PhD degree in Systems Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: BIO/11

Understanding the impact of replication stress on the expression of early genes in mouse embryonic stem cells

Gnocchi Andrea

Istituto FIRC di oncologia molecolare (IFOM), Milan

University of Milan, Milan

Matricola n. R11740

Tutor: Prof. Vincenzo Costanzo, MD-PhD

Istituto FIRC di oncologia molecolare (IFOM), Milan

University of Milan, Milan

PhD Coordinator: Prof. Giuseppe Viale

Anno accademico 2019-2020

Table of Contents

List of abbreviations

2C: 2-cells stage

2i: 2-inhibitors (ERK inhibitor and GSK3 inhibitor)

APH: aphidicolin

ChIP: chromatin immunoprecipitation

cPRC1: canonical Polycomb repressive complex 1

DEGs: differentially expressed genes

DSB: double strand break

ESCs: embryonic stem cells

ExE: extraembryonic ectoderm

FSHD: facioscapulohumeral dystrophy

H2AK119Ub: histone H2A monoubiquitination on lysine 119

H3K27me3: histone H3 trimethylation on lysine 27

HR: homologous recombination

iATR: inhibitor of ATR

ICM: inner cell mass

IPSCs: induced pluripotent stem cells

LIF: leukemia inhibitory factor

LTR: long terminal repeat

MEFs: mouse embryonic fibroblasts

MERVL: mouse endogenous retroviruses with leucine tRNA primer

ORC: origin recognition complex

PRC1: Polycomb repressive complex 1

PRC2: Polycomb repressive complex 2

PrE: primitive endoderm

qPCR: quantitative polymerase chain reaction

RS: DNA replication stress

RT: retrotranscription

TE: trophectoderm

TSS: transcription start sites

VE: visceral endoderm

vPRC1: variant Polycomb repressive complex 1

Figures Index

Abstract

Embryonic stem cells (ESCs) are characterized by a rapid cell cycle, which leads to high DNA replication stress (RS) in otherwise unperturbed conditions. The mechanisms that ESCs adopt to cope with their endogenous RS, however, remain to this day elusive. In our recent work we demonstrated that the activation of the checkpoint kinase ATR in response to RS leads to a broad activation of 2-cells stage specific genes in mouse ESCs. This response relies on the up-regulation of *Dux*, a transcription factor encoded in a macrosatellite sequence repeated in tandem. *Dux* is repressed by variant Polycomb repressive complex 1 (vPRC1) in unperturbed ESCs, independently from PRC2 presence.

Here we demonstrate that RS causes a major rearrangement of both PRC1 and PRC2 in ESCs nuclei, resulting in a major loss of both repressive marks in correspondence to target promoters. Surprisingly, *Dux* undergoes an increase in vPRC1 occupancy upon RS in an ATR-dependent manner, possibly due to PRC1 involvement in the replication of highly repeated DNA sequences. More interestingly, *Dux* activation upon RS requires the presence of PRC2. This result is possibly due to PRC2 proved role in the processing of stalled replication forks, which are the main structure signaling RS. In agreement to this data, also the fork remodeling translocases HLTF and ZRANB3 displayed an effect in *Dux* activation following RS.

Taken together, our results show that the up-regulation of 2-cells genes following RS not only requires ATR activation, but also downstream remodeling processes.

1 - Introduction

1.1 – Early embryonic development

All the complexity of pluricellular organisms originates from a single cell, the zygote. This cell, derived by the fusion of oocyte and sperm, has the potential to generate the entire organism, together with all extraembryonic tissues, a property known as totipotency (1,2). During embryonic development, cells generated from the zygote by subsequent rounds of duplication undergo a series of fate commitment steps, which restrict progressively their developmental potential, namely the ability to give rise to different tissues. Embryonic development differs among species, but here we will focus mostly on mouse (*Mus musculus*), which constitutes the most common model system for embryonic development among placental mammals, for reasons of simplicity and ethical constraints. After fertilization, the newly formed zygote undergoes many rounds of cell division, giving rise to an increasing number of daughter cells called blastomeres. The structure formed by the blastomeres during the first stages of development takes the name of morula (1,2).

In the case of mouse development, the embryo, enveloped in a protective glycoproteic layer called zona pellucida, starts cavitating around 3 days post fertilization (or post coitum, named also stage E3) forming a liquid-filled space surrounded by cells (Fig. 1). Together with the formation of the cavity, named blastocoel, the blastomeres constituting the morula undergo the first step of lineage commitment, forming a compact lump of cells called the inner cell mass (ICM), and a cell layer, named trophectoderm (TE), which encloses both ICM and blastocoel (Fig. 1) (1,2). This new structure is now called blastocyst or blastula. The TE is divided in polar TE, which is in contact with the ICM, and mural TE, which instead surrounds the blastocoel. Inside the TE, the ICM undergoes a subsequent step of lineage commitment, forming an additional epithelial layer in contact with the internal cavity, that is named primitive endoderm (PrE). Enclosed between the PrE and the polar TE

develops the epiblast, a mass of cells that is the main responsible for the following development of embryonic tissues (Fig. 1) (1,2).

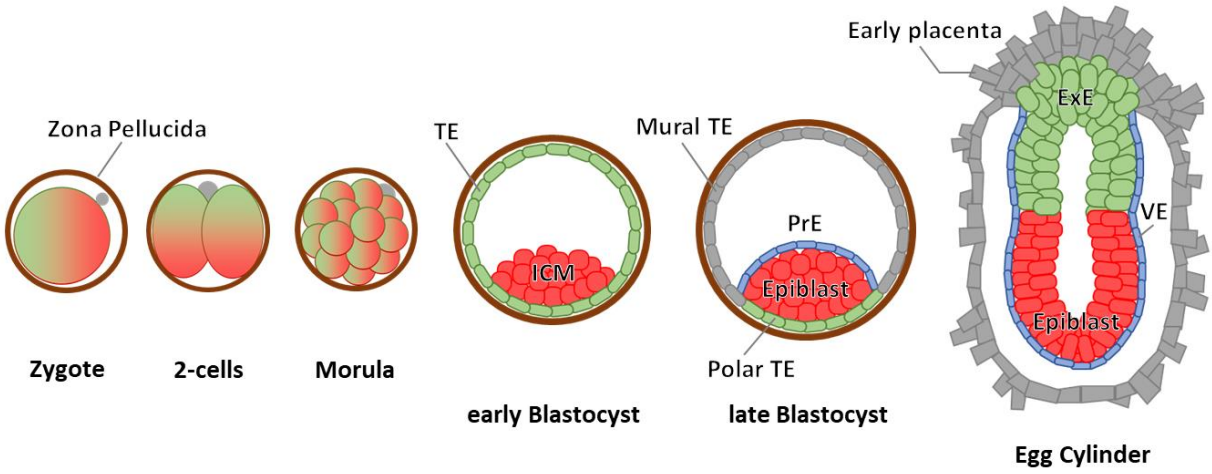

Figure 1 | Early embryonic development: Schematic depicting early embryonic development from Zygote formation to the appearance of the Egg cylinder. From left to right: the zygote undergoes the first cleavage step to become the 2-cells embryo. Following divisions form the morula which starts the first lineage commitment to create the cavitated blastocyst, composed by Inner cell mass (ICM) and trophectoderm (TE). In the blastocyst the TE differentiates in polar TE and mural TE, while the ICM divides in Primitive endoderm (PrE) and Epiblast. Finally, the blastocyst hatches from the zona pellucida and implants in the uterus. The mural TE forms the first placental tissues while polar TE becomes the Extraembryonic ectoderm (ExE). The Epiblast lumen fuses with the ExE lumen and the PrE generates the Visceral endoderm (VE).

Around the stage E4 – E4.5 the blastocyst hatches from the zona pellucida and starts the invasion of the maternal tissues. At this point the cells constituting the polar TE assume a columnar morphology while the mural TE proceeds in the actual invasion of the uterine tissue (3). At the same time, also the epiblast cells, surrounded by the PrE, are reshaped to a columnar conformation, and forms an internal cavity. The polar TE differentiates into extraembryonic ectoderm (ExE), which will contribute to the embryonic part of the placenta, while cells of the mural TE start to grow in size in order to form the invading trophoblast giant cells, which will constitute the main part of the placenta (3). The ExE invaginates, forming a second cavity, while the PrE grows to cover both epiblast and ExE. Finally, the cavities formed by ExE and epiblast fuse to form a single lumen around stage E6.5 – E7.5, resulting in a structure called egg cylinder (Fig. 1), surrounded by the PrE and connected to the maternal tissues by the invading trophoblasts (4). The egg cylinder,

connected to the uterus by the placenta, will then proceed to develop into the entire embryo.

1.1.1 – Zygote genome activation and the 2-cells stage

At the time of fertilization, both parental genomes are transcriptionally silent. All cellular functions indispensable for the survival of the newly formed zygote are carried on by transcripts and proteins accumulated in the oocyte during its maturation (5). The event that sets the start for zygotic transcription is known as zygotic genome activation (ZGA). In different animals the number of cell division cycles before ZGA can vary broadly. As an example, in humans transcription takes place for the first time between the 4 and 8-cells stage, while in mouse ZGA starts much earlier, when the embryo is constituted by only 2 cells (2-cells or 2C stage) (5). ZGA is a fundamental step in the transition from oocyte to zygote, which leads a highly specialized cell to acquire a precise developmental program, responsible for the growth and development of the entire embryo. This process entails profound modifications of the chromatin landscape of both parental genomes. Over few division cycles, the maternal pronucleus loses most repressive histone marks, while the paternal pronucleus presents itself in an almost naïve configuration, due to the substitution of histones with protamines during sperm maturation (5). Similarly, DNA methylation is removed from the genome with the exception of few imprinted loci (5). This wave of demethylation together with loss of repressive chromatin marks leads to a reduction of parental heterochromatin (5). In mouse, ZGA takes place at a very early stage, with some transcription already detectable at the 1-cell stage (5). This early transcription is called minor ZGA and its biological significance is still debated, due to the low level of transcription and the apparent non-specificity of the transcripts produced in this phase. Among the features suggesting that transcription happening during this stage may be the result of non-specific genome de-repression, there is the limited efficiency of splicing and

polyadenylation of the produced transcripts. However, some have hypothesized that poor splicing and impaired nonsense mediated decay may favor the expression of intronless genes, such as the double homeobox transcription factor *Dux*, considered to be the master regulator of ZGA (5). Most of ZGA takes place at the 2-cells stage and is called major ZGA or simply ZGA. Major ZGA involves more than 3000 genes, among which the aforementioned transcription factor *Dux*, its main target *Zscan4* (Zinc finger and SCAN domain containing 4), and others, such as *Tcstv1* and *Tcstv3* (2-Cell-Stage, Variable group, member 1-3) (5,6). One of the peculiarities of major ZGA is the expression of endogenous retroelement called MERVL (Mouse Endogenous Retroviruses with Leucine tRNA primer). These elements constitute the remnants of ancient viruses integrated in the genome of all placental mammals, and are generally tightly repressed in differentiated cells. MERVL are generally incomplete and cannot produce functional viral particles, but in some cases they have been coopted by mammalian genes to be used as promoters (7). The complex regulation of ZGA and the mechanisms that it sets in motion are still for the most part unknown. It is important to remember though that around this stage the first transcripts of genes such as *Pou5f1* (encoding the octamer binding transcription factor OCT4), *Sox2* (SRY-box 2) and *Nanog* appear (5). These genes will drive the next steps of development, setting in place the pathways that will originate the entire embryo.

1.1.2 - The blastocyst

After ZGA, the embryo undergoes several rounds of divisions forming the morula. At this stage the first lineage commitment takes place, separating the group of cells that will produce the extraembryonic tissues and the one that will develop into the embryo. This separation arises progressively through a series of cell divisions that segregate cell fate determinants asymmetrically (1). In the early morula, this results in the concomitant presence of cells expressing both TE determining genes (*Cdx2* and *Elf5*) and ICM-specific

factors (*Pou5f1*, *Nanog* and *Sox2*) (1,2). Around the stage of 8-cells, the morula compacts, and single cells are not visible anymore. This process depends upon the expression of E-cadherin (encoded by *Cdh1*) that forms adherens junctions (8). The formation of adherens junctions is fundamental for the following steps of development, since it allows the establishment of polarity required for the differentiation of the TE, and lays the ground for the formation of the blastocoel in the next stage (8). In the morula, the acquisition of the two cell fates (TE or ICM) is not complete yet, and cells are still believed to be totipotent (1,2). The progressive separation of the two lineages forms a structure characterized by a pluripotent ICM surrounded by a multipotent epithelial layer (TE). This structure is the blastocyst, which, thanks to the ability of the TE to pump liquids in a highly directional manner, will form the blastocoel (1,8). During the formation of the cavity, the ICM remains in contact with a portion of the TE, determining the future differentiation between polar and mural TE. Inside the cavity, the ICM undergoes another lineage commitment step, generating PrE and epiblast, which will then undergo a series of dramatic rearrangements that will lead to the conversion of the invading blastocyst in egg cylinder.

1.1.3 - Inner cell mass, primitive endoderm and epiblast

Around the stage of 32-cells the embryo starts to cavitate, assuming the characteristic structure of the blastocyst. At this point the ICM is visible as a lump of cells gathered against the internal side of the TE layer. The ICM is characterized by the expression of the pluripotency factors OCT4, NANOG and SOX2, which constitute the core of the ICM transcriptional network. Embryos lacking any of these factors are able to form the blastocyst, although after implantation stage they show complete absence of the epiblast (9,10). By stage E4.5 the ICM undergoes a second step of cell fate determination, which separates the so called epiblast, responsible for the formation of the embryo proper, and the PrE, which will originate a second layer of extraembryonic tissues (1). Even in this case,

the exact determinants of the transition are still uncertain. One hypothesis is that it may originate from an initial heterogeneity of the ICM. More specifically, some of the cells forming the ICM express factors which are required for the acquisition of PrE-identity through the up-regulation of GATA4 and GATA6 (GATA binding protein 4/6) and are, therefore, primed for differentiation into PrE. According to another hypothesis, the difference among cells of the PrE and the epiblast could arise through a series of asymmetrical divisions in a time-dependent fashion (1). PrE and epiblast become definitively separated around the stage of 64-cells, when the expression of NANOG and GATA6 become mutually exclusive. Both OCT4 and NANOG expression is restricted to the epiblast, while PrE cells lose the ability to maintain pluripotency and undergo differentiation, acquiring epithelial features (1). After embryo implantation, the PrE develops into visceral endoderm (VE) which will be responsible for the development of both extraembryonic tissues such as the yolk sac, and embryonic tissues, like the gut endoderm (1). On the other hand, around the time of implantation the epiblast organizes itself in a rosette-like structure. Cellular compartmentalization driven by actomyosin fibers and adherens junctions allow the epiblast to form a central lumen (8). This cavity then expands through a process of hollowing, resulting in the final fusion between the epiblast and the ExE lumens in a so called proamniotic cavity (1,8). The epiblast will give rise, during the following step of gastrulation, to the three main lineages of the embryo, namely ectoderm, mesoderm and endoderm (1,8).

1.1.4 - Trophectoderm and early extra-embryonic tissues

The differentiation of the TE is the first lineage commitment step in mammalian development and takes place between the 8 and 32-cells stages. The transcription factor CDX2 (caudal type homeobox 2) is the main determinant of TE development and is necessary for both TE genes expression and pluripotency genes suppression (1). *Cdx2*-null

embryos can develop normally only until early blastocyst, when eventually the blastocyst cavity collapses leading to embryonic death (1). The progressive accumulation of CDX2-positive cells in the external part of the morula seems to be dependent on the portioning defective (Par) group of genes, a family of genes involved in many processes involved in cell polarity (1). Cell polarization increase CDX2 expression in external cells, which in turn furtherly promotes polarization (1). The resulting positive feed-back loop leads to the segregation of the highly polarized CDX2-positive cells in the external part of the embryo, while CDX2-negative cells accumulate in the innermost part. In addition, a mechanism dependent on the Hippo pathway and the transcription factor TEAD4 (TEA Domain transcription factor 4) allows cells to sense their surroundings and regulate *Cdx2* expression in response to their position inside the embryo (1). The newly formed TE highly expresses E-cadherin, which leads to the formation of adherens junctions and then tight junctions (8). Tight junctions form a physical barrier impenetrable to fluids, thus isolating the internal part of the morula from the outside environment. Accumulation of positive ions at the basal side of TE cells then drives fluid import inside the blastocyst and leads to the formation of the blastocoel (2,8). Around stage E4 -E4.5 the blastocyst hatches, breaking the zona pellucida to start the process of implantation. Polar TE cells, close to the epiblast, undergo a drastic morphological change, reshaping in a columnar morphology and forming a multilayered epithelium which gives rise to the ExE (1,8). Since only the polar TE, but not the mural TE, undergoes this change, many have hypothesized a role for the epiblast in this transformation. In the last steps of implantation, the ExE forms a lumen that is then fused with the epiblast cavity to form the proamniotic cavity, while the mural TE, characterized by low CDX2 expression, progresses in the invasion of the uterine wall, giving rise to the first cells of the placenta.

1.2 - Embryonic Stem cells

The first reports that the propagation of undifferentiated cells from mouse embryos was possible date back to 1970. Originally, intratesticular grafting of mouse embryos allowed researcher to observe that embryos could differentiate in a variegated and disorganized ensemble of tissues called teratomas (11). The observation that some of these tumors, named teratocarcinomas, could be transferred and propagated from mouse to mouse, highlighted the presence of an undifferentiated population of cells with the ability to survive for many generations a give rise to multiple lineages (12). In the early 1980s, it was finally possible to isolate this population of undifferentiated cells from mouse blastocysts and culture them *in vitro* (13). These cells proved able to grow for a much higher number of divisions with respect to normal primary cells if maintained in appropriate culture conditions. Subsequent studies demonstrated that these blastocyst-derived cells were able to complement the inability of tetraploid embryos to generate embryonic tissues, leading to the complete formation of a healthy mouse (14). For this ability to maintain their identity almost indefinitely and at the same time give rise to a wide range of cell lineages, these cells of embryonic origin are nowadays known as embryonic stem cells (ESCs). In the course of the 1990s further advances have been made and ESC lines have been driven also from human and other primates (15,16). However, in the following paragraphs we will focus mostly on mouse ESCs, which constitute the most commonly used and best studied type of ESCs.

1.2.1 – ESC lines establishment and culture conditions

Mouse ESCs are derived from the blastocyst-stage embryo and are generally considered a good *in vitro* approximation of the features that characterize the ICM, although differences exist between the two. It is in fact important to remember that, while ESCs can be propagated *in vitro* for a high number of passages, the ICM exists for a very short phase of

embryonic development. The process of ESCs derivation is surprisingly simple considering its importance for developmental research. Cavitating blastocysts are plated in tissue culture dishes where a feeder layer of mouse embryonic fibroblasts (MEFs) has been previously seeded (17). Blastocysts are kept growing on the feeder layer in common cell culture medium supplemented with serum and β-mercaptoethanol to maintain the correct reducing environment. After several days, the blastocysts hatch growing out of the zona pellucida and adhering to the feeder layer. The ICM outgrowths can then be harvested, disaggregated and re-seeded on a new feeder layer in order to grow single colonies of ESCs. The feature of ESCs to grow in compact colonies makes particularly easy to pick and isolate clonally pure lines, which can then be expanded for many generations. MEFs are critical for the maintenance of ESCs in an undifferentiated state through their ability to secrete LIF (Leukemia Inhibitory Factor), a cytokine that acts by binding the gp130 receptor and activates the STAT3 (Signal Transducer and Activator of Transcription 3) pathway (18,19). MEFs are instrumental to the survival of ESCs, however they are a source of technical issues during the performing of complex experiments. In fact, the presence of a contaminating cell line in co-culture with ESCs may affect the results obtained from techniques which rely on the extraction of material from bulk cultures. Many practical precautions can be taken to limit this problem, above all the use of mitomycin-C inactivated MEFs, which due to their inability to grow can be easily overcome in number by the fast-growing ESCs. As mentioned before, the trophic activity of MEFs on ESCs is mostly dependent on the secretion of LIF, as proven by the inability of *Lif*$^{-/-}$ MEFs to sustain the growth of ESCs (20). LIF is by itself able to sustain the ESCs ability to grow in an undifferentiated state and therefore can substitute MEFs feeders (21). However, in the last 20 years, several studies have highlighted the fact that multiple pathways concur to determine ESCs peculiar features. In particular, it has been demonstrated that ESCs are responsive to various members of the Wnt family (22).

Binding of Wnt to their Frizzled receptors leads to the inactivation of the GSK-3 kinase and the consequent accumulation of β-catenin inside the nucleus, where it can activate the transcription of target genes, among which *Stat3* (22). Direct inhibition of GSK-3 through the administration of small chemicals, such as CHIR-98014 (CT), have demonstrated to improve the expression of pluripotency genes in ESCs cultures and prevent their differentiation through the up-regulation of the Wnt/β-catenin pathway (23). In addition, studies using a mutant gp130 receptor, which lacked the ability to activate the ERK/MAPK pathway without affecting ESCs ability to induce STAT3 expression, underlined a peculiar behavior of ESCs. In fact, while in more differentiated cells the MAPK activation leads to sustained growth, in ESCs the inhibition of ERK/MAPK through the administration of the chemical inhibitor PD098059 (PD) improves STAT3 expression and overall increases their growth rate (24). Thanks to these advances, nowadays the medium formulation comprising LIF and the CT and PD inhibitors (known as 2 inhibitors medium or 2i medium)(25) is largely favored with respect to the use of feeder layers and allows easier handling of ESCs in more complex experimental designs.

1.2.2 – Self-renewal, pluripotency and embryonic contribution

ESCs take the definition of stem cells thanks to their ability to both generate identical copies of themselves and differentiate in a wide variety of lineages. The symmetrical generation of two undifferentiated daughter cells from an original stem cell takes the name of self-renewal. ESCs are peculiar in their ability to self-renew indefinitely, which makes them an excellent instrument for developing engineered models for genetical studies. As aforementioned, ESCs self-renewal relies on the signal provided by the LIF/gp130 and Wnt/β-catenin pathways, which act through *Stat3* to maintain ESCs undifferentiated nature. In addition to these extrinsic signals, also intrinsic determinants of self-renewal have been identified. Above all, there is the transcription factor OCT4, encoded by the gene

Pou5f1. As reported earlier, OCT4 is expressed in the ICM of mouse blastocysts while it is suppressed in the TE. *Pou5f1*$^{-/-}$ mice show defective epiblast formation and do not reach term, while *in vitro* culture of *Pou5f1*$^{-/-}$ blastocysts fails to produce ESCs and generates instead only cells with TE features (9). Genetic models able to conditionally shut down OCT4 expression have shown that in its absence cells tend to differentiate towards the TE lineage (26). At the same time, ectopic expression of OCT4 cannot suppress the differentiation due to LIF withdrawal (26), indicating that this transcription factor is necessary but not sufficient to maintain ESCs identity. Another determinant of ESCs self-renewal ability is the transcription factor SOX2, known to play a role in the transcription of many OCT4 targets (27). Embryos lacking SOX2 expression are able to implant, but by stage E6 they display complete loss of OCT4 in the epiblast (28). At the same time, *Sox2*$^{-/-}$ ICMs fail to generate ESCs when cultured *in vitro,* although in this case, together with TE cells, also epithelial cells with PrE features can be detected (28). The last among the key intrinsic determinants of ESCs identity is the transcription factor NANOG. As mentioned before, during embryonic development, NANOG expression appears later than OCT4 and it is restricted from morula to implantation (1,10,29). Different studies indicate that ectopic expression of NANOG is able to provide LIF-independent self-renewal, allowing ESCs to maintain their undifferentiated state in the presence of an antagonist of LIF receptor (29). Interestingly, in agreement to its LIF-independent action, NANOG does not regulate *Stat3* expression, nor it is regulated by LIF, proving the existence of parallel and independent pathways governing ESCs self-renewal ability (29). Finally, NANOG ectopic expression cannot suppress the TE differentiation induced by *Pou5f1* repression, indicating that this transcription factor requires OCT4 to exert its function (29). Together, OCT4, NANOG and SOX2 constitute the core of the so-called pluripotency transcriptional network, which

grants the maintenance of ESCs undifferentiated nature, in coordination with extrinsic cues such as LIF and serum.

As mentioned before, ESCs can be induced to differentiate upon removal of LIF from the culture medium. In these conditions, ESCs generate an heterogeneous mixture of cell types from all the three germ layers, namely ectoderm, mesoderm and endoderm (17,30). This property is known as pluripotency, since it entails the ability to produce all somatic cell types but not the extraembryonic tissues. Pluripotency is not to be confused with totipotency, that is instead the ability to give rise to all embryonic and extraembryonic tissues and is generally referred to cells that precede the ICM-TE separation. The process of differentiation requires a progressive loss in the expression of pluripotency factors, together with the activation of lineage specific factors. As a consequence of ESCs pluripotency, the injection of ESCs into mouse embryos results in their contribution to all embryonic compartments, resulting in the generation of chimeric mice (31). Moreover, ESCs are able to complement the inability of tetraploid embryos to generate embryonic tissues and produce viable offspring (14). This feature makes ESCs able to produce hypothetically any kind of adult tissue even *in vitro*, provided that the right differentiation protocol is developed. Collectively, the ability to self-renew indefinitely, together with the capacity to generate virtually any tissue of the body, makes ESCs an incredibly powerful model for the study of many biological processes well beyond their role as a model for developmental biology.

1.3 - 2C-like cells and expanded developmental potential

ESCs growing in serum/LIF conditions show a certain level of heterogeneity, due most likely to the extreme plasticity of this cell type (32). Interestingly, an early work from Falco and collaborators identified in ESCs colonies a small group of cells expressing *Zscan4*, using *in situ* RNA hybridization (33). As mentioned above, *Zscan4* is one of the genes expressed

during major ZGA at the 2-cells stage of mouse embryonic development, and it was one of the first 2C specific markers to be identified (33). Further work form Macfarlan and colleagues, using a fusion cassette comprising the 5'-LTR portion of a MERVL endogenous retrovirus and the fluorescent protein tdTomato, demonstrated the presence of a MERVL positive population in both ESCs and IPSCs (34). Following studies proved that both *Zscan4* and MERVL are expressed by the same sub-population of ESCs, and that these cells express other transcripts specific for the 2-cells stage. These cells were therefore named 2C-like cells (35). To this day, *Zscan4* and MERVL elements are the most widely used markers for 2C-like cells studies. The use of engineered systems comprising the *Zscan4* promoter or MERVL 5'-LTR portion fused with fluorescent proteins have allowed researchers to better characterize the features of these cells and the reason why they are present into ESCs cultures. In each ESC culture it is possible to find around 1-5% of cells expressing 2-cells stage markers (34,35). Interestingly, 2C-like cells are not a stable population, but each ESCs enters and exits this state cyclically (33,36). Studies taking advantage of the β-galactosidase marker proved that over the course of several cell cycles every ESC enters this state at least once, and that homogeneous ESC or 2C-like cultures are able to reconvert one into the other over the time, returning to the original proportion between ESCs and 2C-like cells (36). For their transient nature, 2C-like cells are sometimes also referred to as ESCs in a 2C-like state, underlying the ability of the two populations to interconvert one into the other. Interestingly, recent works using single cells transcriptomics have suggested the presence of an intermediate step during the passage from ESCs to 2C-like cells which does not coincide with the opposite passage from 2C-like cells back to ESCs (37,38). During ESC to 2C-like transition *Zscan4* expression seems to precede MERVL activation, indicating a precise hierarchy in 2C genes activation (37,38). The opposite passage, instead, appears to be mostly driven by a rapid degradation of *Dux* transcript, which is proposed to be the first

step in the down-regulation of 2C-specific genes (38). While a clear mechanism describing the interconversion of ESCs into 2C-like cells and *vice versa* is still missing, 2C-like cells have already shown interesting features. The cyclic transition through the 2C state has in fact been linked to an improvement of the genomic integrity of ESCs, mostly due to increased telomeres length in response to *Zscan4*, *Tcstv1* and *Tcstv3* expression (36,39,40). More importantly, 2C-like cells have demonstrated expanded developmental potential (34,41,42). Namely, while ESCs can contribute only to embryonic tissues when injected into morula stage embryos, injected 2C-like cells can develop in both embryonic and extraembryonic tissues. This feature has been proposed to depend on the more undifferentiated nature of 2C-like cells in comparison to ESCs, which approximates the nature of the blastomeres at the 2-cells stage. Thanks to this feature, 2C-like cells have been proposed to be totipotent (43). In this case, a single 2C-like cell should in theory be able to give rise to an entire organism. However, since the nature and the fine biology of these cells is still mostly obscure, we prefer to consider 2C-like cells to have an expanded developmental potential, limiting the use of the adjective totipotent for the zygote.

1.3.1 – 2-cells stage markers

Among 2-cells stage genes, *Zscan4* is one of the most studied. Its expression is limited to the 2-cells stage of mouse development, spermatogenesis and oogenesis (33). In ESCs cultures its expression is used as one of the main markers of 2C-like cells. Interestingly, transient bursts of *Zscan4* expression in ESCs have been linked to a general de-repression of heterochromatin, resembling the lose chromatin configuration of 2-cells embryos (35,44). At the same time, ZSCAN4 has been proposed to act at the site of shortened telomeres to favor their lengthening through sister chromatid exchange (36). In agreement with this hypothesis, ZSCAN4 has been shown to prevent DNA damage checkpoint activation at telomeres upon loss of the shelterin component TRF2 (Telomeric Repeat

Binding Factor 2) in ESCs (45). Recent studies have also suggested a more general role for ZSCAN4 in protecting DNA integrity by binding to microsatellite repeats, and stabilizing their interaction with nucleosomes (46). Other reports have instead proposed a role in transcriptional regulation for ZSCAN4, due to its ability to interact with various transcriptional regulators, such as TET2 (Tet Methylcytosine Dioxygenase 2), LSD1 (Lysine Demethylase 1A) and CtBP2 (C-Terminal Binding Protein 2) (47,48). Notwithstanding the uncertainty surrounding ZSCAN4 activity, its expression has proven to improve ESCs genomic stability and increase telomere length (36,39) and, at the same time, its transient expression during cell reprogramming seems to improve the quality of the IPSCs (Induced Pluripotent Stem Cells) obtained (49). Therefore, in addition to its role as a *bona fide* marker for 2-cells stage embryos and 2C-like cells, this gene appears to be of potential high interest for the study of genomic stability during the early stages of embryonic development.

MERVL retroelements are other peculiar markers of the 2-cells stage, useful to identify 2C-like cells. Approximately 40% of mouse and human genomes are constituted by interspersed repetitive elements, many of which are related to transposable elements. MERVL constitute a subclass of retrotransposons, segments of DNA able to copy and transfer themselves into other regions of the genome using an RNA intermediate. These sequences have been proposed to be remnants of ancient retroviral infections, due to their similarity with the basic structure of a retrovirus. MERVL are indeed constituted by a LTR (long terminal repeat) sequence, followed by the retroviral polygene *Gag* (50). In some cases, MERVL LTRs sequences have been coopted by endogenous genes, and are used as promoters for the expression of ZGA genes, generating chimeric transcripts (7). Both ZSCAN4 and DUX have been proposed to activate MERVL transcription in mouse by direct or indirect binding of its sequence (51,52). During ZGA, MERVL-driven gene expression is

considered the main regulator of zygotic transcription and their activation in ESCs is considered sufficient to acquire a 2C-like phenotype (34,35).

In addition to *Zscan4* and MERVL, many other genes are expressed in 2-cell stage embryos and 2C-like cells. A large part of these genes is poorly studied and their role in ZGA remains cryptic. Among the best known are the *Tcstv1* and *Tcstv3* genes (6). Ectopic expression of both these genes in ESCs increases telomeres length through sister chromatid exchange, similarly to what observed in the case of *Zscan4* (40). Differently from *Zscan4* expression though, their expression does not seem to induce transition to 2C-like cells, suggesting a specific role in telomeres metabolism for the products of these genes (40). Other genes associated to both the 2-cells stage and 2C-like cells are the ones from the Eif1a-like family, such as *Gm5662*, *Gm2022*, *Gm4027*, *Gm2016* and *Gm8300* (53). These proteins have been proposed to cause a general suppression of protein synthesis in 2C-like cells in a dominant-negative fashion, although the precise biological significance of this function remains unknown to this time (53). Finally, one of the most relevant markers of both the 2-cells stage and 2C-like cells is the transcription factor *Dux*, that will be better described in the following paragraph.

1.3.2 - The *Dux* gene family

Among all the genes defining the 2-cells stage of development, *Dux* (Double Homeobox) is probably the most peculiar one. In fact, this gene is inserted in a macrosatellite sequence, which is highly repeated in a head-to-tail orientation and organized in tandem. The original interest for this family of genes arose from the evident link between the human member of the family, namely *DUX4*, and a degenerative disease known as facioscapulohumeral muscular dystrophy (FSHD), which causes progressive loss of muscle tissue in the upper body (54). While in healthy individuals the macrosatellite array D4Z4 that contains the copies of *DUX4* is constituted of 11-150 repeats, in FSHD patients this number drops to less

than 11 (54). Similar to other repeated sequences, D4Z4 is highly repressed, and in mature tissues is part of constitutive heterochromatin. Massive loss of repeats leads to the de-repression of the D4Z4 locus and the surrounding genes (54). At first, the role of *DUX4* repeats in FSHD etiology was thought to be only due to its proximity to other genes de-repressed by D4Z4 contraction. In an early study, Ding and colleagues identified a region of D4Z4 among the targets of the protein HLTF (Helicase-like Transcription Factor) (55). Further analysis revealed the presence of a coding sequence containing two homeoboxes and a putative promoter. Following studies identified *DUX4* expression in both FSHD myoblast, where it was linked to an undifferentiated apoptotic phenotype, and testis (56). Interestingly, ectopic expression of *DUX4* in healthy human primary myoblast is sufficient to activate the expression of a wide range of germline genes, together with endogenous LTR retroelements (57). Phylogenetic analyses uncovered a large family of *DUX4*-like genes in many placental mammals (58). Comparison among the members of the family highlighted the presence of both an intronless group of members, such as *DUX4* in human and its paralog *Dux* in rodents, and a more distant group of homologs containing an intron, that took the name of *DUXA* and *DUXB* in human and *Duxbl* in mouse (58). Although *DUX4* and *Dux* share high homology, they display no synteny homology in the surrounding regions of, respectively, human chromosome 4 and mouse chromosome 10 (58,59). Therefore, it has been proposed that these two genes are the result of two different events of retrotransposition from an intron-containing member of the *Dux* family (58). The two genes must then have undergone a similar but independent process of expansion, possibly mediated by recombination events. Recent works from different labs have deepened our understanding of the role of human *DUX4* and mouse *Dux* during the first stages of development (51,60). Whole transcriptome studies on human oocytes and early embryos up to the blastocyst stage identified *DUX4* as a main determinant of ZGA-associated

transcriptome (51). Similarly, murine *Dux* expression is strictly limited to the 2-cells stage during development (51,60,61). Ectopic expression of both *DUX4* and *Dux* in human IPSCs and mouse ESCs, respectively, leads to up-regulation of many ZGA genes and induction of retroelements expression (51,60). In addition, overexpression of *Dux* proved to be sufficient to convert mouse ESCs into 2C-like cells, while siRNA-mediated knock-down of the gene greatly impairs the generation of 2C-like cells (51). Importantly, CRISPR-Cas9-mediated knock-out of *Dux* in mouse zygotes causes arrest of the developing embryo at very early stages, mainly around the time of first division, suggesting that this gene is necessary for a successful the ZGA (51). However, it has been reported that *Dux* knock-out mice are able to develop normally and reach term, although at sub-Mendelian ratios (62), highlighting the fact that very little is actually known about *Dux* role and regulation in early development, and more investigation is required to crack its mysteries.

1.3.3 – Factors involved in 2C-like cells regulation

Over the years, evidence about factors involved in the ESC to 2C-like transition has accumulated, but a comprehensive understanding of the mechanism controlling this passage is still missing. Chromatin remodelers, such as the histone demethylase KDM1a (Lysine-specific histone demethylase 1A), the transcriptional repressor TRIM28 (Transcription intermediary factor 1-beta) and the histone methylase SETDB1 (SET domain bifurcated 1) have been proposed to suppress the passage from ESCs to 2C-like cells (34,63). At the same time, loss of the p150 or p60 subunit of the histone remodeler CAF-1 (Chromatin Assembly Factor 1) leads to massive up-regulation of *Zscan4*, MERVL and other 2-cells stage markers (64). All these proofs support the concept that a more open chromatin conformation could favor 2C-like cells appearance. At the same time, *Zscan4*-expressing cells have been shown to possess a more open chromatin conformation, together with a broad de-repression of heterochromatin (35,44), suggesting a complex

relationship between the expression of 2-cells stage genes and chromatin status. In addition, the miRNA mir-34a has been shown to impair the formation of 2C-like cells, possibly through *Gata2* (GATA binding protein 2) repression, although a clear understanding of its mechanism of action would require further study (41). A recent work has also delineated a role for splicing factors in favoring the appearance of 2C-like cells among ESCs (37), possibly due to a post-transcriptional effect on the mRNA of 2C-specific genes. As already mentioned, *Dux* overexpression is able alone to markedly increase the number of 2C-like cells in culture by directly activating 2-cells stage genes and retroelements (51). Recent studies, looking for upstream regulators of *Dux*, have identified the maternal factor NELFA (Negative Elongation Factor complex member A) and the proteins DPPA2 and DPPA4 (Developmental Pluripotency Associated 2/4) as activators of this gene (65–67). Interestingly, *Dppa2/4* double knock-out completely suppresses the presence of 2C-like cells in culture. This effect is thought to depend only on the lack of *Dux* expression, since both DPPA2 and DPPA4 do not appear able to bind any other among the 2-cells stage genes or retroelements (66,67). On the contrary, a complex formed by the nucleolar protein nucleolin (NCL) with the RNA produced from LINE-1 (Long Interspersed Nuclear Elements 1) retrotransposons has been proposed to act in concert with TRIM28 to repress *Dux* expression in ESCs and impair 2C-like cells formation (68). At the same time, Polycomb proteins, and in particular the vPRC1.6 complex, have been shown to repress *Dux* expression in ESCs, and through this action to suppress the expression of 2-cells stage genes (37,69). Polycomb repressive action on 2C-genes seems to require the activity of the E3 SUMO-conjugating enzyme UBC9 (Ubiquitin Carrier protein 9). Direct deposition of SUMO on Polycomb proteins seems in fact to be required for the stability of vPRC1.6 on *Dux* promoter (69). It is important to notice though, that in the same study, SUMOylation has been shown to act as a general suppression mechanism both during MEFs reprogramming

and trans-differentiation processes (69). It is also worth mentioning that UBC9 is involved in a broad range of processes, among which cell cycle, DNA replication and DNA damage response (70,71). It is then reasonable to look at UBC9 role in *Dux* regulation with caution, since its ablation may lead to a wide range of phenotypes. Finally, although not directly linked to 2C-like cells regulation, the NuRD (Nucleosome Remodeling acetylase) complex has been demonstrated to occupy the human *DUX4* locus in human FSHD myoblasts, and ablation of its components (such as CHD4 or HDAC1/2) proved sufficient to cause de-repression of *DUX4* in both FSHD myoblasts and human IPSCs (72).

1.4 – DNA replication in eukaryotes

At every cell division the entire genome must be duplicated faithfully in order to be equally divided between the two daughter cells. Errors, lack of replication or over-replication involving even a small fraction of the genome can have dire consequences, ranging from the death of single cells to the acquisition of dangerous mutations that can put at risk the entire organism. Faithfull replication is particularly important in the case of fast replicating cells, such as ESCs. In the following paragraphs, a general review of the process of DNA replication will be given, with particular focus on the problems that can arise during the process and the mechanisms enacted to ensure replication completion.

1.4.1 – Origins licensing and firing

Eukaryotes, and in particular metazoans, possess very large genomes. In order to replicate such a vast amount of DNA in a timespan compatible with the normal cell cycle, eukaryotic cells start DNA replication from multiple replication origins at the same moment. No consensus sequences have been identified among metazoans replication origins, although in some studies it was hypothesized they might be enriched in both CpG islands and AT-rich sequences (73,74). Not all replication origins are activated (fired) at each cycle. More precisely, only a small group of constitutive origins starts replication at each cycle, while

most origins fire in a stochastic manner. In addition, a group of dormant origins is fired only in specific cell types or under certain conditions (75). The division between constitutive/stochastic and dormant origins allows eukaryotic cells to regulate the speed of DNA replication by simply changing the number of origins that can be fired at any moment (75). Moreover, data from budding yeast demonstrate that origins are not all fired at once, but their activation is spread throughout the S-phase (76). Origin firing must be tightly controlled in order to ensure complete replication of the genome and, at the same time, to avoid events of re-replication during the same cycle. This control is based on a system of licensing through which only chosen origins can be fired. Origin licensing can take place form late M to the G1-phase of cell cycle and starts with the binding of the ORC (Origin Recognition Complex) to the DNA (77,78). As shown in figure 2, ORC binding is followed by the recruitment of CDC6 (Cell Division Cycle 6) and CDT1 (Chromatin Licensing and DNA Replication Factor 1) which cooperate with ORC to load onto the DNA the hetero-hexameric MCM2-7 (Minichromosome Maintenance) complex (75,78).

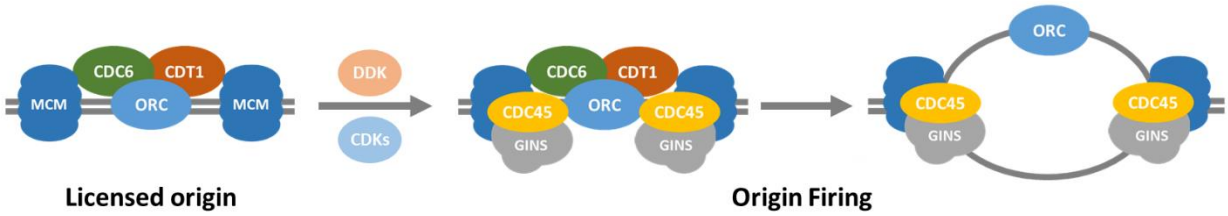

Figure 2 | Scheme of origin licensing and firing in eukaryotes: The ORC complex recruits CDC6 and CDT1 on the site of the origin, recruiting the MCM complex. At the beginning of S-phase, DDK and CDKs activities lead to the loading of CDC45 and the GINS complex, completing the replicative helicase. Finally, The MCM-CDC45-GINS complex starts to unwind the DNA, opening the replication bubble, while the degradation of CDC6 and CDT1 prevents further licensing of replication origins during S-phase.

Loading of MCM2-7 completes the formation of the pre-replicative complex and it is necessary for the following firing of the origins. The MCM2-7 complex is actually loaded on the origin as a double hexamer, as the two helicase-cores will then move in opposite direction, according to the bi-directional replication model. At the onset of S-phase, CDK-

dependent phosphorylation and ubiquitination by the SCF/SKP2 (Skp, Cullin, F-box containing/S-Phase Kinase Associated Protein 2) complex lead to CDT1 degradation, preventing further origin licensing during S-phase (79,80). In addition, in metazoans, the protein Geminin inhibits CDT1 during S and M-phases, providing an additional mechanism to avoid re-replication (81). Origin firing is regulated by many factors and it is still partly obscure. CDK activity in cooperation with DDK (Dbf4-Dependent Kinase) promotes the loading of CDC45 (Cell Division Cycle 45) on the pre-replicative complex, initiating origin firing (Fig. 2) (82). CDC45 is critical for the recruitment of many replication proteins, such as polymerase α, polymerase ε and the processivity factor PCNA (Proliferating Cell Nuclear Antigen) (83). Together with CDC45, also the GINS (go-ichi-ni-san) complex is loaded onto the origin, completing the formation of the replicative helicase (84). The CDC45-MCM-GINS (CMG) complex represents the actual replicative DNA helicase: this now starts to unwind the DNA (Fig. 2), expanding the replication bubble in both directions.

1.4.2 – Replication fork progression

At each fired origin, two replication forks are established, which progress symmetrically in opposite directions (75). A massive protein complex, named replisome, is associated to each fork. Among the proteins of the replisome are the ones composing the aforementioned CMG complex. DNA replication is ensured by various DNA polymerases, mainly polymerase ε and δ, associated with ancillary factors, such as the processivity factor PCNA and the replication factor C (RFC) (85). Due to their biochemical features, these two DNA polymerases can only elongate existing DNA filaments. For this reason, DNA synthesis is always started by the polymerase α/primase complex, which synthesizes a short RNA primer followed by a short DNA tract. The RNA primer will be then removed from the newly replicated DNA (75). Associated to the replisomes can also be found checkpoint proteins,

such as ATR (Ataxia telangiectasia and Rad3 related) and ATRIP (ATR Interacting Protein) which will be discussed in more detail in the following paragraph.

Since DNA biochemistry dictates that new deoxynucleotides can be added only at the 3' extremity of a polynucleotide chain, replication must always proceed in a 5' to 3' direction. Since in the double helix the two filaments run antiparallel, the direction of fork progression is however 5' to 3' for one parental strand and 3' to 5' for the other strand. For this reason DNA replication can progress continuously only on one filament, named leading strand, while the other, called lagging strand, requires constant re-priming (86). In eukaryotes, the fragments (known as Okazaki fragments) formed on the lagging strand by continuous priming by polymerase α/primase, and elongated by Polymerase δ, are usually around 200bp in length, following nucleosomes periodicity, and are spliced in a complex process of maturation (87).

Replication fork progression requires continuous unwinding of the DNA double helix, causing the accumulation of topological stress ahead of the fork (88). Due to their length and disposition, eukaryotic chromosomes cannot simply disperse this tension by the swiveling of their extremities. Torsional stress poses a physical impairment to further unwinding of the double helix and, therefore, its resolution is of paramount importance for the progression of DNA replication. This task is fulfilled by a specific class of enzymes known as DNA Topoisomerases. Topoisomerases of type I (such as TOP1) catalyze the transient formation of single strand nick allowing the DNA molecule to rotate on its longitudinal axis (88). Type II topoisomerases (like TOP2a and TOP2b) generate, instead, a transient double strand break (DSB) in the DNA molecule and pass an intact DNA helix through the gap (88). Topoisomerases are generally not considered to be part of the replisome, but their activity is fundamental for the progression of replication forks, and their inhibition results in fork arrest and DNA breakage (88).

Replication bubbles expand along the genome until the entire parental strand is unwound and replicated. The convergence of two opposite forks generates topological tension that is resolved by topoisomerases. Replication is completed by gap filling on the lagging strand and nick ligation. At the end of S-phase the MCM2-7 complex is polyubiquitinated and unloaded from the chromatin (89). Unloading of PCNA is independent of MCM2-7 ubiquitination and requires instead the ELG1 ATPase (90). Successful unloading of the replisome completes the replication process, leaving two identical sister chromatids ready to be separated during M-phase.

1.4.3 – DNA Replication stress and its sources

Replicating the entire genome is a tremendous enterprise which entails massive processing of the chromatin. DNA replication itself imposes mechanical stresses on the double helix, and the rapid progression of the massive replication machinery must cope with the presence of many other DNA interacting proteins inside the nucleus. Their presence makes the genome particularly vulnerable to both internal and external sources of stress, which may interfere with replication and cause great damage to the stability of the genome.

To this day, no clear definitions of DNA replication stress (RS) exist, but this phenomenon is generally described as any event causing the slowing or the stalling of replication forks. This broad definition is a consequence of the fact that the potential sources of RS are many and very different in nature. Lesions to the DNA either occurring in physiological conditions or caused by external agents can impair the proceeding of the replication fork. Other physiological sources of RS are the presence of DNA-binding proteins, transcriptional units or even unusual DNA structures (91). In particular, the process of transcription can impair DNA replication both due to the direct collision between the replisome and the transcriptional machinery, or through the formation of highly stable DNA-RNA duplexes known as R-loops (88,92,93). Highly repeated sequences such as dinucleotide,

trinucleotide, inverted, mirror and direct tandem repeats are also intrinsically prone to trigger RS (91,94). These and other loci are often referred to as fragile sites, due to their tendency to show gap or breaks formation upon partial inhibition of DNA synthesis (94,95). Similarly, highly heterochromatic regions are highly enriched in RS markers in unperturbed S-phase, suggesting that there may be a conflict between chromatin status and DNA replication (96).

Aberrantly high levels of RS may be caused also by mutations affecting components of the replication machinery or their regulators (97). Furthermore, also mutations that do not directly involve components of the replication machinery can be a source of RS. A clear example is the activation of oncogenes, such as *K-RAS* (Kirsten Rat Sarcoma), *c-MYC* (Myelocytomatosis oncogene) or *CCNE1* (Cyclin E) which has been linked to increased RS, probably due to dysregulated cell duplication or increased replication-transcription conflicts (98). Finally, many external sources of RS have been identified over the years, and their use in laboratories has allowed further understanding of the RS response. For example UV radiation causes RS by generating thymine dimers that physically impair the progression of the forks (99). Instead, Hydroxyurea (HU) acts by depleting the cellular pool of deoxynucleoside-triphosphates (dNTPs) (100). The chemotherapeutic cisplatin forms inter-strand covalent cross-links, physically impairing the separation of the two strands at the fork (101) while camptothecin and etoposide, inhibit type I and type II topoisomerases respectively, preventing the relaxation of topological stress (102,103). Finally, aphidicolin directly inhibits the activity of the replicative DNA polymerases, generating long stretches of unreplicated ssDNA downstream to the proceeding replication helicases (104).

RS caused by any of these internal or external factors can prove extremely detrimental for the survival of the cell. To counter this risk the cell puts in place a refined system to grant the stability of replication forks in the face of stressful events.

1.4.4 – ATR checkpoint activation

During DNA replication the ssDNA formed by the unwinding of the double helix is coated by the RPA (Replication protein A) complex, which prevents the formation of secondary structures and protects ssDNA from degradation (105,106). When replication is impaired, long stretches of RPA coated ssDNA constitute the recruiting platform for the ATR-ATRIP complex at the site of stalled replication forks (105). ATR is the master regulator of the RS response pathway (also referred as checkpoint) and is involved in tackling a broad range of DNA damage and replication problems. ATR activity is essential for correct replication, and complete deletion of the gene have proven to be embryonic lethal in mouse models (107). ATR is primed for activation by the association with its regulatory partner ATRIP, which mediates its recruitment to RPA-coated ssDNA (105,108). Upon binding, ATR activates itself by auto-phosphorylating the residue T1989, critical for the recruitment of the regulatory protein TopBP1 (DNA topoisomerase II binding protein 1) (105,109). TopBP1 increases ATR kinase activity, leading to a feed-forward effect that results in the phosphorylation in *trans* of many ATR molecules. In addition, it has been recently shown that another protein, ETAA1 (Ewing tumor-associated antigen 1), can be recruited at sites of stalled replication forks in a RPA-dependent manner, to activate ATR in a similar fashion to TopBP1 but independently from it (110). ATR activation is then completed by its interaction with the 9-1-1 complex, a clamp-like structure similar to PCNA, that is loaded at the site of the fork by the RFC complex (105). Activated ATR can phosphorylate a wide range of effectors, among which the H2AX isoform of histone H2A. The phosphorylated form of this histone on residue S139 is commonly referred as γH2AX and is considered one of the main markers of RS, although its presence can be detected also at sites of DNA DSBs (111). Phosphorylation of RPA by ATR is required for efficient fork restart, while phosphorylation of other mediators, such as CDC17 (dell division cycle 17), BRCA1 (breast cancer 1, early onset) and

Claspin is involved in the cascade that leads to the activation of ATR main effector: CHK1 (checkpoint kinase 1) (105). CHK1 is activated by phosphorylation by ATR on residues S317 an S345. Activated CHK1 then phosphorylates additional effectors, among which CDC25 (cell division cycle 25) and Treslin, resulting in late origins inhibition and ultimately in G2 arrest (112,113). Therefore, CHK1 activation allows to slow down DNA replication, giving the cell more time to recover stalled forks and overcome RS. In addition ATR phosphorylates replisome components, like the MCM complex, and proteins involved in fork protection, such as SMARCAL1 (SWI/SNF related, Matrix associated, Actin dependent Regulator of Chromatin, subfamily a like 1), or in DNA damage repair, which may be required for the recovery of the fork in a context-dependent manner (105,114).

1.4.5 – Mechanisms of fork protection and replication restoration

After ATR activation, several pathways can engage the stalled replication fork in order to resume replication. All these pathways are complex and partially intertwined, but they can be broadly depicted as three main scenarios. First, the RS source or DNA lesion can be directly bypassed. This can be achieved through translesion synthesis (TLS) by specialized DNA polymerases (such as polymerase κ or η) which insert nucleotides across lesions or physical obstacles with poor sequence specificity (Fig. 3a). These polymerases are recruited on the site by ubiquitinated PCNA and synthesize only short stretches of DNA before being replaced by replicative polymerases (115). In alternative, repriming of DNA synthesis downstream of the lesion by the error-prone polymerase-primase PRIMPOL can bypass the lesion, allowing the progression of DNA replication independently from the repair of the lesion itself (Fig. 3b) (116).

In alternative to lesion bypass, stalled replication forks can undergo a process called fork reversal, as depicted in figure 3c. In this case, reannealing of the parental strands takes place. As a consequence, the nascent DNA strands extrude and pair back, forming a

reversed arm, that is then covered by RAD51 (Radiation resistance 51) in order to protect it from degradation (117,118). Reversed forks have been proposed to limit the amount of ssDNA at the site of stalled forks, promoting their stability, and to replace DNA lesions in a dsDNA context, avoiding for example the conversion of single strand nicks into DSBs upon replication fork encounter (117,118). Moreover, fork reversal could enable error-free bypass of DNA lesions by exposing the newly synthesized leading strand, that could be used as a template in place of the corresponding damaged parental strand. This pathway, known as template switching, has been proposed to depend upon PCNA polyubiquitination (119). The precise mechanism of fork reversal remains still cryptic, but genetic and biochemical studies have demonstrated that it mainly depends on the activity of three chromatin remodelers of the SNF2 family: HLTF, ZRANB3 (Zinc Finger, RAN-Binding Domain Containing 3) and SMARCAL1. HLTF has been proved to catalyze the ATP-dependent formation of reversed forks both *in vitro* and *in vivo* (120–122). It is thought to be recruited to the stalled fork by the interaction with exposed 3'-ssDNA ends, where, in addition to its fork reversal activity, it performs polyubiquitination of PCNA, making it a potential candidate involved in template switching (121,123,124). ZRANB3 has been shown to be recruited at stalled forks by the interaction with polyubiquitinated PCNA, making it another possible candidate for the template switch pathway (125). In addition to its ATP-dependent helicase activity, ZRANB3 possesses a nuclease domain, not necessary for fork reversal, that has been proposed to facilitate excision of DNA lesions at stalled forks (126). Finally, SMARCAL1 is known to associate to unperturbed replication forks and to be enriched at stalled forks (127,128). It is recruited at replication forks through the binding of RPA, where it promotes ATP-dependent fork regression (127). Ablation of any of these translocases results in reduced reversed forks formation and impaired fork restart upon administration of RS (114,121,125). The differences among their mechanisms of recruitment suggests that all

three may work in parallel or partially interdependent pathways, but complete understanding of their regulation and role still require further study. Once the reversed arm is formed, it can be partially resected by nucleases such as MRE11 (MRX complex nuclease subunit), DNA2 (DNA replication helicase/nuclease 2) and EXO1 (Exonuclease 1) to restart the fork (129,130). In alternative, homologous recombination (HR)-mediated strand invasion can restore DNA replication through error-free template switching (119).

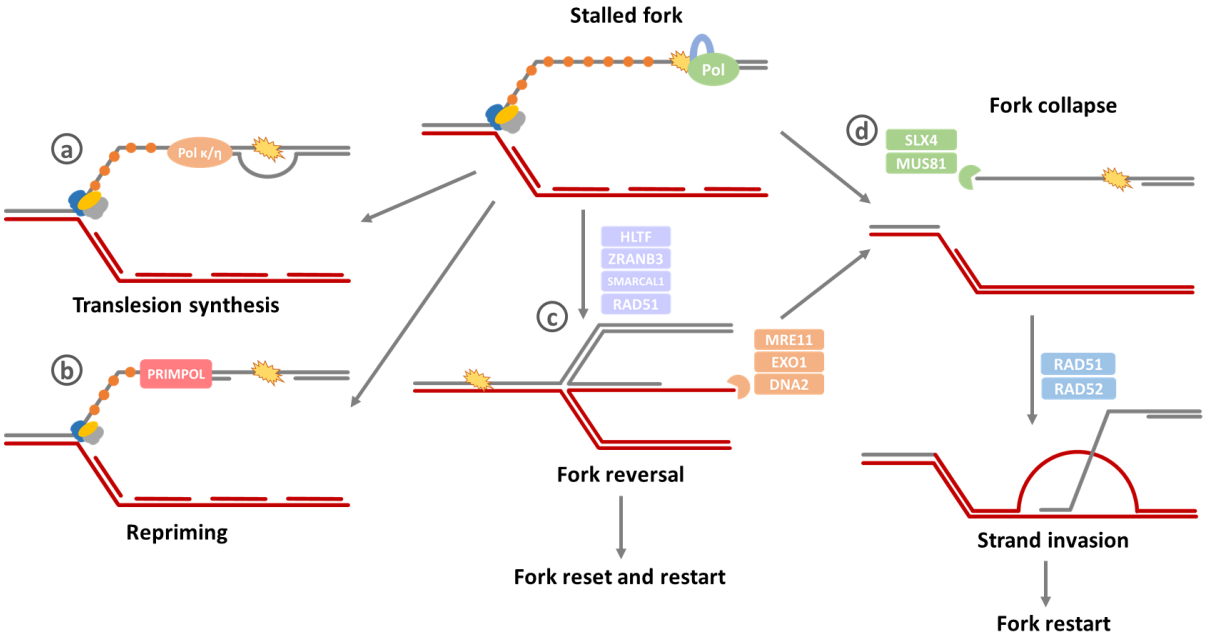

Figure 3 | Main pathways of fork protection and DNA replication restart: a) Replication can bypass physical obstacles on DNA by loading error-prone DNA polymerases to perform translesion synthesis or **b)** Replication can restart downstream to the lesion thanks to repriming by the primase-polymerase PRIMPOL. **c)** Stalled forks can also undergo a process of fork reversal catalyzed by the HLTF, ZRANB3 and/or SMARCAL1 translocases to protect the fork while the source of RS is overcome. In this case, RAD51 protects the regressed arm from degradation by the MRE11, EXO1 and DNA2 nucleases. **d)** Stalled and reversed forks can in alternative be degraded by the action of SLX4-MUS81 nucleases. A process of strand invasion guided by RAD51 and RAD52 can, in this case, regenerate the fork and restart replication.

Importantly, it has been proposed that reversed forks may constitute irreversible lesions, which require MRE11-mediated degradation before replication can restart. This is particularly true in the case of BRCA1/2 defective mutants, which show impaired RAD51 loading at reversed forks, making them optimal substrates for MRE11 (131).

Finally, if the source of stress cannot be overcome or bypassed, the fork undergoes MUS81/SLX4-dependent cleavage, leading to the formation of a DSB (Fig. 3d). In this case,

RAD51 and RAD52 promote HR-mediated strand invasion and DSB repair, which restarts DNA replication (129). In their complex, all these pathways guarantee fork stability and allow restoration of DNA replication in an extremely wide range of situations, protecting one of the most fundamental processes of cellular biology.

1.5 – Polycomb proteins

Developmental processes and cell fate commitment events are tightly regulated at the epigenetic level by a broad range of chromatin modifying factors. Polycomb group proteins are among the most emblematic members of this category. This large family of proteins was discovered for the first time more than 70 years ago in *Drosophila melanogaster* with the identification of the *Polycomb* (*PC*) gene (132). Following studies demonstrated that loss of *Polycomb* could transform anterior embryonic segments in more posterior ones by dysregulating *Hox* genes (133). Decades of studies have identified many other members of the Polycomb family in both animals and plants, while expanding the comprehension of the fundamental importance of this group of proteins, that goes well beyond the simple regulation of the *Hox* cluster. In the following paragraphs will follow a brief description of the structure and main functions of Polycomb complexes, with a particular focus on recent findings regarding the role of these chromatin modifiers in relation to DNA replication.

1.5.1 – Structure of the Polycomb repressive complexes

Polycomb complexes are divided in two main groups, based on original purification experiments in *Drosophila*: Polycomb repressive complex 1 (PRC1) and PRC2. In mammals, the composition of Polycomb complexes, and in particular of PRC1, is more articulated than what originally identified in flies. Mammalian PRC1 is composed by a core subunit (RING1a or RING1b) (RING finger proteins 1 a/b) with E3 ubiquitin ligase activity, which is responsible for the deposition of histone H2A ubiquitination on residue K119 (H2AK119Ub) specific for this complex (132,134). The RING protein is always associated with one of the

six PCGF1-6 Polycomb group ring-finger proteins (Fig. 4). The principal division among PRC1 complexes is between canonical (cPRC1) and variant (vPRC1) complexes. All cPRC1 complexes contain one of the CBX chromobox proteins (namely CBX2/4/6 - 8), which allow these complexes to interact with H3K27me3 (trimethylated histone H3 on residue K27), the histone mark deposited by PRC2 complexes (135).

On the contrary, vPRC1 complexes are completely devoid of CBX subunits, which are substituted by the zinc-finger domain and YY1-binding protein RYBP or its paralog YAF2 (YY1-associated factor 2) (Fig. 4) (135). In addition to chromobox proteins, cPRC1 complexes contain PCGF2/4, a Polyhomeotic homologous protein (PHC1/2/3) essential for cPRC1 repressive function and the SCMH1/2 (Scm Polycomb Group Protein Homolog 1/2) proteins (136). On the opposite, vPRC1 complexes can associate with any of the PCGF proteins, which subdivide them in further subclasses. As shown in figure 4, the composition of the vPRC1 complexes is extremely various among different subclasses, which may imply a difference in specificity or activity, although these differences remain mostly to be elucidated.

The core of PRC2 complexes is constituted by one of the SET domain-containing histone methyltransferases enhancer of zeste (EZH1/2), embryonic ectoderm development (EED) and suppressor of zeste (SUZ12) bound to the RB-Binding Proteins RBBP4 and RBBP7 (Fig. 4) (132). EZH2 is the main catalytic subunit of PRC2, but in order to relieve its autoinhibition and perform histone H3 trimethylation it requires the interaction with SUZ12 and EED (137). This highly conserved core is then regulated by the non-stochiometric binding of a series of accessory subunits which modulate PRC2 activity and binding to its targets. Systematic studies in human cells have demonstrated the existence of two main subgroups of PRC2 complexes, named PRC2.1 and PRC2.2 (138). The first of these subclasses is characterized by the interaction between the PRC2 core and one among the three proteins

PHF1 (PHD finger protein 1), PHF19 (PHD Finger Protein 1) or MTF2 (MRE-binding transcription factor 2) (Fig. 4) (138). In addition, the two poorly characterized proteins LCOR and EPOP have been proven to interact with PRC2.1 in human cells (138). The PHF1/19 or MTF2 direct PRC2.1 towards different targets on the chromatin affecting its function during differentiation processes (139). The PRC2.2 subclass is, instead, defined by the concomitant binding of JARID2 (Jumonji and AT-rich interaction domain containing 2) and AEBP2 (AE Binding Protein 2) which have been shown to regulate PRC2.2 binding and activity (Fig. 4) (134,140,141).

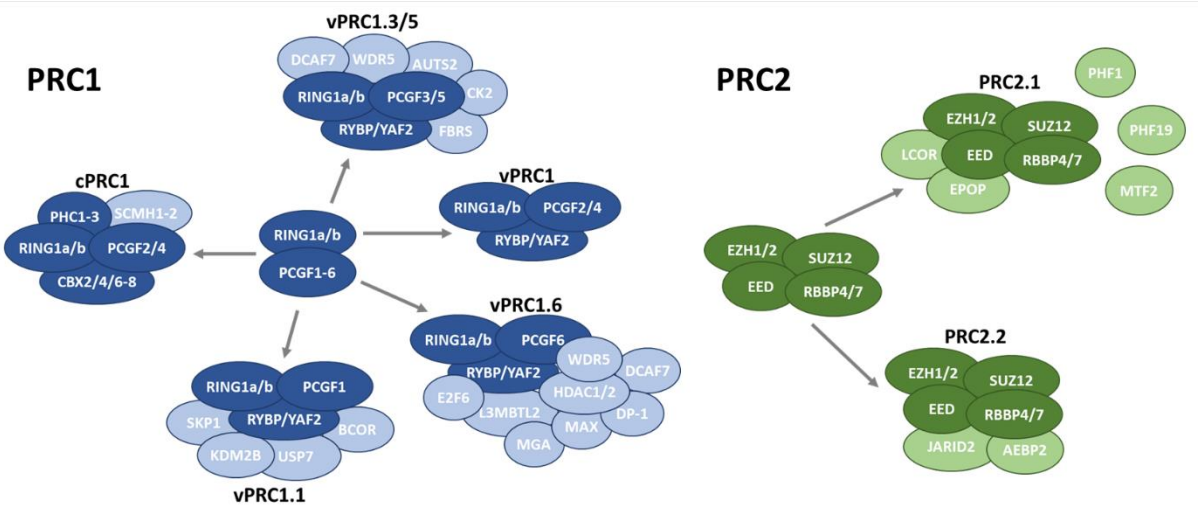

Figure 4 | Composition of Polycomb complexes: All PRC1 complexes are constituted by a core made of one RING1a/b ubiquitin-ligase protein and one among PCGF1-6 Polycomb proteins. Canonical PRC1 (cPRC1) contain either PCGF2/4, one among CBX2/4/6-8 chromobox proteins and SCMH1/2. Variant complexes (vPRC1) contain either PCGF1, PCGF2/4, PCGF3/5 or PCGF6 taking the respective names of vPRC1.1, vPRC1, vPRC1.3/5 or vPRC1.6. All these complexes contain the RYBP protein or its homolog YAF2, but no CBX or PHC proteins. The vPRC1 complexes are completed by a wide range of accessory subunits, as depicted in the figure. PRC2 complexes contain a core made of the catalytic subunit EZH1/2, SUZ12, EED and RBBP4/7. This core is associated either with the LCOR/EPOP proteins and one among PHF1, pHF19 or MTF2 subunits (forming PRC2.1 complexes) or with the JARID2/AEBP2 regulatory subunits (forming the PRC2.2 complex).

This complex and multifaceted structure of Polycomb complexes results in an equally refined function in gene regulation. The understanding of the way in which Polycomb complexes are recruited to target sites and exert their function has progressively changed over the decades. In the following paragraphs is given a brief summary of the current understanding of Polycomb recruitment and function at target sites.

1.5.3 – Polycomb targeting mechanisms

The traditionally accepted model of Polycomb recruitment at target sites was originally proposed from observations made in flies and is known as hierarchical model. According to this model, PRC2 complexes are the first to bind to target sites, where they deposit H3K27me3. cPRC1 complexes are then recruited to the site through the interaction between H3K27me3 and CBX proteins, leading to H2AK119Ub deposition (132,134). The determinants of the initial PRC2 binding are partially still unknown but many indications have been collected over the years. For example, It has been proposed that long non-coding RNAs (lncRNA) may recruit PRC2 to specific loci. This is the case during chromosome X activation, when *Xist*-mediated recruitment of PRC2 is required for the silencing of the chromosome (142). Moreover, it has been shown that the *HOTAIR* lncRNA produced by the *HOXC* locus in human binds and recruits PRC2 to the *HOXD* promoter (143). In addition to lncRNA, also transcription factors, such as SNAI1 (Snail family transcriptional repressor 1), have been proposed to recruit PRC2 at the promoter of genes (144). In all these cases, PRC2 recruitment has been proven only for very specific loci or promoters but also more general systems of recruitment do exist. In fact, PRC2 has been reported to localize at GC-rich sequences in a JARID2-dependent fashion (145). In addition, it has been recently shown that PRC2 binds with little specificity to a wide range of nascent RNAs (146). These observations are currently under debate and a consensus on their biological meaning is still lacking. PRC2-RNA interaction could in fact be a way to reinforce Polycomb repression at sites of repressed genes that experience spurious transcription, but could also be a way in which active genes can sequester PRC2 in order to avoid Polycomb-mediated repression. In addition to these mechanisms, PRC2 can bind histone modifications. It is the case for H3K27me3 itself, that is recognized by EED or H3K36me3 (which has been proposed to facilitate the passage from active to repressed chromatin in lineage transitions), which has

been shown to recruit PRC2 during ESCs differentiation (147,148). This mechanism would promote rapid re-establishment of PRC2 binding to chromatin following DNA replication, ensuring mitotic inheritance and self-propagation of PRC2 marked sites (148).

The discovery of vPRC1 has put in crisis the traditional hierarchical model of Polycomb recruitment. In fact, these complexes cannot interact with H3K27me3 due to lack of CBX proteins. In agreement with this hypothesis, deletion of *Eed* in mouse ESCs cannot completely suppress the deposition of H2AK119Ub at target genes, showing that PRC1 does not always require PRC2 to localize at its targets (149). Further studies have shown the ability of PRC1 to associate with different transcription factors, which guide the complex on specific promoters (150). Moreover, the PRC1 subunit KDM2B (Lysine demethylase 2B) has been shown to bind unmethylated CpG di-nucleotides *in vitro* and CpG islands *in vivo*, proving and additional mechanism for the PRC2-independent recruitment of PRC1 (151).

In a further twist of the traditional model, recruitment of PRC1 complexes to an artificial sequence, using fusion proteins made of PCGF subunits and the tetracycline repressor, have demonstrated that the binding of vPRC1.1/3/5 is sufficient for the recruitment of PRC2 and H3K27me3 deposition (152). A similar result was obtained in mouse ESCs, where KDM2B-mediated recruitment of vPRC1 to pericentromeric chromatin proved to be sufficient for H3K27me3 deposition on the loci (153). These results may be dependent on the ability of both JARID2 and AEBP2 to bind H2AK119Ub, as it has been proved in biochemical studies (154).

1.5.4 – Polycomb functions in gene regulation

The presence of Polycomb on target genes is generally associated with transcriptional repression. Interestingly, several observations suggest that genes down-regulation may precede Polycomb binding on the promoter, questioning the exact role of these complexes (155,156). According to these data, Polycomb would not directly drive gene repression, but

would act as a memory system to maintain a correct transcriptional status of genes regulated by more specific systems. The mechanism through which Polycomb maintains genes in a repressed state is generally associated to a compaction of the chromatin. For example, PRC1 has been shown to condense chromatin through nucleosome bridging mediated by CBX2 (157). Similarly, both RING1b and EZH2 have been shown to mediate chromatin compaction at imprinted loci (158). In addition, cPRC1 subunit PHC2 has been shown to undergo polymerization, driving direct compacting of chromatin (159). Moreover, the histone mark H3K27me3 not only impedes the deposition of the activating acetylation on H3 residue K27, but is also known to recruit PRC2 in a feedback loop that allows further expansion of the repressed region (148). Interestingly, a recent report proved that histone H1 promotes PRC2 methylation activity *in vitro*. At the same time, H1 ablation causes loss of chromatin compaction and gene up-regulation among PRC2 targets, strongly suggesting that H1-dependent chromatin compaction could be an additional mechanism of PRC2 recruitment (160).

Finally, the precise role of H2AK119Ub in gene repression has not yet been fully determined although its absence has been proven to strongly affect the ability of both cPRC1 and PRC2 complexes to bind their targets (161). Additionally, a surprising report linked an AUTS2 (Autism susceptibility gene 2)-containing vPRC1 complex to direct gene activation during mouse neural development (162). This result highlights the complexity of the Polycomb regulatory network and how far we still are from a thorough understanding of its functions and mechanism of action.

1.5.5 – Polycomb role during DNA replication

In addition to Polycomb canonical role in gene regulation, over the years evidence has accumulated linking this group of proteins to DNA replication. Short hairpin-mediated ablation of both SUZ12 and EED, as well as knock-down of EZH2 have proven to strongly

affect MEFs growth rate through a general reduction of replication fork speed (163). Similarly, double knock-down of both *Ring1* (which encodes RING1a) and *Rnf2* (which encodes RING1b) also results in reduced cell growth (163,164). Both loss of PRC2 or PRC1 components have proven to increase the percentage of asymmetrical replication bubbles detected in mouse models, a sign associated with frequent fork stalling and inefficient fork restart (163,164). In addition, deletion of both *Ring1* and *Rnf2* in MEFs results in accumulation of RS markers during S-phase (164). The mechanistic details of this novel Polycomb function remain to this time mostly cryptic. Interestingly, both RING1b and SUZ12 colocalize with PCNA in dividing cells, indicating a direct role for both Polycomb complexes at the fork during replication (163). Moreover, also the L3MBTL1 (Lethal 3 Malignant Brain Tumor L3) protein, a subunit of vPRC1, was demonstrated to interact with components of the replication fork, such as PCNA, MCM2-7 and CDC45 in human cells, and its depletion proved sufficient to cause fork stalling and RS (165). These evidences clearly indicate that both PRC1 and PRC2 are required during normal DNA replication and their absence impairs the progression of the forks. In particular, it has been demonstrated that both RING1a and RING1b, respectively in human and mouse, bind to pericentromeric satellite regions (164,166). RING1a/b are thought to favor the replication of these highly repeated sequences through H2AK119Ub deposition. In fact, expression of a protein construct composed by only the RING domains of RING1b and PCGF4, fused with the methyl-CpG binding protein MBD1, is able to rescue the RS level displayed by *Ring1*$^{-/-}$*Rnf2*$^{-/-}$ double knock-out MEFs (164). In addition to PRC2 role in normal replication, EZH2 has been shown to localize in proximity to ssDNA formed at stalled forks in a *BRCA2*$^{-/-}$ background. EZH2 recruitment at the fork results in H3K27me3 deposition and mediates the binding of MUS81, resulting in the degradation of the stalled fork (167).

Polycomb role in promoting genomic stability is not limited to DNA replication though. Other reports have, in fact, demonstrated that PRC1 components, such as PCGF2/4 and CBX2, or PRC2 component EZH2, as well as H3K27me3, strongly localize at sites of DSBs induced by IR (168,169). These evidences clearly suggest that Polycomb is tightly involved in DNA metabolism and that it plays a major role in the protection of genomic stability.

1.6 – ATR expands ESCs fate potential in response to replication stress

ESCs are characterized by a fast cell cycle, mainly due to the presence of a very short G1-phase. This peculiar cell cycle imposes a high level of RS in ESCs already in unperturbed conditions (170). In our most recent work (42), we investigated the mechanisms that ESCs put in place to try to cope with such a high level of RS without damage. Surprisingly, we observed that different RS inducing agents, such as aphidicolin (APH), hydroxyurea (HU) or UV radiation increase the expression of several genes specific for the 2-cells stage of embryonic development, among which *Dux*, *Zscan4*, *Tsctv1/3* and the retroelements MERVL. Both single cell and bulk transcriptomic analysis confirmed that RS can activate a wide range of 2-cells stage specific genes. Interestingly, we observed that this increase is not driven by an accumulation of 2-cells stage transcripts from 2C-like cells already present in the culture but is due to a general increase in the number of cells with a 2C-like transcriptome. Further investigation demonstrated that ATR inhibition, but not ATM inhibition, can strongly suppress this increase. Depletion of ATR by siRNA-mediated knock-down or low ATR levels due to the Seckel mutation impairs RS-induced activation of 2C genes. Similarly, RS cannot efficiently up-regulate 2C genes in cells expressing low levels of CHK1 or in which CHK1 is chemically inhibited. In addition to these evidences, activation of ATR in a RS-free context, thanks to the overexpression of ETAA1 ATR activating domain

(AAD), leads to the up-regulation of several 2-cells stage markers. Even in this case, 2C genes activation can be efficiently suppressed by inhibition of either ATR or CHK1.

Among the ATR-induced genes in response to RS, *Dux* has proven to have a central role. Its knock-down in fact is sufficient to strongly suppress the expression of main 2C-markers like Zscan4 or MERVL, while the complete deletion of the gene leads to total loss of 2C genes expression already in unperturbed conditions. In our effort to understand the mechanism of *Dux* activation in response to RS, we performed a siRNA-based knock-down screening on 148 candidate genes. Candidates were selected on the basis of an *in silico* analysis of potential binders of *Dux* promoter together with a review of the literature (37,72). Our analysis provided a total of 49 hits, divided between putative activators and repressors of *Dux*. Among the identified genes, there were actors involved in DNA replication, a small group of transcription factors and chromatin remodelers, and a large group of genes involved in various stages of mRNA maturation. Further validation of the members of this group, mostly categorized as repressors, confirmed the ability of GRSF1 knock-down to partially reduce both *Dux* and *Zscan4* activation in response to RS. Moreover, GRSF1 proved to interact with a synthetic RNA carrying the sequence of *Dux* 3'-UTR (untranslated region) *in vitro*. Given the role of GRSF1 in mRNA binding and stabilization, this result lead us to speculate that *Dux* mRNA stabilization could be one of the mechanisms responsible for 2C genes activation in response to RS in ESCs, although our data do not allow us to exclude the presence of concurrent mechanisms of activation.

In agreement to what already proposed in the literature (34), also ATR-induced 2C-like cells have been proved to possess an expanded developmental potential. Activation of ATR through ETAA1-AAD overexpression in ESCs, prior to their injection into morula stage embryo, results indeed in their increased contribution to the extraembryonic compartment both at the late blastocyst stage (E4.5) and in the egg cylinder (E7.5).

The existence of a pathway in ESCs that can drive their conversion in more potent 2C-like cells in response to RS led us to speculate about the existence of a similar process in the embryo (Fig. 5). Blastomeres undergoing fast replication cycles during morula and blastocyst formation may in fact face RS. Failure in proper fork protection can have dramatic effect on the cell, ranging from accumulation of karyotypic abnormalities to cell death. The expression of 2C genes in response to RS may favor genomic stability, as suggested by several publications (36,39,40), and help the early blastomeres to face the consequences of RS. In case of failure, however, the acquisition of a more potent state due to the expression of 2-cells stage genes, could offer the additional advantage to give to the damaged blastomere a second chance to undergo differentiation toward extraembryonic lineages and not participate to the embryo proper. This pathway could act as a safeguard to promote genomic integrity in the embryo while at the same time favoring the contribution of more damaged cells to extraembryonic tissues.

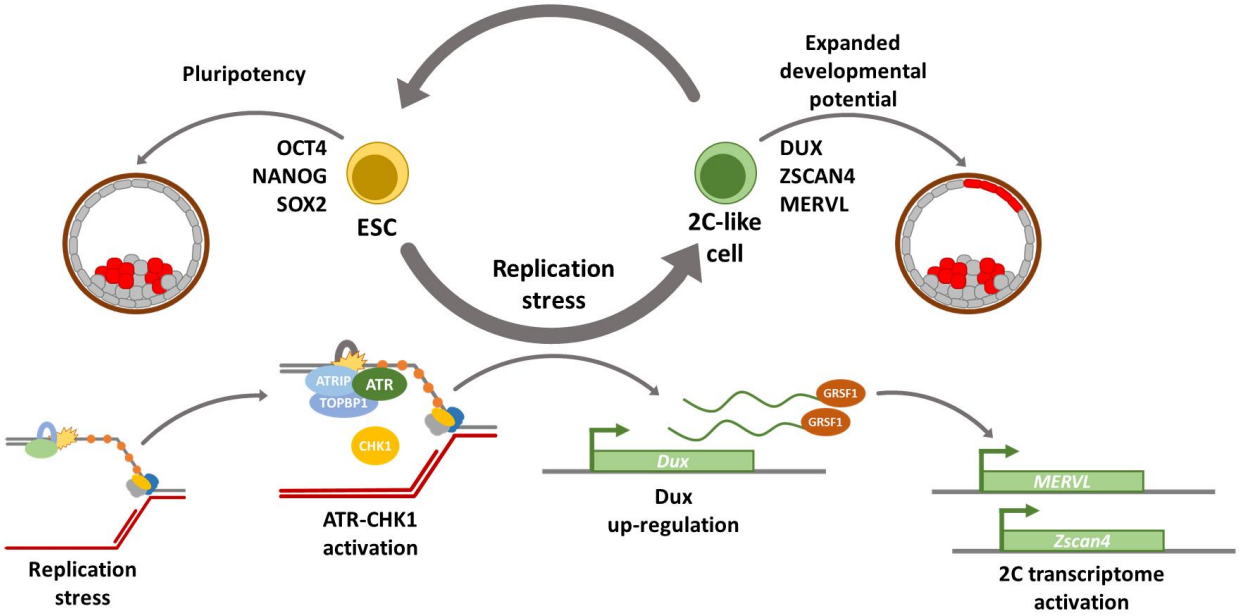

Figure 5 | Replication stress induces 2C-like cells with expanded developmental potential through ATR activation: Schematics depicting the cycle of interconversion of ESCs and 2C-like cells in culture. Replication stress leads to ATR and CHK1 activation, which in turn causes Dux up-regulation. DUX binds to MERVL elements and 2-cells stage specific genes leading to the acquisition of an expanded developmental potential.

1.7 - Purpose of the work

As reported in the previous paragraph, our recent work highlighted a surprising link between RS and the appearance of 2C-like cells in ESCs cultures. Remarkably, this link passes through the action of the ATR-CHK1 pathway, the master regulator of the RS response. The activation of this molecular mechanism results in the transcriptional up-regulation of the gene *Dux*, the main controller of 2C genes expression. Starting from these groundbreaking observations, in this work we are going to focus on the changes affecting the *Dux* locus in response to RS, with particular interest for its regulation at the chromatin level, in order to uncover how RS affects the expression of the gene.

In particular, in the following pages we will focus on the effects of RS on the disposition and genomic occupancy of polycomb proteins, which have been proven to repress *Dux* in mouse ESCs. Our effort aims at uncovering the changes in chromatin composition promoted by RS, which can have an effect on gene regulation in the short term but can also produce long lasting effects on cell identity.

2 – Materials and methods

2.1 – Cell culture

ESCs of three different lines (E14, R1 and MC1) were grown in feeder-free culture condition in order to avoid the confounding factors due to the co-culture. The standard ESC culture medium was composed of KnockOut DMEM (ThermoFisher, 10829-018), 15% ESC qualified Fetal Bovine Serum (ThermoFisher, 16141-079), 2mM L-glutamine, 0.1mM non-essential amino acids, 0,1mM 2-mercaptoethanol, 50units/mL penicillin and 50mg/mL streptomycin. In order to maintain ESCs in their undifferentiated status the medium was supplemented with home-made leukemia inhibitory factor (LIF) at a dilution of 1:500 and the two inhibitors PD-0325901 (at a final concentration of 1µM) and CHIR-99021 (at a final concentration of 3µM) to inhibit respectively the ERK/MAPK pathway and the GSK-3 kinase. Cells were cultured at 37°C at 5% CO_2 tension. The oxygen tension was maintained at 3% to mimic the conditions of the peri-implantation development (171).

For each experiment, ESCs were seeded 48 hours before samples collection either at 15000 cells/cm^2 or 10000 cells/cm^2 (only for microscopy experiments). ESCs were treated with aphidicolin (Sigma, A0781) at a concentration of 6µM to induce high RS (104) or with a combination of aphidicolin 6µM and a specific inhibitor of ATR (VE-822, Axon Medchem 2452) at the concentration of 1µM. All treatments were performed for 16 hours, maintaining the cells at 37°C under 3% O2 tension and 5% CO_2 tension.

2.2 – Transfection and esiRNA-mediated knock-down

Cells were transfected with the cationic lipid approach, using Lipofectamine RNAiMax (Invitrogen), which allows efficient delivery of both ssRNAs and dsRNAs with very low toxicity even in the presence of antibiotics. Knock-downs were performed taking advantage of RNA interference, an endogenous cellular mechanism that degrades mRNA based on the

interaction with sequence specific dsRNAs. We took advantage of endoribonuclease-prepared small interfering RNAs (esiRNAs). These are generated from an *in vitro* synthesized dsRNA 300 – 600bp long with a sequence specific for the target mRNA (non-redundant sequences with high probability of knock-down are identified by a proprietary algorithm). These long dsRNA are then digested by an endoribonuclease of the RNAse III superfamily to obtain a mixture of dsRNAs ranging from 18bp to 25bp (172). The pool of siRNAs obtained targets specifically the selected mRNA, granting high efficiency and low risk of off-target effects. The following esiRNA pools were used: Ring1 (Sigma-Aldrich, EMU056361), Rnf2 (Sigma-Aldrich, EMU003061), Ezh2 (Sigma-Aldrich, EMU014681), Suz12 (Sigma-Aldrich, EMU050451), Chd4 (Sigma-Aldrich, EMU006021), Rybp (Sigma-Aldrich, EMU094241), Kdm6b (Sigma-Aldrich, EMU206511), Hltf (Sigma-Aldrich, EMU088781), Zranb3 (Sigma-Aldrich, EMU186751) and Smarcal1 (Sigma-Aldrich, EMU051461). In the case of Kdm6a a single siRNA was used (Thermo Fisher, silencer select, s75839).

Lipofectamine was diluted 1:50 in Opti-MEM (ThermoFisher) and incubated 5 minutes at room temperature. In the meantime, the aforementioned esiRNAs were diluted in Opti-MEM to reach a final concentration of 60nM in the well. After incubation time, lipofectamine was added 1:1 to each esiRNA and incubated for 20 minutes at room temperature to allow the incorporation of RNAs inside lipidic micelles.

ESCs were counted with a Countess automated cell counter (Thermo Fisher) and seeded at a cell density of 15000 cells/cm^2 in standard ESC medium the same day of transfection (Trypan blue 4% was added to the counting sample to count only live cells). The mixture of RNA and lipofectamine was finally added to the seeded cells and ESCs were then incubated at 37°C under 3% O2 tension and 5% CO_2 tension for 48 hours before samples collection.

2.3 – RNA extraction, cDNA synthesis and real time qPCR

RNA was extracted using RNeasy Mini Kit (QIAGEN, 74104) taking advantage of the binding properties of silica membranes, and quantified by spectrophotometer (NanoDrop, ThermoFisher) after removal of DNA contaminants by DNAse I digestion (QIAGEN, 79254). cDNA was prepared from 1-2µg of total RNA using the iScript kit (BioRad, 1708890), following manufacturer instructions. This kit provides a hybrid approach based on both oligo-dT and random primers, granting homogeneous coverage of each RNA while at the same time favoring the retrotranscription of mRNA over rRNAs.

The qPCR assay was performed based on a standard protocol using final working concentration of 1X SsoFast™ EvaGreen® Supermix (Bio Rad, 1725201), 0.5µM primer mix and 5ng of sample cDNA. Either TBP or GAPDH were used as internal controls to normalize the qPCR data following ΔΔCt method. Briefly, for each sample each gene was tested in technical triplicate using a specific pair of primers. The average threshold cycle (Ct) for each gene was then normalized for the average Ct calculated for either GAPDH or TBP for the same sample obtaining a ΔCt. The ΔCt for each gene of each sample was then normalized again over the ΔCt for that gene on a control sample obtaining a ΔΔCt. The relative level of mRNA expression was then expressed as fold change ($2^{-\Delta\Delta Ct}$). Data was analyzed using either the one-way ANOVA with Tuckey's post-test correction or the unpaired two-tailed T-test functions of the GraphPad Prism software (GraphPad).

Table of Primers used for RT-real time qPCR		
Gene	**Forward Primer**	**Reverse Primer**
Dux	AAAGGAAGAGCATGTGCCAGC	GCAGTAAGCTGTCCTGGGAAC
Zscan4	GAGATTCATGGAGAGTCTGACTGATGAGTG	GCTGTTGTTTCAAAAGCTTGATGACTTC
MERVL	ATCTCCTGGCACCTGGTATG	AGAAGAAGGCATTTGCCAGA
Tcstv3	ACCAGCTGAAACATCCATCC	CCATGGATCCCTGAAGGTAA
Tbp	AAGAAAGGGAGAATCATGGACC	GAGTAAGTCCTGTGCCGTAAG
Pou5f1	TCTTTCCACCAGGCCCCCGGCTC	TGCGGGCGGACATGGGGAGATCC
Nanog	CAAAGGATGAAGTGCAAGCG	CCAGATGCGTTCACCAGATAG
Ring1	GTATCCAAGCGGTCCCTACG	CCTGATGGGCCTCGTATTCC
Rnf2	CCCGAGTAACAAACGGACCA	ACTCAATCTCACTGGCACCG
Rybp	AGACCAGCGAAACAAACCAC	AGGAGGAGCGAGTCTTTTCC

Suz12	CCTGGAAGTCCTGCTTGTGA	ACTGGAAACTGCCAGGGATG
Ezh2	CAGATAAGGGCACCGCAGAA	ACATTCAGGAGGCAGAGCAC
Chd4	TCCTCTGTCCACCATCATCA	ACCCAAGATGGCCATATCAA
Kdm6a	ATGAACACAGCACAGCAGAAT	AACTGACTTTCCAGCTGTTCCA
Kdm6b	ACATACGGTGGAGGTCCGTA	CTGCAGTGACTCCTGGAAGG
Hltf	AGCTTTATCAGGCTGTCTGGG	TCCATACTGGACAGACTGCG
Zranb3	CTCAGCCAGAGATTGGGCAG	GCAGTTCATTGGCTTTCCATCC
Smarcal1	AAGGCTCTGAAAGACGCTCC	ACAACTTCTCAGCTCGACGG
Rex1	CAAAGACAAGTGGCCAGAAAG	GGAGCTTCCACTCTGGTATTC
Gapdh	CTTTGTCAAGCTCATTTCCTGG	TCTTGCTCAGTGTCCTTGC

2.4 – Flow cytometry (FACS) for cell cycle analysis

ESCs were harvested by trypsinization and collected together with their supernatants in order to take both live and dead cells. Cells were washed once in PBS by centrifugation at 3000 rpm for 5 minutes (the same speed and time was used for each washing step) to remove leftover medium and then resuspended in formaldehyde 2% in PBS and incubated for 20 minutes on ice. At each passage, cells were abundantly resuspended in order to obtain single cells.

After fixation, cells were washed with 1% FBS in PBS and then resuspended in PBS. Three parts of pure ethanol were then added dropwise to the cell suspension while vortexing and kept 30 minutes on ice to complete fixation. Cells were then washed with 1% FBS in PBS and resuspended in PBS with DAPI (4',6-diamidino-2-phenylindole) and kept at 4°C overnight before FACS analysis. Data were acquired using a FACSCanto instrument (BD Biosciences). Cell cycle profile analysis was performed with the FolwJo (BD Biosciences) software.

2.5 – Protein extraction and western blot

Cells were harvested by trypsinization, collected by centrifugation and washed in PBS to remove leftover medium. Cell pellets were either stored at -80°C or directly processed. Briefly, cells were resuspended in RIPA buffer (NaCl 150mM, NP-40 1%, Sodium deoxycholate 0.5%, Sodium dodecyl sulphate 0.1% and Tris pH7.4 25mM in water)

supplemented with a protease/phosphatase inhibitor cocktail (Cell Signaling, 5872) and incubated for 30 minutes at 4°C on a rotating wheel to lyse both cell and nuclear membranes. To complete lysis and break large aggregates the suspension was sonicated with a Bioruptor Sonication System (UCD200) at high power for 15 cycles (30 seconds of sonication with 30 seconds breaks) while kept at 4°C. Lysates were cleared by centrifugation at 13000 rpm for 20 minutes and transferred to clean tubes. Proteins were quantified using a Bradford assay (BioRad) measuring the absorbance at 595 nm.

For protein detection 20 µg of each sample were loaded on 4-20% or 4-15% precast gels (BioRad). Proteins were then transferred on nitrocellulose membranes by semi-dry approach (Trans-Blot Turbo Transfer, BioRad). Filters were blocked with either 5% milk in TBST buffer (Tris pH7.4 20mM, NaCl 150mM, 0.1% Tween) or 5% BSA in TBST (for detection of phosphorylated proteins) for 1 hour at room temperature on agitation. Blocking is required to occupy aspecific binding sites for antibodies, which may generate background signal. The high presence of phosphorylated residues in milk casein renders milk not suited for blocking in cases when phosphorylated proteins are to be detected. After blocking filters were incubated with primary antibodies overnight at 4°C on rotation, washed three times with TBST and then incubated with secondary antibodies conjugated with HRP (horse radish peroxidase) for 1 hour at room temperature on agitation. Proteins were detected using the ECL (Amersham) luciferase-based kit and acquired on photographic films.

The following antibodies were used: anti-p-CHK1 (S317) (cell signaling, 12302), anti-ZSCAN4 (Merck Millipore, AB4340), anti-CHK1 (Santa Cruz, sc8408), anti-GAPDH (Sigma-Aldrich, G8795), anti-H2A (Merck Millipore, 07-146), anti-HLTF (abcam, ab17984), anti-SMARCAL1 (Santa Cruz, sc376377), anti-ZRANB3 (Proteintech, 23111-1-AP), anti-H3K27me3 (Cell Signaling, 9733) and anti-H2AK119Ub (Cell Signaling, 8240).

2.6 – Chromatin immunoprecipitation and real time qPCR (ChIP-qPCR)

ChIP assay was performed using the SimpleChIP Enzymatic Chromatin IP Kit (Cell Signaling, #9003) following manifacturer's instructions. Briefly cells were fixed while still in culture by adding formaldehyde to a final concentration of 1% and incubating for 10 minutes at room temperature with mild shacking to create covalent bonds between chromatin proteins and their target DNA loci. Formaldehyde was quenched for 5 minutes with glycine at room temperature to remove excess formaldehyde and stop the fixation process. Medium was removed and plates were washed twice with cold PBS. Cells were harvested by scraping in cold PBS supplemented with protease inhibitors (Cell Signaling, 5872) and centrifuged before removing the supernatant. Cell pellets were then stored at -80°C till the time of processing.

Pellets were then thawed and resuspended in a lysis buffer (Cell Signaling) supplemented with protease inhibitors and DTT 0.5mM. Resuspended pellets were incubated 10 minutes on ice with frequent shaking to dissolve cell membranes without affecting nuclei. Lysis buffer was then removed by centrifugation (2000g for 5 minutes at 4°C) and cells were resuspended in Micrococcal nuclease working buffer (Cell signaling) supplemented with DTT 0.5mM. Chromatin was fragmented by Micrococcal nuclease treatment for 20 minutes at 37°C with frequent shaking. Fragments dimension was verified on an agarose gel and generally ranged from 150bp to 500bp. Fragment size assessment is important, too small fragments result in inefficient PCR amplification in the analysis step and loss of signal, while too large fragments are poorly precipitated in the immunoprecipitation step. Nuclease digestion was stopped by addition of EDTA (Ethylenediaminetetraacetic acid) to sequester bivalent ions necessary for nuclease activity and by rapid transfer to ice. Digested chromatin was pelleted by centrifugation for 2 minutes at 16000g at 4°C and resuspended

in ChIP buffer (Cell Signaling) supplemented with protease inhibitors. Nuclear membranes were dissolved by sonication for 4 cycles of 10 seconds at 10% amplitude with 1 minute breaks on ice using a tip sonicator (Branson, SFX150). Lysates were then cleared by centrifugation at 4900g for 10 minutes at 4°C to remove membranes and large chromatin fragments and supernatants were transferred in clean tubes.

Fragmented chromatin was diluted 1:4 in ChIP buffer supplemented with protease inhibitors and divided in tubes in order to have an amount of chromatin correspondent to about $4*10^6$ original cells for each Immunoprecipitation (IP) reaction. A volume correspondent to 2% of an IP was taken as input for each sample analyzed and kept at 4°C. The following antibodies were added to each IP at the concentration suggested by the manufacturer: anti-RING1b (Cell Signaling, 5694), anti-RYBP (Merck Millipore, AB3637), anti-H3K27me3 (Cell Signaling, 9733), anti-H2AK119Ub (Cell Signaling, 8240) and as a negative control normal rabbit IgG (Cell Signaling, 2729). IPs were incubated with antibodies overnight at 4°C on rotation in order to allow antibodies to reach a binding equilibrium with their targets. Protein-G conjugated magnetic beads (Cell Signaling, 9006) were used to precipitate the antibodies-target complexes by taking advantage of protein-G ability to interact with the constant portion of IgG antibodies. After a 2 hours incubation at 4°C with rotation beads were collected with a magnetic rack and subjected to three low salt washes with ChIP buffer and one high salt wash with ChIP buffer supplemented with 350mM NaCl to displace unspecific binding. Samples were eluted from the beads by incubation in an SDS-containing elution buffer by incubation at 65°C for 30 minutes with vortexing (1200rpm). In this passage, high temperature, SDS presence and shacking cause denaturation and loss of interaction between the beads and the immunoprecipitated complexes which are released in solution. Input samples were processed together with IPs. Eluted chromatin was reverse cross-liked by adding NaCl at a final concentration of 200mM

and 40µg of proteinase K, and incubating for 2 hours at 65°C with vortexing (700rpm). High temperature and salt presence in fact causes the loss of covalent bounds generated by formaldehyde while proteinase K-mediated degradation of proteins furtherly favors DNA release. DNA was then extracted by de-cross-linked chromatin using spin columns, taking advantage of the binding properties of silica membranes. DNA was eluted in Tris-EDTA buffer (TE) to protect it from nuclease degradation and stored at -20°C.

The DNA was analyzed by real time qPCR loading 1 µL from each IP and input for each tested locus. The Ct obtained for each gene in technical triplicate was averaged for inputs and then adjusted to 100% (in the 2% input case the 100% signal is obtained by subtracting 5.644 to the original Ct). The average Ct for each locus of each IP was then normalized on the correspondent adjusted Ct for the input for the same locus by subtracting the IP Ct from the input Ct. Finally the obtained ΔCt is expressed as percentage over input using the formula $100*2^{-\Delta Ct}$. Data was analyzed using the unpaired two-tailed T-test function of the GraphPad Prism software (GraphPad).

2.7 – Confocal microscopy and image analysis

ESCs cultured on standard cell culture dishes were washed with room temperature PBS and rapidly fixed with 4% formaldehyde in PBS for 20 minutes on ice. Fixed cells were washed three times with PBS to remove completely the formaldehyde, waiting 5 minutes on ice between washes. Cells were then incubated for 40 minutes in 10% FBS, 0.1% Triton X-100 in PBS to both permeabilize and block the samples. Cell membranes permeabilization is necessary to allow antibodies to enter the cells and stain intracellular proteins, while blocking impairs protein-protein unspecific interactions, reducing background. Primary antibodies were diluted in a solution of 10% FBS in PBS. Samples were then incubated overnight with primary antibodies at 4°C with mild agitation. After primary antibodies incubation, samples were washed twice with PBS, waiting 5 minutes between washes, and

then incubated with secondary antibodies conjugated with fluorophores and DAPI diluted in 10% FBS in PBS for 1 hour repaired from light. The secondary antibodies used are anti-rabbit conjugated with cyanine 3 and anti-mouse conjugated with Alexa Fluor 647. After secondary antibodies incubation, samples were washed three times with PBS waiting 5 minutes between washes and then mounted using the Vectamount reagent (Vector labs). Samples were covered with microscopy coverglasses (VWR, 631-0174) and let to dry at room temperature, repaired from light, before storing at 4°C till the day of the analysis.

Pictures were acquired using a confocal microscope (Leica, TCS SP5) using a 63X oil immersion objective with numerical aperture 1.4 acquiring Z-stacks. Analysis was then performed using the colocalization threshold function of the Fiji software. ROIs were defined using the DAPI signal for each cell and colocalization between signal of two channels was estimated throughout the entire Z-stack for each cell. The Rcoloc (Pearson coefficient of colocalization after threshold is removed) was used as an indication of colocalization. Total intensity of fluorescence signal for each channel was also calculated for each cell by measuring the sum of the total signal intensity inside the DAPI-defined ROI for the entire Z-stack. Statistical analyses were performed using python packages Numpy (version 1.17.0), Pandas (1.1.3) and Scipy (version 1.4.5) using the built-in unpaired two tailed T-test and the hypergeometric test functions.

2.8 – RNA Sequencing

RNA was extracted using RNeasy Mini Kit (QIAGEN, 74104) and DNA contaminants were removed by DNAse I digestion (QIAGEN, 79254), analogously to what indicated in the case of RNA extraction for qPCR analysis. Each sample was quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity was checked with RNA Screen Tape on an Agilent 2200 TapeStation instrument (Agilent Technologies). The RIN algorithm was used to compare the 18S and 28S peaks in samples electropherogram

to assess eventual degradation of the samples (173). Only samples scoring 10 (corresponding to negligible degradation) were used for the analysis. rRNAs were then removed through the Ribo-Zero kit (Illumina) which takes advantage of targeted hybridization with sequence specific DNA probes and RNAse H-mediated degradation to deplete rRNAs from samples. In this way mRNAs were enriched in the sample and higher sequencing depth could be achieved. Libraries were then prepared using the TruSeq Stranded Total RNA library Prep kit (Illumina, RS-122–2101). Briefly, the rRNA depleted samples were fragmented for 8 minutes at 94°C, before first strand cDNA synthesis with random hexamer primers. RNA was then degraded and a second cDNA strand was synthesized using dUTPs instead of dTTPs. Incorporation of dUTP does not allow the amplification of the second strand in the following passages and consent to distinguish the two strands of the cDNA. The newly synthesized cDNAs were purified with AMPure XP beads (Beckman) and monoadenylated on the 3' end to remove blunt extremities and avoid ligation of multiple fragments during the adapter ligation. Adapters containing a 3'-T overhang were then ligated to the fragments, obtaining blunt cDNAs. A PCR reaction using primer pairs containing barcoded indexes and able to pair with adapters was used to amplify the library (in this step only the first strand could be amplified). In this step amplification was limited to 15 cycles to reduce artifacts due to PCR amplification. After purification of the cDNA libraries these were validated using a DNA Analysis Screen Tape on the Agilent 2200 TapeStation (Agilent Technologies) and quantified by using Qubit 2.0 Fluorometer (Invitrogen) as well as by qPCR (Applied Biosystems).

Barcoded libraries were then normalized to the same concentration, pooled and loaded on a single flow-cell lane. Clusters were generated and sequencing was performed with an Illumina HiSeq4000 sequencer (Illumina) using a 2 x 150 Pair-end high throughput configuration. Raw machine files were de-multiplexed using the bcl2fastq program (version

2.17) to generate single FASTQ files for each sample. One mis-match was allowed for index sequence identification. About 29 million reads were generated for each sample.

Transcript and gene level quantification was performed with the Kallisto package (version 0.43.0) using the RefSeq collection to build the transcriptome index (174,175). Kallisto was run with the --bias option to perform sequence-based bias correction on each sample. Additionally, all reads were mapped with the STAR spliced aligner (version 2.5.2b) (176) to the GRCm38 (mm10) genome after trimming the reads to 100 bp with Trimmomatic (version 0.36) (177). Finally, read counts were imported in the R environment (version 3.2.2) for differential gene expression analysis using the limma package (version 3.24.15) (178). All samples were normalized using the trimmed mean of M-values (TMM) method and voom transformation. A linear model was fitted to the data with the limma lmFit function, and the moderated t-statistics was calculated with the eBayes function. Genes were defined as differentially expressed if they had FDR adjusted p-values<0.05 and |log2 fold change| > 1.

Enrichment analysis was performed using the webtool of the enrichR package (179,180). Comparison between our differentially expressed genes and published lists of Polycomb targets was performed in a python environment using the hypergeometric test function of the Scipy package (version 1.5.4).

2.9 – Cut&Tag

Cut&Tag was performed according to Hatice et al. (181). Cells cultured as previously indicated were harvested by trypsinization and counted with a Countess automated cell counter (Thermo Fisher). Trypan blue 4% was added to the counting sample to count only live cells. Cells were then pelleted by centrifugation at 600g for 3 minutes at room temperature and resuspended in a wash buffer (20mM HEPES pH7.4, 150mM NaCl, 0.5mM spermidine in water, supplemented with protease inhibitors).

In the meantime, concavalin A-coated magnetic beads (Bangs Laboratories, Inc.) were washed twice with binding buffer (20mM HEPES pH7.4, 10mM KCl, 1mM $CaCl_2$ and 1mM $MnCl_2$ in water) using a magnetic rack to separate the beads from the supernatant. Washed beads were resuspended in a volume of binding buffer equal to the original volume of the beads and the added dropwise to the cell suspension at room temperature while gently vortexing. For each 150000 cells, corresponding to one reaction, 15µL of beads were used. The suspension was then placed on rotation for 10 minutes at room temperature to allow binding of the cells to the beads. The binding of cellular membrane glycoproteins to concavalin A allows the cells to stick to the magnetic beads and reduces the loss of material during the samples handling.

The beads were then separated from the liquid using a magnetic rack and resuspended in ice-cold antibody buffer (wash buffer supplemented with 0.05% digitonin, 2mM EDTA and 0.1% BSA) while vortexing, and then aliquoted in as many vials as the number of reactions required. Aliquots of 100µL containing about the equivalent of 150000 cells were made and kept on ice. To each aliquot 1µL of primary antibody was added while gently vortexing before incubating overnight at 4°C on a nutator. During this step, the detergent digitonin acted to permeabilize cell membranes, allowing primary antibodies to penetrate the cells and bind to their targets.

After this step beads were separated from the supernatant using a magnetic rack and then resuspended in 100µL per sample of a dilution 1:100 of secondary antibody (Antibodies online ABIN101961) in Dig-wash buffer (wash buffer supplemented with 0.05% digitonin) while vortexing. Samples were then incubated for 1 hour at room temperature on a nutator. This passage is not strictly required for a successful Cut&Tag, but the secondary antibody amplifies signal and overall improves the signal-to-noise ratio of the technique.

After this passage, the beads were separated from the liquid using a magnetic rack and washed three times with an excess volume of Dig-wash buffer, inverting the tubes to dislodge the beads every time.

Protein A-Tn5 transposase protein fusions (pA-Tn5)(Vazyme biotech, S603) pre-loaded with adapters suited for Illumina library preparation kits was diluted 1:250 in Dig-300 buffer (20mM HEPES pH7.4, 300mM NaCl, 0.5mM spermidine, 0.01% digitonin in water, supplemented with protease inhibitors). Each fusion protein is loaded with two different cDNA tags with 5'-overhangs that can be inserted in the genome by the enzyme.

After the last wash, the beads were resuspended in 100µL per sample of Dig-300 buffer containing the pA-Tn5 complexes, while gently vortexing. Samples were then incubated for 1 hour at room temperature on a nutator. During this passage, the interaction between protein-A and the constant portion of the secondary antibodies anchors the Tn5 transposons in proximity of the target sites recognized by the primary antibody. However, in the absence of bivalent cations the Tn5 transposase is maintained inactive.

After incubation, the samples were washed three times with Dig-300 buffer to remove the excess of transposase and then resuspended in 300µL per sample of Tagmentation buffer (Dig-300 buffer supplemented with 10mM $MgCl_2$) while gently vortexing. Samples were then incubated at 37°C for 1 hour to allow the Tn5 transposase to insert the DNA adapters in proximity to the target sites, tagging the regions and at the same time causing DNA fragmentation, in a process that is referred to as tagmentation. The reaction was stopped by adding 10µL of 0.5M EDTA, 3µL of 10% SDS and 2.5µL of 20 mg/mL proteinase K to each sample. Samples were then vortexed at high speed and incubated for 1 hour at 50°C. The EDTA quenches the bivalent cations, stopping the transposase reaction, while SDS denaturates proteins (among which the Tn5 transposase) and the proteinase K digest most proteins, eliminating the transposase and all contaminating proteins and freeing the DNA.

The solubilized DNA was then isolated using phenol-chloroform extraction. Briefly, 300µL of PCI (phenol-chloroform:isoamyl alcohol in a ratio 24:1) was added to each sample before high speed vortexing. Samples were centrifuged at 16000g for 3 minutes at room temperature before adding 300µL of chloroform per sample and mixing by inverting the tubes. Samples were again centrifuged at 16000g for 3 minutes at room temperature and the aqueous phase, containing the DNA, was transferred in new tubes containing three parts of absolute ethanol. Samples were chilled on ice and then centrifuged for 15 minutes at 16000g and 4°C to pellet DNA. The liquid was removed and the DNA pellets were washed with absolute ethanol before air-dry them and resuspend them in 25µL of TE buffer supplemented with 1:400 RNAse A. Samples were then incubated for 10 minutes at 37°C to degrade potential RNA contaminants before storing them at 4°C until further use.

Libraries were prepared by PCR using universal i5 and uniquely barcoded i7 Nextera primers (Illumina), both at 10µM concentration (182) and a NEBNext HiFi 2x PCR Master mix (New England Biolabs). The PCR cycles were as follows:

Cycle 1: 72°C for 5 min (gap filling)

Cycle 2: 98°C for 30 sec

Cycle 3: 98°C for 10 sec

Cycle 4: 63°C for 10 sec

Repeat Cycles 3-4 13 times

72°C for 1 min and hold at 8°C

Since the DNA tags inserted by the Tn5 transposase can pair specifically either with the i5 or the i7 primers all and only the DNA fragments encompassed by two different Tn5-inserted tags can be amplified in this passage. Moreover, by its nature, the PCR reaction favors the amplification of shorter fragments over longer ones, excluding large fragments that could not be sequenced efficiently. Libraries were then purified using AMPure XP

beads (Beckman), validated and quantified using an Agilent 2100 BioAnalyzer (Agilent Technologies) with HS Dna reagents (Agilent Technologies). Libraries were normalized and pooled before loading on the flow cell. Sequencing was performed with an Illumina NextSeq 550 sequencer (Illumina) using a 2 x 75 Pair-end configuration obtaining around 22 million reads per sample.

Raw reads were aligned to the GRCm38 (mm10) mouse reference genome using bowtie2 (183) using the parameters –local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I 10 -X 700. Reads from samples of interest were compared with input samples in which only normal rabbit IgGs were used. Peaks were called using the macs2 package (184) with the parameters -p 0.01 -g mm --keep-dup 1. Both H3K27me3 and H2AK119Ub peaks were called with the --broad parameter. Target genes were identified according to the presence of a peak in a window of +/- 1.5kb (EED and Ring1b) or +/- 5kb (H2AK119Ub and H3K27me3) and intersections were computed with Bedtools. Bigwig files generated by macs2 were used to produce heatmaps and density plots using deepTools2 (185). Bigwig files were visualized using the Igvtools software (186) and other graphs were generated using ad-hoc scripts in a python environment.

3 - Results

3.1 – RS causes the expression of several PRC target genes in ESCs

ESCs are fast replicating cells which are already subjected to high levels of RS in a normal cell cycle in comparison to more differentiated cells. In our recent work (42) we investigated the consequences of RS in mouse ESCs, focusing mainly on the effects of the ATR checkpoint activation. Importantly, cell cycle profiling by FACS analysis on both E14 and R1 ESCs in unperturbed conditions showed that more than 50% of these cells are in S-phase, while a very small percentage (from 17.5% to 15%) is in G1-phase (Fig. 6), confirming previous reports indicating that ESCs undergo a very fast cell cycle with an extremely short G1-phase (170). Interestingly, in our experimental setup, aphidicolin treatment led to an increase in the percentage of cells in G1 phase to 48.4% and 39.5%, in E14 and R1 respectively (Fig. 6). This is quite surprising considering that ESCs have been shown to have a defective G1 checkpoint due to a high amount of hyperphosphorylated Rb and an uncoupling of the E2F4 activity from Rb regulation (187–189). Notably, this block did not appear to be extremely marked, since around 40% of ESCs had a DNA content compatible with S-phase (Fig. 6), as it is expected from cells affected by aphidicolin-mediated block of DNA polymerases.

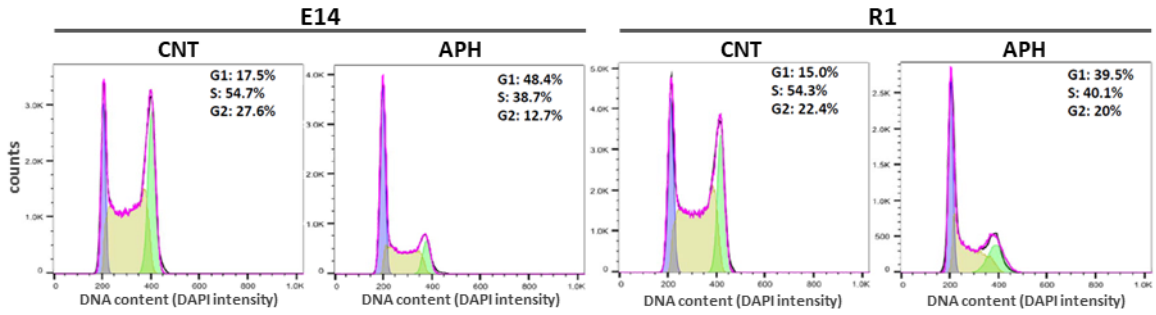

Figure 6 | Cell cycle changes induced by aphidicolin in ESCs: FACS analysis of DAPI-stained E14 and R1 ESCs either in unperturbed conditions (CNT) or after 16 hours aphidicolin 6µM treatment (APH). DAPI signal is used as a measure of DNA content across cell cycle phases.

To exclude effects due to the onset of differentiation programs, we verified the expression of pluripotency markers in E14 and R1 cells treated with aphidicolin 6µM for 16 hours in comparison to their untreated controls at the mRNA level by reverse transcription (RT) coupled with real time qPCR. As shown in figure 7, all the three pluripotency markers *Pou5f1*, *Nanog* and *Zfp42* were not significantly affected by aphidicolin treatment in any of the tested ESC lines at the mRNA level. Similarly, aphidicolin treatment did not completely suppress the expression of the pluripotency markers OCT4 (encoded by the *Pou5f1* gene) and SOX2 at the protein level, although a reduction was observed for both (Fig. 8).

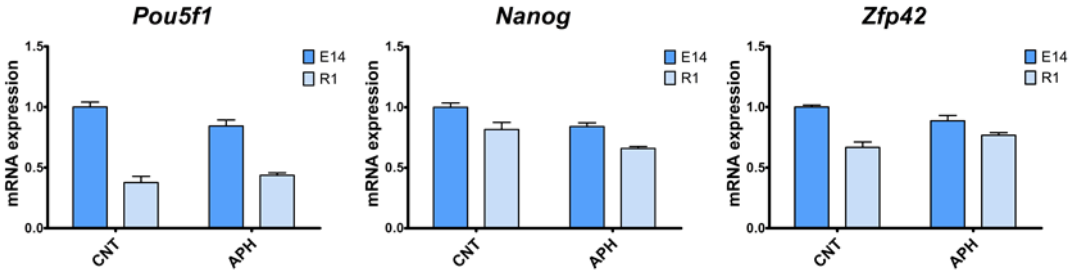

Figure 7 | Replication stress does not affect the expression of pluripotency genes at the mRNA level: RT-real time qPCR analysis of in order *Pou5f1*, *Nanog* and *Zfp42* transcripts in E14 and R1 ESCs either unperturbed (CNT) or after 16 hours aphidicolin 6µM treatment (APH). All data is normalized on *Gapdh* expression. Data is expressed as mean + SEM of three independent experiments and analyzed by unpaired two-tailed t-student test.

However, this reduction is in agreement with the increased presence of 2C-like cells among ESCs treated with aphidicolin (42), confirmed by the higher expression of the 2-cells stage protein ZSCAN4 (Fig. 8). In fact, as previously shown by us and others (35,42), 2C-like cells do not express pluripotency markers such as OCT4, NANOG or SOX2 at the protein level, although almost no change can be observed at the transcript level. These data indicate that the majority of ESCs still retain their undifferentiated nature after aphidicolin treatment and the accumulation of cells in G1 that we observed is therefore most likely due to an increasing number of cells stuck in the early stage of S-phase, unable to efficiently synthesize new DNA due to the inhibitory effect of aphidicolin on DNA polymerases.

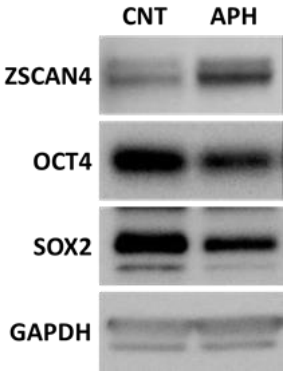

Figure 8 | Replication stress mildly affects the expression of pluripotency genes at the protein level: Western blot analysis of ESCs either unperturbed (CNT) or after 16 hours of aphidicolin 6μM treatment (APH). Antibodies against ZSCAN4, OCT4 and SOX2 were used for the analysis. GAPDH was used as a loading control

In our effort to characterize the effect of RS on ESCs, we subjected three different ESC lines to RS induced by aphidicolin (6μM for 16 hours) and performed whole transcriptome profiling by bulk RNA-Seq (Fig. 9a). We identified 3714 differentially expressed genes (DEGs) upon RS induction, which were common to the three lines. Of these, 3074 genes were up-regulated by RS while 640 were downregulated (Fig. 9a). Our previous analysis identified a significantly enriched set of 2 cells-stage specific genes among the genes controlled by RS (42). Further enrichment analysis using the R-based online tool Enrichr (179,180), comparing our set of RS-induced genes with the ChEA and ESCAPE databases, highlighted a significant enrichment of Polycomb targets among our DEGs (Fig. 9b). In particular, we could observe significant enrichment among the targets of PRC2 components SUZ12 and EZH2, together with the corresponding histone mark H3K27me3 (Fig. 9b). Additionally, we manually compared our list of RS-dependent DEGs with published sets of both PRC1 and PRC2 targets. For this purpose, we created two sets of genes which were reported either as targets of RING1b, a main component of PRC1, or SUZ12, which is part of the core PRC2 complexes, in mouse ESCs cultured under standard conditions. For our comparison we used only genes which appeared to be targets of either RING1b or SUZ12 in more than a dataset, in order to have only *bona fide* candidates (141,190,191).

Interestingly, also in this case our gene-set showed significant overlap with both target lists as per hypergeometric test (Fig. 9c, d) indicating that a significant portion of the genes which we identified as differentially regulated by RS are occupied by Polycomb complexes in unperturbed mouse ESCs. Of note, also the two sets of RING1b and SUZ12 targets showed significant overlap with each other, with 2969 genes resulting to be targets of both Polycomb complexes in ESCs, and so confirming that PRC1 and PRC2 often co-occur on the same promoters (Fig. 9e).

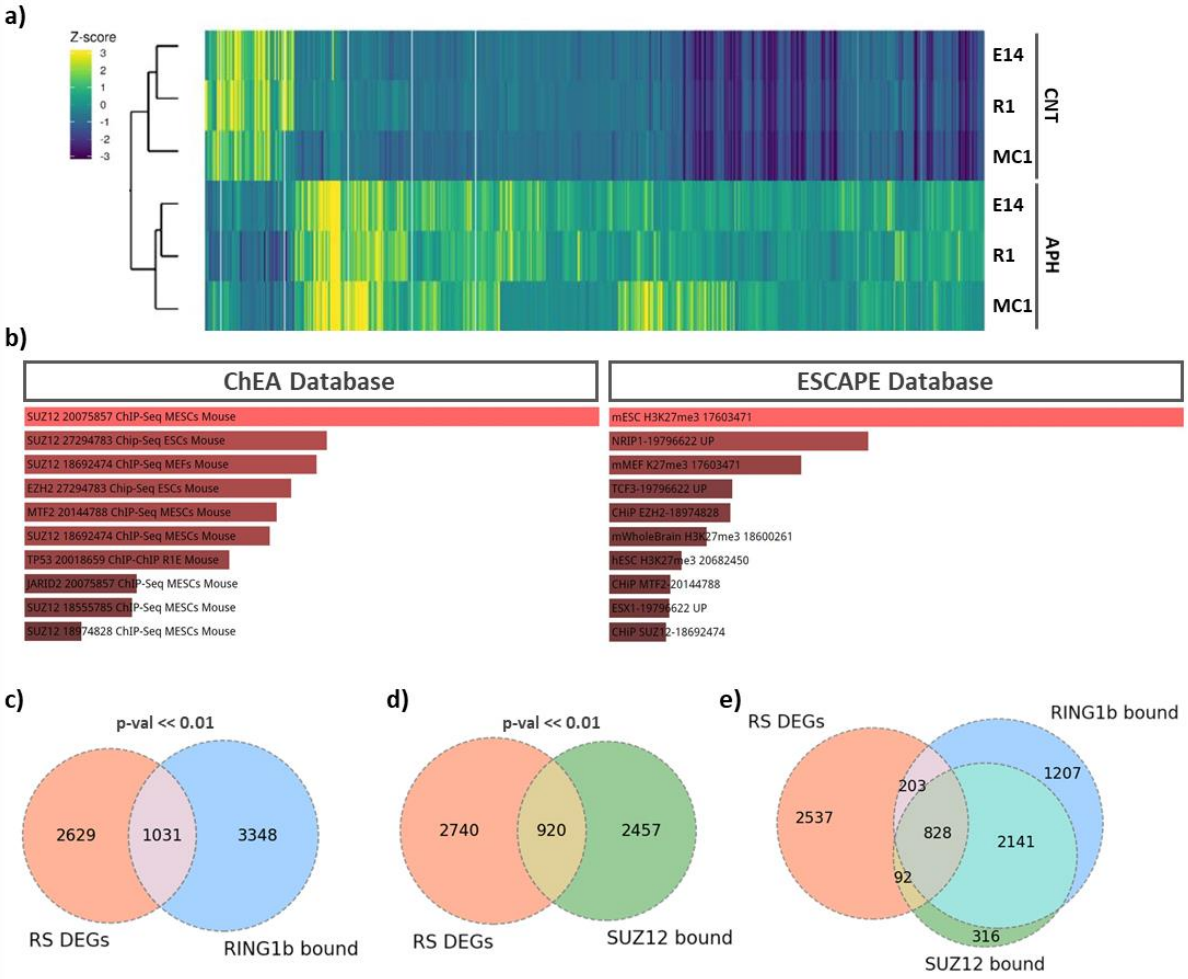

Figure 9 | Polycomb controls a large subset of genes responsive to RS in mouse ESCs: a) Heatmap representing the clustering of three different ESC lines (E14, R1, MC1) in unperturbed conditions (CNT) or treated with aphidicolin 6µM for 16 hours (APH) based on the expression level of genes measured by bulk RNA-Seq. Only genes which display significant variation between CNT and APH are shown. Expression levels are color-coded according to Z-score. **b)** Enrichr analysis of genes differentially regulated in mouse ESCs upon RS induction using the ChEA and the ESCAPE databases for chromatin bound factors. Terms are ordered by ranked p-value. **c-d)** Proportions of genes overlapping among RING1b targets or SUZ12 targets identified by analyses present in literature and our group of RS controlled genes. Significance of the overlap was calculated by hypergeometric test. **e)** comparison among genes controlled by RS in our previous analysis and target genes of RING1b and SUZ12 present in the literature.

Given the notorious repressive action of Polycomb complexes, these evidences suggest that loss of Polycomb repression could be at least in part responsible for the differential regulation of many of the genes responsive to RS in mouse ESCs.

3.2 – PRC1 is responsible for the repression of *Dux* and its targets in mouse ESCs

In previously published works, it has been shown that Polycomb complexes, and in particular PRC1, are responsible for the repression of *Dux* and its target genes in mouse ESCs (192). Interestingly, RYBP has been proven among the PRC components involved in *Dux* repression, demonstrating the importance of variant PRC in 2-cells stage genes repression (192). At the same time, the NuRD complex has been shown to repress the human homolog of *Dux*, *DUX4* in human induced pluripotent stem cells (IPSCs) and myoblasts (72). To confirm this evidence, we tested the effect of the ablation of PRC1 core components *Ring1* (encoding the protein RING1a) and *Rnf2* (encoding RING1b), and the vPRC1 subunit *Rybp* by esiRNA-mediated knock-down in two different ESC lines. As a negative control, we transfected our cells with an esiRNA pool designed against the Renilla luciferase gene (*Rluc*). Both E14 and R1 cells were transfected at seeding and harvested 48 hours post-transfection for RNA extraction, retrotranscription (RT) and real time qPCR analysis.

As shown in figure 10a, ablation of both *Rnf2* and *Rybp* resulted in strong up-regulation of *Dux* in both ESC lines, although the extent of this increase varied broadly between the two lines. Similarly, this induction was mirrored by the activation of *Dux* main target *Zscan4*, together with the other 2-cells stage specific gene *Tcstv3* and the *MERVL* endogenous retroelements (Fig. 10b-d). Interestingly, *Rnf2* knock-down resulted in a much stronger up-regulation of all the 2C genes tested in comparison to *Rybp* depletion, consistently with the much more fundamental role of RING1b in the composition of PRC1 complexes (Fig. 10a-

d). Interestingly, the ablation of the other PRC1 core component *Ring1* did not have any effect on Dux expression, displaying a stark contrast with the depletion of its paralog *Rnf2*. In fact, cells lacking *Ring1* displayed *Dux* mRNA levels very similar to the negative control in both cell lines (Fig. 10a). An analogous lack of effect was observed also in the case of *Dux* targets *Zscan4*, *Tcstv3* and *MERVL* (Fig. 10b-d). The efficiency of the knock-down was verified by measuring the mRNA levels of *Ring1*, *Rnf2* and *Rybp* by RT-real time qPCR in both E14 and R1 cells, transfected with esiRNA targeting each of these genes, and comparing them with the expression level measured in cells transfected with the *Rluc*-targeting esiRNA. The extent of knock-down proved to be extremely efficient for both *Rnf2* (Fig. 10f) and *Rybp* (Fig. 10g), bringing the mRNA level to respectively about 10% and 20% of the negative control levels, with an extremely similar effect in both cell lines. However, in the case of *Ring1,* esiRNA administration resulted only in a reduction of almost 50% of control levels in the case of E14 and 30% for R1 cells (Fig. 10e). The limited extent of knock-down in the case of *Ring1* may be taken as an explanation for its ineffectiveness in activating *Dux* and its targets. However, it is important to observe that, in the case of R1 cells, a reduction of *Rybp* to 20% of control levels results in a clear and significant activation of *Dux* (Fig. 10a), while depletion of *Ring1* to 30% of what observed in control cells corresponds to the same level of *Dux* expression measured in the control (Fig. 10a). According to these data, it is unlikely that the absence of *Dux* activation upon *Ring1* knock-down may be simply due to inefficient depletion of the mRNA. In their complex, these data prove that the composition of PRC1 complexes responsible for 2C genes repression in ESCs is highly specific and involves the action of vPRC1 complexes.

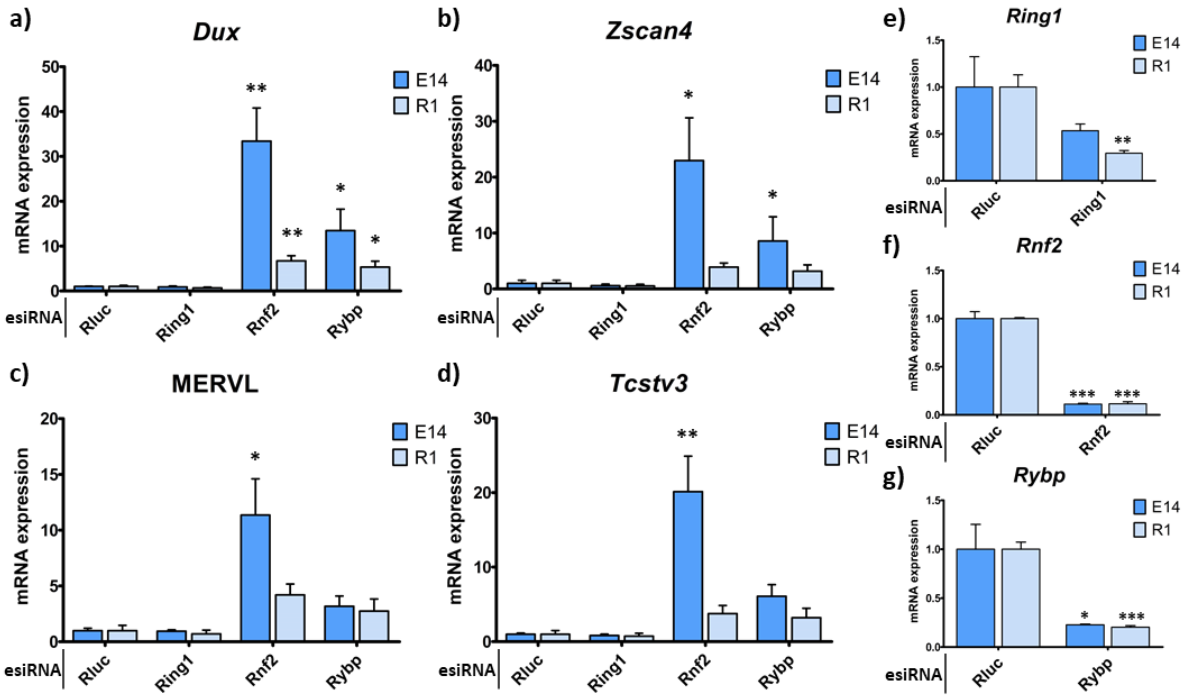

Figure 10 | Variant PRC1 complexes repress *Dux* in mouse ESCs: RT-real time qPCR analysis measuring mRNA expression levels in E14 and R1 cells transfected with the indicated esiRNA pools respectively for the genes **a)** *Dux*, **b)** *Zscan4*, **c)** MERVL retroelements, **d)** *Tcstv3*, **e)** *Ring1*, **f)** *Rnf2* and **g)** *Rybp*. mRNA levels were normalized on *Tbp* expression and expressed as fold change in respect to the Rluc sample for the corresponding cell line. All data are shown as mean + SEM of three independent experiments. mRNA expression data were analyzed by One-way ANOVA followed by Tukey post-test when applicable or by unpaired two-tailed t-student test (* = p-val < 0.05, ** = p-val < 0.01, *** = p-val < 0.001).

Since in ESCs the two Polycomb complexes PRC1 and PRC2 are found often together on gene promoters, we wondered if also PRC2 complexes may be recruited to *Dux* promoter through the presence of PRC1, and therefore be involved in the repression of *Dux* as well. To this purpose, we tested also the effect of the ablation of two main PRC2 components, *Suz12* and *Ezh2*, by esiRNA-mediated knock-down in both E14 and R1 ESC lines. To verify if NuRD may have in mouse ESCs a similar effect to what observed in human IPSCs, we knocked-down in parallel the NuRD component *Chd4*. Also in this case, cells were transfected at seeding and were then harvested 48 hours post-transfection for RNA extraction, retrotranscription and real time qPCR analysis. *Tbp* expression was used for normalization. Differently from what we observed in the case of PRC1 knock-down, ablation of neither PRC2 components *Suz12* and *Ezh2* or NuRD subunit *Chd4* could increase the expression of *Dux* (Fig. 11a) or its target genes (Fig. 11b-d) in the two tested cell lines.

Interestingly, depletion of PRC2 or CHD4 resulted in a slightly lower expression of *Dux* (Fig. 11a), that was even more evident in the case of *Zscan4* (Fig. 11b). This peculiar behavior was observed in both cell lines tested, although it did not result significant.

Also in this case, the efficiency of knock-down was evaluated by comparing the levels of mRNA measured for *Suz12*, *Ezh2* and *Chd4* in cells transfected with the respective esiRNA with their level in cells transfected with the *Rluc* esiRNA. Both *Suz12* and *Ezh2* knock-down resulted in efficient depletion of the respective mRNAs, leading to less than 10% of expression in comparison to the control (Fig. 11e, f) and therefore confirming that the lack of *Dux* activation could not be attributed to inefficient knock-down. *Chd4* knock-down showed instead a slightly different efficiency in E14 and R1 cells, leading to a drop in the mRNA levels to around 30% and 40% of control levels respectively (Fig. 11g).

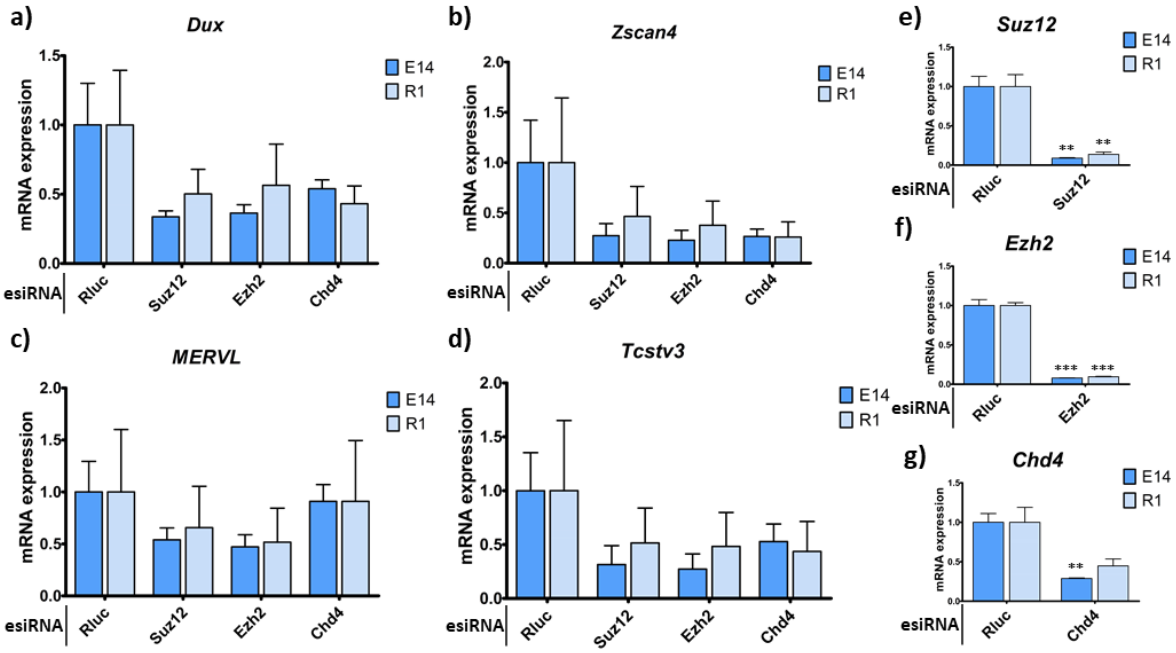

Figure 11 | PRC2 and NuRD complexes do not repress *Dux* in mouse ESCs: RT-real time qPCR analysis measuring mRNA expression levels in E14 and R1 cells transfected with the indicated esiRNA pools respectively for the genes **a)** *Dux*, **b)** *Zscan4*, **c)** MERVL retroelements, **d)** *Tcstv3*, **e)** *Suz12*, **f)** *Ezh2* and **g)** *Chd4*. mRNA levels were normalized on *Tbp* expression and expressed as fold change in respect to the Rluc sample for the corresponding cell line. All data are shown as mean + SEM of three independent experiments. mRNA expression data were analyzed by One-way ANOVA followed by Tukey post-test when applicable or by unpaired two-tailed t-student test (* = p-val < 0.05, ** = p-val < 0.01, *** = p-val < 0.001).

These data show a clear difference in roles between the two families of Polycomb complexes for what regards 2-cells stage genes repression. Interestingly, this evidence suggests also that, differently from humans, in mouse ESCs 2C genes are not repressed by NuRD. These data seems to highlight a species-specific difference among humans and mice, although it cannot be excluded that this difference may be linked to the different developmental stages represented by human and mouse ESCs.

Polycomb complexes have been shown to be involved in replication, and their absence may impair the correct proceeding of replication forks, leading to RS and genome instability (163,165,167). In the light of these evidences we wondered whether the strong up-regulation of *Dux* that we observed upon *Rnf2* and *Rybp* depletion may be due to RS induced by their absence, which in turn would activate the ATR-CHK1 axis, responsible for 2C genes activation, as shown by our previous work (42). To test this hypothesis, we collected proteins from E14 ESCs grown in unperturbed conditions and transfected with esiRNAs targeting *Ring1*, *Rnf2* or *Rybp* 48 hours before harvesting. As a negative control, we transfected cells with an esiRNA pool targeting *Rluc*, while as a positive control we treated ESCs transfected with the *Rluc* esiRNA with aphidicolin at the concentration of 6μM for 16 hours to strongly induce RS. Total protein extracts were produced from the collected samples and analyzed by western blot. Ablation of RING1b was confirmed by complete loss of its protein signal in western blot, while RYBP knock-down resulted only in diminished signal (Fig. 12), in agreement with what we measured at the mRNA level by qPCR. Aphidicolin treatment resulted in a clear activation of the ATR-CHK1 checkpoint in respect to the negative control, as shown by CHK1 phosphorylation at serine 317 (Fig. 12). Knock-down of none of the PRC1 components tested resulted in ATR-CHK1 activation as displayed by the phospho-CHK1 signal (Fig. 12). Interestingly, even if none of the knock-downs performed resulted in ATR-CHK1 activation, both *Rnf2* and *Rybp* depletion led to a stronger

expression of the ZSCAN4 protein in comparison to aphidicolin-induced RS, indicating a much stronger up-regulation of 2C genes. Notably, analogously to what we observed in the case of mRNA levels, the *Rnf2* ablation resulted in stronger ZSCAN4 expression in comparison to *Rybp* knock-down (Fig. 12). This may be due to the more relevant and ubiquitous role of RING1b in the formation of PRC1 complexes, although this result may be affected by the lower efficiency of *Rybp* knock-down in comparison to *Rnf2* depletion.

Taken together, these results prove the role of vPRC1 complexes in 2C genes repression in ESCs, while exclude an analogous involvement of PRC2 complexes or the NuRD complex.

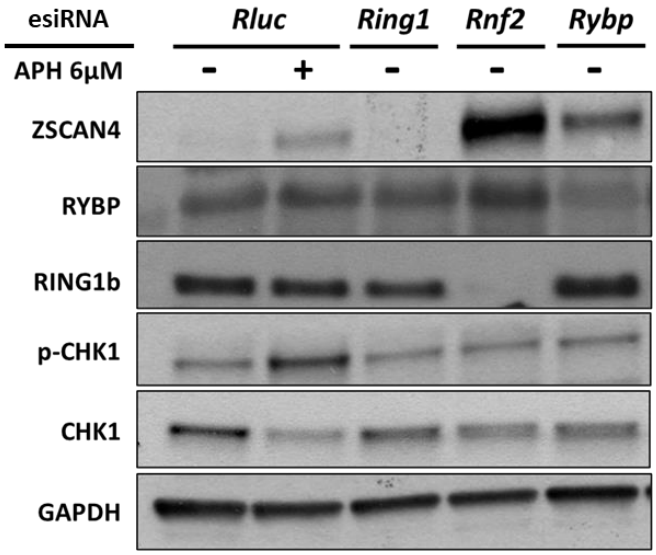

Figure 12 | PRC1 ablation induces 2-cells stage up-regulation without activation of CHK1: Western blot analysis of E14 cells transfected with the indicated esiRNA pools. Treatment with aphidicolin 6µM for 16 hours was performed when indicated. Antibodies against ZSCAN4, RYBP, RING1b, p-CHK1 (Ser317) and total CHK1 were used for the analysis. GAPDH was used as a loading control.

3.3 – RS induces massive reorganization of RING1b and changes in PRC2 marks

Given the importance of PRC1 complexes in the repression of *Dux* and its targets, and since Polycomb proteins have been proven to be involved in replication forks progression, we wondered whether the up-regulation of *Dux* and its targets that we observe upon RS induction may be due to a displacement of these complexes from the promoter of the gene during the processes involved in fork stalling and rescuing. As shown in figure 12, in whole cell protein extract we could not observe any difference in the protein expression of both RING1b and RYBP in response to RS, indicating that the RS response does not affect the expression of main vPRC1 components. Quite peculiarly, the PRC1 deposited histone mark, namely H2AK119Ub, was increased after RS induction by aphidicolin 6µM treatment for 16 hours in E14 cells, but this effect could not be reproduced in R1 cells, suggesting that this effect may be dependent on either cell lines background or cells status, and therefore it is not likely to play any major part in the up-regulation of *Dux* in response to RS. On the contrary, PRC2 deposited histone mark H3K27me3 showed a clear increase in response to RS in both cell lines tested (Fig. 13).

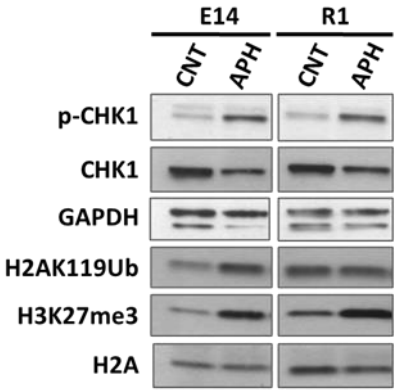

Figure 13 | RS increases H3K27me3 in ESCs: western blot analysis on E14 and R1 cells either in unperturbed (CNT) conditions or treated with aphidicolin 6µM for 16 hours (APH). CHK1 phosphorylation at Ser317 (p-CHK1), CHK1, H2AK119Ub and H3K27me3 levels were measured. H2A and GAPDH were used and loading controls.

This result suggests an increase in Polycomb activity in response to RS, or at least an increase in PRC2 activity, which, given the notorious repressive function of these

complexes, is somehow in contrast with the broad gene overexpression that we observed in response to RS in our previous work. To better characterize this phenomenon, we then investigated the behavior of PRC1 main component RING1b, which we showed to be the main responsible for *Dux* repression in control conditions, at the single cell level by immunofluorescence. We treated MC1 ESCs with aphidicolin 6μM for 16 hours and investigated the distribution of RING1b protein throughout the cell's nuclei by confocal microscopy. Induction of RS was confirmed by both an increase in the total γH2AX intensity per cell and the number of γH2AX foci per nucleus in response to the aphidicolin treatment (Fig. 14e, f). Notably, few γH2AX foci were already visible in control conditions, in agreement with previous work from us and other groups, demonstrating that ESCs are subjected to moderate levels of RS already in unperturbed conditions (170). Very interestingly, RING1b displayed a very peculiar distribution in ESCs nuclei under normal conditions, forming very tight and bright foci, which corresponded to regions of highly condensed chromatin, as showed by stronger DAPI signal (Fig. 14a). These foci could be analogous to the Polycomb associated domains (PADs) recently identified in mouse oocytes, and reported to be fundamental for the maintenance of their transcriptional identity (193). Interestingly, upon RS induction, RING1b foci became broader and less bright, suggesting a profound rearrangement in PRC1 disposition inside the nucleus (Fig. 14a). However, this rearrangement did not correspond to a major change in H2AK119Ub disposition into the nuclei, which maintained a diffused appearance in both control and aphidicolin treated cells (Fig. 14b). Worth of noting, H2AK119Ub signal was consistently excluded form γH2AX foci in both conditions, probably due to the inability of the antibody used to recognize the ubiquitination of H2AX isoform. Differently from PRC1 core component RING1b, EED, a fundamental subunit of PRC2 complexes core, did not show any clear localization in foci in control ESC but was, instead, distributed homogeneously inside

nuclei (Fig. 14c). This diffused distribution was not affected by RS (Fig. 14c). Interestingly, histone H3 trimethylation signal increased in response to RS at the single cell level (Fig. 14e) analogously to what observed by western blot in bulk protein extracts (Fig. 13).

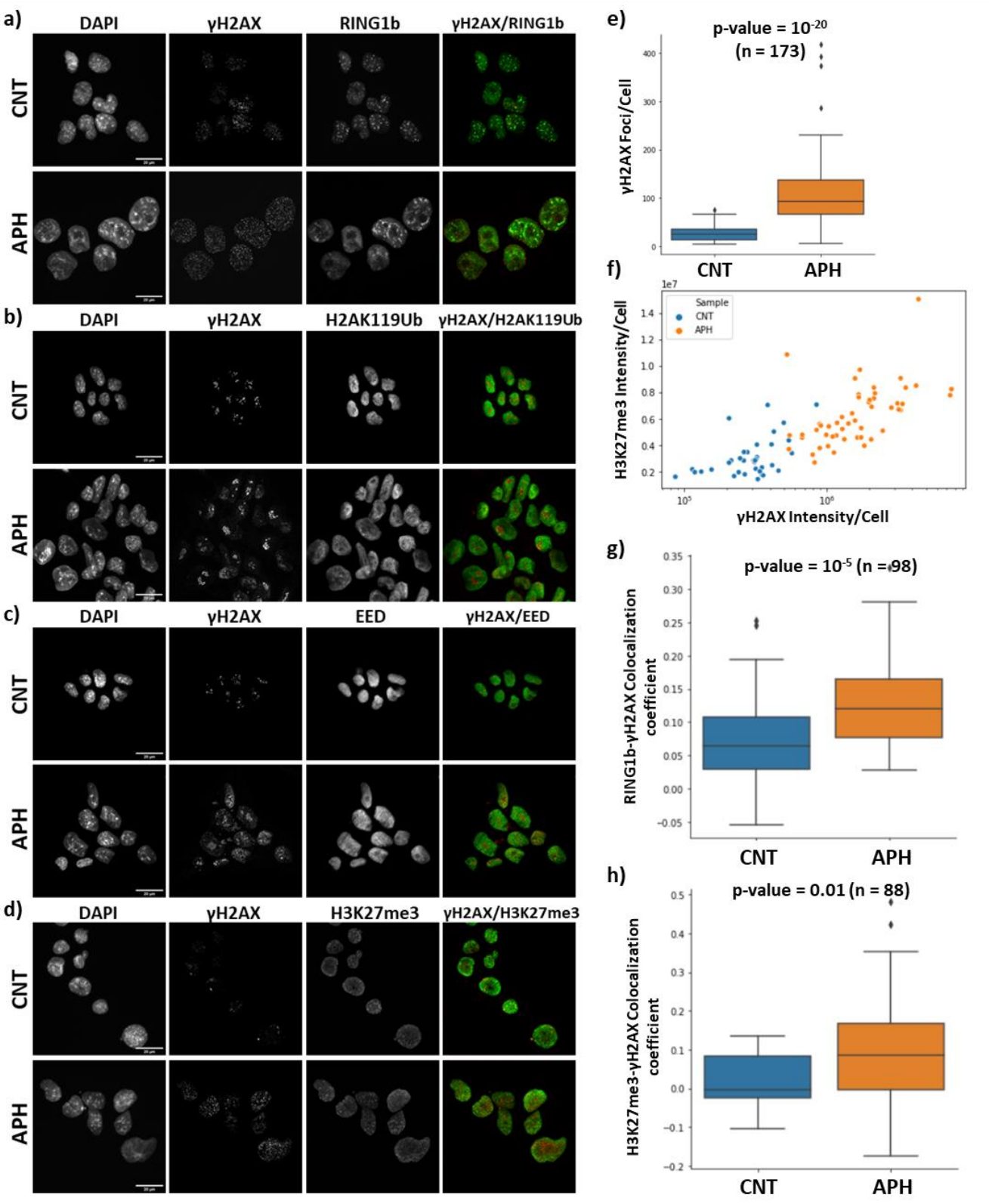

Figure 14 | RS causes a rearrangement of Polycomb proteins and their associated marks in the nucleus of ESCs: Representative pictures of unperturbed MC1 ESCs (CNT) or MC1 ESCs treated with aphidicolin 6μM for 16 hours (APH) showing the disposition of respectively **a)** γH2AX and RING1b, **b)** γH2AX and H2AK119Ub, **c)** γH2AX and EED and **d)** γH2AX and H3K27me3. **a-d)** pictures are the sum of several focal planes acquired by confocal microscopy. Nuclei are marked by DAPI staining. Scale bars correspond to 20 μm. **e)** Quantification of the number of γH2AX foci per cell nucleus counted by confocal microscopy in MC1 cells either in CNT or APH conditions. The two conditions were compared by unpaired two-tailed t-student test. **f)** Scatterplot representing the relation between the total level of H3K27me3 and γH2AX signal per cell either in CNT or APH conditions as measured by confocal microscopy (n = 88). **g-h)** Boxplot showing the colocalization coefficient for signal measured in confocal microscopy stacks for each cell acquired either in CNT or APH conditions for γH2AX with either **g)** RING1b or **h)** H3K27me3. The two distributions were compared using an unpaired two-tailed t-student test.

In addition, total H3K27me3 signal intensity also correlated with γH2AX signal intensity at the single cell level inside both control and RS groups (Fig. 14e), indicating that H3K27me3 amount increases together with the increase in H2AX phosphorylation. Notably, while in control conditions H3K27me3 appeared mostly at the nuclei borders, a region known for higher chromatin condensation, after RS induction this histone mark formed clearly visible foci throughout the nuclei (Fig. 14d). Interestingly, colocalization analysis of the fluorescence signal from RING1b and γH2AX showed a certain degree of colocalization, which was clearly increased by RS (Fig. 14g). The higher degree of colocalization between RING1b and γH2AX upon RS may be due to the more disperse nature of RING1b in cells subjected to RS and may indicate a recruitment of PRC1 complexes to sites of RS. The same analysis showed a similar increase in the median correlation coefficient between H3K27me3 signal and γH2AX at the single cell level (Fig. 14h). However, in this case we observed a wider spread of this value around the median, suggesting a variegated response to RS at the single cell level for what regards H3K27me3 localization. These results clearly indicate that upon RS the ESCs nuclei undergo a massive rearrangement of Polycomb complexes, with a concurrent modification in their correspondent histone marks. This result may explain at least in part the extent of gene up-regulation in response to RS, which

could be dependent on extensive loss of repressed chromatin, or rearrangement of active and repressed chromatin domains.

3.4 – RS affects PRC1 and PRC2 occupancy on target genes promoters

To better characterize the response of PRC1 and PRC2 complexes to RS at the genome-wide scale, we took advantage of a recently developed technique named Cut&Tag (181). This protocol is based on the ability of Tn5 bacterial transposases to insert short pieces of DNA inside the genome, causing at the same time the formation of a double strand break (DSB), and therefore efficiently fragmenting genomic DNA. Briefly, an antibody is used to bind the target protein of the analysis directly on the chromatin, without need of fixation (Fig. 15a). Secondary antibodies targeted at the constant region of the primary antibody are then used to increase the binding sites for an engineered protein constituted by the fusion of protein A and the Tn5 transposase. Protein A allows the fusion protein to bind to the antibodies on the target site and therefore brings in position the Tn5 transposase, loaded with two different DNA tags with short overhangs (Fig. 15a). Working as a dimer, the transposases insert these short sequences in the genome, leading to the formation of DNA DSBs at the sites of insertion. Various events of insertion lead to the formation of fragments of different sizes surrounding the original binding site of the protein of interest. These fragments are tagged at the extremities by the selected DNA tags. The two tags possess short overhangs which allow binding and amplification in a common PCR reaction using standard Illumina Nextera barcoded primers. Only fragments bound at the two extremities by the combination of the two different DNA tags are then amplified by a standard PCR reaction using barcoded sequencing primers (Fig. 15a). Through this process is possible to select and amplify sequences in close proximity to the original binding site of the target protein thus increasing the signal to noise ratio and therefore allowing analysis

from less abundant starting material in comparison to a traditional ChIP-Seq analysis. The process of PCR amplification has also the additional advantage of favoring the amplification of smaller fragments, ranging from about 100 to 300 bp, and thus granting a good resolution of peaks for the following analysis.

To investigate the effect of RS on Polycomb genomic disposition we performed Cut&Tag in E14 mouse ESCs either unperturbed or subjected to aphidicolin 6μM for 16 hours in order to induce RS. The experiment was performed in duplicate and normal rabbit IgG were used as negative control to subtract the signal due to the natural tendency of Tn5 transposases to react with sites of particularly open chromatin. We investigated genomic occupancy of both RING1b (for PRC1 complexes) and EED (for PRC2 complexes) and included the two histone marks H2AK119Ub and H3K27me3. Only peaks significantly enriched over the negative control in both replicates were kept has target specific peaks.

Our analysis showed narrow peaks of RING1b around the transcription starts sites (TSS) of 2972 genes in unperturbed ESCs (Fig. 15b). Upon RS, RING1b peaks showed substantially a similar appearance in terms of disposition around TSS of genes, however we could observe a clear decrease in RING1b signal surrounding genes promoters in response to RS if compared to control ESCs (Fig. 15b, f). As expected, the histone mark H2AK119Ub showed broader peaks in comparison to RING1b. Interestingly this mark presented a stronger signal downstream to TSS in comparison to what measured upstream of gene promoters (Fig. 15c). Quite surprisingly, in opposition to what we observed for RING1b, RS induction resulted in accumulation of H2AK119Ub signal around genes TSS (Fig. 15c, g). Even more interestingly, H2AK119ub accumulation did not lead to an increase in the peak intensity of the signal, which corresponded in both unperturbed and RS-treated cells with the TSS of genes, but led to broader peaks, extending both downstream in the gene body and upstream of the TSS (Fig. 15c, g). This result, apparently contradictory with our RING1b

results, could maybe be explained by the proven role of H2A ubiquitination in DNA damage and RS checkpoints activation (168,194).

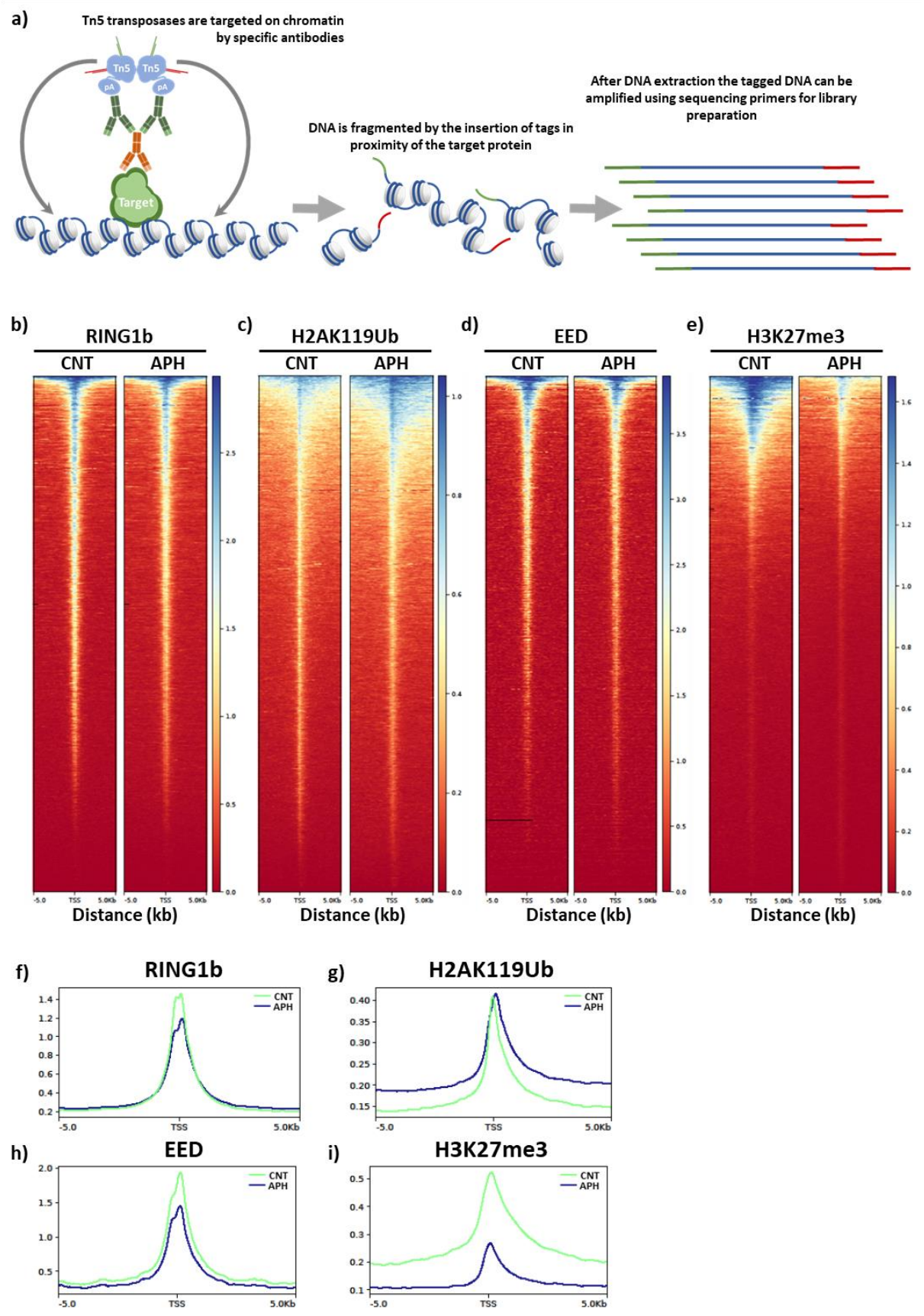

Figure 15 | RS affects the genome wide disposition of Polycomb complexes and their histone marks: a) Schematic representing the basic principle of the Cut&Tag assay. **b-e)** Heatmaps showing the binding signal detected in proximity of target genes TSS in both unperturbed (CNT) or aphidicolin treated (APH) conditions for antibodies targeting **b)** RING1b, **c)** H2AK119Ub, **d)** EED and **e)** H3K27me3. **f-i)** Density plots comparing the level of binding detected in correspondence to gene TSS genome wide for antibodies targeting **f)** RING1b, **g)** H2AK119Ub, **h)** EED and **i)** H3K27me3.

In fact, the observed accumulation of H2AK119Ub around genes TSS upon RS could be due to a general increase in H2A ubiquitination all over the genome due to the activation of the RS response. Analogous analysis of EED signal showed, instead, a depletion of this PRC2 core component from target TSS upon RS (Fig. 15d, h) in a similar fashion to what we observed in the case of RING1b. However, differently from H2AK119Ub, the PRC2 associated histone mark H3K27me3 displayed a dramatic decrease around TSS upon RS (Fig. 15e, i), which is in agreement with the massive gene up-regulation reported in our previous work (42). A more in-depth analysis showed that, of 2972 genes bound by significant peaks of RING1b in control ESCs, 1600 lost RING1b binding upon RS induction, while a core of 1372 gene promoters maintained the associated RING1b peaks even after aphidicolin treatment (Fig. 16a). Quite interestingly, in the case of the histone mark H2AK119Ub, RS resulted in a massive loss of bound domains in proximity of gene promoters, with 4645 genes losing their H2AK119Ub associated signal (Fig. 16b). At the same time, while only 126 genes acquired RING1b peaks upon RS (Fig. 16a), as many as 711 genes showed newly formed H2AK119Ub domains in response to RS induction (Fig. 16b). The massive loss of H2AK119Ub in proximity to gene promoters is, somehow, in contrast to our previous observation that the overall signal associated to H2AK119Ub on TSS genome wide increases in strength upon RS (Fig. 16c, g). However, it is important to notice that, while H2AK119Ub signal extends further from TSS upon RS, its maximum signal in proximity of the TSS does not show on average an analogous increase. This result may be the effect of broad accumulation of low level signal throughout the genome, forming

extended domains in proximity of genes TSS (Fig. 15c) which may impair the ability of the used algorithms to identify this low intensity extended domains as significantly enriched

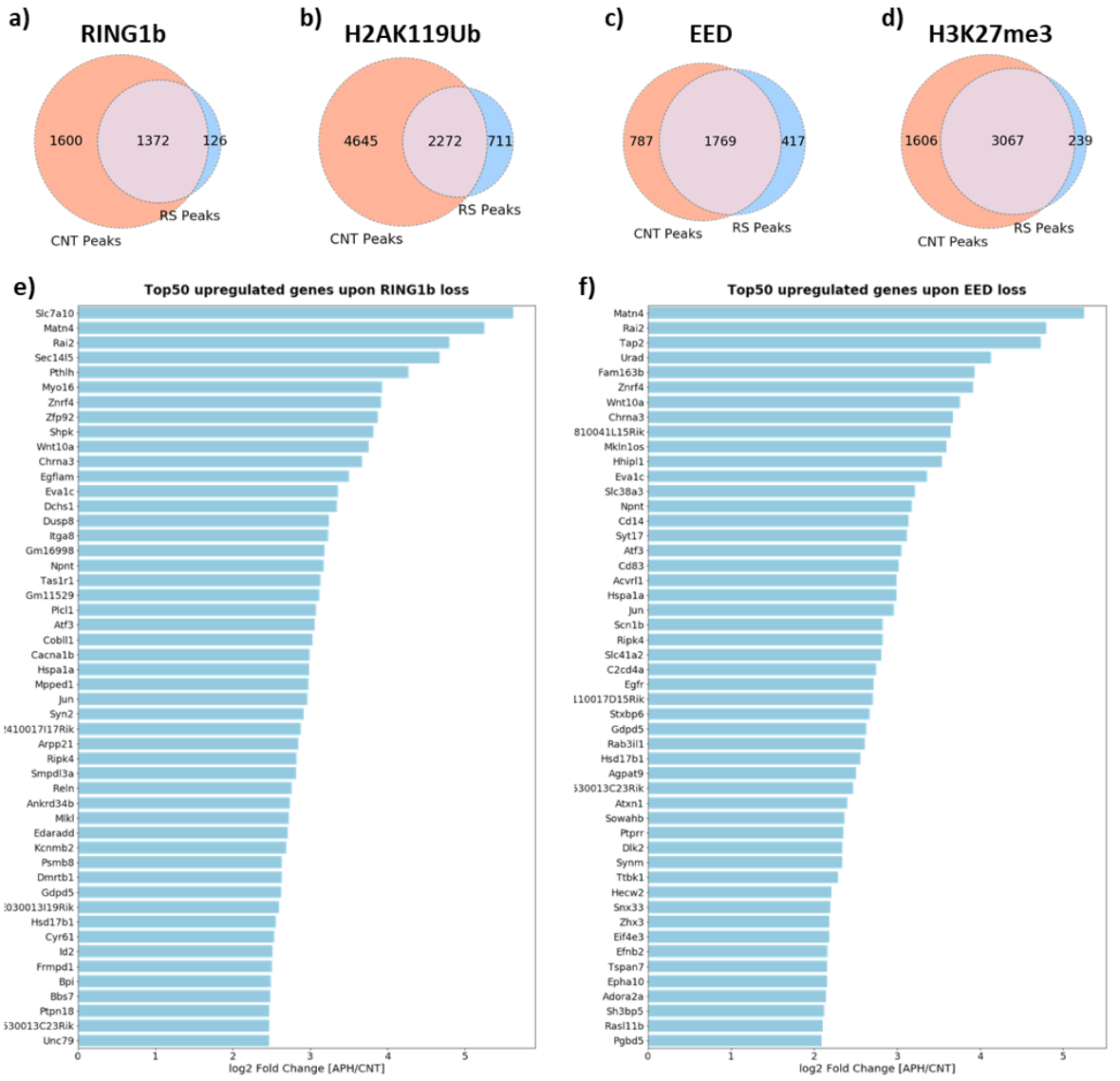

Figure 16 | Loss of Polycomb complexes affects genes involved in several pathways: a) Venn diagram showing the overlap between genes whose promoters present peaks identified by Cut&Tag analysis in unperturbed (CNT) or aphidicolin treated (APH) conditions for either **a)** RING1b, **b)** H2AK119Ub, **c)** EED or **d)** H3K27me3. **e-f)** Barplot showing, among genes which lose either **e)** RING1b or **f)** EED binding in our Cut&Tag analysis, the top 50 most overexpressed genes according to our previous RNA Seq analysis. **g)** Enrichr analysis for the GO term biological function for genes which lose peaks for RING1b, H2AK119Ub, EED or H3K27me3. Terms are ordered by ranked p-value.

over the background noise. Interestingly, of the 2556 genes bound by EED in unperturbed ESCs, 787 genes lost this signal upon RS induction while 417 gene promoters displayed newly formed peaks after treatment with aphidicolin (Fig. 16c). Differently from what we observed in the case of RING1b, for which the loss of gene associated peaks greatly outnumbered the newly formed ones, suggesting a general loss of PRC1 repression, in the case of EED, the number of exclusively RS associated gene peaks was closer to the number of peaks observed only in unperturbed cells, indicating a reshaping of the PRC2 repressive landscape in response to RS, with many genes losing repression and many others acquiring PRC2 *de novo*. At the same time, as many as 1606 genes lost H3K27me3 domains in proximity to their promoters, while 3067 maintained significantly enriched peaks (Fig. 16d), although the intensity of the measured signal for each peak was greatly reduced upon aphidicolin treatment (Fig. 15e). In their complex, these data show a general reduction of Polycomb associated repression throughout the genome, although at the same time they highlight the presence of striking differences between the behavior of PRC complexes and the presence on the chromatin of their associated histone marks.

To better understand the effects of Polycomb loss on the expression level of their targets, we compared the lists of genes which lost either RING1b or EED presence in response to RS as they result from our Cut&Tag data with the expression levels of genes up-regulated by RS according to our previous RNA Seq analysis (42). Among the genes which lost RING1b peaks upon RS, we observed 238 genes whose expression was up-regulated by RS in our previous RNA Seq analysis, corresponding to almost 15% of this group. At the same time,

among genes which lost EED signal, we counted 138 genes which increased their expression upon RS, accounting for more than 17% of these genes.

In evaluating these results, it is important to remember that our RNA Seq data consider the variability among three different ESC lines and, therefore, are reasonably more restrictive than our Cut&Tag analysis performed on E14 ESCs. Therefore, it is quite interesting to observe that a relevant proportion of genes, which show loss of PRC associated repression in E14 cells, display a significant increase in expression level, which is reproducible even in different cell lines. In figures 16e and 16f, the extent of the up-regulation for the most representative of these genes is shown. Interestingly, some of these genes, such as the extracellular matrix gene *Matn4*, the retinoic acid induced *Rai2* or the proto-oncogene *c-Jun* displayed both loss of RING1b and EED associated peaks, suggesting a loss of both PRC1 and PRC2 repression, which was in agreement with their up-regulation in response to RS. Quite interestingly, among these genes we could find actors involved in many different processes. For example, *Matn4* is involved in limb development and in particular in cartilage formation (195) while *Rai2*, a gene expressed during early brain formation, has been shown to have a role in chromosomal stability maintenance, both in early breast cancer development and metastasis formation (196,197). Moreover, also the transcription factor *Atf3*, involved in both MAPK and c-JUN pathways, and controversially proposed as both oncogene and tumor suppressor, was among the genes which showed loss of both Polycomb proteins occupancy (Fig. 16e, f). These evidences suggested that loss of Polycomb repression in response to RS may affect several pathways through the deregulation of various transcription factors and developmental genes, such as *Atf3*, *Id2* or *Wnt10a* (Fig. 16e, f). To better characterize this possibility, we decided to perform a gene ontology enrichment analysis for the term biological function on the lists of genes which lost peaks of RING1b, EED or their associated histone marks H2AK119Ub and H3K27me3

upon RS induction. As expected, many of the identified terms were involved in various developmental processes. In particular, among genes unbound by RING1b upon RS we observed mainly terms referring to nervous system development, such as voltage-gated channels regulation, axon development and axon guidance (Fig. 16g). Similarly, we identified the term nervous system development also in genes which lost H2AK119Ub domains in response to RS, although in this case the most common and high scoring terms were linked to the development of myeloid dendritic cells (Fig. 16g), a class of antigen presenting cells being part of the innate immune system. However, as previously mentioned, a certain degree of caution has to be applied when evaluating the genes which lost H2AK119Ub domains upon RS, due to the general increase in the breadth of H2AK119Ub domains upon RS, which may affect our algorithms to correctly separate peak signals from background noise. Also in the case of genes which lost EED occupancy in response to RS we could identify terms linked to neural development, although also terms referring to epithelial development were present (Fig. 16g). Interestingly, also the canonical Wnt pathway scored significantly among these genes (Fig. 16g), suggesting that RS may affect this fundamental pathway, which is involved in many developmental but also pathological processes, including cancer (198). Finally, among genes which lost H3K27me3 repressive mark upon RS, we observed a broad range of GO terms, mainly referring to developmental processes, ranging from terms linked to cardiac development to axonogenesis and angiogenesis, but also, quite surprisingly, terms related to the maintenance of an undifferentiated state (Fig. 16g).

Since the Enrichr tool allows for easy recovery of the genes associated to each identified GO term, we investigated the status of Polycomb occupancy close to the promoter of each gene associated to some of the most relevant GO terms identified. For each of the selected terms and for each of the associated genes we investigated if our analysis showed the loss

or gain of RING1b, EED peaks or their associated marks H2AK119Ub and H3K27me3, or if no change was observed. Through this analysis for the term “Nervous system development”, which was identified from genes that lost RING1b peaks upon RS, we observed that almost half of these genes showed concomitant loss of the associated mark H2AK119Ub (Fig. 17a). Moreover, 13 genes showed concomitant loss of EED signal in response to RS and 15 displayed a loss in H3K27me3 domains, indicating that genes involved in the neural developmental pathway are subjected to a broad loss of Polycomb repression in response to RS that involves not only PRC1 but also PRC2 (Fig. 17a). To better visualize this effect, we submitted the list of genes associated to the nervous system development term to the STRING database, looking for a potential protein interaction network among the products of these genes. This investigation allowed us to find a strongly interconnected network centered around *Shh*, a gene involved in many steps of early embryo patterning, and the neural developmental factors *Bdnf* and *Nes* (Fig. 17b), proving that RS leads to a loss of Polycomb repressing signal on a functional cluster of neural developmental genes. The extent of this de-repression was particularly clear when the disposition of Polycomb signal around the promoters of some of the selected genes was visualized, as shown in figures 17c and 17d. An analogous investigation on genes linked to the term “Negative regulation of cell differentiation” showed that the loss of H3K27me3 on these genes was generally accompanied by similar loss of H2AK119Ub, and in some cases concomitant loss of either RING1b or EED (Fig. 17e), highlighting how, also in this case, the loss of one Polycomb associated mark was often paired with the loss of others.

Also in this case, among the gene products associated with this GO term, was possible to identify an interconnected group of proteins whose promoters underwent loss of Polycomb marks upon RS (Fig. 17f-h). Finally, is worth of mentioning the presence, among the GO terms identified for loss of EED signal, of the “Canonical Wnt signaling pathway”, a

major player in both stem cells biology and early embryonic development, represented by a small by highly interconnected group of genes, which in many cases showed, in addition to EED loss, also depletion of RING1b, H2AK119Ub or H3K27me3 on their promoters (Fig. 17i-k).

Taken together, these data suggest that high RS leads to a global rearrangement of the disposition pattern of Polycomb complexes on the genome of mouse ESCs, affecting in various extents different pathways, and therefore having the potential to majorly affect the cells ability to maintain their correct chromatin landscape.

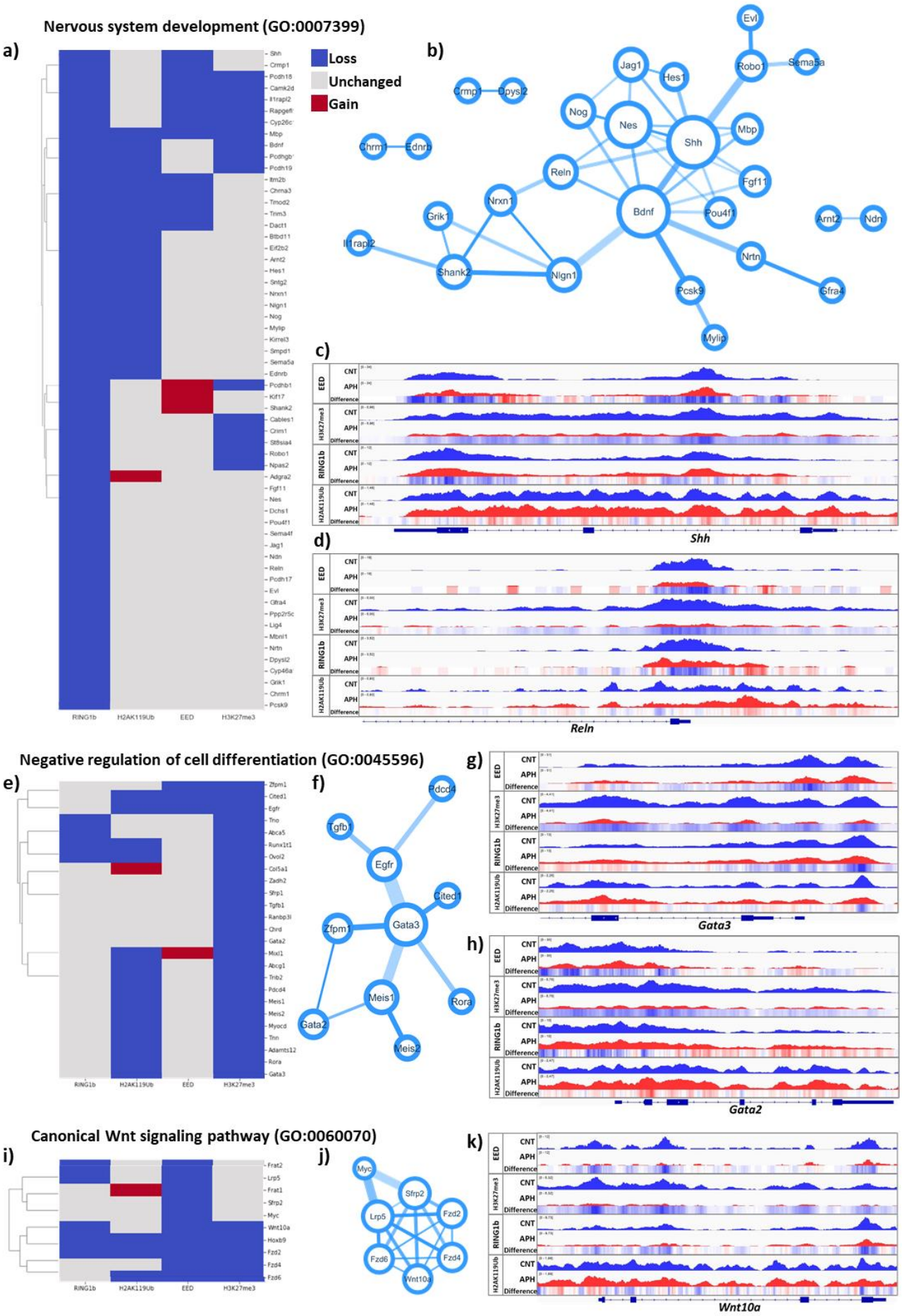
Nervous system development (GO:0007399)
a)
Loss
Unchanged
Gain
RING1b
H2AK119Ub
EED
H3K27me3
b)
c)
d)
CNT
APH
Difference
Shh
Reln
Negative regulation of cell differentiation (GO:0045596)
e)
f)
g)
h)
Gata3
Gata2
Canonical Wnt signaling pathway (GO:0060070)
i)
j)
k)
Wnt10a

Figure 17 | Multiple functional networks lose Polycomb repression upon RS: a) Heatmap showing which among the genes linked to the term «Nervous system development» lose, gain or show no change in peaks for each of the protein tested by Cut&Tag. **b)** STRING network of protein interactions among the genes belonging to the «Nervous system development» term. **c-d)** Reads alignment for each of the tested proteins by Cut&Tag on either the **c)** *Shh* or **d)** *Reln* loci for both unperturbed (CNT, blue) conditions and aphidicolin treated cells (APH, red). **e)** Heatmap showing which among the genes linked to the term «Negative regulation of cell differentiation» lose, gain or show no change in peaks for each of the protein tested by Cut&Tag. **f)** STRING network of protein interactions among the genes belonging to the «Negative regulation of cell differentiation» term. **g-h)** Reads alignment for each of the tested proteins by Cut&Tag on either the **g)** *Gata3* or **h)** *Gata2* loci for both CNT and APH conditions. **i)** Heatmap showing which among the genes linked to the term «Canonical Wnt signaling pathway» lose, gain or show no change in peaks for each of the protein tested by Cut&Tag. **j)** STRING network of protein interactions among the genes belonging to the «Canonical Wnt signaling pathway» term. **k)** Reads alignment for each of the tested proteins by Cut&Tag on the *Wnt10a* locus for both CNT and APH conditions. In all interaction networks only proteins showing at least one link are shown. The dimension of each node is proportional to the number of interactors. Edges dimension is proportional to edge betweeness while edge intensity is proportional to the combined score for each interaction as reported by the STRING database. In plots c-d-g-h-k difference between APH and CNT conditions is represented as a heatmap (blue = higher binding in CNT, red = higher binding in APH).

3.5 – RS causes an increase in variant PRC1 occupancy on *Dux* promoter

To answer our question regarding the behavior of PRC1 complexes upon RS, and their relation with the activation of *Dux* and its target genes, we used our Cut&Tag data to investigate the presence of Polycomb marks in proximity to the *Dux* locus either in unperturbed conditions or after RS induction. Our analysis highlighted the presence of PRC1 marks, RING1b and H2AK119Ub on the promoter of the *Dux* gene in unperturbed ESCs (Fig. 18). This observation is consistent with data already present in the literature, showing RING1b and RYBP presence on the *Dux* promoter in mouse ESCs (192) and our data (shown in the previous paragraphs) linking PRC1 to the repression of *Dux* in normal mouse ESCs. Surprisingly, notwithstanding our observation that PRC2 main components SUZ12 and EZH2 are not involved in *Dux* repression in ESCs, we could observe a clear presence of PRC2 structural component EED on *Dux* promoter, accompanied by a strong deposition of PRC2 histone mark H3K27me3 (Fig. 18). Interestingly, not only RS could not

reduce the signal of EED in proximity of *Dux* promoter, but it resulted in a mild accumulation of this PRC2 main component (Fig. 18). Surprisingly, in complete opposition to what observed for EED, we noticed a stark decrease in H3K27me3 signal in response to RS over the entire *Dux* locus (Fig. 18). Even more surprisingly, both RING1b and its associated histone mark H2AK119Ub were clearly accumulated on the *Dux* locus in response to RS (Fig. 18). Given the role of PRC1 complexes in repressing *Dux* and its target 2C genes, this peculiar behavior is in complete opposition to what we could expect.

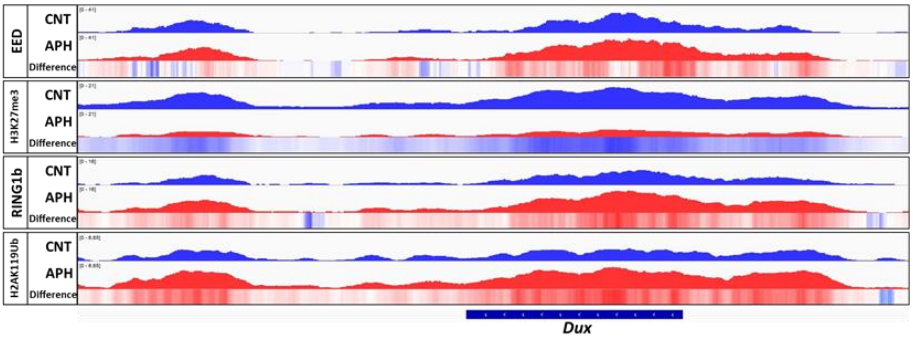

Figure 18 | Cut&Tag reveals PRC1 accumulation on *Dux* locus upon RS: Reads alignment for each of the tested proteins by Cut&Tag on the *Dux* locus for both unperturbed (CNT, blue) conditions and aphidicolin treated cells (APH, red). For each protein the difference between signal measured in APH compared to CNT is given as an heatmap (negative values = blue, positive values = red).

To confirm PRC1 accumulation on *Dux* upon RS, we performed chromatin immunoprecipitation (ChIP) coupled with real time qPCR on both E14 and R1 ESCs either cultured in unperturbed conditions or after inducing high RS by aphidicolin treatment for 16 hours at 6µM. After fixation and chromatin extraction, the samples were immunoprecipitated with antibodies against RING1b, RYBP or their associated histone mark H2AK119Ub. To confirm the H3K27me3 decrease observed in Cut&Tag, an antibody against this histone modification was also added to the panel.

As a negative control for the immunoprecipitation, part of each sample was incubated with normal rabbit IgGs. Alongside *Dux* promoter, we also measured the percentage of enrichment over input for each of the tested proteins also for *Cdx2* promoter, as a positive control for Polycomb repression. The promoter of *Pou5f1*, which is instead highly active in

unperturbed ESCs, and a random intergenic sequence on chromosome 14 have been used as a control for background signal. As expected, chromatin immunoprecipitated with an antibody against RING1b showed poor enrichment of both *Pou5f1* promoter and our intergenic negative control for both cell lines tested (Fig. 19a, b). This level was not sensibly increased upon RS induction and remained in the range measured for samples incubated with IgGs (Fig. 19a, b, i, l). On the contrary, *Cdx2* promoter displayed strong RING1b signal, both in control conditions and after RS induction. Interestingly, R1 cells showed weaker signal in comparison to E14 ESCs, although they followed an analogous pattern (Fig. 19a, b). Of note, *Dux* promoter displayed less RING1b signal in comparison to *Cdx2* promoter (Fig. 19a, b). However, both E14 and R1 ESCs showed increased RING1b signal on *Dux* promoter in response to RS, confirming our Cut&Tag results (Fig. 19a, b). This data was further strengthened by RYBP ChIP-qPCR. RYBP is the other principal actor involved in *Dux* repression in ESCs according to our data, and is a main component of vPRC1 complexes. Also in this case, immunoprecipitation led to a low enrichment for the intergenic negative control region and for the promoter of the active gene *Pou5f1*, in the range of IgG immunoprecipitated samples for both cell lines (Fig. 19c, d, i, l). RYBP levels on these regions was not affected by RS in any of the lines tested (Fig. 19c, d). Instead, the extent of enrichment was higher in the case of the Polycomb repressed gene *Cdx2*, and this level of enrichment appeared unchanged in both conditions tested (Fig. 19c, d). RYBP immunoprecipitation led to a slightly superior enrichment for *Dux* in comparison to the negative control region or the *Pou5f1* promoter in unperturbed conditions, but still lower than the level displayed by the *Cdx2* promoter (Fig. 19c, d). Interestingly *Dux* levels increased to the range of *Cdx2* in response to RS in both E14 and R1, indicating a clear accumulation of variant Polycomb complexes on the locus (Fig. 19c, d). Differently from what we observed in Cut&Tag, the deposition of H2A ubiquitination on lysine 119 was not

markedly affected by RS in any of the cell lines analyzed (Fig. 19e, f). In this case is important to notice how, while our ChIP-qPCR analysis measured a lower level of enrichment for both the intergenic negative region and *Pou5f1* promoter in comparison to *Dux* and *Cdx2* (Fig. 19e, f), the extent of this difference was much inferior to what we observed in the case of RING1b and RYBP immunoprecipitation, indicating how H2AK119Ub could be a not optimal marker to discriminate among PRC1 repressed loci in ESCs. It is important to remember though, that in all the tested loci H2AK119Ub was clearly higher than the IgG background (Fig. 19e, f, i, l), showing that the presence of this histone mark on intergenic sequences or

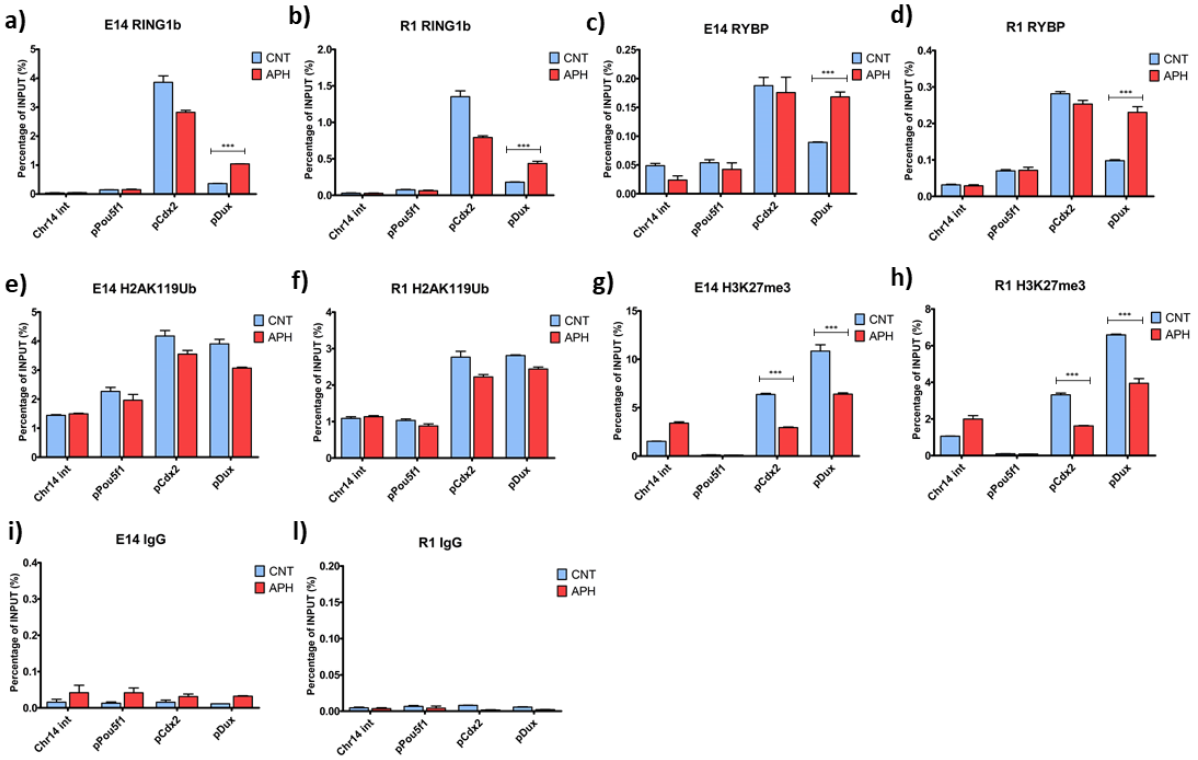

Figure 19 | RS leads to an accumulation of variant PRC1 components on *Dux* locus: a-b) ChIP-qPCR analysis showing the percentage of enrichment over input for the indicated loci immunoprecipitated (IP) with RING1b antibody in unperturbed (CNT) or aphidicolin treated (APH) conditions for either **a)** E14 or **b)** R1 cells. **c-d)** ChIP-qPCR analysis showing the percentage of enrichment over input for the indicated loci immunoprecipitated (IP) with RYBP antibody in CNT or APH conditions for either **c)** E14 or **d)** R1 cells. **e-f)** ChIP-qPCR analysis showing the percentage of enrichment over input for the indicated loci immunoprecipitated (IP) with H2AK119Ub antibody in CNT or APH conditions for either **e)** E14 or **f)** R1 cells. **g-h)** ChIP-qPCR analysis showing the percentage of enrichment over input for the indicated loci immunoprecipitated (IP) with H3K27me3 antibody in CNT or APH conditions for either **g)** E14 or **h)** R1 cells. **i-l)** ChIP-qPCR analysis showing the percentage of enrichment over input for the indicated loci immunoprecipitated with control IgGs in CNT or APH for either **i)** E14 or **l)** R1 cells. All data are shown as mean + SEM of a triplicate. CNT and APH were compared using unpaired two-tailed t-student test (* = p-val < 0.05, ** = p-val < 0.01, *** = p-val < 0.001).

active genes is not likely due to a technical issue. Interestingly, our ChIP-qPCR data confirmed the decrease in H3K27me3 on the *Dux* locus in response to RS in both E14 and R1 cells (Fig. 19g, h). In particular, the *Dux* locus displayed higher enrichment in comparison to the *Cdx2* promoter in unperturbed conditions and this level was reduced for both loci in response to RS (Fig. 19g, h). Quite surprisingly, these results together show a complex repression system on the master regulator of the 2C stage in response to RS.

Increased expression of *Dux* in response to RS in fact corresponds to a concurrent increase in the presence of PRC1 components on the locus, proving that there is not a linear correlation between Polycomb presence on the promoter and repression level. Moreover, the loss of H3K27me3 signal on *Dux* locus upon RS may be related to the dilution of marked nucleosomes with newly synthesized ones which occurs during replication (199,200). Arise of RS may affect the redeposition of the H3K27me3 mark on the locus following replication, leading to a decrease methylation level of histones. Notably, H3K27me3 dilution was not mirrored by a similar EED reduction as shown by our Cut&Tag data, suggesting that binding of PRC2 to its targets is not sufficient for efficient redeposition of H3 trimethylation.

Since, in our previous work, we proved that 2C genes activation in response to RS is strongly dependent on the activity of the ATR-CHK1 axis, we wondered whether the peculiar behavior of PRC1 at *Dux* locus in response to RS may be dependent on ATR activity as well. In order to answer this question, we repeated our ChIP-qPCR analysis on E14 and R1 cells either in unperturbed conditions, or treated with aphidicolin 6µM for 16 hours to induce RS, or treated with both aphidicolin 6µM and a specific ATR inhibitor (VE-822) for 16 hours, in order to observe the effect of RS in the absence of ATR activity. Interestingly, in both cell lines, although with a different extent, the increase in RING1b occupancy of *Dux* promoter upon RS was reverted by ATR inhibition (Fig. 20a, b), indicating that this increase is not simply due to collateral effects of RS, but requires the activity of the ATR checkpoint. The

same behavior was observed in the case of the other characteristic component of vPRC1 complexes, RYBP. Also in this case, in fact, the increase of RYBP signal on *Dux* promoter in response to RS induction was suppressed by concurrent ATR inhibition (Fig. 20c, d), highlighting a concerted behavior of both PRC1 components. Worth of mention, in the two cell lines tested, RING1b and RYBP did not show any conserved pattern of deposition on *Cdx2* promoter in response to ATR inhibition, proving that ATR inhibition does not result in a generalized reduction of PRC1 presence on its target loci (Fig. 20a-d).

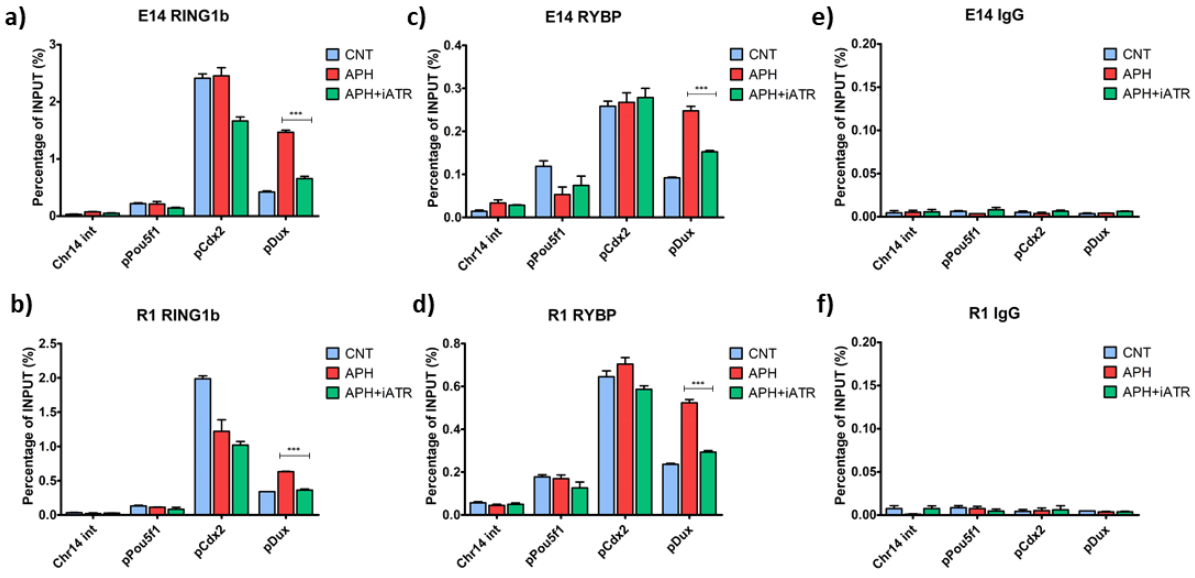

Figure 20 | PRC1 accumulation on *Dux* in response to RS requires ATR activity: a-b) ChIP-qPCR analysis showing the percentage of enrichment over input for the indicated loci immunoprecipitated (IP) with RING1b antibody in unperturbed conditions (CNT), after treatment with aphidicolin 6µM for 16 hours (APH) or after aphidicolin treatment in the presence of an ATR inhibitor (APH+iATR) for either **a)** E14 or **b)** R1 cells. **c-d)** ChIP-qPCR analysis showing the percentage of enrichment over input for the indicated loci in RYBP in CNT, APH or APH+iATR for either **c)** E14 or **d)** R1 cells. **e-f)** ChIP-qPCR analysis showing the percentage of enrichment over input for the indicated loci immunoprecipitated with control IgGs in CNT, APH or APH+iATR for **e)** E14 or **f)** R1 cells. All data are shown as mean + SEM of a triplicate. CNT, APH and APH+iATR conditions were compared using one-way ANOVA (* = p-val < 0.05, ** = p-val < 0.01, *** = p-val < 0.001).

These results show a very peculiar behavior of PRC1 on *Dux* promoter in response to RS. The increase in Polycomb presence on the promoter does not, in fact, result in increased repression, and does not affect the general level of H2AK119Ub deposition. Very interestingly, PRC1 presence on the promoter seems to be controlled by ATR activity and

provides further proof for a model in which PRC1 components may be involved in RS response.

3.6 - PRC2 but not PRC1 is required for efficient activation of 2C genes in response to RS

As we have shown before, RS leads to an increase in the level of H3K27me3 in ESCs nuclei, both in bulk and at the single cell levels. However, this increase was accompanied by a reduction in the number of gene promoters which showed H3K27me3 peaks in our Cut&Tag experiment, as exemplified by *Dux*. We wondered then if the reduction in H3K27me3 may be responsible for the loss of *Dux* repression in response to RS. To understand if *Dux* expression may be dependent on active erasure of histone H3 trimethylation we focused on the role of the two main histone demethylases, responsible for the removal of H3K27me3 from the chromatin, namely KDM6a and KDM6b (Lysine demethylase 6a/b) (201). These enzymes, and in particular KDM6a, have been shown to be necessary for differentiation of ESCs and their loss leads to severe defects in mouse embryonic development (202). To investigate if the active removal of H3K27me3 by these two demethylases could be involved in *Dux* overexpression upon RS induction we knocked-down *Kdm6b* using an esiRNA pool, while *Kdm6a* was ablated through the administration of a single siRNA. All experiments were performed in E14 cells. To better investigate the possibility that one demethylase may compensate for the absence of the other we performed also a knock-down on both demethylases at the same time. As a negative control cells were transfected with an esiRNA pool targeting the *Rluc* gene. ESCs were kept either in unperturbed conditions or subjected to RS induction through a 16 hours long treatment with aphidicolin at 6µM. The expression level of both *Dux* and its main target gene *Zscan4* mRNAs were measured by RT-real time qPRC 48 hours after transfection for all conditions. As shown in figure 21a, neither *Kdm6a* or *Kdm6b* knock-down had any effect

on *Dux* expression in unperturbed conditions, and similarly, no effect was observed in the case of *Zscan4* (Fig. 21b). Analogously, not even the concomitant ablation of both demethylases had any effect on *Dux* or *Zscan4* mRNA levels (Fig. 21a, b). Upon RS induction, *Kdm6a* knock-down not only did not impair *Dux* up-regulation, but resulted in a slight increase in respect to cells transfected with the *Rluc* esiRNA, although this increase was not statistically significant (Fig. 21a). The effect of *Kdm6a* ablation was even less relevant in the case of *Zscan4* expression, where we did not observe any difference in response to RS for cells transfected with the control esiRNA or *Kdm6a* siRNA (Fig. 21b). An even milder effect was observed in the case of *Kdm6b* depletion, which did not lead to a change of *Dux* expression in response to RS in comparison to cells transfected with the *Rluc*-targeting esiRNA (Fig. 21a). *Kdm6b* depletion corresponded to a small and not significant drop in *Zscan4* overexpression in response to RS (Fig. 21b).

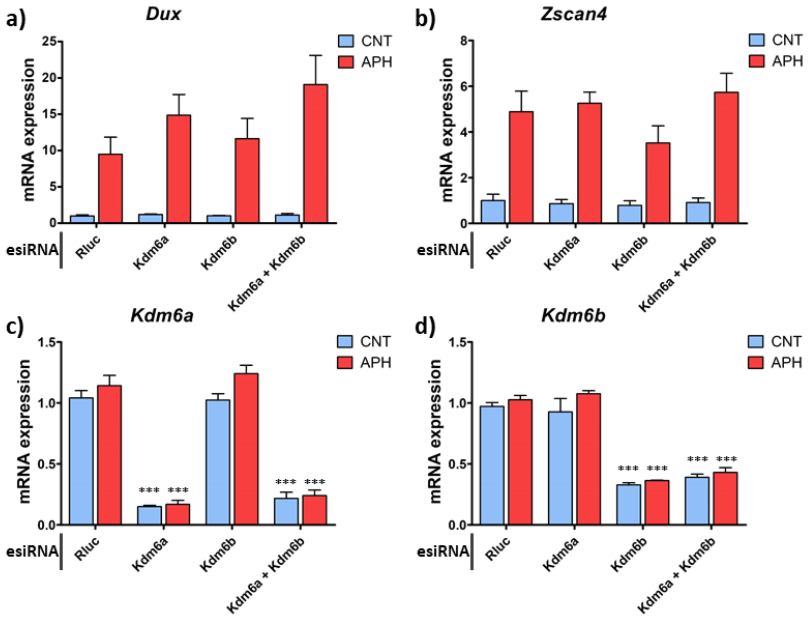

Figure 21 | Active demethylation does not influence *Dux* up-regulation upon RS: RT-real time qPCR analysis of mRNA levels in E14 cells either in unperturbed conditions (CNT) or after aphidicolin 6µM treatment for 16 hours (APH) and transfected with the indicated esiRNAs for **a)** *Dux*, **b)** *Zscan4*, **c)** *Kdm6a* and **d)** *Kdm6b* genes. All data are normalized on *Tbp* expression and shown as fold change in respect to the CNT *Rluc* sample. All data are shown as mean + SEM of three independent experiments. Comparison among all conditions was performed by one-way ANOVA with Tukey post-test (* = p-val < 0.05, ** = p-val < 0.01, *** = p-val < 0.001).

Surprisingly, ablation of both demethylases did not result in a reduced ability of ESCs to upregulate *Dux* in response to RS, but it instead caused a higher expression, although not significant, of the gene in comparison with cells transfected with the control esiRNA (Fig. 21a). This increase however did not correspond to an equal increase in *Zscan4* levels and therefore it most likely did not result in a general activation of 2C genes (Fig. 21b). To confirm that the demethylases knock-down inability to revert *Dux* up-regulation was not due to inefficient knock-down, or a change in *Kdm6a* or *Kdm6b* expression upon RS, we measured the expression of both genes in all the tested conditions. Both *Kdm6a* and *Kdm6b* proved to be efficiently knocked-down, as shown in figures 21c and 21d. Worth of mentioning, RS did not affect the expression of any of the demethylases, resulting in a comparable level of knock-down between control and treated conditions (Fig. 21c, d). Additionally, ablation of any of the two genes did not interfere with the expression of the other, and combined knock-down of both resulted in similar depletion in comparison to the single knock-down configuration (Fig. 21c, d). These results in their complex demonstrate that the activation of *Dux* in response to RS is not dependent upon active demethylation of the locus, although our data does not allow us to exclude that the reduction in H3K27me3 at *Dux* locus observed upon RS induction may be instrumental for its de-repression.

To better characterize the role of PRC2 in response to RS, we decided to perform RS induction by aphidicolin treatment in E14 ESCs, in which the two PRC2 main components SUZ12 and EZH2 were ablated by esiRNA-based knock-down. Even in this case cells were treated with aphidicolin 6μM for 16 hours in order to induce RS, and samples were harvested 48 hours post transfection for RNA extraction and RT-real time qPCR analysis. ESCs transfected with an esiRNA targeting *Rluc* were used as negative control. As already mentioned, both *Suz12* and *Ezh2* ablation caused a mild reduction in *Dux* expression under

control conditions, which was more pronounced in the case of its target genes *Zscan4* and *Tcstv3* (Fig. 22a, b, d). Interestingly, RS could not increase the expression of *Dux* in cells transfected with either *Suz12* or *Ezh2*-targeting esiRNAs to the level observed in cells transfected with the control esiRNA (Fig. 22a). An analogous effect was observed in the case of the other 2C stage markers *Zscan4*, *MERVL* and *Tcstv3* (Fig. 22b-d), indicating that indeed the depletion of PRC2 components strongly impaired the up-regulation of 2C stage genes in response to RS.

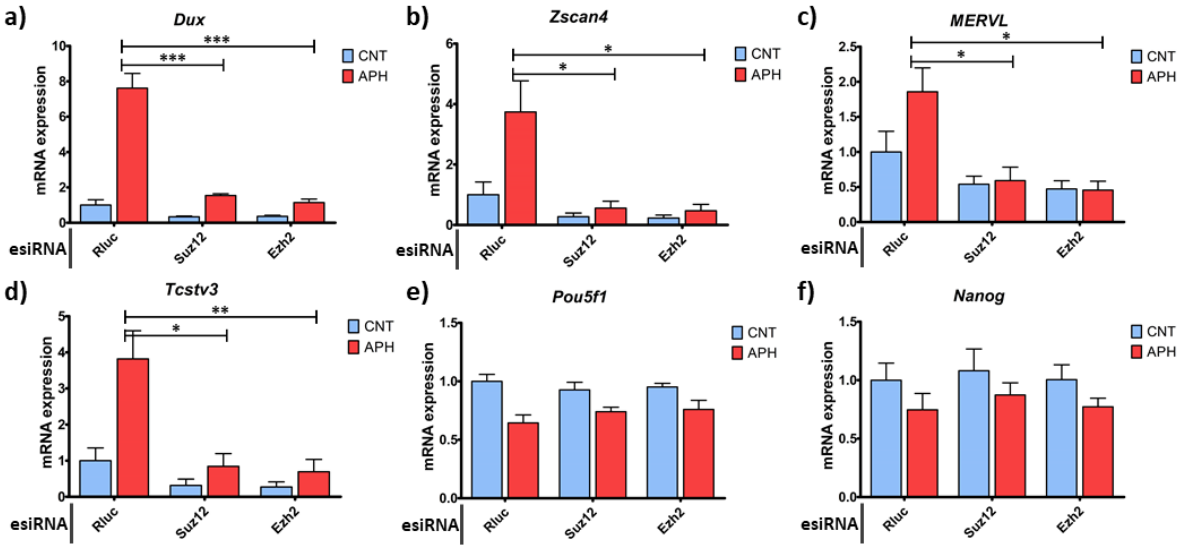

Figure 22 | PRC2 is required for *Dux* up-regulation upon RS: RT-real time qPCR analysis of mRNA levels in E14 cells either in unperturbed ESCs (CNT) after aphidicolin 6µM treatment for 16 hours (APH) and tranfected with the indicated esiRNAs for the **a)** *Dux*, **b)** *Zscan4*, **c)** MERVL retroelements, **d)** *Tcstv3*, **e)** *Pou5f1* and **f)** *Nanog* genes. All data are normalized on *Tbp* expression and shown as fold change in respect to the CNT *Rluc* sample. All data are shown as mean + SEM of three independent experiments. Comparison among all conditions was performed by one-way ANOVA with Tukey post-test (* = p-val < 0.05, ** = p-val < 0.01, *** = p-val < 0.001).

Since in our previous work we demonstrated that RS could not induce the activation of 2C stage genes in differentiated cells, such as MEFs, we wondered if the inability of ESCs depleted for PRC2 components to effectively increase 2C genes expression may be due to their differentiation, induced by the loss of PRC2-dependent repression of main developmental genes. To verify this hypothesis, we measured the expression of two key pluripotency markers *Pou5f1* and *Nanog* in all our tested conditions. As shown in figures 22e and 22f, neither *Suz12* or *Ezh2* knock-down affected the expression of *Pou5f1* and

Nanog in unperturbed conditions. RS led to a small and not significant reduction in *Pou5f1* and almost no change in *Nanog* expression, but this variation was not affected in any extent by either *Suz12* or *Ezh2* depletion (Fig. 22e, f). Therefore, these data indicate that the inability of ESCs to activate the expression of *Dux* and its targets in response to RS upon ablation of *Suz12* or *Ezh2* is not due to a generic loss of stemness induced by the weakened PRC2-mediated repression.

Given the surprising effect of PRC2 ablation on the RS-induced activation of *Dux* and its targets, and since the two PRC1 components RING1b and RYBP showed ATR dependent accumulation on *Dux* locus in response to RS, we wondered whether also PRC1 could be involved in RS-mediated up-regulation of *Dux*. To this purpose, we combined the knock-down of *Ring1*, *Rnf2* (encoding RING1b) and *Rybp* with overnight (16 hours) treatment with aphidicolin 6μM, in order to induce RS in E14 mouse ESCs and measured the expression level of *Dux* and its main targets mRNAs by RT-real time qPCR, comparing it to the expression measured in cells transfected with an esiRNA targeting the Rluc gene. Also in this case, the samples either in unperturbed or treated conditions were harvested 48 hours post transfection for RNA extraction. Since both RING1b and RYBP proved in our hands to be involved in *Dux* repression in unperturbed conditions and, as previously shown, their ablation led to a stark increase in *Dux* levels over basal expression, the combination of PRC1 components knock-downs and RS induction could very likely lead to high levels of *Dux* expression in respect to unperturbed control cells. However, we reasoned that if PRC1 was involved in RS-mediated activation of *Dux*, RS would had not been able to increase *Dux* expression in ESCs subjected to PRC1 components ablation over the level of unperturbed cells subjected to the same knock-downs. However, as shown in figure 23a, RS induction in cells depleted for either *Rnf2* or *Rybp* led to a strong increase in *Dux* expression, well above the level measured in cells transfected with the *Rluc* targeting esiRNA when subjected to

the same treatment. Moreover, the extent of *Dux* up-regulation due to the combination of RS induction and *Rnf2* or *Rybp* depletion was well above the level observed in the presence of the respective knock-downs alone, indicating that even in the absence of PRC1 main components RS was still able to induce strong up-regulation of *Dux* (Fig. 23a). Interestingly, *Ring1* knock-down did not have any effect on *Dux* expression in response to RS, similarly to what we already observed in unperturbed conditions (Fig. 23a).

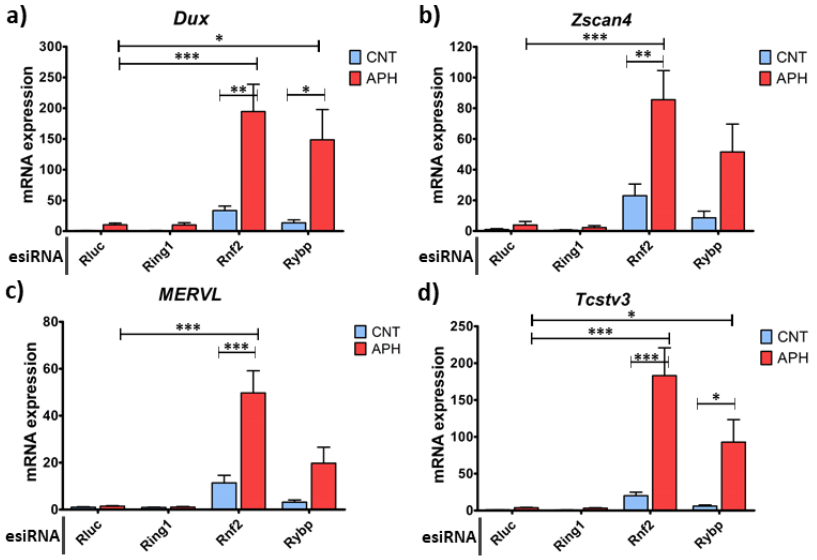

Figure 23 | PRC1 is not required for *Dux* activation upon RS: RT-real time qPCR analysis of mRNA levels in E14 cells either in unperturbed ESCs (CNT) after aphidicolin 6µM treatment for 16 hours (APH) and transfected with the indicated esiRNAs for the genes **a)** *Dux*, **b)** *Zscan4*, **c)** MERVL retroelements and **d)** *Tcstv3*. All data are normalized on *Tbp* expression and shown as fold change in respect to the CNT *Rluc* sample. All data are shown as mean + SEM of three independent experiments. Comparison among all conditions was performed by one-way ANOVA with Tukey post-test (* = p-val < 0.05, ** = p-val < 0.01, *** = p-val < 0.001).

A remarkably similar behavior was observed for all the tested 2C genes, showing also in this case a concerted regulation among *Dux* and its targets. In particular, with various extents, RS led to a strong increase in *Zscan4*, *MERVL* and *Tcstv3* expression both in *Rnf2* and *Rybp* depleted cells, if compared with cells subjected to transfection with the same esiRNA but kept in unperturbed conditions (Fig. 23b-d). At the same time, the combination of either *Rnf2* and *Rybp* knock-downs with aphidicolin treatment resulted in markedly higher expression of all three genes mRNAs in comparison to cells treated with aphidicolin alone (Fig. 23b-d). Finally, also in the case of *Zscan4*, *MERVL* and *Tcstv3* expression, *Ring1*

depletion did not have any effect either in unperturbed or RS-induced conditions (Fig. 23b-d).

These surprising data show that, not only PRC2 complexes are not responsible for *Dux* repression in unperturbed ESCs, but their presence is required for its efficient activation in response to RS, although the understanding of their role requires further investigation. At the same time, the accumulation of PRC1 components on *Dux* locus after RS induction observed in our Cut&Tag and ChIP-qPCR experiments, not only do not lead to a repression of the gene, but are also not instrumental to *Dux* activation in response to RS.

3.7 – Fork remodeling translocases are required for efficient activation of *Dux* and its targets in response to RS

In our previous work, we linked the activation of *Dux* and its target 2-cells stage genes in response to RS to the activation of the ATR kinase, the master regulator of RS response, and its main downstream effector CHK1. In our effort to expand our knowledge on how the activity of the checkpoint is linked to the transcriptional regulation of a main developmental gene, we decided to investigate the role of some of the most important genes known to have a role in replication fork stabilization and recovery, acting downstream to the ATR-CHK1 axis. Among the proteins of this group the better known actors are the helicase-like transcription factor HLTF (also known as SMARCA3) known for its ability to ubiquitinate PCNA at stalled replication forks, its putative downstream effector ZRANB3 and the ATP-dependent annealing helicase SMARCAL1 (121,125,203). All these proteins work downstream of the ATR checkpoint in order to protect and recover stalled or collapsed forks caused by high levels of RS, through the pathway of fork reversal. We wondered then, whether the activation of *Dux* in response to RS may not be directly due to the activity of the ATR-CHK1 checkpoint, but could be a product of fork remodeling activities promoted by this pathway. To answer this question, we transfected E14 mouse

ESCs with esiRNAs targeting either *Hltf*, *Zranb3* or *Smarcal1* mRNAs and kept these cells in either unperturbed conditions or treated them with aphidicolin 6µM for 16 hours to induce high RS. As a negative control we used an esiRNA pool targeting the gene *Rluc*. At 48 hours post transfection, cells were collected for both RNA extraction and subsequent RT-real time qPCR analysis and protein extraction for western blot analysis. As shown in figure 24a, none of the tested knock-downs had an effect on *Dux* basal expression level, although *Smarcal1* depletion resulted in a very mild and not significant increase.

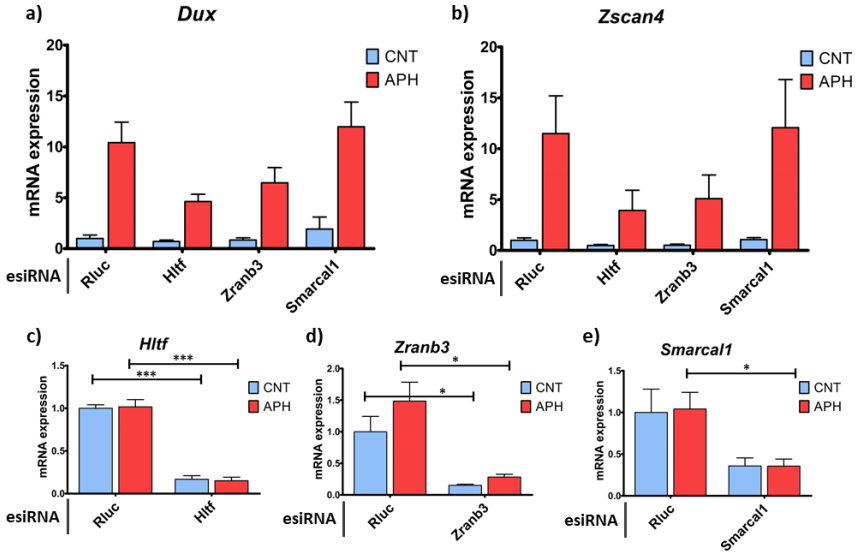

Figure 24 | Fork remodeling translocases are required for efficient up-regulation of *Dux* and its targets in response to RS: RT-real time qPCR analysis of mRNA levels in E14 cells either in unperturbed conditions (CNT) or after aphidicolin 6µM treatment for 16 hours (APH) and transfected with the indicated esiRNAs for the genes **a)** *Dux*, **b)** *Zscan4*, **c)** *Hltf*, **d)** *Zranb3* and **e)** *Smarcal1*. All data are normalized on *Tbp* expression and shown as fold change in respect to the CNT *Rluc* sample. All data are shown as mean + SEM of three independent experiments. Comparison among all conditions was performed by one-way ANOVA with Tukey post-test (* = p-val < 0.05, ** = p-val < 0.01, *** = p-val < 0.001).

Interestingly both *Hltf* and *Zranb3* ablation were able to impair *Dux* activation in response to RS, with *Hltf* displaying the strongest effect, although neither of these reductions resulted statistically significant (Fig. 24a). On the contrary, knocking-down *Smarcal1* did not have any effect on *Dux* expression upon RS (Fig. 24a), highlighting perchance a difference between the role of HLTF-ZRANB3 and SMARCAL1 in response to RS in ESCs. The expression of *Zscan4* mirrored what we observed in the case of *Dux*, with only *Hltf* and *Zranb3* knock-

downs showing a reduction in RS dependent activation of the gene, while no effect was observed in the case of *Smarcal1* depletion (Fig. 24b). Also in this case, the effect of the knock-downs on the basal expression of *Zscan4* was negligible, with only a minor reduction in the expression of the gene observed in unperturbed ESCs subjected to either *Hltf* or *Zranb3* knock-down (Fig. 24b). To confirm the effectiveness of our knock-down approach we measured the mRNA levels of *Hltf*, *Zranb3* and *Smarcal1* by RT-real time qPCR in cells transfected with either the *Rluc* esiRNA or the esiRNA targeting any of these genes, in both unperturbed or aphidicolin treated conditions. This analysis confirmed that all three genes were efficiently knocked-down, with *Hltf* and *Zranb3* showing a higher knock-down efficiency than *Smarcal1* (Fig. 24c-e). The effectiveness of our approach was confirmed also by our western blot analysis, showing that each of the used esiRNA led to an almost complete depletion of the corresponding protein in comparison to the control level, but no change in any of the others (Fig. 25). This analysis confirmed that also in the case of *Smarcal1*, where our qPCR analysis could still detect almost 40% of the control levels of mRNA after knock-down, the corresponding protein level was almost undetectable (Fig. 24e and 25). Importantly, our western blot analysis of ZSCAN4 protein expression showed the same pattern observed for *Dux* and *Zscan4* mRNAs, with only *Hltf* and *Zranb3* knock-downs being able to revert the up-regulation of ZSCAN4 due to RS induction. Interestingly, neither *Hltf* or *Zranb3* ablation could impair CHK1 phosphorylation in response to RS (Fig. 25), indicating that loss of the corresponding protein did not block the activation of the ATR-CHK1 axis, and therefore their role in *Dux* activation is most likely exerted downstream of the RS checkpoint.

These data move our understanding of the link between RS and 2C stage genes activation one step further, highlighting how the very process of fork remodeling and restart may be involved in *Dux* de-repression.

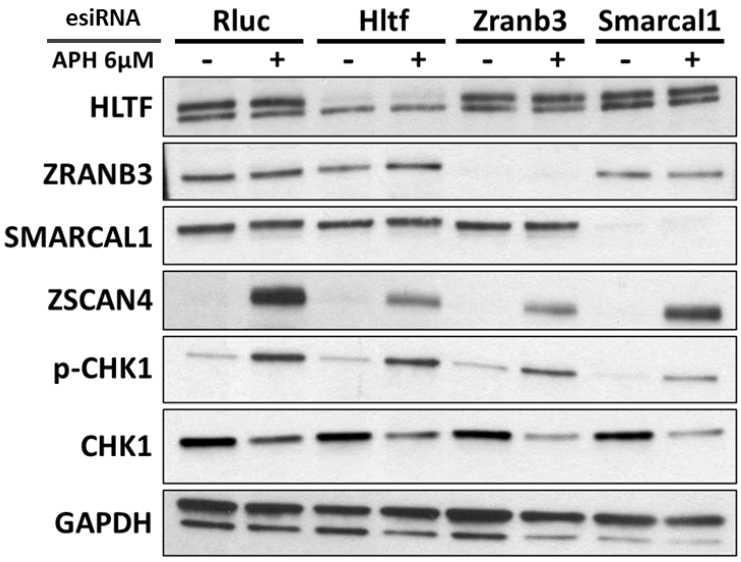

Figure 25 | HLTF or ZRANB3 depletion does not impair CHK1 activation upon RS induction: Western blot analysis for E14 cells either in unperturbed conditions or treated with aphidicolin 6µM for 16 hours. Cells were transfected with the indicated esiRNAs. Protein expression was tested for HLTF, ZRANB3, SMARCAL1, ZSCAN4, CHK1. Phosphorylation of CHK1 (p-CHK1) at Ser317 was tested as a marker of ATR-CHK1 activation. GAPDH was used as a loading control.

4 – Discussion

4.1 - PRC1 represses 2-cell stage genes in unperturbed ESCs

ESCs are characterized by a fast replicative cycle, due to the high number of replication origins firing at each cycle, combined with a short G1 phase. These features are the source of a great deal of RS in otherwise unperturbed ESCs (204,205), posing a risk for the stability of their genome. It is therefore surprising that ESCs are considered to be remarkably stable at genomic level (206). This indicates the presence of peculiar systems of dealing with RS in ESCs. The work of several labs has highlighted the link between the physiological process of transcription and RS, due to the risk of collision between the replication machinery and RNA polymerases, and the formation of highly recombinogenic structures known as R-loops (207). Many have reported how the cell puts in place various mechanisms to suppress transcription in response to RS, in order to reduce the risk of fork stalling and collapse (92,208). Given these premises, in our previous work, we surprisingly observed that, in response to RS, ESCs upregulate a vast group of genes (42). Such a broad increase in gene expression may in fact pose in even greater danger the stability of the genome, by widening the portion of DNA occupied by the transcriptional machinery. Among the genes overexpressed upon RS in our dataset, we observed a substantial percentage (ranging from 25% to 28%) of Polycomb targets (141,190,191). In ESCs, members of the Polycomb family, belonging to either PRC1 or PRC2, occupy the promoters of many developmental genes, keeping them under various degrees of repression (209,210). Interestingly, among the genes activated by RS and identified in our previous work, a wide group of 2-cells stage specific genes resulted particularly relevant, with their master regulator *Dux* occupying a key position (42). Worth of noting, *Dux* appeared to be among Polycomb targets in only one of the works taken into consideration by us to build sets of Polycomb targets (141). However, this is not surprising, since *Dux* repetitive nature affects the ability of sequencing-

based techniques to correctly identify it (58). Other more recent works have in fact shown the presence of the variant vPRC1.6 complex on the promoter of *Dux* (69), while depletion of vPRC1.6 components RING1b, RYBP and PCGF6 have proven to increase the number of 2C like cells among the ESCs population in unperturbed culture conditions (37). Here we confirm these results, showing that the role of vPRC1 in the repression of 2C-like genes is due to its repressive effect on *Dux* promoter. Interestingly, our results show that only RING1b is involved in *Dux* repression in ESCs, while its paralog RING1a is not involved either in *Dux* repression under unperturbed conditions or in its activation upon RS. This clearly indicates a non-redundant function of RING1b, highlighting a difference between the two paralogs. The two core components of PRC1 have in fact already been shown to have, at least partially, non-redundant functions, as highlighted by the broadly different phenotypes displayed by developmental models lacking either of the two (211,212). The binding of vPRC1.6 to *Dux* promoter has been suggested to require SUMOylation, as shown by the ability of *Ubc9* knock-downs to increase *Dux* expression and reduce the presence of PRC1 components on its promoter (69). However, given the broad role of UBC9 as an E2 enzyme involved in many cellular processes, from cell cycle regulation to fork stabilization and DNA damage repair (70,71,213), it is important to consider this data with caution. UBC9 loss may in fact affect the correct proceeding of the replicative forks, resulting in RS and therefore activation of 2-cells stage specific genes in ESCs through the ATR-CHK1 axis. Importantly, also RING1b and other PRC1 components, such as L3MBTL1, have been shown to associate with replication forks, and their deletion leads to impaired fork progression and RS (163,165). However, in our hands, a 48-hour long depletion of variant PRC1 components did not lead to strong activation of CHK1, comparable to what we could observe in the case of aphidicolin treatment. Notably, the levels of ZSCAN4 protein expression displayed by these knock-down experiments were markedly higher than the

expression levels shown by aphidicolin-treated ESCs, indicating that they cannot be explained by a mild activation of CHK1.

4.2 - Replication stress influences the binding of Polycomb to its targets

Our data shows that in ESCs RING1b is present throughout the nucleus, but it is particularly enriched in dense foci. These dense spots are generally coincident with the chromocenters, regions of condensed chromatin, which appear as brighter spots inside the nuclei of ESCs in standard DAPI staining. The presence of compact regions of chromatin, enriched in Polycomb proteins, is compatible with recent findings that indicate the presence of discrete chromatin domains associated with Polycomb marks (or PADs) in the nuclei of both ESCs and oocytes (193,214). These domains have been proposed to be critical for the maintenance of cell identity, and their loss leads to massive rearrangements of chromatin compartments (193,214). Here we observe that, in response to RS, ESCs rearrange the disposition of RING1b enriched domains, which assume a more disperse disposition in the nucleus. These change in localization corresponds to a parallel increase in the colocalization coefficient between RING1b and γH2AX, a common marker of RS, which is enriched at stalled forks due to the activation of the ATR checkpoint (111,215). This behavior is in agreement with a role of RING1b at replication forks. Interestingly, components of PRC1, such as PCGF2 (MEL18) and PCGF4 (BMI1), have been shown to be recruited to sites of DNA damage in a PARP dependent manner (165,168,169), while RING1b and RING1a themselves have been shown to be involved in replication of highly complex loci in unperturbed conditions (164,166). Our data cannot distinguish between a displacement of RING1b due to its accumulation in proximity of undamaged replication forks or to its recruitment at sites of collapsed forks, which have caused the formation of DSBs. RING1b disposition, however, was not mirrored by an analogous positioning of the PRC1 deposited mark

H2AK119Ub, which instead appeared more homogeneously distributed inside nuclei. This broad occupancy of the genome by H2AK119Ub may be due to the multiple functions of this histone mark. In addition to its role as a Polycomb-associated repressive mark, H2A ubiquitination is also increased in S-phase and is found in proximity of replication forks, where it is proposed to function as a scaffold for the recruitment of various fork stabilization factors, such as BAP1 and INO80 (216,217). At the same time, H2AK119Ub is known to be present in pericentromeric regions (164) and, given the association of PRC1 proteins with sites of DNA damage, it is possible that this histone mark is present at sites of damage throughout the nucleus (165,168,169). It is interesting to notice, that in our hands H2AK119Ub signal did not co-localize with γH2AX. However, given the fact that the H2AX isoform of the H2A histone has been proven, in previous works, to be ubiquitinated at lysine 119 (218), it is possible that our antibody could not efficiently recognize both isoforms, or that binding of the anti-γH2AX antibody interfered with the accessibility of the H2AK119Ub antibody binding site. Our Cut&Tag data confirmed that RING1b occupancy of its target promoters is affected by RS. Consistently, the loss of RING1b foci observed by immunofluorescence correlated with a decrease in the intensity of RING1b signal measured in proximity of the TSS of its targets. Loss of RING1b in the promoter region coincided in many cases with genes that were identified as up-regulated by our previous RNA-Seq experiment. Similarly, also EED presence on its target gene promoters was affected by RS, resulting in a broad loss of EED peaks, which, also in this case, partially overlapped with genes up-regulated in our previous work. However, it is important to notice that RING1b or EED loss is not sufficient to directly explain the broad up-regulation of genes that we observed in our published work, since the number of direct targets of Polycomb which lose EED or RING1b signal in response to RS is far inferior than the number of genes that experience up-regulation in the same conditions. This discrepancy could be explained by

the fact that, among the targets of Polycomb that are upregulated by RS, are present transcription factors, such as c-JUN, ID2 or ATF3, which in turn could activate a cohort of genes in their respective pathways. It is reasonable to think though, that RS may exert an effect on many different chromatin remodeling factors, either by not allowing their correct repositioning on the chromatin due to the presence of stalled replication forks, or by coopting them in the processes required for the protection and restart of the forks. Polycomb targets may be then only a group of a larger number of genes affected by the remodeling of the surrounding chromatin landscape in response to RS. Differently from RING1b, H2AK119Ub did not show any major decrease in peak intensity in proximity of TSS genome-wide upon RS induction. Instead, H2AK119Ub signal accumulated both upstream and downstream of genes TSS in response to RS. The expansion of H2AK119Ub covered domains may be in part responsible for the loss of significance that we observed among peaks in cells treated with aphidicolin in comparison to unperturbed cells, through a widespread increase in background signal. The genome-wide deposition of H2AK119Ub upon RS may be explained by extensive presence of stalled forks throughout the genome. As we show, exposure of mouse ESCs to high doses of aphidicolin leads to accumulation of these cells mostly in G1/S-phase, although a remarkable number of them is well into S-phase at the end of treatment. Given the fact that mouse ESCs are reported to have a very brief G1-phase, and to not undergo a Rb-dependent arrest in G1 in response to DNA damage (188,189,219), it is reasonable to think that in ESCs a high number of stalled replication forks is present in conditions of marked RS. Therefore, the broad deposition of H2AK119Ub that we observe may be reasonably due to a high number of slowly moving or stalled replication forks throughout the genome. In contrast to what we observed in the case of H2AK119Ub, the H3K27me3 histone mark displayed a dramatic loss of signal throughout the genome in proximity of gene TSS. The drop in H3K27me3 was even more

marked compared to RING1b and EED. The extent of this difference is striking, considering that the presence of Polycomb complexes at the sites of repressed genes is fairly stable throughout the process of replication (220–222). One simple hypothesis to explain the wide loss of H3K27me3 at gene promoters would be that widespread stalling of replication forks, due to aphidicolin inhibition of DNA polymerases, would leave large portions of single stranded DNA, devoid of nucleosomes and therefore of histone marks. However, this simple explanation would not fit our observation regarding the broad presence of H2AK119Ub around the genome, which shows that the amount of ssDNA resulting from RS is not of an extent that would affect the genome-wide association between nucleosomes and DNA.

During replication, nucleosomes are evicted by the replication machinery and re-deposited behind the fork. However, the doubling of DNA requires the deposition of newly synthesized histones, mostly devoid of histone marks, which are then replaced by various chromatin remodeling complexes, although a clear model for this process is still lacking (223). Our protein expression and immunofluorescence data show that H3K27me3 is increased upon RS, indicating that the global loss at gene promoters cannot be explained by a general suppression of H3K27 trimethylation activity. From a qualitative observation of our immunofluorescence data, it is also possible to see that H3K27me3 in unperturbed cells is preferentially disposed in proximity to the nuclear membrane, in agreement with the higher presence of heterochromatin at the nucleus borders (224). Upon RS instead, H3K27me3 disposition becomes less compartmentalized, and bright H3K27me3 foci appear in various positions inside nuclei. Notably, the presence of the H3K27me3 histone mark has been shown to decrease during DNA replication in both *Drosophila* and human cells, in correspondence to an parallel increase in H3K27 monomethylation, whose role remains elusive (225–227). The replacement of H3K27me3 is then achieved after cell division and

throughout the G1-phase (228). However, data from human cellular models indicates that RS affects the redeposition of nucleosomes after replication, by increasing their retention by histone chaperones (199). It is possible that the elevated level of RS may interfere with the correct recycling of histones behind forks. In addition, PRC2, similarly to PRC1, has been proven to be involved in DNA replication through association with the replication machinery (163), and to act at stalled replication forks to recruit remodeling factors through the deposition of H3K27me3 (167). Therefore, the coopting of PRC2 components at sites of stalled forks could reduce their ability to deposit H3K27me3 on newly synthesized histones, concurring to its general reduction on target genes TSS. In this case, the dilution of PRC2 due to its recruitment at replication forks may affect its ability to properly maintain the repressive marks at target sites in conditions of RS.

4.3 - The highly repeated *Dux* locus accumulates vPRC1 in response to replication stress

Interestingly, RING1b and its ability to ubiquitinate histone H2A has been shown to be required for the correct replication of highly complex sequences, such as pericentromeric satellites (166). Both RING1b and H2AK119Ub have been shown to accumulate at pericentromeric satellites during normal replication, and H2A ubiquitination was proven to suppress the RS shown by *Ring1*$^{-/-}$*Rnf2*$^{-/-}$ MEFs (164). The peculiarity of *Dux* genomic region is that its coding sequence is encased in a macrosatellite sequence, which is repeated an unknown number of times. Studies regarding its human homolog *DUX4* proved that the macrosatellite can be repeated more than 100 times in tandem, and data from mouse cells indicate a similar disposition (58,229). It is therefore reasonable to hypothesize that the accumulation of RING1b, that we observed in our Cut&Tag experiment over the *Dux* region after RS induction, could be due to RING1b role in the replication of complex sequences. The macrosatellite structure of *Dux* may be responsible for RING1b recruiting to its locus in

order to properly replicate this region, and fork stalling due to RS could be responsible for retaining high amounts of RING1b on the site. Quite interestingly, also RYBP showed enrichment on the *Dux* locus in our ChIP-qPCR experiments, demonstrating that RING1b is not recruited on *Dux* alone in response to RS, but it is most likely part of variant PRC1. Moreover, the increased localization of both vPRC1 components was abolished upon chemical inhibition of ATR, suggesting that the accumulation of these proteins on the locus requires a functional checkpoint. This data, in particular, offers an interesting starting point for further analysis. The increased presence of PRC1 in proximity of *Dux* upon RS may be the result of the accumulation of stalled replication forks on the locus, which would bring Polycomb proteins on the site. In this case, ATR inhibition would lead to increased collapse of the stalled forks, in this way reducing the presence of Polycomb on the site, together with the rest of the replicative complex. Alternatively, ATR activity may be directly required for proper recruitment of Polycomb to stalled replication forks, in an attempt to protect and restart the fork. It is however important to keep in mind that these two scenarios are not necessary mutually exclusive. Notably, H2AK119Ub only slightly accumulated over *Dux* in our Cut&Tag experiment, and this mild increase was not confirmed by ChIP-qPCR. However, it is important to notice that H2AK119Ub levels over *Dux* in unperturbed conditions are already high, due to its PRC1 dependent repression, and therefore increased RING1b presence on the locus may not result in further increase in this histone mark for a simple saturation effect. Notably, also EED showed a moderate increase over *Dux* locus in response to RS in our Cut&Tag experiment, possibly due to the reported role of PRC2 proteins in DNA replication (163). Interestingly, H3K27me3, in contrast to the behavior of EED, and similarly to what observed genome-wide, displayed a marked loss upon RS in both our Cut&Tag and ChIP-qPCR data. This shows that the presence of PRC2 on the locus is not *per se* sufficient to efficiently deposit H3K27me3. Differently from H2AK119Ub, H3K27me3

has been mostly linked with the remodeling of stalled forks (167). It is therefore possible that H3K27me3 is preferentially deposited at sites of forks that fail to be recovered, and is not broadly associated with forks in RS. However, given the sustained presence of PRC1 and PRC2 proteins at the *Dux* locus upon RS, the marked loss of H3K27me3 is still somehow counterintuitive and a better understanding of the fine mechanisms linking Polycomb to replication forks is required for a complete understanding of these results. It is possible that, due to its repetitive nature, the *Dux* macrosatellites tend to accumulate PRC1 marks in a similar way to pericentromeric loci (164,166). However, in the presence of a functional RS checkpoint it is also possible that this region is not significantly more prone to fork remodeling in comparison to the rest of the genome, and therefore does not show sensible accumulation of H3K27me3.

4.4 - *Dux* activation upon replication stress requires PRC2

The loss of H3K27me3 on *Dux* promoter in response to RS could be responsible of its overexpression, or at least enable its activation through other independent mechanisms. Our data shows that active demethylation of H3K27 by KDM6a and KDM6b is not required for proper *Dux* activation in response to RS. These data suggest that the loss of H3K27me3 observed on *Dux* promoter after RS induction is not dependent upon an active process of demethylation, although a more in-depth analysis is required to establish if KDM6a and KDM6b ablation has any direct effect on the level of H3K27me3 at *Dux* promoter, either in unperturbed conditions or in response to RS. Notably, knock-down of PRC2 components *Suz12* and *Ezh2* did not increase the expression of *Dux* and its targets in unperturbed conditions. These data, taken together with the inability of *Kdm6a*/*Kdm6b* knock-down to reduce *Dux* activation in response to RS, suggests that PRC2-dependent deposition of H3K27me3 is not the main controller of *Dux* status. However, our data do not allow us to exclude that loss of H3K27me3 on *Dux* promoter in response to RS may be instrumental to

its activation, by making the region available for binding by activating transcription factors. Surprisingly, ablation of either SUZ12 or EZH2 led to a mild decrease of *Dux* expression in unperturbed ESCs, which is in agreement with previously published work showing that knock-down of PRC2 proteins leads to a reduction in the 2C-like cell population among mouse ESCs (37). The reduction in *Dux* expression caused by *Suz12* and *Ezh2* knock-downs was even more marked in response to RS, where we could observe an almost complete suppression of the RS effect on the activation of *Dux* and its targets. Although this apparent activating effect of PRC2 proteins in response to RS may seem surprising at first, this result may be explained by the proven role of these proteins in DNA replication (163,167). Similarly to RING1b, in fact, also SUZ12 has been shown to colocalize with PCNA during replication (163), and EZH2 has proven to act at reversed forks in BRCA2 deficient cells to promote their degradation (167). MEFs deficient for various components of PRC2 have shown to grow at a slower pace (163). This suggests that the lower expression of *Dux* in PRC2 knock-down cells may be due to the inability of these cells to perform DNA replication. However, ablation of PRC2 proteins has been proven to cause RS in cancer cells and other model organisms (230,231), which would *per se* cause activation of *Dux* expression through the ATR-CHK1 axis in our cell model, in opposition to our data. Moreover, we could not see any sensible difference in the growth rate of *Suz12* and *Ezh2* knocked-down ESCs in comparison to their control in the time span observed (not shown). It is therefore possible that ESCs respond to ablation of PRC2 components differently from more differentiated cells, given also their inability to arrest in G1 in response to DNA damage (219,232–234). It is however important to verify the ability of PRC2 ablated cells to properly activate the ATR-CHK1 axis, to better characterize the role of this complex in the activation of *Dux*. Differently from PRC2 proteins, ablation of single components of PRC1 could not suppress the effect of RS on *Dux* levels. In particular, neither *Ring1* nor *Rnf2*

knock-down impaired RS-induced expression of *Dux* and its targets, indicating that these proteins are not required for *Dux* up-regulation. Since the involvement of RING1a and RING1b in replication of pericentromeric satellites seems to be redundant, it is possible that an eventual role of this process in *Dux* de-repression upon RS may not be observed through single knock-downs. However, given the complete lack of suppression shown by both *Ring1* and *Rnf2* knock-downs, and the dramatic effect of both *Suz12* and *Ezh2* knock-downs, it is reasonable to hypothesize that PRC1 and PRC2 play very different roles in response to RS, at least for what regards *Dux* activation.

4.5 - Fork remodeling is involved in *Dux* activation during replication stress

PRC2 role in the RS-dependent activation of *Dux* elicited the hypothesis that up-regulation of the gene may not be a direct effect of the ATR-CHK1 axis activation, but may be a consequence of fork remodeling processes, which are themselves regulated by the RS checkpoint. When replication forks progression is impaired, the cell enacts several mechanisms to ensure the protection of the fork and its recovery, such as translesion DNA synthesis, repriming and fork reversal (235). In particular, the formation of reversed forks promotes the stabilization of the replicative machinery on the fork, allowing the cell to overcome the stress source before proceeding with DNA replication. In this case, the newly synthesized strands are reannealed by the fork remodeling ATP-dependent helicase activity of proteins such as SMARCAL1, HLTF and ZRANB3, in order to reduce the amount of exposed ssDNA (117). Our data show that the fork remodeling proteins HLTF and ZRANB3, but not SMARCAL1, are required for efficient up-regulation of *Dux* transcript in response to RS. Unfortunately, a defined model for the action of each one of these proteins at the stalled fork is still lacking, although all three have been shown to promote fork reversal in conditions of RS (120,124,236). Interestingly, HLTF has been proven to promote

polyubiquitination of PCNA in conditions of RS (122,237) while ZRANB3 is known to be recruited at stalled forks through its ability to interact with polyubiquitinated PCNA (125). This evidence suggests that the similar suppressive effect shown by both proteins on RS-induced *Dux* expression may be explained by them acting in coordination through the same pathway. Further analysis, comprising the combination of both genes knock-downs is necessary to elucidate this point. In the light of these data and in combination to the proven role of ATR-CHK1 in the control of *Dux* expression, it is clear that replication forks remodeling is a requirement for *Dux* activation upon RS in ESCs. It would be therefore relevant to further the understanding of this phenomenon investigating the epistatic relation between ATR, HLTF and ZRANB3. In fact, while the action of HLTF and ZRANB3 is generally considered to happen downstream to ATR activation, a definitive model of their relation is still lacking. The ability of *Hltf* and *Zranb3* knocked-down ESCs to still activate CHK1 in response to RS is however an indication that the role of these proteins in RS-induced *Dux* expression takes place most probably downstream of ATR, or the two pathways concur independently to determine the same phenomenon.

4.6 - Final remarks

Taken together, our data depict a very complex portrait of *Dux* regulation in response to RS. Quite surprisingly, notwithstanding the clear role of vPRC1 in *Dux* repression under unperturbed conditions, their increased presence over *Dux* promoter in response to RS is not sufficient for its repression. It is possible that PRC1 complexes are recruited on *Dux* locus during replication due to its repeated nature, in order to grant its efficient replication, similarly to what has been demonstrated in the case of pericentromeric satellites (164). However, their presence on the locus does not affect the activity of the gene, possibly indicating a different composition of the PRC1 complexes acting in replication and in gene repression. At the same time, we show that PRC2 components are required for efficient

activation of *Dux*. In addition to their role in gene repression and in replication under normal conditions, PRC2 proteins have been shown to be also involved in the processing of reverse forks in concert with MUS81 and MRE11 (167). This evidence, together with our data indicating an involvement of fork remodeling translocases such as HLTF and ZRANB3 in *Dux* activation upon RS, suggests that the formation and processing of reverse forks may be required for the de-repression of the gene. It is important to underline that the role of PRC2 complexes at reversed forks has been linked to their ability to deposit H3K27me3, however, our observation that this repressive mark is clearly depleted on *Dux* locus upon RS, and the proven involvement of EZH2 in other DNA repair pathways in a methylase independent manner (238), suggest that PRC2 complexes may act at reversed forks also independently from their ability to methylate histone H3. According to our data, ATR and CHK1-induced fork pausing in response to RS, and the consequent formation of reversed forks, may be the main responsible for *Dux* de-repression in response to RS, although the fine molecular mechanism remains elusive (Fig. 26a). It is possible that the rearrangement of Polycomb complexes due to their recruitment to replication forks may lift their repressive action on the locus, due to a different composition of the Polycomb complexes involved in gene repression and replication (Fig. 26a). Alternatively, active transcription has been shown to be required for the restart of reverse forks caused by R-loops (93). However, since the main source of RS in aphidicolin treated cells is due to inhibition of DNA polymerases, which is not a source of R-loops, a deeper investigation is required in order to understand if transcription may be involved in fork restart also in case of different sources of stress. In addition, cooptation of PRC1 and PRC2 by stalled and reversed forks leads to a macroscopic rearrangement of their positioning throughout the nucleus, as shown by our Cut&Tag data (Fig. 26b). The loss of repressive marks on a genome-wide level constitutes a major risk for the ability of the cell to maintain its identity. Transcriptomic

data published in our previous work suggest that Polycomb proteins may not be the only chromatin remodeling actors whose positioning is affected by RS, given the large set of genes deregulated in these conditions. Progressive erosion of the chromatin landscape has been linked with the process of aging, and has been shown to be facilitated by DNA damaging factors (239). In addition, mouse models prone to unusually high RS have been proven to undergo accelerated aging and rapid exhaustion of the adult stem cell pools (240,241). Our data suggests a path through which RS may facilitate the progressive erosion of the chromatin landscape by coopting chromatin remodelers and repressor complexes such as Polycomb (Fig. 26b). Mild but protracted RS, or conditions of high RS due to mutations in key components of the DNA replication pathway, could accelerate aging by destabilizing the epigenome, which may have particularly dire consequences in highly plastic cells, such as stem cells (242). The effort of the cell to duplicate faithfully the information stored in the genome, and at the same time to regulate its expression to perform complex functions, all revolve around refined molecular machineries acting on the DNA. Our work offers new insight on how these two functions of the genome may come into conflict, and offers an insight on the dramatic consequences that may derive from such a conflict.

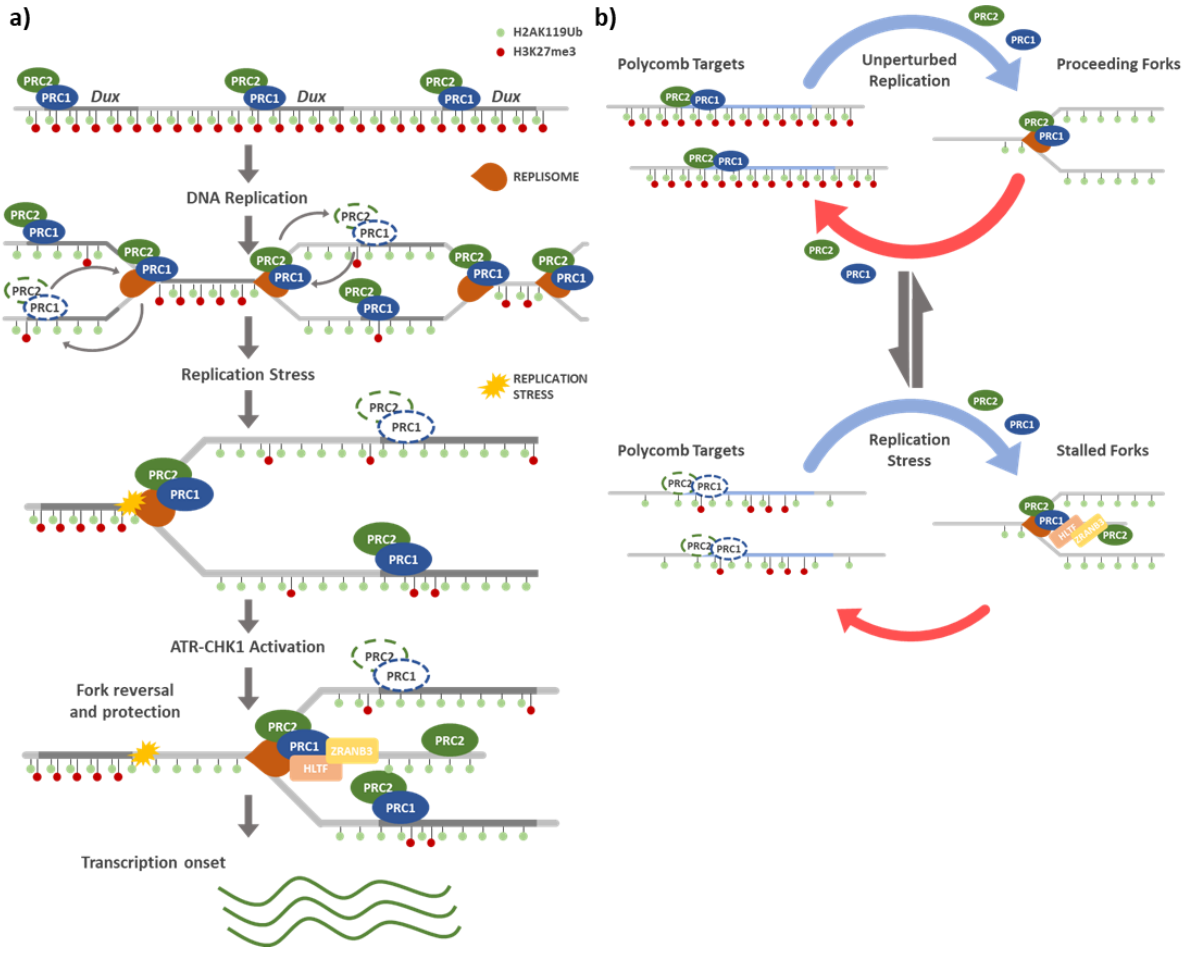

Figure 26 | Effects of Replication stress at the *Dux* locus and throughout the genome: a) Schematic of the proposed mechanism of *Dux* de-repression upon RS. *Dux* copies are disposed in tandem repeats along the locus and bound by PRC1 and PRC2 and marked with H2K119Ub and H3K27me3. During replication in unperturbed conditions H3K27me3 marked histones are evicted and replaced overtime after replication completion. The H2K119Ub is maintained at the macrosatellite for efficient replication. Both PRC1 and PRC2 are likely associated to the proceeding replication forks. Upon RS insurgence the stalled replication forks are stabilized by the activation of the ATR-CHK1 pathway. Upon recruitment of HLTF and ZRANB3 forks are reversed and additional PRC2 components are recruited at the site to process the reversed arm. Transcription is increased either as in correspondence to the processing of the forks or subsequently due to the rearrangement of Polycomb complexes on the locus. **b)** Proposed model of Polycomb targets de-repression upon RS. In unperturbed conditions Polycomb proteins are both required for gene repression and in association to replication forks for efficient DNA replication. The balance of this equilibrium shifts towards replication forks during S-phase and returns in favor of gene repression after replication completion. In conditions of RS slowed and stalled forks retain for longer time Polycomb proteins. In addition, reversed forks recruit further PRC2 components (and possibly PRC1). The prolonged presence of Polycomb proteins at fork sites shifts the balance from gene repression and repressive marks are depleted from target promoters.

Bibliography

1. Zernicka-Goetz M, Morris SA, Bruce AW. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. Nat Rev Genet. 2009 Jul;10(7):467–77.

2. Arias AM, Nichols J, Schröter C. A molecular basis for developmental plasticity in early mammalian embryos. Development. 2013 Sep 1;140(17):3499–510.

3. Soncin F, Natale D, Parast MM. Signaling pathways in mouse and human trophoblast differentiation: a comparative review. Cell Mol Life Sci. 2015 Apr 1;72(7):1291–302.

4. Beddington RSP, Robertson EJ. Axis Development and Early Asymmetry in Mammals. Cell. 1999 Jan 22;96(2):195–209.

5. Svoboda P. Mammalian zygotic genome activation. Semin Cell Dev Biol. 2018 Dec 1;84:118–26.

6. Zeng F, Baldwin DA, Schultz RM. Transcript profiling during preimplantation mouse development. Dev Biol. 2004 Aug 15;272(2):483–96.

7. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, et al. Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. Dev Cell. 2004 Oct 1;7(4):597–606.

8. Saiz N, Plusa B. Early cell fate decisions in the mouse embryo in: Reproduction Volume 145 Issue 3 (2013). Reproduction. 2013;145(3):R65–80.

9. Bin GCL, Muñoz-Descalzo S, Kurowski A, Leitch H, Lou X, Mansfield W, et al. Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. Development. 2014 Mar 1;141(5):1001–10.

10. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, et al. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. Cell. 2003 May 30;113(5):631–42.

11. Solter D, Škreb N, Damjanov I. Extrauterine Growth of Mouse Egg-cylinders results in Malignant Teratoma. Nature. 1970 Aug;227(5257):503–4.

12. Stevens LC. The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. Dev Biol. 1970 Mar 1;21(3):364–82.

13. Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature. 1981 Jul;292(5819):154–6.

14. Nagy A, Gocza E, Diaz EM, Prideaux VR, Ivanyi E, Markkula M, et al. Embryonic stem cells alone are able to support fetal development in the mouse. Development. 1990 Nov 1;110(3):815–21.

15. Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Becker RA, et al. Isolation of a primate embryonic stem cell line. Proc Natl Acad Sci. 1995 Aug 15;92(17):7844–8.

16. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, et al. Embryonic Stem Cell Lines Derived from Human Blastocysts. Science. 1998 Nov 6;282(5391):1145–7.

17. Martin GR, Evans MJ. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. Proc Natl Acad Sci U S A. 1975 Apr;72(4):1441–5.

18. Yoshida K, Chambers I, Nichols J, Smith A, Saito M, Yasukawa K, et al. Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. Mech Dev. 1994 Feb 1;45(2):163–71.

19. Niwa H, Burdon T, Chambers I, Smith A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. 1998 Jul 1;12(13):2048–60.

20. Stewart CL, Kaspar P, Brunet LJ, Bhatt H, Gadi I, Köntgen F, et al. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. Nature. 1992 Sep;359(6390):76–9.

21. Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, et al. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature. 1988 Dec;336(6200):684–7.

22. Hao J, Li T-G, Qi X, Zhao D-F, Zhao G-Q. WNT/β-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. Dev Biol. 2006 Feb 1;290(1):81–91.

23. Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med. 2004 Jan;10(1):55–63.

24. Burdon T, Stracey C, Chambers I, Nichols J, Smith A. Suppression of SHP-2 and ERK Signalling Promotes Self-Renewal of Mouse Embryonic Stem Cells. Dev Biol. 1999 Jun 1;210(1):30–43.

25. Ying Q-L, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, et al. The ground state of embryonic stem cell self-renewal. Nature. 2008 May;453(7194):519–23.

26. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet. 2000 Apr;24(4):372–6.

27. Ambrosetti D-C, Schöler HR, Dailey L, Basilico C. Modulation of the Activity of Multiple Transcriptional Activation Domains by the DNA Binding Domains Mediates the Synergistic Action of Sox2 and Oct-3 on the Fibroblast Growth Factor-4Enhancer. J Biol Chem. 2000 Jul 28;275(30):23387–97.

28. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 2003 Jan 1;17(1):126–40.

29. Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, et al. Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. Cell. 2003 May 30;113(5):643–55.

30. Kleinsmith LJ, Pierce GB. Multipotentiality of Single Embryonal Carcinoma Cells. Cancer Res. 1964 Oct 1;24(9):1544–51.

31. Brinster RL. The effect of cells transferred into the mouse blastocyst on subsequent development. J Exp Med. 1974 Oct 1;140(4):1049–56.

32. Torres-Padilla M-E, Chambers I. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. Development. 2014 Jun 1;141(11):2173–81.

33. Falco G, Lee S-L, Stanghellini I, Bassey UC, Hamatani T, Ko MSH. Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. Dev Biol. 2007 Jul 15;307(2):539–50.

34. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature. 2012 Jul 5;487(7405):57–63.

35. Eckersley-Maslin MA, Svensson V, Krueger C, Stubbs TM, Giehr P, Krueger F, et al. MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. Cell Rep. 2016 Sep;17(1):179–92.

36. Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee S-L, et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature. 2010 Apr;464(7290):858–63.

37. Rodriguez-Terrones D, Gaume X, Ishiuchi T, Weiss A, Kopp A, Kruse K, et al. A molecular roadmap for the emergence of early-embryonic-like cells in culture. Nat Genet. 2018 Jan;50(1):106–19.

38. Fu X, Djekidel MN, Zhang Y. A transcriptional roadmap for 2C-like–to–pluripotent state transition. Sci Adv. 2020 May 1;6(22):eaay5181.

39. Amano T, Hirata T, Falco G, Monti M, Sharova LV, Amano M, et al. Zscan4 restores the developmental potency of embryonic stem cells. Nat Commun. 2013 Jun 6;4.

40. Zhang Q, Dan J, Wang H, Guo R, Mao J, Fu H, et al. Tcstv1 and Tcstv3 elongate telomeres of mouse ES cells. Sci Rep. 2016 Jan 27;6.

41. Choi YJ, Lin C-P, Risso D, Chen S, Kim TA, Tan MH, et al. Deficiency of microRNA *miR-34a* expands cell fate potential in pluripotent stem cells. Science. 2017 Feb 10;355(6325):eaag1927.

42. Atashpaz S, Samadi Shams S, Gonzalez JM, Sebestyén E, Arghavanifard N, Gnocchi A, et al. ATR expands embryonic stem cell fate potential in response to replication stress. Aguilera A, Tyler JK, Bartek J, editors. eLife. 2020 Mar 12;9:e54756.

43. Iturbide A, Torres-Padilla M-E. A cell in hand is worth two in the embryo: recent advances in 2-cell like cell reprogramming. Curr Opin Genet Dev. 2020 Oct 1;64:26–30.

44. Akiyama T, Xin L, Oda M, Sharov AA, Amano M, Piao Y, et al. Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. DNA Res. 2015 Oct;22(5):307–18.

45. Markiewicz-Potoczny M, Lobanova A, Loeb AM, Kirak O, Olbrich T, Ruiz S, et al. TRF2-mediated telomere protection is dispensable in pluripotent stem cells. Nature. 2020 Nov 25;1–6.

46. Srinivasan R, Nady N, Arora N, Hsieh LJ, Swigut T, Narlikar GJ, et al. Zscan4 binds nucleosomal microsatellite DNA and protects mouse two-cell embryos from DNA damage. Sci Adv. 2020 Mar 1;6(12):eaaz9115.

47. Cheng Z-L, Zhang M-L, Lin H-P, Gao C, Song J-B, Zheng Z, et al. The Zscan4-Tet2 Transcription Nexus Regulates Metabolic Rewiring and Enhances Proteostasis to Promote Reprogramming. Cell Rep. 2020 Jul 14;32(2).

48. Storm MP, Kumpfmueller B, Bone HK, Buchholz M, Sanchez Ripoll Y, Chaudhuri JB, et al. Zscan4 Is Regulated by PI3-Kinase and DNA-Damaging Agents and Directly Interacts with the Transcriptional Repressors LSD1 and CtBP2 in Mouse Embryonic Stem Cells. Kyba M, editor. PLoS ONE. 2014 Mar 3;9(3):e89821.

49. Hirata T, Amano T, Nakatake Y, Amano M, Piao Y, Hoang HG, et al. Zscan4 transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells. Sci Rep. 2012 Dec;2(1).

50. Bénit L, De Parseval N, Casella JF, Callebaut I, Cordonnier A, Heidmann T. Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. J Virol. 1997 Jul;71(7):5652–7.

51. Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim J-W, Wike CL, et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet. 2017 May 1;49(6):925–34.

52. Zhang W, Chen F, Chen R, Xie D, Yang J, Zhao X, et al. Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. Nucleic Acids Res. 2019 Sep 19;47(16):8485–501.

53. Hung SSC, Wong RCB, Sharov AA, Nakatake Y, Yu H, Ko MSH. Repression of global protein synthesis by Eif1a-like genes that are expressed specifically in the two-cell embryos and the transient Zscan4-positive state of embryonic stem cells. DNA Res Int J Rapid Publ Rep Genes Genomes. 2013 Aug;20(4):391–402.

54. Richards M, Coppée F, Thomas N, Belayew A, Upadhyaya M. Facioscapulohumeral muscular dystrophy (FSHD): an enigma unravelled? Hum Genet. 2012 Mar;131(3):325–40.

55. Ding H, Beckers MC, Plaisance S, Marynen P, Collen D, Belayew A. Characterization of a double homeodomain protein (DUX1) encoded by a cDNA homologous to 3.3 kb dispersed repeated elements. Hum Mol Genet. 1998 Oct;7(11):1681–94.

56. Kowaljow V, Marcowycz A, Ansseau E, Conde CB, Sauvage S, Mattéotti C, et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord. 2007 Aug 1;17(8):611–23.

57. Geng LN, Yao Z, Snider L, Fong AP, Cech JN, Young JM, et al. DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy. Dev Cell. 2012 Jan 17;22(1):38–51.

58. Clapp J, Mitchell LM, Bolland DJ, Fantes J, Corcoran AE, Scotting PJ, et al. Evolutionary Conservation of a Coding Function for D4Z4, the Tandem DNA Repeat Mutated in Facioscapulohumeral Muscular Dystrophy. Am J Hum Genet. 2007 Aug;81(2):264–79.

59. Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, Felicetti L. Physical Mapping Evidence for a Duplicated Region on Chromosome 10qter Showing High Homology with the Facioscapulohumeral Muscular Dystrophy Locus on Chromosome 4qter. Eur J Hum Genet. 1995 May;3(3):155–67.

60. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet. 2017 May 1;49(6):941–5.

61. Whiddon JL, Langford AT, Wong C-J, Zhong JW, Tapscott SJ. Conservation and innovation in the DUX4-family gene network. Nat Genet. 2017 May 1;49(6):935–40.

62. Chen Z, Zhang Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. Nat Genet. 2019 Jun;51(6):947–51.

63. Wu K, Liu H, Wang Y, He J, Xu S, Chen Y, et al. SETDB1-Mediated Cell Fate Transition between 2C-Like and Pluripotent States. Cell Rep. 2020 Jan 7;30(1):25-36.e6.

64. Ishiuchi T, Enriquez-Gasca R, Mizutani E, Bošković A, Ziegler-Birling C, Rodriguez-Terrones D, et al. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat Struct Mol Biol. 2015 Sep;22(9):662–71.

65. Hu Z, Tan DEK, Chia G, Tan H, Leong HF, Chen BJ, et al. Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. Nat Cell Biol. 2020 Feb;22(2):175–86.

66. De Iaco A, Coudray A, Duc J, Trono D. DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. EMBO Rep. 2019 May 1;20(5):e47382.

67. Eckersley-Maslin M, Alda-Catalinas C, Blotenburg M, Kreibich E, Krueger C, Reik W. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. Genes Dev. 2019 01;33(3–4):194–208.

68. Percharde M, Lin C-J, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, et al. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell. 2018 12;174(2):391-405.e19.

69. Cossec J-C, Theurillat I, Chica C, Búa Aguín S, Gaume X, Andrieux A, et al. SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. Cell Stem Cell. 2018 Nov;23(5):742-757.e8.

70. Branzei D, Sollier J, Liberi G, Zhao X, Maeda D, Seki M, et al. Ubc9- and Mms21-Mediated Sumoylation Counteracts Recombinogenic Events at Damaged Replication Forks. Cell. 2006 Nov 3;127(3):509–22.

71. Moschos SJ, Mo Y-Y. Role of SUMO/Ubc9 in DNA damage repair and tumorigenesis. J Mol Histol. 2006 Sep;37(5–7):309–19.

72. Campbell AE, Shadle SC, Jagannathan S, Lim J-W, Resnick R, Tawil R, et al. NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. eLife. 2018 Mar 13;7:e31023.

73. Delgado S, Gómez M, Bird A, Antequera F. Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J. 1998 Apr 15;17(8):2426–35.

74. Vashee S, Cvetic C, Lu W, Simancek P, Kelly TJ, Walter JC. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes Dev. 2003 Aug 1;17(15):1894–908.

75. Méchali M. Eukaryotic DNA replication origins: many choices for appropriate answers. Nat Rev Mol Cell Biol. 2010 Oct;11(10):728–38.

76. Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, et al. Replication dynamics of the yeast genome. Science. 2001 Oct 5;294(5540):115–21.

77. Rowles A, Chong JPJ, Brown L, Howell M, Evan GI, Blow JJ. Interaction between the Origin Recognition Complex and the Replication Licensing Systemin Xenopus. Cell. 1996 Oct 18;87(2):287–96.

78. Gillespie PJ, Li A, Blow JJ. Reconstitution of licensed replication origins on Xenopus sperm nuclei using purified proteins. BMC Biochem. 2001 Dec 5;2(1):15.

79. Nishitani H, Sugimoto N, Roukos V, Nakanishi Y, Saijo M, Obuse C, et al. Two E3 ubiquitin ligases, SCF-Skp2 and DDB1-Cul4, target human Cdt1 for proteolysis. EMBO J. 2006 Mar 8;25(5):1126–36.

80. Liu E, Li X, Yan F, Zhao Q, Wu X. Cyclin-dependent Kinases Phosphorylate Human Cdt1 and Induce Its Degradation. J Biol Chem. 2004 Apr 23;279(17):17283–8.

81. Melixetian M, Ballabeni A, Masiero L, Gasparini P, Zamponi R, Bartek J, et al. Loss of Geminin induces rereplication in the presence of functional p53. J Cell Biol. 2004 May 24;165(4):473–82.

82. Zou L, Stillman B. Assembly of a Complex Containing Cdc45p, Replication Protein A, and Mcm2p at Replication Origins Controlled by S-Phase Cyclin-Dependent Kinases and Cdc7p-Dbf4p Kinase. Mol Cell Biol. 2000 May 1;20(9):3086–96.

83. Mimura S, Masuda T, Matsui T, Takisawa H. Central role for Cdc45 in establishing an initiation complex of DNA replication in Xenopus egg extracts. Genes Cells. 2000;5(6):439–52.

84. Moyer SE, Lewis PW, Botchan MR. Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc Natl Acad Sci. 2006 Jul 5;103(27):10236–41.

85. Moldovan G-L, Pfander B, Jentsch S. PCNA, the Maestro of the Replication Fork. Cell. 2007 May 18;129(4):665–79.

86. Okazaki R, Okazaki T, Sakabe K, Sugimoto K, Sugino A. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. Proc Natl Acad Sci U S A. 1968 Feb;59(2):598–605.

87. Balakrishnan L, Bambara RA. Okazaki Fragment Metabolism. Cold Spring Harb Perspect Biol. 2013 Jan 2;5(2):a010173.

88. Branzei D, Foiani M. Maintaining genome stability at the replication fork. Nat Rev Mol Cell Biol. 2010 Mar;11(3):208–19.

89. Maric M, Maculins T, Piccoli GD, Labib K. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. Science. 2014 Oct 24;346(6208).

90. Kubota T, Nishimura K, Kanemaki MT, Donaldson AD. The Elg1 Replication Factor C-like Complex Functions in PCNA Unloading during DNA Replication. Mol Cell. 2013 Apr 25;50(2):273–80.

91. Mirkin SM. Expandable DNA repeats and human disease. Nature. 2007 Jun;447(7147):932–40.

92. Castel SE, Ren J, Bhattacharjee S, Chang A-Y, Sánchez M, Valbuena A, et al. Dicer Promotes Transcription Termination at Sites of Replication Stress to Maintain Genome Stability. Cell. 2014 Oct 23;159(3):572–83.

93. Chappidi N, Nascakova Z, Boleslavska B, Zellweger R, Isik E, Andrs M, et al. Fork Cleavage-Religation Cycle and Active Transcription Mediate Replication Restart after Fork Stalling at Co-transcriptional R-Loops. Mol Cell. 2020 Feb 6;77(3):528-541.e8.

94. Durkin SG, Glover TW. Chromosome fragile sites. Annu Rev Genet. 2007;41:169–92.

95. Debatisse M, Le Tallec B, Letessier A, Dutrillaux B, Brison O. Common fragile sites: mechanisms of instability revisited. Trends Genet. 2012 Jan 1;28(1):22–32.

96. Rozenzhak S, Mejía-Ramírez E, Williams JS, Schaffer L, Hammond JA, Head SR, et al. Rad3ATR Decorates Critical Chromosomal Domains with γH2A to Protect Genome Integrity during S-Phase in Fission Yeast. PLOS Genet. 2010 Jul 22;6(7):e1001032.

97. Hossain M, Stillman B. Meier-Gorlin syndrome mutations disrupt an Orc1 CDK inhibitory domain and cause centrosome reduplication. Genes Dev. 2012 Aug 15;26(16):1797–810.

98. Kotsantis P, Petermann E, Boulton SJ. Mechanisms of Oncogene-Induced Replication Stress: Jigsaw Falling into Place. Cancer Discov. 2018 May 1;8(5):537–55.

99. Setlow RB, Swenson PA, Carrier WL. Thymine Dimers and Inhibition of DNA Synthesis by Ultraviolet Irradiation of Cells. Science. 1963 Dec 13;142(3598):1464–6.

100. Bianchi V, Pontis E, Reichard P. Changes of deoxyribonucleoside triphosphate pools induced by hydroxyurea and their relation to DNA synthesis. J Biol Chem. 1986 May 12;261(34):16037–42.

101. Eastman A. The formation, isolation and characterization of DNA adducts produced by anticancer platinum complexes. Pharmacol Ther. 1987 Jan 1;34(2):155–66.

102. Hsiang Y-H. DNA Topoisomerase I-mediated DNA Cleavage and Cytotoxicity of Camptothecin Analogues. Cancer Res. 1989;49(16).

103. Edwards CM, Glisson BS, King CK, Smallwood-Kentro S, Ross WE. Etoposide-induced DNA cleavage in human leukemia cells. Cancer Chemother Pharmacol. 1987;20(2):162–8.

104. Ikegami S, Taguchi T, Ohashi M, Oguro M, Nagano H, Mano Y. Aphidicolin prevents mitotic cell division by interfering with the activity of DNA polymerase-α. Nature. 1978 Oct;275(5679):458–60.

105. Yazinski SA, Zou L. Functions, Regulation, and Therapeutic Implications of the ATR Checkpoint Pathway. Annu Rev Genet. 2016 Nov 23;50:155–73.

106. Chen H, Lisby M, Symington LS. RPA coordinates DNA end resection and prevents formation of DNA hairpins. Mol Cell. 2013 May 23;50(4):589–600.

107. Brown EJ, Baltimore D. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. Genes Dev. 2000;14:397–402.

108. Zou L, Elledge SJ. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science. 2003 Jun 6;300(5625):1542–8.

109. Liu S, Shiotani B, Lahiri M, Maréchal A, Tse A, Leung CCY, et al. ATR autophosphorylation as a molecular switch for checkpoint activation. Mol Cell. 2011 Jul 22;43(2):192–202.

110. Haahr P, Hoffmann S, Tollenaere MAX, Ho T, Toledo LI, Mann M, et al. Activation of the ATR kinase by the RPA-binding protein ETAA1. Nat Cell Biol. 2016 Nov;18(11):1196–207.

111. Chanoux RA, Yin B, Urtishak KA, Asare A, Bassing CH, Brown EJ. ATR and H2AX Cooperate in Maintaining Genome Stability under Replication Stress. J Biol Chem. 2009 Feb 27;284(9):5994–6003.

112. Busino L, Donzelli M, Chiesa M, Guardavaccaro D, Ganoth D, Dorrello NV, et al. Degradation of Cdc25A by beta-TrCP during S phase and in response to DNA damage. Nature. 2003 Nov 6;426(6962):87–91.

113. Guo C, Kumagai A, Schlacher K, Shevchenko A, Shevchenko A, Dunphy WG. Interaction of Chk1 with Treslin negatively regulates the initiation of chromosomal DNA replication. Mol Cell. 2015 Feb 5;57(3):492–505.

114. Couch FB, Bansbach CE, Driscoll R, Luzwick JW, Glick GG, Bétous R, et al. ATR phosphorylates SMARCAL1 to prevent replication fork collapse. Genes Dev. 2013 Jul 15;27(14):1610–23.

115. Sale JE, Lehmann AR, Woodgate R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. Nat Rev Mol Cell Biol. 2012 Mar;13(3):141–52.

116. Bianchi J, Rudd SG, Jozwiakowski SK, Bailey LJ, Soura V, Taylor E, et al. PrimPol Bypasses UV Photoproducts during Eukaryotic Chromosomal DNA Replication. Mol Cell. 2013 Nov 21;52(4):566–73.

117. Joseph SA, Taglialatela A, Leuzzi G, Huang J-W, Cuella-Martin R, Ciccia A. Time for remodeling: SNF2-family DNA translocases in replication fork metabolism and human disease. DNA Repair. 2020 Nov 1;95:102943.

118. Bhat KP, Cortez D. RPA and RAD51: fork reversal, fork protection, and genome stability. Nat Struct Mol Biol. 2018 Jun;25(6):446–53.

119. Minca EC, Kowalski D. Multiple Rad5 activities mediate sister chromatid recombination to bypass DNA damage at stalled replication forks. Mol Cell. 2010 Jun 11;38(5):649–61.

120. Bai G, Kermi C, Stoy H, Schiltz CJ, Bacal J, Zaino AM, et al. HLTF Promotes Fork Reversal, Limiting Replication Stress Resistance and Preventing Multiple Mechanisms of Unrestrained DNA Synthesis. Mol Cell. 2020 Jun 18;78(6):1237-1251.e7.

121. Motegi A, Liaw H-J, Lee K-Y, Roest HP, Maas A, Wu X, et al. Polyubiquitination of proliferating cell nuclear antigen by HLTF and SHPRH prevents genomic instability from stalled replication forks. Proc Natl Acad Sci. 2008 Aug 26;105(34):12411–6.

122. Unk I, Hajdú I, Blastyák A, Haracska L. Role of yeast Rad5 and its human orthologs, HLTF and SHPRH in DNA damage tolerance. DNA Repair. 2010 Mar 2;9(3):257–67.

123. Kile AC, Chavez DA, Bacal J, Eldirany S, Korzhnev DM, Bezsonova I, et al. HLTF's Ancient HIRAN Domain Binds 3' DNA Ends to Drive Replication Fork Reversal. Mol Cell. 2015 Jun 18;58(6):1090–100.

124. Vujanovic M, Krietsch J, Raso MC, Terraneo N, Zellweger R, Schmid JA, et al. Replication Fork Slowing and Reversal upon DNA Damage Require PCNA Polyubiquitination and ZRANB3 DNA Translocase Activity. Mol Cell. 2017 Sep 7;67(5):882-890.e5.

125. Ciccia A, Nimonkar AV, Hu Y, Hajdu I, Achar YJ, Izhar L, et al. Polyubiquitinated PCNA Recruits the ZRANB3 Translocase to Maintain Genomic Integrity after Replication Stress. Mol Cell. 2012 Aug 10;47(3):396–409.

126. Weston R, Peeters H, Ahel D. ZRANB3 is a structure-specific ATP-dependent endonuclease involved in replication stress response. Genes Dev. 2012 Jul 15;26(14):1558–72.

127. Yusufzai T, Kong X, Yokomori K, Kadonaga JT. The annealing helicase HARP is recruited to DNA repair sites via an interaction with RPA. Genes Dev. 2009 Oct 15;23(20):2400–4.

128. Bétous R, Mason AC, Rambo RP, Bansbach CE, Badu-Nkansah A, Sirbu BM, et al. SMARCAL1 catalyzes fork regression and Holliday junction migration to maintain genome stability during DNA replication. Genes Dev. 2012 Jan 15;26(2):151–62.

129. Lemaçon D, Jackson J, Quinet A, Brickner JR, Li S, Yazinski S, et al. MRE11 and EXO1 nucleases degrade reversed forks and elicit MUS81-dependent fork rescue in BRCA2-deficient cells. Nat Commun. 2017 Oct 16;8(1):860.

130. Thangavel S, Berti M, Levikova M, Pinto C, Gomathinayagam S, Vujanovic M, et al. DNA2 drives processing and restart of reversed replication forks in human cells. J Cell Biol. 2015 Mar 2;208(5):545–62.

131. Kolinjivadi AM, Sannino V, De Antoni A, Zadorozhny K, Kilkenny M, Técher H, et al. Smarcal1-Mediated Fork Reversal Triggers Mre11-Dependent Degradation of Nascent DNA in the Absence of Brca2 and Stable Rad51 Nucleofilaments. Mol Cell. 2017 Sep 7;67(5):867-881.e7.

132. Schuettengruber B, Bourbon H-M, Di Croce L, Cavalli G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell. 2017 Sep;171(1):34–57.

133. Lewis EB. A gene complex controlling segmentation in Drosophila. Nature. 1978 Dec;276(5688):565–70.

134. Blackledge NP, Rose NR, Klose RJ. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. Nat Rev Mol Cell Biol. 2015 Nov;16(11):643–9.

135. Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, et al. PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. Mol Cell. 2012 Feb 10;45(3):344–56.

136. Robinson AK, Leal BZ, Chadwell LV, Wang R, Ilangovan U, Kaur Y, et al. The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. J Biol Chem. 2012 Mar 16;287(12):8702–13.

137. Antonysamy S, Condon B, Druzina Z, Bonanno JB, Gheyi T, Zhang F, et al. Structural Context of Disease-Associated Mutations and Putative Mechanism of Autoinhibition Revealed by X-Ray Crystallographic Analysis of the EZH2-SET Domain. PLOS ONE. 2013 Dec 19;8(12):e84147.

138. Hauri S, Comoglio F, Seimiya M, Gerstung M, Glatter T, Hansen K, et al. A High-Density Map for Navigating the Human Polycomb Complexome. Cell Rep. 2016 Oct 4;17(2):583–95.

139. Kloet SL, Makowski MM, Baymaz HI, van Voorthuijsen L, Karemaker ID, Santanach A, et al. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. Nat Struct Mol Biol. 2016 Jul;23(7):682–90.

140. Son J, Shen SS, Margueron R, Reinberg D. Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. Genes Dev. 2013 Dec 15;27(24):2663–77.

141. Pasini D, Cloos PAC, Walfridsson J, Olsson L, Bukowski J-P, Johansen JV, et al. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature. 2010 Mar;464(7286):306–10.

142. Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, et al. Role of histone H3 lysine 27 methylation in X inactivation. Science. 2003 Apr 4;300(5616):131–5.

143. Davidovich C, Cech TR. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA N Y N. 2015 Dec;21(12):2007–22.

144. Herranz N, Pasini D, Díaz VM, Francí C, Gutierrez A, Dave N, et al. Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. Mol Cell Biol. 2008 Aug;28(15):4772–81.

145. Li G, Margueron R, Ku M, Chambon P, Bernstein BE, Reinberg D. Jarid2 and PRC2, partners in regulating gene expression. Genes Dev. 2010 Feb 15;24(4):368–80.

146. Davidovich C, Zheng L, Goodrich KJ, Cech TR. Promiscuous RNA binding by Polycomb repressive complex 2. Nat Struct Mol Biol. 2013 Nov;20(11):1250–7.

147. Brien GL, Gambero G, O'Connell DJ, Jerman E, Turner SA, Egan CM, et al. Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat Struct Mol Biol. 2012 Dec;19(12):1273–81.

148. Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury Iii WJ, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. Nature. 2009 Oct;461(7265):762–7.

149. Tavares L, Dimitrova E, Oxley D, Webster J, Poot R, Demmers J, et al. RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. Cell. 2012 Feb 17;148(4):664–78.

150. Yu M, Mazor T, Huang H, Huang H-T, Kathrein KL, Woo AJ, et al. Direct Recruitment of Polycomb Repressive Complex 1 to Chromatin by Core Binding Transcription Factors. Mol Cell. 2012 Feb 10;45(3):330–43.

151. He J, Shen L, Wan M, Taranova O, Wu H, Zhang Y. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat Cell Biol. 2013 Apr;15(4):373–84.

152. Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LLP, et al. Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. Cell. 2014 Jun 5;157(6):1445–59.

153. Cooper S, Dienstbier M, Hassan R, Schermelleh L, Sharif J, Blackledge NP, et al. Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. Cell Rep. 2014 Jun 12;7(5):1456–70.

154. Kalb R, Latwiel S, Baymaz HI, Jansen PWTC, Müller CW, Vermeulen M, et al. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat Struct Mol Biol. 2014 Jun;21(6):569–71.

155. Hosogane M, Funayama R, Nishida Y, Nagashima T, Nakayama K. Ras-Induced Changes in H3K27me3 Occur after Those in Transcriptional Activity. PLOS Genet. 2013 Aug 29;9(8):e1003698.

156. Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, Helin K. Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. Mol Cell. 2014 Aug 7;55(3):347–60.

157. Lau MS, Schwartz MG, Kundu S, Savol AJ, Wang PI, Marr SK, et al. Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. Science. 2017 Mar 10;355(6329):1081–4.

158. Terranova R, Yokobayashi S, Stadler MB, Otte AP, van Lohuizen M, Orkin SH, et al. Polycomb Group Proteins Ezh2 and Rnf2 Direct Genomic Contraction and Imprinted Repression in Early Mouse Embryos. Dev Cell. 2008 Nov 11;15(5):668–79.

159. Isono K, Endo TA, Ku M, Yamada D, Suzuki R, Sharif J, et al. SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. Dev Cell. 2013 Sep 30;26(6):565–77.

160. Willcockson MA, Healton SE, Weiss CN, Bartholdy BA, Botbol Y, Mishra LN, et al. H1 histones control the epigenetic landscape by local chromatin compaction. Nature. 2020 Dec 9;1–6.

161. Tamburri S, Lavarone E, Fernández-Pérez D, Conway E, Zanotti M, Manganaro D, et al. Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. Mol Cell. 2020 20;77(4):840-856.e5.

162. Gao Z, Lee P, Stafford JM, von Schimmelmann M, Schaefer A, Reinberg D. An AUTS2-Polycomb complex activates gene expression in the CNS. Nature. 2014 Dec 18;516(7531):349–54.

163. Piunti A, Rossi A, Cerutti A, Albert M, Jammula S, Scelfo A, et al. Polycomb proteins control proliferation and transformation independently of cell cycle checkpoints by regulating DNA replication. Nat Commun. 2014 May;5(1):3649.

164. Bravo M, Nicolini F, Starowicz K, Barroso S, Calés C, Aguilera A, et al. Polycomb RING1A- and RING1B-dependent histone H2A monoubiquitylation at pericentromeric regions promotes S-phase progression. J Cell Sci. 2015 Oct 1;128(19):3660–71.

165. Gurvich N, Perna F, Farina A, Voza F, Menendez S, Hurwitz J, et al. L3MBTL1 polycomb protein, a candidate tumor suppressor in del(20q12) myeloid disorders, is essential for genome stability. Proc Natl Acad Sci. 2010 Dec 28;107(52):22552–7.

166. Saurin AJ, Shiels C, Williamson J, Satijn DPE, Otte AP, Sheer D, et al. The Human Polycomb Group Complex Associates with Pericentromeric Heterochromatin to Form a Novel Nuclear Domain. J Cell Biol. 1998 Aug 24;142(4):887–98.

167. Rondinelli B, Gogola E, Yücel H, Duarte AA, van de Ven M, van der Sluijs R, et al. EZH2 promotes degradation of stalled replication forks by recruiting MUS81 through histone H3 trimethylation. Nat Cell Biol. 2017 Nov;19(11):1371–8.

168. Ginjala V, Nacerddine K, Kulkarni A, Oza J, Hill SJ, Yao M, et al. BMI1 Is Recruited to DNA Breaks and Contributes to DNA Damage-Induced H2A Ubiquitination and Repair. Mol Cell Biol. 2011 May 15;31(10):1972–82.

169. Chou DM, Adamson B, Dephoure NE, Tan X, Nottke AC, Hurov KE, et al. A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. Proc Natl Acad Sci. 2010 Oct 26;107(43):18475–80.

170. Ahuja AK, Jodkowska K, Teloni F, Bizard AH, Zellweger R, Herrador R, et al. A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells. Nat Commun. 2016 Feb 15;7.

171. Catt JW, Henman M. Toxic effects of oxygen on human embryo development. Hum Reprod. 2000 Jul 1;15(suppl_2):199–206.

172. Yang D, Buchholz F, Huang Z, Goga A, Chen C-Y, Brodsky FM, et al. Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. Proc Natl Acad Sci. 2002 Jul 23;99(15):9942–7.

173. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006 Jan 31;7(1):3.

174. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, et al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 2014 Jan;42(Database issue):D756-763.

175. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016 May;34(5):525–7.

176. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15–21.

177. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

178. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47–e47.

179. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013 Apr 15;14(1):128.

180. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016 Jul 8;44(W1):W90–7.

181. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019 Apr 29;10(1):1930.

182. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015 Jul;523(7561):486–90.

183. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr;9(4):357–9.

184. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008 Sep 17;9(9):R137.

185. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016 Jul 8;44(W1):W160–5.

186. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.

187. Chuykin IA, Lianguzova MS, Pospelova TV, Pospelov VA. Activation of DNA damage response signaling in mouse embryonic stem cells. Cell Cycle. 2008 Sep 15;7(18):2922–8.

188. Hsu J, Arand J, Chaikovsky A, Mooney NA, Demeter J, Brison CM, et al. E2F4 regulates transcriptional activation in mouse embryonic stem cells independently of the RB family. Nat Commun. 2019 Jul 3;10(1):2939.

189. Stead E, White J, Faast R, Conn S, Goldstone S, Rathjen J, et al. Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. Oncogene. 2002 Nov;21(54):8320–33.

190. Morey L, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah SA, et al. Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. Cell Stem Cell. 2012 Jan 6;10(1):47–62.

191. Morey L, Aloia L, Cozzuto L, Benitah SA, Di Croce L. RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells. Cell Rep. 2013 Jan 31;3(1):60–9.

192. Cossec J-C, Theurillat I, Chica C, Búa Aguín S, Gaume X, Andrieux A, et al. SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. Cell Stem Cell. 2018 Nov;23(5):742-757.e8.

193. Du Z, Zheng H, Kawamura YK, Zhang K, Gassler J, Powell S, et al. Polycomb Group Proteins Regulate Chromatin Architecture in Mouse Oocytes and Early Embryos. Mol Cell. 2020 Feb 20;77(4):825-839.e7.

194. Lee H-S, Lee S-A, Hur S-K, Seo J-W, Kwon J. Stabilization and targeting of INO80 to replication forks by BAP1 during normal DNA synthesis. Nat Commun. 2014 Oct 6;5(1):5128.

195. Wilson R, Norris EL, Brachvogel B, Angelucci C, Zivkovic S, Gordon L, et al. Changes in the Chondrocyte and Extracellular Matrix Proteome during Post-natal Mouse Cartilage Development. Mol Cell Proteomics [Internet]. 2012 Jan 1 [cited 2020 Nov 13];11(1). Available from: https://www.mcponline.org/content/11/1/M111.014159

196. Esposito M, Kang Y. RAI2: Linking Retinoic Acid Signaling with Metastasis Suppression. Cancer Discov. 2015 May 1;5(5):466–8.

197. Werner S, Borgmann K, Pantel K, Wikman H. Abstract 2733: Novel function of the RAI2 protein in genomic integrity of breast cancer cells. Cancer Res. 2016 Jul 15;76(14 Supplement):2733–2733.

198. Reya T, Clevers H. Wnt signalling in stem cells and cancer. Nature. 2005 Apr;434(7035):843–50.

199. Jasencakova Z, Scharf AND, Ask K, Corpet A, Imhof A, Almouzni G, et al. Replication Stress Interferes with Histone Recycling and Predeposition Marking of New Histones. Mol Cell. 2010 Mar 12;37(5):736–43.

200. Petruk S, Cai J, Sussman R, Sun G, Kovermann SK, Mariani SA, et al. Delayed Accumulation of H3K27me3 on Nascent DNA Is Essential for Recruitment of Transcription Factors at Early Stages of Stem Cell Differentiation. Mol Cell. 2017 Apr 20;66(2):247-257.e5.

201. Hong S, Cho Y-W, Yu L-R, Yu H, Veenstra TD, Ge K. Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. Proc Natl Acad Sci. 2007 Nov 20;104(47):18439–44.

202. Tran N, Broun A, Ge K. Lysine Demethylase KDM6A in Differentiation, Development, and Cancer. Mol Cell Biol [Internet]. 2020 Sep 28 [cited 2021 Jan 20];40(20). Available from: https://mcb.asm.org/content/40/20/e00341-20

203. Puccetti MV, Adams CM, Kushinsky S, Eischen CM. Smarcal1 and Zranb3 Protect Replication Forks from Myc-Induced DNA Replication Stress. Cancer Res. 2019 Apr 1;79(7):1612–23.

204. Ahuja AK, Jodkowska K, Teloni F, Bizard AH, Zellweger R, Herrador R, et al. A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells. Nat Commun. 2016 Feb 15;7:10660.

205. Fujii-Yamamoto H, Kim JM, Arai K, Masai H. Cell Cycle and Developmental Regulations of Replication Factors in Mouse Embryonic Stem Cells. J Biol Chem. 2005 Jan 4;280(13):12976–87.

206. Stambrook PJ. An ageing question: Do embryonic stem cells protect their genomes? Mech Ageing Dev. 2007 Jan 1;128(1):31–5.

207. Huertas P, Aguilera A. Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. Mol Cell. 2003 Sep 1;12(3):711–21.

208. Poli J, Gerhold C-B, Tosi A, Hustedt N, Seeber A, Sack R, et al. Mec1, INO80, and the PAF1 complex cooperate to limit transcription replication conflicts through RNAPII removal during replication stress. Genes Dev. 2016 Jan 2;30(3):337–54.

209. Aloia L, Stefano BD, Croce LD. Polycomb complexes in stem cells and embryonic development. Development. 2013 Jun 15;140(12):2525–34.

210. Schuettengruber B, Bourbon H-M, Di Croce L, Cavalli G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell. 2017 Sep 21;171(1):34–57.

211. Lorente M del M, Marcos-Gutierrez C, Perez C, Schoorlemmer J, Ramirez A, Magin T, et al. Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice. Development. 2000 Dec 1;127(23):5093–100.

212. Voncken JW, Roelen BAJ, Roefs M, Vries S de, Verhoeven E, Marino S, et al. Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. Proc Natl Acad Sci. 2003 Mar 4;100(5):2468–73.

213. Prudden J, Perry JJP, Nie M, Vashisht AA, Arvai AS, Hitomi C, et al. DNA Repair and Global Sumoylation Are Regulated by Distinct Ubc9 Noncovalent Complexes. Mol Cell Biol. 2011 Jun 1;31(11):2299–310.

214. Kundu S, Ji F, Sunwoo H, Jain G, Lee JT, Sadreyev RI, et al. Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Mol Cell. 2017 Feb 2;65(3):432-446.e5.

215. Ward IM, Chen J. Histone H2AX Is Phosphorylated in an ATR-dependent Manner in Response to Replicational Stress. J Biol Chem. 2001 Dec 21;276(51):47759–62.

216. Lee HS, Lee SA, Hur SK, Seo JW, Kwon J. Stabilization and targeting of INO80 to replication forks by BAP1 during normal DNA synthesis. Nat Commun. 2014;5:5128.

217. Vassilev AP, Rasmussen HH, Christensen EI, Nielsen S, Celis JE. The levels of ubiquitinated histone H2A are highly upregulated in transformed human cells: partial colocalization of uH2A clusters and PCNA/cyclin foci in a fraction of cells in S-phase. J Cell Sci. 1995 Mar 1;108(3):1205–15.

218. Ikura T, Tashiro S, Kakino A, Shima H, Jacob N, Amunugama R, et al. DNA Damage-Dependent Acetylation and Ubiquitination of H2AX Enhances Chromatin Dynamics. Mol Cell Biol. 2007 Oct 15;27(20):7028–40.

219. Chuykin IA, Lianguzova MS, Pospelova TV, Pospelov VA. Activation of DNA damage response signaling in mouse embryonic stem cells. Cell Cycle. 2008 Sep 15;7(18):2922–8.

220. Lengsfeld BM, Berry KN, Ghosh S, Takahashi M, Francis NJ. A Polycomb complex remains bound through DNA replication in the absence of other eukaryotic proteins. Sci Rep. 2012 Sep 17;2(1):661.

221. Lo SM, Follmer NE, Lengsfeld BM, Madamba EV, Seong S, Grau DJ, et al. A Bridging Model for Persistence of a Polycomb Group Protein Complex through DNA Replication In Vitro. Mol Cell. 2012 Jun 29;46(6):784–96.

222. Francis NJ, Follmer NE, Simon MD, Aghia G, Butler JD. Polycomb Proteins Remain Bound to Chromatin and DNA during DNA Replication In Vitro. Cell. 2009 Apr 3;137(1):110–22.

223. Stewart-Morgan KR, Petryk N, Groth A. Chromatin replication and epigenetic cell memory. Nat Cell Biol. 2020 Apr;22(4):361–71.

224. Padeken J, Heun P. Nucleolus and nuclear periphery: Velcro for heterochromatin. Curr Opin Cell Biol. 2014 Jun 1;28:54–60.

225. Zee BM, Britton L-MP, Wolle D, Haberman DM, Garcia BA. Origins and Formation of Histone Methylation across the Human Cell Cycle. Mol Cell Biol. 2012 Jul 1;32(13):2503–14.

226. Petruk S, Sedkov Y, Johnston DM, Hodgson JW, Black KL, Kovermann SK, et al. TrxG and PcG Proteins but Not Methylated Histones Remain Associated with DNA through Replication. Cell. 2012 Aug 31;150(5):922–33.

227. Petruk S, Black KL, Kovermann SK, Brock HW, Mazo A. Stepwise histone modifications are mediated by multiple enzymes that rapidly associate with nascent DNA during replication. Nat Commun. 2013 Nov 26;4(1):2841.

228. Aoto T, Saitoh N, Sakamoto Y, Watanabe S, Nakao M. Polycomb Group Protein-associated Chromatin Is Reproduced in Post-mitotic G1 Phase and Is Required for S Phase Progression. J Biol Chem. 2008 Apr 7;283(27):18905–15.

229. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, et al. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystropothhy. Hum Mol Genet. 1994 Aug 1;3(8):1287–95.

230. León TE, Rapoz-D'Silva T, Bertoli C, Rahman S, Magnussen M, Philip B, et al. EZH2-Deficient T-cell Acute Lymphoblastic Leukemia Is Sensitized to CHK1 Inhibition through Enhanced Replication Stress. Cancer Discov. 2020 Jul 1;10(7):998–1017.

231. Basenko EY, Sasaki T, Ji L, Prybol CJ, Burckhardt RM, Schmitz RJ, et al. Genome-wide redistribution of H3K27me3 is linked to genotoxic stress and defective growth. Proc Natl Acad Sci. 2015 Nov 17;112(46):E6339–48.

232. Savatier P, Huang S, Szekely L, Wiman KG, Samarut J. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. Oncogene. 1994 Mar;9(3):809–18.

233. Dannenberg J-H, Rossum A van, Schuijff L, Riele H te. Ablation of the Retinoblastoma gene family deregulates G1 control causing immortalization and increased cell turnover under growth-restricting conditions. Genes Dev. 2000 Jan 12;14(23):3051–64.

234. Momčilović O, Choi S, Varum S, Bakkenist C, Schatten G, Navara C. Ionizing Radiation Induces Ataxia Telangiectasia Mutated-Dependent Checkpoint Signaling and G2 But Not G1 Cell Cycle Arrest in Pluripotent Human Embryonic Stem Cells. STEM CELLS. 2009;27(8):1822–35.

235. Berti M, Cortez D, Lopes M. The plasticity of DNA replication forks in response to clinically relevant genotoxic stress. Nat Rev Mol Cell Biol. 2020 Oct;21(10):633–51.

236. Bétous R, Couch FB, Mason AC, Eichman BF, Manosas M, Cortez D. Substrate-Selective Repair and Restart of Replication Forks by DNA Translocases. Cell Rep. 2013 Jun 27;3(6):1958–69.

237. Saugar I, Ortiz-Bazán MÁ, Tercero JA. Tolerating DNA damage during eukaryotic chromosome replication. Exp Cell Res. 2014 Nov 15;329(1):170–7.

238. Koyen AE, Madden MZ, Park D, Minten EV, Kapoor-Vazirani P, Werner E, et al. EZH2 has a non-catalytic and PRC2-independent role in stabilizing DDB2 to promote nucleotide excision repair. Oncogene. 2020 Jun;39(25):4798–813.

239. Yang J-H, Griffin PT, Vera DL, Apostolides JK, Hayano M, Meer MV, et al. Erosion of the Epigenetic Landscape and Loss of Cellular Identity as a Cause of Aging in Mammals [Internet]. Rochester, NY: Social Science Research Network; 2019 Sep [cited 2020 Nov 27]. Report No.: ID 3461780. Available from: https://papers.ssrn.com/abstract=3461780

240. Murga M, Bunting S, Montaña MF, Soria R, Mulero F, Cañamero M, et al. A mouse model of ATR-Seckel shows embryonic replicative stress and accelerated aging. Nat Genet. 2009 Aug;41(8):891–8.

241. Ruzankina Y, Pinzon-Guzman C, Asare A, Ong T, Pontano L, Cotsarelis G, et al. Deletion of the Developmentally Essential Gene ATR in Adult Mice Leads to Age-Related Phenotypes and Stem Cell Loss. Cell Stem Cell. 2007 Jun;1(1):113–26.

242. Ruzankina Y, Asare A, Brown EJ. Replicative stress, stem cells and aging. Mech Ageing Dev. 2008 Jul 1;129(7):460–6.