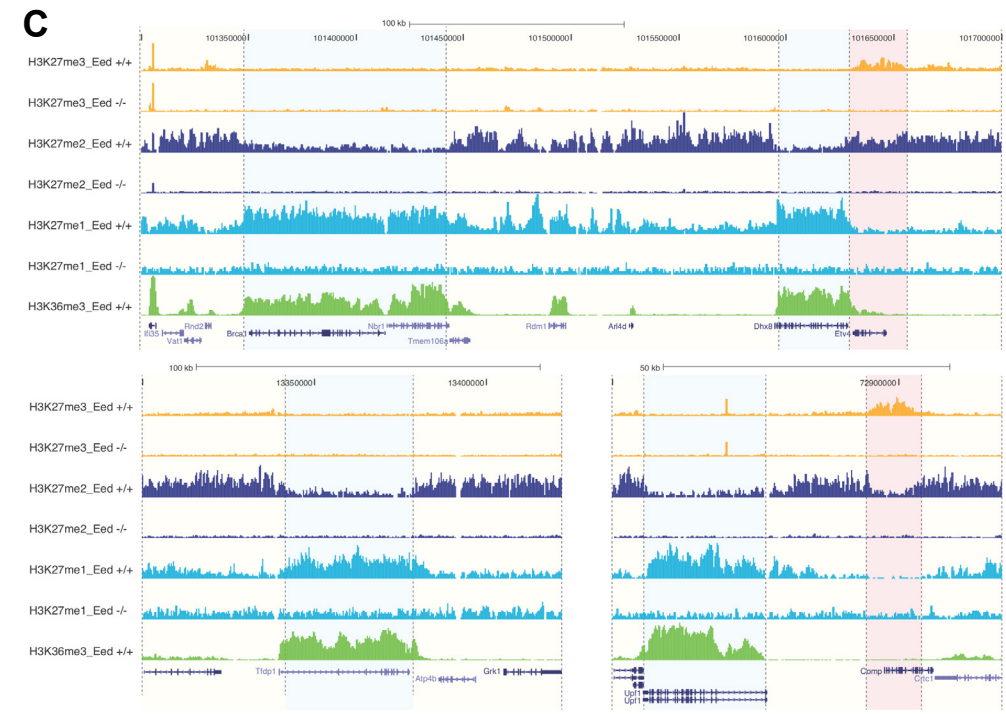
Fig.2.1.4 B-D. H3K27me2 is linearly deposited during cell cycle progession. (B) FACS analysis using H3K27me2 specific antibodies in WT (Eed +/+) and *Eed* KO (Eed -/-) mESC. Rabbit IgG were used as negative control. (C) FACS analysis of BrdU and Propidium Iodide (PI) incorporation in WT mESC. Gray boxes indicate the gates within the G1, S and G2 phases of the cell cycle, from which mean fluorescence values, used in (D), were retrieved. (D) Correlation plot between the average fluorescence intensities of H3K27me2 staining and Propidium Iodide of the gates defined in (C).

2.1.5. Deposition of H3K27me1 fully depends on PRC2 activity and it is linked to active transcription.

In order to provide evidences that H3K27me1 and H3K27me2 depositions are dependent on PRC2 enzymatic activity, we performed specific ChIP analyses in mESC WT or KO for PRC2 protein Eed ($Eed^{+/+}$ and $Eed^{-/-}$, respectively). ChIPs were analysed by real time quantitative PCR (qRT-PCR) using primer pairs specific for both intergenic and intragenic regions in correspondence of domains of H3K27me1 and H3K27me2 enrichment. ChIP results showed that intragenic H3K27me1 was lost in Eed KO cells (figure 2.1.5A). To further validate these results, we carried out the same experiments in Suz12, Ezh2 KO mESC and mESC in which *Eed* expression was impaired by shRNA transduction; also in these cell lines, the intragenic deposition of H3K27me1 was dependent on the enzymatic activity of the PRC2 complex (figure 2.1.5B). To observe this phenomenon at a genome-wide level, we performed H3K27me1 and H3K27me2 ChIPseq analyses in WT and Eed KO mESC. As shown in figure 2.1.5C, the genomic snapshots validated the qRT-PCR results, unveiling additional genomic loci loosing H3K27me1 enrichment. By means of bioinformatic analyses, performed to quantify genome wide H3K27me1 enrichment signal, we demonstrated that in Eed KO mESC, K27 monomethylation was lost from all the genes bearing H3K27me1 in WT cells (figure 2.1.5D). At the light of these evidences, we confirmed that H3K27me1 deposition at genes with a high transcriptional activity is dependent on the enzimatic activity of the PRC2 complex.


A
Histone H3
H3K27me3
H3K27me2
H3K27me1
H3K36me3
Utp6
H11x
Gm5577
Rab7
Rpn1
Gata2
% of Histone H3
IgG
H3K27me1
H3K27me2
WT
KO


B
ChIP in mouse ES cells Eed WT and KO
ChIP in mouse ES cells Suz12 WT and KO
ChIP in mouse ES cells Ezh2 WT and KO
ChIP in mouse ES cells shCtrl and Eed KD
% of Histone H3
IgG
H3K27me1
WT
KO
Ctrl
Eed
Hjurp
Copg
Rab7A
YY1
Fbxl11
Suz12
Utp6



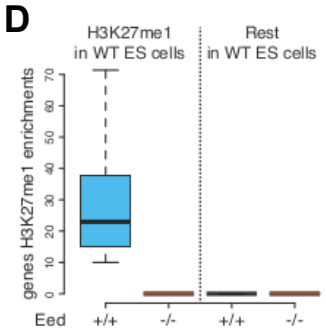


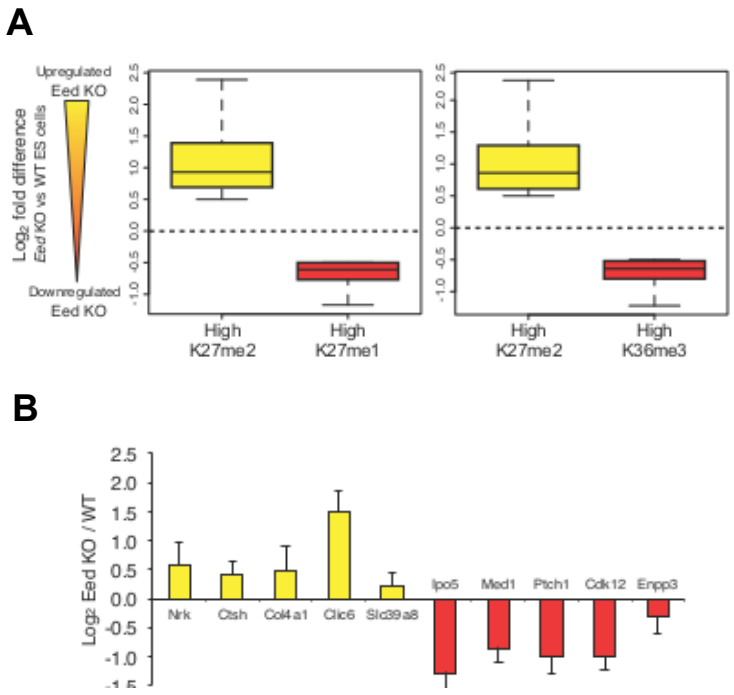
Fig.2.1.5 A-D. Genome wide H3K27me1 deposition is PRC2 dependent. (A) qRT-PCR of ChIP analyses in WT and Eed KO mESC performed with the indicated antibodies for the loci shown by USCS genome browser snapshots in the top panels. Black boxes indicate primers position within genomic loci. ChIPs with IgG rabbit were made as negative control. ChIP enrichments are normalized to histone H3 density. (B) qRT-PCR using primers for the indicated intragenic regions on DNA purified from ChIP assays performed in the specified mESC, made using the indicated antibodies. ChIPs with IgG rabbit were performed as negative control. ChIP enrichments are normalized to histone H3 density.
(C) Genomic snapshots of H3K27me1, H3K27me2 and H3K27me3 ChIPseq analyses in WT (Eed +/+) and Eed KO (Eed -/-) mESC together H3K36me3 ChIPseq analyses from E14 mESC. H3K27me1 domains are highlighted in blue while H3K27me3 domains are highlighted in red. (D) Box plot analyses of H3K27me1 ChIPseq enrichment intensities between WT (+/+) and Eed KO (-/-) mESC for all the annotated RefSeq genes that were divided in two groups based on their H3K27me1 levels in WT mESC (10Log p-value cut off = 10).

2.1.6. PRC2 dependent H3K27me1 correlates with genomic loci characterized by active transcription.

As described in paragraph 2.1.3, H3K27me1 and H3K27me2 correlate with the transcriptional status of the genes at which they are deposited. For this reason we envisaged a possible role for intragenic H3K27me1 in promoting gene transcription. To prove this hypothesis, we analysed the transcription data retrieved from WT and Eed KO mESC (as described in 2.1.3 section)[159] correlating the transcriptional

changes of genes that were differentially expressed with the enrichment of H3K27 methylation forms at intragenic regions found in WT mESC. Such correlations, shown in figure 2.1.6A as box plot graphs, evidence that genes enriched in H3K27me2 had increased transcription, while genes with significant H3K27me1 enrichment were characterized by a reduced expression in Eed KO mESC. To confirm these results, we carried out expression analyses and ChIP experiments followed by qRT-PCR for selected target genes (figures 2.1.6B, C).

Another useful tool to assess the link between gene expression changes and H3K27 methylation enrichment was luciferase assay. We took advantage of a 293T cell line stably expressing Luciferase under the control of a stably integrated promoter[101]. ChIP analyses showed H3K27me1 accumulation within Luciferase coding region, but upon forced recruitment of PRC2 to Luciferase promoter driven by Gal4-Jarid2 fusion protein, the levels of both H3K27me1 and H3K36me3 drastically reduced in favour of H3K27me3 deposition as demonstrated by ChIP-qPCR analyses (figure 2.1.6D).



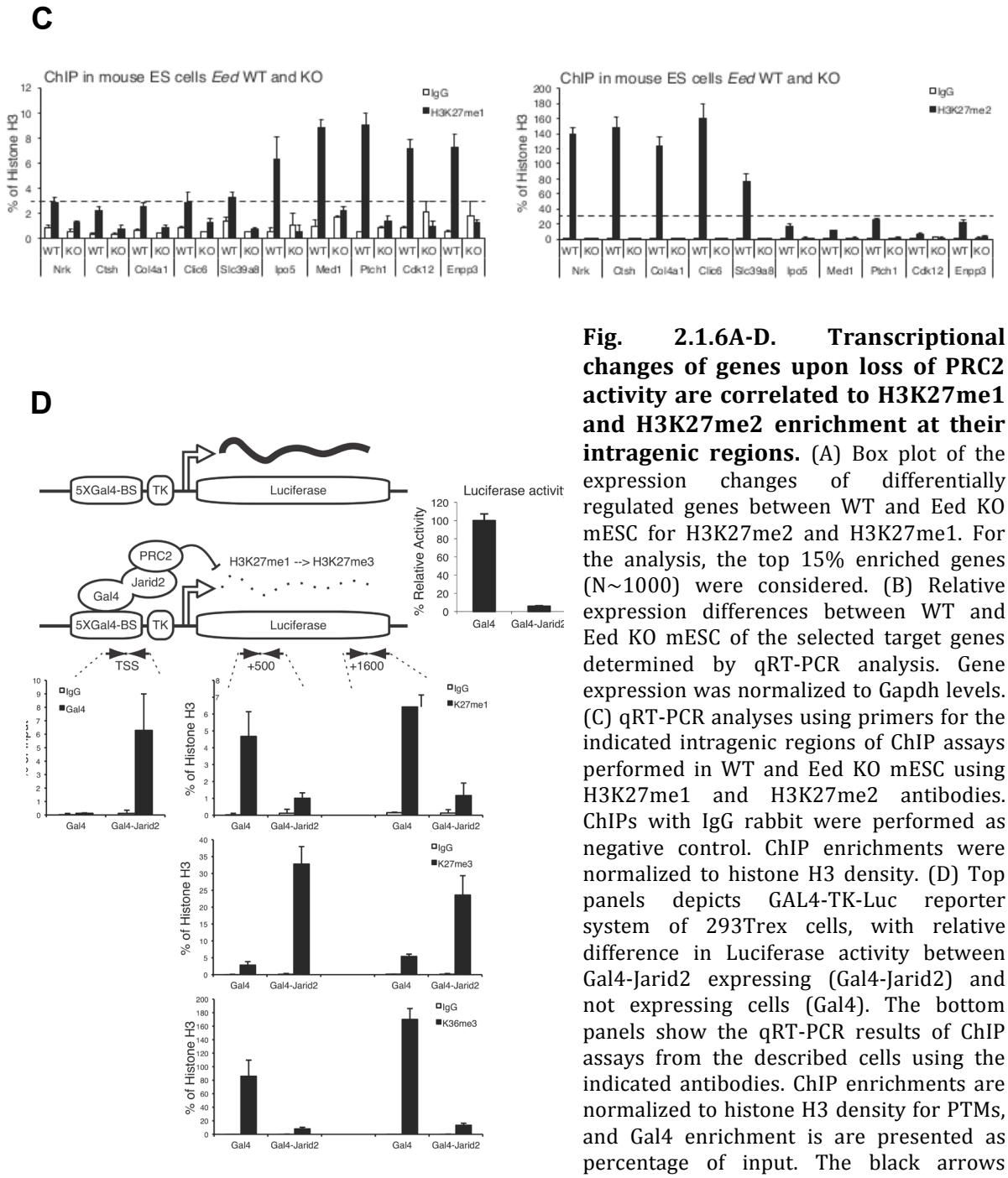


Fig. 2.1.6A-D. Transcriptional changes of genes upon loss of PRC2 activity are correlated to H3K27me1 and H3K27me2 enrichment at their intragenic regions. (A) Box plot of the expression changes of differentially regulated genes between WT and Eed KO mESC for H3K27me2 and H3K27me1. For the analysis, the top 15% enriched genes (N~1000) were considered. (B) Relative expression differences between WT and Eed KO mESC of the selected target genes determined by qRT-PCR analysis. Gene expression was normalized to Gapdh levels. (C) qRT-PCR analyses using primers for the indicated intragenic regions of ChIP assays performed in WT and Eed KO mESC using H3K27me1 and H3K27me2 antibodies. ChIPs with IgG rabbit were performed as negative control. ChIP enrichments were normalized to histone H3 density. (D) Top panels depicts GAL4-TK-Luc reporter system of 293Trex cells, with relative difference in Luciferase activity between Gal4-Jarid2 expressing (Gal4-Jarid2) and not expressing cells (Gal4). The bottom panels show the qRT-PCR results of ChIP assays from the described cells using the indicated antibodies. ChIP enrichments are normalized to histone H3 density for PTMs, and Gal4 enrichment is are presented as percentage of input. The black arrows indicate the position of the qRT-PCR primers relative to the GAL4-TK-Luc construct.

The observations made in Gal4-Luciferase reporter system also support that H3K27me1 is transiently deposited by PRC2, while conversion to H3K27me3 form requires PRC2 to stably bind target promoter to be established.

Overall, these result further confirm the correlation between H3K27me1 with active transcription and suggest that the stable recruitment of the PRC2 complex to DNA induces the full conversion of H3K27 to its tri-methylated form.

2.1.7. PRC2-dependent H3K27me1 is necessary for correct gene transcription.

The transcriptional changes observed upon loss of PRC2 enzymatic activity, correlated to the deposition of H3K27me1, suggest that H3K27me1 is required for correct activation of transcription of PRC2 target genes. To further prove this relation, we tested if it occurred also during the differentiation process of WT and Eed KO mESC, using cell culture differentiation into embryoid bodies (EBs) as model of study. This protocol allows the formation of three-dimensional aggregates of cells differentiating toward the three germ layers upon LIF removal and addition of retinoic acid. Eed KO cells were unable to properly differentiate respect to wild type cells, as confirmed by the lack of activation of canonical differentiation genes (figure 2.1.7A).

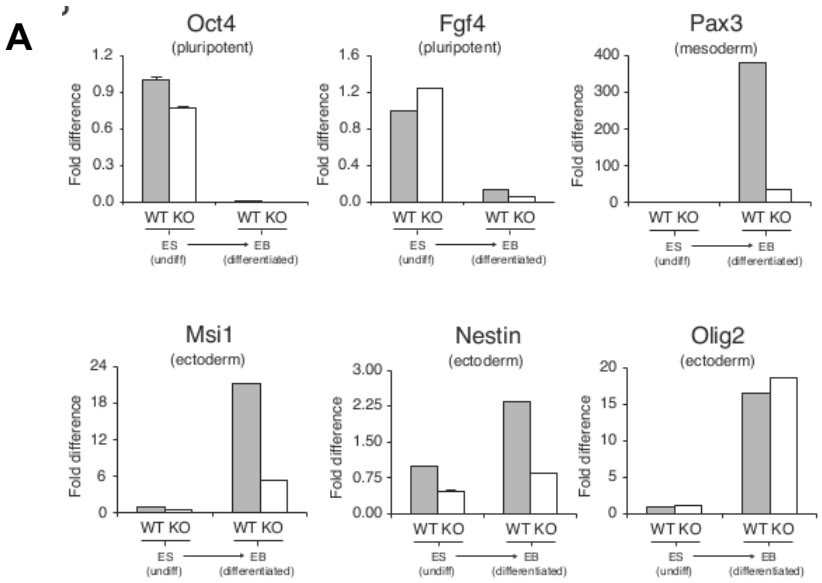


Fig.2.1.7A. PRC2 deficient mESC fail to properly differentiate. Relative expression of the indicated differentiation markers determined by qRT-PCR in WT and *Eed* KO mESC before (ES) and after 9 day (EB) differentiation. Gene expression is normalized to *Gapdh* levels.

We took advantage of previously published genome wide expression profiling[86] of both wild type and PRC2 depleted (e.g. sh control and sh Jarid2, respectively) EBs to identify genes whose activation is impaired upon differentiation in absence of PRC2 activity. RT-qPCR analyses on these selected genes confirmed their failure to be activated in Eed KO EBs respect to the wild type counterpart (figure 2.1.7B, top panel). To generate insights about the H3K27me1 status at intragenic regions of these targets, we carried out specific ChIP-qPCR analyses which showed a H3K27me1 loss from intragenic regions in absence of PRC2 (figure 2.1.7B, bottom panels). Although those genes failed to be expressed, the levels of H3K36me3, a histone PTM associated to active transcription, on the same regions remained unaltered, suggesting that this modification could function upstream of H3K27me1 deposition.

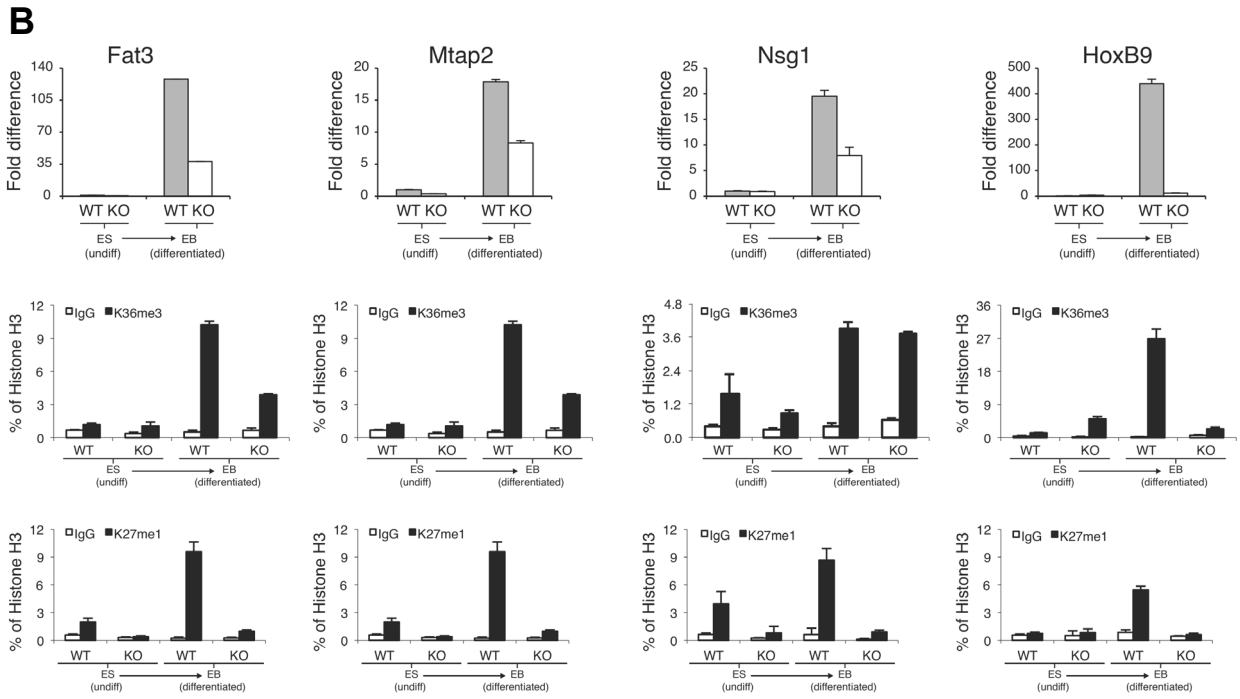


Fig.2.1.7B PRC2 dependent H3K27me1 intragenic deposition is necessary for the activation of differentiation genes in mESC. Top panels show relative expression of the indicated genes determined by qRT-PCR in WT and Eed KO mESC before (ES) and after (EB) differentiation. Gene expression is normalized to Gapdh levels. Middle and bottom panels show qRT-PCR analyses made using primers for the indicated intragenic regions on DNA purified from ChIP assays for H3K27me1 and H3K36me3 performed in the same cells. ChIPs with IgG rabbit served as negative control. ChIP enrichments were normalized to histone H3 density.

In addition, we carried out global transcriptome profiling by RNA sequencing analyses on wild type and Eed KO EBs, as well as on un-differentiated counterparts. Bioinformatic analyses demonstrated a global impairment in the activation of genes upon differentiation in PRC2 KO mESC, and this correlated with lack of deposition of H3K27me1 at their gene bodies. The deposition of H3K36me3 mark resulted quite unaltered (figure 2.1.7C, D).

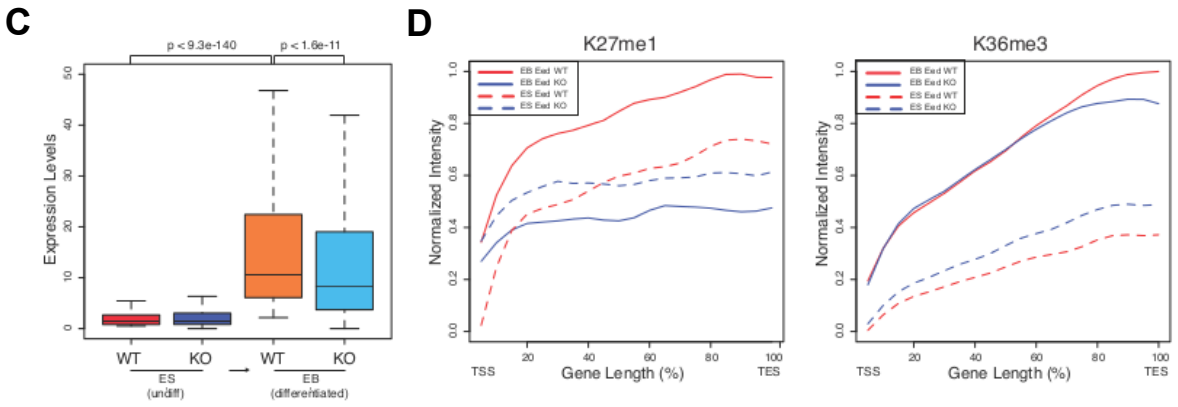


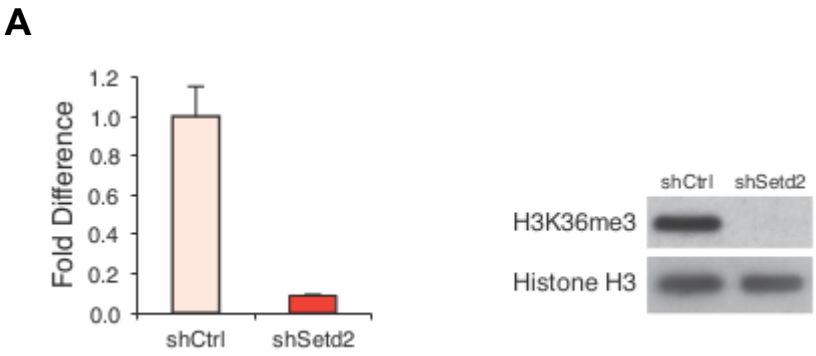
Fig.2.1.7 C, D. Upregulation of genes during differentiation process correlates with gain of H3K27me1 deposition at their intragenic regions. (C) RNAseq analyses showing the expression levels of up-regulated genes during differentiation process in WT and Eed KO samples. The total number of upregulated genes is 844 (N=844). (D) Average deposition profiles of H3K27me1 and H3K36me3 through the intragenic regions (length expressed in %) of genes activated upon EB differentiation (N=844).

At the light of these evidences, we could conclude that, also during differentiation process, H3K27me1 is deposited by PRC2 enzymatic activity and it is required for proper transcriptional activation of PRC2 target genes. The results also highlighted that H3K36me3 is a PTMs that temporally precedes H3K27me1 deposition.

2.1.8. Mechanisms regulating H3K27me1 deposition: the role of Setd2 dependent H3K36me3.

An interesting issue to investigate was the mechanism that allows PRC2 to be prevented from converting the monomethylated state in order to establish precise and mutually exclusive chromatin domains. Several reports present in the literature

explain how histone PTMs are able to cross-talk each other, thus adding another layer of complexity in the regulation of their functional outcomes[261]. Given the correlation between H3K27me1 and H3K36me3 deposition patterns, we envisaged a possible role for H3K36me3 in the regulation of H3K27me1 switch to H3K27me2. The H3K36me3 modification is the results of the enzymatic activity of the Setd2 methyltransferase, which is also able to recruit RNA Pol II. In order to decrease H3K36me3 levels in mESC, we developed shRNAs specifically targeting Setd2, as shown by the reduced protein expression level measured by RT-qPCR (figure 2.1.8A). We used these lentiviral transduced mESC lines to perform ChIP analyses for H3K27me1 and H3K27me2 and H3K36me3 as control. Enrichments of these histone PTMs at intragenic regions of selected loci, measured by RT-qPCR analyses, showed a loss of H3K36me3 levels accompanied with a decrease in H3K27me1 in favour of H3K27me2 accumulation (figure 2.1.8B).



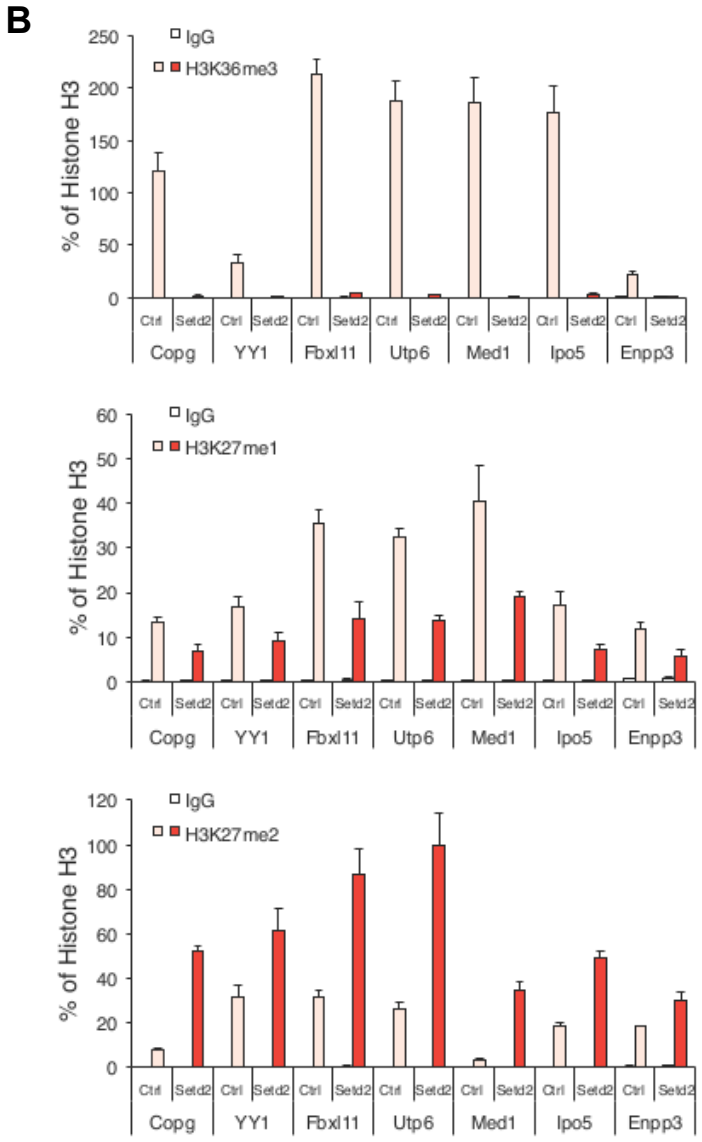


Fig. 2.1.8A, B. Setd2-dependent H3K36me3 coordinates the balance between H3K27me1 and H3K27me2. (A) Left panel shows the expression level of Setd2 protein measured by RT-qPCR in mESC interfered with control shRNA and shRNA specifically targeting *Setd2* coding sequence. Expression levels are normalized to Gapdh. Right panel show immunoblot analysis with the indicated antibodies of extracted histones from mESC lines described before. Total H3 was used as loading control. (B) qRT-PCR analyses made using primers for the indicated intragenic regions on DNA purified from ChIP assays for H3K27me1, H3K27me2 and H3K36me3 performed in the same cells described in (A). ChIPs with IgG rabbit were made as negative control. ChIP enrichments were normalized to histone H3 densitv.

To investigate the transcriptional changes caused by the loss of Setd2 activity, in comparison with those observed upon PRC2 depletion, we carried out genome wide transcriptome profiling through RNAseq analysis of mESC lines described before, and we determined the fraction of downregulated genes induced by the loss of

H3K36me3. Interestingly, a significative fraction of these downregulated genes was in common with those that were downregulated in Eed KO mESC (figure 2.1.8C). This result is consistent with the concomitant loss of H3K36me3 and H3K27me1 upon knock down of Setd2 protein.

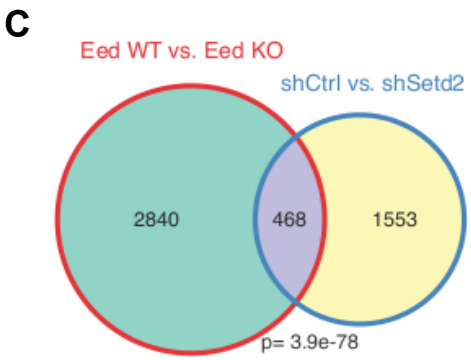


Fig.2.1.8C. Overlap of downregulated genes upon loss of Setd2 protein and PRC2. Venn diagram depicting the extend of overlap of regulated genes between the indicated conditions. p value was calculated by hypergeometric distribution.

The loss of Setd2, although reducing H3K27me1 levels at intragenic regions, did not affect the expression of PRC2 component. It is known that H3K36me3 can be recognized by the Tudor domain retained by Phf19 protein, which in turn can act as PRC2 recruiter to a set of target loci in mESC[110]. To exclude the possibility that Setd2 could be required for the localization to chromatin and/or the regulation of the catalytic properties of PRC2, we tested the localization of PRC2 to DNA replication foci during S phase in mESC lacking Setd2. Immunofluorescence staining revealed that, in absence of Setd2, colocalization between PRC2 core member Suz12 and BrdU foci was unaltered respect to control cells (figure 2.1.8D).

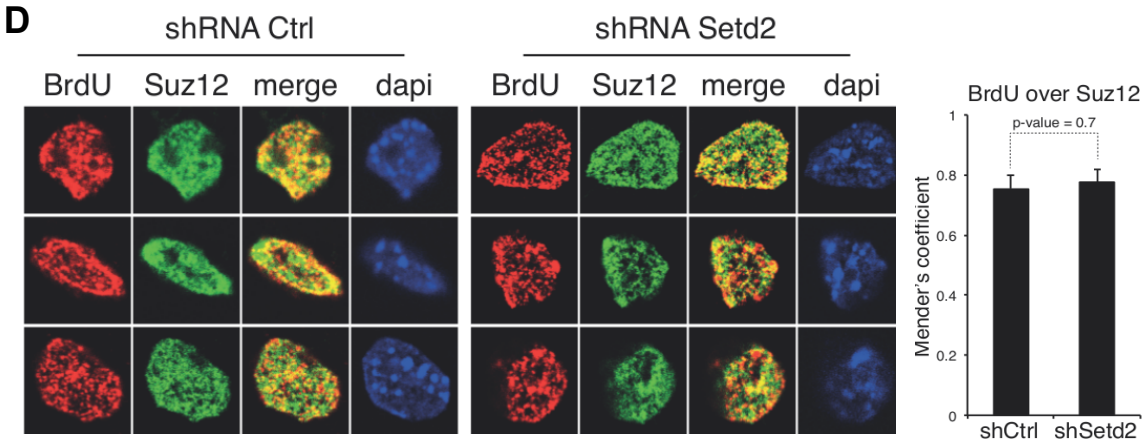
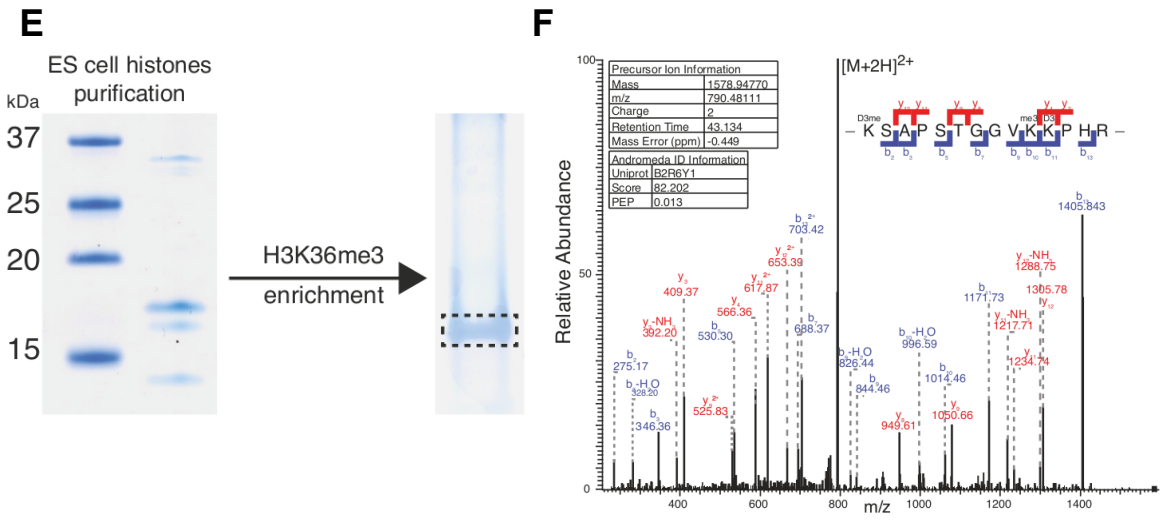


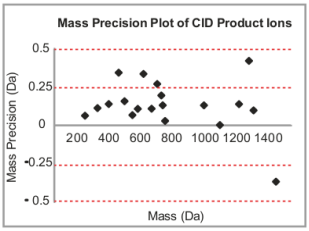
Fig.2.1.8D. Setd2 depletion does not impair PRC2 localization to DNA replication foci. Confocal immunofluorescence images of mESC expressing scrambled (shRNA Ctrl) or *Setd2* specific shRNAs (shRNA Setd2) 3 h after release in presence of BrdU after a 12 h long thymidine block, stained with the indicated antibodies. Bar plots indicate the average Mender's co-localization coefficient among all Z-stacks for each cell. Error bars indicate SD. P value was calculated by student t-test.

Overall, these data support the existence of a mechanism by which the presence of H3K36me3 prevents PRC2 from further methylate H3K27me1 to H3K27me2. To futher endorse the existence of an *in cis* mechanism of inhibition of H3K27me2 formation mediated by H3K36me3, we wanted to demonstrate that H3K36me3 and H3K27me1 co-exist on the same histone N-terminal tails. To address this issue we purified histones by acidic extraction from mESC which were further immunopurified with antibody specific for H3K36me3 (figure 2.1.8E) and then subjected to tandem mass spectrometry analysis to unveil PTMs. Analyzed histone peptides revealed that H3K27me1 and H3K36me3 reside on the same histone N-terminal tails, in particular at those belonging to H3.3 variant (figures 2.1.8.F, G). Mass spectrometry analyses were performed in collaboration with members of Dr. Tiziana Bonaldi's group at European Institute of Oncology (IEO). This suggested an *in cis* mechanism of regulation, and in addition, it was coherent with the prevalence of histone H3.3 isoform at genes characterized by high expression levels[11].



G

b ion-H_2O	b^{2+} ion	b ion		seq		y ion	y^{2+} ion	y ion-NH_3
	mass	mass				mass	mass	
--	--	188.1473	1	K(D3me)	14	--	--	--
--	--	275.1793	2	S	13	1392.815	--	--
328.2059	--	346.2164	3	A	12	1305.783	653.3951	1288.756
--	--	443.2692	4	P	11	1234.746	617.8765	1217.719
--	--	530.3012	5	S	10	1137.693	--	--
--	--	631.3489	6	T	9	1050.661	525.8341	--
--	--	688.3704	7	G	8	949.6133	--	--
--	--	745.3918	8	G	7	892.5919	--	--
826.4497	--	844.4602	9	V	6	835.5704	--	--
996.5916	--	1014.602	10	K(me3)	5	736.502	--	--
--	--	1171.732	11	K(D3)	4	566.3601	--	--
--	--	1268.784	12	P	3	409.2306	--	392.2041
--	703.4253	1405.843	13	H	2	312.1779	--	--
--	--	--	14	R	1	175.119	--	--



Annotation	
Protein	B2R6Y1
Protein Localisation	27-40
AminoAcid Coverage	79%
Method	ITMS, CID

Andromeda Search Configuration			
Variable Modification			
Title	Description	Specificity	ΔM (Da)
Lysines D3-acetylation (D3)	deuterate acetylation	K	45.0294
Lysines monomethylation (D3me)	D3-acetylation+monomethylation	K	59.0454
Lysines dimethylation (me2)	dimethylation	K	28.0310
Lysines trimethylation (me3)	trimethylation	K	42.0460

Fig.2.1.8 E-G. H3K36me3 and H3K27me1 PTMs reside on the same N-terminal tail of histone H3.3 variant. (E) Coomassie gel stainings of acidic extracted histones from mESC before and after immunopurification with H3K36me3 specific antibody.(F) Annotated MS/MS spectrum of modified (27-40) KSAPSTGGVKKPHR peptide from H3.3 generated from the fragmentation of the precursor ion with m/z 790.4811 and charge +2 assigned to the tetra-methylated form of peptide H3.3 (27-40); upon chemical modification (D3-acetylation) reveals coexistence of mono-methylated K27 and tri-methylated K36 within the same peptide. Uniprot accession number, score and posterior error probability (PEP) relative to peptide identification by MaxQuant are reported in the left panel. (G) Annotation of product ions from CID fragmentation of tetra-methylated and D3-acetylation peptide (27-40) KSAPSTGGVKKPHR from H3.3. The table summarizing the b- and y-ions produced by CID-fragmentation is reported, as generated by Viewer.exe software. The b- and y ions experimentally detected are reported in blue and red, respectively. Mass deviation for each experimental ion in the fragmentation spectrum is less than 0.5 Da (bottom left panel). The different modification upon *in vitro* D3-acetyl alkylation and the corresponding ΔM that were specified as variable modifications in Andromeda Search Configuration are annotated (bottom right panel).

Previous studies performed *in vitro* on nucleosomes demonstrated that H3K36me3 blocks the deposition of H3K27me2 and H3K27me3, while H3K27me1 can be efficiently deposited[262,263]. To validate this inhibition in our model of study, we performed *in vitro* methyltransferase assay. Recombinant PRC2 complex was assayed for its efficiency in H3K27me2 deposition on recombinant histones both unmodified or fully modified as H3K36me3 (Methyl-lysine analogue (MLA) histones). Enzymatic products were subjected to MS analyses, unveiling the PRC2 inefficiency in H3K27me2 formation when in presence of H3K36me3 (figure 2.1.8H).

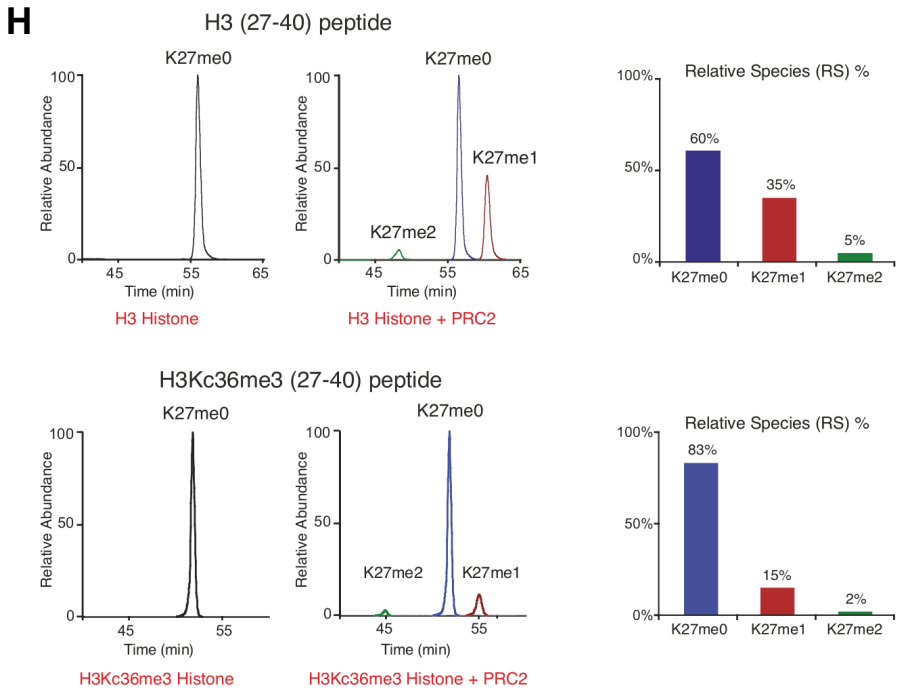


Fig.2.1.8H. H3K36me3 inhibits *in vitro* PRC2-mediated H3K27me1 to H3K27me2 conversion. Extract ion chromatograms relative to (27-40) KSAPSTGGVKKPHR peptide from unmodified recombinant histone H3 (H3 Histone) and recombinant histone H3 with K36me3 (H3Kc36me3 Histone) and estimation of relative percentage species (RS%). In the upper panel: extract ion chromatograms of precursor ions corresponding to the recombinant H3 (27-40) peptide upon enzymatic reaction were identified as unmodified (K27me0), mono- (K27me1) and di-methylated (K27me2) on K27 and depicted in blue, red and green, respectively (H3 Histone + PRC2). Extract ion chromatograms of the same peptide in absence of enzyme was also reported as a control (H3 Histone). In the lower panel: extract ion chromatograms of precursor ions of recombinant H3Kc36me3 (27-40) peptide upon enzymatic reaction (H3Kc36me3 Histone + PRC2). Extract ion chromatograms of the same peptide in absence of enzyme was also reported as a control (H3Kc36me3 Histone). In each case relative species percentage (RS%) was also displayed in bar plot. Shift in retention time among the different degree of modification is due to D3-acetyl chemical alkylation on lysine residues.

2.1.9. Genome-wide H3K27me2 deposition ensures enhancer fidelity in a cell type specific manner.

The observations that H3K27me2 accounts for approximately the 70% of total K27 PTMs of histone H3 and that it is widely deposited throughout the genome in mESC, led us to envisage that this methylation mark had important functions. We expected a role in the context of genomic control, rather than merely being as a mark that is present between functional domains marked by H3K27me1 and H3K27me3. Previous published works reported an upregulation of H3K27ac upon loss of PRC2 enzymatic activity[27,86]; we were able to observe the same phenomenon in mESC lines KO or interfered for PRC2 components compared to wild type or control mESC (figure 2.1.9A). H3K27ac has been reported to be enriched, in combination with other PTMs, at transcription start sites (TSS) of genes (in association with high levels of H3K4me3) and at enhancer elements (in association to low levels of H3K4me3 and high H3K4me1 enrichment).

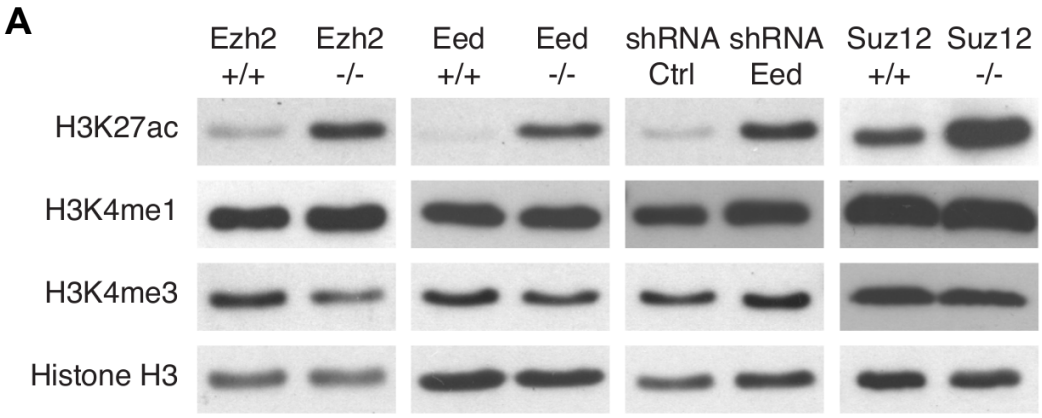


Fig.2.1.9 A. Global levels of H3K27ac increase upon loss of PRC2 activity. WB analyses of histones extracted from the indicated mESC lines with indicated antibodies. Total H3 was used as loading control.

The fact that H3K27me2 accumulates at intergenic regions made us to think about a possible connection between H3K27me2 and enhancer regulatory elements. Enhancer elements can be found in two functional states: poised or active on the basis of H3K27ac levels (low and high, respectively)[264,265]. To investigate a possible involvement for H3K27me2 in enhancer control, we carried out ChIPseq analyses to assess the genome wide behavior of H3K27ac in mESC in absence of PRC2 activity respect to the wild type counterpart. Consistently to the global increase of H3K27ac observed in the immunoblot analyses, ChIPseq highlighted a differential enrichment of H3K27ac between Eed KO and WT mESC. Indeed, as shown by the Venn diagrams presented in figure 2.1.9B, a number of H3K27ac peaks were specifically gained in Eed KO (called "Eed KO unique"), while other peaks were present only in Eed wt mESC ("Eed wt unique"). Examples of "unique" peaks are shown in figure 2.1.9C, which were validated by ChIP-qPCR analyses using primer sets specific for the selected genomic regions (figure 2.1.9D). Peaks were annotated relatively to ±. 2.5 Kb region across RefSeq annotated TSS. We analyzed the peaks distribution among TSS and not TSS regions, and we noted that common peaks shared by both Eed WT and Eed KO were equally distributed among TSS and not TSS regions; "unique" H3K27ac peaks, instead, showed a preferential enrichment at not TSS regions (figure 2.1.9B). These results suggested that PRC2 activity can induce variations in H3K27ac deposition in correspondence of potential regulatory elements, implying that the broad H3K27me2 deposition could exert protective functions in preventing the activation of non-cell type specific enhancer elements.

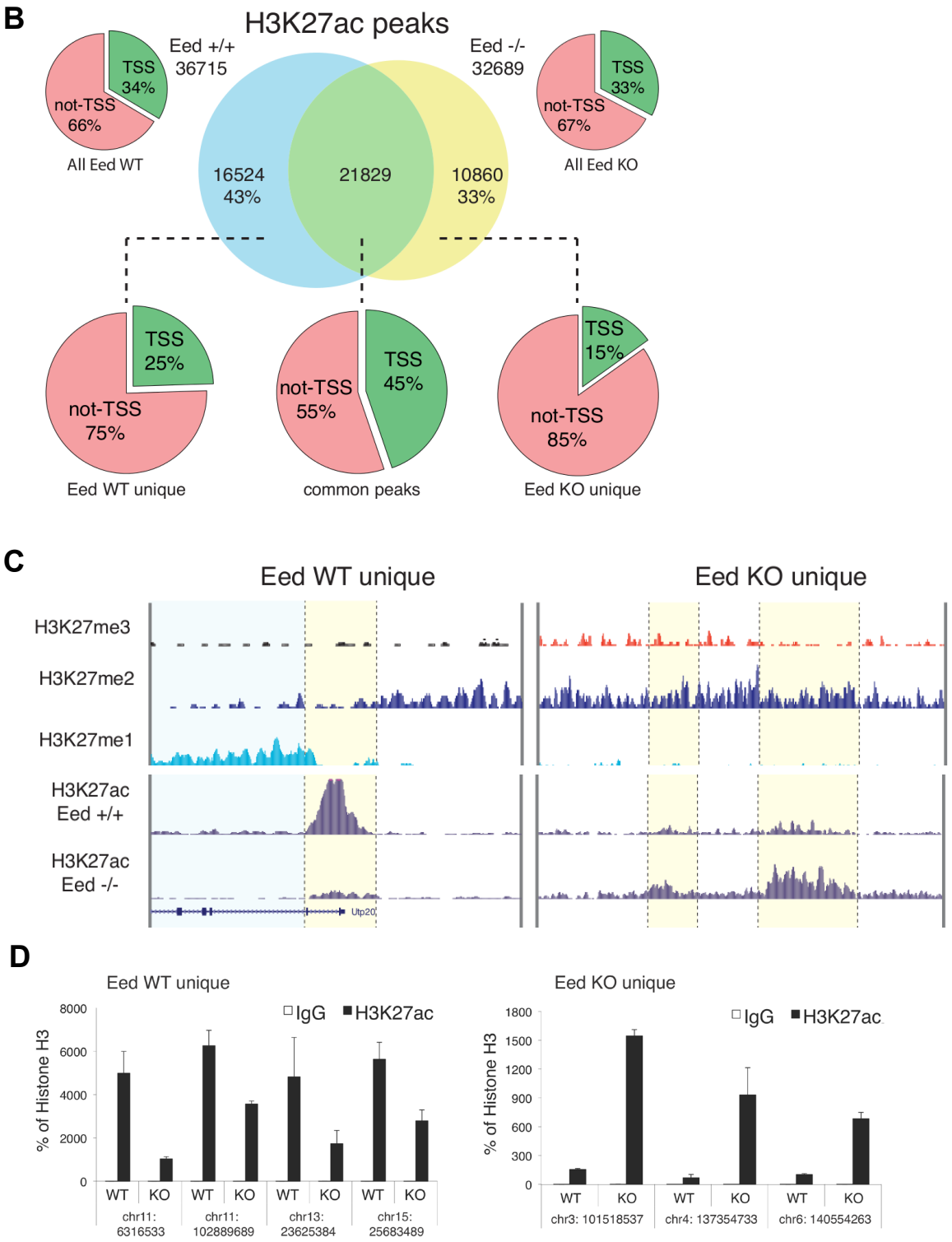


Fig.2.1.9 C-D. Differential distribution of H3K27ac peaks upon loss of PRC2 activity in mESC. (B) Genome-wide overlap of H3K27ac peaks between WT and Eed KO mESC. The pies depict the percentage distribution of the different groups of H3K27ac peaks relative to promoter region of all RefSeq genes defined as a ±. 2.5kb region around centered the TSS. (C) Genomic snapshots of ChIPseq analyses for H3K27me1, H3K27me2, H3K27me3, H3K27ac, in WT (Eed +/+) and Eed KO (Eed -/-) mESC. Regions of regulated H3K27ac are highlighted in yellow. (D) qRT-PCR analyses of DNA purified from H3K27ac ChIP in WT and *Eed* KO mESC using primers amplifying the indicated genomic loci. Chromosomal coordinates correspond to the center of the amplified genomic region. Purified rabbit IgG were used as negative control. ChIP enrichments were normalized to histone H3 density.

In order to map enhancer elements in Eed wt and Eed KO mESC we carried out ChIPseq analyses for H3K4me1 and H3K4me3. As shown by snapshots from the UCSC genome browser visualizing ChIPseq analyses, and by the global quantification reported in the box plot analysis, the sites of novel H3K27ac deposition (unique peaks) are enriched in H3K4me1 but lack H3K4me3 (figure 2.1.9E, F).

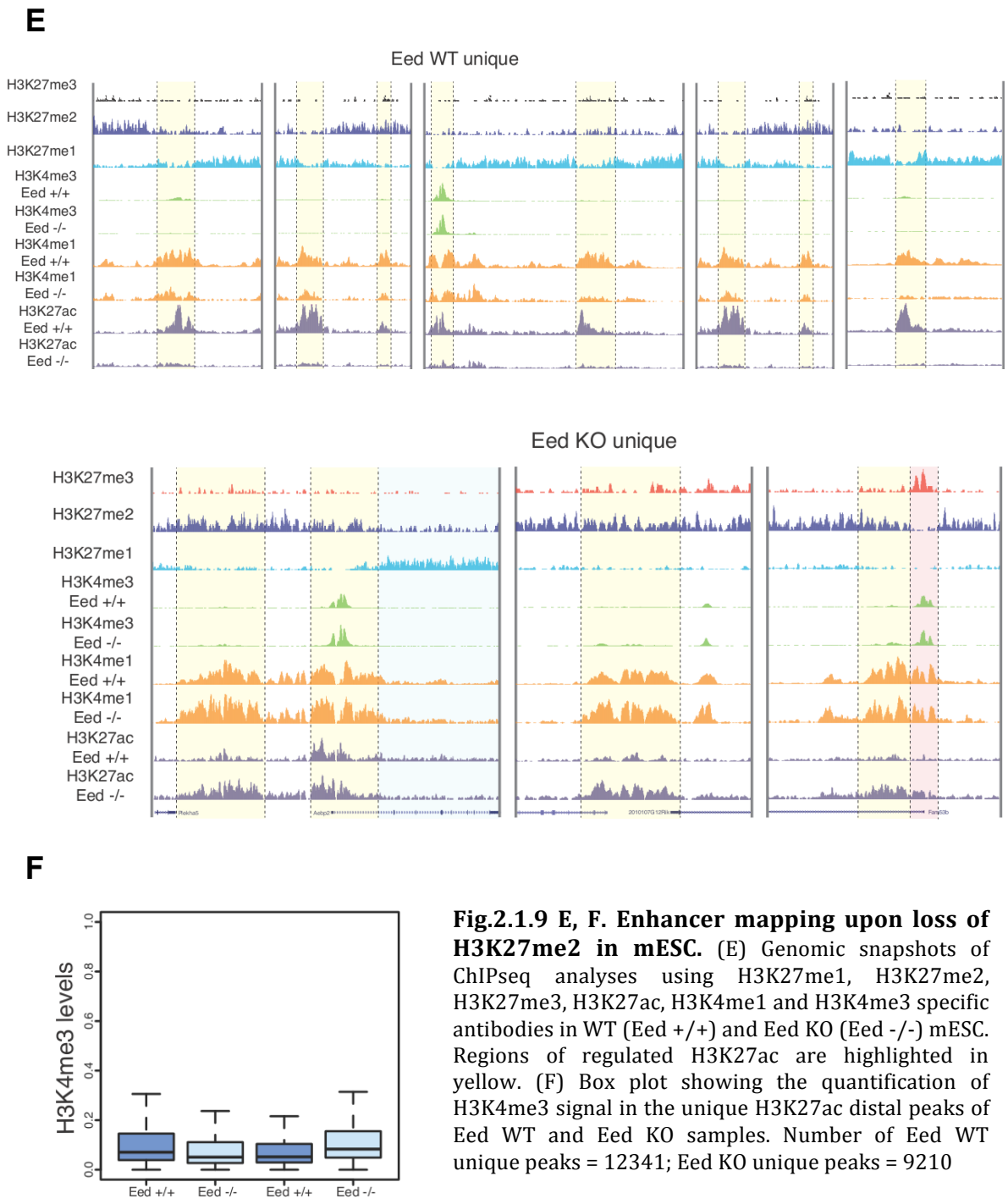
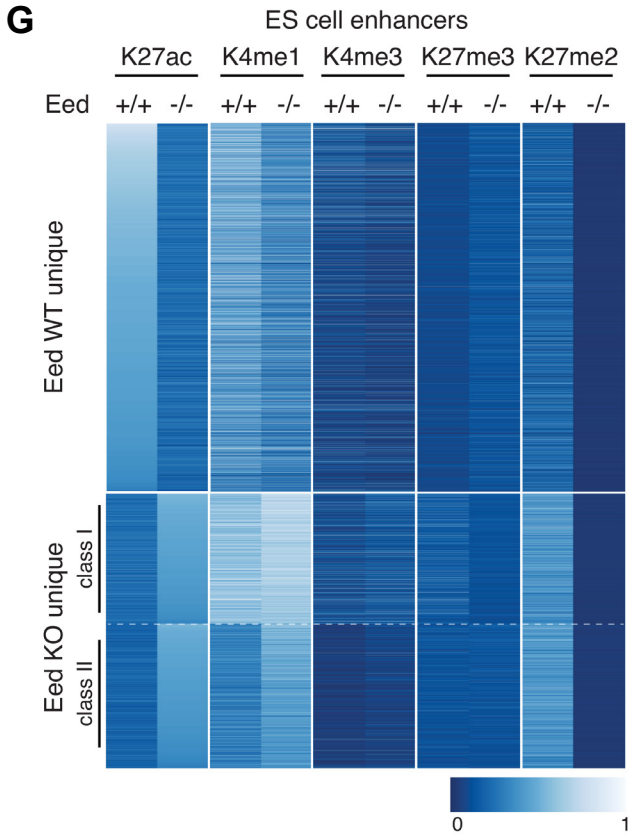


Fig.2.1.9 E, F. Enhancer mapping upon loss of H3K27me2 in mESC. (E) Genomic snapshots of ChIPseq analyses using H3K27me1, H3K27me2, H3K27me3, H3K27ac, H3K4me1 and H3K4me3 specific antibodies in WT (Eed +/+) and Eed KO (Eed -/-) mESC. Regions of regulated H3K27ac are highlighted in yellow. (F) Box plot showing the quantification of H3K4me3 signal in the unique H3K27ac distal peaks of Eed WT and Eed KO samples. Number of Eed WT unique peaks = 12341; Eed KO unique peaks = 9210

To quantify global signal intensities relative to ChIPseq performed to map enhancers in both Eed WT and Eed KO mESC, we computed heat maps showing the normalized intensities of analyzed histone PTMs at distal H3K27ac peaks. Interestingly, this analysis unveiled that the activated enhancers in Eed KO cluster into two different groups, on the basis of H3K4me1 level as shown in figure 2.1.9G. As reported in the heat map, activated enhancers belonging to class I were pre marked by H3K4me1 in the Eed WT, while class II enhancers gained H3K4me1 contemporarily with H3K27ac deposition upon loss of PRC2 dependent H3K27me2 (figure 2.1.9G). Class I and class II enhancers accounted for approximately 60% and 40% of total, respectively.



Results of enhancer elements analyses represented as histone PTMs signal intensities heat map were further quantified by box plot analyses (figure 2.1.9H).

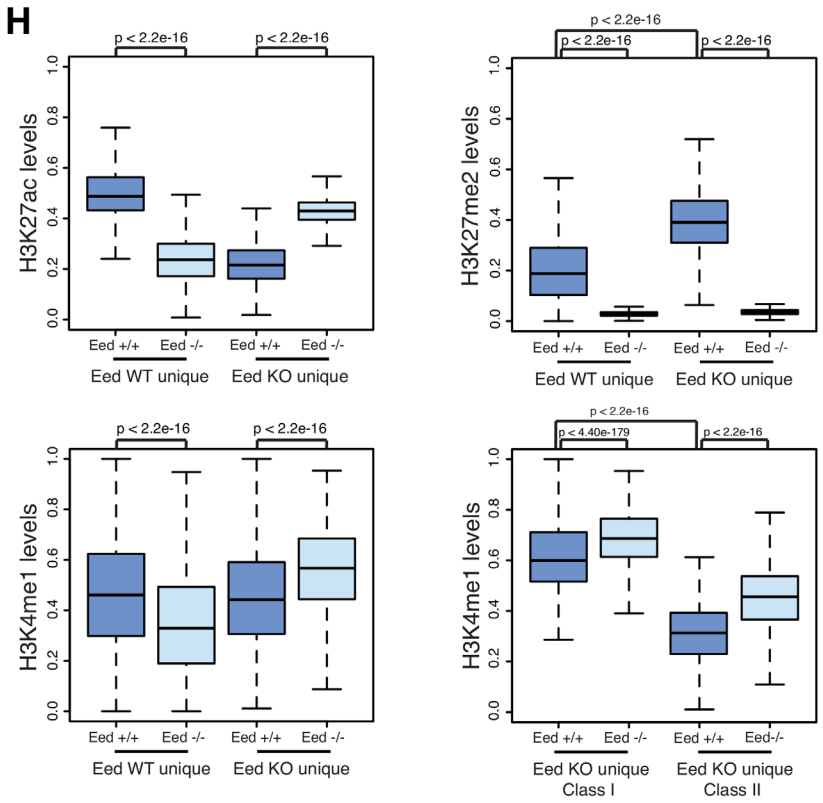


Fig.2.1.9 G, H. Disruption of proper enhancer regulation upon loss of H3K27me2 in mESC. (G) Heatmap of the normalized intensities of H3K27ac, H3K4me1, H3K4me3, H3K27me3, H3K27me2 in WT (+/+) and Eed KO (-/-) mESC for all H3K27ac peaks distal from a TSS uniquely found in either WT (Eed WT unique peaks) or Eed KO (Eed KO unique peaks). Classification into Class I (n = 4,391) and Class II (n = 4,819) was applied on the basis of pre-existence of H3K4me1 in Eed WT sample. Grouping was based on k mean clustering (k = 2) with respect to the H3K4me1 normalized intensities in Eed WT ESCs. (H) Boxplot analyses quantifying the data shown in figure 2.1.9G. p value was calculated by Wilcoxon rank test.

Importantly, in support of our protective model exerted by H3K27me2, we could observe that both class of enhancers gained H3K27me2 while losing H3K27ac levels; moreover, unique enhancer sites in Eed WT mESC showed a negative correlation between H3K27ac and H3K27me2 deposition (figure 2.1.9I).

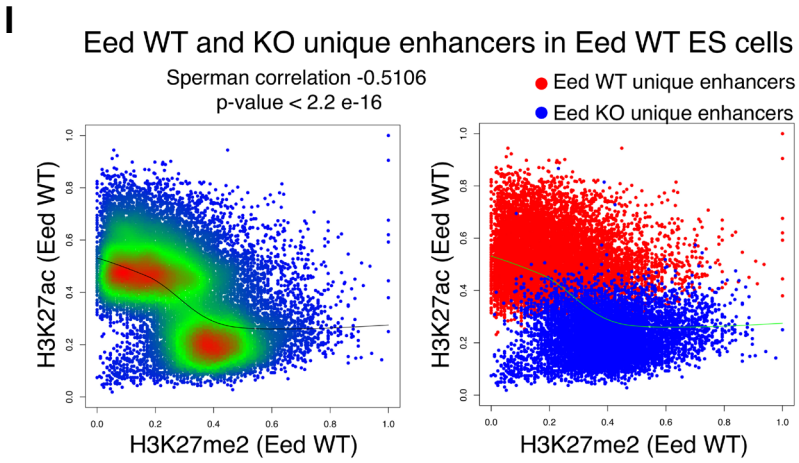


Fig.2.1.9 I. Anti-correlation of H3K27ac and H3K27me2 deposition at unique enhancers sites in mESC. Scatter plot correlation of the H3K27ac versus H3K27me2 levels in WT ESCs for unique enhancers identified in WT and Eed KO ESCs. Left panel shows whole density distributions. Right panel shows the modification levels of Eed WT unique (red) and Eed KO unique (blue) enhancers. The Spearman correlation value is indicated (rs = 0.5106). p value was calculated by asymptotic t approximation

As previous works described a class of H3K27me3 enriched enhancers in human ESC[264], we further tested H3K27me3 levels at unique enhancers sites, and showed that in both Eed WT unique and Eed KO unique sites, H3K27me3 was similarly present at very low levels (figure 2.1.9J). Consistently with this observation, we found that Eed KO unique enhancer sites had no Ezh2 association and did not overlap with CpGi, which are known to be PRC2 target and sites of H3K27me3 deposition (figure 2.1.9K)

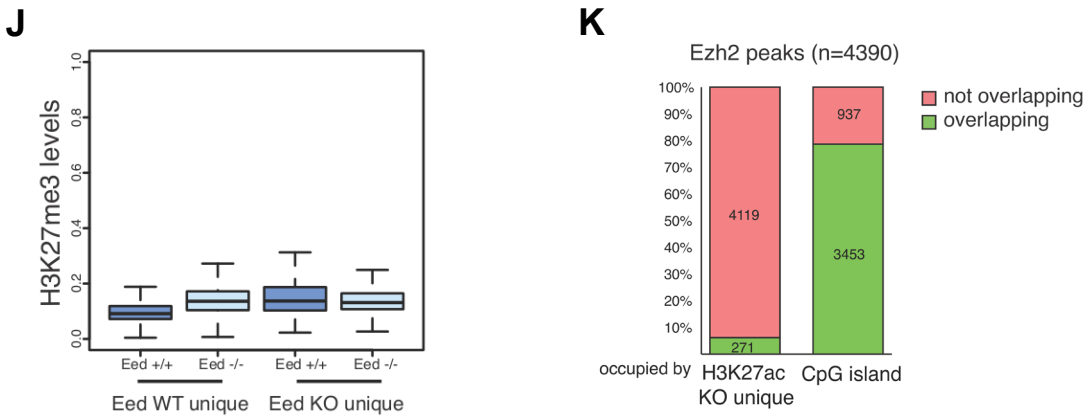


Fig.2.1.9 J, K. Unique enhancers sites in mESC are not enriched in H3K27me3 and do not reside on CpG islands. (J) Box plot showing the quantification of H3K27me3 signal in the unique H3K27ac distal peaks of Eed WT and Eed KO samples. (K) Percentage of Ezh2 peaks occupancy (determined by ChIPseq analysis in mouse E14 ES cells) and of CpG islands respect to genomic regions corresponding to H3K27ac peaks uniquely found in Eed KO mESC.

To have insights about the correlation between activated enhancers in Eed KO mESC and gene activation, we carried out bioinformatic analyses linking these enhancer sites with the closest activated genes in Eed KO mESC. As shown by the box plot in figure 2.1.9L, the distance between enhancers that acquired H3K27ac in Eed KO mESC and the closest upregulated genes was reduced respect to the distance observed analyzing enhancers that lost H3K27ac upon PRC2 depletion (Eed WT unique). This result did not depend on H3K27me3 levels deposition at TSS of analyzed genes in WT mESC (figure 2.1.9M). This observation further endorses our protective model by which H3K27me2 controls enhancer activation by preventing aberrant H3K27ac deposition at these regulatory elements.

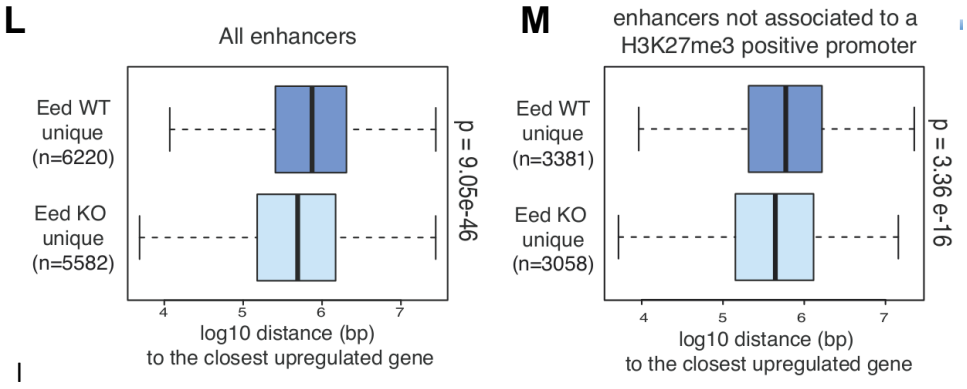


Fig.2.1.9 L, M. Activation of enhancers upon loss of PRC2 dependent H3K27me2 in mESC correlates with gene activation of the closest genes (L) Representation of the relative distance between putative WT and Eed KO enhancers and the up-regulated genes in Eed KO ES cells.All identified enhancers are included in the analysis. (M) The panel represents the same analysis in which the enhancers associated to a H3K27me3 positive gene in WT ES cells were excluded from the analysis. H3K27me3 enriched genes were defined by the presence of a H3K27me3 peak within +/- 2.5kb from the TSS. p-values were calculated by Mann-Whitney Test.

According to these data, we envisaged that the loss of H3K27ac at enhancers sites should be replaced by PRC2-dependent H3K27me2 deposition. Thus, as additional prove supporting our mechanism, we inhibited the enzymatic activity of histone acetyl transferase (HAT) Cbp and p300 treating WT mESC with the chemical compound C646 for 48 hours[266]. The levels of global H3K27ac were reduced respect

to control cells (figure 2.1.9N), and we performed ChIPseq analyses for H3K27me2 and H3K27ac in treated and control mESC to test the behavior of such histone PTMs. As depicted by the average profile plots and the relative quantification by box plots, we found the loss of H3K27ac at ca. 4800 enhancers sites upon C646 treatment, showing rapid deposition of H3K27me2 (figures 2.1.9O,P). This observation reinforces the proposed model, demonstrating that loss of H3K27ac from regulatory elements is followed by H3K27me2 deposition, defining H3K27me2 as a default state of these regions.

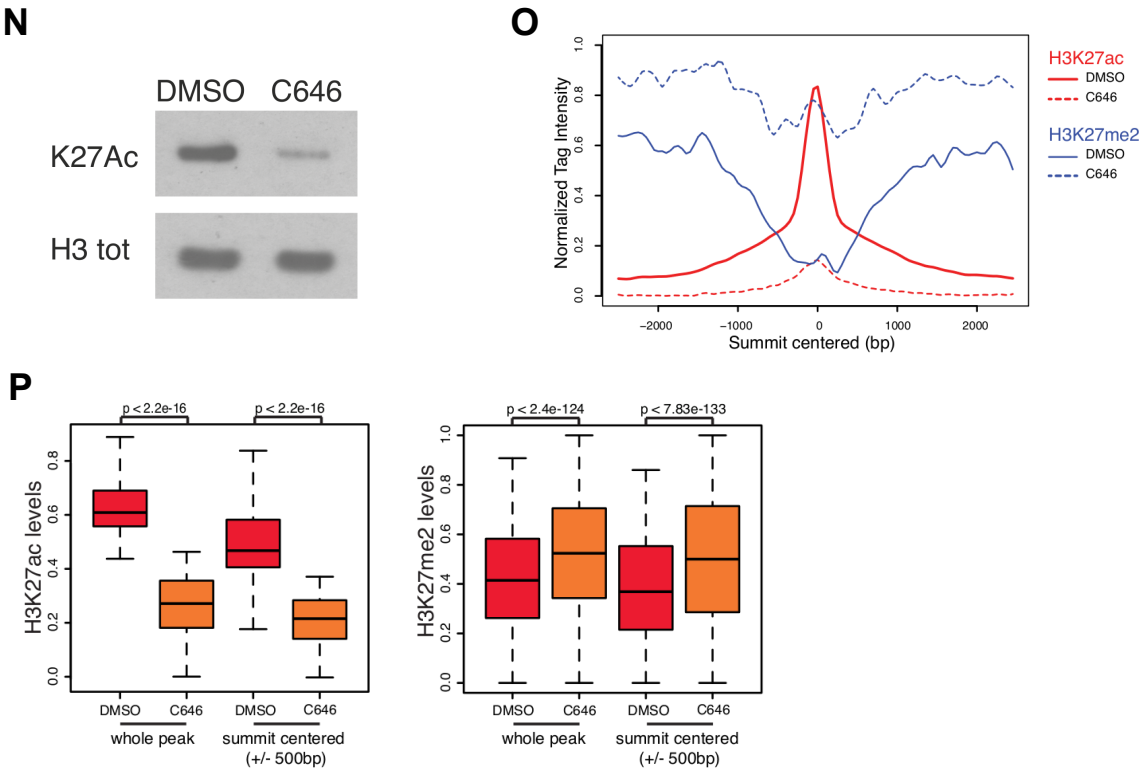


Fig.2.1.9 N-P.Loss of H3K27ac deposition at enhancer sites is replaced by H3K27me2 enrichment. (N) Immunoblot analysis for H3K27ac antibody of histones extracted from mouse E14 mESC treated with 35 µM C646 p300 inhibitor for 48 h. DMSO was used as vehicle control. Histone H3 was used as loading control. (O) Average profiles of H3K27ac and H3K27me2 deposition across a genomic window of +/- 2500 bp (centered to the H3K27ac peak summit) for active enhancers that lose H3K27ac deposition upon treatment with 35 µM C646 compound for 48 h (N=4838). DMSO was used as vehicle control. (P) Box plots analyzing the quantification of the H3K27ac and H3K27me2 levels at the same enhancer sites of figure 2.1.9O upon treatment with C646 for the whole H3K27ac peak region or for a 1kb genomic region surrounding the summit of each H3K27ac peak.

Overall, these data suggest that PRC2-dependent H3K27me2 ensures correct activation of enhancers through a mechanism that prevents aberrant H3K27ac deposition at these regulatory elements.

2.2. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells.

2.2.1. Identification of Ogt interacting partners in embryonic stem cells.

2.2.1.1. Generation and characterization of Ogt-tagged mESC lines.

In order to address the role of Ogt protein in the regulation of transcription in mESC; we thought the identification of Ogt interacting proteins within nuclei of mESC very worthy and interesting to pursue. To accomplish this goal, we took advantage of a system, already successfully used in our laboratory[146], which allowed us to generate mESC lines expressing tagged form of Ogt (Flag-Avi-Ogt, hereafter named bioOgt) in mESC constitutively expressing the bacterial enzyme BirA (birA-ES). Using this method we obtained a birA-ES clone in which bioOgt was efficiently biotinilated and expressed at sub physiological level (figure 2.2.1A). Ogt can localize in both cytoplasm and nucleus[267], thus we tested if the bioOgt localization resembled its endogenous counterpart in the different cellular fractions. As shown by the immunoblot analysis of hypotonic extracted cellular fractions, bioOgt had the same localization respect to the endogenous Ogt (figure 2.2.1B). We then tested the Ogt extractability by lysing at different salt concentrations the bioOgt expressing mESC; we obtained the maximum Ogt solubility from chromatin at 300mM NaCl concentration (figure 2.2.1C). We applied this lysis condition to all the following experiments, including the purification of the Ogt complex to identify its interacting partners.

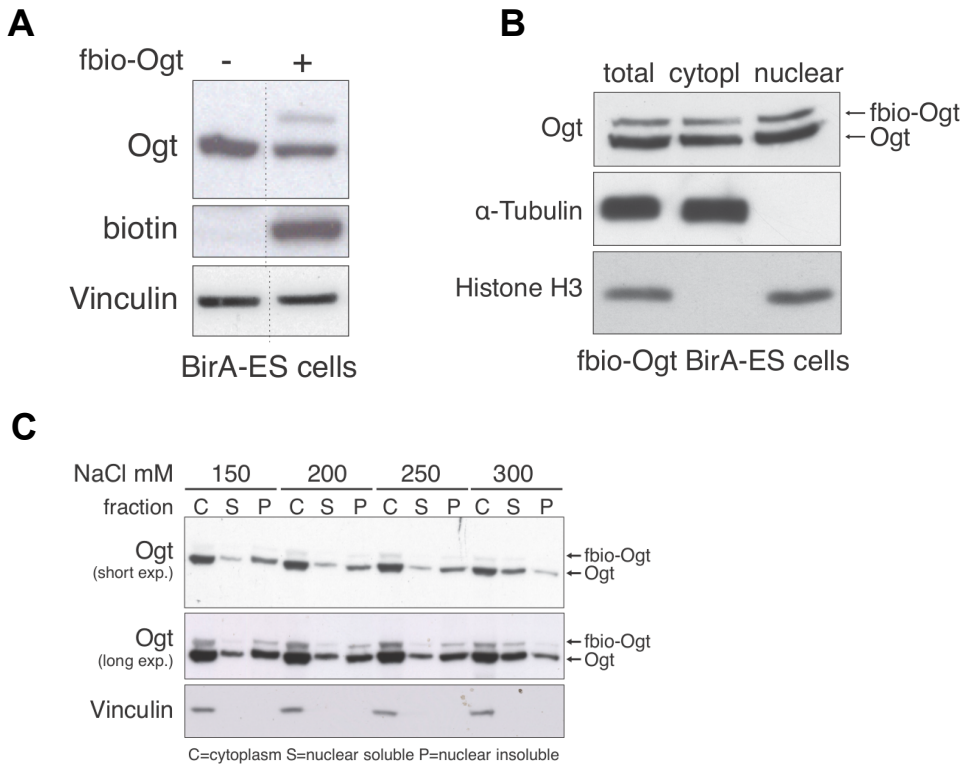
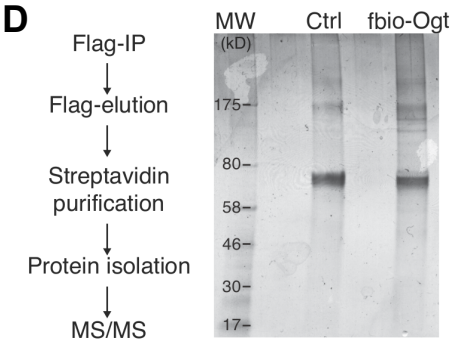


Fig.2.2.1 A-C. Generation and characterization of Ogt tagged mESC. (A) Immunoblot analysis with the indicated antibodies of protein extracts obtained from BirA ES cells expressing (+) or not (–) bioOgt. Vinculin was used as loading control. (B) Western blot (WB) of cytoplasm versus nuclear fractionation with the indicated antibodies in bioOgt-expressing BirA ES cells. a-tubulin and histone H3 are presented as fractionation controls. (C) WB analyses using the indicated antibodies of nuclear lysates of bioOgt-expressing BirA ES cells using the reported NaCl concentrations to solubilize nuclear proteins. The figure shows cytoplasmic (C), nuclear soluble (S) and nuclear insoluble (P) fractions.

2.2.1.2. Assessing the Ogt interactome by mass spectrometry analysis: Tet proteins as major Ogt interactors in mESC.

To identify Ogt interactors, we purified nuclei from BirA and BirA-Ogt cells which were subjected to mass spectrometry analysis (MS/MS) after flag and streptavidin sequential immunoprecipitations (figure 2.2.1D). Eluted proteins were separated and silver stained on a polyacrilamide gel under denaturing conditions, highlighting the enrichment of several protein in the fbio-Ogt purification respect to its related control (figure 2.2.1D). Mass spectrometry analyses led to the identification of known previously identified Ogt partners such as Sin3a and Hcfc1 proteins[213,268] (figure

2.2.1E). Interestingly we found, among the most abundant identified interactors, two members of the ten-eleven translocation (Tet) protein, Tet1 and Tet2 (figure 2.2.1E). Notably, Sin3a protein was found to interact with Tet1 and Hdac1[229,269] , suggesting the existence of a large multiprotein complex. In order to validate the MS output, we carried out specific streptavidin immunoprecipitations in fbioOgt BirA mESC which confirmed the interactions observed by MS analysis (figure 2.2.1F).



E

	number of peptides			
Protein	fbio-OGT	fbio-Ctrl	PEP	iBAQ
Ogt	289	2	0.0	49412000
Hcfc1	145	0	0.0	4164900
Tet1	182	3	0.0	2907400
Tet2	120	0	0.0	1119100
Sin3a	16	0	4.0E-85	93391
Hdac1	4	0	6.0E-05	57279

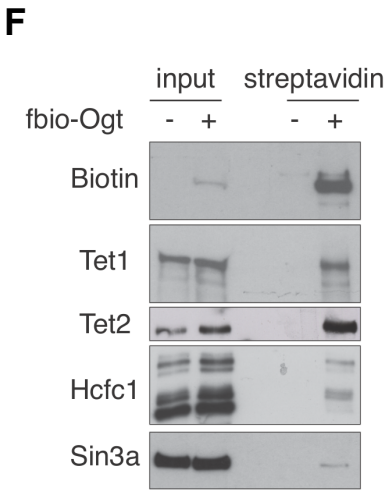


Fig.2.2.1 D-F. Identification and validation of Ogt interactome within mESC nuclei by mass spectrometry analysis. (D) Scheme of the purification procedure and silver-staining of 10% of the eluted proteins from tandem purifications with control or bioOgt-expressing BirA ES cells. (E) Table showing the most relevant proteins identified by mass spectrometry analysis and the number of peptides detected in both control or bioOgt purifications together with the posterior probability error (PEP) and the intensity based absolute quantification score (iBAQ) determined with MaxQuant. (F) WB using the indicated antibodies of streptavidin purifications from protein extracts derived from fbio-Ogt expressing (+) or not expressing (-) BirA mESC.

In addition, we validated these interactions at endogenous level, performing immunoprecipitations with specific antibodies for Ogt, Tet1 and Tet2 in nuclear extracts from WT mESC. Moreover, Tet1 and Tet2 were able to co-immunoprecipitate endogenous Ogt, reinforcing the observed interactions (figure 2.2.1G). Remarkably, Tet1 IPs performed on nuclear extracts prepared from mESC depleted of Ogt

expression by transfection of endoribonuclease-prepared siRNAs (esiRNAs) showed loss of Tet1-Tet2 interaction upon depletion of Ogt expression (figure 2.2.1H).

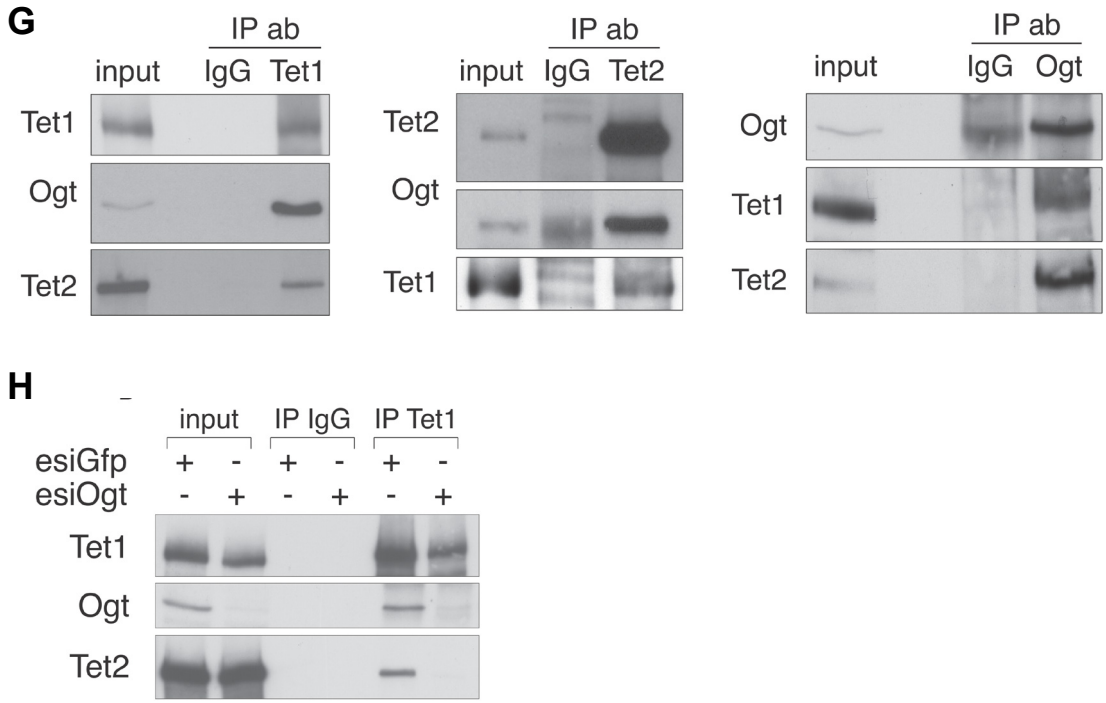
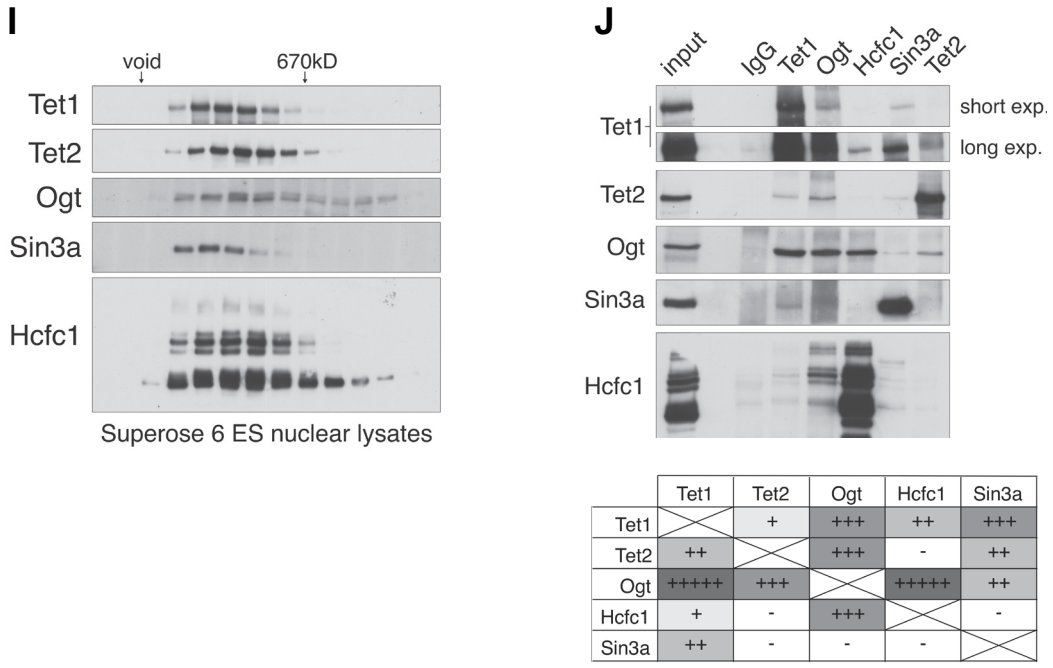


Fig.2.2.1 G,H. Endogenous Ogt interacts with Tet1 and Tet2. (G) WB with the indicated antibodies of immunoprecipitated (IP) proteins from nuclear E14 ES cell lysates with Tet1 (left panel), Tet2- (middle panel), and Ogt (right panel) specific antibodies. IgG served as unrelated IP control.t (–) bioOgt. (H) WB with the indicated antibodies of immunoprecipitated proteins with a Tet1-specific antibody from protein lysates of cells transfected for 48 hr with Ogt- and Gfp-(control) specific esiRNAs.

The interactions observed among Hcfc1, Tet proteins and Ogt, made us to envisage that these proteins are part of a multiprotein complex. To prove this we performed size exclusion chromatography on mESC nuclear lysates, and the immunoblot analyses on eluted fractions revealed that Tet1, Tet2, Ogt and Hcfc1 co-eluted at high molecular weight, suggesting the existence of a large multiprotein complex (figure 2.2.1I). To assess the characteristic of this complex, we carried out co-IP analyses on mESC nuclear extracts using antibodies for Tet1, Tet2, Ogt, Hcfc1, and Sin3a (figure 2.2.1J): This result strongly suggested that Tet1 and Hcfc1 are part of two distinct Ogt containing complexes that can still be in weak contact with each other.



	Tet1	Tet2	Ogt	Hcfc1	Sin3a
Tet1		+	+++	++	+++
Tet2	++		+++	-	++
Ogt	+++++	+++		+++++	++
Hcfc1	+	-	+++		-
Sin3a	++	-	-	-	

Fig.2.2.1 I,J. Ogt forms distinct multiprotein complexes with tet proteins and Hcfc1. (I) WBs with the indicated antibodies of different fractions of E14 ES cell nuclear lysates separated on a Superose 6 column by size-exclusion chromatography. (J) upper panel: WB with the indicated antibodies of immunoprecipitated proteins from nuclear E14 ES cell lysates with Tet1-, Ogt-, Hcfc1-, Sin3a-, and Tet2-specific antibodies. IgG served as unrelated IP control; lower panel: table summarizing the strength of interaction in the IPs presented by WBs.

Further, in light of the glycosiltransferase property of Ogt, using a wheat germ agglutinin (WGA) agarose resin which allows the isolation of glycoproteins under denaturing conditions, we were able to show that Tet1, Tet2, and Sin3a are enriched on the WGA resin, as well as Hcfc1 and Ogt itself (two known Ogt substrates). This suggests that Ogt binding to Tet1 and Tet2 induces their O-GlcNAcylation (Figures 2.2.1K). Consistently with this, proteomic studies found that Tet1 was one of the nuclear O-GlcNAcylated proteins in mESC[270]. Taken together, these results demonstrate that Tet1 and Tet2 are physiological stable partners of Ogt in the nucleus of mESC and that they are targets of its enzymatic activity.

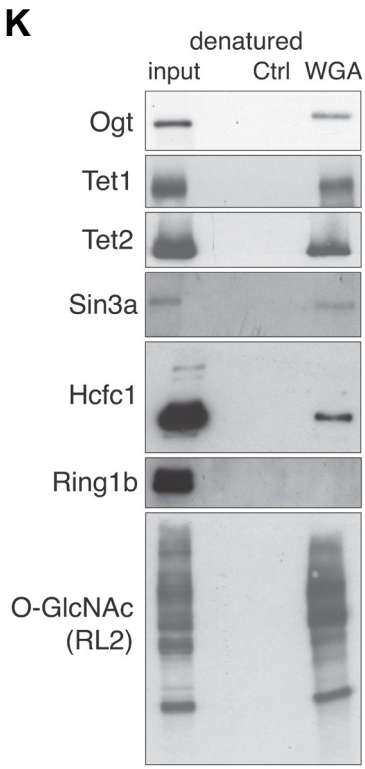


Fig.2.2.1 K. Ogt O-glycosilates its interacting proteins.. WB with the indicated antibodies of purified proteins with WGA agarose resin from the denatured E14 ES cells protein lysates. A non-conjugated agarose resin served as purification control (ctrl).

2.2.2. Analysis of genome-wide Ogt localization.

We wanted to assess whether Ogt had chromatin binding capacity in mESC, thus we performed high throughput DNA sequencing (ChIP-seq) analysis using specific antibody to unravel genome wide Ogt localization to chromatin. This allowed us to identify 11552 Ogt binding sites, which enriched at promoter, intron and exon regions (figure 2.2.2A). Notably, 62% of these binding sites coincided with transcription start sites (TSS) of annotated RefSeq genes with TSS defined as 5 kilobase region centered over TSS of genes (figure 2.2.2B). This was further confirmed by the observed distribution of Ogt peaks over a 20 kb window around TSSs of RefSeq genes (figure 2.2.2C). According to these data, Ogt preferentially binds to promoter regions, precisely in correspondence of their TSS.

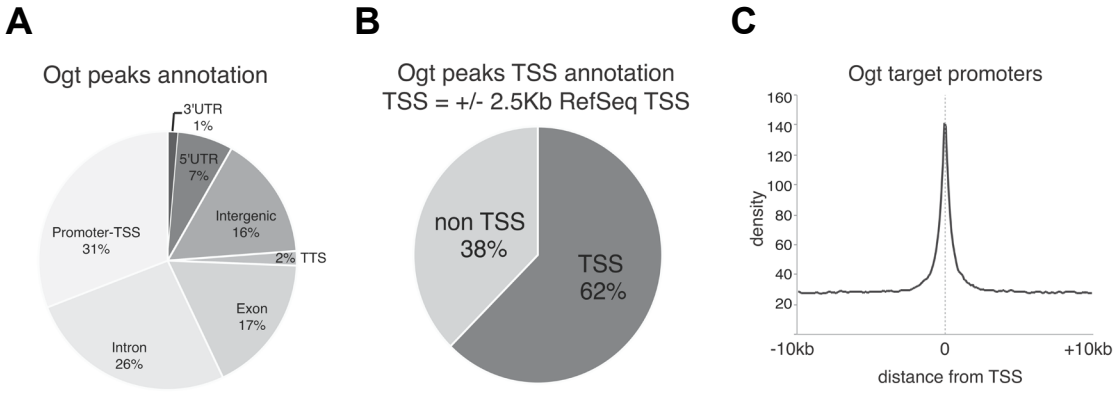


Fig. 2.2.2 A-C. Ogt binds chromatin in correspondence of TSS regions of target genes. (A) Ogt peak annotation relative to the indicated genomic feature. Ogt ChIPseq was performed in E14 ES cells, and peak call was generated by comparison with an unrelated ChIP-seq experiment with rabbit IgG. (B) Ogt peak annotation relative to genomic regions included within ± 2 kb from RefSeq TSSs. (C) Density distribution of the Ogt ChIPseq signals over a window of 20 Kb centered on RefSeq gene TSSs.

2.2.3. Ogt controls transcription of metabolic genes in mESC.

The genome-wide association of Ogt to gene promoters made us to envisage a possible role for Ogt in transcriptional regulation. At this purpose we took advantage of esiRNA to deplete Ogt expression in mESC. A transient knock down of the protein was performed to avoid lethal effects, given the importance of Ogt activity for mESC viability[271]. This allowed us to downregulate Ogt expression and global levels of related enzymatic activity (O-GlcNAcylation). Upon knock down conditions, neither the expression of Tet proteins expression, nor of pluripotency markers Oct4 and Nanog were affected (figure 2.2.3A).

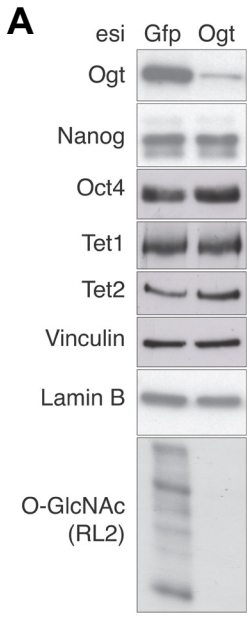


Fig.2.2.3A. Ogt down-regulation upon esiRNA delivery. WB with the indicated antibodies from protein extracts obtained from E14 ES cells transfected with Gfp- or Ogt-specific esiRNAs. Vinculin and lamin B served as loading controls.

To have insights about transcriptional changes upon Ogt depletion, we performed microarray expression analysis, which led to the identification of 1056 regulated genes. A similar number of genes were both upregulated or downregulated upon Ogt loss (figure 2.2.3B). In independent experiments we were able to validate microarray results on a set of genes assayed by RT-qPCR analysis (figure 2.2.3C). Regulated genes were subjected to functional annotation taking advantage of Ingenuity pathway analysis (IPA) tool, highlighting a strong increase in the expression of genes involved in glycan biosinthesis and glycerolipids methabolism (figure 2.2.3D). Genes that became overexpressed upon Ogt deletion were mainly involved in metabolic signaling pathways and protein biosynthesis (figure 2.2.3D). Further, we analyzed Ogt occupancy at promoters of regulated genes, showing that a significant fraction of these genes retained Ogt association at relative TSSs (figure 2.2.3E). However, a significant part of these regulated genes are not bound by Ogt, indicating an indirect regulation which is potentially related to broad Ogt functions (figure 2..2.3E). Altogether, these evidences underline the pleiotropic functions exerted by Ogt within mESC, as shown by the transcriptional changes mediated by the Ogt depletion and

consistently with the strict link between Ogt activity and cellular metabolism. This also agreed with previous studies conducted on *D. melanogaster* and *C.elegans* showing a role for Ogt in gene expression control[221,272].

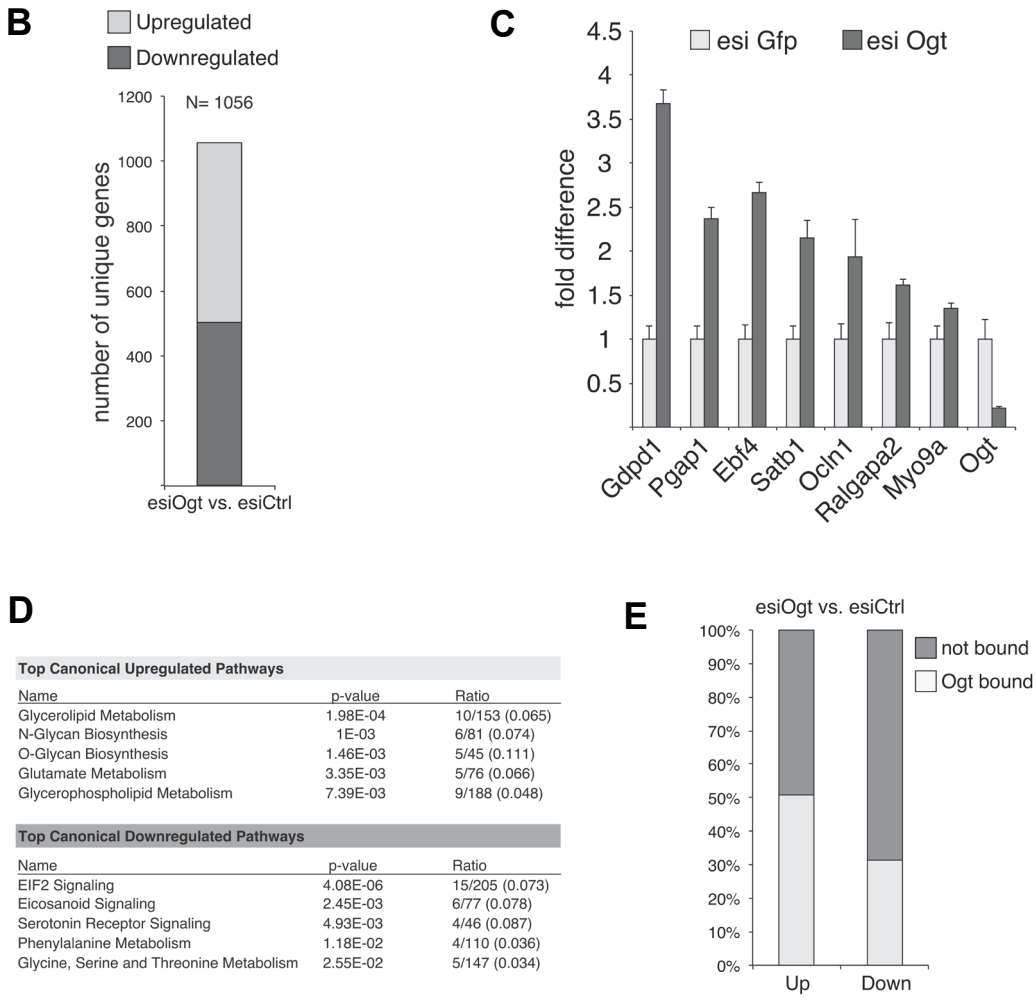
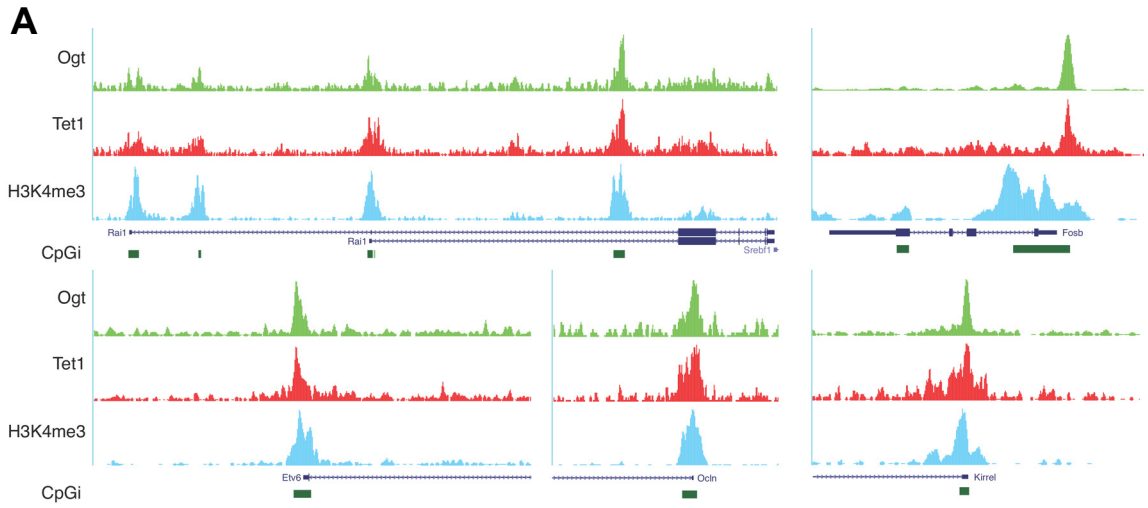


Fig.2.2.3 B-E. Ogt regulates expression of metabolic genes in mESC. (B) Gene expression regulation in two independent Ogt-interference experiments displaying the total number of genes up- or down-regulated (R1.3-fold, 1056 genes in total) (B) Gene Ontology analysis performed with IPA of up-regulated (top panel) and down-regulated (bottom panel) genes identified by the analyses presented in (B). (D) Annotation of the Ogt peaks identified in the ChIP-seq analyses relative to the promoters of the up-regulated (up) or down-regulated (down) genes.

2.2.4. Ogt and Tet1 proteins bind chromatin in mESC.

In light of the biochemical interaction between Ogt and Tet proteins, and once assessed Ogt binding to chromatin in proximity of TSS regions, we also expected a genome wide colocalization between Ogt and Tet activity. To understand this issue,

we made bioinformatic analyses to overlap our Ogt ChIP-seq data with available ChIP-seq datasets for Tet1 and H3K4me3 which were performed in the same mESC line[229,257]. As shown by genome wide snapshot from UCSC genome browser, Ogt and Tet1 binding profiles share a perfect overlap at H3K4me3 positive promoters (figure 2.2.4A). More in detail, overlap analyses showed that more than 90% of Ogt binding sites colocalize with Tet1 ones; 70% of overlap is observed if also H3K4me3 positive regions were taken into account (figure 2.2.4B). In order to validate colocalization data, we carried out RT-qPCR analysis on ChIP made using Tet1 antibody and two independent Ogt antibodies (figure 2.2.4C). These overlap analyses were additionally made using Tet1 genome wide occupancy dataset generated with an "in house" made antibody (Tet1-C). The extend of overlap between Ogt and Tet1 peaks was even broader than the previous analysis, showing 98% of overlap (figure 2.2.4D). To further characterize the Ogt-Tet1 colocalization, we measured the coverage between Ogt and Tet1 peaks denoting a high degree of overlay (80%–100%) that is consistent with their physical interaction (figure 2.2.4E).



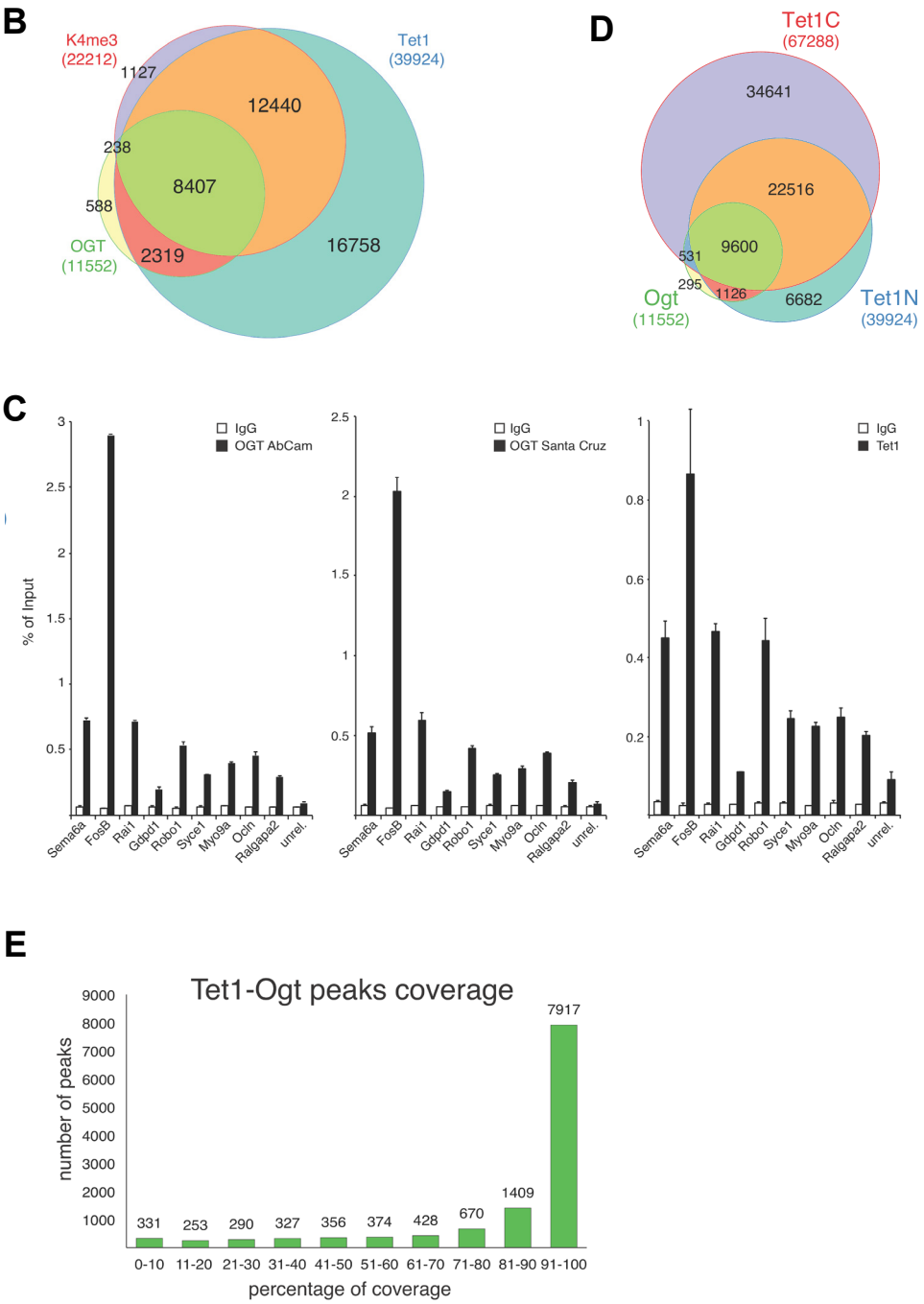


Fig 2.2.4 A-E. Genome-wide Ogt and Tet1 colocalization. (A) Genomic snapshots from UCSC genome browser for Ogt, Tet1, and H3K4me3 ChIP-seq analyses. (B) Overlap between the genomic positions of the peaks identified in Ogt, Tet1, and H3K4me3 ChIP-seq analyses. (C) Validation of ChIP results by RT-qPCR with specific primers for the indicated antibodies. Data are represented as mean ± SEM. (D) Overlap among the genomic positions of the peaks identified in ChIPseq assays using Ogt, and two distinct Tet1 antibodies (Tet1C and Tet1N). (E) Coverage between Ogt and Tet1 peaks expressed in percentage over total identified Ogt peaks,

Ogt and Tet1 unique binding sites identified in the analysis presented in figure 2.2.4B showed signal accumulation of Tet1 and Ogt, respectively (figure 2.2.4F), thus strengthening the overlay between the two protein at chromatin, which was further confirmed by RT qPRC analyses on independent ChIP experiments (figure 2.2.4G).

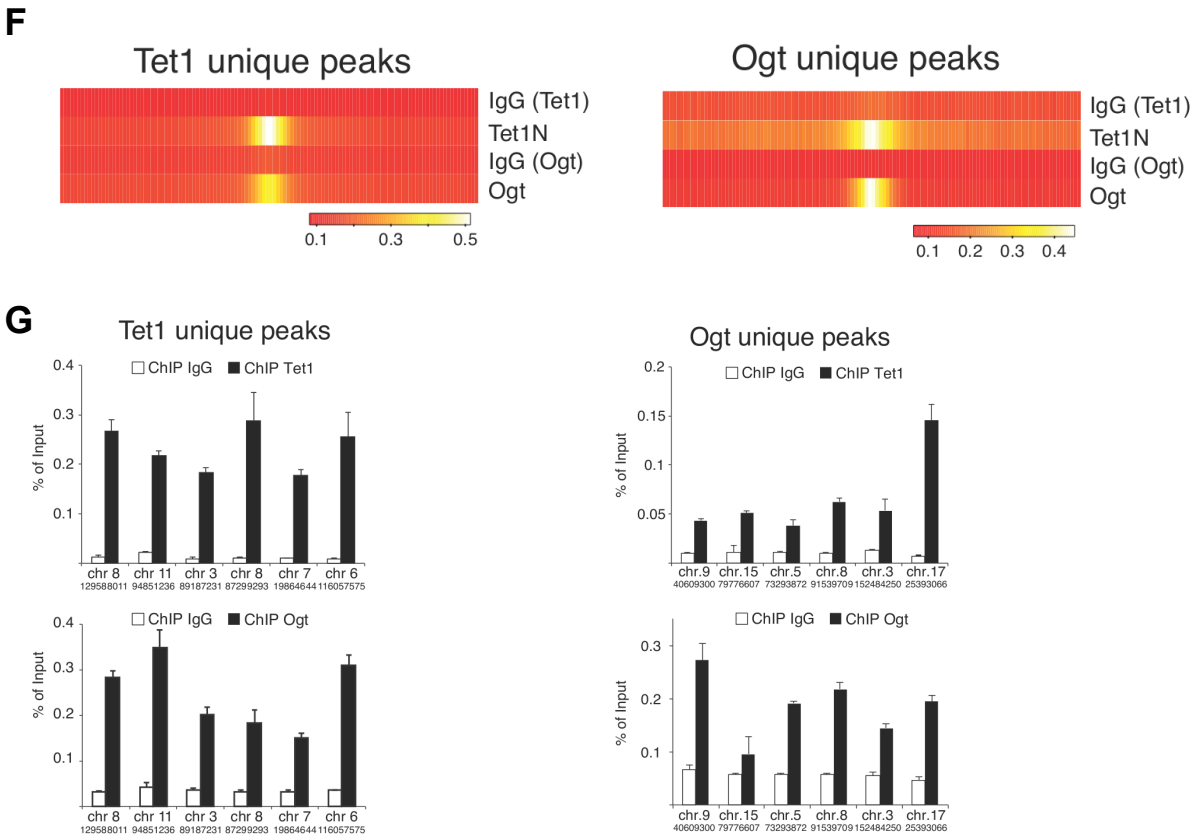


Fig.2.2.4 F, G. Occupancy of Tet1 and Ogt at their unique genomic regions. (F) Heat map of the indicated ChIPseq signal densities in a window of +/-3kb centered on the summit of Tet1 unique peaks (left) and Ogt unique peaks (right). Data are related to those presented in figure 2.2.4B (G) RT qPCR analyses with indicated primers specific to indicated unique region (left panels show Tet1 unique regions; right panels show ogt unique regions) for ChIP performed with the indicated antibodies. Genomic coordinates (centered on primers annealing loci) of unique regions are indicated. IgG were used as negative control.

Moreover, loss of function experiments allowed us to strengthen these observation: as shown in figure 2.2.4H, depletion of Tet1 led to the loss of Ogt occupancy at both Ogt unique and Tet1 unique binding sites.

H

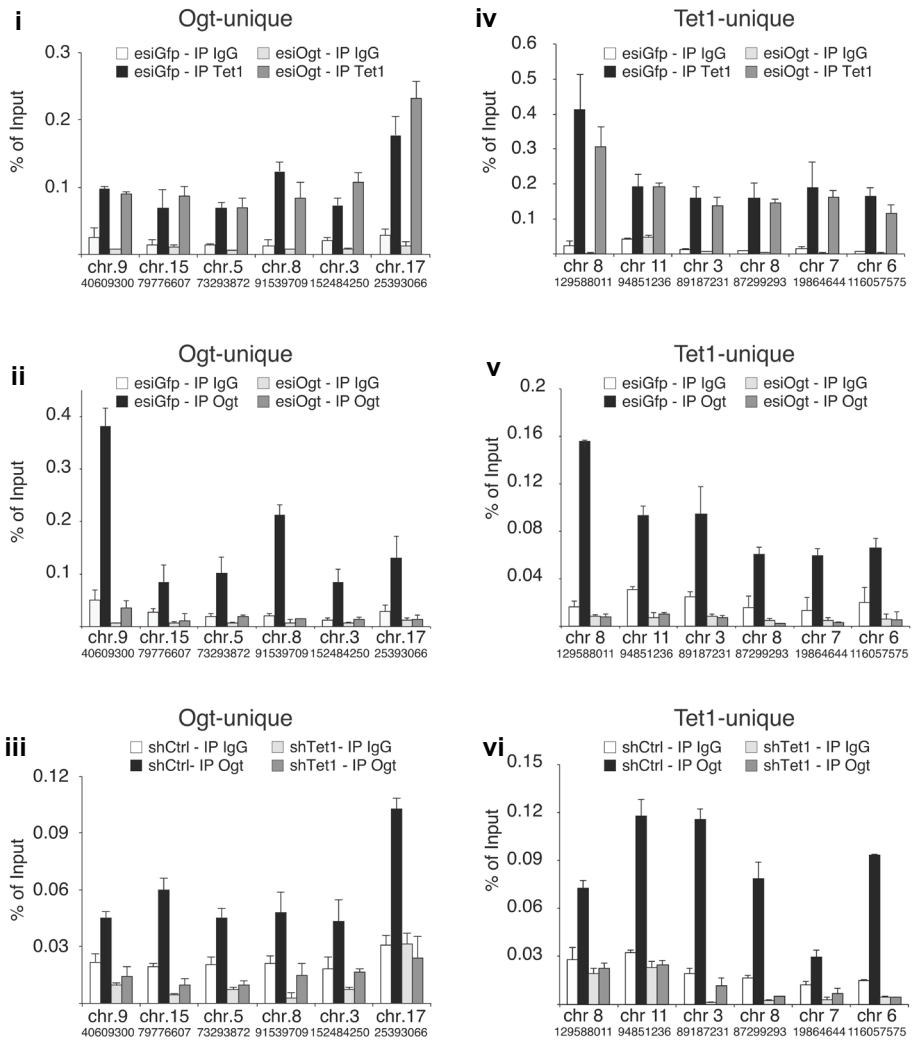


Fig.2.2.4 H. Specificity of Ogt and Tet1 signals at unique regions. Real-time quantitative PCR analyses using primers for the indicated genomic positions (coordinates indicate the central point of PCR amplification) corresponding to Ogt-unique (i-iii) and Tet1-unique (iv, vi) regions using DNA purified from ChIP assays performed using Tet1 (i, iv) and Ogt (ii, iii, v, vi) specific antibodies in E14 ES cells that stably express scrambled (Ctrl) or Tet1 specific shRNAs (iii, vi) or from E14 ES cells transfected for 48h with Ogt and Gfp specific esiRNAs (i, ii, iv, v). IgG are presented as a negative control for all ChIP assays. Data are represented as mean +/- SEM.

Finally, in agreement with our biochemical interaction data, Ogt also showed a broad overlap with Sin3a binding sites as demonstrated by genome wide ChIP-seq analysis (figure 2.2.4I) and relative independent validations (figure 2.2.4K). Specificity of Sin3a signal was proven performing specific knock down with shRNA transduction of

mESC (figure 2.2.4J). Overall, these data indicate an extensive genome-wide co-localization between Ogt and Tet1, particularly at promoter regions also enriched for H3K4me3.

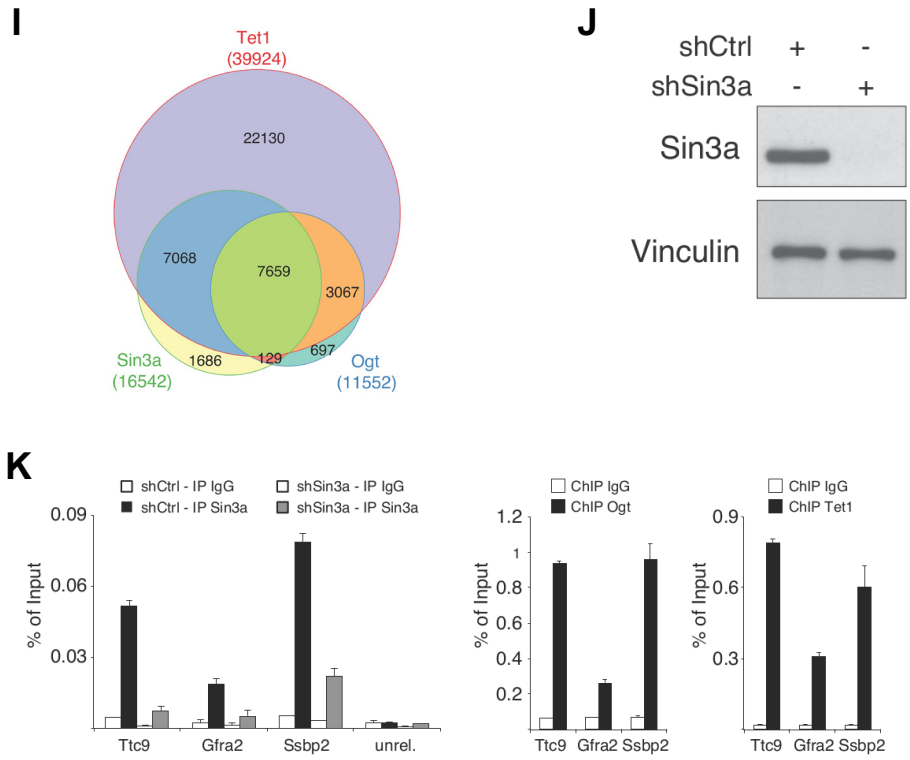


Fig.2.2.4 I-K. Ogt, Sin3a and Tet1 genome-wide overlap. (I) Venn diagram depicting the extend of overlap among genome-wide binding sites relative to the indicated chIPseq datasets. (J) WB analyses using the indicated antibodies of E14 ES cells extracts that stably express Sin3a specific shRNAs. A scrambled shRNA sequence was used as negative control (shCtrl). Vinculin served as fractionation control. (K) Real-time quantitative PCR analyses using primers for the indicated gene promoters using DNA purified from ChIP assays. ChIPs were performed using Sin3a, Ogt, Tet1 specific antibodies in E14 ES cells as in (J). IgG were used as a negative control for ChIP. Data are represented as mean ± SEM.

2.2.5. Characterization of genome-wide colocalization between Ogt and Tet1 proteins in mESC.

Simultaneous binding of Tet1 and Ogt to chromatin mainly occurs at promoter regions (Figure 2.2.4A); to strengthen this observation we clustered Tet1 positive regions with both TSS and not TSS loci which were positive (+Ogt) or negative (-Ogt) for Ogt occupancy, and *vice-versa*. This analysis showed that Tet1-Ogt co occupancy

occurs preferentially at TSSs rather than at distal regions (figure 2.2.5A). A similar result was obtained when, instead of TSSs, CpG islands (CpGi) were analyzed (figure 2.2.5B).

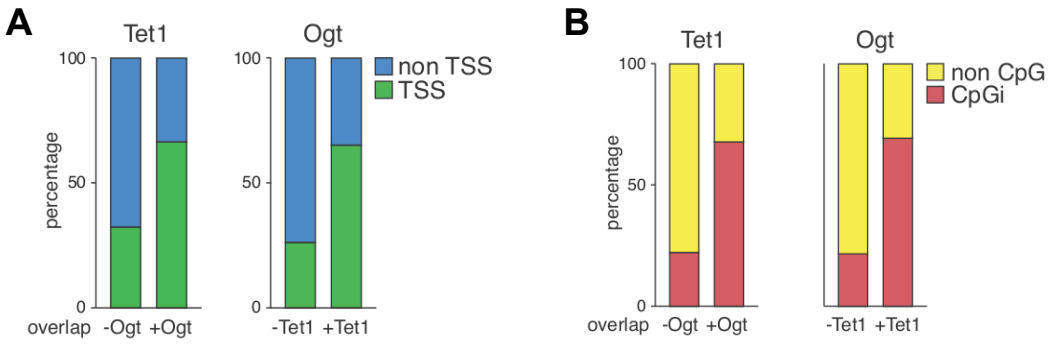


Fig.2.2.5 A, B. Ogt and Tet1 colocalization at promoter regions and CpGi. (A) Left panel: annotation relative to RefSeq gene TSSs of Tet1 peaks that colocalize (+Ogt) or do not colocalize (–Ogt) with Ogt. Right panel: annotation relative to the RefSeq gene TSSs of Ogt peaks that colocalize (+Tet1) or do not colocalize (–Tet1) with Tet1. (B) Annotations as in (A) relative to CpG islands (CpGi).

Tipical features of CpGi is their localization at promoter regions and that they are devoid of methylation due to a mechanism depending on the inhibition of *de novo* DNA methyltransferases controlled by H3K4me3 deposition[273]. We analyzed the correlation between Tet1 occupancy and CpGi distribution, showing that Tet1 binding at CpG rich promoters correlate with lack of 5-hydroxy methyl cytosine (5hmC) modification, while Tet1 binding at distal regions correlates with 5hmC deposition (figure 2.2.5C-D). Ogt signal intensity increased simultaneously with CpG density, correlating with loss of 5hmC and 5mC (figure 2.2.5C-D). Moreover we took into account DNA methylation levels comparing them to Ogt, Tet1 and CpG enrichments. On the basis of available single base methylome dataset[235] we identified genomic regions with low levels of DNA methylation (~ 30%) which were characterized as regulatory elements with poor CpG content (LMRs)[235]. These LMR regions show 5hmC and Tet1 binding and low level of Ogt (figure2.2.5E), consistently with observation made in figure 2.2.5C. Unmethylated regions (UMRs) negatively

correlated with 5hmC enrichment, while showing Ogt binding (figure 2.2.5E). In summary these data highlight the Ogt association to CpG rich Tet1 binding sites devoid of cytosine modifications on DNA.

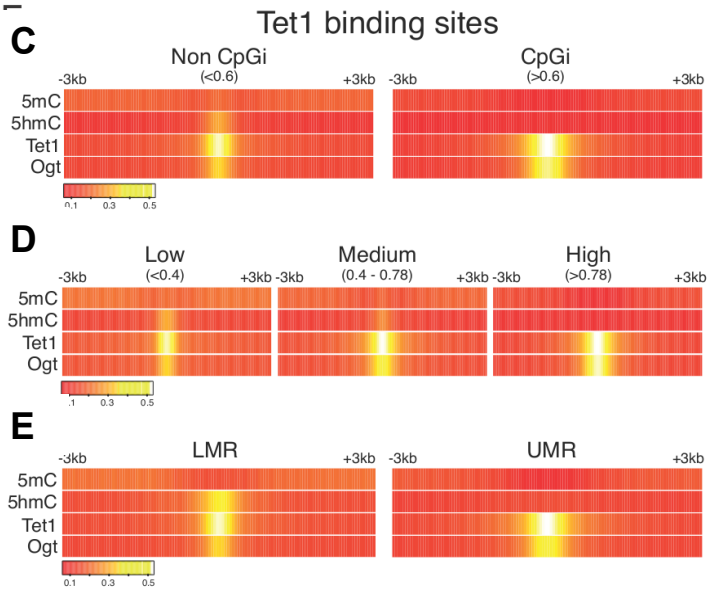


Fig.2.2.5 C-E. Genome wide Ogt – Tet1 colocalization negatively correlates with 5hmC deposition. (C) Heat map of the indicated ChIP-seq signal densities in a window of ± 3 kb centered on the summit of Tet1 peaks respect to CpGi. (D) As in (C), considering Tet1 binding sites relative to their indicated CpG densities. (E) As in (D), considering Tet1 binding sites annotated relative to low-methylated regions (LMR) and unmethylated regions (UMR).

2.2.6. Tet1 recruits Ogt to promoter regions in mESC.

From the ChIP data shown in figure2.2.4H, it is evident that Ogt binding to chromatin is significantly decreased in absence of Tet1: this suggests that Tet1 could mediate Ogt association to DNA. To prove this, we performed loss of function experiments by depleting Tet1 expression with shRNAs in mESC. A recent work reported a partial locus specific reduction of 5hmC levels[229], thus we envisaged a compensatory effect mediated by Tet2. To avoid this we also performed depletion of Tet2 protein in mESC previously knocked down for Tet1 (figure 2.2.6A), and we used these cell lines to perform ChIP analyses. ChIP analysis for Tet2 binding to chromatin were unsuccessful due to antibody inefficiency in ChIP assay or to its instable association to chromatin. We extracted proteins from chromatin fraction of Tet1-Tet2 KD mESC, and immunoblot analyses highlighted a significant reduction of Ogt association to chromatin (figure 2.2.6A). Further evidences came from immunofluorescence (IF)

staining performed in the same mESC lines, which showed a de-localization of Ogt from nuclei to cytoplasm (figure 2.2.6C). The specifity of the Ogt antibody in immunofluorescence (IF) procedure is shown in figure 2.2.6B, using Ogt depleted mESC as negative control.

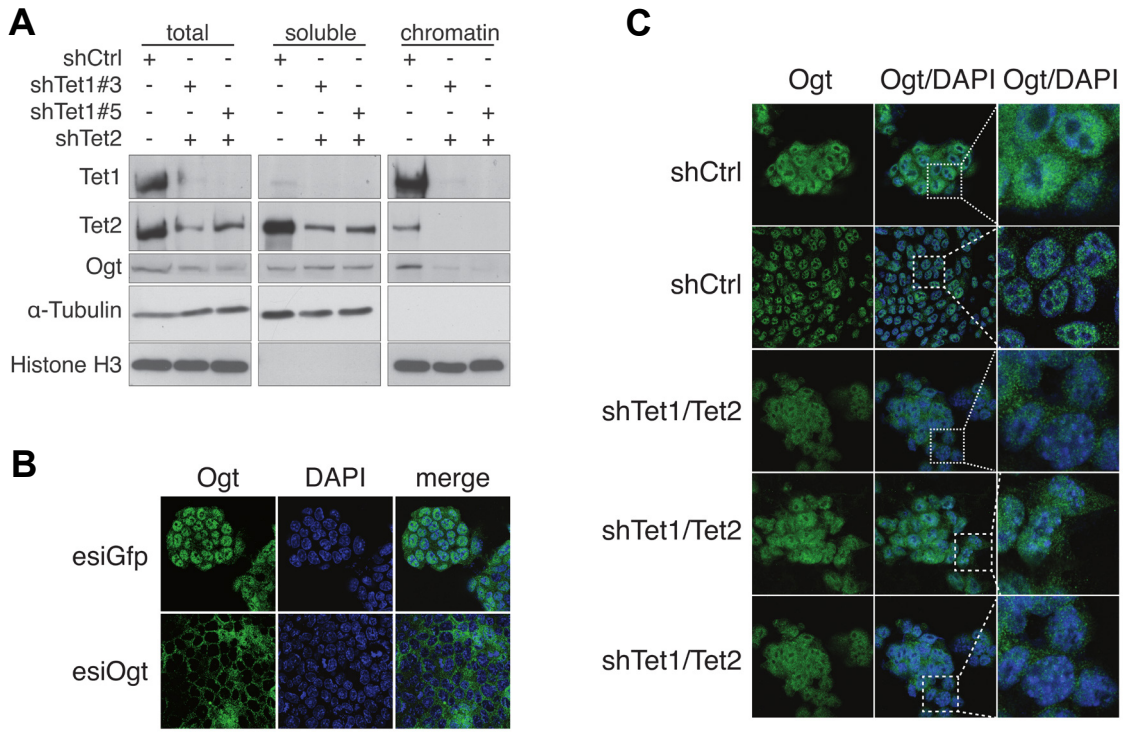


Fig.2.2.6 A-C. Ogt delocalization from nuclei upon Tet proteins depletion. (A) WBs with the indicated antibodies of E14 ES cell-soluble and chromatin-associated proteins that stably express the indicated shRNAs. A scrambled shRNA sequence is used as negative control (shCtrl). a-tubulin and histone H3 are presented as fractionation controls. (B) Confocal pictures of Immuno-fluorescent staining with Ogt specific antibodies in ES cells expressing Ogt and Gfp specific esiRNAs. (C) Confocal pictures of Immuno-fluorescent staining with Ogt specificantibodies in ES cells expressing the indicated shRNAs.

Fractionation assay also revealed that Tet2 enriched in the soluble fraction, thus explaining our failure in performing Tet2 ChIP and suggesting a dispensable role in recruiting Ogt to chromatin. Indeed, Tet2 depletion in mESC led to unaltered Tet1 and Ogt association to chromatin (figure 2.2.6D). To prove that Ogt binding to chromatin is dependent only from Tet1, we carried out fractionation assays in two mESC

independently transduced with two shRNA specific for Tet1, showing that Tet1 depletion led to Ogt displacement from chromatin (figure 2.2.6E); in addition, a minor degree of Tet2-chromatin association was observed, suggesting a possible role for Ogt-Tet2 interaction in the stabilization of Tet2 binding to chromatin (figure 2.2.6E).

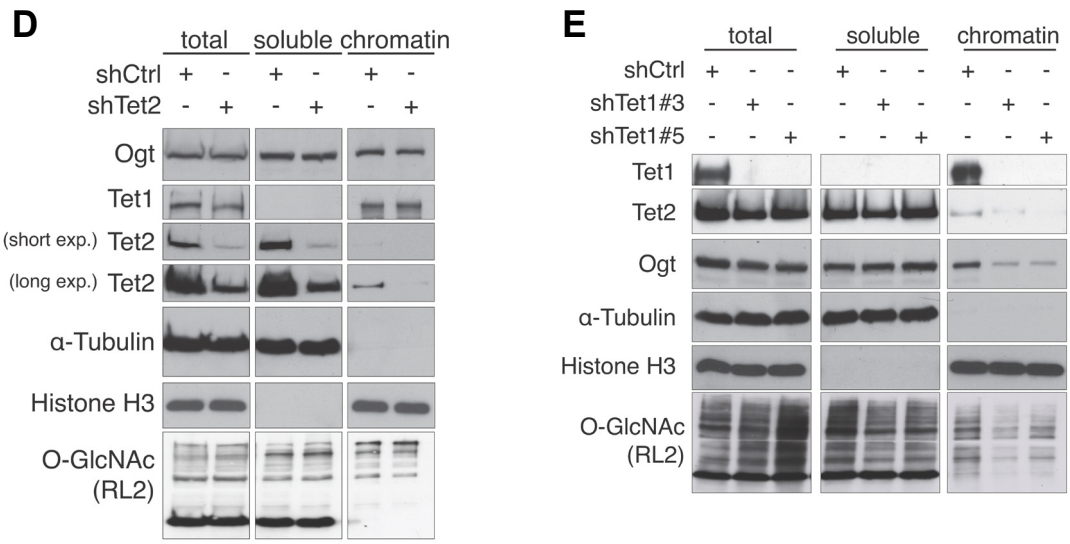


Fig.2.2.6 D, E. Ogt localization to chromatin depends on Tet1. (D) WBs as in figure 2.2.6A with E14 ES cells stably expressing scrambled or Tet2-specific shRNAs. (C) WBs as in figure 2.2.6A with E14 ES cells stably expressing scrambled or Tet1-specific shRNAs.

We were able to strengthen these evidences by ChIP analyses: ChIP done with specific Ogt antibody in mESC interfered for Tet1 expression and relative control, reported Ogt loss from promoter regions of its target genes (figure 2.2.6F, and figure 2.2.4H iii, vi). Moreover, under this condition, the expression of Ogt-Tet1 common target genes was increased (figure 2.2.6G).

Taken together, these evidences highlight that Ogt is recruited to chromatin in a Tet1-dependent manner.

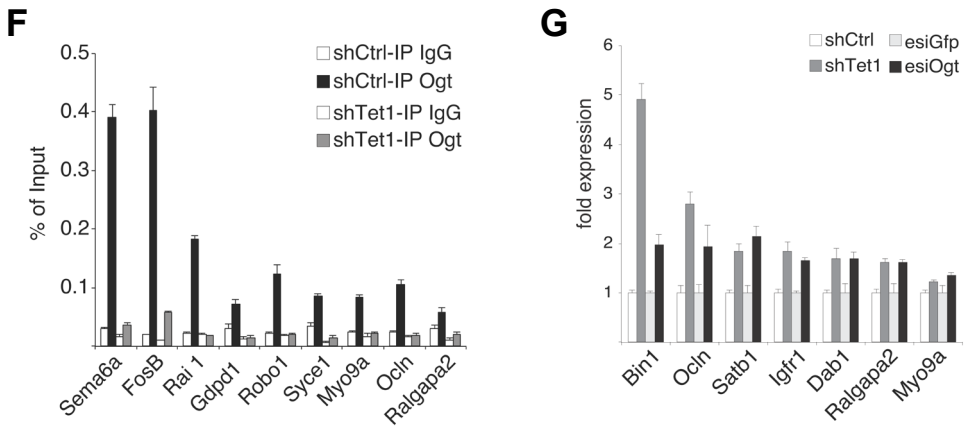
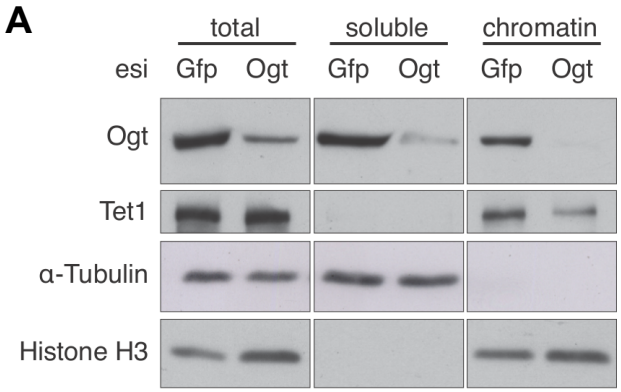


Fig.2.2.6 F, G. Tet1 depletion leads to Ogt displacement from Ogt-Tet1 common target loci and upregulation of target genes. (F) RT-qPCR analyses with primers for the indicated gene promoters on DNA purified from ChIP assays. ChIPs were performed with an Ogt-specific antibody in E14 ES cells that stably express scrambled (Ctrl) or Tet1-specific shRNAs. IgG were used as a negative control for ChIP assays. Data are represented as mean ± SEM. (G) Real time quantitative PCR analyses, using specific primers for the indicated genes, of RNA extracted from E14 ES cells stably expressing scrambled (Ctrl) or Tet1 specific shRNAs and E14 ES cells transfected for 48h with Gfp or Ogt specific esiRNAs. Data are represented as mean +/- SEM.

2.2.7. Effects of Ogt on Tet1 chromatin binding.

As proven by figure 2.2.1J, Tet1 was target of Ogt O-GlcNAcylation, thus we conceived a possible role for Ogt in stabilizing Tet1 binding to chromatin. To address this topic we fractionated protein extracts derived from mESC upon Ogt depletion by esiRNA, and immunoblot analyses showed a decrement of Tet1 level in chromatin fraction (figure 2.2.7A); similar results were achieved by ChIP analyses at target promoters. (figure 2.2.7B, C).



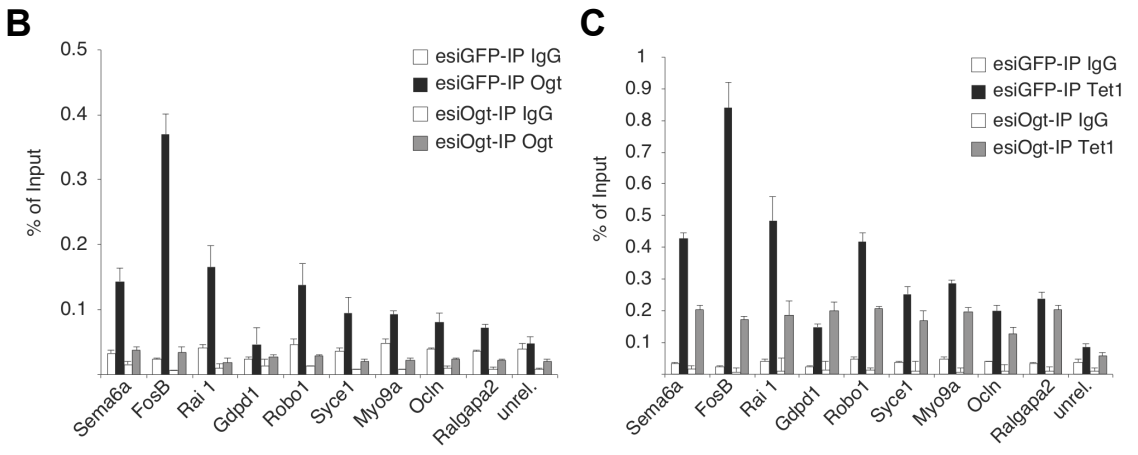


Fig.2.2.7 A-C. Ogt regulates Tet1 stability at chromatin. (A) WB with the indicated antibodies of E14 ES cells soluble and chromatin-associated proteins in cells transfected for 48 hr with Ogt- and Gfp-specific esiRNAs. a-tubulin and histone H3 are presented as fractionation controls. (B) (C) RT-qPCR analyses with primers for the indicated gene promoters with DNA purified from ChIP assays performed with Ogt (B) and Tet1 (C) specific antibody in E14 ES cells transfected for 48 hours with Ogt and Gfp-specific esiRNAs (B) or transduced with specific Tet1 targeting shRNA and sh Ctr (C). IgG are presented as a negative control for ChIP assays. Data are represented as mean ± SEM.

In spite of this, levels of 5hmc at Tet1 target sites (above all those bearing low 5hmC and high Ogt occupancy) increased upon Ogt depletion (figure 2.2.7D), suggesting that Ogt loss favours 5hmC accumulation. Conversely, stabilization of O-GlcNAcylation (figure 2.2.7E), achieved by chemical inhibition of Oga -which catalyzes the removal of such modification-, led to a decrease of 5hmc levels at the same target loci (figure 2.2.7F).

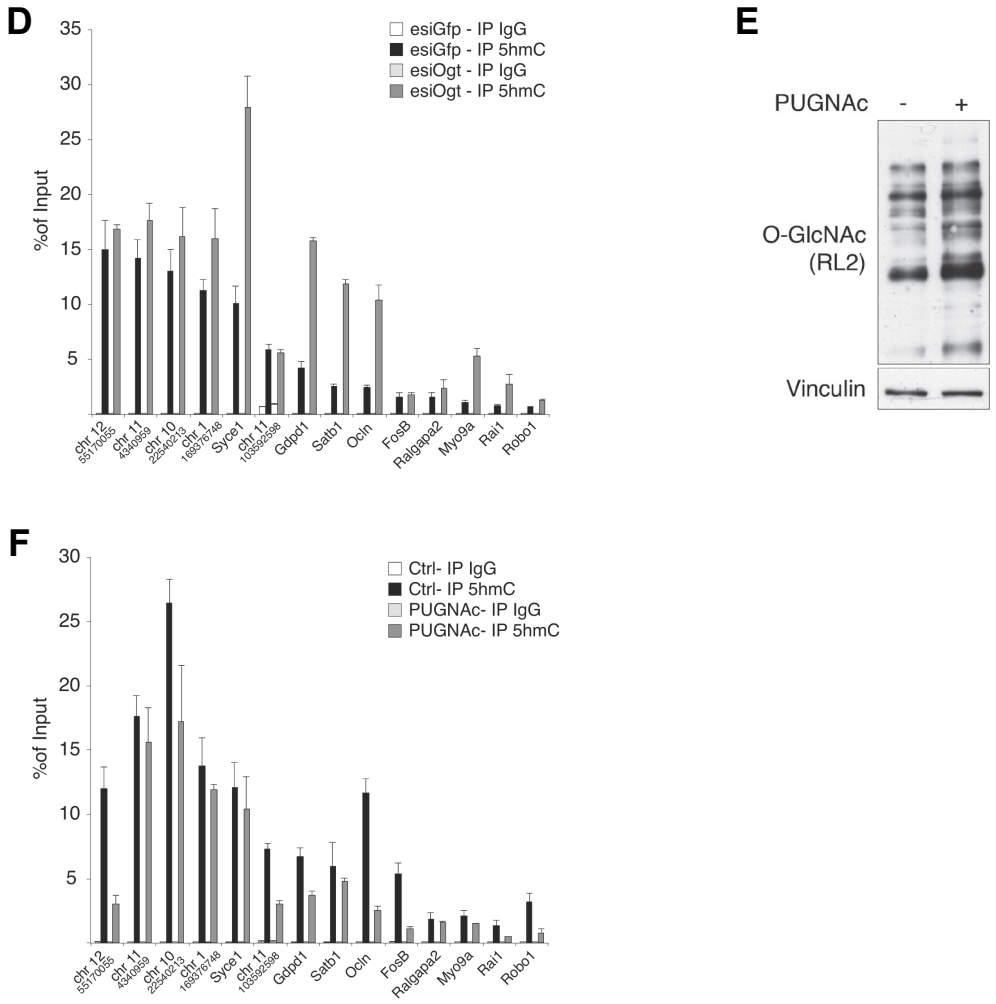


Fig.2.2.7 D-F. Ogt regulates locus specific Tet1 activity. (D) RT-qPCR analyses with primers for the indicated gene promoters and genomic sites (coordinates indicate the central point of PCR amplification) on DNA purified from DNA IP (DIP) assays performed with 5hmC specific antibodies on DNA extracted from E14 ES cells transfected for 48 hr with Ogt- and Gfp-specific esiRNAs. IgG were used as a negative control for the DIP assay. Data are represented as mean± SEM. (E) WBs using the indicated antibodies of protein lysate from E14 ES cells treated for 48 hours with 40μM PUGNAc. Vinculin served as loading control. (F) RT-qPCR analyses with primers for the gene promoters as figure (D) on DNA purified from DIP assays performed with 5hmC specific antibodies on DNA extracted from E14 ES cells treated for 48 hours with 40 μM PUGNAc. IgG was used as negative control for the DIP assay. Data are represented as mean± SEM.

These data endorsed us to conclude that Tet1 recruits Ogt to chromatin and that Ogt can exert a stabilizing function of Tet1 binding to its target promoters, thus regulating its enzymatic activity, i.e. 5hmC levels.

Chapter 3 – MATHERIAL AND METHODS

3.1. Plasmids.

3.1.1. Generation of pCAG-FLAG-AVI expression plasmid.

Coding sequence of mouse Ogt was amplified by polymerase chain reaction (PCR) using specific primers and relative product was cloned into a pCR8 Gateway entry vector by pCR8/Gateway/TOPO/TA cloning kit (Invitrogen). The reaction mix was used to transform TOP10 competent cells, plated on Luria Bertani (LB) agar plates containing 100 μg/mL spectinomycin. The expression construct for Flag-Avi-Ogt was obtained by recombination of pCR8-Ogt Gateway entry vector into a pCAG-Flag-Avi-IRES-Puromycin Gateway compatible destination vector, using LR recombinase (Invitrogen). The resulting constructs was checked by sequencing at Nucleic Acid Unit of the European Institute of Oncology.

3.1.2. Generation of lentiviral pLKO.1 vectors.

pLKO.1 lentiviral vectors were used to transduce mESC in order to perform efficient knock down of protein of interest. The target sequences, added with specific oligonucleotides allowing the insertion and the formation of the hairpin structure, were cloned into pLKO.1 backbone previously co-digested by AgeI and EcoRI restriction enzymes by means of T4 DNA ligase (Invitrogen). The following target sequences for the indicated proteins were used to generate the specific pLKO.1. Tet2: 5'-GCTCTGAACAGTATTCAAAGC-3'; Setd2 (TRCN0000238533): 5'-ACTTTGTGAGGATAGTATAAA-3'; Eed (TRCN0000095719): 5'-TCTTGCTAGTAAGGGCACATA-3'; Utx (TRCN0000096242): 5'-GCTACGAATCTCTAATCTTAA-3'; Jmjd3 (TRCN0000095265): 5'-

CCTCTGTTCTTGAGGGACAAA-3'; vector control was purchased from Sigma (SHC202). pLKO.1-Tet1#3 and #5 were described in[229].

3.2. Cell culture and manipulation.

3.2.1. Mouse embryonic stem cells: description and treatments.

Mouse embryonic stem cells (mESC) of ES-E14TG2a cell line were used. These cells, commonly called E14, are derivative of one of several embryonic stem cell (ES) lines developed by Smith and Hooper[274]. In some experiments, the line E36 was used. All mESC lines were grown on 0.1% gelatinized tissue culture dishes in GMEM supplemented with 15% fetal calf serum (Euroclone), 2 mM glutamine (Gibco), 100 U/ml penicillin and 0.1 mg/ml streptomycin (Gibco), 0.1 mM non-essential aminoacids (Gibco), 1 mM Na-Pyruvate (Gibco), 50 µM ß-mercaptoethanol-phosphate-buffered saline (PBS; Gibco) and Leukemia Inhibitory Factor (produced in house). *Ezh2* conditional (loxP/loxP) mouse ES cells, *Eed* WT and KO mouse ES cells and *Suz12* KO mouse ES cells were described previously[86,275,276]. BirA expressing ES cell clones were generated from an E14 ES cell line described elsewhere [277] by removing the puromycin selection coding sequence used for targeting purposes by transient CRE recombinase expression. Stable cell lines were obtained by transient transfection of pCAG-Flag-Avi empty, and Ogt expression constructs in BirA-ES cells using Lipofectamine 2000 (Invitrogen), followed by stable selection with 2µg/ml of puromycin.

Gal4-Jarid2-inducible 293Trex cells are described elsewhere[86]. The expression of Gal4-Jarid2 was induced by the addition of 1 µg/mL doxycycline (Sigma) to the cell media for 48 h before collecting for Luciferase and ChIP analyses. Cell lysates were prepared with Passive Lysis Buffer (Promega) and processed for Bradford Protein

Assay (Biorad) and Luciferase Assay (Promega) according to the manufacturer's protocol.

When indicated, wild type mouse ES cells were treated with 100 μM $CoCl_2$ (Sigma Aldrich) for 48 h. H_2O was used as vehicle control.

Inhibition of histone acetylation was achieved by treatment of wild type mouse ES cells with 35 μM C646 p300 inhibitor (EMD Millipore, Cat. 382113) for 48 h. DMSO was used as vehicle control.

Treatment with O-(2-Acetamido-2-deoxy-D-glucopyranosylidenamino) N-phenylcarbamate (PUGNAc, Sigma, catalog number A7229) was performed by adding the compound at 40 μM final concentration in proper medium for 48 hours.

3.2.2. 293T cell line.

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 CO_2 incubator (5% CO_2) with standard oxygen tension (21% oxygen).

3.2.3. RNA interference.

Stable RNA interference of target proteins were obtained with mESC transduction with viral particles (VSV-G pseudotyped lentivirus) delivering the pLKO.1 vectors described before. Viral particles were produced by calcium-phosphate transfection of 293T cells in a 10 cm-dish using 10 μg of lentiviral delivery vector, 3 μg of VSV-G and 6μg of Δ8.2 packaging vectors per dish. 36 hours post transfection, the supernatant containing viral particles was collected, 0.45μm filtered, LIF supplemented and added to $5x10^6$ mESC plated on a 10cm dish. 3 rounds of infection (~8 hours per day) were carried out in the presence of 5ug/mL of polybrene (Sigma) followed by selection

with the appropriate antibiotic selection (puromycin: 2 μg/mL; neomycin: 0.5 mg/mL). When indicated, RNA interfering was obtained by transfection of final 0.32 μg/ml endoribonuclease- prepared siRNA (esiRNA). Glp (EMU068911), esiRNA G9a (EMU050951), esiRNA GFP (EHUEGFP) were purchased from Sigma-Aldrich, and transfected into E14 mESC using Lipofectamine 2000 (Invitrogen). mESC were harvested after 48 hours post transfection.

3.2.4. Embryoid bodies generation.

Embryoid bodies formation was exploited as differentiation model of mESC. Indeed, this method allows the formation of cell aggregates differentiating towards the three germ layers. Undifferentiated mESC were induced to differentiate into embryoid bodies (EBs) by LIF removal in hanging drops containing 1,000 cell/20 μl drop on the lid of 15 cm petri dish for 48 h. EBs were then collected from the drops and stimulated between day 2 and 5 with 0.5 μM all-trans-retinoic acid (ATRA). EBs were left in culture in non-coated petri dishes up to day 9 in ES medium without LIF. Medium was replaced every second day.

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

3.3.1. Immunoblot analysis.

This method, commonly known as Western blot analysis (WB), allows the detection of protein of interest and relative post translational modification 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, 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 20-30 μg of protein extracts (and 1-5 μg of histones) were loaded onto each lane of a acrylamide, bisacrylamide gel, and a sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed. Gel-separated proteins were transferred to a Protran Nitrocellulose Membrane (Whatman), one hour and 20 minutes at 4 °C, at 100 Volts. Membranes were blocked with a solution of in TRIS-buffered saline (TBS: 20mM TRIS/HCl, pH 7.4, 137 mM NaCl, 2.7 mM KCl) plus 0.1% Tween (TBS-T) containing 5% non-fat dried milk. The same milk/TBS-T solution was prepared to dilute primary antibodies, which were incubated for 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. Although, histones were extracted from unsoluble fraction obtained after high salt lysis by the use of 8M urea, subsequently sonicated and boiled for 5 minutes.

3.3.2. Cellular fractionation.

Cytoplasmic and nuclear cell fractions were isolated re-suspending whole cell pellets in nuclear prep buffer, followed by centrifugation. Then, Laemmli sample buffer was added to supernatants, while nuclear pellets were washed once with nuclear prep buffer, resuspendeded in Laemmli sample buffer, briefly sonicated and boiled 10 minutes at 95 °C. All the fractions were then analyzed by Western blot.

3.3.3. Immunoprecipitation and streptavidin recovery.

Immunoprecipitations were performed by incubation of protein extracts (0.5-1 mg) for 2 hours with specific antibodies at 4°C on a rotating platform. Then, slurry protein A sepharose beads (GE Healthcare, Cat. 170780-01; 30 µl per milligram of protein extract) or protein A magnetic beads (Invitrogen, Cat. 10001D) were added to the lysates and incubated 1 hour at 4°C. After six washes with lysis buffer the immunoprecipitated proteins were eluted by Laemmli sample buffer addition to the beads and analyzed by WB. Direct streptavidin purifications were performed with the indicated BirA ES cell extracts by ON incubation of 25 µl of Streptavidin-coated magnetic beads (Invitrogen, Cat. 656-01) for each milligram of protein extract. Beads were then washed six times with lysis buffer and the proteins eluted with Laemmli sample buffer and analyzed by WB.

3.3.4. Non-nucleosomal histone H3 enrichment.

This method allows the purification of histones which are not incorporated imto chromatin. Whole cell extracts from E14 mouse ES cells were obtained by swelling cell pellets in high salt Lysis Buffer, then incubated for 30 min on ice and centrifuged at 13000rpm for 30 min at 4°C. To enrich soluble histones, 1 mg of total extract was incubated overnight with 3 µg of rabbit anti-Histone H3 antibody on a rotating platform at 4°C. Immunoprecipitated proteins were recovered by addition of 30 µL protein A-sepharose beads (GE Healthcare) for 1h at 4°C. Beads were washed six times with Lysis Buffer, then samples were denatured in Laemmli sample buffer for 5 min at 95°C, loaded on a polyacrylamide gel, and analyzed by WB.

3.3.5. Enrichment of glycosylated proteins through WGA-conjugated agarose resin.

To detect glycosylated proteins, agarose resin conjugated with wheat germ agglutinin is (WGA) available. WGA has a selective affinity for carbohydrate moieties. WGA purifications under denaturing conditions were carried out as described elsewhere [278]. Briefly, cells were resuspended in 5 M urea, 10 mM HEPES pH 7.4, 100 mM NaCl, and 1 mM EDTA; wash twice in 0.1 M TRIS/HCl pH 7.6 and the pellets lysed and sonicated in 0.1 M Tris/HCl pH 7.6, 2% SDS and 50 mM DTT. Supernatants were boiled for 10 minutes at 95ºC and diluted 3 times in 0.1 M Tris/HCl pH 7.6. Denatured supernatants were either used for IP analyses or incubated ON at 4°C with agarose bound, succinylated Wheat Germ Agglutinin (WGA) resin (Vector Laboratories, Cat. AL-1023S; 40 µl slurry for milligram of protein extract). The resin was then washed four times with lysis buffer, proteins eluted with Laemmli sample buffer and then analyzed by Western blot.

3.3.6. 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 colocalization of 2 or more antigens on the same focal plane with a resolution of 200-400nm.

Indicated mESC 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 and, when cells were treated with BrdU, further incubated with 20 mU/µl of DNAseI (NEB)

for 30 min at 37°C to unmask incorporated BrdU. Then, fixed cells or nuclei were incubated with primary antibodies diluted in 0.1% tween-TBS for 1 h 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.

Mouse E14 ES cells interfered for Setd2 protein and scrambled control were previously synchronized in G1/S phase by addition of 2 mM thymidine for 12 h; after 3h of release in normal medium, the cultured cells were pulsed with 33 μM BrdU for 30 min.

3.3.7. Size exclusion chromatography.

Size exclusion chromatography allows separation of molecules on the basis of their molecular size and conformation. The tool is a column containing loosely packed resin particles. The total volume (Vt) of the column is divided in three components: the volume of the external solvent (Vo or void volume), the solid volume constituted by the gel particles (Vg) and the internal volume of the porous particles (Vi). When a solute is added to the column, it will separate between these three parts. A solute excluded by the resin will be eluted after a volume equal to the void; conversely, a solute small enough to enter the bead channels slowed down and eluted later. On the basis of this principle, if proteins belong to the same complex, they will be eluted in the same fractions. Hence, nuclei from E14 mESC were lysed in high salt buffer and loaded on a Superose 6 10/30GL column (Amersham) and fractionated in high salt buffer. 0.5 mL fractions were collected, Laemmli-added, boiled for 10 minutes at 95 °C and WB-analyzed.

3.3.8. Tandem affinity purification.

This type of protein purification is used to improve the quality and the clearness of single step tag-protein purification. The FLAG tag is a commonly used to identify and isolate overexpressed proteins. The Avi tag consists in a peptidic sequence that could be identify by a bacterial enzyme called Bir-A that specifically biotinylate a lysine residue of the Avi peptidic sequence. This allows the efficient biotynilation of your Avi-tagged protein in cells that express the BirA enzyme. Furthermore, performing the second purification step using streptavidin coated beads eliminate antibodies chains which are usually responsible of the noise in later analyses (i.e. mass spectrometry).

To identify the interactome of Ogt protein, tandem (double-step) affinity purifications were performed with Flag-Avi-Ogt BirA mESC, compared to Flag-Avi-empty BirA ES cells used as negative control.Protein purifications were carried out on nuclei of the correspondent cell lines, prepared by 20 min swelling in nuclear prep buffer (10mM Tris, 100mM NaCl, 2mM MgCl2, 0.3 M Sucrose, 0.25 % v/v Igepal) at 4 °C. Nuclei were lysed in high salt buffer (50mM Tris–HCl pH 7.5, 300mM NaCl, 10% glycerol, 0.25% Igepal) with fresh addition of a protease inhibitor cocktail (Roche). The tandem affinity purifications were performed by incubating about 20 mg of nuclear protein extracts with 200 µl of packed agarose beads covalently linked with anti-Flag antibody (anti-Flag agarose beads, Sigma-Aldrich, cat. A2220) over night (ON) at 4 °C on a rotating platform. Beads were washed six times in minimum ten beads volumes of high salt buffer at 4°C and protein complexes eluted for 30 min with 0.5 mg/ml of Flag peptide (DYKDDDDK) in high salt buffer at 20°C four times. Eluates were pulled and further precipitated with 100 µl of streptavidin magnetic beads (Invitrogen, cat. 656-01) ON at 4 °C. Streptavidin beads were washed six times as before at 4°C and

protein complexes eluted with Laemmli sample buffer (Invitrogen). Purification efficiency was tested by separation of 10% of the eluate on a gel and stained using SilverQuest staining kit (Invitrogen, cat. LC6070). The remaining elauted purification was separated on an independent gel and stained with Coomassie Blue using a Colloidal Blue Staining Kit (Invitrogen, Cat. LC60nor25) for subsequent in gel trypsinization in order to perform Mass Spectrometry analysis.

3.3.9. 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 buy an ion source, 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 step 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[279].

3.3.10. Data analysis for Ogt interactome assessment.

This bioinformatic methodology allows retrieving information on the identities of the proteins that were interacting with your protein of interest using specific algorithm and protein databases.

Protein identification was performed using Andromeda search engine for purifications done with Flag-Avi-empty and Flag-Avi-Ogt BirA ES cells. The raw data from LTQ-FT Ultra were converted to mgf files using Raw2MSM software. Raw data files were analyzed using MaxQuant software (version 1.3.0.1) as described [280]. Parent ion and MSMS spectra were searched against a database containing 87,061 protein sequences obtained from the human IPI protein database version 3.68 and 248 protein sequences of commonly observed contaminants using Andromeda search engine [281]. Spectra were searched with a mass tolerance of 7 ppm in MS mode and 0.5

Da in MS/MS mode, strict trypsin specificity and allowing up to two missed cleavage sites. Cysteine 35 carbamidomethylation was searched as a fixed modification, whereas N-terminal protein acetylation, methionine oxidation were searched as variable modifications. We accepted peptides and proteins with a false discovery rate (FDR) of <1%, estimated based on the number of accepted reverse hits [281]. Intensity based absolute quantification score (iBAQ) intensity values as calculated by MaxQuant were used to estimate relative abundance of proteins in each experiment[282].

3.3.11. Acidic extraction of histones and Immunopurification.

E14 mESC 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 $125x10^6$ cells/ml and lysed 10 min at 4ºC with gentle stirring. Lysate was centrifuged onto a 10% sucrose cushion (10^7 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 Histone Buffer (50 mM Tris/HCl, pH 8.0, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100) and incubated at 4ºC for 4 hours with 10 µg of H3K36me3 antibody followed by 3 h incubation with protein A-sepharose beads (GE Healthcare). Beads were washed twice in 10-20 beads volumes with Histone Buffer, twice in 1xPBS and eluted in Laemmli sample buffer (Invitrogen).

3.3.12. In-gel digestion of histones for MS analysis

Immunopurified sample was separated by 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 immunopurified 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[258]. 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-C_{18}/Carbon and a strong cation exchange (SCX) solid phase extraction (SPE) StageTip[283]. Briefly, digested peptides loaded on combined RP-C_{18}/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) as previously described.

3.3.13. Data analysis for histone PTMs MS/MS

The mass spectrometric raw data were analyzed with the MaxQuant software (version 1.1.1.25)[284]. 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 [285]. 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 [286]. 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[258-260]. Output table from MaxQuant were filtered with the following criteria: peptides with a low score (cut-off score value, 60) [287] 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 both from H3(C110A) and H3(MLA) was 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(MLA) was further included in Andromeda configuration module (AndromedaConfig.exe) and for the specific delta masses we referred to manufacturer specification (http://www.activemotif.com/).

3.4. Cell cycle analysis and flow cytometry.

Indicated mESC 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. To assess antibody specificity, *Eed* WT and KO mouse ES cells were stained first with H3K27me2 antibody (Cell Signaling, Cat. 9728) and unrelated rabbit IgG as isotype control, then with APC-conjugated secondary antibody. Cell events were acquired at FACSCalibur flow cytometer (BD Biosciences) and analyzed with FlowJo 8.5.3 software (Tree Star). For cell cycle analysis, E14 mouse ES cells were cultured as previously described and pulsed with 33 μM 5-bromo-2-deoxyuridine (BrdU) for 15 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 and rabbit anti-H3K27me2 as primary antibodies; anti-mouse IgG-FITC conjugate and anti-rabbit IgG-APC conjugate were 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,

excluding doublets and non-viable cells. 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.

3.5. In vitro methyltransferase assay.

This method is widely used to demonstrate *in vitro* efficiency of methyltrasferase in catalyzing the transfer of methyl group from SAM to a substrate. *In vitro* histone methyltransferase assays were performed incubating 0.5 μg recombinant MLA histone H3K36me3 (Active Motif, Cat. 31219), or recombinant not modified histone H3 (Active Motif, Cat. 31207) with 1.5 μg human recombinant PRC2 complex (Sigma Aldrich, Cat. SRP0134) in 50 mM TRIS HCl pH 8.6, 5 mM $MgCl_2$, 4 mM DTT, 20 μM S-adenosylmethionine (New England Biolabs, Cat. B9003). Reactions were performed at 30°C for 1 h, then blocked with Laemmli sample buffer (Invitrogen), and denatured for 5 min at 95°C. Samples were loaded on a 4-12% NuPAGE Novex Bis-Tris gel (Invitrogen) to separate histones in order to perform subsequent in-gel digestion for MS analysis as described before.

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

3.6.1. 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 to an average size of 200–350 bp and immunoprecipitated ON with 1-5 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; 0.5-1 mg of chromatin was used for each precipitation; 100 µl 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).

3.6.2. DNA immunoprecipitation (DIP)

This method allows studies about DNA modifications and was performed as in [288] with little adaptations. Briefly, 1 µg of denatured and sonicated genomic DNA was immunoprecipitated overnight with 1 µg of rabbit anti 5hmC antibody in 100µL of IP buffer (10mM Na-Phosphate pH7, 140mM NaCl, 0,05% TritonX-100). Then, samples were incubated with 10 µL of slurry sepharose protein A beads for two hours, washed four times with IP buffer, and bound DNA eluted by incubation with 20 µg proteinase K (Roche) at 50°C in proper digestion buffer. Eluted DNA was recovered and purified using QIAquick PCR purification kit (Qiagen).

3.6.3. 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 pltaform (Illumina) following the protocol of the manufacturer.

3.6.4. Quantitative Real Time PCR (RT-qPCR)

The quantitative polymerase chain reaction (qPCR), also called real-time quantitative PCR RT–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 RT-qPCRs were carried out using Fast SYBR Green (Applied Biosystem) as dye. Lists of primers used are available online: http://dx.doi.org/10.1016/j.molcel.2012.12.019; http://dx.doi.org/10.1016/j.molcel.2013.10.030.

3.6.5. ChIP sequencing data analysis.

Sequencing data generated from the Illumina platforms related to the first project described, were aligned to mouse reference genome (mm9) using Bowtie version 0.12.7. Only reads with unique alignment were retained for downstream analysis.

Peak calling and bigWig files were generated using MACS version 1.4. Only peaks with 10x-Log p-value ≥70 are considered for further processing. bigWig files were visualized using the UCSC browser (http://genome.ucsc.edu). The list of mm9 annotated RefSeq genes used for the different analyses was downloaded from the UCSC database. Intragenic reads density for histone H3, H3K27me1, H3K27me2 and H3K36me3 were determined by computing the aligned reads within each RefSeq genes normalized for sequencing depth. PTM enrichments relative to histone H3 density were determined for each gene as the 10x–Log p-value computed using a chi-square test (PTM vs. H3) and adjusted using Bonferroni correction. The corrected p-values between different PTMs were compared using Pearson correlation test. Genome wide correlation among H3K27me1, H3K27me2 and H3K36me3 modifications with the read intensities in gene bodies was computed using PCA method in R factorMineR package. For generating the average occupancy in gene bodies, the gene body of each RefSeq gene was split into 20 bins where each bin represents 5% of the gene length. The occupancy was computed in each bin, normalized to the length of the bin and to the sequencing depth and then averaged over all genes. Intensity of each modification in all set of classes was then scaled down to a 0 to 1 range.

TSS vs. non-TSS location of H3K27ac peaks was determined by overlapping H3K27ac peaks with a 5 kb region centered on TSS for each mm9 RefSeq annotated gene. For capturing real intensities at the sites of active enhancers marked by K27ac peaks distal from TSS in comparison with genome wide distribution, we partitioned genome into small non-overlapping bins of 500 bp in size. Within these regions the histone H3 modifications K27ac, K4me1, K4me3, K27me3 and K27me2 from both WT and *Eed* KO samples were quantified by counting the number of aligned reads in the regions

followed by sequencing depth normalization and adjustment of outlier values to 99th quintile. The intensities generated using the same antibody in WT and *Eed* KO samples were then scaled down to a 0 to 1 range. Finally, the bins overlapping with K27ac distal peaks were merged together and the intensities obtained for all modifications between WT and *Eed* KO samples were represented as heatmap.

Each H3K27ac KO distal peak was assigned to the closest TSS RefSeq gene. These genes were then classified accordingly to their expression levels between WT and *Eed* KO and classified as up regulated (FC > 1.5) or down-regulated (FC < 1.5). For the genes belonging to each class as well as in the entire RNASeq dataset, we determined if the observed frequencies of up-regulated and down-regulated genes under putative control of the H3K27ac distal peaks were significantly different respect to the expected frequencies determined by analyzing the whole RNAseq dataset. Accordingly, we determined the relative distance of each H3K27ac distal peak identified in either WT or *Eed* KO samples respect to the closest up-regulated gene in Eed KO ES cells.

Active enhancers were classified on the basis of presence of both H3K27ac and H3K4me1 peaks, the absence of H3K4me3 and a minimal distance of 2.5Kb from annotated TSSs. Poised enhancers were defined by the absence of H3K27ac using the same criteria. The relative intensities of all the indicated histone marks were determined at H3K27me3 positive promoters, at poised and at active enhancers in mESC.

Ogt ChIPseq data was mapped onto the mm9 release of the mouse genome using Bowtie. Prior to alignment process short reads, which failed filtering (tagged as "Y") by Illumina sequencer was ignored. Duplicate reads with same alignment start position were removed with Picard (http://picard.sourceforge.net) markduplicates

tool. Samples were analyzed further using Model-based Analysis for Chip-Seq [289] for calling enriched regions w.r.t control. Wiggle tracks for the visualization on the UCSC genome browse (http://genome.ucsc.edu/) were generated using MACS. Only peaks with p-value > 70 were considered for downstream analysis. Annotation for Ogt binding sites was carried out with AnnotatePeaks utility of HOMER ChIP Seq suite (http://biowhat.ucsd.edu/homer/chipseq/index.html). Venn plot showing overlaps between Ogt, Tet1, K4me3 and Sin3a was generated by ChIPpeakAnno package[290]. CpG islands and gene sets of mouse mm9 assembly were downloaded from UCSC genome browser. Ogt and Tet1 binding sites were overlapped with CpG islands and promoter regions (2.5 kb up and downstream from TSS) for understanding its occupancy at empirical level. Density profile for target genes of all Ogt binding sites was generated for a region of 10kb up and downstream around transcription start site with a window size of 100 bp. Total tags within each window were normalized to sequencing depth. Raw data from previously published ChIPseq datasets GSE24843 and GSE12241 were aligned to the mm9 release following the same criteria.

For heat map generation, observed to expected (o/e) ratio was computed for all Tet1 binding sites, which were further classified into three classes namely low (o/e < 0.4), medium (0.4 < o/e < 0.78) and high (o/e > 0.78) CpG densities. Each binding site was tagged as CpG island positive if the o/e ratio was greater than 0.6. Composite heatmap profiles for all the classified groups were generated for a region of around 3kb up and downstream from the centre of the Tet1 binding site with a window size of 50 bp. Total tags within each window were normalized to sequencing depth. Similar profiles were also generated for low and unmethylated regions as previously reported[235].

3.7. Methods for RNA analysis.

3.7.1. 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. Differently from microarray analysis, this method is not limited by the number of probes present on the chip.

Total RNAs from mESC lines and embryoid bodies were extracted using TRIzol reagent (Invitrogen, Cat. 15596) according to manufacturer's instruction. Retrieved RNA was checked for integrity on a Bioanalyzer instrument by picoRNA Chip (Agilent), 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).

3.7.2. RNA sequencing data analysis.

RNASeq data generated for ES WT, ES *Eed* KO, Ebs WT, Ebs *Eed* KO samples were aligned to mouse reference genome using tophat. Differentially expressed genes were identified with cuffdiff. Microarray row data were retrieved from the Gene Omnibus Database (http://www.ncbi.nlm.nih.gov/geo/) at the accession number GSE19076. Data were RNA normalized using the affy package in R and probeset with a >1.3 fold change (FC) difference and a 95% confidence determined by T-test were selected for the analyses.

3.7.3. Real Time quantitative PCR

Total RNA was extracted with the Qiagen RNeasy Plus RNA extraction kit and retro-transcribed with M-MuLV Reverse Transcriptase RNase H (Finnzyme) according to the manufacturer's instructions. Real-time quantitative PCR (qPCR) was carried out using Fast Sybergreen as previously described. Link for primer detailsis also indicated.

3.7.4. Microarray analysis and gene ontology annotation.

Microarray technology allows the assessment and the evaluation of transcriptome variation among biological samples. It relies on chip containing arrays of RNA or DNA molecules, depending on the aims. The microarray contains a collection of DNA spots attached to a solid surface. Each DNA spot resembles a specific DNA sequence, known as probe, that can be for example a short portion of a gene. Fluorophore-, silver-, or chemiluminescence-labeled probes are used to hybridize a cDNA or an anti-sense RNA (cRNA) samples, which can be detected and quantified. Since such experiments are designed to compare different samples, data analysis mainly identifies statistically significant differences in gene expression and whether the regulated

genes are functionally related, after filtering out the background signal. RNA from two independent Ogt RNAi experiments was hybridized independently to Mouse Gene 1.0 ST Affymetrix Arrays. Signals were RMA normalized and analyzed using Affy and limma bioconductor packages in R. Affy IDs were annotated using mogene10sttranscriptcluster.db package. Probeset with a 1.4 fold expression difference and a p-value less than 0.05 were considered as differentially expressed. Both up and down regulated genes were overlapped with target genes of Ogt binding sites. Ingenuity (www.ingenuity.com) was used for functional annotation of up and downregulated genes. Data were independently validated bt RT-qPCR analyses as previously described.

3.8. Antibodies characterization.

3.8.1. Antibody specificity test

Antibody specificity for mouse anti-H3K27me1 (Active Motif, Cat. 61015), rabbit anti-H3K27me2 (Cell Signaling, Cat. 9728), rabbit anti-H3K27me3 (Active Motif, Cat. 39155), rabbit anti-H3K4me1 (Abcam, Cat. 8895), rabbit anti-H3K36me3 (Cell Signaling, Cat. 4909) was tested by immunoblot on a MODified™ Histone Peptide Array (Active Motif, Cat. 13005) according to manufacturer's instructions.

Mouse anti-H3K27me1 (Active Motif, Cat. 61015), rabbit anti-H3K27me2, and rabbit anti-H3K27me3 were further tested by immuno-dot blot assay, spotting 1 μg of biotinylated not methylated, Lys27 mono, di or tri-methylated Histone H3 peptide (JPT Peptide Technologies GmbH, custom made) on a nitrocellulose membrane, then analyzed by western blot. Biotin was used as loading control.

3.8.2. Antibodies used for Immunoblot and immunoprecipitation analyses.

For Western blots the following antibodies were used: mouse anti-Vinculin (Sigma-Aldrich, Cat. V9131), rabbit anti-HA (Santa Cruz, Cat. sc-805), rabbit anti-Oct4 (Abcam, Cat. ab19857), goat anti-Biotin (Pierce, Cat. 31852), mouse anti-Eed[78], rabbit anti-Ogt (Santa Cruz, Cat. sc-32921), rabbit anti-Tet1 C-term. (against the C-terminus of the protein) , rabbit anti-Tet1 N-term. [229], rabbit anti-Tet2 (produced in house), rabbit anti-Hcf1 (Bethyl Laboratories, Cat. A301-400A-1), rabbit anti-Sin3a (Santa Cruz, Cat. sc-994), mouse monoclonal anti-Ring1b (MBL, Cat. D139-3), mouse monoclonal RL2 anti-O-linked N-acetylglucosamine (Abcam, Cat. ab2739), mouse anti-Oct3-4 (Santa Cruz, Cat. Sc5279), rabbit anti-Nanog (Novus Biologicals, Cat. 100-587A), mouse anti-α Tubulin (Sigma-Aldrich, Cat. T9026), goat anti-Lamin B (Santa Cruz, Cat. sc6216), rabbit anti-Hif2 (Novus Biologicals, Cat. NB 100-122), mouse anti-S6 Ribosomal protein (Cell Signaling, Cat. 2317), rabbit anti-phospho S6 Ribosomal protein (Cell Signaling, Cat. 4857), rabbit anti-G9a (Cell Signaling, Cat. 3306S), rabbit anti-Glp (Abcam, Cat. 41969), rabbit anti-GAL4 (Santa Cruz, Cat. Sc-577). For histone PTMs detection the following antibodies were used: rabbit anti-H3 (Abcam, Cat. 1791), rabbit anti-H3K27me1 (Upstate, Cat. 07-448), mouse anti-H3K27me1 (Active Motif, Cat. 61015), rabbit anti-H3K27me2 (Cell Signaling, Cat. 9728), rabbit anti-H3K27me3 (Active Motif, Cat. 39155), rabbit anti-H3K27ac (Active Motif, Cat. 39133), rabbit anti-H3K36me3 (Cell Signaling, Cat. 4909), rabbit anti-H3K4me1 (Abcam, Cat. 8895), rabbit anti-H3K4me3 (Active Motif, Cat. 39159).

For immunoprecipitation experiments, the following antibodies were used: rabbit anti-Ogt, (Santa Cruz Cat. sc-32921), rabbit anti-Tet1 C-term. [229], rabbit anti-Tet1 (Millipore, Cat. 09-872) and rabbit anti-Tet2 (produced in house); rabbit anti-H3

(Abcam, Cat. 1791), rabbit anti-H3K36me3 (Cell Signaling, Cat. 4909), rabbit IgG (Sigma-Aldrich, Cat. I5006) were used as negative control.

3.8.3. Antibodies used for ChIP analyses.

In ChIP and DIP experiments, the following antibodies were used: rabbit anti-Ogt (Abcam Cat. ab50273), rabbit anti-Ogt (Santa Cruz Cat. sc-32921), rabbit anti-Sin3a (Abcam ab3479), rabbit anti-Tet1 N-term and rabbit anti-5hmC [229]. For histone PTMs ChIP the following antibodies were used: rabbit anti-H3 (Abcam, Cat. 1791), rabbit anti-H3K27me1 (Upstate, Cat. 07-448), mouse anti-H3K27me1 (Active Motif, Cat. 61015), rabbit anti-H3K27me2 (Cell Signaling, Cat. 9728), rabbit anti-H3K27me3 (Active Motif, Cat. 39155), rabbit anti-H3K27ac (Abcam, Cat. 4729), rabbit anti-H3K36me3 (Cell Signaling, Cat. 4909), rabbit anti-H3K4me1 (Abcam, Cat. 8895), rabbit anti-H3K4me3 (Active Motif, Cat. 39159). Rabbit IgG (Sigma, Cat. I5006) were used as negative control.

3.8.4. Antibodies used for immunofluorescence staining and flow cytometry.

For immunofluorescence, the following antibodies were used: rabbit anti-Ogt, (Santa Cruz Cat. sc-32921), and rabbit anti-Tet2 (produced in house); rabbit anti-Suz12 (Cell Signaling, Cat. 3737S), 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). rabbit IgG (Sigma-Aldrich, Cat. I5006) were used as negative control. For FACS analysis the following antibodies were employed: rabbit anti-H3K27me2 antibody (Cell Signaling, Cat. 9728), mouse anti-BrdU (BD Biosciences, Cat. 347580); anti-mouse IgG-FITC conjugate and anti-rabbit IgG-APC conjugate were used as secondary antibodies.

Chapter 4 - DISCUSSION

4.1. Polycomb dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity.

4.1.1. Methylation states of H3K27 in mouse embryonic stem cells are dependent on PRC2 activity.

Our data based on MS approach show how about 80% of total H3K27 is methylated in a PRC2 dependent manner; these observation are endorsed by genome wide approaches which unraveled how methylation forms of H3K27 are differentially deposited throughout the genome in a mutual exclusive manner. Notably, the diffused H3K27me2 distribution along the genome reflects its abundance in the MS quantification. H3K27me1 enriches at gene bodies of active genes; this seems to be a conserved phenomenon among species and tissues, as this correlation was already reported in human $CD4^+$ T cells[291].

4.1.2. PRC2 enzymatic activity is influenced by the extent of binding to DNA.

While it is widely demonstrated that H3K27me3 mark is established at target loci upon stable PRC2 binding to chromatin, the other methylation forms of K27 are deposited, with high efficiency, when PRC2 transiently binds the DNA. *In vitro* assays on histone H3 tails demonstrated how PRC2 is more efficient in catalyzing H3K27me0 to me1 and me2 than in converting H3K27me2 into me3[122]. Thus, H3K27me1 and H3K27me2 are deposited upon transient association of PRC2 to DNA, with the H3K27me2 being the main activity of PRC2, as suggested by its diffused genome wide distribution and its considerable high levels on bulk histones. Previous reports unraveled the PRC2 (EZH2) localization at sites on ongoing replication marked by

BrdU[101], but the number of H3K27me3 peaks identified in mESC[257] were relatively low respect to the high extent of overlap between replication foci and EZH2; thus, this association could rely on H3K27me2 rather than to H3K27me3. Our finding that correlates H3K27me2 deposition with DNA synthesis during S phase endorses this hypothesis. Furthermore, other reports showed H3K27me1 and H3K27me2 to be readily deposited into novel nucleosomal histones, while H3K27me3 deposition appeared to be delayed[292]. The nature of PRC2 association to DNA replication machinery would be an important topic to address in order to further elucidate the mechanisms of methylation deposition and PRC2 recruitment. PRC2 accessory proteins, able to bind histone proteins, could be a fundamental role in the recognition of pre existing mark within histone tails, thus driving PRC2 enzymatic activity.

4.1.3. Establishment of discrete H3K27me1 and H3K27me2 domains is regulated by H3K36me3.

In order to understand how H3K27me1 and H3K27me2 are deposited upon transient PRC2 interaction with chromatin in distinct genomic domain, we generated data showing that loss of H3K36me3 impaired intragenic H3K27me1 deposition causing consequent H3K27me2 accumulation at the same genomic loci. Nevertheless, the loss of Setd2-mediated H3K36me3 did not affect PRC2 localization at DNA replication foci. In addition, the observed co-existence of H3K27me1 (and not H3K27me2) and H3K36me3 on the same histone tail, support the model by which H3K36me3 is able to control the ratio between H3K27me1/me2. Another study, in agreement with our results, reported PRC2 inefficiency to catalyze H3K27me2 on histone tail bearing H3K36me3[36]. Overlall, we propose a model in which H3K36me2 is the main PRC2 activity taking place along with DNA synthesis, H3K27me1 domains are formed in a

H3K36me3 dependent manner inhibiting further methylation into dimethylated form. H3K27me3 is deposited at CpG-rich loci upon stable PRC2 binding.

4.1.4. Deposition of H3K27me1 and H3K36me3 regulates transcription.

Decreased expression of H3K27me1 positive genes in Eed KO mESC, and its correlation with H3K36me3, strongly suggests an implication of PRC2 enzymatic activity in active gene expression. Different mechanisms underlying this correlation can be envisaged. For example, H3K36me3 and H3K27me1 deposited at gene bodies of active genes could lead to nucleosome mobility, or H3K27me1 could be involved in the RNA splicing-dependent recruitment of Setd2[293], thus functioning as a permissive modification for elongation or splicing, while H3K27me2 could inhibit this process.

4.1.5. H3K27me2 ensures proper enhancer regulation in mESC.

Upon loss of PRC2 activity, a global increase in H3K27ac levels was observed, indicating that in these conditions histone acetyltransferases (HATs) are more accessible to chromatin. H3K27ac changes occur at enhancer elements, and this histone PTM discriminates active enhancers from poised ones[28]. Thus, we showed that loss of PRC2 activity causes the activation of poised enhancers in mESC, possibly through increased chromatin accessibility to HAT activities; we then conceived that the broad ‘‘unspecific’’ deposition of H3K27me2 protects H3K27 from HATs activity. This is supported by the rapid deposition of H3K27me2 at sites that lose H3K27ac upon inhibition of global HAT activity. In addition, 60% of novel H3K27ac accumulation at enhancer sites takes place in sites that are pre marked by H3K4me1, highlighting a link between HATs recruitment and H3K4me1. Aberrantly activated enhancers upon PRC2 loss could highly contribute to the several defects in development and lineage specification characterizing PRC2 deficient mouse models

and mESC lines. In summary, these novel characterized functions of H3K27me1 and H3K27me2 propose new structural and protective activities of PRC2 in transcriptional regulation and cell type specification. Figure 4.1.1 summarizes the model that we propose.

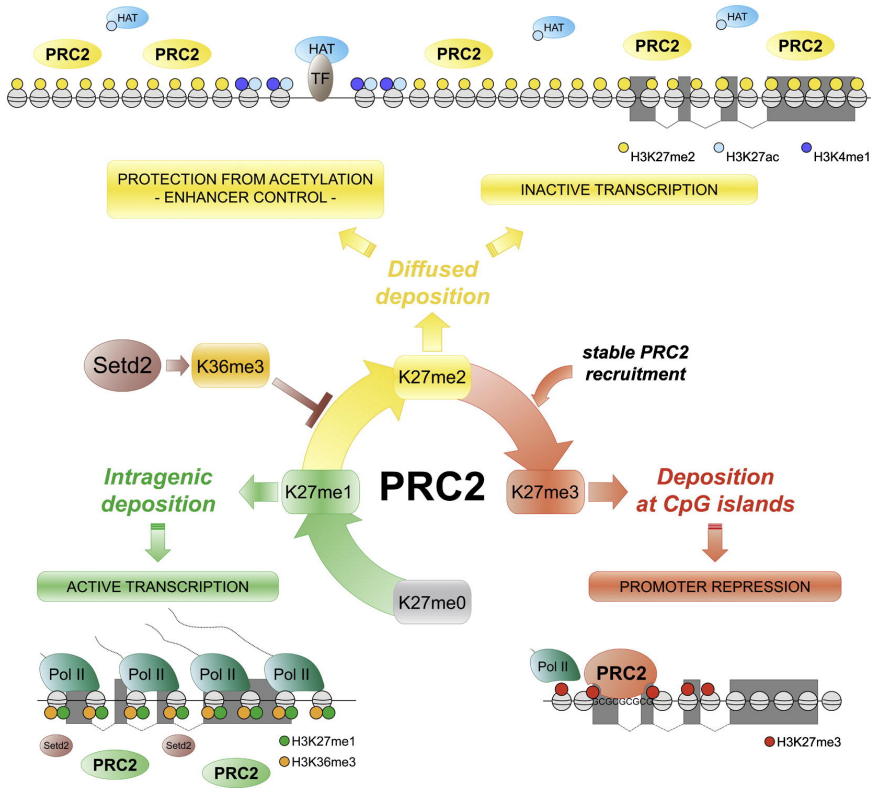


Figure 4.1.1 Model summarizing the different functions of methylated forms of Lys27 on histone H3.

4.1.6. Open issues.

The worthy and interesting results that we obtained certainly fuel further questions about the roles of methylation forms of H3K27, and the mechanisms underlying their deposition.

As previously discussed, the enrichment of H3K27me1 at intragenic regions of transcribed genes seems to be a conserved mechanism; indeed, it would be interesting to know if the differential and mutually exclusive deposition of methylated forms takes place also in other cellular models, above all in terminal

differentiated cells, or if the well defined patterns of deposition of the K27 methylated forms are perturbed in pathologic conditions, such as upon cancer onset. This could further confirm that the different methylated forms of K27 cooperates in the regulation of gene transcription.

Another interesting topic to further investigate would be to understand the mechanisms and factors underlying transient PRC2 localization to DNA replication sites. Most probably, still undefined PRC2 subunits governs this process. This aspect could be addressed by means of unbiased MS approach aiming to identify PRC2 associated protein in the S phase of the cell cycle upon short release after thymidine block. This would eventually also contribute to the identification of protein recruiting PRC2 at selected target sites.

Understanding the mechanisms by which Setd2-dependent H3K36me3 is able to control the ratio between H3K27me1 and me2 would be very worthy and challenging to pursue, as well as to understand the nature of H3K27me1-H3K36me3 link with active transcription. In this context we could expect a role for such histone PTMs in facilitating mRNA splicing and RNA Pol II elongation.

Finally, of important note is the protective activity of H3K27me2 ensuring proper enhancer element regulation. Several evidences present in the literature show how the activity of enhancer elements can be disrupted in cancer by means of genetic and epigenetic alterations[294]. Moreover, also PcG proteins are hugely implicated in cancer onset: thus it should be worthy to investigate how PRC2-dependent H3K27me2 domain are altered in cancer cell lines, and how this alteration could have consequences in aberrant enhancer firing, potentially leading to oncogene activation.

4.2. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells.

4.2.1. Ogt is part of a multiprotein complex containing Tet proteins and Hcfc1.

Out data reported the presence of a stable complex on chromatin between two enzymatic activities, the O-GlcNAc transferase Ogt and the hydroxylases Tet1 and Tet2. We also identified Sin3A and Hcfc1 as Ogt binding partners, previously shown to associate with Tet1: this suggests the existence of a multiprotein complex, gathering enzymatic activities linked to transcriptional repression. In particular, in the light of the weak interactions found between Hcfc1 and Tet1, we envisage that Ogt associates to Tets and Hcfc1 in different complexes staying in close contact through Ogt ability to form multimers[295]. Within this complex, Hcfc1 seems to have a key role, explaining the observed correlation between Tet1-Ogt interaction and their localization to CpG rich promoters enriched for H3K4me3. Indeed, Hcfc1 interacts, with different domains, with both Sin3A/NURD complex and Set1/Ash2 H3K4 methyltransferase[296]. The Tet-Ogt co-localization at H3K4me3 positive promoters further correlates with unmethylated cytosine, as H3K4me3 deposition inhibits DNA methylation. Therefore, our data suggested a functional connection between Ogt-Tet-Hcfc1, the accumulation of H3K4me3 and the prevention of cytosine methylation at promoter regions. In this context, Ogt was shown to regulate Hcfc1 activity[297], while we demonstrated that Ogt is recruited to chromatin to exert its functions by Tet1.

4.2.2. Genome-wide Ogt binding correlate with unmethylated CpG rich promoters.

Although our ChIP seq data evidence a large overlap between Tet1 and Ogt genome-wide localization, Ogt associates quite exclusively to TSS regions, while Tet1 is found also at regulatory elements. However, this could be the consequence of the lower Tet1 binding at CpG poor elements that, owing to the lower Ogt ChIP sensitivity, could result in a loss of Ogt detected signal at those sites. Tet2, as a consequence of the lack of a CXXC domain, is loosely associated to chromatin, and this could explain our failure in obtaining good Tet2 ChIP results. The weak interaction that we observed between Tet1 and Tet2 (even further decreased upon Ogt depletion) could indicate that Ogt functions as a bridge between these proteins. Tet1 and Tet2 have been shown to independently regulate 5hmC levels[229,242,298,299], thus indicating additional mechanisms that control Tet2 enzymatic activity. Our results showed Ogt binds Tet1 at CpG rich promoters with unmodified cytosines, suggesting that, in general, the maintenance of 5hmC levels are independent from Ogt. However, Ogt depletion led to local 5hmC enrichment, thus we can hypothesize that Ogt is required for locally maintain low 5hmC levels.

Overall, we proposed a summarizing model shown in figure 4.2.1.

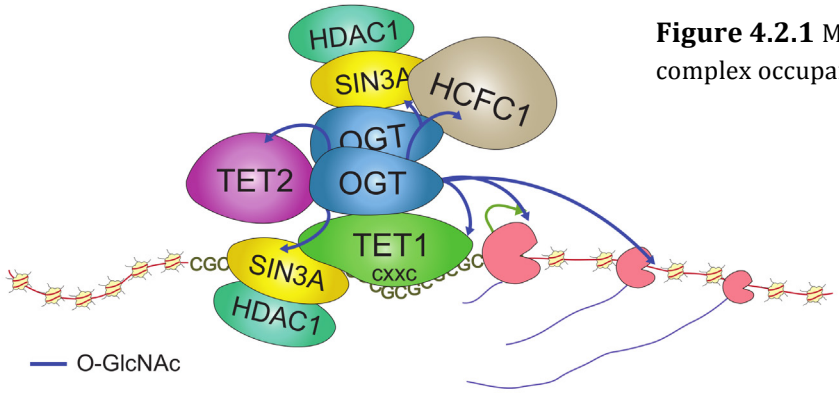


Figure 4.2.1 Model of the TETs-Ogt complex occupancy at target sites.

4.2.3. Ogt roles as transcriptional regulator.

In light of its localization at TSS regions, it is likey that Ogt could have transcriptional regulation activity. Indeed, upon Ogt depletion, upregulated and downregulated genes, involved in metabolic and signaling pathways, were detected. To assess the direct or indirect role for Ogt in transcriptional regulation is though, due to its pleiotropic function within the cell. We reported a preferential association of Ogt to TSSs of upregulated genes, thus suggesting a repressive function. This is in line with what reported for Tet1[229,240] and the repressive function of Tet1–Ogt interacting partners (Sin3A and HDAC1). A key role in this context could be exerted by O-GlcNAcylation of RNA Pol II CTD and histone PTMs. Ogt role as transcriptional activator, as highlighted by gene down regulation upon Ogt loss, is consistent with both previous reports on Tet1 activity and the role of Ogt in regulating H2B ubiquitylation through the GlcNAcylation of H2B S112[300]. Moreover, Ogt could contribute to maintain CpG rich promoters devoid of methylation in cooperation with Tet1 through Tet1 O-GlcNAcylation.

In spite of the possible Ogt roles in transcriptional regulation, the principal functions of Ogt are related to its several cytoplasmic activities. Indeed, the severe effects observed in Ogt KO embryos, which die at early developmental stages, cannot be ascribed to chromatin related Ogt functions, also at the light of the lack of striking developmental defects of Tets KO mice[242,243,245]. This is further endorsed by the deregulation of metabolic and signaling pathways driven by the Ogt loss observed in our transcriptome analysis; furthermore, such deregulation was not observed upon loss of Tet proteins function.

4.2.4. Open issues.

Recent works reported also Ogt-Tet3 association in HEK293T cells where Tet3 was ectopically overexpressed[301]; we were not able to observe uch interaction due to the absence of Tet3 in mESC. Indeed, Tet3 is mainly expressed in the zygote and oxydize 5mC of paternal genome, regulating embryonic development. Assessing the role of possible Ogt-Tet3 complex in the regulation of 5hmC levels during embryonic development would be very interesting to pursue.

Moreover, a still open question is the understanding of the roles of the catalytic activity of Ogt on Tet1 and 5hmC deposition. Tet1 enzymatic proficiency could be tested in cells retaining Tet1 mutant at O-glycosylation site; thus, if this modification were fundamental, mutant Tet1 activity would be impaired.

As O-GlcNAcylation is uniquely reverted by Oga enzyme, is likely that it could be involved in TETs-Ogt partnership, recruited to chromatin thus participating in chromatin regulation processes.

From a pathological point of view, it should be intriguing to investigate the roles of Ogt-Tet1 partnership in neurodegenerative disorders like Alzheimer's disease (AD). Ogt, which is highly expressed in the brain, exerts O-GlcNAcylation of proteins (namely tau protein and β-amyloid fibrils) that form aggregates causing neuronal degeneration[302]. Similarly, TET activity is high in mammalian adult brain[303], where the highest ratio between 5hmC and 5mC has been found. Tet1 KO mouse show altered neurological phenotypes, suggesting 5hmC to be relevant in this context. Although OGT role in AD does not seem associated to its chromatin-related functions, it will be interesting to characterize the role of Ogt-Tet proteins interaction in neurodegenerative disorders.

REFERENCES

1 Wu, H. & Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45-68, doi:10.1016/j.cell.2013.12.019 (2014).

2 Hanover, J. A., Krause, M. W. & Love, D. C. Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. *Nature reviews. Molecular cell biology* **13**, 312-321, doi:10.1038/nrm3334 (2012).

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

4 Woodcock, C. L. & Ghosh, R. P. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology* **2**, a000596, doi:10.1101/cshperspect.a000596 (2010).

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

6 Luger, K. Structure and dynamic behavior of nucleosomes. *Current opinion in genetics & development* **13**, 127-135 (2003).

7 Ausio, J. The shades of gray of the chromatin fiber: Recent literature provides new insights into the structure of chromatin. *BioEssays : news and reviews in molecular, cellular and developmental biology*, doi:10.1002/bies.201400144 (2014).

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

9 Loyola, A., Bonaldi, T., Roche, D., Imhof, A. & Almouzni, G. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Molecular cell* **24**, 309-316, doi:10.1016/j.molcel.2006.08.019 (2006).

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

11 Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678-691, doi:10.1016/j.cell.2010.01.003 (2010).

12 Filipescu, D., Muller, S. & Almouzni, G. Histone h3 variants and their chaperones during development and disease: contributing to epigenetic control. *Annual review of cell and developmental biology* **30**, 615-646, doi:10.1146/annurev-cellbio-100913-013311 (2014).

13 Wallrath, L. L., Lu, Q., Granok, H. & Elgin, S. C. Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures. *BioEssays : news and reviews in molecular, cellular and developmental biology* **16**, 165-170, doi:10.1002/bies.950160306 (1994).

14 Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **51**, 786-794 (1964).

15 Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016-1028, doi:10.1016/j.cell.2011.08.008 (2011).

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

17 Cosgrove, M. S., Boeke, J. D. & Wolberger, C. Regulated nucleosome mobility and the histone code. *Nature structural & molecular biology* **11**, 1037-1043, doi:10.1038/nsmb851 (2004).

18 Voigt, P. *et al.* Asymmetrically modified nucleosomes. *Cell* **151**, 181-193, doi:10.1016/j.cell.2012.09.002 (2012).

19 Phillips, D. M. The presence of acetyl groups of histones. *The Biochemical journal* **87**, 258-263 (1963).

20 Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nature structural & molecular biology* **20**, 259-266, doi:10.1038/nsmb.2470 (2013).

21 Unnikrishnan, A., Gafken, P. R. & Tsukiyama, T. Dynamic changes in histone acetylation regulate origins of DNA replication. *Nature structural & molecular biology* **17**, 430-437, doi:10.1038/nsmb.1780 (2010).

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

23 Zeng, L. & Zhou, M. M. Bromodomain: an acetyl-lysine binding domain. *FEBS letters* **513**, 124-128 (2002).

24 Suka, N., Suka, Y., Carmen, A. A., Wu, J. & Grunstein, M. Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Molecular cell* **8**, 473-479 (2001).

25 Garcia, B. A. *et al.* Organismal differences in post-translational modifications in histones H3 and H4. *The Journal of biological chemistry* **282**, 7641-7655, doi:10.1074/jbc.M607900200 (2007).

26 Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics* **40**, 897-903, doi:10.1038/ng.154 (2008).

27 Tie, F. *et al.* CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* **136**, 3131-3141, doi:10.1242/dev.037127 (2009).

28 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

29 Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Molecular cell* **33**, 1-13, doi:10.1016/j.molcel.2008.12.013 (2009).

30 Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599, doi:10.1038/35020506 (2000).

31 Nguyen, A. T. & Zhang, Y. The diverse functions of Dot1 and H3K79 methylation. *Genes & development* **25**, 1345-1358, doi:10.1101/gad.2057811 (2011).

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

33 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 & development* **15**, 3286-3295, doi:10.1101/gad.940201 (2001).

34 Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell* **25**, 15-30, doi:10.1016/j.molcel.2006.12.014 (2007).

35 Schwartz, Y. B. & Pirrotta, V. Polycomb silencing mechanisms and the management of genomic programmes. *Nature reviews. Genetics* **8**, 9-22, doi:10.1038/nrg1981 (2007).

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

37 Schneider, R. *et al.* Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature cell biology* **6**, 73-77, doi:10.1038/ncb1076 (2004).

38 Santos-Rosa, H. *et al.* Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Molecular cell* **12**, 1325-1332 (2003).

39 Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R., 3rd & Grant, P. A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**, 434-438, doi:10.1038/nature03242 (2005).

40 Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88, doi:10.1016/j.cell.2007.05.042 (2007).

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

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

43 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics* **39**, 311-318, doi:10.1038/ng1966 (2007).

44 Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nature reviews. Molecular cell biology* **13**, 115-126, doi:10.1038/nrm3274 (2012).

45 Kizer, K. O. *et al.* A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol Cell Biol* **25**, 3305-3316, doi:10.1128/MCB.25.8.3305-3316.2005 (2005).

46 Li, B., Howe, L., Anderson, S., Yates, J. R., 3rd & Workman, J. L. The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *The Journal of biological chemistry* **278**, 8897-8903, doi:10.1074/jbc.M212134200 (2003).

47 Li, J., Moazed, D. & Gygi, S. P. Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *The Journal of biological chemistry* **277**, 49383-49388, doi:10.1074/jbc.M209294200 (2002).

48 Xiao, T. *et al.* Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes & development* **17**, 654-663, doi:10.1101/gad.1055503 (2003).

49 Sun, X. J. *et al.* Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *The Journal of biological chemistry* **280**, 35261-35271, doi:10.1074/jbc.M504012200 (2005).

50 Yuan, W. *et al.* Heterogeneous nuclear ribonucleoprotein L Is a subunit of human KMT3a/Set2 complex required for H3 Lys-36 trimethylation activity in vivo. *The Journal of biological chemistry* **284**, 15701-15707, doi:10.1074/jbc.M808431200 (2009).

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

52 Carrozza, M. J. *et al.* Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581-592, doi:10.1016/j.cell.2005.10.023 (2005).

53 Joshi, A. A. & Struhl, K. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Molecular cell* **20**, 971-978, doi:10.1016/j.molcel.2005.11.021 (2005).

54 Keogh, M. C. *et al.* Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**, 593-605, doi:10.1016/j.cell.2005.10.025 (2005).

55 Pradeepa, M. M., Sutherland, H. G., Ule, J., Grimes, G. R. & Bickmore, W. A. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS genetics* **8**, e1002717, doi:10.1371/journal.pgen.1002717 (2012).

56 Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature genetics* **41**, 376-381, doi:10.1038/ng.322 (2009).

57 Watt, F. & Molloy, P. L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & development* **2**, 1136-1143 (1988).

58 Rothbart, S. B. & Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochimica et biophysica acta* **1839**, 627-643, doi:10.1016/j.bbagrm.2014.03.001 (2014).

59 Blackledge, N. P. & Klose, R. CpG island chromatin: a platform for gene regulation. *Epigenetics : official journal of the DNA Methylation Society* **6**, 147-152 (2011).

60 Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends in biochemical sciences* **31**, 89-97, doi:10.1016/j.tibs.2005.12.008 (2006).

61 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935, doi:10.1126/science.1170116 (2009).

62 Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303, doi:10.1126/science.1210597 (2011).

63 Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692-706, doi:10.1016/j.cell.2013.04.002 (2013).

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

65 Brunk, B. P., Martin, E. C. & Adler, P. N. Drosophila genes Posterior Sex Combs and Suppressor two of zeste encode proteins with homology to the murine bmi-1 oncogene. *Nature* **353**, 351-353, doi:10.1038/353351a0 (1991).

66 van Lohuizen, M., Frasch, M., Wientjens, E. & Berns, A. Sequence similarity between the mammalian bmi-1 proto-oncogene and the Drosophila regulatory genes Psc and Su(z)2. *Nature* **353**, 353-355, doi:10.1038/353353a0 (1991).

67 van Lohuizen, M. *et al.* Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* **65**, 737-752 (1991).

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

69 Schumacher, A. & Magnuson, T. Murine Polycomb- and trithorax-group genes regulate homeotic pathways and beyond. *Trends in genetics : TIG* **13**, 167-170 (1997).

70 Piunti, A. & Pasini, D. Epigenetic factors in cancer development: polycomb group proteins. *Future oncology* **7**, 57-75, doi:10.2217/fon.10.157 (2011).

71 Schwartz, Y. B. & Pirrotta, V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nature reviews. Genetics* **14**, 853-864, doi:10.1038/nrg3603 (2013).

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

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

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

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

76 Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Molecular cell* **20**, 845-854, doi:10.1016/j.molcel.2005.12.002 (2005).

77 Elderkin, S. *et al.* A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. *Molecular cell* **28**, 107-120, doi:10.1016/j.molcel.2007.08.009 (2007).

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

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

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

81 Orlando, V. & Paro, R. Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. *Cell* **75**, 1187-1198 (1993).

82 Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Molecular cell* **49**, 808-824, doi:10.1016/j.molcel.2013.02.013 (2013).

83 Laugesen, A. & Helin, K. Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. *Cell Stem Cell* **14**, 735-751, doi:10.1016/j.stem.2014.05.006 (2014).

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

85 Wu, X., Johansen, J. V. & Helin, K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Molecular cell* **49**, 1134-1146, doi:10.1016/j.molcel.2013.01.016 (2013).

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

87 He, J. *et al.* Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nature cell biology* **15**, 373-384, doi:10.1038/ncb2702 (2013).

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

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

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

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

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

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

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

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

96 O'Carroll, D. *et al.* The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol* **21**, 4330-4336, doi:10.1128/MCB.21.13.4330-4336.2001 (2001).

97 Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313, doi:10.1016/j.cell.2006.02.043 (2006).

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

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

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

101 Hansen, K. H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nature cell biology* **10**, 1291-1300, doi:10.1038/ncb1787 (2008).

102 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). *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2989-2994, doi:10.1073/pnas.1116418109 (2012).

103 Kim, H., Kang, K. & Kim, J. AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic acids research* **37**, 2940-2950, doi:10.1093/nar/gkp149 (2009).

104 Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V. & Reinberg, D. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol* **28**, 2718-2731, doi:10.1128/MCB.02017-07 (2008).

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

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

107 Zhang, Z. *et al.* PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. *Stem cells* **29**, 229-240, doi:10.1002/stem.578 (2011).

108 Boulay, G., Rosnoblet, C., Guerardel, C., Angrand, P. O. & Leprince, D. Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes. *The Biochemical journal* **434**, 333-342, doi:10.1042/BJ20100944 (2011).

109 Wang, S., Robertson, G. P. & Zhu, J. A novel human homologue of Drosophila polycomblike gene is up-regulated in multiple cancers. *Gene* **343**, 69-78, doi:10.1016/j.gene.2004.09.006 (2004).

110 Ballare, C. *et al.* Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nature structural & molecular biology* **19**, 1257-1265, doi:10.1038/nsmb.2434 (2012).

111 Li, G. *et al.* Jarid2 and PRC2, partners in regulating gene expression. *Genes & development* **24**, 368-380, doi:10.1101/gad.1886410 (2010).

112 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. *Nature cell biology* **12**, 618-624, doi:10.1038/ncb2065 (2010).

113 Shen, X. *et al.* Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* **139**, 1303-1314, doi:10.1016/j.cell.2009.12.003 (2009).

114 Peng, J. C. *et al.* Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**, 1290-1302, doi:10.1016/j.cell.2009.12.002 (2009).

115 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. *Nature cell biology* **12**, 618-624, doi:10.1038/ncb2065 (2010).

116 Li, G. *et al.* Jarid2 and PRC2, partners in regulating gene expression. *Genes & development* **24**, 368-380, doi:10.1101/gad.1886410 (2010).

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

118 Bracken, A. P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *Embo Journal* **22**, 5323-5335, doi:Doi 10.1093/Emboj/Cdg542 (2003).

119 Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature genetics* **42**, 181-U124, doi:Doi 10.1038/Ng.518 (2010).

120 Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298-303, doi:10.1038/nature10351 (2011).

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

122 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). *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2989-2994, doi:10.1073/pnas.1116418109 (2012).

123 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. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20980-20985, doi:10.1073/pnas.1012525107 (2010).

124 Wigle, T. J. *et al.* The Y641C mutation of EZH2 alters substrate specificity for histone H3 lysine 27 methylation states. *FEBS letters* **585**, 3011-3014, doi:10.1016/j.febslet.2011.08.018 (2011).

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

126 Vandamme, J., Volkel, P., Rosnoblet, C., Le Faou, P. & Angrand, P. O. Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Molecular & cellular proteomics : MCP* **10**, M110 002642, doi:10.1074/mcp.M110.002642 (2011).

127 Levine, S. S. *et al.* The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* **22**, 6070-6078 (2002).

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

129 Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445-1459, doi:10.1016/j.cell.2014.05.004 (2014).

130 Barrero, M. J. & Izpisua Belmonte, J. C. Polycomb complex recruitment in pluripotent stem cells. *Nature cell biology* **15**, 348-350, doi:10.1038/ncb2723 (2013).

131 He, J., Kallin, E. M., Tsukada, Y. & Zhang, Y. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). *Nature structural & molecular biology* **15**, 1169-1175, doi:10.1038/nsmb.1499 (2008).

132 Trojer, P. *et al.* L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Molecular cell* **42**, 438-450, doi:10.1016/j.molcel.2011.04.004 (2011).

133 Qin, J. *et al.* The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. *Cell Stem Cell* **11**, 319-332, doi:10.1016/j.stem.2012.06.002 (2012).

134 Ang, Y. S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183-197, doi:10.1016/j.cell.2011.03.003 (2011).

135 Trievel, R. C. & Shilatifard, A. WDR5, a complexed protein. *Nature structural & molecular biology* **16**, 678-680, doi:10.1038/nsmb0709-678 (2009).

136 Amati, B. *et al.* Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. *Nature* **359**, 423-426, doi:10.1038/359423a0 (1992).

137 Attwooll, C. *et al.* A novel repressive E2F6 complex containing the polycomb group protein, EPC1, that interacts with EZH2 in a proliferation-specific manner. *The Journal of biological chemistry* **280**, 1199-1208, doi:10.1074/jbc.M412509200 (2005).

138 Hayakawa, T. & Nakayama, J. Physiological roles of class I HDAC complex and histone demethylase. *Journal of biomedicine & biotechnology* **2011**, 129383, doi:10.1155/2011/129383 (2011).

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

140 Mohd-Sarip, A., Cleard, F., Mishra, R. K., Karch, F. & Verrijzer, C. P. Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes & development* **19**, 1755-1760, doi:10.1101/gad.347005 (2005).

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

142 Schwartz, Y. B. *et al.* Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nature genetics* **38**, 700-705, doi:10.1038/ng1817 (2006).

143 Tolhuis, B. *et al.* Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. *Nature genetics* **38**, 694-699, doi:10.1038/ng1792 (2006).

144 Wang, L. *et al.* Hierarchical recruitment of polycomb group silencing complexes. *Molecular cell* **14**, 637-646, doi:10.1016/j.molcel.2004.05.009 (2004).

145 Squazzo, S. L. *et al.* Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome research* **16**, 890-900, doi:10.1101/gr.5306606 (2006).

146 Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T. & Pasini, D. Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic acids research* **40**, 3403-3418, doi:10.1093/nar/gkr1290 (2012).

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

148 Ren, X. & Kerppola, T. K. REST interacts with Cbx proteins and regulates polycomb repressive complex 1 occupancy at RE1 elements. *Mol Cell Biol* **31**, 2100-2110, doi:10.1128/MCB.05088-11 (2011).

149 Yu, M. *et al.* Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Molecular cell* **45**, 330-343, doi:10.1016/j.molcel.2011.11.032 (2012).

150 Herranz, N. *et al.* Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. *Mol Cell Biol* **28**, 4772-4781, doi:10.1128/MCB.00323-08 (2008).

151 Villa, R. *et al.* Role of the polycomb repressive complex 2 in acute promyelocytic leukemia. *Cancer cell* **11**, 513-525, doi:10.1016/j.ccr.2007.04.009 (2007).

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

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

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

155 Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes & development* **25**, 1010-1022, doi:10.1101/gad.2037511 (2011).

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

157 Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & development* **17**, 1870-1881, doi:10.1101/gad.1110503 (2003).

158 Min, J., Zhang, Y. & Xu, R. M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & development* **17**, 1823-1828, doi:10.1101/gad.269603 (2003).

159 Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & development* **24**, 265-276, doi:10.1101/gad.544410 (2010).

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

161 Hunkapiller, J. *et al.* Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. *PLoS Genet* **8**, e1002576, doi:10.1371/journal.pgen.1002576 (2012).

162 Long, H. K., Blackledge, N. P. & Klose, R. J. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society transactions* **41**, 727-740, doi:10.1042/BST20130028 (2013).

163 Riising, E. M. *et al.* Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Molecular cell*, doi:10.1016/j.molcel.2014.06.005 (2014).

164 Cooper, S. *et al.* Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell reports* **7**, 1456-1470, doi:10.1016/j.celrep.2014.04.012 (2014).

165 Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nature structural & molecular biology* **21**, 569-571, doi:10.1038/nsmb.2833 (2014).

166 Endoh, M. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**, 1513-1524 (2008).

167 Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a polycomb group protein complex. *Science* **306**, 1574-1577, doi:10.1126/science.1100576 (2004).

168 Bantignies, F. *et al.* Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell* **144**, 214-226, doi:10.1016/j.cell.2010.12.026 (2011).

169 Endoh, M. *et al.* Histone H2A Mono-Ubiquitination Is a Crucial Step to Mediate PRC1-Dependent Repression of Developmental Genes to Maintain ES Cell Identity. *Plos Genetics* **8**, e1002774, doi:ARTN e1002774

DOI 10.1371/journal.pgen.1002774 (2012).

170 Francis, N. J., Saurin, A. J., Shao, Z. & Kingston, R. E. Reconstitution of a functional core polycomb repressive complex. *Molecular cell* **8**, 545-556 (2001).

171 Yuan, W. *et al.* Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* **337**, 971-975, doi:10.1126/science.1225237 (2012).

172 Min, I. M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes & development* **25**, 742-754, doi:10.1101/gad.2005511 (2011).

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

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

175 Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-956, doi:10.1016/j.cell.2005.08.020 (2005).

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

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

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

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

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

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

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

183 Wang, S. *et al.* Polycomblike-2-deficient mice exhibit normal left-right asymmetry. *Developmental dynamics : an official publication of the American Association of Anatomists* **236**, 853-861, doi:10.1002/dvdy.21070 (2007).

184 Voncken, J. W. *et al.* Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2468-2473, doi:10.1073/pnas.0434312100 (2003).

185 Posfai, E. *et al.* Polycomb function during oogenesis is required for mouse embryonic development. *Genes & development* **26**, 920-932, doi:10.1101/gad.188094.112 (2012).

186 Hisada, K. *et al.* RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. *Mol Cell Biol* **32**, 1139-1149, doi:10.1128/MCB.06441-11 (2012).

187 Katoh-Fukui, Y. *et al.* Male-to-female sex reversal in M33 mutant mice. *Nature* **393**, 688-692, doi:10.1038/31482 (1998).

188 Koren, A. & McCarroll, S. A. Random replication of the inactive X chromosome. *Genome research* **24**, 64-69, doi:10.1101/gr.161828.113 (2014).

189 Akasaka, T. *et al.* A role for mel-18, a Polycomb group-related vertebrate gene, during theanteroposterior specification of the axial skeleton. *Development* **122**, 1513-1522 (1996).

190 van der Lugt, N. M. *et al.* Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes & development* **8**, 757-769 (1994).

191 Akasaka, T. *et al.* Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development* **128**, 1587-1597 (2001).

192 Su, I. H. *et al.* Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nature immunology* **4**, 124-131, doi:10.1038/ni876 (2003).

193 Xie, H. *et al.* Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* **14**, 68-80, doi:10.1016/j.stem.2013.10.001 (2014).

194 Park, I. K. *et al.* Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302-305, doi:10.1038/nature01587 (2003).

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

196 Wang, L., Jin, Q., Lee, J. E., Su, I. H. & Ge, K. Histone H3K27 methyltransferase Ezh2 represses Wnt genes to facilitate adipogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 7317-7322, doi:10.1073/pnas.1000031107 (2010).

197 Caretti, G., Di Padova, M., Micales, B., Lyons, G. E. & Sartorelli, V. The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes & development* **18**, 2627-2638, doi:10.1101/gad.1241904 (2004).

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

199 Pereira, C. F. *et al.* ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. *Cell Stem Cell* **6**, 547-556, doi:10.1016/j.stem.2010.04.013 (2010).

200 Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483**, 598-602, doi:10.1038/nature10953 (2012).

201 Fragola, G. *et al.* Cell reprogramming requires silencing of a core subset of polycomb targets. *PLoS Genet* **9**, e1003292, doi:10.1371/journal.pgen.1003292 (2013).

202 Supic, G., Jagodic, M. & Magic, Z. Epigenetics: a new link between nutrition and cancer. *Nutrition and cancer* **65**, 781-792, doi:10.1080/01635581.2013.805794 (2013).

203 Haltiwanger, R. S., Blomberg, M. A. & Hart, G. W. Glycosylation of nuclear and cytoplasmic proteins. Purification and characterization of a uridine diphospho-N-acetylglucosamine:polypeptide beta-N-acetylglucosaminyltransferase. *The Journal of biological chemistry* **267**, 9005-9013 (1992).

204 Holt, G. D. & Hart, G. W. The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, O-linked GlcNAc. *The Journal of biological chemistry* **261**, 8049-8057 (1986).

205 Kreppel, L. K. & Hart, G. W. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. *The Journal of biological chemistry* **274**, 32015-32022 (1999).

206 Gao, Y., Wells, L., Comer, F. I., Parker, G. J. & Hart, G. W. Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *The Journal of biological chemistry* **276**, 9838-9845, doi:10.1074/jbc.M010420200 (2001).

207 Issad, T. & Kuo, M. O-GlcNAc modification of transcription factors, glucose sensing and glucotoxicity. *Trends in endocrinology and metabolism: TEM* **19**, 380-389, doi:10.1016/j.tem.2008.09.001 (2008).

208 Fardini, Y., Dehennaut, V., Lefebvre, T. & Issad, T. O-GlcNAcylation: A New Cancer Hallmark? *Frontiers in endocrinology* **4**, 99, doi:10.3389/fendo.2013.00099 (2013).

209 McClain, D. A. *et al.* Altered glycan-dependent signaling induces insulin resistance and hyperleptinemia. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10695-10699, doi:10.1073/pnas.152346899 (2002).

210 Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Science's STKE : signal transduction knowledge environment* **2005**, re13, doi:10.1126/stke.3122005re13 (2005).

211 Goto, H. *et al.* A single-point mutation in HCF causes temperature-sensitive cell-cycle arrest and disrupts VP16 function. *Genes & development* **11**, 726-737 (1997).

212 Hanover, J. A. A versatile sugar transferase makes the cut. *Cell* **144**, 321-323, doi:10.1016/j.cell.2011.01.025 (2011).

213 Yang, X., Zhang, F. & Kudlow, J. E. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* **110**, 69-80 (2002).

214 Jang, H. *et al.* O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. *Cell Stem Cell* **11**, 62-74, doi:10.1016/j.stem.2012.03.001 (2012).

215 Love, D. C., Krause, M. W. & Hanover, J. A. O-GlcNAc cycling: emerging roles in development and epigenetics. *Seminars in cell & developmental biology* **21**, 646-654, doi:10.1016/j.semcdb.2010.05.001 (2010).

216 Kim, H. S. *et al.* Excessive O-GlcNAcylation of proteins suppresses spontaneous cardiogenesis in ES cells. *FEBS letters* **583**, 2474-2478, doi:10.1016/j.febslet.2009.06.052 (2009).

217 Comer, F. I. & Hart, G. W. Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* **40**, 7845-7852 (2001).

218 Sakabe, K., Wang, Z. & Hart, G. W. Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19915-19920, doi:10.1073/pnas.1009023107 (2010).

219 Zhang, S., Roche, K., Nasheuer, H. P. & Lowndes, N. F. Modification of histones by sugar beta-N-acetylglucosamine (GlcNAc) occurs on multiple residues, including histone H3

serine 10, and is cell cycle-regulated. *The Journal of biological chemistry* **286**, 37483-37495, doi:10.1074/jbc.M111.284885 (2011).

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

221 Gambetta, M. C., Oktaba, K. & Muller, J. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science* **325**, 93-96, doi:10.1126/science.1169727 (2009).

222 Chu, C. S. *et al.* O-GlcNAcylation regulates EZH2 protein stability and function. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1355-1360, doi:10.1073/pnas.1323226111 (2014).

223 O'Donnell, N., Zachara, N. E., Hart, G. W. & Marth, J. D. Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol Cell Biol* **24**, 1680-1690 (2004).

224 Shafi, R. *et al.* The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5735-5739, doi:10.1073/pnas.100471497 (2000).

225 Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell cycle* **8**, 1698-1710 (2009).

226 Lorsbach, R. B. *et al.* TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* **17**, 637-641, doi:10.1038/sj.leu.2402834 (2003).

227 Ono, R. *et al.* LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer research* **62**, 4075-4080 (2002).

228 Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes & development* **25**, 2436-2452, doi:10.1101/gad.179184.111 (2011).

229 Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343-348, doi:10.1038/nature10066 (2011).

230 Ko, M. *et al.* Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**, 122-126, doi:10.1038/nature12052 (2013).

231 Xu, Y. *et al.* Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell* **151**, 1200-1213, doi:10.1016/j.cell.2012.11.014 (2012).

232 He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307, doi:10.1126/science.1210944 (2011).

233 Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-717, doi:10.1038/nature05987 (2007).

234 Szulwach, K. E. *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet* **7**, e1002154, doi:10.1371/journal.pgen.1002154 (2011).

235 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-495, doi:10.1038/nature10716 (2011).

236 Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature reviews. Molecular cell biology* **14**, 341-356, doi:10.1038/nrm3589 (2013).

237 Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome biology* **12**, R54, doi:10.1186/gb-2011-12-6-r54 (2011).

238 Serandour, A. A. *et al.* Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic acids research* **40**, 8255-8265, doi:10.1093/nar/gks595 (2012).

239 Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368-1380, doi:10.1016/j.cell.2012.04.027 (2012).

240 Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389-393, doi:10.1038/nature09934 (2011).

241 Williams, K., Christensen, J. & Helin, K. DNA methylation: TET proteins-guardians of CpG islands? *EMBO reports* **13**, 28-35, doi:10.1038/embor.2011.233 (2012).

242 Dawlaty, M. M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**, 166-175, doi:10.1016/j.stem.2011.07.010 (2011).

243 Dawlaty, M. M. *et al.* Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Developmental cell* **24**, 310-323, doi:10.1016/j.devcel.2012.12.015 (2013).

244 Rudenko, A. *et al.* Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109-1122, doi:10.1016/j.neuron.2013.08.003 (2013).

245 Ko, M. *et al.* Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14566-14571, doi:10.1073/pnas.1112317108 (2011).

246 Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606-610, doi:10.1038/nature10443 (2011).

247 Ficz, G. & Gribben, J. G. Loss of 5-hydroxymethylcytosine in cancer: Cause or consequence? *Genomics*, doi:10.1016/j.ygeno.2014.08.017 (2014).

248 Huang, H. *et al.* TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11994-11999, doi:10.1073/pnas.1310656110 (2013).

249 Langemeijer, S. M. *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nature genetics* **41**, 838-842, doi:10.1038/ng.391 (2009).

250 Kudo, Y. *et al.* Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer science* **103**, 670-676, doi:10.1111/j.1349-7006.2012.02213.x (2012).

251 Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer cell* **19**, 17-30, doi:10.1016/j.ccr.2010.12.014 (2011).

252 Losman, J. A. & Kaelin, W. G., Jr. What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes & development* **27**, 836-852, doi:10.1101/gad.217406.113 (2013).

253 Jung, H. R., Pasini, D., Helin, K. & Jensen, O. N. Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Molecular & cellular proteomics : MCP* **9**, 838-850, doi:10.1074/mcp.M900489-MCP200 (2010).

254 Wu, H. *et al.* Histone methyltransferase G9a contributes to H3K27 methylation in vivo. *Cell research* **21**, 365-367, doi:10.1038/cr.2010.157 (2011).

255 Tachibana, M. *et al.* Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes & development* **19**, 815-826, doi:10.1101/gad.1284005 (2005).

256 Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *The Journal of biological chemistry* **276**, 25309-25317, doi:10.1074/jbc.M101914200 (2001).

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

258 Bonaldi, T., Regula, J. T. & Imhof, A. The use of mass spectrometry for the analysis of histone modifications. *Methods Enzymol* **377**, 111-130, doi:10.1016/S0076-6879(03)77006-2 (2004).

259 Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protoc* **2**, 933-938, doi:10.1038/nprot.2007.106 (2007).

260 Loyola, A., Bonaldi, T., Roche, D., Imhof, A. & Almouzni, G. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Molecular cell* **24**, 309-316, doi:10.1016/j.molcel.2006.08.019 (2006).

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

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

263 Yuan, W. *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *The Journal of biological chemistry* **286**, 7983-7989, doi:10.1074/jbc.M110.194027 (2011).

264 Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283, doi:10.1038/nature09692 (2011).

265 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

266 Bowers, E. M. *et al.* Virtual ligand screening of the p300/CBP histone acetyltransferase: identification of a selective small molecule inhibitor. *Chemistry & biology* **17**, 471-482, doi:10.1016/j.chembiol.2010.03.006 (2010).

267 Kreppel, L. K., Blomberg, M. A. & Hart, G. W. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *The Journal of biological chemistry* **272**, 9308-9315 (1997).

268 Chou, T. Y., Hart, G. W. & Dang, C. V. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *The Journal of biological chemistry* **270**, 18961-18965 (1995).

269 Hassig, C. A., Fleischer, T. C., Billin, A. N., Schreiber, S. L. & Ayer, D. E. Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell* **89**, 341-347 (1997).

270 Myers, S. A., Panning, B. & Burlingame, A. L. Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 9490-9495, doi:10.1073/pnas.1019289108 (2011).

271 Shafi, R. *et al.* The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc Natl Acad Sci U S A* **97**, 5735-5739, doi:10.1073/pnas.100471497 (2000).

272 Love, D. C. *et al.* Dynamic O-GlcNAc cycling at promoters of Caenorhabditis elegans genes regulating longevity, stress, and immunity. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 7413-7418, doi:10.1073/pnas.0911857107 (2010).

273 Hashimoto, H., Vertino, P. M. & Cheng, X. Molecular coupling of DNA methylation and histone methylation. *Epigenomics* **2**, 657-669 (2010).

274 Smith, A. G. & Hooper, M. L. Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Developmental biology* **121**, 1-9 (1987).

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

276 Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *Embo J* **25**, 3110-3122, doi:10.1038/sj.emboj.7601187 (2006).

277 de Boer, E. *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proceedings of the*

National Academy of Sciences of the United States of America **100**, 7480-7485, doi:10.1073/pnas.1332608100 (2003).

278 Wisniewski, J. R., Zougman, A. & Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *Journal of proteome research* **8**, 5674-5678, doi:10.1021/pr900748n (2009).

279 Olsen, J. V. *et al.* Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular & cellular proteomics : MCP* **4**, 2010-2021, doi:10.1074/mcp.T500030-MCP200 (2005).

280 Olsen, J. V. *et al.* Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular & cellular proteomics : MCP* **4**, 2010-2021, doi:10.1074/mcp.T500030-MCP200 (2005).

281 Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805, doi:10.1021/pr101065j (2011).

282 Arike, L. *et al.* Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. *Journal of proteomics* **75**, 5437-5448, doi:10.1016/j.jprot.2012.06.020 (2012).

283 Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906, doi:10.1038/nprot.2007.261 (2007).

284 Cox, J. *et al.* A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc* **4**, 698-705, doi:10.1038/nprot.2009.36 (2009).

285 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).

286 Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805, doi:10.1021/pr101065j (2011).

287 Thakur, S. S. *et al.* Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Molecular & cellular proteomics : MCP* **10**, M110 003699, doi:10.1074/mcp.M110.003699 (2011).

288 Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics* **39**, 457-466, doi:10.1038/ng1990 (2007).

289 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

290 Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC bioinformatics* **11**, 237, doi:10.1186/1471-2105-11-237 (2010).

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

292 Scharf, A. N. *et al.* Monomethylation of lysine 20 on histone H4 facilitates chromatin maturation. *Mol Cell Biol* **29**, 57-67, doi:10.1128/MCB.00989-08 (2009).

293 de Almeida, S. F. *et al.* Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nature structural & molecular biology* **18**, 977-983, doi:10.1038/nsmb.2123 (2011).

294 Herz, H. M., Hu, D. & Shilatifard, A. Enhancer malfunction in cancer. *Molecular cell* **53**, 859-866, doi:10.1016/j.molcel.2014.02.033 (2014).

295 Kreppel, L. K. & Hart, G. W. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. *J Biol Chem* **274**, 32015-32022 (1999).

296 Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N. & Herr, W. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes & development* **17**, 896-911, doi:10.1101/gad.252103 (2003).

297 Capotosti, F. *et al.* O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* **144**, 376-388, doi:10.1016/j.cell.2010.12.030 (2011).

298 Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**, 200-213, doi:10.1016/j.stem.2011.01.008 (2011).

299 Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & development* **25**, 679-684, doi:10.1101/gad.2036011 (2011).

300 Fujiki, R. *et al.* GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature* **480**, 557-560, doi:10.1038/nature10656 (2011).

301 Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *Embo J* **32**, 645-655, doi:10.1038/emboj.2012.357 (2013).

302 Griffith, L. S. & Schmitz, B. O-linked N-acetylglucosamine is upregulated in Alzheimer brains. *Biochemical and biophysical research communications* **213**, 424-431, doi:10.1006/bbrc.1995.2149 (1995).

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